Posters

August 29, 2024 (Thr.) [Poster Session III]

Quantum Network Tomography via Learning Isometries on Stiefel Manifold

Ze-Tong Li[1 2 4] Xin-Lin He[1 2] Cong-Cong Zheng[1 2] Yu-Qian Dong[5] Tian Luan[5]
Xu-Tao Yu[1 2 4 *] Zai-Chen Zhang[2 3 4 †]

[1] *State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China.*
[3] *National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China.*
[2] *Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, China.*
[4] *Purple Mountain Lab, Nanjing 211111, China.*
[5] *Yangtze Delta Region Industrial Innovation Center of Quantum and Information Technology, Suzhou 215100, China.*

Abstract. Explicit mathematical reconstructions of quantum networks play a significant role in developing quantum information science. However, tremendous parameter requirements and physical constraint implementations have become computationally non-ignorable encumbrances. In this work, we propose an efficient method for quantum network tomography by learning isometries on the Stiefel manifold. Tasks of reconstructing quantum networks are tackled by solving a series of unconstrained optimization problems with significantly less parameters. The stepwise isometry estimation shows the capability for providing information of the truncated quantum network while processing the tomography. Remarkably, this method enables the compressive quantum network tomography by specifying the dimensions of isometries. As a result, our proposed method exhibits high accuracy and efficiency.

Keywords: Tomography, Quantum Network, Quantum Comb, Stiefel Manifold

1 Introduction

Quantum networks are extremely important in quantum information science [1, 2] with capabilities of performing complex tasks that require multiple input-output states at different time steps, as a non-Markovian quantum process [3]. Furthermore, the quantum network is competent to model non-Markovian quantum noise resulting from indispensable system-environment correlations, and promotes development of clean quantum computers [2].

A prevalent way to model a quantum network is the quantum comb [4]. As shown in Fig. 1, an N-time-step quantum comb constructs a completely positive (CP) map from N input states to N output states, labeled by even and odd numbers, respectively, with causality that later input systems cannot influence previous output systems. Recent quantum comb tomography (QCT) methods requires tremendous $\mathcal{O}(\prod_{i=0}^{2N-1} d_i^2)$ parameters (d_i is the dimension of i-th state) to represent an arbitrary quantum network with CP and causality (CPC) constraints, which are computationally intractable.

In this work, inspired by the isometry realization of the quantum comb [5], we propose an efficient isometry-based QCT (iQCT) optimized by the adaptive moment estimation (ADAM) on the Stiefel manifold. Our proposed method parameterizes the target quantum comb by a list of isometries with significantly fewer parameters, which is $\mathcal{O}(\prod_{k=0}^{N-1} d_{2k}^2 d_{2N-1})$, and inherent satisfaction of CPC constraints. Then, the original QCT task is transformed into solving N unconstrained optimization problems on the Stiefel manifold. As a result, our proposed method exhibits high accuracy and efficiency. Furthermore, the stepwise optimization determines one isometry at each time step. Hence, our method is capable of providing information about the truncated quantum comb while processing the tomography. Remarkably, this method enables the compressive QCT by specifying the dimensions of the isometries, especially in cases where the experimenter has prior information that the time correlations are limited and can be characterized by low-dimensional isometries, or for characterizing non-Markovian quantum noise with mild system-environment correlations. A technical version can be found in arXiv:2404.06988.

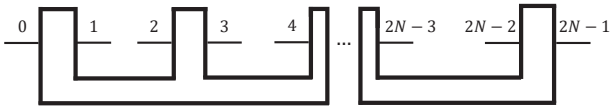


Figure 1: A quantum comb with N time steps. Wires labeled by $2k$ and $2k+1$ represent the input and output systems, respectively, at time step k, $k = 0, 1, \ldots, N-1$. Causality of the quantum comb indicates that the information flows along with the time step, which implies that the input system at time step l cannot influence the output system m if $l > m$.

2 Framework of Isometry based QCT

A quantum comb $\mathcal{C}^{(N)}$ with N time steps as shown in Fig. 1, that represents an N-time-step quantum network, maps N input systems $\rho^{(2k)} \in \mathrm{Lin}(\mathcal{H}_{2k})$ to N output systems $\rho^{(2k+1)} \in \mathrm{Lin}(\mathcal{H}_{2k+1})$, $k = 0, 1, \ldots, N-1$, where $\mathcal{H}_i$ is a d_i-dimensional Hilbert space and $\mathrm{Lin}(\mathcal{H}_i)$ is the space of linear operator on $\mathcal{H}_i$. Let $\mathcal{H}_{\mathrm{in}}^{(N)} := \bigotimes_{k=0}^{N-1} \mathcal{H}_{2k}$

*yuxutao@seu.edu.cn
†zczhang@seu.edu.cn

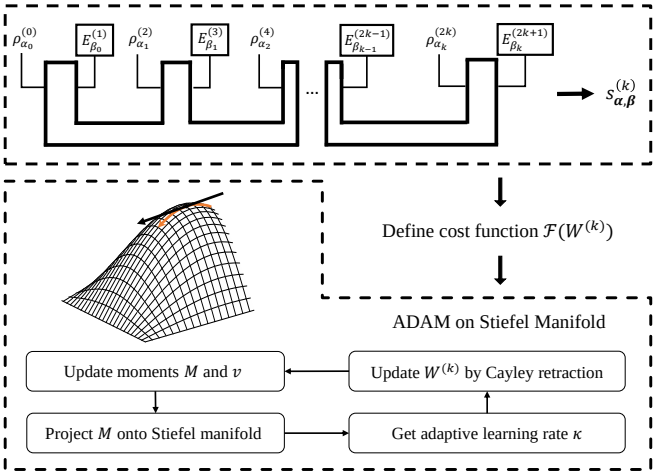


Figure 2: Workflow for estimating $V^{(k)}$ in the target N-time-step quantum network. The experimenter tests the truncated quantum network with time steps $t = 0, \ldots, k$, $k \leq N-1$, from the target quantum network, as shown at the top of the figure. Then, the cost function is defined by measurement results $s^{(k)}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ w.r.t. $W^{(k)}$. Finally, we use the ADAM on the Stiefel manifold to determine $V^{(k)} = \arg\min_{W^{(k)} \in \mathrm{St}^{(k)}} \mathcal{F}(W^{(k)})$.

and $\mathcal{H}^{(N)}_{\mathrm{out}} := \bigotimes_{k=0}^{N-1} \mathcal{H}_{2k+1}$.

An arbitrary quantum comb $\mathcal{C}^{(N)}$ can be implemented by isometries $V^{(k)} : \mathcal{H}_{2k} \otimes \mathcal{H}_{A_k} \mapsto \mathcal{H}_{2k+1} \otimes \mathcal{H}_{A_{k+1}}$, $k = 0, 1, \ldots, N-1$, where $\mathcal{H}_{A_k} = \bigotimes_{j=0}^{k-1} \mathcal{H}_{2j}$, $k > 0$, and $\mathcal{H}_{A_0} = \mathbb{C}$ are Hilbert ancillary spaces. Output states are computed by

$$\mathcal{C}^{(N)}(\rho) = \mathrm{Tr}_{A_N}[V^{(N-1)} \ldots V^{(0)} \rho V^{(0)\dagger} \ldots V^{(N-1)\dagger}], \quad (1)$$

where $\rho \in \mathrm{Lin}(\mathcal{H}_{\mathrm{in}})$, $\mathcal{C}^{(N)}(\rho) \in \mathrm{Lin}(\mathcal{H}_{\mathrm{out}})$.

This indicates that $V^{(k)}$ can be adequately represented using $\prod_{t=0}^{k} d^2_{2t} d_{2k+1}$ complex parameters. Hence, $\mathcal{O}(\prod_{k=0}^{N-1} d^2_{2k} d_{2N-1})$ complex parameters are sufficient to represent entire quantum comb $\mathcal{C}^{(N)}$, which is significantly more efficient than the Choi state based QCT methods.

To perform tomography on the quantum comb, the experimenter prepares known tomographically complete state sets $\Gamma^{(2k)} := \{\rho^{(2k)}_i\}_{i=0}^{d^2_{2k}-1}$ and measurement sets $\Xi^{(2k+1)} := \{E^{(2k+1)}_j\}_{j=0}^{d^2_{2k+1}-1}$ for input systems $\rho^{(2k)}$ and output system $\rho^{(2k+1)}$ that span $\mathrm{Lin}(\mathcal{H}_{2k})$ and $\mathrm{Lin}(\mathcal{H}_{2k+1})$, respectively, for $k = 0, 1, \ldots, N-1$.

The causality indicates that the input systems at later time steps cannot influence previous output systems, which enables the stepwise optimization in the iQCT. At step k, only the isometry $V^{(k)}$ is reconstructed with the known $V^{(t)}$, $t < k$. The workflow for estimating $V^{(k)}$ is summarized in Fig. 2. The experimenter conducts experiments that combine the input states $\{\rho^{(t)}_{\alpha_t}\}_{t=0}^{k}$ and measurements on output states $\{E^{(2t+1)}_{\beta_t}\}_{t=0}^{k}$ and records the results $s^{(k)}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, where $\boldsymbol{\alpha} := [\alpha_0, \ldots, \alpha_k]$, $\boldsymbol{\beta} := [\beta_0, \ldots, \beta_k]$. The criteria for selecting $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are that $\{\eta^{(k-1)}_{\boldsymbol{\alpha},\boldsymbol{\beta}}\}$ consists of at least $d^2_{2k} d^2_{A_k}$ linear independent matrices and that β_k spans $\{0, \ldots, d^2_{2k+1} - 1\}$, where

$$\eta^{(t)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \mathrm{Tr}_{2t+1}[\rho^{(2t+2)}_{\alpha_{t+1}} E^{(2t+1)}_{\beta_t} V^{(t)} \eta^{(t-1)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} V^{(t)\dagger}], t \geq 0, \quad (2)$$

and $\eta^{(-1)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \rho^{(0)}_{\alpha_0}$. From (1), the recovered probability is

$$p_{\boldsymbol{\alpha},\boldsymbol{\beta}}(W^{(k)}) = \mathrm{Tr}[E^{(2k+1)}_{\beta_k} W^{(k)} \eta^{(k-1)}_{\boldsymbol{\alpha},\boldsymbol{\beta}} W^{(k)\dagger}]. \quad (3)$$

Then, the isometry $V^{(k)}$ is reconstructed by optimizing the cost function $\mathcal{F}$ on the Stiefel manifold without constraints

$$\min_{W^{(k)} \in \mathrm{St}^{(k)}} \mathcal{F}(W^{(k)}) = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} |\tilde{p}_{\boldsymbol{\alpha},\boldsymbol{\beta}} - p_{\boldsymbol{\alpha},\boldsymbol{\beta}}(W^{(k)})|^2, \quad (4)$$

where $\tilde{p}_{\boldsymbol{\alpha},\boldsymbol{\beta}} = s^{(k)}_{\boldsymbol{\alpha},\boldsymbol{\beta}}/n_s$ represents the measurement probability, n_s is the total number of samples, $\mathrm{St}^{(k)} := \{X \in \mathbb{C}^{(k)} : X^\dagger X = I\}$ represents the Stiefel manifold on which $V^{(k)}$ lies, and $\mathbb{C}^{(k)} := \mathbb{C}^{d_{2k+1} d_{A_{k+1}} \times d_{2k} d_{A_k}}$. This optimization problem is solved by ADAM on the Stiefel manifold.

Note that the stepwise optimization determines one isometry at each time step. Hence, the iQCT has the capability of providing isometries of $C^{(k)}$, $k \leq N$, while performing tomography to $C^{(N)}$. The causality indicates that the isometries of $C^{(k)}$, $k \leq N$, completely characterize the truncated quantum network from time step 0 to $k-1$. This property facilitates experimenters to analyze the currently determined information of the truncated quantum network when the iQCT is determining later isometries. Furthermore, this method enables compressive tomography of quantum networks by reducing the dimensions of ancillary spaces. This compressive method is both efficient and effective when the experimenter has prior information about the required ancillary dimensions.

3 Experimental Results

We first simulate iQCT to reconstruct a series of random 2-time-step quantum networks defined by isometries. We use the reconstruction fidelity $F(\Upsilon, \Upsilon')$ between the Choi states of reconstructed and ideal quantum networks Υ and Υ' to represent the accuracy of QCT methods. The fidelity becomes 1 when two quantum networks construct same map between input and output states, and 0 when they are totally different such that their Choi states are orthogonal. We adopt the absolute running time Δ_T (s) running in the same computer to fairly showcase the efficiency.

In Fig. 3, we show the fidelity and absolute running time w.r.t. dimensions of input and output states. We set the dimensions of input and output states at the same time step are identical. Labels 'n-m' henceforth represent that states at time step 0 and 1 consist of n and m qubits, respectively. From the results, the iQCT efficiently reconstructs quantum networks with fidelity $F > 99\%$. The increasing fidelities to dimensions of isometries as shown in Fig. 3(a) result from that the termination condition

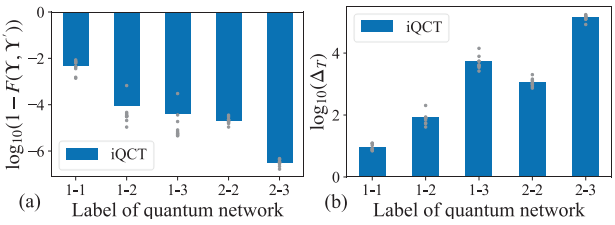


Figure 3: Results of QCT for 2-time-step quantum networks. For each label, we conduct the QCT to 10 random quantum networks. Bars represent average values, while gray points are values of individual results of the random quantum networks. (a) Fidelity gaps to 1. (b) Absolute running times.

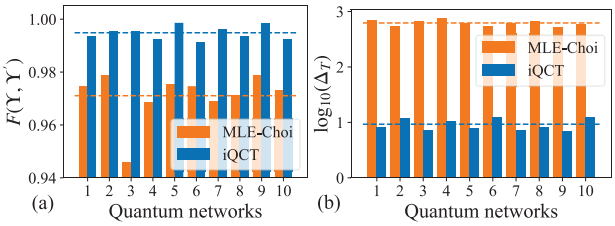


Figure 4: Results of QCT for 10 random '1-1' quantum networks. Dashed lines represent average values, while bars are values of individual results of networks. (a) Fidelities. (b) Absolute running times.

act on the gradients of isometries instead of differences of cost function values. The low variance of fidelities and absolute running times in the same settings indicates the stability of the iQCT in ideal circumstances.

We further compare the iQCT to the recent state-of-the-art QCT method which construct the Choi state based on the maximum likelihood estimation (MLE) with physical constraints implemented by Dikstra projection [6], labeled by MLE-Choi. In Fig. 4, we show the fidelity and absolute running time of the two schemes in estimating 10 random '1-1' quantum networks. Our proposed iQCT achieves 82.3% and 98.5% improvements to the average fidelity gap to 1 and absolute running time, respectively, in this situation.

To showcase the capability of compressive tomography, we apply the iQCT with specified ancillary dimensions to '2-2' random quantum networks, where full ancillary dimensions are d_{A_0}=4 and d_{A_1}=16. In Fig.5, we exhibit fidelities w.r.t. ancillary dimensions. The compressive iQCT achieves high fidelities when ancilary dimensions are sufficient. This implies that the compressive iQCT is both efficient and accurate when we know the requirement of ancillary dimensions.

Furthermore, we apply the iQCT to a single-qubit 3-time-step real quantum computer. We utilize the relative cost $\mathcal{L}(\mathcal{C}_{\text{cmp}}, \mathcal{C}_{\text{full}})$ to measure distances between reconstructed compressive quantum network $\mathcal{C}_{\text{cmp}}$ and full-ancillary-dimension quantum network $\mathcal{C}_{\text{full}}$. $\mathcal{L}$=0 when $\mathcal{C}_{\text{cmp}}$ is equivalent to $\mathcal{C}_{\text{full}}$ with unitary non-measurement operations. From Fig. 6, the iQCT achieves significant similarity between $\mathcal{C}_{\text{cmp}}$ and $\mathcal{C}_{\text{full}}$. This indicates

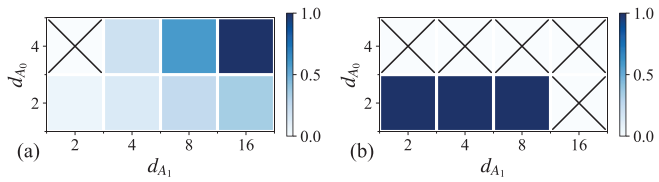


Figure 5: Average fidelities with specified ancillary dimensions. Blocks marked by '×' means that we do not conduct simulations in corresponding situations. Random '2-2' quantum networks are generated with (a) full and (b) specified ancillary dimensions. The specified dimensions in generating and reconstructing quantum networks are equal.

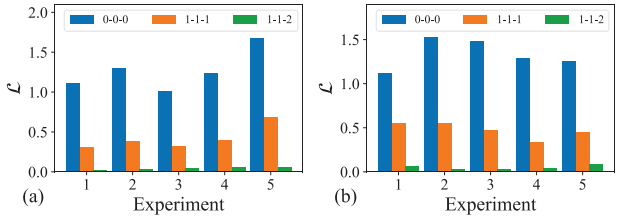


Figure 6: Results on real quantum chips. Labels in legends represent $\log_2 d_{A_0}$-$\log_2 d_{A_1}$-$\log_2 d_{A_2}$. Instruments are performed with gap (a) 10ns and (b) 20ns.

the potentiality of iQCT to efficiently reconstruct non-Markovian quantum noise with less computational resources but high fidelities.

acknowledgments This work is supported by the Fundamental Research Funds for the Central Universities 2242022k60001, Jiangsu Key R&D Program Project BE2023011-2, National Natural Science Foundation of China No.61960206005, and National Natural Science Foundation of China No.61871111.

References

[1] S.H. Wei, *et al.* Towards real-world quantum networks: A review. *Laser & Photonics Reviews*, 16(3):2100219, 2022.

[2] G.A.L. White, *et al.* Unifying non-Markovian characterisation with an efficient and self-consistent framework. *arXiv:2312.08454*, 2023.

[3] Gus Gutoski and John Watrous. Toward a general theory of quantum games. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, pages 565–574, 2007.

[4] Giulio Chiribella, *et al.* Theoretical framework for quantum networks. *Phys. Rev. A*, 80(2):022339, 2009.

[5] Alessandro Bisio, *et al.* Minimal computational-space implementation of multiround quantum protocols. *Physical Review A*, 83(2):022325, 2011.

[6] G.A.L. White, *et al.* Non-Markovian Quantum Process Tomography. *PRX Quantum*, 3(2):020344, 2022.

Witnessing Non-Gaussian Entanglement in cQED Devices With Conditional Displacement Gates

Lin Htoo Zaw[1*]

[1] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*

Abstract. In weakly-dispersive cQED devices, conditional displacement (CD) gates are used to probe the characteristic function of cavity states, while Wigner function measurements are difficult and quadrature measurements are unavailable. As such, past demonstrations of entanglement in such architectures have resorted to state tomography. I recently proposed a non-Gaussian entanglement witness that uses only CD gates and qubit readouts [1]. The witness arises from a result from harmonic analysis and a surprising connection between two negativities: that of the reduced Wigner function, and that of the partial transpose. It requires as few as four points of the characteristic function, and simultaneously lower bounds the Wigner negativity volume and a measure conjectured to be the partial transpose negativity.

Keywords: circuit and cavity quantum electrodynamics, characteristic function, Wigner negativity, entanglement witness

Background and Motivation. In both circuit and cavity quantum electrodynamics (cQED), operations on high quality-factor cavities are mediated by qubits dispersively coupled to them. If the coupling between the qubit and the cavity is weakly dispersive, it is possible to perform conditional displacement (CD) gates, of the form $U_{\text{CD}} := D(\vec{\xi}/2)\,|e\rangle\langle g| + D(-\vec{\xi}/2)\,|g\rangle\langle e|$ where $D(\vec{\alpha})$ is the displacement operator on the cavities, with high fidelity and low gate times [2, 3, 4]. The pointwise characteristic function of cavity states can be probed with CD gates and qubit measurements, while displaced parity gates have gate times orders of magnitude longer, and quadrature measurements are unavailable.

As most continuous variable entanglement witnesses are based on quadrature statistics, past demonstrations of entanglement in such architectures have resorted to violating Bell inequalities with Wigner function measurements or by computing the entanglement fidelity from the tomographically-reconstructed state. The former is difficult in weakly-coupled systems due to the necessity of parity gates, while the latter is an expensive operation.

Contributions. The first contribution of this work is a proof that a Wigner negativity witness based on Bochner's theorem provides a lower bound to the Wigner negativity volume. To the best of my knowledge, existing Wigner negativity witnesses have only been shown to lower bound the trace distance of Wigner negativity, which, unlike the Wigner logarithmic volume, has not been shown to be a non-Gaussian monotone.

The second contribution of this work is a method to directly detect non-Gaussian entanglement between cavities using only CD gates and qubit readouts. In most cases, as few as four settings of the CD gates are needed for non-Gaussian entanglement to be certified.

The Protocol. Let $D_A(\vec{\xi})$ ($D_B(\vec{\xi})$) be the displacement operator for partition A (B). The CD entanglement witness can be implemented with the following steps:

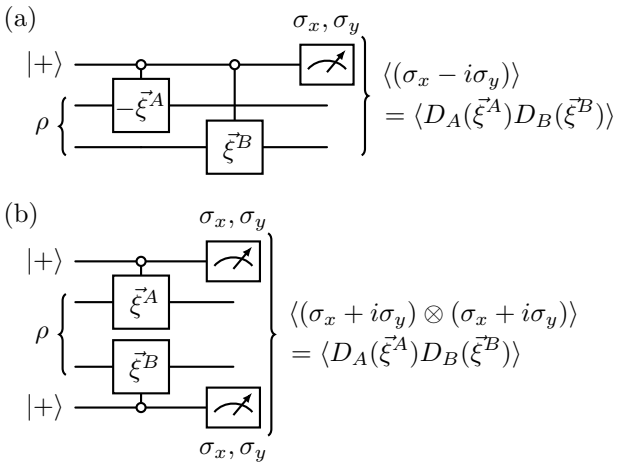


Figure 1: Measurement of $\langle D_A(\vec{\xi}_j^A - \vec{\xi}_k^A) D_B(\vec{\xi}_j^B - \vec{\xi}_k^B)\rangle$ with (a) one or (b) two auxiliary qubits.

(1) Choose N phase-space pairs $\Xi = \{(\vec{\xi}_k^A, \vec{\xi}_k^B)\}_{k=1}^N$ with $\vec{\xi}_k^A = \vec{\xi}_k^B$.

(2) Using CD gates and qubit measurements (see Fig. 1), measure $\langle D_A(\vec{\xi}_j^A - \vec{\xi}_k^A) D_B(\vec{\xi}_j^B - \vec{\xi}_k^B)\rangle$ for all $j < k$. Denote the experimental error bars as $\delta_{j,k}$.

(3) Construct $\mathbf{C}_2$ with the matrix elements $[\mathbf{C}_2]_{j,k} = \langle D_A(\vec{\xi}_j^A - \vec{\xi}_k^A) D_B(\vec{\xi}_j^B - \vec{\xi}_k^B)\rangle / N$ for $j < k$ from the previous step, while the other elements are given by $[\mathbf{C}_2]_{j,j} = 1$ and $[\mathbf{C}_2]_{j,k} = [\mathbf{C}_2]_{k,j}^*$ for $j > k$.

(4) Calculate λ_-, the minimum eigenvalue of $\mathbf{C}_2$. Define $\mathcal{E}_C := \max(0, -\lambda_-)$ and $\delta := \max_j \sum_{k \neq j} \delta_{j,k}/N$. If $\mathcal{E}_C > \delta$, then the system is entangled.

Lower Bounds of Measures. The expectation value of the witness $\mathcal{E}_C \pm \delta$ is a lower bound to the Wigner negativity volume $\mathcal{N}_V$ and the positive-partial-transpose trace distance $\mathcal{E}_{\text{PPT}}$, where

$$\begin{aligned} \mathcal{N}_V &= \frac{1}{2} \int \mathrm{d}\vec{\alpha}\, [|W_\rho(\vec{\alpha})| - W_\rho(\vec{\alpha})], \\ \mathcal{E}_{\text{PPT}} &= \min_{\sigma \in \text{PPT}} \operatorname{tr} \left|\sigma^{T_B} - \rho^{T_B}\right|. \end{aligned} \tag{1}$$

*htoo@zaw.li

Here W_ρ is the Wigner function of ρ, T_B is the partial transpose over partition B, and PPT is the set of positive-partial-transpose states. $\mathcal{E}_{\mathrm{PPT}}$ is conjectured to be equivalent to the partial transpose negativity $\mathrm{tr}\,|\rho^{T_B}|-1$ [5]. The Wigner negativity volume is a non-Gaussian monotone, while the partial transpose negativity is an entanglement monotone.

States detected by the witness. While the detected states are not yet fully characterised, the witness can detect common families of non-Gaussian entangled states—entangled Fock states, photon-subtracted two-mode squeezed vacua, and entangled cats. Most of them can be detected with just four measurement settings.

Timeliness. The CD gate on a single cavity was first proposed and demonstrated four years go in cavity QED [2] and three years ago in circuit QED [3], while a CD gate on two cavities coupled to a single qubit, exactly the type of measurement needed for the CD witness, was demonstrated just last year [4]. As this architecture gains in popularity—especially since the control scheme for CD gates can be applied to other weakly-dispersive cQED devices—this CD witness is a very timely contribution for certifying non-Gaussian entanglement on such devices with low overheads and a simple implementation.

References

[1] Lin Htoo Zaw. Certifiable lower bounds of wigner negativity volume and non-gaussian entanglement with conditional displacement gates, 2024.

[2] C. Flühmann and J. P. Home. Direct Characteristic-Function Tomography of Quantum States of the Trapped-Ion Motional Oscillator. *Phys. Rev. Lett.*, 125:043602, Jul 2020. doi: 10.1103/PhysRevLett.125.043602.

[3] Alec Eickbusch, Volodymyr Sivak, Andy Z. Ding, Salvatore S. Elder, Shantanu R. Jha, Jayameenakshi Venkatraman, Baptiste Royer, S. M. Girvin, Robert J. Schoelkopf, and Michel H. Devoret. Fast universal control of an oscillator with weak dispersive coupling to a qubit. *Nature Physics*, 18 (12):1464–1469, Dec 2022. ISSN 1745-2481. doi: 10.1038/s41567-022-01776-9.

[4] Asaf A. Diringer, Eliya Blumenthal, Avishay Grinberg, Liang Jiang, and Shay Hacohen-Gourgy. Conditional-NOT Displacement: Fast Multioscillator Control with a Single Qubit. *Phys. Rev. X*, 14:011055, Mar 2024. doi: 10.1103/PhysRevX.14.011055.

[5] Ray Ganardi, Marek Miller, Tomasz Paterek, and Marek Żukowski. Hierarchy of correlation quantifiers comparable to negativity. *Quantum*, 6:654, February 2022. ISSN 2521-327X. doi: 10.22331/q-2022-02-16-654.

Group twirling and noise tailoring for multi-qubit-controlled phase gates

Guoding Liu[1] Ziyi Xie[1] Zitai Xu[1] Xiongfeng Ma[1] *

[1] *Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084 China*

Abstract. Group twirling is crucial in quantum information processing, particularly in randomized benchmarking and randomized compiling. While Pauli twirling has been utilized to transform arbitrary noise channels into Pauli channels for Clifford gates, the lack of practical twirling groups for multi-qubit non-Clifford gates remains a challenge. To address this gap, we study the issue of finding twirling groups for generic quantum gates. Interestingly, for multi-qubit-controlled phase gates, which are essential in quantum algorithms and directly implementable in practice, we determine optimal twirling groups within a large gate set. We propose associated benchmarking procedures for such gates and numerically identify their practicality.

Keywords: group twirling, noise tailoring, quantum benchmarking, multi-qubit-controlled phase gates

1 Introduction

Group-twirling-based noise tailoring is an essential step to deal with noise in quantum information processing. Group twirling symmetrizes the noise channel [1, 2], allowing accurate and efficient extraction of noise channel parameters. This principle underlies the randomized benchmarking (RB) methodology [3–6], which stands out as a major quantum benchmarking technique due to its low sample complexity and resilience against state preparation and measurement (SPAM) errors. Moreover, group twirling is essential in randomized compiling [7], which turns generic noise into a Pauli channel, reducing the worst-case error of quantum gates and facilitating Pauli channel learning protocols [8–11].

While group twirling facilitates many tasks, the additional overhead to implement twirling gates should be considered. It is favoured to use a more compact and facilely implementable twirling gate set without compromising task performance. Focusing on applying group twirling in tailoring a quantum gate in randomized benchmarking or random compiling, currently, the landscape is dominated mainly by efficient noise tailoring protocols for Clifford gates, primarily achieved through Pauli group twirling [7, 12, 13]. In contrast, practical twirling groups for multi-qubit non-Clifford gates are lacking. This raises the pertinent question of finding suitable twirling groups for tailoring generic quantum gates and investigating their optimality. This research is essential not only for experimental advancements in non-Clifford gate benchmarking and applications but also for theoretical insights that underscore the need for tailoring quantum gates.

In this work, we investigate noise tailoring strategies for generic quantum gates. Central to noise tailoring is the selection of appropriate twirling gates for the twirled gate, namely the gate undergoing twirling. We study this question within a frequently used circuit structure – the twirled and twirling gates are intertwined. Within this structure, we summarize the constraints between the twirled and twirling gates, and we find that any quantum gate tailoring demands a twirling gate set comparable in magnitude to the Pauli group, implying the optimality of existing noise tailoring schemes designed for Clifford gates. In addition to the well-studied Clifford gates, for multi-qubit-controlled phase gates in the form of

$$C^nZ_\theta = \begin{pmatrix} \mathbb{I}_{2^n-1} & \mathbf{0} \\ \mathbf{0} & e^{i\theta} \end{pmatrix}, \tag{1}$$

where n is a positive integer and θ is a real number, we found optimal twirling groups within the realm of classically replaceable unitary operations [14]. These gates are the key components in quantum algorithms [15–17] and directly implementable in quantum processors [18–21]. The optimal twirling groups found in this work are subgroups of the CNOT dihedral group used to benchmark C^nZ_θ in previous works [22]. Unlike the relatively straightforward tailoring process for Clifford gates, the optimal twirling gate set for C^nZ_θ grows exponentially with the qubit count.

We further conducted numerical simulations of benchmarking C^nZ_θ gates using various noisy twirling groups. Our findings indicate that the optimal twirling group delivers superior performance compared to the CNOT dihedral group and the Pauli group in a small-scale quantum system. We believe these results will contribute to the broader use of native non-Clifford gates in quantum computing and facilitate the practical implementation of a wide range of quantum algorithms.

2 Twirling groups in RB

In this extended abstract, we mainly present the results of RB and leave the part of randomized compiling to the technical version of our work in the appendix.

We consider the task of RB that estimates the fidelity of an individual target gate, U, robust to state preparation and measurement errors. The noisy quantum gate is expressed as $\widetilde{\mathcal{U}} = \mathcal{U}\Lambda$ where Λ is the noise channel, $\mathcal{U}$ denotes the Pauli-Liouville representation of U, and $\widetilde{\cdot}$ represents the noisy version of quantum gates or observables. The fidelity of U is namely the process fidelity of Λ [23, 24]. Based on the result of character RB [25], if

*xma@tsinghua.edu.cn

one can obtain the powers of the G-twirled noise channel, Λ_G^m, where $m \in \mathbb{Z}_+$ and $\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger$, and Λ_G is diagonal in the Pauli-Liouville representation up to a unitary transformation independent of Λ, then one can obtain the fidelity of Λ accurately with single-exponential fitting. Here, G is a prefixed twirling group. If Λ_G is not diagonal, we need to overcome the notorious matrix exponential fitting problem for getting fidelity [26], which would cause inaccurate estimation.

In this work, to obtain Λ_G^m, we utilize the circuit in Fig. 1. One independently and randomly samples m twirling gates G_i from group G, and implements them interleaved with target gate U. The circuit ends with the inverse gate $G_{\text{inv}} = (\prod_{i=1}^m UG_i)^\dagger$. We prove that, as long as $U\mathsf{G}U^\dagger = \mathsf{G}$, this circuit allows us to obtain $\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m$ along with the fidelity of $\mathcal{U}' = (\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}$. We further prove that, in this case, the fidelity of $\mathcal{U}'$ is a lower bound of the fidelity of $\mathcal{U}$. Furthermore, the two are equal when U is a multi-qubit-controlled phase gate.

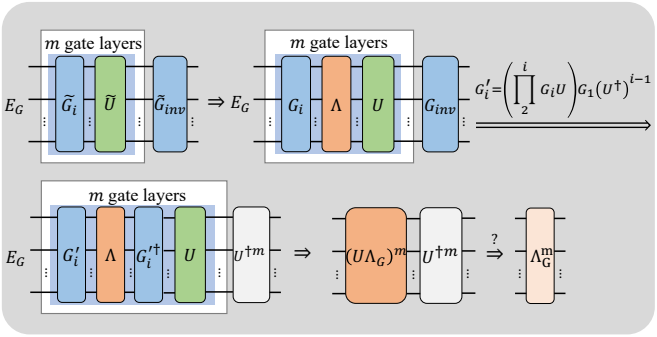


Figure 1: Random twirling gates $G_1, G_2, \cdots, G_m$ interleaved with U in RB. The inverse gate $G_{\text{inv}} = (\prod_{i=1}^m UG_i)^{-1}$. $\tilde{\cdot}$ represents the noisy version of a quantum gate. Λ is the noise channel of U and Λ_G is the G-twirled noise channel. This circuit actually measures the noise from both U and G but the noise effect of G can be removed with the technique of interleaved RB [27].

Based on the discussion above, we summarize the requirements of the twirling group G to tailor U in RB:

$$\begin{aligned}&\text{for any quantum channel } \Lambda,\\ &\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \text{ is diagonal up to a}\\ &\text{unitary transformation independent of } \Lambda,\end{aligned} \tag{2}$$

and,

$$U\mathsf{G}U^\dagger = \mathsf{G}. \tag{3}$$

The first condition is required to symmetrize the noise channel, and the second condition means that the action of U does not destroy this symmetry. A good solution of G should be small and easily implementable.

In this work, we study the diagonalizability of Λ_G and prove two associated lemmas, restricting the choice of G and helping us to find the optimal twirling group for multi-qubit-controlled phase gates.

Lemma 1 *If a finite n-qubit unitary subgroup, G, satisfies Eq. (2), then the Pauli-Liouville representation of G is multiplicity-free. As a corollary, the cardinality of the twirling group $|\mathsf{G}| \geq 4^n$.*

Lemma 2 *If a finite n-qubit CRU subgroup, G, satisfies Eq. (2), then G can interchange any two computational basis states. That is, for any two computational basis states $|\mathbf{i}\rangle$ and $|\mathbf{j}\rangle$ where $\mathbf{i},\mathbf{j} \in \{0,1\}^n$, there exists a gate $G \in \mathsf{G}$ such that $|\mathbf{j}\rangle = G|\mathbf{i}\rangle$.*

Lemma 1 implies that any twirling gate set must be comparable in magnitude to the Pauli group, showing the superiority of Clifford gates in noise tailoring, thanks to Clifford gates normalizing the Pauli group.

The set of classically replaceable unitary operations comprises all gates that can be moved after computational basis measurements and become classical post-processing. This gate set is large and will be universal after adding Hadamard gates. Lemma 2 puts a very strong restriction on the twirling group in this set and indicates that the twirling group should contain a group like X, the group generated by Pauli X gates on all qubits. With Lemma 2, we further prove the theorem below, revealing the optimal twirling groups for multi-qubit-controlled phase gates.

Theorem 3 *The optimal twirling group G in CRU for the multi-qubit-controlled phase gate, $U = C^nZ_m$, with $n \geq 1, m \geq 2$, is given by*

$$\mathsf{G} = \{\Pi(\prod_{i=1}^t(\Pi_i^\dagger U\Pi_i U^\dagger)^{l_i})|\Pi \in \mathsf{X}, t \in \mathbb{N}, \forall i, l_i \in \pm 1, \Pi_i \in \mathsf{X}\}. \tag{4}$$

Note that any CRU subgroup is decomposable into the semi-product of a permutation group and a diagonal group, $\mathsf{G}_{\text{CRU}} = \Pi \ltimes \mathsf{W} = \{\Pi W|\Pi \in \Pi, W \in \mathsf{W}\}$. The optimal group can be written as $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ where $\mathsf{W}_X = \{\Pi^\dagger U\Pi U^\dagger, \Pi \in \mathsf{X}\}$ only comprises diagonal gates. The optimality here means that any G_{CRU} satisfying Eqs (2) and (3) implies $\mathsf{W} \supseteq \mathsf{W}_X$ and $|\Pi| \geq |\mathsf{X}|$. The optimal group is clearly the smallest option. In practice, one normally determines the permutation part and the diagonal part separately to sample a gate in CRU. The gate is realized by sequentially implementing two parts. In this sense, the optimal group is also the most easily implementable option.

We list specific forms of twirling groups for C^nZ and CZ_m in Table 1. Our results are better than previous ones [22] in terms of group size and computational complexity. Nonetheless, even for the optimal solution, the twirling group would be large and highly non-local with the increased controlled qubit number. The sample complexity and the computational complexity would be unacceptable for large control qubit numbers. The results indicate the fundamental difficulty of benchmarking multi-qubit non-Clifford gates.

Table 1: Scaling of the group size and computational complexity of the twirling group for tailoring C^nZ and CZ_m in our work and [22]; CXD represents the CNOT dihedral group. The complexity is expressed with the qubit number N, controlled qubit number n, and phase angle index m ($\theta = \frac{2\pi}{m}$ in Eq. (1)).

C^nZ			
	Group	Size	Complexity
This work	$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle$	$O(N^n)$	$O(N^n)$
CXD [22]	$\langle CX, Z_{2^{n+1}}, X\rangle$	$O(N^{n+1})$	$O(N^{3n+1})$
CZ_m			
	Group	Size	Complexity
This work	$\langle CZ_m, Z_m, X\rangle$	$O(N^2 \log m)$	$O(N^2 \log m)$
CXD	$\langle CX, Z_{2m}, X\rangle$	$O(N^{\log m})$	$O(N^{3\log m+1})$

3 Simulation

Here, we show part of the simulation results of benchmarking CS and CCZ gates with the full results in the technical version of our work. We first propose a benchmarking procedure using the optimal twirling group and a random Pauli SPAM setting. And we further enhance this procedure by using a group, which add the phase gate $S = |0\rangle\langle 0| + i\,|1\rangle\langle 1|$ to the optimal twirling group, and a SPAM setting that we name as ZX-SPAM.

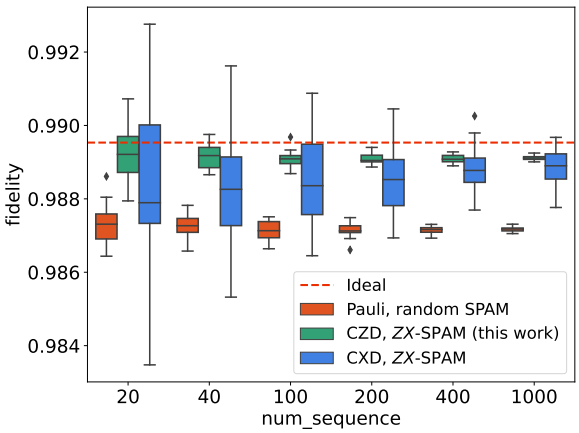
In the random Pauli SPAM setting, for a randomly sampled Pauli operator, P, one inputs the eigenstate of P, implements the circuit in Fig. 1 and measures P to estimate the noise channel parameter $\mathrm{tr}(P\Lambda(P))/2^N$, where Λ is the noise channel and N is the qubit number. The process is repeated for a constant number of Pauli observables and the fidelity is estimated by averaging $\mathrm{tr}(P\Lambda(P))/2^N$ from different P. In the ZX-SPAM setting, one prepares $|0\rangle^{\otimes N}$ and measures in $Z^{\otimes N}$ to extract $\mathrm{tr}(P_Z\Lambda(P_Z))/2^N$ for all $P_Z \in \{\mathbb{I}, Z\}^{\otimes N}$. Similarly, by preparing $|+\rangle^{\otimes N}$ and measuring in $X^{\otimes N}$, one extracts $\mathrm{tr}(P_X\Lambda(P_X))/2^N$ for all $P_X \in \{\mathbb{I}, X\}^{\otimes N}$. The ZX-SPAM setting is sufficient to obtain all different diagonal terms of the twirled noise channel as long as the twirling group includes the CZ dihedral group $\langle CZ, Z, S\rangle$.

Below, in Fig. 2, we compare our scheme, which uses the CZ dihedral group $\mathsf{G}^Z = \langle CZ, X, S\rangle$ and the ZX-SPAM, the CNOT dihedral group with the ZX-SPAM, and the Pauli group with a random SPAM in benchmarking CS and CCZ gates. Note that the Pauli group does not satisfy Eqs. (2) and (3) to benchmark multi-qubit controlled phase gates, and in this case, the inverse gate in Fig. 1 is outside the Pauli group.

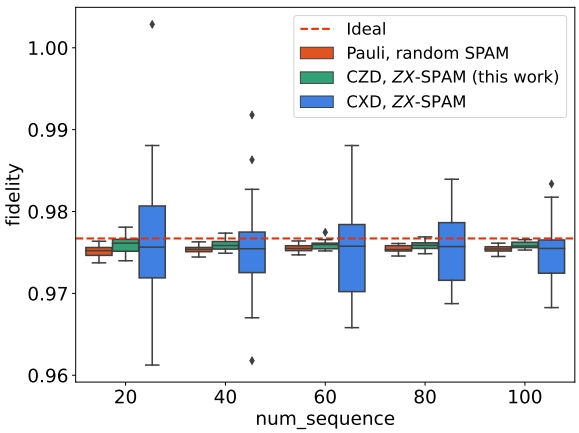
The twirling gates are simulated with gate-dependent noise. We simulate both the fidelity of the composite noise channel and the fidelity of the twirling group and get the fidelity of the target gate by interleaved RB technique. We only show the target gate fidelity in Fig. 2. The results show our scheme performs better than other two methods in terms of both the precision and accuracy for benchmarking CS and CCZ gates, providing a practical scheme for tailoring multi-qubit controlled phase gates in small-scale systems.

4 Outlook

Identifying optimal twirling groups for gate tailoring is intriguing in both theoretical and experimental contexts. Our findings provide an initial glimpse into this area. Future research can explore optimal twirling groups for various quantum gates and investigate the feasibility of tailoring gates solely with local twirling gates. Additionally, by seeking benchmarking methodologies that do not rely on a group structure, we may find smaller twirling gate sets. Beyond the scope of noise tailoring, the properties of twirling groups explored in this work may prove valuable in other quantum information processing tasks, such as Pauli error rate learning and shadow tomography.



(a) CS



(b) CCZ

Figure 2: Benchmarking results for the CS and CCZ gates in Figures (a) and (b), respectively, with the Pauli group, the CZ dihedral group (CZD), and the CNOT dihedral group (CXD). The red dashed line is the theoretical fidelity value. Each box plot contains 20 fidelities, and each fidelity is estimated with circuit depths $\{2, 4, 6, 8, 10\}$, and the total number of different gate sequences for each depth is specified by the horizontal axis.

References

[1] J. Emerson, R. Alicki, and K. Zyczkowski, Journal of Optics B: Quantum and Semiclassical Optics **7**, S347 (2005), URL https://doi.org/10.1088/1464-4266/7/10/021.

[2] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Science **317**, 1893 (2007), https://www.science.org/doi/pdf/10.1126/science.1145699, URL https://www.science.org/doi/abs/10.1126/science.1145699.

[3] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008), URL https://link.aps.org/doi/10.1103/PhysRevA.77.012307.

[4] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011), URL https://link.aps.org/doi/10.1103/PhysRevLett.106.180504.

[5] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012), URL https://link.aps.org/doi/10.1103/PhysRevA.85.042311.

[6] R. Harper, S. T. Flammia, and J. J. Wallman, Nature Physics **16**, 1184 (2020), ISSN 1745-2481, URL https://doi.org/10.1038/s41567-020-0992-8.

[7] J. J. Wallman and J. Emerson, Phys. Rev. A **94**, 052325 (2016), URL https://link.aps.org/doi/10.1103/PhysRevA.94.052325.

[8] S. T. Flammia and J. J. Wallman, ACM Transactions on Quantum Computing **1** (2020), URL https://doi.org/10.1145/3408039.

[9] R. Harper, W. Yu, and S. T. Flammia, PRX Quantum **2**, 010322 (2021), URL https://link.aps.org/doi/10.1103/PRXQuantum.2.010322.

[10] S. T. Flammia and R. O'Donnell, Quantum **5**, 549 (2021), ISSN 2521-327X, URL https://doi.org/10.22331/q-2021-09-23-549.

[11] C. Rouzé and D. S. França, *Efficient learning of the structure and parameters of local pauli noise channels* (2023), 2307.02959.

[12] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Nature Communications **10**, 5347 (2019), ISSN 2041-1723, URL https://doi.org/10.1038/s41467-019-13068-7.

[13] Y. Zhang, W. Yu, P. Zeng, G. Liu, and X. Ma, Photon. Res. **11**, 81 (2023), URL https://opg.optica.org/prj/abstract.cfm?URI=prj-11-1-81.

[14] G. Liu, X. Zhang, and X. Ma, Quantum **6**, 845 (2022), ISSN 2521-327X, URL https://doi.org/10.22331/q-2022-10-24-845.

[15] P. Shor, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994), pp. 124–134.

[16] L. K. Grover, in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (1996), pp. 212–219.

[17] A. M. Childs and N. Wiebe, Quantum Info. Comput. **12**, 901–924 (2012), ISSN 1533-7146.

[18] A. Fedorov, L. Steffen, M. Baur, M. P. da Silva, and A. Wallraff, Nature **481**, 170 (2012), ISSN 1476-4687, URL https://doi.org/10.1038/nature10713.

[19] S. Li, A. D. Castellano, S. Wang, Y. Wu, M. Gong, Z. Yan, H. Rong, H. Deng, C. Zha, C. Guo, et al., npj Quantum Information **5**, 84 (2019), ISSN 2056-6387, URL https://doi.org/10.1038/s41534-019-0202-7.

[20] T. Monz, K. Kim, W. Hänsel, M. Riebe, A. S. Villar, P. Schindler, M. Chwalla, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **102**, 040501 (2009), URL https://link.aps.org/doi/10.1103/PhysRevLett.102.040501.

[21] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuletić, H. Pichler, et al., Phys. Rev. Lett. **123**, 170503 (2019), URL https://link.aps.org/doi/10.1103/PhysRevLett.123.170503.

[22] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, npj Quantum Information **2**, 16012 (2016), ISSN 2056-6387, URL https://doi.org/10.1038/npjqi.2016.12.

[23] M. A. Nielsen, Physics Letters A **303**, 249 (2002), ISSN 0375-9601, URL https://www.sciencedirect.com/science/article/pii/S0375960102012720.

[24] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Phys. Rev. A **71**, 062310 (2005), URL https://link.aps.org/doi/10.1103/PhysRevA.71.062310.

[25] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, npj Quantum Information **5**, 71 (2019), ISSN 2056-6387, URL https://doi.org/10.1038/s41534-019-0182-7.

[26] J. Helsen, I. Roth, E. Onorati, A. Werner, and J. Eisert, PRX Quantum **3**, 020357 (2022), URL https://link.aps.org/doi/10.1103/PRXQuantum.3.020357.

[27] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, et al., Phys. Rev. Lett. **109**, 080505 (2012), URL https://link.aps.org/doi/10.1103/PhysRevLett.109.080505.

Group Twirling and Noise Tailoring for Multi-Qubit-Controlled Phase Gates

Guoding Liu, Ziyi Xie, Zitai Xu, and Xiongfeng Ma*
Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084 China

Group twirling is crucial in quantum information processing, particularly in randomized benchmarking and randomized compiling. While protocols based on Pauli twirling have been effectively crafted to transform arbitrary noise channels into Pauli channels for Clifford gates — thereby facilitating efficient benchmarking and mitigating worst-case errors — the lack of practical twirling groups for multi-qubit non-Clifford gates remains a challenge. To address this gap, we study the issue of finding twirling groups for generic quantum gates, focusing on a widely used circuit structure in randomized benchmarking or randomized compiling. Specifically, for multi-qubit-controlled phase gates, which are essential in quantum algorithms and directly implementable in practice, we determine optimal twirling groups within the realm of classically replaceable unitary operations. Contrasting with the local Pauli twirling group for Clifford gates, the optimal groups for such gates contain nonlocal operations, highlighting the overhead of tailoring noise in global non-Clifford contexts. We design new benchmarking procedures for multi-qubit controlled phase gates based on the optimal twirling groups. Our simulation results show that our scheme can improve the precision and accuracy of benchmarking in small-scale systems.

I. INTRODUCTION

There has been an increased interest in quantum information processing due to its potential revolution in both science and technology. While quantum computing holds the promise of quantum advantages, its practical implementation faces significant challenges, primarily due to the inherent noise of quantum systems. When dealing with quantum noise, group-twirling-based noise tailoring is an essential step. Group twirling symmetrizes the noise channel [1, 2], allowing accurate and efficient extraction of noise channel parameters. Then, it enables efficient noise characterization and subsequent quantum control optimization [3, 4]. This principle underlies the randomized benchmarking (RB) methodology [5–8], which stands out as a major quantum benchmarking technique due to its low sample complexity and resilience against state preparation and measurement (SPAM) errors. Moreover, group twirling is essential in randomized compiling [9], which turns generic noise into a Pauli channel, reducing the worst-case error of quantum gates and facilitating Pauli channel learning protocols [10–13].

While randomized benchmarking and randomized compiling have seen considerable advancements, efficient noise tailoring protocols for multi-qubit gates mainly focus on the Clifford case, primarily achieved through Pauli group twirling [9, 14, 15]. For multi-qubit non-Clifford gates, researchers also made some progress and proposed benchmarking protocols for two kinds of quantum gate sets: CNOT dihedral group [16] and matchgate group [17, 18]. Nonetheless, these groups are both large and highly non-local, posing challenges to practical implementation in expansive quantum systems. Experimental undertakings have, so far, been limited to the two-qubit domain for the CNOT dihedral group [19] and none for the matchgate group. This raises the question of whether more compact and easily implementable twirling groups might suffice for twirling and benchmarking tasks. Similar issues exist for randomized compiling. The lack of methodologies for tailoring multi-qubit non-Clifford gates hinders their applications, though many of which are directly implementable or native in quantum processors [20, 21].

In this work, we investigate noise tailoring strategies for generic quantum gates within RB and randomized compiling. Central to noise tailoring is the selection of appropriate twirling gates for the twirled gate, namely the gate undergoing twirling. We study this question within a frequently used circuit structure – the twirled and twirling gates are intertwined. Within this structure, we summarize the constraints between the twirled and twirling gates, and we find that any quantum gate tailoring demands a twirling gate set comparable in magnitude to the Pauli group, implying the optimality of existing noise tailoring schemes designed for Clifford gates.

In addition to the well-studied Clifford gates, our study emphasizes multi-qubit-controlled phase gates, given by

$$C^nZ_m = \begin{pmatrix} \mathbb{I}_{2^{n+1}-1} & \mathbf{0} \\ \mathbf{0} & e^{i\frac{2\pi}{m}} \end{pmatrix}, \tag{1}$$

where n and m are both positive integers, representing the number of control qubits and the phase, respectively. The matrix $\mathbb{I}_{2^{n+1}-1}$ denotes the identity of dimension $2^{n+1}-1$ and $\mathbf{0}$ denotes the zero matrix. Our study can be further generalized by replacing the phase from $\frac{2\pi}{m}$ to any angle $\theta \in [0, 2\pi]$. These gates are the key components in quantum information processing, featuring in seminal algorithms such as Shor's algorithm [22], Grover's algorithm [23], linear combinations of unitary operations [24], and the preparation of hypergraph states for universal quantum computing [25]. Moreover, the C^nZ_m gate is native to superconducting [26, 27], ion trap [28], and Ry-

* xma@tsinghua.edu.cn

dberg [29] quantum systems. It is also a key ingredient to construct fault-tolerant non-Clifford gates in quantum error correction [30–32]. Thus, devising efficient noise tailoring schemes for such gates is crucial in both the theory and experiment.

Within our framework, we introduce an optimal noise tailoring scheme for the C^nZ_m gate if the twirling group falls within the class of classically replaceable unitary operations (CRU) [33] or incoherent unitary operations [34, 35]. CRU comprises gates that can be moved after computational basis measurements, typically the Z-basis measurements, and replaced by classical post-processing. Unlike the relatively straightforward tailoring process for Clifford gates, the optimal twirling gate set for C^nZ_m grows exponentially with the number of qubits.

We further conducted numerical simulations to assess the benchmarking performance of the CS gate and C^nZ gates using various noisy twirling groups. Our findings indicate that the optimal twirling group delivers superior performance compared to the CNOT dihedral group and the Pauli group in a small-scale quantum system. We believe these results will contribute to the broader use of native non-Clifford gates in quantum computing and facilitate the practical implementation of a wide range of quantum algorithms.

The structure of this paper is organized as follows. In Section II, we rigorously describe the problem of finding the twirling group for a target gate and show the main results. In Section III, we extend the results to randomized compiling. In Section IV, we show simulation results of RB protocols with noisy twirling groups. Finally, we conclude in Section V.

II. TWIRLING GROUPS IN RANDOMIZED BENCHMARKING

We start with clarifying the algebraic relations between the twirled gate and the twirling gates in RB, following the ideas of [14, 36–38]. More details about RB and the technical derivations of the results in this work are available in the Appendices A and B.

The task of RB is estimating the fidelity of an individual gate, U, robust to SPAM errors. Consider a noisy quantum gate, $\widetilde{\mathcal{U}} = \mathcal{U}\Lambda$, where $\mathcal{U}$ is the Pauli-Liouville representation of the noiseless gate U, and $\widetilde{\cdot}$ represents the noisy version. The noise channel is denoted by Λ. The key to enabling RB is obtaining the powers of the G-twirled noise channel, Λ_G^m, where $m \in \mathbb{Z}_+$ and $\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger$. Here, G represents a twirling group that renders Λ_G diagonal in the Pauli-Liouville representation up to a unitary transformation independent of Λ. The diagonalizability ensures reliable and SPAM-error-free extraction of the diagonal elements and hence the fidelity of Λ using single-exponential fitting via character RB techniques [37]. In cases where Λ_G is not diagonal, a matrix exponential fitting challenge occurs, and accurate fidelity estimation becomes problematic.

To obtain Λ_G^m, we consider the circuit in Figure 1. One independently and randomly samples m twirling gates G_i from group G and implements them interleaved with target gate U. The circuit ends with the inverse gate $G_{\text{inv}} = (\prod_{i=1}^m UG_i)^\dagger$. This circuit structure is first devised in interleaved RB [38], but here U is not necessarily in G. The target gate U is the one whose noise channel is twirled, so we call U the twirled gate. The gates from G are named the twirling gates.

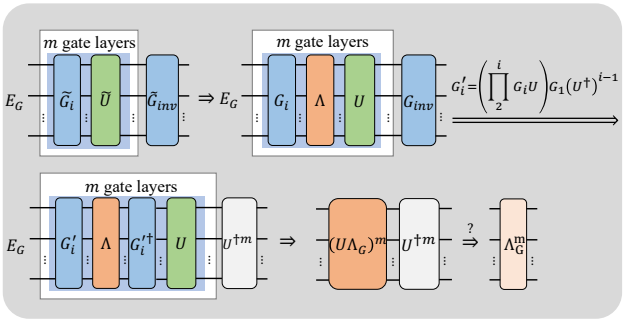


FIG. 1. Random twirling gates $G_1, G_2, \cdots, G_m$ interleaved with U in RB. The inverse gate $G_{\text{inv}} = (\prod_{i=1}^m UG_i)^\dagger$; $\widetilde{\cdot}$ represents the noisy version of a quantum gate; Λ denotes the composite noise channel of U and the twirling group, which reduces to the noise channel of U in the case of noiseless twirling group; Λ_G is the G-twirled noise channel.

Note that in practice, the interleaved circuit measures noise from both the target gate U and the twirling group G. One can set the target gate as the identity to measure the average noise of the twirling group solely. While this requires a gate-independent noise for the twirling group, a higher-order fitting [7, 38] can alleviate this constraint. By comparing the fidelities measured with and without the twirled gate, one can derive an interval estimation of the fidelity of the target gate. The accuracy of this estimation improves as the fidelity of the twirling group increases [38]. Later in the simulation part, we will demonstrate the influence of the gate-dependent noise of the twirling group on the fidelity estimation of the target gate. It turns out that, in this case, the interleaved technique still works in a small-scale system and enables reliable fidelity estimation. In the following, we omit the noise of the twirling group for brevity and use the noise channel of U, Λ, to represent the composite noise channel of U and the twirling group.

Focusing on the interleaved circuit, we mathematically describe it as follows,

$$\begin{aligned}\widetilde{\mathcal{S}}_m &= \mathcal{G}_{inv}\prod_{i=1}^m \mathcal{U}\Lambda\mathcal{G}_i \\ &= \mathcal{U}^{\dagger m}\prod_{i=1}^m \mathcal{U}\mathcal{G}_i'^\dagger\Lambda\mathcal{G}_i',\end{aligned} \tag{2}$$

where $G_i' = (\prod_{j=2}^i G_jU)G_1U^{\dagger i-1}$ for $1 \le i \le m$. To get Λ_G^m from $\widetilde{\mathcal{S}}_m$, we just need $U\mathsf{G}U^\dagger = \mathsf{G}$ as shown below.

If $U\mathsf{G}U^\dagger = \mathsf{G}$, then $\{G_i', 1 \le i \le m\}$ are independent and random elements from G. After taking expectation, Eq. (2) would become $\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m$. For multi-qubit-controlled phase gates, we found that $U\mathsf{G}U^\dagger = \mathsf{G}$ ensures $\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G$. Then, Eq. (2) would further reduce to Λ_G^m as we want. For other quantum gates, one can at least obtain the fidelity of $(\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}$, which we proved to be a lower bound of the fidelity of Λ in Lemma 11 in the Appendix. The lower bound of the fidelity is known to be fidelity witness [39] and is also useful in quantum benchmarking. When $U\mathsf{G}U^\dagger = \mathsf{G}$, one can also choose m such that $U^m = \mathbb{I}$ to ensure the inverse gate belonging to G and implement the inverse gate without additional difficulty. Below, we summarize the requirements of the twirling group to tailor U in RB.

Question 1. *Given a gate, U, find a twirling group, G such that,*

$$\begin{aligned} &\textit{for any quantum channel } \Lambda, \\ &\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \textit{ is diagonal up to a} \\ &\textit{unitary transformation independent of } \Lambda, \end{aligned} \tag{3}$$

and,

$$U\mathsf{G}U^\dagger = \mathsf{G}. \tag{4}$$

Equation (3) implies that the twirling group G must make the twirled noise channel Λ_G sufficiently symmetric, irrespective of the characteristics of the original noise channel. Additionally, Eq. (4) ensures that the operation of U does not destroy this symmetry. Identifying the smallest and most easily implementable groups within Question 1 is not only pivotal for devising practical benchmarking schemes but also crucial for understanding the varying levels of difficulty in benchmarking different quantum gates. To find the optimal group for a gate, we examine the diagonalizability of Λ_G, which imposes constraints on G as follows.

Lemma 1. *If a finite n-qubit unitary subgroup, G, satisfies Eq. (3), then the Pauli-Liouville representation of G is multiplicity-free. As a corollary, the cardinality of the twirling group $|\mathsf{G}| \ge 4^n$.*

In the proof of Lemma 1, we leverage the arbitrariness of Λ to demonstrate that any trace-preserving map can be twirled into a diagonal form. From this, we prove G multiplicity-free, which means that any irreducible representation of G appears at most once in the decomposition of the Liouville representation of G. Then, with Burnside's theorem [40], which relates the group size to irreducible representations, we obtain that $|\mathsf{G}| \ge 4^n$. Interestingly, Lemma 1 is a converse proposition of a result in character RB [37], which asserts that a multiplicity-free group G can twirl any channel into a diagonal form.

The above lemma reveals that the twirling group for noise tailoring cannot be smaller than the (projective) Pauli group, whose cardinality achieves 4^n. Pauli gates are also local and easily implementable. Thus, we show the superiority of Clifford gates in noise tailoring, for which the Pauli group is a solution to Question 1. For a non-Clifford gate, this solution fails, for which we present a simple but may not be an optimal solution in Appendix B 3.

Lemma 1 shows requirements for generic unitary twirling groups. Below, in Lemma 2, we consider twirling groups belonging to CRU and show much stronger requirements other than the cardinality constraint for twirling groups. CRU comprises all gates in the product of a permutation matrix, like Pauli X and Toffoli gates, and a diagonal matrix on the computational basis, like Pauli Z gate. This set is large and becomes universal after adding Hadamard gates. Using CRU subgroups for twirling lets us replace the inverse gate with classical post-processing when measurements are under the computational basis [33], bringing additional advantages in practical implementation. We provide more discussions about CRU twirling groups in Appendix B 4.

Lemma 2 (Informal Version)**.** *If a finite n-qubit CRU subgroup, G, satisfies Eq. (3), then G can interchange any two computational basis states. That is, for any two computational basis states $|\mathbf{i}\rangle$ and $|\mathbf{j}\rangle$ where $\mathbf{i},\mathbf{j} \in \{0,1\}^n$, there exists a gate $G \in \mathsf{G}$ such that $|\mathbf{j}\rangle = G|\mathbf{i}\rangle$.*

To prove Lemma 2, we leverage the decomposable property of CRU into a permutation matrix and a diagonal matrix, facilitating the analysis of twirling. We also introduce Burnside's lemma [41] to transform the issue of calculating multiplicities into a problem of orbit counting. More specifically, we translate the multiplicity of the trivial representation into the orbit of G acting on computational basis states. Then, Lemma 2 follows from Lemma 1.

Note that diagonal gates, CNOT gates, and Toffoli gates do not affect $|\mathbf{0}\rangle$. Lemma 2 implies that the CRU subgroup G must include a gate set like $\mathsf{X} = \langle X\rangle^{\otimes n}$. Combining this with Eq. (4), we can deduce the necessary quantum gates that G must contain for tailoring a gate, U. Furthermore, when U is a multi-qubit-controlled phase gate, we derive its optimal twirling group as detailed in Theorem 1.

Theorem 1. *The optimal twirling group G in CRU for the multi-qubit-controlled phase gate, $U = C^nZ_m$, with $n \ge 1, m \ge 2$, is given by*

$$\mathsf{G} = \{\Pi(\prod_{i=1}^{t}(\Pi_i^\dagger U\Pi_i U^\dagger)^{l_i})|\Pi \in \mathsf{X}, t \in \mathbb{N}, \forall i, l_i \in \pm 1, \Pi_i \in \mathsf{X}\}. \tag{5}$$

Group G is the smallest group containing X and normalized by U, which we prove to satisfy Question 1. Note that any CRU subgroup is decomposable into the semi-product of a permutation group and a diagonal group,

$$\mathsf{G}_{\mathrm{CRU}} = \Pi \ltimes \mathsf{W} = \{\Pi W|\Pi \in \Pi, W \in \mathsf{W}\}. \tag{6}$$

The optimal group can be written as $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ where $\mathsf{W}_X = \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \mathsf{X}\}$ only comprises diagonal gates. The optimality here means that any $\mathsf{G}_{\mathrm{CRU}}$ satisfying Question 1 implies $\mathsf{W} \supseteq \mathsf{W}_X$ and $|\mathsf{\Pi}| \geq |\mathsf{X}|$. The optimal group is clearly the smallest option. In practice, one normally determines the permutation part and the diagonal part separately to sample a gate in CRU. The gate is realized by sequentially implementing two parts. In this sense, the optimal group $\mathsf{X} \ltimes \mathsf{W}_X$ is also the most 'local' and easily implementable option since X and W_X are the most 'local' choices of the permutation part and the diagonal part, respectively.

Specific forms of twirling groups for C^nZ and CZ_m are summarized in Table I. Group G reduces to $\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle$ and $\langle CZ_m, Z_m, X\rangle$ for $U = C^nZ$ and $U = CZ_m$ with odd m, respectively. For $U = CZ_m$ with even m, $\mathsf{G} \leq \langle CZ_{m/2}, Z_m, X\rangle$. Within group generator $\langle\cdot\rangle$, we use X to denote $\{X_1, X_2, \cdots, X_n\}$, the Pauli X gates on all qubits, and similarly omit subscripts associated with the acting qubits when referring to $C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z$ and CZ_m, Z_m.

C^nZ			
	Group	Size	Complexity
This work	$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle$	$O(N^n)$	$O(N^n)$
CXD [16]	$\langle CX, Z_{2^{n+1}}, X\rangle$	$O(N^{n+1})$	$O(N^{3n+1})$
CZ_m			
	Group	Size	Complexity
This work	$\langle CZ_m, Z_m, X\rangle$	$O(N^2 \log m)$	$O(N^2 \log m)$
CXD	$\langle CX, Z_{2m}, X\rangle$	$O(N^{\log m})$	$O(N^{3\log m+1})$

TABLE I. Scaling of the group size and computational complexity of the twirling group for tailoring C^nZ and CZ_m in this work and Ref. [16]; CXD represents the CNOT dihedral group. The term "Complexity" refers to the complexity of computing the multiplication and the inverse of group elements. The size and the complexity are expressed with the qubit number N, controlled qubit number n, and phase angle index m. Unlike Ref. [16], which specifically addresses computational complexity for $m = 2^k$, our results apply to generic positive integer m.

Note that enabling noise tailoring requires sampling from the twirling group and performing group element multiplication. These tasks introduce considerations for sample complexity, directly related to group size and computational complexity, contingent on the algorithm used for group multiplication. Table I provides complexity results for C^nZ and CZ_m and shows that our method surpasses previous results in [16]. Nonetheless, even with the optimal approach, the twirling group becomes large and highly non-local as the controlled qubit number increases. This leads to unfavorable sample and computational complexity for large quantum gates, highlighting an inherent challenge in benchmarking multi-qubit non-Clifford gates.

III. RANDOMIZED COMPILING FOR MULTI-QUBIT NON-CLIFFORD GATES

In this part, we discuss the application of previous results in randomized compiling. The task of randomized compiling is turning the noisy quantum gate $\widetilde{\mathcal{U}} = \mathcal{U}\Lambda$ to $\mathcal{U}\Lambda_G$, where the twirled noise channel Λ_G is a Pauli channel. More specifically, suppose that we implement a quantum circuit, $\widetilde{C} = \cdots\widetilde{\mathcal{U}}_2\widetilde{\mathcal{U}}_1 = \cdots\mathcal{U}_2\Lambda_2\mathcal{U}_1\Lambda_1$, as shown in Figure 2(a). The aim is tailoring $\widetilde{C}$ into $\cdots\mathcal{U}_2\Lambda_{2G}\mathcal{U}_1\Lambda_{1G}$ as shown in Figure 2(c). The quantum gates $U_i, i \in \mathbb{Z}_+$ are tailored gates. To tailor them, we add twirling gates G_i and $G_i' = U_iG_i^\dagger U_i^\dagger$ beside the tailored gates as shown in Figure 2(b), where G_i is randomly sampled from a twirling group, G. In reality, the gates G_i and G_{i-1}' are merged and implemented together to reduce quantum resource consumption. Thus, the twirling gates in general belong to $\mathsf{V} = \bigcup_{U\in\mathsf{U}} \mathsf{G}U\mathsf{G}U^\dagger$.

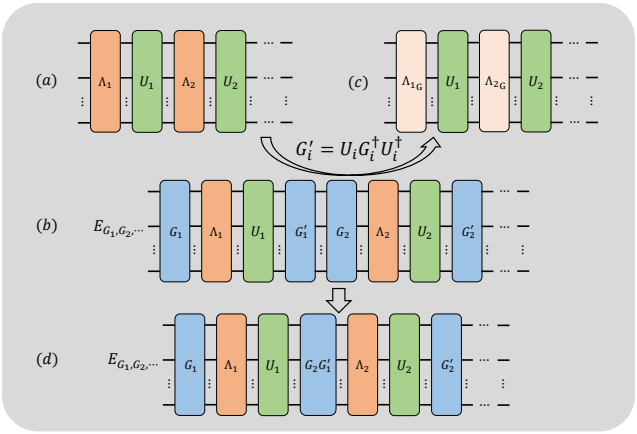


FIG. 2. Randomized compiling for U_1, U_2, $\cdots$, by inserting twirling gates G_i and $G_i' = U_iG_i^\dagger U_i^\dagger$ beside U_i with $i \in \mathbb{Z}_+$; Λ_{iG} is the G-twirled noise channel required to be a Pauli channel. In reality, G_i and G_{i-1}' are merged and implemented together as quantum gate G_iG_{i-1}'.

Similar to the discussion in RB, we investigate the requirements of the twirling gates for tailoring a gate set $\mathsf{U} = \{U_1, U_2, \cdots\}$. More specifically, we should optimize the choice of G to make the twirling gate set $\mathsf{V} = \bigcup_{U\in\mathsf{U}} \mathsf{G}U\mathsf{G}U^\dagger$ smaller and more easily implementable in practice. We summarize the question below.

Question 2. *Given a gate set, $\mathsf{U} = \{U_1, U_2, \cdots\}$, find a twirling gate set, V, such that $\mathsf{V} = \bigcup_{U\in\mathsf{U}} \mathsf{G}U\mathsf{G}U^\dagger$ and the twirling group G satisfies*

$$\Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \text{ is a Pauli channel.} \tag{7}$$

When the tailored gate set U only comprises Clifford gates and single-qubit T gates, the twirling gate set suffices to be chosen as the dihedral group $\langle X, S\rangle$ [9]. Here, we consider that U is composed of a multi-qubit controlled phase gate $U = C^nZ_m$. Question 2 reduces to finding gate set $\mathsf{G}U\mathsf{G}U^\dagger$ while Λ_G is a Pauli channel.

When considering G as a CRU subgroup, $\mathsf{G}U\mathsf{G}U^\dagger$ can be written in a form,

$$\mathsf{V} = \mathsf{G}\mathsf{W} = \{GW | G \in \mathsf{G}, W \in \mathsf{W}\}, \tag{8}$$

where $\mathsf{W} = \{G^\dagger U G U^\dagger, G \in \mathsf{G}\}$ is a set of diagonal gates. Similar to the results in RB, for $U = C^n Z_m$, $\mathsf{W} \supseteq \mathsf{W}_X = \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \mathsf{X}\}$ based on Lemma 2, proved in the proof of Theorem 1. Meanwhile, based on Lemma 1, G cannot be smaller and more easily implementable than the Pauli group P_n. On the other hand, by choosing G as Pauli group P_n, we would obtain a solution $\mathsf{V} = \mathsf{P}_n U \mathsf{P}_n U^\dagger = \mathsf{P}_n \mathsf{W}_X$ to Question 2. This is the optimal solution for tailoring a multi-qubit controlled phase gate with G restricted in CRU if one separately implements the permutation matrix and the diagonal matrix for a CRU gate.

Note that the twirling gate set $\mathsf{P}_n \mathsf{W}_X$ is simpler than the twirling group $\mathsf{X} \ltimes \mathsf{W}_X$ for $C^n Z_m$. Generally speaking, the requirements for the twirling gates in randomized compiling are much simpler than those in RB. A direct observation to Question 2 is that as long as $U\mathsf{P}_n U^\dagger$ only comprises local gates, U can be tailored with only local twirling gates in randomized compiling. We leave the problem of tailoring more kinds of quantum gates in the setting of randomized compiling to the future.

IV. SIMULATION OF BENCHMARKING MULTI-QUBIT-CONTROLLED PHASE GATES

Below, we present the numerical benchmarking results for CS and $C^n Z$ gates with noisy twirling groups. To simulate practical benchmarking experiments, we consider the noise of the twirling group to be gate-dependent. Here, all twirling gates and the target gate belong to CRU, enabling us to sample and simulate a gate by separately determining its permutation part and diagonal part as follows.

- The permutation part is generated by Pauli X and CNOT gates. We consider unitary errors $e^{i\delta X}$ and controlled-$e^{i\delta X}$ for the two gates, respectively, where δ is small.

- The diagonal part is generated by multi-qubit controlled phase gates and single-qubit Z rotation. The unitary error of the former is considered to be a diagonal gate where the diagonal elements depend on the control qubit number. The unitary error of the latter is a small Z rotation.

We also consider local dephasing and local amplitude-damping noise, where the noise strength depends on the time to implement the gate. The SPAM is also simulated with flip errors. More details about the noise model are shown in Appendix C 1.

Notice that one needs first to obtain the fidelity of the composite noise channel, denoted as F_1, via the interleaved circuit and the fidelity of the twirling group, denoted as F_2, via the reference circuit. The fidelity of the target gate is evaluated by

$$F = \frac{d^2 F_1 - 1}{d^2 F_2 - 1}\left(1 - \frac{1}{d^2}\right) + \frac{1}{d^2}. \tag{9}$$

The derivation is given in Appendix A 5. In this section, we only present the fidelity of the target gate, while the other two kinds of fidelities are available in Appendix C 3. We also simulate the case of the noiseless twirling group, and the results are shown in Appendix C 5.

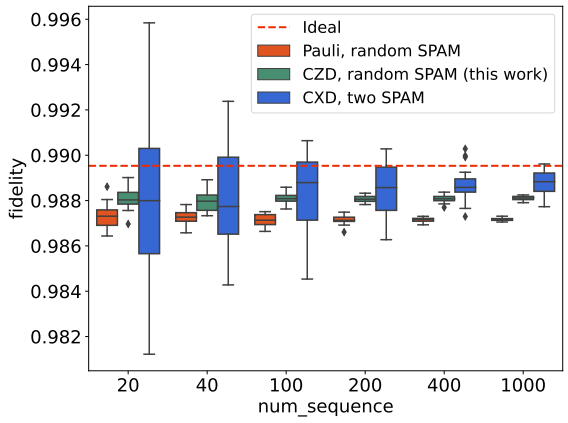
Depending on the SPAM settings, we simulate two different benchmarking procedures and show the results in Sections IV A and IV B, respectively. We compare our scheme with the Pauli group and the CNOT dihedral group. Note that the Pauli group does not satisfy the requirements in Question 1 to benchmark multi-qubit controlled phase gates. In this case, the inverse gate in Figure 1 has to be outside the Pauli group to make the ideal circuit equal to the identity. In general, this inverse gate can take any element within the optimal group given by Eq. (5), which may be nonlocal.

A. Benchmarking with random SPAM settings

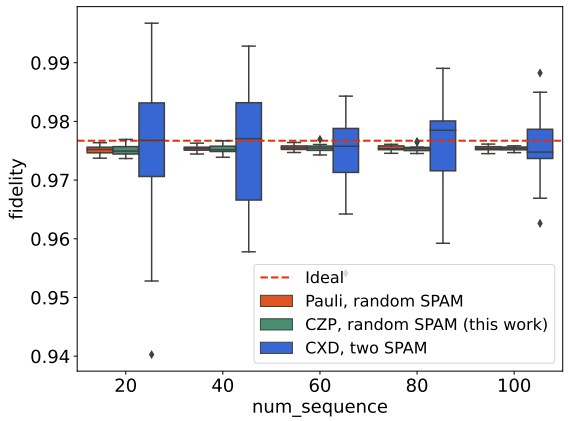
Recall that the gate fidelity is estimated by evaluating different diagonal terms of twirled noise channels in the Pauli-Liouville representation. The number of different diagonal terms of the twirled noise channel depends on the twirling group. This number increases exponentially with respect to the qubit number for the Pauli group and the optimal group and is 2 for the CNOT dihedral group. In the simulation, we sample and evaluate 20 different diagonal terms instead of all of them for the former two groups. For the CNOT dihedral group, we only evaluate 2 different diagonal terms.

To evaluate a diagonal term, $\mathrm{tr}(P\Lambda(P))/2^N$, where Λ is the noise channel, N is the qubit number, and P is a non-identity Pauli observable, we simulate the circuit in Figure 1 with the eigenstate of P as input and P as measurement observable. By simulating randomly sampled gate sequences and averaging the results over different sequences, we estimate $\mathrm{tr}\left(\widetilde{P}\Lambda_{\mathsf{G}}^m(\widetilde{|\psi\rangle\langle\psi|})\right)$ where Λ_{G} is the twirled noise channel and $P|\psi\rangle = |\psi\rangle$. The circuit depth m is taken from $\{2, 4, 6, 8, 10\}$, and the number of different gate sequences for each circuit depth varies. The total number of gate sequences for different groups is the same to make a fair comparison. By fitting $\mathrm{tr}\left(\widetilde{M}\Lambda_{\mathsf{G}}^m(\widetilde{\rho})\right)$ to $A\lambda^m$ one can get λ and estimate $\mathrm{tr}(P\Lambda(P))/2^N$.

In Figure 3, we present the simulation results for the CS and CCZ gates. The results demonstrate that the fluctuation of the benchmarking results for the optimal group is as small as the Pauli group and is much smaller than the CNOT dihedral group for benchmarking two gates. The large fluctuation of the CNOT dihedral group mainly results from a more severe gate-dependent noise.



(a) CS



(b) CCZ

FIG. 3. Benchmarking results for the CS and CCZ gates in Figures (a) and (b), respectively, with the optimal twirling group, the Pauli group, and the CNOT dihedral group. The optimal twirling group is the CZ dihedral group $\langle CZ, Z, S\rangle$ for CS, and the CZ Pauli group $\langle CZ, Z, P\rangle$ for CCZ. The red dashed line is the theoretical value of the noise channel fidelity. Each box plot contains 20 fidelities, and each fidelity is estimated with circuit depths $\{2, 4, 6, 8, 10\}$, and the total number of different gate sequences for each depth is specified by the horizontal axis. Here, for the Pauli group and the optimal group, we randomly sample and estimate 20 different diagonal terms of the twirled noise channel. We mark this setting with 'random SPAM' on the label. For the CNOT dihedral group, we estimate only two different diagonal terms of the twirled noise channel. We mark this setting with 'two SPAM' on the label. Each SPAM setting prepares an eigenstate of a Pauli observable with eigenvalue 1 and measures this Pauli observable. In 'two SPAM', the two Pauli observables are chosen as $Z^{\otimes N}$ and $X^{\otimes N}$.

Concerning the bias, the performance of the optimal group is between the other two groups for benchmarking CS gate, while the performance of the three groups is close when benchmarking CCZ gates. The bias of the Pauli group results from a weaker twirling action than the other two groups. Meanwhile, the sampling of diagonal terms also introduces bias to the results of the Pauli group and the optimal group. The gate-dependent noise also influences the accuracy of the results. The CNOT dihedral group is influenced the most, followed by the optimal group and the Pauli group.

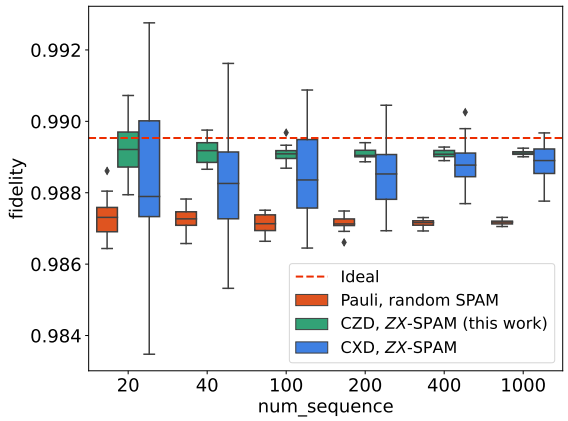
B. Benchmarking with the ZX-SPAM setting

We notice that when the twirling group contains the CZ dihedral group, $\langle CZ, Z, S\rangle$, another SPAM setting can be employed to enhance the performance, which we name ZX-SPAM. Instead of extracting the diagonal term $\mathrm{tr}(P\Lambda(P))/2^N$ once at a time, by preparing $|0\rangle^{\otimes N}$ and measuring in $Z^{\otimes N}$, one can extracting $\mathrm{tr}(P_Z\Lambda(P_Z))/2^N$ for all $P_Z \in \{\mathbb{I}, Z\}^{\otimes N}$. Similarly, by preparing $|+\rangle^{\otimes N}$ and measuring in $X^{\otimes N}$, one can extract $\mathrm{tr}(P_X\Lambda(P_X))/2^N$ for all $P_X \in \{\mathbb{I}, X\}^{\otimes N}$. These two SPAM settings are sufficient to obtain all different diagonal terms as long as the twirling group includes the CZ dihedral group. Compared to extracting the diagonal term once at a time, the procedure with ZX-SPAM estimates the fidelity more precisely and accurately, thanks to more efficient extraction of diagonal terms and the absence of sampling errors. More details about this simulation procedure are available in Appendix C 2.

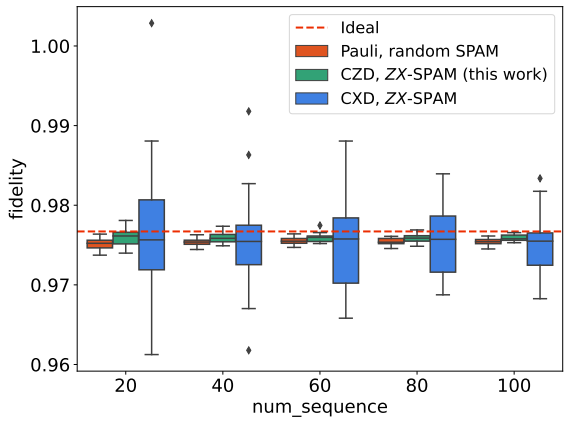
We simulate this enhanced benchmarking procedure and present the results in Figure 4. For C^nZ gates, we use group $\mathsf{G}^Z = \langle C^{n-1}Z, \cdots, CZ, X, S\rangle$ instead of the optimal one $\langle C^{n-1}Z, \cdots, CZ, X, Z\rangle$ for twirling to make the twirling group contain the CZ dihedral group. A little overhead of adding phase gate S simplifies the noise channel more and helps to improve the benchmarking performance. For the CS gate, the optimal group is the CZ dihedral group. To make a fair comparison, in this simulation, we adopt the ZX-SPAM setting for the CNOT dihedral group as well.

Figure 4 shows improved benchmarking results for G^Z and the CNOT dihedral group with the enhanced benchmarking procedure. Among all benchmarking methods, utilizing G^Z with the ZX-SPAM enjoys the best precision and accuracy for benchmarking CS and CCZ gates, providing a practical scheme for tailoring multi-qubit controlled phase gates in small-scale systems.

When the system becomes larger, the CNOT dihedral group no longer works and cannot give a faithful fidelity estimation due to the severe gate-dependent noise of the twirling group. In Figure 5, we present the results of benchmarking C^nZ gates below 6 qubits, comparing the Pauli group, G^Z, and the optimal group. Due to a large gate-dependent noise, the results of the latter two have a large fluctuation for 5 qubits. Nonetheless, the result of G_Z with the ZX-SPAM is still the least deviated from the ideal value. This is also true for 6 qubits, which we demonstrate in Appendix C 4. The simulation emphasizes the importance of identifying small and easily implementable twirling groups with sufficient twirling 'ability.' All of the results are extendable to Tof-



(a) CS



(b) CCZ

FIG. 4. Benchmarking results for the CS and CCZ gates in Figures (a) and (b), respectively, with the Pauli group, the CZ dihedral group $\mathsf{G}^Z = \langle CZ, X, S\rangle$, and the CNOT dihedral group. The group G^Z is denoted as CZD in the label. The red dashed line is the theoretical fidelity value. Each box plot contains 20 fidelities. The setting of circuit depths and sampling is the same as in Figure 3. Nonetheless, for G_Z and the CNOT dihedral group, the SPAM setting and the postprocessing differ from that in Figure 3, which is preparing $|0\rangle^{\otimes N}$ and measuring in Z basis and preparing $|+\rangle^{\otimes N}$ and measuring in X basis to get all different diagonal terms of the twirled noise channel. We mark this setting with 'ZX-SPAM' on the label.

foli gates, controlled-$\sqrt{Y}$ gates, and various useful non-diagonal gates via the technique of local gauge transformation [15].

V. OUTLOOK

In this work, we study noise tailoring from a distinct perspective of identifying suitable twirling groups and show nearly optimal results for multi-qubit-controlled phase gates C^nZ_m. Besides benchmarking gate fidelity,

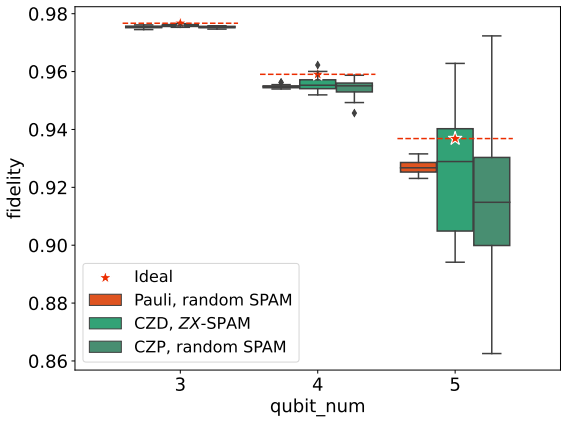


FIG. 5. Benchmarking results for C^nZ gates with the Pauli group, $\mathsf{G}^Z = \langle C^{n-1}Z, \cdots, CZ, X, S\rangle$, and the optimal group $\langle C^{n-1}Z, \cdots, CZ, X, Z\rangle$. The three methods are labeled with 'Pauli, random SPAM', 'CZD, ZX-SPAM', and 'CZP, random SPAM', respectively. The benchmarking setting is the same in Figures 3 and 4. The red five-pointed star with a line denotes the ideal fidelity value. Each box plot contains 20 fidelities.

the tailoring scheme can help extract more information, such as Pauli eigenvalues, and study the learnability of Pauli noise [42]. The results also benefit fault-tolerant protocols and quantum algorithms utilizing multi-qubit controlled phase gates.

In the future, it is crucial to explore optimal twirling groups for practical quantum gates and assess the feasibility of using only local twirling gates for efficient noise tailoring. The gate-dependent noise and the unfavorable size and computational complexity of the optimal twirling groups for large-scale non-Clifford gate benchmarking pose a significant challenge. Potential solutions include exploring circuit structures beyond those shown in Figure 1 and seeking benchmarking methodologies that do not rely on a group structure [36, 43–45]. Based on our simulation results, a twirling gate set between the Pauli group and the optimal group may be an intermediate option to twirl multi-qubit controlled phase gates, like the set of gates implemented within a short depth in the optimal group.

Beyond noise tailoring, the insights gained on CRU subgroups in this study can further enhance random matrix protocols involving them, such as classical shadow [46] and gate-set shadow [47, 48], and contribute to the exploration of the classical simulation capabilities of various groups.

ACKNOWLEDGMENTS

We thank Boyang Chen, Zhuo Chen, Daojin Fan, and Shaowei Li for the helpful discussions. This work was supported by the National Natural Science Founda-

tion of China Grants No. 12174216 and the Innovation Program for Quantum Science and Technology Grants No. 2021ZD0300804 and No. 2021ZD0300702.

[1] J. Emerson, R. Alicki, and K. Zyczkowski, Journal of Optics B: Quantum and Semiclassical Optics **7**, S347 (2005).
[2] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Science **317**, 1893 (2007), https://www.science.org/doi/pdf/10.1126/science.1145699.
[3] L. M. K. Vandersypen and I. L. Chuang, Rev. Mod. Phys. **76**, 1037 (2005).
[4] S. Chu, Nature **416**, 206 (2002).
[5] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008).
[6] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).
[7] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012).
[8] R. Harper, S. T. Flammia, and J. J. Wallman, Nature Physics **16**, 1184 (2020).
[9] J. J. Wallman and J. Emerson, Phys. Rev. A **94**, 052325 (2016).
[10] S. T. Flammia and J. J. Wallman, ACM Transactions on Quantum Computing **1** (2020), 10.1145/3408039.
[11] R. Harper, W. Yu, and S. T. Flammia, PRX Quantum **2**, 010322 (2021).
[12] S. T. Flammia and R. O'Donnell, Quantum **5**, 549 (2021).
[13] C. Rouzé and D. S. França, "Efficient learning of the structure and parameters of local pauli noise channels," (2023), arXiv:2307.02959 [quant-ph].
[14] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Nature Communications **10**, 5347 (2019).
[15] Y. Zhang, W. Yu, P. Zeng, G. Liu, and X. Ma, Photon. Res. **11**, 81 (2023).
[16] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, npj Quantum Information **2**, 16012 (2016).
[17] J. Claes, E. Rieffel, and Z. Wang, PRX Quantum **2**, 010351 (2021).
[18] J. Helsen, S. Nezami, M. Reagor, and M. Walter, Quantum **6**, 657 (2022).
[19] S. Garion, N. Kanazawa, H. Landa, D. C. McKay, S. Sheldon, A. W. Cross, and C. J. Wood, Phys. Rev. Res. **3**, 013204 (2021).
[20] L. B. Nguyen, Y. Kim, A. Hashim, N. Goss, B. Marinelli, B. Bhandari, D. Das, R. K. Naik, J. M. Kreikebaum, A. N. Jordan, D. I. Santiago, and I. Siddiqi, "Programmable heisenberg interactions between floquet qubits," (2022), arXiv:2211.10383 [quant-ph].
[21] Y. Kim, A. Morvan, L. B. Nguyen, R. K. Naik, C. Jünger, L. Chen, J. M. Kreikebaum, D. I. Santiago, and I. Siddiqi, Nature Physics **18**, 783 (2022).
[22] P. Shor, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994) pp. 124–134.
[23] L. K. Grover, in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (1996) pp. 212–219.
[24] A. M. Childs and N. Wiebe, Quantum Info. Comput. **12**, 901–924 (2012).
[25] M. Rossi, M. Huber, D. Bruß, and C. Macchiavello, New Journal of Physics **15**, 113022 (2013).
[26] A. Fedorov, L. Steffen, M. Baur, M. P. da Silva, and A. Wallraff, Nature **481**, 170 (2012).
[27] S. Li, A. D. Castellano, S. Wang, Y. Wu, M. Gong, Z. Yan, H. Rong, H. Deng, C. Zha, C. Guo, L. Sun, C. Peng, X. Zhu, and J.-W. Pan, npj Quantum Information **5**, 84 (2019).
[28] T. Monz, K. Kim, W. Hänsel, M. Riebe, A. S. Villar, P. Schindler, M. Chwalla, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **102**, 040501 (2009).
[29] H. Levine, A. Keesling, G. Semeghini, A. Omran, T. T. Wang, S. Ebadi, H. Bernien, M. Greiner, V. Vuletić, H. Pichler, and M. D. Lukin, Phys. Rev. Lett. **123**, 170503 (2019).
[30] A. Paetznick and B. W. Reichardt, Phys. Rev. Lett. **111**, 090505 (2013).
[31] T. Jochym-O'Connor and T. J. Yoder, Phys. Rev. Res. **3**, 013118 (2021).
[32] Y.-F. Wang, Y. Wang, Y.-A. Chen, W. Zhang, T. Zhang, J. Hu, W. Chen, Y. Gu, and Z.-W. Liu, "Efficient fault-tolerant implementations of non-clifford gates with reconfigurable atom arrays," (2023), arXiv:2312.09111 [quant-ph].
[33] G. Liu, X. Zhang, and X. Ma, Quantum **6**, 845 (2022).
[34] T. Baumgratz, M. Cramer, and M. B. Plenio, Phys. Rev. Lett. **113**, 140401 (2014).
[35] E. Chitambar and G. Gour, Phys. Rev. Lett. **117**, 030401 (2016).
[36] J. Helsen, I. Roth, E. Onorati, A. Werner, and J. Eisert, PRX Quantum **3**, 020357 (2022).
[37] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, npj Quantum Information **5**, 71 (2019).
[38] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 080505 (2012).
[39] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Nature Reviews Physics **2**, 382 (2020).
[40] C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Vol. 356 (American Mathematical Soc., 1966).
[41] W. Burnside, *Theory of groups of finite order* (University, 1911).
[42] S. Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, and L. Jiang, Nature Communications **14**, 52 (2023).
[43] J. Chen, D. Ding, and C. Huang, PRX Quantum **3**, 030320 (2022).
[44] M. Heinrich, M. Kliesch, and I. Roth, "Randomized benchmarking with random quantum circuits," (2023), arXiv:2212.06181 [quant-ph].
[45] Y. Gu, W.-F. Zhuang, X. Chai, and D. E. Liu, Nature Communications **14**, 5880 (2023).
[46] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**,

1050 (2020).
[47] J. Helsen, M. Ioannou, J. Kitzinger, E. Onorati, A. Werner, J. Eisert, and I. Roth, Nature Communications **14**, 5039 (2023).
[48] Y. Wang, G. Liu, Z. Liu, Y. Tang, X. Ma, and H. Dai, "Robust estimation of nonlinear properties of quantum processes," (2023), arXiv:2312.09643 [quant-ph].
[49] M. Horodecki, P. Horodecki, and R. Horodecki, Physical Review A **60**, 1888 (1999).
[50] B. Steinberg, *Representation theory of finite groups:an introductory approach* (Springer, 2012).
[51] W. Fulton and J. Harris, *Representation Theory: A First Course* (Springer New York, New York, NY, 2004).
[52] A. Winter and D. Yang, Phys. Rev. Lett. **116**, 120404 (2016).
[53] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Phys. Rev. X **4**, 011050 (2014).
[54] T. G. d. Brugière, M. Baboulin, B. Valiron, S. Martiel, and C. Allouche, IEEE Transactions on Quantum Engineering **2**, 1 (2021).

Here, we briefly introduce the content of the Appendices. In Appendix A, we introduce the notations used in this work and preliminaries about quantum channels, group theory, randomized benchmarking, and classically replaceable unitary operations. In Appendix B, we show the proof of the main results in this work, including restating the requirements of the twirling groups in randomized benchmarking, proof of main lemmas and theorems in the main text, and the structure analysis of the optimal twirling groups for multi-qubit controlled phase gates. In Appendix C, we show the simulation details, including the noise model, the benchmarking protocol, and additional benchmarking results.

CONTENTS

Appendix A: Preliminary

In this part, we introduce some related works and basic mathematical tools. Below, we first introduce basic notations. The Hilbert space for n qubits is denoted as $\mathcal{H}$ and the set of linear operators on $\mathcal{H}$ is denoted as $\mathcal{L}(\mathcal{H})$. Same in the main text, we denote a quantum gate, or a unitary transformation acting on $\mathcal{H}$, in the standard representation with a capital letter, like U and G. A set composed of quantum gates is denoted with sans serif fonts like S. A quantum gate set satisfying group condition under the gate composition is always denoted as G. Note that, in this work, we distinguish two concepts: the twirling gate set and the twirling group. The twirling group is a twirling gate set with a group structure. We normally use V and G to represent the twirling gate set and the twirling group, respectively. The Liouville representation of U is denoted as its calligraphic letter, $\mathcal{U}$. In this paper, the Liouville representation refers to the Pauli-Liouville representation, which will be reviewed in detail in the following subsection. Quantum channels are defined as completely positive and trace-preserving (CPTP) linear maps on $\mathcal{L}(\mathcal{H})$. The noise channel of U is normally denoted as Λ, and we use the same expression for the map representation and the Liouville representation for Λ. We use a wavy line to represent the noisy version of U like $\widetilde{U}$ and $\widetilde{\mathcal{U}}$. Note that a quantum gate is always a quantum channel, but a quantum channel is generally not a quantum gate. We summarize the frequently-used notations in Table II. Note that we sometimes reuse part of notations, and the meanings of these notations are relevant to the context.

TABLE II. Notation

Quantity	Notation
n	number of controlled qubit in C^nZ_m or qubit number
m	phase of controlled qubit in C^nZ_m or circuit depth
N	qubit number
$\mathcal{H}$	Hilbert space
$\mathcal{L}(\mathcal{H})$	the set of linear operators on $\mathcal{H}$
U	target gate
V	twirling gate
G	twirling gate in group G
$\mathcal{U}$	Pauli Liouville representation of U
Λ	noise channel
Λ_G	G-twirled noise channel
$\widetilde{U}$	noisy version of U
V	twirling gate set
G	twirling gate group
$C^nZ_m = \begin{pmatrix} \mathbb{I}_{2^{n+1}-1} & \mathbf{0} \\ \mathbf{0} & e^{i\frac{2\pi}{m}} \end{pmatrix}$	multi-qubit controlled phase gate
X_j	Pauli X on qubit j
Y_j	Pauli Y on qubit j
Z_j	Pauli Z on qubit j
$\mathbb{I}$	identity operator
I	identity channel
$\langle G_1, G_2, ... \rangle$	Group generated by $G_1, G_2, ...$
P_n	n-qubit Pauli group $\langle X, Z \rangle^{\otimes n}$
D_n^m	n-qubit local dihedral group $\langle X, Z_m \rangle^{\otimes n}$
χ	character function of a group
Π	projector or permutation matrix

1. Pauli-Liouville representation

Here, we introduce the Pauli-Liouville representation of quantum channels. Normally, a quantum channel, Λ, is defined within its Kraus representation. For any $O \in \mathcal{L}(\mathcal{H})$, the action of channel Λ is defined with:

$$\Lambda(O) = \sum_{l=1}^{k} K_l O K_l^\dagger, \tag{A1}$$

where $k \in \mathbb{Z}_+$ and $\{K_l, 1 \le l \le k\}$ is the set of Kraus operators satisfying the condition:

$$\sum_{l=1}^{k} K_l^\dagger K_l = \mathbb{I}, \tag{A2}$$

where $\mathbb{I}$ represents the identity operator.

The Kraus representation is not a matrix representation and, hence, not convenient for our work. For further elaboration, we introduce the Liouville representation, which is defined on a set of trace-orthonormal basis elements in $\mathcal{L}(\mathcal{H})$. In this work, we choose the basis to be the normalized and projective Pauli group. In this case, the representation is named Pauli-Liouville representation. The n-qubit Pauli group, denoted as P_n, is given by:

$$\mathsf{P}_n = \bigotimes_{j=1}^{n} \langle X, Z \rangle = \langle X_1, Z_1, X_2, Z_2, \cdots, X_n, Z_n \rangle = \{\pm 1, \pm i\} \times \{\bigotimes_{j=1}^{n} P_j | P_j \in \{\mathbb{I}, X, Y, Z\}\}, \tag{A3}$$

where $\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is the identity operator with dimension 2. X, Y, and Z are the single-qubit Pauli matrices given by

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A4}$$

X_j is the single-qubit Pauli X matrix acting on j-th qubit given by $(\bigotimes_{i=1}^{j-1} \mathbb{I}_i) \otimes X (\bigotimes_{i=j+1}^{n} \mathbb{I}_i)$. The notation $\langle \cdot \rangle$ denotes the group generated by $\cdot$. The elements in $\langle \rangle$ are called group generators. For instance, $\langle X_j, Z_j \rangle = \{X_j^{m_1} Z_j^{n_1} X_j^{m_2} Z_j^{n_2} \cdots | \forall i, m_i, n_i \in \mathbb{Z}\} = \{\pm 1, \pm i\} \times \{\mathbb{I}_j, X_j, Y_j, Z_j\}$. In the main text and below, for brevity we sometimes use notation $\langle X, Z \rangle^{\otimes n}$ or even $\langle X, Z \rangle$ to represent $\langle X_1, Z_1, X_2, Z_2, \cdots, X_n, Z_n \rangle$ where in this case, X and Z denote Pauli X and Pauli Z gates on all qubits, respectively. We would always omit subscripts, representing which qubits the gates act on, in $\langle \rangle$ when the group generators contain the same gates acting on all qubits.

In literature, people always use another definition for the Pauli group, which is, in fact, the projective Pauli group. Projective Pauli group is the quotient of the Pauli group by its center $\{\pm 1, \pm i\}$, $\langle X_1, Z_1, X_2, Z_2, \cdots, X_n, Z_n \rangle / \{\pm 1, \pm i\}$. Thus, we normally use $\otimes_{j=1}^{n} \{\mathbb{I}, X, Y, Z\}$ to represent it. As in quantum systems, the overall phase is not important. The quotient by the phase $\{\pm 1, \pm i\}$ does not influence the quantum state. Below we simply use the Pauli group to denote the projective Pauli group and use the definition below.

$$\mathsf{P}_n = \bigotimes_{j=1}^{n} \{\mathbb{I}, X, Y, Z\}. \tag{A5}$$

The normalized Pauli group is obtained by normalizing the elements of the n-qubit Pauli group by a factor of $1/\sqrt{2^n}$ as shown below.

$$\mathsf{P}'_n = \{\sigma_i = \frac{1}{\sqrt{2^n}} P_i | P_i \in \mathsf{P}_n\}. \tag{A6}$$

These normalized Pauli operators are complete and satisfy the orthonormality under the Hilbert-Schmidt inner product,

$$\mathrm{tr}\left(\sigma_i^\dagger \sigma_j\right) = \delta_{ij}, \tag{A7}$$

where δ_{ij} is the Kronecker delta symbol. Thus, we can represent the quantum state and the quantum channel with normalized Pauli operators as a set of bases.

Any n-qubit operator O in $\mathcal{L}(\mathcal{H})$ can be decomposed with the 4^n normalized Pauli operators,

$$O = \sum_{\sigma_i \in \mathsf{P}'_n} \mathrm{tr}\left(\sigma_i^\dagger O\right) \sigma_i. \tag{A8}$$

Thus, we can use the expansion coefficients $\mathrm{tr}\left(\sigma_i^\dagger O\right)$ to represent O and in Pauli-Liouville representation, we express Eq. (A8) as

$$|O\rangle\rangle = \sum_{\sigma_i \in \mathsf{P}'_n} \mathrm{tr}\left(\sigma_i^\dagger O\right) |\sigma_i\rangle\rangle, \tag{A9}$$

where $|\sigma_i\rangle\rangle$ in Pauli-Liouville representation is a length-4^n vector with only one non-zero element 1.

Furthermore, any quantum channel Λ can be represented as a matrix in the Liouville representation. The action of the channel on an operator O is given by

$$|\Lambda(O)\rangle\rangle = \Lambda|O\rangle\rangle, \tag{A10}$$

and the elements of the Liouville representation of Λ are given by

$$\Lambda_{ij} = \langle\langle\sigma_i|\Lambda|\sigma_j\rangle\rangle = \mathrm{tr}(\sigma_i\Lambda(\sigma_j)). \tag{A11}$$

Specifically, the diagonal terms of Pauli-Liouville representation are named Pauli fidelity [15], defined as below.

$$\lambda_i = \langle\langle\sigma_i|\Lambda|\sigma_i\rangle\rangle = \frac{1}{d}\,\mathrm{tr}(P_i\Lambda(P_i)), \tag{A12}$$

where P_i is a Pauli operator, and d is the dimension of the Hilbert space.

In the Liouville representation, the concatenation of two channels can be represented as the product of their corresponding matrices,

$$|\Lambda_2 \circ \Lambda_1(O)\rangle\rangle = \Lambda_2|\Lambda_1(O)\rangle\rangle = \Lambda_2\Lambda_1|O\rangle\rangle. \tag{A13}$$

The Liouville representation also allows us to vectorize the measurement operators using the Liouville bra-notation. The measurement probability of a state ρ using a positive operator-valued measure (POVM) F_i is given by

$$p_i = \langle\langle F_i|\rho\rangle\rangle = \mathrm{tr}\big(F_i^\dagger\rho\big), \tag{A14}$$

where $\langle\langle F_i|$ is the Liouville bra-notation of F_i and is equal to the conjugate transpose of $|F_i\rangle\rangle$.

We also briefly introduce another representation called the χ-matrix representation for quantum channel Λ. Given an n-qubit state, ρ, $\Lambda(\rho)$ can be expanded as

$$\Lambda(\rho) = d\sum_{i,j}\chi_{ij}\sigma_i\rho\sigma_j^\dagger, \tag{A15}$$

where the process matrix χ is uniquely determined by the orthonormal Pauli operator basis σ_j and $d = 2^n$ represents the dimension of the quantum system. Note that $\sigma_0 = \frac{\mathbb{I}_d}{\sqrt{d}}$. If a channel is diagonal in this representation, it is called a Pauli channel. It is easy to verify that a Pauli channel is also diagonal in the Pauli-Liouville representation. The diagonal terms of Pauli-Liouville representation and that of χ-matrix representation are related by Walsh-Hadamard transformation as shown in the following equation.

$$\lambda_j = \sum_i(-1)^{\langle i,j\rangle}\chi_{ii}, \tag{A16}$$

where $\langle i,j\rangle = 0$ if P_i commutes with P_j, and $\langle i,j\rangle = 1$ otherwise. The inverse Walsh-Hadamard transformation is given by,

$$\chi_{jj} = \frac{1}{d^2}\sum_i(-1)^{\langle i,j\rangle}\lambda_i. \tag{A17}$$

2. Quantum channel fidelity

To facilitate the understanding of randomized benchmarking (RB), we introduce the concepts of process fidelity and average fidelity. These two fidelities are equivalent to each other by linear transformation. The task of RB is estimating the fidelity of a given quantum gate or a given gate set.

The process fidelity of a channel, Λ, can be defined with its χ-matrix representation as follows.

$$F(\Lambda) = \chi_{00}(\Lambda). \tag{A18}$$

The process fidelity can be obtained from Eq. (A17) as:

$$F(\Lambda) = \chi_{00}(\Lambda) = \frac{1}{d^2}\sum_j\lambda_j = \frac{1}{d^2}\,\mathrm{tr}(\Lambda), \tag{A19}$$

which is equal to the trace of Λ in Liouville representation divided by d^2. Eq. (A19) provides an alternative definition of process fidelity.

There exists a relation between the commonly-used average fidelity F_{ave} and the process fidelity [49],

$$F_{ave} = \frac{dF + 1}{d + 1}. \tag{A20}$$

The average fidelity is defined as:

$$F_{ave} = \int d\psi \, \mathrm{tr}(|\psi\rangle\langle\psi| \, \Lambda(|\psi\rangle\langle\psi|)), \tag{A21}$$

where the integral is taken over the Haar measure, it means the average fidelity of the ideal final state and the realistic final state over all pure states. Both the process fidelity and the average fidelity are well-defined metrics for quantifying the closeness of a quantum channel to the identity. Below, we refer to the fidelity to the process fidelity without further explanation. Also, note that in reality, a quantum gate, U, is noisy, and its noisy version can be expressed as the composite channel of $\mathcal{U}$ and its noise channel Λ. Then, the fidelity of U refers to the fidelity of its noise channel Λ.

3. Group twirling and representation theory

Representation theory is an effective tool for analyzing abstract groups and is especially useful in randomized benchmarking and randomized compiling. Below, we briefly introduce basic concepts in representation theory and refer to [50, 51] for systematic introduction. Let G be a finite group and $G \in \mathsf{G}$ be an element of the group. The representation of G is defined as follows.

Definition 1 (Group representation)**.** *A map, ϕ, is called a representation of the group G on a linear space, V, if it is a group homomorphism mapping from G to $GL(V)$, where*

$$\begin{aligned} &\phi : \mathsf{G} \to GL(V), \\ &g \mapsto \phi(G), \ \ \forall G \in \mathsf{G}. \end{aligned} \tag{A22}$$

Here, $GL(V)$ denotes the general linear group of V. The representation ϕ satisfies the condition of preserving multiplication that for all $G_1, G_2 \in \mathsf{G}$,

$$\phi(G_1)\phi(G_2) = \phi(G_1 G_2). \tag{A23}$$

Intuitively, representation is just using a matrix group to represent G. Below, we introduce the concept of irreducible representation. Given a representation, ϕ, on V, a linear subspace, $W \subseteq V$, is referred to as invariant if for all $w \in W$ and for all $G \in \mathsf{G}$,

$$\phi(G)w \in W. \tag{A24}$$

The restriction of ϕ to the invariant subspace W is called the subrepresentation of G on W. Note that any representation has a subrepresentation mapping all elements in G to 1 where $W = \{0\}$ and has itself as a subrepresentation where $W = V$. These two subrepresentations are both trivial. With the concept of subrepresentation, we define the irreducible representation as below.

Definition 2 (Irreducible representation)**.** *A representation ϕ of the group G on the linear space V is said to be irreducible if it only possesses trivial subrepresentations.*

Maschke's theorem provides an interesting property stating that every representation ϕ of a finite group, G, can be decomposed into the direct sum of irreducible representations. For all $G \in \mathsf{G}$, the decomposition can be expressed as follows.

$$\phi(G) \simeq \bigoplus_{\phi_i \in R_{\mathsf{G}}} \mathbb{I}_{n_i \times n_i} \otimes \phi_i(G), \tag{A25}$$

Here, R_{G} denotes the set of all irreducible representations, and n_i represents the multiplicity of the equivalent irreducible representations of ϕ_i in ϕ.

The trace of a representation is called a character, which is defined below.

Definition 3 (Character function). *Let ϕ be a representation over the group G. The character of ϕ is defined as the function $\chi_\phi : \mathsf{G} \to \mathbb{C}$ such that for every $G \in \mathsf{G}$,*

$$\chi_\phi(G) = \mathrm{tr}[\phi(G)]. \tag{A26}$$

Character function is a powerful tool in representation theory. We can define the inner product of two character functions as below.

Definition 4 (Inner product of character function). *Let χ_1 and χ_2 be two character functions of a finite group, G, their inner product is defined as*

$$\langle \chi_1 | \chi_2 \rangle = \frac{1}{|\mathsf{G}|} \sum_{G \in \mathsf{G}} \chi_1^*(G) \chi_2(G). \tag{A27}$$

There is a useful result for the inner product between two characters when χ_1 and χ_2 are both irreducible representations. Then $\langle \chi_1 | \chi_2 \rangle = 1$ if χ_1 and χ_2 are equivalent and $\langle \chi_1 | \chi_2 \rangle = 0$ otherwise. Note that Eq. (A25) tells us that

$$\chi_\phi(G) = \mathrm{tr}(\phi(G)) = \sum_{\phi_i \in R_\mathsf{G}} n_i \chi_i(G), \tag{A28}$$

where $\chi_i(G) = \mathrm{tr}(\phi_i(G))$. With the orthonormality of irreducible representations, we obtain $n_i = \langle \chi_i | \chi_\phi \rangle$. Then, we have the following lemma.

Lemma 3 (Multiplicity of irreducible representations). *For any irreducible representation ϕ_i with character χ_i, the multiplicity of ϕ_i in ϕ is given by*

$$n_i = \langle \chi_j | \chi_\phi \rangle . \tag{A29}$$

We also introduce a concept named multiplicity-free representation for further elaboration.

Definition 5 (Multiplicity-free representation). *Given a representation, ϕ, of group G, if for any irreducible representation, ϕ_i, the multiplicity of ϕ_i in ϕ is 1, we call ϕ a multiplicity-free representation.*

Using the character function, we can also obtain the generalized projection formula utilized in character randomized benchmarking [37].

Lemma 4 (Generalized projection formula). *Consider a finite group, G, and its representation ϕ. Let ϕ_i be an irreducible representation contained in ϕ with character χ_i. The projector onto the support space of ϕ_i can be expressed as:*

$$\Pi_i = \frac{d_i}{|\mathsf{G}|} \sum_{G \in \mathsf{G}} \chi_i(G) \phi(G), \tag{A30}$$

where $d_i = \dim \phi_i$ denotes the dimension of ϕ_i.

The representation theory is also useful for us to analyze the group twirling on a channel, Λ, over a group, G.

Definition 6 (Group Twirling). *For a representation ϕ of the group G on the linear space V, the twirling of a linear map, $\Lambda : V \to V$ over G is defined as:*

$$\Lambda_\mathsf{G} = \frac{1}{|\mathsf{G}|} \sum_{G \in \mathsf{G}} \phi(G)^\dagger \Lambda \phi(G). \tag{A31}$$

To analyze the result of group twirling, we introduce Schur's lemma, which is essential in our further elaboration.

Lemma 5 (Schur's lemma). *Let $\phi_1 : \mathsf{G} \to GL(V_1)$ and $\phi_2 : \mathsf{G} \to GL(V_2)$ be two arbitrary irreducible representations of group G and $A : V_1 \to V_2$ be a linear map from V_1 to V_2. Suppose for any $G \in \mathsf{G}$,*

$$A\phi_1(G) = \phi_2(G)A. \tag{A32}$$

Then, A equals 0, mapping all vectors in V_1 to zero vector in V_2, if ϕ_1 and ϕ_2 are inequivalent irreducible representations; A is proportional to the identity operator if ϕ_1 and ϕ_2 are equivalent irreducible representations. Note that when ϕ_1 and ϕ_2 are equivalent, V_1 and V_2 are equivalent linear space, so the identity operator is well-defined.

Using Schur's Lemma, we can establish the following proposition.

Proposition 1. *Let $\phi_1 : \mathsf{G} \to GL(V_1)$ and $\phi_2 : \mathsf{G} \to GL(V_2)$ be two arbitrary irreducible representations of group G and $A : V_1 \to V_2$ be an arbitrary linear map from V_1 to V_2. Define*

$$A_G = \mathbb{E}_{G\in\mathsf{G}}\phi_2^\dagger(G)A\phi_1(G). \tag{A33}$$

Then,

$$A_G = 0, \tag{A34}$$

if ϕ_1 and ϕ_2 are inequivalent, and

$$A_G = \frac{\mathrm{tr}(A)}{\dim V_1}\mathbb{I}, \tag{A35}$$

if ϕ_1 and ϕ_2 are equivalent, where $\mathbb{I}$ is the identity operator mapping from V_1 to V_2.

For further elaboration and proof, we provide the Burnside theorem in the representation theory.

Lemma 6. *Consider a finite group, G, and its all inequivalent irreducible representations $R_G = \{\phi_i\}$, then the cardinality of the twirling group, $|\mathsf{G}|$ can be given by*

$$|\mathsf{G}| = \sum_{\phi_i\in R_G} \dim \phi_i^2. \tag{A36}$$

4. Randomized benchmarking and diagonalizability of twirled noise channel

Below, we introduce randomized benchmarking (RB) and character randomized benchmarking [37]. Let us start with a brief review of the standard RB, which aims to estimate the fidelity of a quantum gate group. We consider an n-qubit gate group, G, where each gate $G \in \mathsf{G}$ is assumed to be with a noise channel Λ in implementation. In reality, the noise channels for different gates can be different. Below, we simply consider the first-order approximation [7], which is valid and useful in experiments, and regard the noise channels for different gates in G as the same. One can express the noisy quantum gate $\widetilde{\mathcal{G}} = \Lambda\mathcal{G}$ as the composite of the noiseless gate G and its noise channel Λ for any gate $G \in \mathsf{G}$. Then, the task of RB is estimating the fidelity of Λ robust to SPAM error.

To achieve that, one always performs random gate sequences composed of m random gates from G and an inverse quantum gate where m is a prefixed integer called circuit depth,

$$\begin{aligned}\widetilde{S} &= \widetilde{\mathcal{G}}_{inv}\prod_{i=1}^{m}\widetilde{\mathcal{G}}_i \\ &= \Lambda\mathcal{G}_{inv}\prod_{i=1}^{m}\Lambda\mathcal{G}_i,\end{aligned} \tag{A37}$$

where $G_{inv} = (\prod_{i=1}^m G_i)^\dagger$, $\mathcal{G}_i$ denotes the Liouville representation of G_i, and $\widetilde{\cdot}$ means the quantum gate is noisy.

In the sense of expectation, the random gate sequence $\widetilde{S}$ is equal to

$$\begin{aligned}\mathbb{E}_{\forall i,G_i\in\mathsf{G}}\widetilde{S} &= \Lambda\mathbb{E}_{\forall i,G_i\in\mathsf{G}}\prod_{i=1}^{m}(\prod_{j=1}^{i}\mathcal{G}_j)^\dagger\Lambda(\prod_{j=1}^{i}\mathcal{G}_j) \\ &= \Lambda(\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}^\dagger\Lambda\mathcal{G})^m \\ &= \Lambda\Lambda_G^m.\end{aligned} \tag{A38}$$

The second line utilizes the fact that G is a group. Thus, $\prod_{j=1}^i \mathcal{G}_j$ independently and identically satisfy the uniform distribution on G for any $1 \le j \le m$.

If we use the same state preparation and measurement for all random sequences, then in the expectation, we can get

$$\begin{aligned}f(m) &= \mathbb{E}_{\forall i,G_i\in\mathsf{G}}\langle\langle M|\widetilde{S}|\rho\rangle\rangle \\ &= \langle\langle M|\Lambda\Lambda_G^m|\rho\rangle\rangle \\ &= \langle\langle M'|\Lambda_G^m|\rho\rangle\rangle,\end{aligned} \tag{A39}$$

where ρ and M are prefixed initial state and measurement, respectively, and M' is the measurement absorbing the noise of the inverse gate, satisfying $\langle\langle M'| = \langle\langle M|\Lambda$.

The next step is obtaining the trace of Λ_G robust to state preparation and measurement (SPAM) error. Note that the fidelity $F(\Lambda) = \frac{1}{d^2}\operatorname{tr}(\Lambda) = \frac{1}{d^2}\operatorname{tr}(\Lambda_G)$. Normally, Λ_G has few parameters as it has been twirled and becomes symmetric. These parameters can be evaluated robustly to SPAM error via obtaining $f(m)$ with different circuit depths m and employing exponential fitting. Then, the trace of Λ_G, or the fidelity, can be evaluated with the parameters of Λ_G.

Before the work of character RB [37], researchers mainly focus on a large group, G, like the Clifford group, to make Λ_G highly symmetric with few parameters. For Clifford RB [6, 7], Λ_G is a depolarizing channel with a single parameter p satisfying $\Lambda_G(\rho) = p\rho + (1-p)\frac{\mathbb{I}_d}{d}$. In Pauli-Liouville representation, $\Lambda_G = |\sigma_0\rangle\rangle\langle\langle\sigma_0| + p\sum_{i\geq 1}|\sigma_i\rangle\rangle\langle\langle\sigma_i|$. Then, $f(m)$ has a single exponential decay expression

$$f(m) = Ap^m + B, \tag{A40}$$

where p is only relevant to Λ, and A and B are only relevant to SPAM. Thus, via single exponential fitting, one can get p along with the trace of Λ_G and the fidelity of Λ.

For CNOT dihedral RB [16], $\Lambda_G = \Pi_0 + p_Z\Pi_Z + p_X\Pi_X$ has two undetermined parameters, p_Z and p_X where $\Pi_0 = |\sigma_0\rangle\rangle\langle\langle\sigma_0|$, $\Pi_Z = \sum_{\sigma_z\in\{\frac{\mathbb{I}}{\sqrt{2}},\frac{Z}{\sqrt{2}}\}^{\otimes n}/\sigma_0}|\sigma_z\rangle\rangle\langle\langle\sigma_z|$, and $\Pi_X = \sum_{\sigma_x\in\mathsf{P}'_n/\{\frac{\mathbb{I}}{\sqrt{2}},\frac{Z}{\sqrt{2}}\}^{\otimes n}}|\sigma_x\rangle\rangle\langle\langle\sigma_x|$, then $f(m)$ is a double exponential decay function with parameters p_Z and p_X. One has to employ double exponential fitting to obtain $\operatorname{tr}(\Lambda_G)$ and its fidelity.

For a general group G, $f(m)$ is complex and generally not an exponential decay function [36]. But if Λ_G can be written as

$$\Lambda_G = \sum_i p_i\Pi_i, \tag{A41}$$

where $\{\Pi_i\}$ are mutually orthogonal projectors only dependent on G in the space of Liouville representation, we can obtain all decay parameters p_i with only single exponential fitting via the technique of character RB [37]. Note that Eq. (A41) can also be interpreted as that Λ_G is diagonal in the Pauli-Liouville representation up to a unitary transformation $\mathcal{T}$,

$$\Lambda_G = \mathcal{T}(\sum_i p_i\Pi_i^P)\mathcal{T}^\dagger, \tag{A42}$$

where Π_i^P is a projector, equaling the summation of the Pauli operator bases. The projector Π_i^P and the unitary transformation $\mathcal{T}$ are independent of the channel Λ and are only related to group G. We consider the case that $\mathcal{T}$ is a Liouville representation of a unitary gate T.

Note that in this case,

$$\begin{aligned} f(m) &= \langle\langle M'|\Lambda_G^m|\rho\rangle\rangle \\ &= \sum_i \langle\langle M'|\Pi_i|\rho\rangle\rangle p_i^m, \end{aligned} \tag{A43}$$

is a multiple exponential decay function. If we directly fit $f(m)$ with a multiple exponential decay function, $\sum_i A_i p_i^m$, the fitting process would consume massive computational resources, and the result is normally inaccurate. The key step of character RB is utilizing Lemma 4. Instead of implementing the gate sequence in Eq. (A37), in character RB, we select a projector Π_j from $\{\Pi_i\}$ and implement

$$\begin{aligned} \widetilde{S}' &= \widetilde{\mathcal{G}}_{inv}(\prod_{i=1}^m \widetilde{\mathcal{G}}_i)\widetilde{\mathcal{G}}' \\ &= \Lambda\mathcal{G}_{inv}(\prod_{i=1}^m \Lambda\mathcal{G}_i)\mathcal{G}', \end{aligned} \tag{A44}$$

where G' is named character gate and is independently sampled from a predetermined gate set, named character group, G'. We select a character function χ' associated with an irreducible representation ϕ' of G' and obtain a projector according to Lemma 4,

$$\Pi' = \frac{\dim\phi'}{|\mathsf{G}'|}\sum_{G'\in\mathsf{G}'}\chi'(G')\mathcal{G}', \tag{A45}$$

such that for any projector $\Pi_i \in \{\Pi_i\}$,

$$\Pi'\Pi_i = \delta_{ij}\Pi'. \tag{A46}$$

Meanwhile, we consider $\langle\langle M|\widetilde{S}' \dim \phi' \chi'(G')|\rho\rangle\rangle$ and in the expectation, this quantity equals

$$\begin{aligned} f_j(m) &= \mathbb{E}_{\forall i, G_i \in \mathsf{G}, G' \in \mathsf{G}'} \langle\langle M|\widetilde{S}' \dim \phi' \chi'(G')|\rho\rangle\rangle \\ &= \langle\langle M|\Lambda\Lambda_G^m \Pi'|\rho\rangle\rangle \\ &= \langle\langle M'|\Lambda_G^m \Pi'|\rho\rangle\rangle \\ &= \sum_i p_i^m \langle\langle M'|\Pi_i\Pi'|\rho\rangle\rangle \\ &= \langle\langle M'|\Pi'|\rho\rangle\rangle p_j^m. \end{aligned} \tag{A47}$$

Thus, p_j can be obtained via single exponential fitting and due to the arbitrariness of Π_j, all parameters of Λ_G can be obtained via single exponential fitting, and hence the fidelity can be evaluated accurately.

Note that in character RB, the key point is realizing projector Π'. Below we show that we can always realize Π' satisfying Eq. (A46). First, each projector can be decomposed as

$$\Pi_j = \sum_{k=1}^{\mathrm{tr}(\Pi_j)} |\mu_{j_k}\rangle\rangle\langle\langle\mu_{j_k}|, \tag{A48}$$

where $\{\mu_{j_k}, 1 \le k \le \mathrm{tr}(\Pi_j), \forall \Pi_j \in \{\Pi_i\}\}$ forms an orthonormal operator basis in Liouville representation. As normalized Pauli operators P'_n is also an orthonormal basis, there exists a unitary transformation linking the two bases. Then, given Π_j from $\{\Pi_i\}$, we select $|\mu_{j_1}\rangle\rangle\langle\langle\mu_{j_1}|$, which is equal to $\mathcal{T}|\sigma\rangle\rangle\langle\langle\sigma|\mathcal{T}^\dagger = |T\sigma T^\dagger\rangle\rangle\langle\langle T\sigma T^\dagger|$ where $|\sigma\rangle\rangle\langle\langle\sigma|$ is a Pauli operator basis and T is the unitary transformation linking two bases. For Pauli group P_n, we have the following equation,

$$|\sigma\rangle\rangle\langle\langle\sigma| = \mathbb{E}_{P\in\mathsf{P}_n} \chi_\sigma(P)\mathcal{P}, \tag{A49}$$

where $\chi_\sigma = (-1)^{\langle P,\sigma\rangle}$ equals 1 when P and σ commute and -1 otherwise. Then, we only need to choose $\mathsf{G}' = T\mathsf{P}_n T^\dagger$ and $\chi' = \chi_\sigma$ to realize

$$\Pi' = \mathcal{T}|\sigma\rangle\rangle\langle\langle\sigma|\mathcal{T}^\dagger = \mathbb{E}_{P\in\mathsf{P}_n} \chi'(P)\mathcal{T}\mathcal{P}\mathcal{T}^\dagger. \tag{A50}$$

Then, Π' satisfies Eq. (A46).

In summary, as long as Λ_G has an expression of Eq. (A41) or Eq. (A42), then with the technique of character RB, one can obtain $\mathrm{tr}(\Lambda_G)$ and the fidelity of Λ accurately with only single exponential fitting. In Appendix B, we would prove that if the Liouville representation of G is not multiplicity-free, as defined in Definition 5, then Λ_G cannot be diagonal for *arbitrary* noise channel Λ. If the Liouville representation of G is multiplicity-free, then [37] has shown that the character group G' can be chosen as a subgroup of G. In this case, each projector Π_i in Eq. (A41) relates to an irreducible representation subspace of G. Denote this irreducible representation as ϕ_i with character χ_i, then Π_i can be realized with

$$\Pi_i = \frac{\dim \phi_i}{|\mathsf{G}|} \sum_{G'\in\mathsf{G}} \chi_i(G')\mathcal{G}'. \tag{A51}$$

Thus, implementing character gate G' is not harder than the twirling group.

Besides adding character gates, there is another method to effectively realize the projector Π_i in Eq. (A41). If one can realize a measurement M such that $\sum_i p_i \langle\langle M'|\Pi_i = p_j\langle\langle M'|\Pi_j$, then Eq. (A43) will also reduce to $p_j^m\langle\langle M'|\Pi_j|\rho\rangle\rangle$. An accurate initial state ρ, satisfying $\sum_i p_i\Pi_i|\rho\rangle\rangle = p_j\Pi_j|\rho\rangle\rangle$, can also achieve that. Thus, if we have some information about SPAM, we can realize the projection without the need to implement character gates.

5. Interleaved randomized benchmarking

Above, we only introduce how to evaluate the fidelity of a quantum gate group G via randomized benchmarking. In order to obtain the gate fidelity of an individual target gate, U, one needs to utilize the technique of interleaved RB [38]. In [38], the target gate U is embedded into a group, G. Below, we call them the twirled gate and the twirling group, respectively. To enable interleaved RB, one implements two kinds of circuits. The first is just a random gate

sequence in regular RB, extracting the average gate fidelity of the twirling group G. The second type of circuit is composed of random twirling gates from G interleaved with the target gate U, extracting the composition gate fidelity of the twirling group and the target gate. Comparing the two results, one can get the individual gate fidelity of the target gate. Specifically, suppose the noise channel for twirling group G is $\mathcal{E}$ and the noise channel for the target gate is Λ. Then with the RB methods introduced before one can obtain $F(\mathcal{E}) = \frac{\mathrm{tr}(\mathcal{E})}{d^2}$. After that, one implements an interleaved random gate sequence,

$$\begin{aligned}\widetilde{S}_i &= \widetilde{\mathcal{G}}_{inv}\prod_{i=1}^{m}\widetilde{\mathcal{U}}\widetilde{\mathcal{G}}_i \\ &= \mathcal{E}\mathcal{G}_{inv}\prod_{i=1}^{m}\mathcal{U}\Lambda\mathcal{E}\mathcal{G}_i,\end{aligned} \tag{A52}$$

where $G_{inv} = (\prod_{i=1}^{m} UG_i)^\dagger$, $\mathcal{G}_i$ and $\mathcal{U}$ denote the Liouville representation of G_i and U, respectively, and $\widetilde{\cdot}$ means the quantum gate is noisy. Note that $\widetilde{\mathcal{U}} = \mathcal{U}\Lambda$ while $\widetilde{\mathcal{G}} = \mathcal{E}\mathcal{G}$. The noise channels for U and gates in G are put in different positions, but due to the arbitrariness of Λ and $\mathcal{E}$, the difference does not put any restriction on the noise. In [38], the target gate U belongs to twirling group G, then under the expectation of sampling of G_i, Eq. (A52) is equal to,

$$\begin{aligned}\mathbb{E}_{\forall i, G_i\in\mathsf{G}}\widetilde{S}_i &= \mathcal{E}\mathbb{E}_{\forall i, G_i\in\mathsf{G}}\prod_{i=1}^{m}((\prod_{j=2}^{i}\mathcal{G}_j\mathcal{U})\mathcal{G}_1)^\dagger\Lambda\mathcal{E}((\prod_{j=2}^{i}\mathcal{G}_j\mathcal{U})\mathcal{G}_1) \\ &= \mathcal{E}(\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}^\dagger\Lambda\mathcal{E}\mathcal{G})^m \\ &= \mathcal{E}(\Lambda\mathcal{E})_G^m.\end{aligned} \tag{A53}$$

The second line utilizes the fact that G is a group. Thus, same as the discussion in the previous subsection, interleaved quantum circuits allow us to estimate the fidelity of $\Lambda\mathcal{E}$, that is, $F(\Lambda\mathcal{E}) = \frac{\mathrm{tr}(\Lambda\mathcal{E})}{d^2}$. In [38], the authors show how to estimate $F(\Lambda)$ from $F(\Lambda\mathcal{E})$ and $F(\mathcal{E})$. We review this technique as elaborated below.

Denote $\mathcal{E}_d$ to be the $\mathcal{E}$ twirled by the whole unitary group. It is known that $\mathcal{E}_d$ is a mixture of the identity channel and the depolarizing channel,

$$\mathcal{E}_d(\rho) = \frac{d^2F(\mathcal{E})-1}{d^2-1}\rho + (1-\frac{d^2F(\mathcal{E})-1}{d^2-1})\frac{\mathbb{I}}{d}. \tag{A54}$$

Now, let us investigate the quantity $|F(\Lambda\mathcal{E}) - F(\Lambda\mathcal{E}_d)|$, which is equal to

$$|F(\Lambda\mathcal{E}) - F(\Lambda\mathcal{E}_d)| = \left|F(\Lambda\mathcal{E}) - \frac{(d^2F(\Lambda)-1)(d^2F(\mathcal{E})-1)+d^2-1}{d^2(d^2-1)})\right|. \tag{A55}$$

Assuming $|F(\Lambda\mathcal{E}) - F(\Lambda\mathcal{E}_d)|$ is upper bounded by E, we can deduce $F(\Lambda)$ belongs to the following interval,

$$[\frac{d^2(F(\Lambda\mathcal{E})-E)-1}{d^2F(\mathcal{E})-1}(1-\frac{1}{d^2})+\frac{1}{d^2}, \frac{d^2(F(\Lambda\mathcal{E})+E)-1}{d^2F(\mathcal{E})-1}(1-\frac{1}{d^2})+\frac{1}{d^2}]. \tag{A56}$$

This interval has a mean value $\frac{d^2F(\Lambda\mathcal{E})-1}{d^2F(\mathcal{E})-1}(1-\frac{1}{d^2})+\frac{1}{d^2}$ and length $\frac{2(d^2-1)E}{d^2F(\mathcal{E})-1}$. With matrix analysis, the quantity E can be upper bounded by [38],

$$E \le \min\begin{cases}(4(d+1)\sqrt{1-F(\mathcal{E})}+2\frac{d+1}{d}(1-F(\mathcal{E})) \\ \frac{\left|d^2(F(\Lambda\mathcal{E})-F(\mathcal{E}))+2F(\mathcal{E})-F(\Lambda\mathcal{E})-1\right|+(d^2F(\mathcal{E})-1)(1-F(\mathcal{E}))}{d^2-1} \\ \frac{\left|d^2(F(\Lambda\mathcal{E})-F(\mathcal{E}))+2F(\mathcal{E})-F(\Lambda\mathcal{E})-1\right|+(d^2F(\mathcal{E})-1)(F(\mathcal{E})-F(\Lambda\mathcal{E}))}{d^2-1}\end{cases}. \tag{A57}$$

In general, the higher $F(\mathcal{E})$ is, the smaller E is, and the more accurate the estimation of $F(\Lambda)$ is. In [38], the target gate U belongs to G so the noisy levels, or the fidelities, of Λ and $\mathcal{E}$ are close, which is not beneficial for estimating $F(\Lambda)$. This is also one of the reasons that finding a small and practical twirling group G matters. In the main text, we focus on the interleaved circuit to estimate $F(\Lambda\mathcal{E})$, and $F(\Lambda)$ can be estimated via Eqs. (A56) and (A57). Note that though this estimation assumes the noise channel of the twirling group to be gate-independent, this assumption is good enough in a high-fidelity region and could also be relieved by using a higher-order fitting formula by more sophisticated analysis like Ref. [7].

6. Dihedral group and classically replaceable unitary operations

For further elaboration, in this part, we introduce the local dihedral group and classically replaceable unitary operations (CRU) [33] along with their properties. The local dihedral group on n qubits is similar to the n-qubit Pauli group, which is defined as below.

$$\mathsf{D}_n^m = \langle X, Z_m\rangle^{\otimes n} = \langle X_1, (Z_m)_1, X_2, (Z_m)_2, \cdots, X_n, (Z_m)_n\rangle, \tag{A58}$$

where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the Pauli X gate and $Z_m = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{2\pi}{m}} \end{pmatrix}$ is a phase gate with phase $\frac{2\pi}{m}$. Here, m is a positive integer. In general, the phase on different qubits can be different and we can define $\langle X_1, (Z_{m_1})_1, X_2, (Z_{m_2})_2, \cdots, X_n, (Z_{m_n})_n\rangle$, but below we only consider a simple case that $m_1 = m_2 = \cdots = m_n = m$. Same with the discussion of the Pauli group, we can define the projective local dihedral group via quotient by the center of $\langle X, Z_m\rangle^{\otimes n}$,

$$\mathsf{D}_m^m = \langle X, Z_m\rangle^{\otimes n}/\langle\omega_m\rangle, \tag{A59}$$

where $\omega_m = e^{i\frac{2\pi}{m}}$ and $\langle\omega_m\rangle = \{e^{i\frac{2k\pi}{m}}, 0 \le k \le m-1\}$. As the overall phase is not important, below we do not distinguish the local dihedral group and the projective local dihedral group and use the definition $\mathsf{D}_n^m = \langle X, Z_m\rangle^{\otimes n}/\langle\omega_m\rangle$.

Classically replaceable unitary operations, or incoherent unitary operations, are all unitary gates that can be moved after computational basis measurements and be replaced with classical post-processing. Given the computational basis $|i\rangle$ on the Hilbert space $\mathcal{H}$, we can define the dephasing operation Δ. For any $\rho \in \mathcal{D}(\mathcal{H})$,

$$\Delta(\rho) = \sum_i |i\rangle\langle i|\,\rho\,|i\rangle\langle i|\,. \tag{A60}$$

Then, the CRU gate set has an expression,

$$\{U \text{ is unitary} | \mathcal{U}\Delta = \Delta\mathcal{U}\}. \tag{A61}$$

Note that we use the same notation Δ to represent its map representation and Liouville representation. It is shown that a CRU can be given by [33, 52]

$$U = \sum_j e^{i\theta_j}\,|\sigma(j)\rangle\langle j|\,, \tag{A62}$$

where σ is a permutation over computational basis and θ_j is a phase in $[0, 2\pi)$. And any unitary gate having an expression of (A62) is a CRU. We can decompose U as

$$U = \sum_j |\sigma(j)\rangle\langle j| \sum_j e^{i\theta_j}\,|j\rangle\langle j|\,. \tag{A63}$$

Set $\Pi = \sum_j |\sigma(j)\rangle\langle j|$ and $W = \sum_j e^{i\theta_j}\,|j\rangle\langle j|$ and we can see any CRU can be written as the multiplication of a permutation matrix, Π, and a diagonal matrix, W, $U = \Pi W$. Note that any matrix with the form $\sum_j |\sigma(j)\rangle\langle j|$ for arbitrary permutation σ is defined to be a permutation matrix. Also, the multiplication of a permutation matrix and a diagonal matrix is always expressed as Eq. (A62) and is a CRU.

Lemma 7. *Any classically replaceable unitary operation U can be decomposed as the multiplication of a permutation matrix, Π, and a diagonal matrix, W, $U = \Pi W$. Vice versa.*

From the above lemma, we can obtain that the computational basis gate set is invariant under the action of a CRU.

Lemma 8. *Diagonal matrices set or computational basis gate set is invariant under the action of any CRU.*

Proof. For any CRU $U = \sum_j e^{i\theta_j}\,|\sigma(j)\rangle\langle j|$ and diagonal matrix $W = \sum_j e^{\phi_j}\,|j\rangle\langle j|$,

$$\begin{aligned} UWU^{-1} &= \sum_{jj'k} e^{i(\phi_k+\theta_j-\theta_{j'})}\,|\sigma(j)\rangle\,\langle j|k\rangle\,\langle k|j'\rangle\,\langle\sigma(j')| \\ &= \sum_j e^{i\phi_j}\,|\sigma(j)\rangle\langle\sigma(j)| \\ &= \sum_j e^{i\phi_{\sigma^{-1}(j)}}\,|j\rangle\langle j|\,, \end{aligned} \tag{A64}$$

which is a diagonal matrix. Proof is done. □

In fact, for any CRU subgroup $\mathsf{G} \leq \mathrm{CRU}$, all diagonal matrices in G forms a subgroup, G_Z of G. Then, G_Z is invariant under the action of any gate in G, or equivalently, G_Z is a normal subgroup of G.

Lemma 9. *Given an n-qubit CRU subgroup* G*, set the computational basis subgroup of* G *as* $\mathsf{G}_Z = \{U \in \mathsf{G} | U \text{ is diagonal}\}$*. Then,* G_Z *is a normal subgroup of* G*.*

Proof. Obviously, G_Z is a subgroup of G. We only need to prove that under the conjugate action of any quantum gate G in G, G_Z is invariant. For any gate $G \in \mathsf{G}$, $G \in \mathrm{CRU}$, then $G\mathsf{G}_Z G^\dagger$ only contains diagonal matrices. As all diagonal matrices in G are contained in G_Z, $G\mathsf{G}_Z G^\dagger \subseteq \mathsf{G}_Z$. Also, $\left|G\mathsf{G}_Z G^\dagger\right| = |\mathsf{G}_Z|$, so $G\mathsf{G}_Z G^\dagger = \mathsf{G}_Z$. As G is an arbitrary gate in G, we prove that G_Z is the normal subgroup of G. Here, we complete the proof. □

Below, we present a result about permutation matrices for further elaboration. The lemma tells us the structure of the permutation matrices, which can always be decomposed as the multiplication of Toffoli gates, CNOT gates, and Pauli X gates.

Lemma 10. *All permutation matrices on n qubits can be generated by Pauli X gate on each qubit and all Toffoli gates* $C^{n-1}X$ *with* $n-1$ *control qubits and 1 target qubit.*

Proof. The expression of permutation matrices on n qubits can be unified as below.

$$\Pi_\sigma = \sum_{j \in \{0,1\}^n} |\sigma(j)\rangle\langle j| , \tag{A65}$$

where σ is an arbitrary permutation on $\{0,1\}^n$. The permutations themself form a permutation group with cardinality $2^n!$. As transpositions can generate the permutation group, we only need to prove that any transposition (s_1, s_2) can be generated by Pauli X gates and $C^{n-1}X$ gates for $s_1 \neq s_2 \in \{0,1\}^n$.

Denote X_k to be the Pauli X gate acting on k-th qubit and set $X^s = \bigotimes_{i=1}^n X_i^{s_i}$ for $s \in \{0,1\}^n$. We also denote $C^{n-1}X_k$ to be the Toffoli gate with k-th qubit as the target qubit and other qubits as the control qubits. Then, for any $s \in \{0,1\}^n$, $X^s C^{n-1}X_k (X^s)^\dagger$ would swap the basis $|1^n \oplus s\rangle$ and $|1^n \oplus 1_k \oplus s\rangle$ while keeping other bases fixed. Here, $1^n = 11\cdots1$ is the all-1 bit string and $1_k = 0^{k-1}10^{n-k}$ is the bit string with 1 in k-th position and 0 in other positions. Thus, by taking k over 1 to n and s over $\{0,1\}^n$ we would obtain any transposition $(s, s \oplus 1_k)$. As $(s, s \oplus 1_{k1})(s, s \oplus 1_{k2})(s, s \oplus 1_{k1}) = (s \oplus 1_{k1}, s \oplus 1_{k2})$, we would also obtain any transposition $(s, s \oplus 1_{k1} \oplus 1_{k2} \ldots \oplus 1_{kl})$, which suffices to produce any transposition (s_1, s_2) for $s_1 \neq s_2 \in \{0,1\}^n$.

Note that the Toffoli gates with less than $n-1$ control qubits are also permutation matrices. It means that with single qubit Pauli X gates and $C^{n-1}X$ gates, one can construct any controlled-X gates C^kX with $1 \leq k \leq n-2$. □

Appendix B: Optimal twirling groups for multi-qubit controlled phase gates in randomized benchmarking

In original interleaved RB [38], to benchmark an individual target quantum gate, U, one always embeds U in a large group, G, with a number of global quantum gates so that G has a strong twirling effect. However, the large size of the twirling group would make group sampling and computing inverse gates difficult, and plenty of global quantum gates in G would make the twirling gates hard to realize. The former difficulty is a classical computational problem, and the latter is a gate implementability problem. In this work, we focus on finding the smallest and the most easily implementable twirling group for a target gate.

Fortunately, [14, 15, 37] shows that embedding the target gate U in the twirling group G is not necessary to characterize U. A Clifford gate can be effectively characterized with local Clifford twirling or Pauli twirling instead of global Clifford twirling. It gives hope that maybe we can choose a small twirling group for benchmarking a generic target quantum gate. Normally, we would like to choose a twirling group as small as possible, but the twirling group can not be arbitrarily small as well. Otherwise, the twirled noise channel is not symmetric enough, resulting in hard post-processing and inaccurate fidelity estimation. In the main text, we have briefly introduced the requirements for twirling groups in RB and proposed Question 1. Below, we will present the requirements for twirling groups in RB more detailedly and completely and rederive Question 1. Moreover, in this part, we will provide the formal versions of the theorems and lemmas in the main text along with their proof. It is worth mentioning that when we write 'a channel is diagonal', it always means that 'the channel is diagonal up to a unitary transformation'.

1. Requirements for twirling groups in randomized benchmarking

Below, we focus on how to estimate the fidelity of a target gate, U, with methods of RB. Same with the notations in the main text and in Appendix A, we express the noisy quantum gate $\widetilde{\mathcal{U}} = \mathcal{U}\Lambda$ as a composite of the noiseless gate

U and its noise channel Λ, where $\mathcal{U}$ denotes the Pauli-Liouville representation of U, and $\widetilde{\cdot}$ represents the noisy version of a quantum gate. As mentioned in Appendix A 4, the key to enable RB is obtaining the powers of the G-twirled noise channel, Λ_G^m, where $m \in \mathbb{Z}+$, and ensuring Λ_G is diagonal up to a unitary transformation, or can be written as Eq. (A42). Then, with the technique of character RB, one can obtain $\mathrm{tr}(\Lambda_G)$ and the fidelity of Λ. Here, G is the twirling group for tailoring Λ.

Different from the elaboration in the main text, below we distinguish two concepts of twirling gate set V and twirling group G. In the main text, we directly select a twirling group, G, and implement a random gate sequence composed of gates from G interleaved with target gate U to obtain Λ_G^m. But in fact, we have another way to obtain Λ_G^m as shown below.

To obtain Λ_G^m, we implement m twirling gates $V_i = G_i U G_{i-1}^\dagger U^\dagger$, sampled from the twirling gate set $\mathsf{V} = \mathsf{G}U\mathsf{G}U^\dagger$, interleaved with the target gate U as shown in Fig. 6. Here, G_i is uniformly and randomly sampled from the group G, and we set $G_0 = \mathbb{I}$. The circuit ends with the inverse gate $V_{\mathrm{inv}} = (\prod_{i=1}^m UV_i)^\dagger = U^{\dagger m-1} G_m^\dagger U^\dagger$. The circuit ideally equals the identity, corresponding to $\Lambda_G = I$. Note that the twirling group G and the twirling gate set V are in general different and $\mathsf{G} \subseteq \mathsf{V}$. Although we only require Λ to be twirled by G, we need to realize gates in V to achieve that since the circuit involves gate U. The twirling gates should first eliminate the effect of the twirled gate U and then influence the noise channel Λ. Note that the inverse gate V_{inv} can always belong to V as long as we choose m such that $U^m = \mathbb{I}$. The cost to implement the inverse gate is nearly the same as other twirling gates.

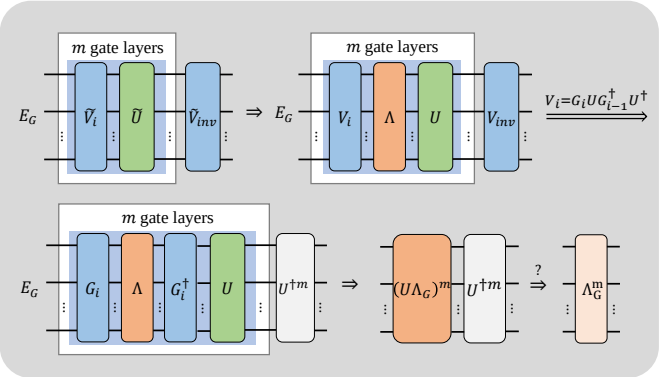


FIG. 6. Random twirling gates $V_1, V_2, \cdots, V_m$ interleaved with U in RB. The inverse gate $V_{\mathrm{inv}} = (\prod_{i=1}^m UV_i)^\dagger$. $\widetilde{\cdot}$ represents the noisy version of a quantum gate. Λ is the composite noise channel of V and U, and Λ_G is the G-twirled noise channel.

Same with the main text and the arguments in Appendix A 5, we omit the noise from the twirling gates. Then, the circuit is represented as

$$\begin{aligned}\widetilde{\mathcal{S}}_m &= \mathcal{V}_{\mathrm{inv}} \prod_{i=1}^m \mathcal{U}\Lambda\mathcal{V}_i \\ &= \mathcal{U}^{\dagger m-1}\mathcal{G}_m^\dagger \mathcal{U}^\dagger \prod_{i=1}^m \mathcal{U}\Lambda\mathcal{G}_i\mathcal{U}\mathcal{G}_{i-1}^\dagger\mathcal{U}^\dagger \\ &= \mathcal{U}^{\dagger m} \prod_{i=1}^m \mathcal{U}\mathcal{G}_i^\dagger\Lambda\mathcal{G}_i.\end{aligned} \tag{B1}$$

In terms of expectation over sampling of G_i, Eq. (B1) becomes $\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m$. Compared to Λ_G^m, this formulation only requires the commutation relation,

$$\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G, \tag{B2}$$

which means that the symmetry of the group G is preserved under the action of U in the Liouville representation. Thus, to tailor U in RB, our task is finding the twirling group G satisfying Eq.(B2) and ensuring the diagonalizability of Λ_G while minimizing the size of the twirling gate set $\mathsf{V} = \mathsf{G}U\mathsf{G}U^\dagger$. The question is summarized below.

Question 3. *Given a gate, U, find V such that $\mathsf{V} = \mathsf{G}U\mathsf{G}U^\dagger$ and the twirling group G satisfies*

$$\begin{aligned}&\textit{for any quantum channel } \Lambda, \Lambda_G = \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \textit{ is diagonal} \\ &\textit{up to a unitary transformation independent of } \Lambda,\end{aligned} \tag{B3}$$

and,

$$\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G. \tag{B4}$$

Question 3 is a more refined problem for finding a twirling group in RB. However, in general, Eq.(B2) is challenging to characterize, so we substitute it with a more easily handled condition, that is,

$$U\mathsf{G}U^\dagger = \mathsf{G}. \tag{B5}$$

In this case, $\mathsf{V} = \mathsf{G}$ and Question 3 reduces to Question 1 in the main text, as shown below.

Question 1. *Given a gate, U, find a twirling group, G such that,*

$$\begin{aligned} &\textit{for any quantum channel } \Lambda, \Lambda_G = \mathbb{E}_{G\epsilon\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \textit{ is diagonal} \\ &\textit{up to a unitary transformation independent of } \Lambda, \end{aligned} \tag{B6}$$

and,

$$U\mathsf{G}U^\dagger = \mathsf{G}. \tag{B7}$$

As mentioned in the main text, the solution to Question 1 allows us to obtain the fidelity of $(\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}$, which is a lower bound of the fidelity of Λ. Below, we present the proof.

Lemma 11. *Given a target gate, U, and a twirling group, G satisfying Question 1, then for any positive integer m, the fidelity of $(\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}$ is a lower bound of the fidelity of Λ. Mathematically,*

$$F((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}) \le F(\Lambda), \tag{B8}$$

or equivalently,

$$\mathrm{tr}\Big((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}\Big) \le \mathrm{tr}(\Lambda). \tag{B9}$$

Proof. As $\mathrm{tr}(\Lambda) = \mathrm{tr}(\Lambda_G)$, we only need to prove $\mathrm{tr}\Big((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}\Big) \le \mathrm{tr}(\Lambda_G)$.

Note that Λ_G has the expression of Eq. (A41) and can be written as

$$\Lambda_G = \sum_{i=1}^{k} p_i\Pi_i. \tag{B10}$$

Also, the condition $U\mathsf{G}U^\dagger = \mathsf{G}$ ensures that $\mathcal{U}^\dagger\Lambda_G\mathcal{U} = (\mathcal{U}^\dagger\Lambda_G\mathcal{U})_G$ as proven in Lemma 12, which means we can express $\mathcal{U}^\dagger\Lambda_G\mathcal{U}$ as

$$\mathcal{U}^\dagger\Lambda_G\mathcal{U} = \sum_{i=1}^{k} q_i\Pi_i. \tag{B11}$$

As $\mathcal{U}$ is unitary and does not change the spectrum of Λ_G, the value of q_i must be equal to one of elements from $\{p_i, 1 \le i \le k\}$. Thus, we can construct a map $f_1 : \{p_i, 1 \le i \le k\} \to \{q_i, 1 \le i \le k\}$. Similarly, as $\mathcal{U}(\mathcal{U}^\dagger\Lambda_G\mathcal{U})\mathcal{U}^\dagger = \Lambda_G$, we can construct a map $f_2 : \{q_i, 1 \le i \le k\} \to \{p_i, 1 \le i \le k\}$ so that the compositions of f_1 and f_2 are identity maps, $f_1 \circ f_2(q_i) = q_i$ and $f_2 \circ f_1(p_i) = p_i$. Then, f_1 is just a permutation on $\{p_i, 1 \le i \le k\}$ and $\{q_i, 1 \le i \le k\}$ is equivalent to $\{p_i, 1 \le i \le k\}$ after permutation. Thus, the diagonal terms of $\mathcal{U}^\dagger\Lambda_G\mathcal{U}\Lambda_G$ are $\{p_ip_{\sigma(i)}\}$ where σ is a permutation on $\{1, 2, \cdots, k\}$.

Similarly, the diagonal terms of $\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m = (\mathcal{U}^\dagger\cdots(\mathcal{U}^\dagger(\mathcal{U}^\dagger\Lambda_G\mathcal{U})\Lambda_G\mathcal{U})\Lambda_G\cdots\mathcal{U})\Lambda_G$ are $\{p_ip_{\sigma_1(i)}\cdots p_{\sigma_{m-1}(i)}\}$ where $\sigma_1, \cdots, \sigma_{m-1}$ are permutations on $\{1, 2, \cdots, k\}$. Thus, with rearrangement inequality,

$$\begin{aligned} \mathrm{tr}\Big((\mathcal{U}^{\dagger m}(\mathcal{U}\Lambda_G)^m)^{\frac{1}{m}}\Big) &= \sum_{i=1}^{k}(p_ip_{\sigma_1(i)}\cdots p_{\sigma_{m-1}(i)})^{\frac{1}{m}} \\ &\le \sum_{i=1}^{k} p_i \\ &= \mathrm{tr}(\Lambda_G). \end{aligned} \tag{B12}$$

Here is the proof. □

Thus, solve Question 1, and then we can provide a lower bound of the fidelity of the target gate. Based on the proof of Lemma 11, this lower bound is close to the real fidelity as long as the diagonal terms of the noise channel in the Pauli-Liouville representation are close to each other. And the lower bound is saturated for a global depolarizing noise. More importantly, if the target gate U is a multi-qubit controlled phase gate, the twirling group satisfying $U\mathsf{G}U^\dagger = \mathsf{G}$ would also satisfy $\Lambda_G\mathcal{U} = \mathcal{U}\Lambda_G$, which we will show in Theorem 2. As a consequence, we can accurately estimate the fidelities of multi-qubit controlled phase gates instead of only providing a lower bound.

Lemma 12. *Given a unitary gate, U, and a unitary subgroup, G, if $U\mathsf{G}U^\dagger = \mathsf{G}$, then for any quantum channel Λ, $(\mathcal{U}\Lambda_G\mathcal{U}^\dagger)_G = \mathcal{U}\Lambda_G\mathcal{U}^\dagger$.*

Proof. Provided with $U\mathsf{G}U^\dagger = \mathsf{G}$, we have

$$\begin{aligned}(\mathcal{U}\Lambda_G\mathcal{U}^\dagger)_G &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\mathcal{U}\Lambda_G\mathcal{U}^\dagger\mathcal{G}^\dagger \\ &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{U}(\mathcal{U}^\dagger\mathcal{G}\mathcal{U})\Lambda_G(\mathcal{U}^\dagger\mathcal{G}^\dagger\mathcal{U})\mathcal{U}^\dagger \\ &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{U}\mathcal{G}\Lambda_G\mathcal{G}^\dagger\mathcal{U}^\dagger \\ &= \mathcal{U}\Lambda_G\mathcal{U}^\dagger.\end{aligned} \tag{B13}$$

The equality in the third line comes from the condition $U\mathsf{G}U^\dagger = \mathsf{G}$. □

2. Proof of Lemma 1

In the main text, we present the lemma that for a finite n-qubit unitary subgroup G, if for any quantum channel Λ, its G–twirled channel, Λ_G, is diagonal in the Pauli-Liouville representation up to a unitary transformation independent of Λ, then the Pauli-Liouville representation of G is multiplicity-free. Also, the cardinality of the twirling group $|\mathsf{G}| \geq 4^n$. Below, we present the proof of this lemma. It is worth mentioning that, in the proof, we consider a more generic scenario: if Λ_G is diagonal up to an invertible matrix transformation rather than a unitary transformation, then the Pauli-Liouville representation of G is multiplicity-free. The result is more generic, and the proof is stronger here.

For convenience, we rewrite the Lemma 1 below.

Lemma 1. *For a finite n-qubit unitary subgroup, G, if for any quantum channel Λ, its G–twirled channel, Λ_G, is diagonal in the Pauli-Liouville representation up to an invertible matrix transformation independent of Λ, then the Pauli-Liouville representation of G is multiplicity-free. As a corollary, the cardinality of the twirling group $|\mathsf{G}| \geq 4^n$.*

Proof. Let us begin with the irreducible representation decomposition of the Pauli Liouville representation of G. For each element $G \in \mathsf{G}$, $\mathcal{G}$ can be decomposed as the direct sum of irreducible representations up to an isomorphism $\mathcal{V}$,

$$\mathcal{V}\mathcal{G}\mathcal{V}^{-1} = \bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G), \tag{B14}$$

where $\mathcal{V}$ is an invertible matrix, ϕ_i denotes the irreducible representation of G, and n_i denotes its multiplicity in $\mathcal{G}$. The term k records the number of inequivalent irreducible representations that $\mathcal{G}$ contains. It is worth mentioning that the basis making $\mathcal{G}$ block-diagonal may not be Pauli operators $\{\mathbb{I}, X, Y, Z\}$. That is why we put $\mathcal{V}$ and $\mathcal{V}^{-1}$ in the left side of Eq. (B14). As any unitary channel has an invariant subspace $\{\mathbb{I}\}$, the Pauli-Liouville representation of a unitary channel must be block-diagonal as below.

$$\mathcal{G} = \begin{pmatrix}1 & \mathbf{0}\\ \mathbf{0} & \mathbf{T}_\mathcal{G}\end{pmatrix}. \tag{B15}$$

Thus, the form of the basis transformation matrix, $\mathcal{V}$, can also be constrained like Eq. (B15). That means $\mathcal{V}$ only changes the non-identity basis while keeping the basis $\mathbb{I}$ invariant. In addition, without loss of generality, we set the irreducible representation in the subspace spanned by $\{\mathbb{I}\}$ as the trivial representation, that is, mapping all group elements to 1.

Given a channel, Λ, its twirling over group G is

$$\begin{aligned}\Lambda_G &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^{-1} \\ &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{V}^{-1}(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))\mathcal{V}\Lambda\mathcal{V}^{-1}(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))^{-1}\mathcal{V} \\ &= \mathcal{V}^{-1}[\mathbb{E}_{G\in\mathsf{G}}(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))\mathcal{V}\Lambda\mathcal{V}^{-1}(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))^{-1}]\mathcal{V}\end{aligned} \tag{B16}$$

Focusing on the calculation in the square brackets in Eq. (B16), we denote $\Lambda' = \mathcal{V}\Lambda\mathcal{V}^{-1}$ and decompose it corresponding to the block partition of the right-hand side in Eq. (B14). That is,

$$\Lambda' = \begin{pmatrix} \Lambda'^{11} & \Lambda'^{12} & \cdots & \Lambda'^{1k} \\ \Lambda'^{21} & \Lambda'^{22} & \cdots & \Lambda'^{2n_k} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda'^{k1} & \Lambda'^{k2} & \cdots & \Lambda'^{kk} \end{pmatrix}, \tag{B17}$$

where we set

$$\Lambda'^{ii} = \begin{pmatrix} \Lambda'^{i}_{11} & \Lambda'^{i}_{12} & \cdots & \Lambda'^{i}_{1n_i} \\ \Lambda'^{i}_{21} & \Lambda'^{i}_{22} & \cdots & \Lambda'^{i}_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda'^{i}_{n_i1} & \Lambda'^{i}_{n_i2} & \cdots & \Lambda'^{i}_{n_in_i} \end{pmatrix}. \tag{B18}$$

Then,

$$\begin{aligned}
&\mathbb{E}_{G\in\mathsf{G}}(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))\Lambda'(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))^{-1} \\
=&\mathbb{E}_{G\in\mathsf{G}}\bigoplus_{i=1}^{k}[\begin{pmatrix} \phi_i(G) & 0 & \cdots & 0 \\ 0 & \phi_i(G) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_i(G) \end{pmatrix}\begin{pmatrix} \Lambda'^{i}_{11} & \Lambda'^{i}_{12} & \cdots & \Lambda'^{i}_{1n_i} \\ \Lambda'^{i}_{21} & \Lambda'^{i}_{22} & \cdots & \Lambda'^{i}_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda'^{i}_{n_i1} & \Lambda'^{i}_{n_i2} & \cdots & \Lambda'^{i}_{n_in_i}. \end{pmatrix}\begin{pmatrix} \phi_i^{-1}(G) & 0 & \cdots & 0 \\ 0 & \phi_i^{-1}(G) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_i^{-1}(G) \end{pmatrix}] \\
=&\mathbb{E}_{G\in\mathsf{G}}\bigoplus_{i=1}^{k}\begin{pmatrix} \phi_i(G)\Lambda'^{i}_{11}\phi_i^{-1}(G) & \phi_i(G)\Lambda'^{i}_{12}\phi_i^{-1}(G) & \cdots & \phi_i(G)\Lambda'^{i}_{1n_i}\phi_i^{-1}(G) \\ \phi_i(G)\Lambda'^{i}_{21}\phi_i^{-1}(G) & \phi_i(G)\Lambda'^{i}_{22}\phi_i^{-1}(G) & \cdots & \phi_i(G)\Lambda'^{i}_{2n_i}\phi_i^{-1}(G) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_i(G)\Lambda'^{i}_{n_i1}\phi_i^{-1}(G) & \phi_i(G)\Lambda'^{i}_{n_i2}\phi_i^{-1}(G) & \cdots & \phi_i(G)\Lambda'^{i}_{n_in_i}\phi_i^{-1}(G). \end{pmatrix} \\
=&\bigoplus_{i=1}^{k}\begin{pmatrix} \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{11}\phi_i^{-1}(G) & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{12}\phi_i^{-1}(G) & \cdots & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{1n_i}\phi_i^{-1}(G) \\ \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{21}\phi_i^{-1}(G) & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{22}\phi_i^{-1}(G) & \cdots & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{2n_i}\phi_i^{-1}(G) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{n_i1}\phi_i^{-1}(G) & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{n_i2}\phi_i^{-1}(G) & \cdots & \mathbb{E}_{G\in\mathsf{G}}\phi_i(G)\Lambda'^{i}_{n_in_i}\phi_i^{-1}(G). \end{pmatrix}
\end{aligned} \tag{B19}$$

By Schur's lemma, for any irreducible representation ϕ_i and matrix A, the twirling of A over G would be proportional to identity,

$$\mathbb{E}_{G\in\mathsf{G}}\phi_i(G)A\phi_i^{-1}(G) = \mathrm{tr}(A)\frac{\mathbb{I}_{d_i}}{d_i}, \tag{B20}$$

where $d_i = \dim\phi_i$. We set

$$\Lambda'_i = \begin{pmatrix} \mathrm{tr}\left(\Lambda'^{i}_{11}\right) & \mathrm{tr}\left(\Lambda'^{i}_{12}\right) & \cdots & \mathrm{tr}\left(\Lambda'^{i}_{1n_i}\right) \\ \mathrm{tr}\left(\Lambda'^{i}_{21}\right) & \mathrm{tr}\left(\Lambda'^{i}_{22}\right) & \cdots & \mathrm{tr}\left(\Lambda'^{i}_{2n_i}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}\left(\Lambda'^{i}_{n_i1}\right) & \mathrm{tr}\left(\Lambda'^{i}_{n_i2}\right) & \cdots & \mathrm{tr}\left(\Lambda'^{i}_{n_in_i}\right). \end{pmatrix} \tag{B21}$$

Thus we conclude that

$$\mathbb{E}_{G\in\mathsf{G}}(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))\Lambda'(\bigoplus_{i=1}^{k}\mathbb{I}_{n_i\times n_i}\otimes\phi_i(G))^{-1} = \bigoplus_{i=1}^{k}\Lambda'_i\otimes\frac{\mathbb{I}_{d_i}}{d_i}, \tag{B22}$$

and

$$\Lambda_G = \mathcal{V}^{-1}(\bigoplus_{i=1}^{k}\Lambda'_i\otimes\frac{\mathbb{I}_{d_i}}{d_i})\mathcal{V}. \tag{B23}$$

Note that Λ_G is diagonal up to an invertible matrix transformation independent of Λ. There exists a fixed invertible matrix transformation $\mathcal{V}'$ such that

$$\mathcal{V}'\Lambda_G\mathcal{V}'^{\dagger} = \sum_i p_i\Pi_i, \tag{B24}$$

where Π_i is a projector independent of Λ and $\mathcal{V}'$ is also independent of Λ.

Notice that Eq. (B24) is a linear function acting on Λ. If for any channel Λ, Λ_G is diagonal, the linear combination of any set of quantum channels would also be diagonal after G-twirling. Note that the linear span of quantum channels is the set of all trace-preserving (TP) maps [53]. Thus, for any TP map with form $\Lambda = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{t} & \mathbf{T}_\Lambda \end{pmatrix}$, $\mathcal{V}'\Lambda_G\mathcal{V}'^\dagger$ must be diagonal. It further requires that for any index i, Λ_i' defined in Eq. (B21) is diagonal up to an invertible matrix transformation. Recall that Λ_i' is defined from Λ', which equals $\mathcal{V}\Lambda\mathcal{V}^{-1}$. As $\mathcal{V}$ has the form $\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_\mathcal{V} \end{pmatrix}$ and $\Lambda = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{t} & \mathbf{T}_\Lambda \end{pmatrix}$ is an arbitrary TP map, $\Lambda' = \mathcal{V}\Lambda\mathcal{V}^{-1} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{T}_\mathcal{V}\mathbf{t} & \mathbf{T}_\mathcal{V}\mathbf{T}_\Lambda\mathbf{T}_\mathcal{V}^{-1} \end{pmatrix}$ is also an arbitrary TP map. Denote $\mathbf{t}' = \mathbf{T}_\mathcal{V}\mathbf{t}$ and $\mathbf{T}_{\Lambda'} = \mathbf{T}_\mathcal{V}\mathbf{T}_\Lambda\mathbf{T}_\mathcal{V}^{-1}$, the elements of $\mathbf{t}'$ and $\mathbf{T}_{\Lambda'}$ would be arbitrary. Back to considering Λ_i', when $i = 1$, ϕ_1 is the trivial irreducible representation, then $\Lambda_1' = \Lambda'^i$ can be written as below,

$$\Lambda_1' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \mathbf{t}_1' & \mathbf{T}_{\Lambda_{11}'} & \cdots & \mathbf{T}_{\Lambda_{1,n_1-1}'} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{t}_{n_1-1}' & \mathbf{T}_{\Lambda_{n_1-1,1}'} & \cdots & \mathbf{T}_{\Lambda_{n_1-1,n_1-1}'} \end{pmatrix}, \tag{B25}$$

where the matrix elements are arbitrary except for the first line. If $n_1 > 1$, there exists a matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \oplus \mathbb{I}_{n_1-2}$ that cannot be diagonalized. Thus, n_1 must take 1. Similarly, when $i \geq 2$, Λ_i' can take any matrix as Λ' is an arbitrary TP map. If $n_i > 1$ in this case, we can also find a matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \oplus \mathbb{I}_{n_i-2}$ that cannot be diagonalized. The above arguments indicate that $\forall i$, $n_i = 1$. The irreducible representation decomposition of Pauli-Liouville representation of G must be multiplicity-free to make Λ_G diagonal for any channel Λ. This completes the proof of the first conclusion in the lemma.

A direct corollary of the non-multiplicity condition of G in Liouville representation is the cardinality of the twirling group $|\mathsf{G}| \geq 4^n$. This can be obtained by the Burnside theorem, as shown below.

$$\begin{aligned} |\mathsf{G}| &= \sum_{\phi_i \in R_G} \dim^2 \phi_i \\ &\geq \sum_{i=1}^{k} \dim^2 \phi_i \\ &\geq \sum_{i=1}^{k} \dim \phi_i \\ &= 4^n, \end{aligned} \tag{B26}$$

where R_G records all inequivalent irreducible representations of G. □

3. Systematic twirling group construction for generic quantum gates

Before we go through the proof of the main theorem, we discuss how to construct twirling groups for generic quantum gates satisfying the conditions in Question 1.

Notice that if G contains a subgroup H that would make arbitrary channels diagonal, up to a unitary transformation, via twirling, then G would also enjoy this property as $\Lambda_G = (\Lambda_G)_H$, which is shown below.

Lemma 13. *If G contains a subgroup H that would twirl any noise channel into a diagonal channel, up to a unitary transformation, in the Pauli-Liouville representation, then G would also enjoy this property. Mathematically,*

$$\mathsf{H} \subset \mathsf{G}, \forall \Lambda, \Lambda_H \text{ is diagonal} \Rightarrow \forall \Lambda, \Lambda_G \text{ is diagonal}. \tag{B27}$$

Proof.

$$\begin{aligned} (\Lambda_G)_H &= \mathbb{E}_{G_h \in \mathsf{H}} \mathbb{E}_{G \in \mathsf{G}} \mathcal{G}_h \mathcal{G} \Lambda (\mathcal{G}_h \mathcal{G})^\dagger \\ &= \mathbb{E}_{G \in \mathsf{G}} \mathcal{G} \Lambda \mathcal{G}^\dagger \\ &= \Lambda_G. \end{aligned} \tag{B28}$$

Proof is done. □

As Pauli group P_n can twirl any channel into a Pauli channel, a simple solution to Question 1 is just the smallest group containing Pauli group and normalized by target U, which is constructed and spanned by continuously applying U on P_n until no new element is generated. It is worth noting that, except for Pauli group P_n, local dihedral group $\mathsf{D}_n^m = \langle X, Z_m\rangle^{\otimes n}$ can also twirl any channel into a Pauli channel. In fact, P_n is a special case of D_n^m when $m = 2$. For specific target gate U like Z_m, substituting P_n with D_n^m may lead to a smaller twirling group G though D_n^m is in general larger than P_n. In reality, one can select an optimal m to obtain a better choice of G. The above discussions can be summarized below.

Corollary 1. *If n-qubit Pauli group $\mathsf{P}_n \subseteq \mathsf{G}$, then Λ_G is a Pauli channel and is diagonal in the Pauli-Liouville representation.*

Corollary 2. *If n-qubit local dihedral group $\mathsf{D}_n^m \subseteq \mathsf{G}$, then Λ_G is a Pauli channel and is diagonal in the Pauli-Liouville representation.*

Example 1 (Simple twirling group construction)**.** *Given an n-qubit unitary, U, a simple solution to Question 1 is the smallest group containing D_n^m and normalized by U where $m \geq 2$ is a positive integer. Concretely, the twirling group G can be constructed as follows.*

$$\begin{aligned}\mathsf{G} &= \langle \bigcup_{l\in\mathbb{N}} U^l \mathsf{D}_n^m (U^\dagger)^l \rangle \\ &= \{(U^{l_1} D_1 (U^\dagger)^{l_1})^{k_1} (U^{l_2} D_2 (U^\dagger)^{l_2})^{k_2} \cdots, \forall i, l_i, k_i \in \mathbb{Z}, D_i \in \mathsf{D}_n^m\} \\ &= \{U^{l_1} D_1 (U^\dagger)^{l_1} U^{l_2} D_2 (U^\dagger)^{l_2} \cdots, \forall i, l_i \in \mathbb{Z}, D_i \in \mathsf{D}_n^m\}.\end{aligned} \tag{B29}$$

Here, D_n^m is the n-qubit local dihedral group $\langle X, Z_m\rangle^{\otimes n}$. In practice, we select an optimal m to make $|\mathsf{G}|$ as small as possible.

4. Proof of Lemma 2

Below, we focus on the case that the twirling group is a CRU subgroup and prove Lemma 2 and Theorem 1 in the main text. We first provide a more formal and more mathematical version of Lemma 2.

Lemma 2 (Formal Version)**.** *For a finite n-qubit CRU subgroup G, set the Z-basis subgroup of G as $\mathsf{G}_Z = \{U \in \mathsf{G} | U \text{ is } Z \text{ basis}\}$, which is a normal subgroup of G by Lemma 9. If for any quantum channel Λ, its twirling over group G, Λ_G, is diagonal up to a unitary transformation in the Pauli-Liouville representation, then the quotient group G/G_Z can interchange any two computational basis states. In another word, G/G_Z contains set $\mathsf{S} = \{\Pi_{\mathbf{i}} | \Pi_{\mathbf{i}} = \Pi_{\mathbf{i}}^X \Pi_{\mathbf{i}}^C, \mathbf{i} = i_1 i_2 ... i_n \in \{0,1\}^n, \Pi_{\mathbf{i}} = X_1^{i_1} X_2^{i_2} ... X_n^{i_n} \in \mathsf{X}, \Pi_{\mathbf{i}}^C \in \mathsf{C}^{[n-1]}\mathsf{X}\}$. Here, $\mathsf{X} = \langle X\rangle$ is the group generated by Pauli X gates on all qubits, $\mathsf{C}^{[n-1]}\mathsf{X} = \langle CX, CCX, C^{n-1}X\rangle$ is the group generated by CNOT and multi-qubit Toffoli gates on all qubits.*

Note that $\langle\cdot\rangle$ denotes the group generated by $\cdot$. Same with before, we simply use $\langle X\rangle$ to represent group $\langle X_1, X_2, \cdots, X_n\rangle$ and $\langle CX, CCX, C^{n-1}X\rangle$ to represent $\langle CX_{12}, \cdots CCX_{123}, \cdots, C^{n-1}X_{1,2,\cdots n}\rangle$ while subscripts label the qubits acted upon. Below we provide the proof of Lemma 2.

Proof. From Lemma 8, we obtain that the space spanned by $\{\mathbb{I}, Z\}^{\otimes n}$ is an invariant subspace of G. By rearranging Pauli operator bases and put $\{\mathbb{I}, Z\}^{\otimes n}$ forward, the Pauli-Liouville representation of G would be in a block-diagonal form

$$\mathcal{G} = \mathcal{G}_Z \bigoplus \mathcal{G}_\perp, \tag{B30}$$

where $\dim \mathcal{G}_Z = 2^n$ and $\dim \mathcal{G}_\perp = 4^n - 2^n$. Correspondingly, we represent channel Λ in a block-diagonal form in the rearranged basis,

$$\Lambda = \begin{pmatrix} \Lambda_Z & \Lambda_{Z\perp} \\ \Lambda_{\perp Z} & \Lambda_\perp \end{pmatrix}. \tag{B31}$$

Thus, the twirling of Λ over G is

$$\begin{aligned}\Lambda_G &= \mathbb{E}_{G\in\mathsf{G}} \mathcal{G}\Lambda\mathcal{G}^{-1} \\ &= \mathbb{E}_{G\in\mathsf{G}} \begin{pmatrix} \mathcal{G}_Z \Lambda_Z \mathcal{G}_Z^{-1} & \mathcal{G}_Z \Lambda_{Z\perp} \mathcal{G}_\perp^{-1} \\ \mathcal{G}_\perp \Lambda_{\perp Z} \mathcal{G}_Z^{-1} & \mathcal{G}_\perp \Lambda_\perp \mathcal{G}_\perp^{-1} \end{pmatrix}\end{aligned} \tag{B32}$$

Due to the direct sum structure, the irreducible representation decomposition of $\mathcal{G}$ is composed of decompositions of $\mathcal{G}_Z$ and $\mathcal{G}_\perp$. From Lemma 1, all of the irreducible representations contained in $\mathcal{G}$ must be inequivalent to make Λ_G diagonal. Therefore, $\mathcal{G}_Z$ and $\mathcal{G}_\perp$ do not have any equivalent irreducible representation in common, which leads to $\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_Z\Lambda_{Z\perp}\mathcal{G}_\perp^{-1}=0$, $\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_\perp\Lambda_{\perp Z}\mathcal{G}_Z^{-1}=0$, and $\Lambda_G=(\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_Z\Lambda_Z\mathcal{G}_Z^{-1})\oplus(\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_\perp\Lambda_\perp\mathcal{G}_\perp^{-1})$. To make Λ_G diagonal, we require that two blocks in it are both diagonal. Below, we focus on the first block.

Note that G_Z is a normal subgroup of G. Consider its coset, or quotient group $\mathsf{G}/\mathsf{G}_Z=\{\Pi_i\mathsf{G}_Z, 1\le i\le|\mathsf{G}/\mathsf{G}_Z|\}$, where $\forall i$, Π_i is a representative element. Without loss of generality, we set Π_1 as $\mathbb{I}_{2^n}$. Then the other representative elements must be outside G_Z. Then we separate the twirling of Λ_Z over G into two parts. One is the twirling over G_Z, and the other is the twirling over quotient group G/G_Z. As Z-basis gates are all identity in the basis of $\{\mathbb{I},Z\}^{\otimes n}$ in Liouville representation, they do not influence the twirled channel in that subspace. For instance, the Liouville representation of the CS gate is shown in Fig. 7. In the subspace spanned by $\{\mathbb{I},Z_1,Z_2,Z_1Z_2\}$, CS gate is equal to identity. Certainly, if the CS gate is a twirling gate, it contributes nothing to the twirling in this subspace. The same are the gates in G_Z.

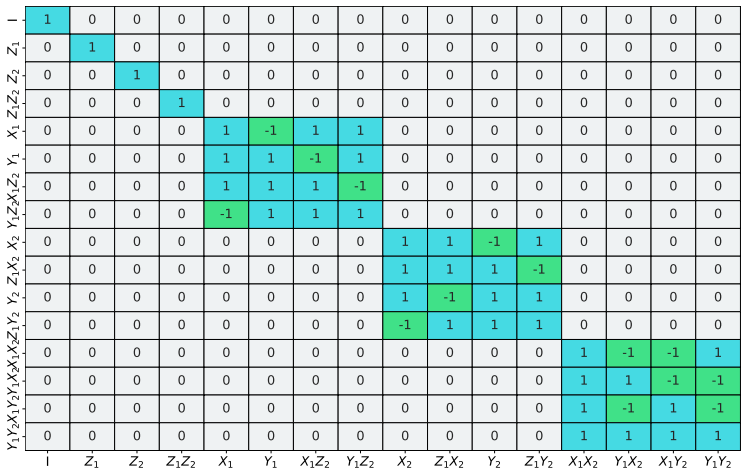


FIG. 7. Pauli Liouville representation of CS gate.

Thus, the only contribution for twirling in this subspace comes from the quotient group. With the similar arguments in Lemma 1, we can directly obtain that $|\mathsf{G}/\mathsf{G}_Z|\ge\dim\Lambda_Z=2^n$, the result in Corollary 3. Specifically,

$$\begin{aligned}\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_Z\Lambda_Z\mathcal{G}_Z^{-1}&=\mathbb{E}_{1\le i\le|\mathsf{G}/\mathsf{G}_Z|}\Pi_{iZ}(\mathbb{E}_{W\in\mathsf{G}_Z}\mathcal{W}_Z\Lambda_Z\mathcal{W}_Z^{-1})\Pi_{iZ}^{-1}\\&=\mathbb{E}_{1\le i\le|\mathsf{G}/\mathsf{G}_Z|}\Pi_{iZ}\Lambda_Z\Pi_{iZ}^{-1},\end{aligned}\tag{B33}$$

where the subscript Z denotes the sub-representation of Pauli-Liouville representation in $\{\mathbb{I},Z\}^{\otimes n}$. It can be verified that $\{\Pi_{iZ}, 1\le i\le|\mathsf{G}/\mathsf{G}_Z|\}$ is a representation of quotient group G/G_Z. Therefore, one can decompose Π_{iZ} with irreducible representations of G/G_Z,

$$\Pi_{iZ}=\bigoplus_{j=1}^{k}\phi_j(\Pi_i).\tag{B34}$$

The twirled channel $\mathbb{E}_{G\in\mathsf{G}}\mathcal{G}_Z\Lambda_Z\mathcal{G}_Z^{-1}$ would be block diagonal corresponding to the irreducible representation decom-

position of Π_{iZ}. With the same arguments in Lemma 1, Π_{iZ} must be multiplicity-free. Then

$$\begin{aligned} |\mathsf{G}/\mathsf{G}_Z| &= \sum_{j\in R_{G/Z}} \dim^2 \phi_j \\ &\geq \sum_{j=1}^{k} \dim^2 \phi_j \\ &\geq \sum_{j=1}^{k} \dim \phi_j \\ &= 2^n, \end{aligned} \tag{B35}$$

where $R_{G/Z}$ records all inequivalent irreducible representations of G/G_Z.

To obtain the result in Lemma 2, we further study the irreducible representation decomposition of quotient group G/G_Z in the space spanned by $\{\mathbb{I}, Z\}^{\otimes n}$. As mentioned before, Π_{iZ} must be multiplicity-free, which means each irreducible representation can appear at most once in Π_{iZ}. Focusing on the trivial irreducible representation, its multiplicity in Π_{iZ} can be obtained via Lemma 3 and is given by

$$m_t = \mathbb{E}_{\Pi\in\mathsf{G}/\mathsf{G}_Z} 1 \cdot \mathrm{tr}(\Pi_Z), \tag{B36}$$

where 1 and $\mathrm{tr}(\Pi_Z)$ are characters of trivial irreducible representation and representation in space $\{\mathbb{I}, Z\}^{\otimes n}$, respectively. Through direct calculation, Eq. (B36) can be simplified to

$$\begin{aligned} m_t &= \mathbb{E}_\Pi \frac{1}{2^n} \sum_{W\in\{\mathbb{I},\mathbb{Z}^{\otimes n}\}} \mathrm{tr}\big(W\Pi W\Pi^\dagger\big) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{i_1,i_2,\cdots,i_n=0}^{1} \mathrm{tr}\left((\prod_{j=1}^{n} Z_j^{i_j})\Pi(\prod_{j=1}^{n} Z_j^{i_j})\Pi^\dagger \right) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{i_1,i_2,\cdots,i_n=0}^{1} \mathrm{tr}\left((\prod_{j=1}^{n} Z_j^{i_j})(\prod_{j=1}^{n} \Pi Z_j^{i_j}\Pi^\dagger) \right) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{i_1,i_2,\cdots,i_n=0}^{1} \mathrm{tr}\left((\prod_{j=1}^{n} Z_j^{i_j})(\prod_{j=1}^{n} \Pi Z_j^{i_j}\Pi^\dagger) \right) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{i_1,i_2,\cdots,i_n=0}^{1} \mathrm{tr}\left((\prod_{j=1}^{n} Z_j^{i_j}\Pi Z_j^{i_j}\Pi^\dagger) \right) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \,\mathrm{tr}\left((\prod_{j=1}^{n} \sum_{i_j=0}^{1} Z_j^{i_j}\Pi Z_j^{i_j}\Pi^\dagger) \right) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \,\mathrm{tr}\left(\prod_{j=1}^{n} (\mathbb{I} + Z_j\Pi Z_j\Pi^\dagger) \right). \end{aligned} \tag{B37}$$

As $\Pi \in \mathsf{G}/\mathsf{G}_Z$ is a permutation matrix, $\Pi Z_j \Pi^\dagger$, $Z_j \Pi Z_j \Pi^\dagger$, and $\mathbb{I} + Z_j \Pi Z_j \Pi^\dagger$ are all diagonal matrices. Therefore,

Eq. (B37) can be further simplified to

$$\begin{aligned} m_t &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n}\prod_{j=1}^{n}\langle\mathbf{i}|\mathbb{I}+Z_j\Pi Z_j\Pi^\dagger|\mathbf{i}\rangle \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n}\prod_{j=1}^{n}(1+\langle\mathbf{i}|Z_j\Pi Z_j\Pi^\dagger|\mathbf{i}\rangle) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n}\prod_{j=1}^{n}(1+(-1)^{\mathbf{i}_j}\langle\mathbf{i}|\Pi Z_j\Pi^\dagger|\mathbf{i}\rangle) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n}\prod_{j=1}^{n}(1+(-1)^{\mathbf{i}_j+\pi(\mathbf{i})_j}) \\ &= \frac{1}{2^n}\mathbb{E}_\Pi \sum_{\mathbf{i}\in\{0,1\}^n}2^n\delta_{\pi(\mathbf{i})=\mathbf{i}} \\ &= \mathbb{E}_\Pi N_\Pi \\ &= O_{\mathsf{G}/\mathsf{G}_Z\to\{0,1\}^n}. \end{aligned} \tag{B38}$$

In the fourth line, we utilize that a permutation matrix would transform a bit string into another bit string and denote $|\pi(\mathbf{i})\rangle = \Pi^\dagger|\mathbf{i}\rangle$. Here, π is a permutation acting on $\{0,1\}^n$ and can be viewed as a representation of Π. In the sixth line, $N_\Pi = \sum_{\mathbf{i}\in\{0,1\}^n}\delta_{\pi(\mathbf{i})=\mathbf{i}}$ denotes the number of fixed points for π acting on $\{0,1\}^n$. Via Burnside's lemma, $\mathbb{E}_\Pi N_\Pi$ equals to the number of orbits for group G/G_Z acting on $\{0,1\}^n$ which we denote as $O_{\mathsf{G}/\mathsf{G}_Z\to\{0,1\}^n}$.

As trivial irreducible representation can appear at most once, m_t cannot be larger than 1, or the number of orbits cannot be larger than 1. It implies that gates of G/G_Z can interchange any two computational states or any two bit strings in $\{0,1\}^n$. Then, we obtain the results in the informal version of Lemma 2.

Now, we obtain that G/G_Z can transform 0^n to any other bit string in $\{0,1\}^n$. Recall that G/G_Z is a subgroup of $\langle X, C^{n-1}X\rangle$, each element in G/G_Z can be written as $\Pi_i^X\Pi_i^C$, where $\Pi_i^X \in \mathsf{X} = \langle X\rangle$, $\Pi_i^C \in \langle CX, CCX, \cdots, C^{n-1}X\rangle$. As any element in $\langle CX, CCX, \cdots, C^{n-1}X\rangle$ has no effect on bit string 0^n, $\Pi_i^X\Pi_i^C$ would simply transform $|0^n\rangle$ to $\Pi_i^X|0^n\rangle$. Thus, Π_i^X must take over all elements in X to transform 0^n to all bit strings in $\{0,1\}^n$. Here, we complete the proof of Lemma 2. □

In the main text, we have discussed the advantages of choosing the CRU subgroup as a twirling group in virtually implementing the inverse gate. Below, we discuss that it might be sufficient only to consider CRU twirling gates for tailoring diagonal gates. We notice that in most RB protocols, the twirling groups contain a CRU subgroup that can twirl the noise channel diagonal, including Clifford group [6], generalized matchgate group [18], and CNOT dihedral group [16]. We conjecture that if any finite group G can make arbitrary noise channels diagonal via twirling, then under the equivalence of unitary transformation, it has a CRU subgroup G_C also achieving that. A concrete example is G as the Clifford group and G_C as the Pauli group. Thus, for diagonal gate U in the computational basis, considering G in CRU highly likely suffices to find the optimal solution. It is worth noting that the above discussion is only conjecture, and we hope in the future, people can find the smallest group in the whole unitary group for tailoring generic quantum gates.

From Lemma 2, we can directly obtain a lower bound for the cardinality of the quotient group, as shown in the following corollary.

Corollary 3. *For a finite n-qubit CRU subgroup G, set the computational basis subgroup of G as $\mathsf{G}_Z = \{U \in \mathsf{G} | U \text{ is computational basis}\}$. By Lemma 8, G_Z is a normal subgroup of G. If for any quantum channel Λ, its twirling over group G, Λ_G, is diagonal up to a unitary transformation in the Pauli-Liouville representation, then the cardinality of the quotient group $|\mathsf{G}/\mathsf{G}_Z| \geq 2^n$.*

5. Proof of Theorem 1

Below, we provide the proof of Theorem 1. The key point is utilizing that G contains specific permutation gates and G is normalized by target gate U. We first present a lemma, telling us what is the smallest group containing a given permutation group and normalized by a given computational basis diagonal gate U.

Lemma 14. *Given a permutation group, Π, and a diagonal gate, U, in the computational basis, the smallest group containing Π and normalized by U is given by*

$$\mathsf{G} = \Pi \ltimes \mathsf{W}, \tag{B39}$$

where $\ltimes$ *denotes semi-product and* $\mathsf{W} = \langle\{\Pi^\dagger U \Pi U^\dagger, \Pi \in \mathsf{\Pi}\}\rangle$.

Proof. We first clarify the meaning of semi-product. It means that

1. Any element in $\mathsf{\Pi} \ltimes \mathsf{W}$ has an expression of ΠW where $\Pi \in \mathsf{\Pi}$ and $W \in \mathsf{W}$.
2. W is a normal subgroup of $\mathsf{\Pi} \ltimes \mathsf{W}$.

It is easy to verify the second condition that W is a normal subgroup with the first condition. We only need to verify that for any Π and Π', $\Pi'^\dagger \Pi^\dagger U \Pi U^\dagger \Pi' \in \mathsf{W}$. This can be seen via the following equation.

$$\Pi'^\dagger \Pi^\dagger U \Pi U^\dagger \Pi' = (\Pi\Pi')^\dagger U (\Pi\Pi') U^\dagger (\Pi'^\dagger U \Pi' U^\dagger)^\dagger. \tag{B40}$$

As $(\Pi\Pi')^\dagger U(\Pi\Pi')U^\dagger$ and $\Pi'^\dagger U \Pi' U^\dagger$ are both elements in W, we successfully show the soundness of the second condition.

Now we turn to the proof of the lemma. Obviously, $\mathsf{\Pi} \ltimes \mathsf{W}$ contains $\mathsf{\Pi}$. It is also normalized by U:

Given an element ΠW where $\Pi \in \mathsf{\Pi}$ and $W \in \mathsf{W}$,

$$\begin{aligned} U \Pi W U^\dagger &= U \Pi U^\dagger W \\ &= \Pi (\Pi^\dagger U \Pi U^\dagger) W. \end{aligned} \tag{B41}$$

The first line comes from the fact that W and U are both diagonal gates. As $\Pi^\dagger U \Pi U^\dagger$ and W both belong to W, $U \Pi W U^\dagger$ has an expression of $\Pi W'$ where $W' \in \mathsf{W}$. Combining the condition 1, we obtain that $\mathsf{\Pi} \ltimes \mathsf{W}$ is normalized by U.

Meanwhile, for any group G containing $\mathsf{\Pi}$ and normalized by U, $U \Pi U^\dagger \in \mathsf{G}$ and $\Pi^\dagger U \Pi U^\dagger \in \mathsf{G}$. Then we know $\mathsf{\Pi}$ and W both belong to G. As a consequence, G must contain $\mathsf{\Pi} \ltimes \mathsf{W}$. Combining with the arguments before, we prove that $\mathsf{\Pi} \ltimes \mathsf{W}$ is the smallest group containing $\mathsf{\Pi}$ and normalized by U. □

With Lemma 14, we can easily prove Theorem 1 in the main text. For convenience, we rewrite the theorem below.

Theorem 1. *The optimal twirling group* G *in CRU for the multi-qubit controlled phase gate,* $U = C^n Z_m$*, with* $n \geq 1, m \geq 2$*, is the smallest group containing* X *and normalized by* U*, given by*

$$\mathsf{G} = \{\Pi(\prod_{i=1}^{t}(\Pi_i^\dagger U \Pi_i U^\dagger)^{l_i}) | \Pi \in \mathsf{X}, t \in \mathbb{N}, \forall i, l_i \in \pm 1, \Pi_i \in \mathsf{X}\}. \tag{B42}$$

Proof. From Lemma 2, we can deduce that G at least contains permutation group $\langle\mathsf{S}\rangle$ where $\mathsf{S} = \{\Pi_i^X \Pi_i^C, \forall \Pi_i^X \in \mathsf{X}, \exists \Pi_i^C \in \mathsf{C}^{[n-1]}\mathsf{X}\}$. Using Lemma 14, we obtain that any twirling group G as a solution to Question 1 must satisfy

$$\langle\mathsf{S}\rangle \ltimes \langle\{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle\mathsf{S}\rangle\}\rangle \leq \mathsf{G}. \tag{B43}$$

If $U = C^{n-1}Z_m = \begin{pmatrix} \mathbb{I}_{2^n-1} & \mathbf{0} \\ \mathbf{0} & e^{i\frac{2\pi}{m}} \end{pmatrix}$, we can show that $\langle\{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle\mathsf{X}\rangle\}\rangle \leq \langle\{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle\mathsf{S}\rangle\}\rangle$. For brevity, we denote $\mathsf{W} = \langle\{\Pi'^\dagger U \Pi' U^\dagger, \Pi' \in \langle\mathsf{S}\rangle\}\rangle$ and $\mathsf{W}_X = \langle\{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle\mathsf{X}\rangle\}\rangle$. For any generator $\Pi^\dagger U \Pi U^\dagger$ in W_X where $\Pi \in \mathsf{X}$, $\Pi^\dagger U \Pi$ is a diagonal matrix while only one diagonal element is not 1 but equals $e^{i\frac{2\pi}{m}}$. As the permutation elements in S can interchange any two computational bases, there must exist an element $\Pi' \in \mathsf{S}$ such that

$$\Pi'^\dagger U \Pi' = \Pi^\dagger U \Pi. \tag{B44}$$

Then,

$$\Pi'^\dagger U \Pi' U^\dagger = \Pi^\dagger U \Pi U^\dagger. \tag{B45}$$

Thus, any generator of W_X belongs to W. Then, we obtain that $\mathsf{W}_X \leq \mathsf{W}$ for $C^{n-1}Z_m$. This result also applies to the diagonal matrix in which only one diagonal element differs from the others.

Note that the smallest group containing X and normalized by U is $\mathsf{X} \ltimes \mathsf{W}_X$. As X is no larger and no global than $\langle\mathsf{S}\rangle$ and W_X is a subset of W, $\mathsf{X} \ltimes \mathsf{W}_X$ is obviously smaller and more easily implementable than $\langle\mathsf{S}\rangle \ltimes \langle\{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle\mathsf{S}\rangle\}\rangle$, which implies that the twirling group cannot be better than $\mathsf{X} \ltimes \mathsf{W}_X$. Below, we would show that for $U = C^n Z_m$ with $n \geq 1, m \geq 2$, $\mathsf{X} \ltimes \mathsf{W}_X$ suffices to tailor $C^n Z_m$. Combining the two sides, we would obtain that the optimal twirling group for $C^n Z_m$ is $\mathsf{X} \ltimes \mathsf{W}_X$, just shown in Theorem 1.

In the following, we show $\mathsf{X} \ltimes \mathsf{W}_X$ satisfies Question 1 for multi-qubit controlled phase gate $U = C^n Z_m$. From the definition of $\mathsf{X} \ltimes \mathsf{W}_X$, we know that $\mathsf{X} \ltimes \mathsf{W}_X$ is normalized by U. Thus, we only need to verify that $\mathsf{X} \ltimes \mathsf{W}_X$ can make arbitrary channels diagonal via twirling. With corollaries 1 and 2, it is sufficient to show that Pauli Z gate or Z_m gate on each qubit belongs to W_X. Below, we show that this is right for $U = C^n Z_m$ with $n \geq 1, m \geq 2$.

Recall that $\mathsf{W}_X = \langle \{\Pi^\dagger U \Pi U^\dagger, \Pi \in \langle \mathsf{X} \rangle\} \rangle$, in the following we analyze the generators of W_X, or $\Pi^\dagger U \Pi U^\dagger$, in detail. For certain $\Pi \in \langle X \rangle$, let us define its pattern to be a 0-1 bit string, s_j^Π, such that $s_j^\Pi = 1$ if the j-th qubit of Π is I and $s_j^\Pi = 0$ if the j-th qubit of Π is X. Then, the matrix representation of $\Pi^\dagger U \Pi U^\dagger$ for $U = C^n Z_m$ is

$$\Pi^\dagger C^n Z_m \Pi C^n Z_m^\dagger = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & e^{i\frac{2\pi}{m}} & & \\ & & & \ddots & \\ & & & & e^{-i\frac{2\pi}{m}} \end{pmatrix}, \tag{B46}$$

where the (s^Π, s^Π) entry is $e^{i\frac{2\pi}{m}}$, the $(2^{n+1}-1, 2^{n+1}-1)$ entry is $e^{-i\frac{2\pi}{m}}$, and other diagonal entries are 1. As generators are all diagonal, all elements in W_X are diagonal in the computational basis. Moreover, the diagonal elements would be power of $e^{i\frac{2\pi}{m}}$. Define an injective map $\phi : \mathsf{W}_X \to \mathbb{Z}_m^{2^{n+1}}$ where $\mathbb{Z}_m = \{0, 1, \cdots, m-1\}$,

$$\phi(w) = -i\frac{m}{2\pi} \begin{pmatrix} \ln[w]_{0,0} \\ \vdots \\ \ln[w]_{2^{n+1}-1, 2^{n+1}-1} \end{pmatrix}. \tag{B47}$$

Note that the entries of $\phi(w)$ will always belong to $\{0, 1, \cdots, m-1\}$. With map ϕ, we express the gate in W_X with a vector. For example, $\phi(\Pi^\dagger C^n Z_m \Pi C^n Z_m^\dagger) = e_{s^\Pi} - e_{2^{n+1}-1}$, where we define e_i to be the basis vector whose i-th entry is 1 and other entries are 0. In this way, the group multiplication is turned into integer vector addition, and the group generation is equivalent to the linear combination of vectors. As Π can take any element in X, vectors $v_0 = e_0 - e_{2^{n+1}-1}, \cdots, v_{2^{n+1}-2} = e_{2^{n+1}-2} - e_{2^{n+1}-1}$, are all basis vectors. Note that the overall phase of a quantum gate is not important, so any vectors differing by multiples of $v_{2^{n+1}-1} = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}^T$ are equivalent. In the next, we show how to construct vectors associated with Z or Z_m gates with vectors $v_0, \cdots, v_{2^{n+1}-1}$.

As different qubits are symmetric under arbitrary permutation, we only need to construct Z gate or Z_m gate on the first qubit. If m is odd, we show $Z_m \in \mathsf{W}_X$. Gate Z_m on first qubit corresponds to vector $\phi((Z_m)_1) = \begin{pmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix}^T$. Define

$$u = -(v_0 + v_1 + \cdots + v_{2^{n+1}-2}) + (2^{n+1}-1)v_{2^{n+1}-1} = \begin{pmatrix} 0 & \cdots & 2^{n+1} \end{pmatrix}^T. \tag{B48}$$

When m is odd, $\gcd(2^{n+1}, m) = 1$, vector u is equivalent to $\begin{pmatrix} 0 & \cdots & 1 \end{pmatrix}^T = e_{2^{n+1}-1}$. Then, we can also obtain $e_i = v_i + u$ for $0 \leq i \leq 2^{n+1} - 2$. With e_i for $0 \leq i \leq 2^{n+1} - 1$, $\phi((Z_m)_1)$ can certainly be constructed by

$$\phi((Z_m)_1) = e_{2^{n+1}-2^n} + e_{2^{n+1}-2^n+1} + \cdots + e_{2^{n+1}-1}. \tag{B49}$$

Thus, if m is odd, we can ensure that phase gates Z_m on all qubits belong to W_X. In this case, $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ definitely satisfies the requirements in Question 1.

When m is even, we show $Z \in \mathsf{W}_X$. We first express $m = q2^k$ where $q \geq 1$ is odd and k is a positive integer. Gate Z on first qubit corresponds to vector $\phi(Z_1) = \begin{pmatrix} 0 & \cdots & 0 & q2^{k-1} & \cdots & q2^{k-1} \end{pmatrix}^T$. As $\gcd(2^{n+1}, m) = 2^{\min(n+1,k)}$, u can be expressed as $\begin{pmatrix} 0 & \cdots & 2^{\min(n+1,k)} \end{pmatrix}^T$. Then, $\phi(Z_1)$ can be constructed by

$$\begin{aligned} \phi(Z_1) &= q2^{k-1}(e_{2^{n+1}-2^n} + e_{2^{n+1}-2^n+1} + \cdots + e_{2^{n+1}-1}) \\ &= q2^{k-1}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) + q2^{k-1}(2^n - 1)e_{2^{n+1}-1} + q2^{k-1}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) + q2^{n+k-1}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) + q2^{n+k-1-\min(n+1,k)}u. \end{aligned} \tag{B50}$$

As $n \geq 1$, $n + k - 1 - \min(n+1, k) = \max(n-1, k-2)$ is always a non-negative integer. Thus, $q2^{n+k-1-\min(n+1,k)}$ is an integer. As a consequence, $\phi(Z_1)$ can be constructed with linear combination of $v_0, \cdots, v_{2^{n+1}-2}$, and u, which is equivalent to $Z_1 \in \mathsf{W}_X$. Thus, if m is even, we can ensure that Pauli Z gates on all qubits belong to W_X. In this case, $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ also satisfies the requirements in Question 1. Proof is done. □

6. Twirling groups for multi-qubit controlled phase gates

Below, we prove the following theorem, showing the twirling groups constructed for multi-qubit controlled phase gates in the previous subsection not only satisfies $U\mathsf{G}U^\dagger = \mathsf{G}$, but also satisfies $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$.

Theorem 2. *For $U = C^nZ_m$ with $n \geq 1, m \geq 2$ and $nm \neq 2$, the twirling group $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$ satisfies $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$.*

For such quantum gates, our protocol can provide their fidelity estimation accurately instead of providing lower bounds. To prove Theorem 2, we present the following lemma.

Lemma 15. *As long as $\langle X, CZ, S\rangle \leq \mathsf{G}$ and U is diagonal in the computational basis, the equation $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ holds.*

Proof. In Pauli-Liouville representation, $\mathcal{U}$ is block-diagonal. Below, we investigate the diagonal blocks of $\mathcal{U}$. By definition,

$$\begin{aligned}\mathcal{U}_{ij} =& \operatorname{tr}(\sigma_i U(\sigma_j))\\ =& d\operatorname{tr}\big(\sigma_i U\sigma_j U^\dagger\sigma_j^\dagger\sigma_j\big)\\ =& d\operatorname{tr}\big(\sigma_j\sigma_i U\sigma_j U^\dagger\sigma_j^\dagger\big).\end{aligned} \tag{B51}$$

As U is diagonal in the computational basis or Z-basis and σ_j is proportional to a Pauli operator, the right part $U\sigma_jU^\dagger\sigma_j^\dagger$ is also diagonal in the computational basis. Thus, when the left part $\sigma_j\sigma_i$ is not diagonal, the trace of the above expression is 0. In general, we can express that $\sigma_i = \frac{1}{\sqrt{d}}X_{\mathbf{a}_i}Z_{\mathbf{b}_i}$ and $\sigma_j = \frac{1}{\sqrt{d}}X_{\mathbf{a}_j}Z_{\mathbf{b}_j}$ where $X_{\mathbf{a}_i}, X_{\mathbf{a}_j} \in \{\mathbb{I}, X\}^{\otimes n}$ and $Z_{\mathbf{b}_i}, Z_{\mathbf{b}_j} \in \{\mathbb{I}, Z\}^{\otimes n}$. The notation $\mathbf{a}_i$ is a bit string from $\{0,1\}^n$ meaning $X_{\mathbf{a}_i} = \prod_{k=1}^n X_k^{(\mathbf{a}_i)_k}$. The same are for $\mathbf{a}_j$, $\mathbf{b}_i$, and $\mathbf{b}_j$. If $X_{\mathbf{a}_i} \neq X_{\mathbf{a}_j}$, then $\sigma_j\sigma_i$ is not diagonal and hence $\mathcal{U}_{ij} = 0$. Thus, for any operator $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}$, there is a corresponding block with dimension $2^n \times 2^n$ in the Liouville representation of the diagonal matrix U. The linear space associated with the block is spanned by bases $\{X_{\mathbf{a}}\cdot I^{\otimes n}, X_{\mathbf{a}}\cdot(I^{\otimes n-1}\otimes Z), \cdots, X_{\mathbf{a}}\cdot Z^{\otimes n}\}$ and we denote the set of this bases as $X_{\mathbf{a}}\mathsf{Z}$. It is worth noting that the block spanned by Z is equal to identity with dimension 2^n, which means this block can be further decomposed into the direct sum of 2^n small blocks with dimensions 1.

Recall that Λ_G is also diagonal in Liouville representation, and all of its blocks are proportional to identities. If Λ_G has the same blocks as $\mathcal{U}$, they will commute with each other and hence $\mathcal{U}\Lambda_G\mathcal{U}^{-1} = \Lambda_G$ holds. Now we show that $\langle X, CZ, S\rangle \leq \mathbf{G}$ is enough for Λ_G to have the same blocks as $\mathcal{U}$. Since $\langle X, Z\rangle \leq \langle X, CZ, S\rangle \leq \mathsf{G}$, Λ_G must be diagonal in Pauli-Liouville representation, $\Lambda_G = \sum_i \lambda_i|\sigma_i\rangle\!\rangle\langle\!\langle\sigma_i|$. Note that if there exists an element $G \in \mathsf{G}$ such that $\sigma_i = G\sigma_jG^\dagger$, then σ_i and σ_j are symmetric under the group action of G, which results in $\lambda_i = \lambda_j$, or equivalently, σ_i and σ_j are in the same diagonal block [16]. Note that any $G \in \langle CZ, S\rangle \leq \mathsf{G}$ can be decomposed as $G = WV$ where $W \in \langle CZ\rangle, V \in \langle S\rangle$, then for any $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}/\mathbb{I}^{\otimes n}$,

$$\begin{aligned}GX_{\mathbf{a}}G^\dagger =& WVX_{\mathbf{a}}V^\dagger W^\dagger\\ =& X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger WX_{\mathbf{a}})(X_{\mathbf{a}}^\dagger VX_{\mathbf{a}})V^\dagger W^\dagger\\ =& X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger WX_{\mathbf{a}}W^\dagger)(X_{\mathbf{a}}^\dagger VX_{\mathbf{a}}V^\dagger).\end{aligned} \tag{B52}$$

Below we express $X_{\mathbf{a}} = \bigotimes_{i=1}^n X_i^{a_i}$ where $a_i \in \{0,1\}$ and we simply denote $\mathbf{a} = (a_1, a_2, \cdots, a_n)^T \in \{0,1\}^n$. As $X_{\mathbf{a}}$ is not equal to identity $\mathbb{I}^{\otimes n}$, there is at least an element in $\{a_1, a_2, \cdots, a_n\}$ to be nonzero. Suppose $a_i = 1$. For any $Z_{\mathbf{b}} \in \{\mathbb{I}, Z\}^{\otimes n}$ where $\mathbf{b} = (b_1, b_2, \cdots, b_n)^T \in \{0,1\}^n$, we can choose $W = \prod_{1\leq j\leq n, j\neq i} CZ_{ij}^{b_j}$, $V = S_i^{\sum_{1\leq j\leq n} a_jb_j}$ such that

$$\begin{aligned}X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger WX_{\mathbf{a}}W^\dagger)(X_{\mathbf{a}}^\dagger VX_{\mathbf{a}}V^\dagger) &= X_{\mathbf{a}}\Big(\prod_{1\leq j\leq n, j\neq i} Z_j^{b_j}Z_i^{a_jb_j}\Big)\Big(Z_i^{\sum_{1\leq j\leq n} a_jb_j}\Big)\\ &= X_{\mathbf{a}}\Big(\prod_{1\leq j\leq n, j\neq i} Z_j^{b_j}\Big)\Big(Z_i^{a_ib_i+2\sum_{1\leq j\leq n, j\neq i} a_jb_j}\Big)\\ &= X_{\mathbf{a}}\prod_{1\leq j\leq n} Z_j^{b_j}\\ &= X_{\mathbf{a}}Z_{\mathbf{b}}.\end{aligned} \tag{B53}$$

Here we utilize the following two identities,

$$X_iCZ_{i,j}X_iCZ_{i,j}^\dagger = Z_j, \tag{B54}$$

$$X_iS_iX_iS_i^\dagger = Z_i. \tag{B55}$$

Then, for any $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}/\mathbb{I}^{\otimes n}$ and $Z_{\mathbf{b}} \in \{\mathbb{I}, Z\}^{\otimes n}$, we can find a unitary $G = WV$, such that

$$GX_{\mathbf{a}}G^\dagger = X_{\mathbf{a}}Z_{\mathbf{b}}. \tag{B56}$$

It means that the linear space spanned by $X_{\mathbf{a}}\mathsf{Z}$ is a diagonal block for Λ_G, which is the same as $\mathcal{U}$. In the space spanned by Z, Λ_G and $\mathcal{U}$ are both fully diagonal with 2^n one-dimensional blocks. Thus, we prove that $\langle X, CZ, S\rangle \le \mathsf{G}$ suffices to make Λ_G commute with the diagonal matrix $\mathcal{U}$ and complete the proof of Lemma 15. □

Now we present the proof of Theorem 2.

Proof. The situation of $m = 2$ and that of $m \ge 3$ are different. Below, we first discuss the case that $m \ge 3$. The main technique is following the discussion in the proof of Theorem 1 and mapping W_X to $\mathbb{Z}_m^{2^{n+1}}$ via injective map ϕ in Eq. (B47).

Suppose $m \ge 3$. In the case that m is odd for $U = C^n Z_m$, from the proof of Theorem 1, we obtain that $\phi(\mathsf{W}_X)$ is spanned by $\{e_i, 0 \le i \le 2^{n+1} - 1\}$. It means that $\phi(\mathsf{W}_X)$ is equal to $\mathbb{Z}_m^{2^{n+1}}$. Note that $\phi(C^n Z_m) = e_{2^{n+1}-1}$, so in this case the target gate U itself belongs to W_X and hence U belongs to the twirling group $\mathsf{G} = \mathsf{X} \ltimes \mathsf{W}_X$. When $U \in \mathsf{G}$, then

$$\begin{aligned}\mathcal{U}\Lambda_G\mathcal{U}^\dagger &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{U}\mathcal{G}\Lambda\mathcal{G}^\dagger\mathcal{U}^\dagger \\ &= \mathbb{E}_{G\in\mathsf{G}}\mathcal{G}\Lambda\mathcal{G}^\dagger \\ &= \Lambda_G.\end{aligned} \tag{B57}$$

The commutation between $\mathcal{U}$ and Λ_G is satisfied automatically.

Below, we discuss the case that m is even and $m \ge 3$ for $U = C^n Z_m$. Based on Lemma 15, we study whether CZ and S gates belong to W_X or not. By symmetry, we only need to study that for the phase gate on the first qubit S_1 and the controlled-phase gate on the first two qubits CZ_{12}. Note that the phase gate S can be expressed as $S = Z_m^{m/4}$ and CZ can be expressed as $CZ = CZ_m^{m/2}$. We express $m = q2^k$ where $q \ge 1$ is odd and $k \ge 1$ is a positive integer. Note that the bases of $\phi(\mathsf{W}_X)$ are $\{v_i = e_i - e_{2^{n+1}-1}, 0 \le i \le 2^{n+1} - 2\}$ along with $u = 2^{\min(n+1,k)}e_{2^{n+1}-1}$. Then, $\phi(S_1)$ can be constructed by

$$\begin{aligned}\phi(S_1) &= q2^{k-2}(e_{2^{n+1}-2^n} + e_{2^{n+1}-2^n+1} + \cdots + e_{2^{n+1}-1}) \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) + q2^{k-2}(2^n - 1)e_{2^{n+1}-1} + q2^{k-2}e_{2^{n+1}-1} \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) + q2^{n+k-2}e_{2^{n+1}-1} \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) + q2^{n+k-2-\min(n+1,k)}u \\ &= q2^{k-2}(v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2}) + q2^{\max(k-3,n-2)}u.\end{aligned} \tag{B58}$$

Meanwhile, $\phi(CZ_{12})$ can be constructed by

$$\begin{aligned}\phi(CZ_{12}) &= q2^{k-1}(e_{2^{n+1}-2^{n-1}} + e_{2^{n+1}-2^{n-1}+1} + \cdots + e_{2^{n+1}-1}) \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) + q2^{k-1}(2^{n-1} - 1)e_{2^{n+1}-1} + q2^{k-1}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) + q2^{n+k-2}e_{2^{n+1}-1} \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) + q2^{n+k-2-\min(n+1,k)}u \\ &= q2^{k-1}(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) + q2^{\max(k-3,n-2)}u.\end{aligned} \tag{B59}$$

The constructions Eq. (B58) and Eq. (B59) are valid only when the coefficients of the bases are integers. Obviously, $k \ge 3$, and $k = 2$ and $n \ge 2$ can make the coefficients to be integers. In this case, S_1 and CZ_{12} belong to W_X and furthermore, $\langle CZ, X, S\rangle \le \mathsf{G}$. Based on Lemma 15, $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ is satisfied. The exceptions are the case that $k = 1$ and that $n = 1$ and $k = 2$. Below, we discuss the two kinds of exceptions in detail.

When $k = 1$, the target gate $U = C^n Z_m$ and $m = 2q$. In this case, $u = 2e_{2^{n+1}-1}$, and $CZ_{m/2}$ and Z_m gates belong to G, which can be derived through the following equations.

$$\begin{aligned}\phi((Z_m)_1) &= e_{2^{n+1}-2^n} + e_{2^{n+1}-2^n+1} + \cdots + e_{2^{n+1}-1} \\ &= v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2} + (2^n - 1)e_{2^{n+1}-1} + e_{2^{n+1}-1} \\ &= v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2} + 2^n e_{2^{n+1}-1} \\ &= v_{2^{n+1}-2^n} + v_{2^{n+1}-2^n+1} + \cdots + v_{2^{n+1}-2} + 2^{n-1}u.\end{aligned} \tag{B60}$$

$$\begin{aligned}\phi((CZ_{m/2})_{12}) &= 2(e_{2^{n+1}-2^{n-1}} + e_{2^{n+1}-2^{n-1}+1} + \cdots + e_{2^{n+1}-1}) \\ &= 2(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) + 2(2^{n-1}-1)e_{2^{n+1}-1} + 2e_{2^{n+1}-1} \\ &= 2(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) + 2^n e_{2^{n+1}-1} \\ &= 2(v_{2^{n+1}-2^{n-1}} + v_{2^{n+1}-2^{n-1}+1} + \cdots + v_{2^{n+1}-2}) + 2^{n-1}u.\end{aligned} \tag{B61}$$

Thus, when $k = 1$, $\langle CZ_{m/2}, Z_m\rangle \leq \mathsf{W}_X$ and $\mathsf{X} \ltimes \langle CZ_{m/2}, Z_m\rangle \leq \mathsf{G}$. Notice that $m \geq 3$ and $m/2$ is odd, so $m/2 \geq 3$ holds. It can be verified that the irreducible representation decomposition of the Liouville representation for group $\langle CZ_m, Z_m, X\rangle$, where $m \geq 3$ and m is odd, and group $\langle CZ, S, X\rangle$ are the same. As $\mathsf{X} \ltimes \langle CZ_{m/2}, Z_{m/2}\rangle \leq \mathsf{X} \ltimes \langle CZ_{m/2}, Z_m\rangle \leq \mathsf{G}$ and $m/2 \geq 3$, after being twirled by G, the twirled noise channel Λ_G would have the same blocks as $\mathcal{U}$, and $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ is satisfied.

When $n = 1$ and $k = 2$, the target gate $U = CZ_m$ and $m = 4q$. The bases for $\phi(\mathsf{W}_X)$ are $v_0 = (1,0,0,-1)^T$, $v_1 = (0,1,0,-1)^T$, $v_2 = (0,0,1,-1)^T$, and $u = (0,0,0,4)$. Note that $\phi(CZ \cdot S_1) = (0,0,q,3q)^T = v_2 + u$ and $\phi(CZ \cdot S_2) = (0,q,0,3q)^T = v_1 + u$. Thus, in this case, $\langle CZ \cdot S_1, CZ \cdot S_2\rangle \in \mathsf{W}_X \leq \mathsf{G}$. In the proof of Theorem 1, we have already shown that Pauli group $\mathsf{P}_2 \in \mathsf{G}$. Thus, the twirled noise channel would be diagonal in Liouville representation, $\Lambda_G = \sum_i \lambda_i |\sigma_i\rangle\rangle\langle\langle\sigma_i|$. Also, we have the following identities,

$$CZ \cdot S_1 X_1 (CZ \cdot S_1)^\dagger = X_1 Z_1 Z_2; \tag{B62}$$

$$CZ \cdot S_2 X_1 (CZ \cdot S_2)^\dagger = X_1 Z_2; \tag{B63}$$

$$(CZ \cdot S_2 \cdot CZ \cdot S_1) X_1 (CZ \cdot S_2 \cdot CZ \cdot S_1)^\dagger = X_1 Z_1. \tag{B64}$$

Under the twirling of $\langle CZ \cdot S_1, CZ \cdot S_2\rangle$, X_1, X_1Z_1, X_1Z_2, and $X_1Z_1Z_2$ would be symmetric. Their corresponding Pauli fidelities λ_i would be the same after the twirling. As the twirling group contains $\langle CZ \cdot S_1, CZ \cdot S_2\rangle$, the twirled noise channel Λ_G would be diagonal and proportional to identity in the space spanned by $X_1\mathsf{Z}$. The cases for spaces spanned by $X_2\mathsf{Z}$ or $X_1X_2\mathsf{Z}$ are the same. In summary, Λ_G has diagonal blocks in spaces spanned by $X_1\mathsf{Z}$, $X_2\mathsf{Z}$ and $X_1X_2\mathsf{Z}$, respectively. Thus, in this case, $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ is also satisfied.

At last, we analyze the case that $m = 2$ and $n \geq 2$. The case that $m = 2$ and $n = 1$ is just CZ and obviously in this case $\mathcal{U}\Lambda_G\mathcal{U}^\dagger \neq \Lambda_G$. When the target gate is C^nZ with $n \geq 2$, the twirling group is $\mathsf{G} = X \ltimes W_X = \langle C^{n-1}Z, \cdots, Z, X\rangle$. Different from a generic diagonal gate, the diagonal blocks spanned by $X_{\mathbf{a}}\mathsf{Z}$ of C^nZ gate in Pauli-Liouville representation can be further divided into two small blocks. We will show that any channel twirled by group $\langle CZ, Z, X\rangle$ would have the same diagonal blocks as C^nZ in Liouville representation. As $\langle C^{n-1}Z, \cdots, Z, X\rangle$ contains $\langle CZ, Z, X\rangle$, $\Lambda_{\langle C^{n-1}Z,\cdots,Z,X\rangle}$ would commute with C^nZ.

Similar to Eq. (B51), by setting $\sigma_i = \frac{1}{\sqrt{d}} X_{\mathbf{a}_i} Z_{\mathbf{b}_i}$ and $\sigma_j = \frac{1}{\sqrt{d}} X_{\mathbf{a}_j} Z_{\mathbf{b}_j}$ where $X_{\mathbf{a}_i}, X_{\mathbf{a}_j} \in \{\mathbb{I}, X\}^{\otimes n}$ and $Z_{\mathbf{b}_i}, Z_{\mathbf{b}_j} \in \{\mathbb{I}, Z\}^{\otimes n}$, we get the matrix element of the Liouville representation of $U = C^nZ$,

$$\begin{aligned}\mathcal{U}_{ji} &= d\,\mathrm{tr}\big(\sigma_i\sigma_j C^nZ\sigma_i C^nZ\sigma_i^\dagger\big) \\ &= \frac{1}{d}\,\mathrm{tr}\big(X_{\mathbf{a}_i} Z_{\mathbf{b}_i} X_{\mathbf{a}_j} Z_{\mathbf{b}_j} W\big) \\ &\propto \frac{1}{d}\,\mathrm{tr}\big(X_{\mathbf{a}_i} X_{\mathbf{a}_j} Z_{\mathbf{b}_i} Z_{\mathbf{b}_j} W\big); \\ W &= C^nZ X_{\mathbf{a}_i} C^nZ X_{\mathbf{a}_i}.\end{aligned} \tag{B65}$$

Notice that to make $\mathcal{U}_{ji}$ nonzero, we at least require $X_{\mathbf{a}_i} = X_{\mathbf{a}_j}$. Define support of $X_{\mathbf{a}_i}$ to be the set, $\mathrm{supp}(X_{\mathbf{a}_i}) = \{i | X_{\mathbf{a}_i} \text{ on qubit } i \text{ is } X\}$, which is essentially the location of 1 in $\mathbf{a}_i$. Notice that two of the diagonal elements of W are -1 while others are all 1. Suppose these two elements are W_{i_1,i_1} and W_{i_2,i_2}. Without loss of generality, i_2 can be set as $2^{n+1} - 1$ as the element of the last row and last column is -1. To make $\mathcal{U}_{ji}$ nonzero, we furthermore require that $(Z_{\mathbf{b}_i} Z_{\mathbf{b}_j})_{i_1,i_1} = (Z_{\mathbf{b}_i} Z_{\mathbf{b}_j})_{i_2,i_2}$, or $\langle i_1| Z_{\mathbf{b}_i} Z_{\mathbf{b}_j} |i_1\rangle = \langle i_2| Z_{\mathbf{b}_i} Z_{\mathbf{b}_j} |i_2\rangle$. Note that $|i_2\rangle = X_{\mathbf{a}_i} |i_1\rangle$. Thus, to make $\mathcal{U}_{ji}$ nonzero, we require that $X_{\mathbf{a}_i}$ and $Z_{\mathbf{b}_i} Z_{\mathbf{b}_j}$ commute, which is equivalent to that $Z_{\mathbf{b}_i} Z_{\mathbf{b}_j}$ has even number of Z gates on qubits of $\mathrm{supp}(X_{\mathbf{a}_i})$. Hence, the block spanned by $X_{\mathbf{a}_i}\mathsf{Z}$ can be further split into two blocks when $m = 2$. One block is spanned by $X_{\mathbf{a}_i}\mathsf{Z}_e = \{X_{\mathbf{a}_i} Z_{\mathbf{b}_k},\ \mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 0\}$ and the other block is spanned by $X_{\mathbf{a}_i}\mathsf{Z}_o = \{X_{\mathbf{a}_i} Z_{\mathbf{b}_k}, \mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 1\}$. Here, $\mathbf{a}_i \cap \mathbf{b}_k$ is the bitwise and operation between $\mathbf{a}_i$ and $\mathbf{b}_k$, $|\mathbf{a}_i \cap \mathbf{b}_k|$ is the weight of $\mathbf{a}_i \cap \mathbf{b}_k$, and $\mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 0$ means that there are even number of Z gates of $Z_{\mathbf{b}_k}$ acting on qubits of $\mathrm{supp}(X_{\mathbf{a}_i})$. Notice that $\mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 0$ is equivalent to that $\sum_{1\leq j\leq n} (\mathbf{a}_i)_j (\mathbf{b}_k)_j$ is even and $\mod(|\mathbf{a}_i \cap \mathbf{b}_k|, 2) = 1$ is equivalent to that $\sum_{1\leq j\leq n} (\mathbf{a}_i)_j (\mathbf{b}_k)_j$ is odd. These relations are useful in the following proof.

Now we obtain that the Liouville representation of $\mathcal{U}$ is spanned by blocks with bases $X_{\mathbf{a}}\mathsf{Z}_e$ and $X_{\mathbf{a}}\mathsf{Z}_o$. To show $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$, we also need to check whether Λ_G has the same diagonal blocks with $\mathcal{U}$ and whether Λ_G is proportional to identity when restricting in these blocks. The proof is similar to that in Lemma 15. Recall that $\Lambda_G = \sum_i \lambda_i |\sigma_i\rangle\rangle\langle\langle\sigma_i|$

in Liouville representation. Also, if there exists an element $G \in \mathsf{G}$ such that $\sigma_i = G\sigma_j G^\dagger$, then $\lambda_i = \lambda_j$, and σ_i and σ_j would be in the same blocks. Then, we only need to check whether any element in $X_{\mathbf{a}}\mathsf{Z}_e$ can be generated from $X_{\mathbf{a}}$ by conjugate action of elements in G.

Given $X_{\mathbf{a}} \in \{\mathbb{I}, X\}^{\otimes n}/\mathbb{I}^{\otimes n}$, choose i from $\mathrm{supp}(X_{\mathbf{a}})$. For any $Z_{\mathbf{b}} \in \{\mathbb{I}, Z\}^{\otimes n}$ where $X_{\mathbf{a}}Z_{\mathbf{b}} \in X_{\mathbf{a}}\mathsf{Z}_e$, we can choose $W = \prod_{j\neq i} CZ_{ij}^{b_j}$, such that

$$\begin{aligned} WX_{\mathbf{a}}W^\dagger =& X_{\mathbf{a}}(X_{\mathbf{a}}^\dagger W X_{\mathbf{a}} W^\dagger) \\ =& X_{\mathbf{a}}(\prod_{1\le j\le n, j\neq i} Z_j^{b_j}) Z_i^{\sum_{1\le j\le n, j\neq i} a_j b_j} \\ =& X_{\mathbf{a}}(\prod_{1\le j\le n, j\neq i} Z_j^{b_j}) Z_i^{(\sum_{1\le j\le n} a_j b_j) - a_i b_i} \\ =& X_{\mathbf{a}}(\prod_{1\le j\le n, j\neq i} Z_j^{b_j}) Z_i^{-a_i b_i} \\ =& X_{\mathbf{a}}(\prod_{1\le j\le n, j\neq i} Z_j^{b_j}) Z_i^{b_i} \\ =& X_{\mathbf{a}} Z_{\mathbf{b}}. \end{aligned} \tag{B66}$$

Here, in the fourth line, we utilize $X_{\mathbf{a}}Z_{\mathbf{b}} \in X_{\mathbf{a}}\mathsf{Z}_e$, which means $\sum_{1\le j\le n} a_j b_j$ is even. In the fifth line, we use the condition that $a_i = 1$. Thus, we show that Λ_G is proportional to identity when restricting in the block spanned by $X_{\mathbf{a}}\mathsf{Z}_e$. From a similar argument, it can be verified that this is also true for block spanned by $X_{\mathbf{a}}\mathsf{Z}_o$. Then, we successfully show that $\mathcal{U}$ and Λ_G have the same diagonal blocks, and $\mathcal{U}\Lambda_G\mathcal{U}^{-1} = \Lambda_G$ when $m = 2$.

As a conclusion, the twirling group $X \ltimes \mathsf{W}_X$ satisfies $\mathcal{U}\Lambda_G\mathcal{U}^\dagger = \Lambda_G$ for $U = C^n Z_m$ with $n = 1, m \ge 3$ or $n \ge 2, m \ge 2$. Proof is done. □

7. Structure of W_X

In this part, we present the concrete form of W_X. Following the discussion of W_X in previous subsections, we map W_X into $\mathbb{Z}_m^{2^{n+1}}$ via Eq. (B47).

We first discuss the cardinality of W_X. The image $\phi(\mathsf{W}_X)$ is the linear span of bases $\{v_0 = e_0 - e_{2^{n+1}-1}, \cdots, v_{2^{n+1}-2} = e_{2^{n+1}-2} - e_{2^{n+1}-1}\}$. If considering the factor of global phase, the image would be the linear span of bases $\{v_0 = e_0 - e_{2^{n+1}-1}, \cdots, v_{2^{n+1}-2} = e_{2^{n+1}-2} - e_{2^{n+1}-1}, u = 2^{\min(n+1,k)} e_{2^{n+1}-1}\}$ and we denote this enlarged image to be $\phi'(\mathsf{W}_X)$. In $\phi'(\mathsf{W}_X)$, any two vectors differing by $v_{2^{n+1}-1} = \begin{pmatrix}1 & \cdots & 1\end{pmatrix}^T$ correspond to the same gate. If $n+1 \ge k$, u is $\begin{pmatrix}0 & \cdots & 2^k\end{pmatrix}^T$. Now we consider the cardinality of $\phi'(\mathsf{W}_X)$. One can take arbitrary values from $\mathbb{Z}_m$ to fix the coefficients of the first $2^{n+1}-1$ bases. After that, the coefficient of the last basis u only has $\frac{m}{2^k}$ inequivalent choices from $\mathbb{Z}_m$ as $(a + \frac{m}{2^k})u = au$. Hence, the cardinality of $\phi'(\mathsf{W}_X)$ is

$$|\phi'(\mathsf{W}_X)| = m^{2^{n+1}-1} \cdot \frac{m}{2^k} = m^{2^{n+1}}/2^k. \tag{B67}$$

For any $\phi(w)$ belongs to $\phi'(\mathsf{W}_X)$, we can also find $\phi(w) + \begin{pmatrix}1 & \cdots & 1\end{pmatrix}^T, \cdots, \phi(w) + (m-1)\cdot\begin{pmatrix}1 & \cdots & 1\end{pmatrix}^T$ in $\phi'(\mathsf{W}_X)$. All these m terms correspond to the same gate. Thus, the cardinality of $\phi(\mathsf{W}_X)$ is just $|\phi'(\mathsf{W}_X)|$ divided by m, excluding the redundacy of the global phase,

$$|\mathsf{W}_X| = |\phi(\mathsf{W}_X)| = \frac{1}{m}|\phi'(\mathsf{W}_X)| = m^{2^{n+1}-1}/2^k. \tag{B68}$$

Now we investigate the structure of $\mathsf{W}_X = \langle \Pi^\dagger C^n Z_m \Pi C^n Z_m^\dagger \rangle$. We first analyze the simple case of $m = q$ and $m = 2q$ where q is an odd number. Then, we use induction to analyze the general case that $m = q2^k$.

1) Case 1: $m = q$.

In this case, $k = 0$, $\gcd(m, 2^{n+1}) = 1$, the vector $u = e_{2^{n+1}-1}$, and $\phi'(\mathsf{W}_X) = \mathbb{Z}_m^{2^{n+1}}$. Clearly, for any $0 \le l \le n$, $\phi(C^l Z_m) \in \phi'(\mathsf{W}_X)$, so we can obtain that $\langle C^n Z_m, C^{n-1}Z_m, \cdots, CZ_m, Z_m\rangle \le \mathsf{W}_X$. Since $\langle C^n Z_m, C^{n-1}Z_m, \cdots, CZ_m, Z_m\rangle = \langle C^n Z_m\rangle \times \langle C^{n-1}Z_m\rangle \times \cdots \times \langle Z_m\rangle$, $\left|\langle C^n Z_m, C^{n-1}Z_m, \cdots, CZ_m, Z_m\rangle\right| = m^{\binom{n+1}{n+1}} \times m^{\binom{n+1}{n}} \times \cdots \times m^{\binom{n+1}{1}} = m^{2^{n+1}-1} = |\mathsf{W}_X|$. The cardinalities of W_X and its subgroup $\langle C^n Z_m, C^{n-1}Z_m, \cdots, CZ_m, Z_m\rangle$ are the same, hence, we can conclude that

$$\mathsf{W}_X = \langle C^n Z_m, C^{n-1}Z_m, \cdots, CZ_m, Z_m\rangle = \langle C^n Z_m\rangle \times \langle C^{n-1}Z_m\rangle \times \cdots \times \langle Z_m\rangle. \tag{B69}$$

2) Case 2: $m = 2q$.
In this case, $k = 1$, $\gcd(m, 2^{n+1}) = 2$, and $u = 2e_{2^{n+1}-1}$. We can use the following formula to construct all $C^l Z_m$ with $0 \le l \le n-1$,

$$C^l Z_m = \phi^{-1}\Big(v_{(2^{l+1}-1)*2^{n-l}} + v_{(2^{l+1}-1)*2^{n-l}+1} + \cdots + v_{2^{n+1}-2} + 2^{n-l-1} \cdot u\Big). \tag{B70}$$

But this construction is not valid for $C^n Z_m$ as it requires $\frac{1}{2}u$. We can only construct $C^n Z_m^2 = C^n Z_{m/2}$. This implies that $\langle C^n Z_{m/2}, C^{n-1} Z_m, C^{n-2} Z_m, \cdots\rangle \le \mathsf{W}_X$. Since $\langle C^n Z_{m/2}, C^{n-1} Z_m, \cdots\rangle = \langle C^n Z_{m/2}\rangle \times \langle C^{n-1} Z_m\rangle \times \cdots \times \langle Z_m\rangle$, $\big|\langle C^n Z_{m/2}, C^{n-1} Z_m, \cdots, C Z_m, Z_m\rangle\big| = (m/2)^{\binom{n+1}{n+1}} \times m^{\binom{n+1}{n}} \times \cdots \times m^{\binom{n+1}{1}} = m^{2^{n+1}-1}/2 = |\mathsf{W}_X|$. Thus, in this case,

$$\mathsf{W}_X = \langle C^n Z_{m/2}, C^{n-1} Z_m, \cdots, C Z_m, Z_m\rangle = \langle C^n Z_{m/2}\rangle \times \langle C^{n-1} Z_m\rangle \times \cdots \times \langle Z_m\rangle. \tag{B71}$$

3) Case 3.1: $m = q2^k$ with $k \le n$.
We first give the result and prove it by induction.

Theorem 3. *If $m = q2^k$ and $k \le n$, the structure of W_X can be expressed in the following recursion,*

$$\begin{aligned} \mathsf{W}_X &= \langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots, Z_m\rangle \times \langle A_k\rangle; \\ A_0 &= \{\mathbb{I}\}, A_1 = \{C^n Z_{m/2}\}, \\ \forall 2 \le k \le n, A_k &= \{C^{n-k+1} Z_{m/2}, g C^{n-k+1} Z_m | g \in A_{k-1}\}. \end{aligned} \tag{B72}$$

Proof. The proof idea is showing $|\mathsf{W}_X| = |\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots, Z_m\rangle \times \langle A_k\rangle|$ and $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots, Z_m\rangle \times \langle A_k\rangle \le \mathsf{W}_X$.

Let us start with studying the relation between $\langle A_{k-1}\rangle$ and $\langle A_k\rangle$. For any $g \in \langle A_{k-1}\rangle$, it can be decomposed as $g = g_1^{\lambda_1} g_2^{\lambda_2} \cdots$ where $g_1, g_2, \cdots$ all belong to A_{k-1} and $\lambda_1, \lambda_2, \cdots$ are integers. Since $g_1 C^{n-k+1} Z_m, g_2 C^{n-k+1} Z_m, \cdots \in A_k$, we can conclude that $g' = (g_1 C^{n-k+1} Z_m)^{\lambda_1} (g_2 C^{n-k+1} Z_m)^{\lambda_2} \cdots = g C^{n-k+1} Z_m^{\sum_i \lambda_i} \in \langle A_k\rangle$. Suppose $\sum_i \lambda_i$ is an even number, since $C^{n-k+1} Z_{m/2} = C^{n-k+1} Z_m^2 \in A_k$, the set $\{g, g C^{n-k+1} Z_m^2, g C^{n-k+1} Z_m^4, \cdots\} \subseteq \langle A_k\rangle$. Now we want to argue that any gate has the form of $g C^{n-k+1} Z_m^{2p+1}$ cannot be in $\langle A_k\rangle$. This claim can be proved by contradiction. If $g C^{n-k+1} Z_m^{2p+1} \in \langle A_k\rangle$, combining with $g C^{n-k+1} Z_m^{2p} \in \langle A_k\rangle$, it is clear that $C^{n-k+1} Z_m = (g C^{n-k+1} Z_m^{2p})^\dagger g C^{n-k+1} Z_m^{2p+1} \in \langle A_k\rangle$. It leads to a contradiction as $C^{n-k+1} Z_m$ cannot be constructed with $\{v_i, 0 \le i \le 2^{n+1}-2\}$ and $u = 2^k e_{2^{n+1}-1}$. Similarly, if $\sum_i \lambda_i$ is an odd number, we can see that all elements have the form $g C^{n-k+1} Z_m^{2p+1}$ will belong to $\langle A_k\rangle$ while elements within form of $g C^{n-k+1} Z_m^{2p}$ will not. In either case, one element g in $\langle A_{k-1}\rangle$ will contribute to $|\langle C^{n-k+1} Z_{m/2}\rangle| = \frac{1}{2}|\langle C^{n-k+1} Z_m\rangle|$ elements in $\langle A_k\rangle$, which implies that $|\langle A_k\rangle| = \frac{1}{2}|\langle A_{k-1}\rangle| \times |\langle C^{n-k+1} Z_m\rangle|$. By iteration, the cardinality of $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots\rangle \times \langle A_k\rangle$ is

$$\begin{aligned} |\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots\rangle \times \langle A_k\rangle| =& \frac{1}{2}|\langle C^{n-k+1} Z_m, C^{n-k} Z_m, \cdots\rangle \times \langle A_{k-1}\rangle| \\ =& \frac{1}{2^k}|\langle C^n Z_m, C^{n-1} Z_m, \cdots\rangle| \\ =& m^{2^{n+1}-1}/2^k. \end{aligned} \tag{B73}$$

This is exactly equal to $|\mathsf{W}_X|$. To complete the proof, we only need to argue that $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots\rangle \times \langle A_k\rangle \le \mathsf{W}_X$. Note that $\phi'(\mathsf{W}_X)$ is spanned by bases $v_i, 1 \le i \le 2^{n+1}-2$ and u. We only need to investigate whether the generators of $\langle C^{n-k} Z_m, C^{n-k-1} Z_m, \cdots\rangle \times \langle A_k\rangle$ belong to the preimage set of ϕ.

For $C^{n-k-l} Z_m$ with $0 \le l \le n-k$, it can be expressed in the form of

$$\begin{aligned} C^{n-k-l} Z_m &= \phi^{-1}\Big(2^l \begin{pmatrix} 0 & \cdots & 2^k \end{pmatrix}^T + \sum_{i=(2^{n+1-k-l}-1)\cdot 2^{k+l}}^{2^{n+1}-2} v_i\Big) \\ &= \phi^{-1}\Big(2^l u + \sum_{i=(2^{n+1-k-l}-1)\cdot 2^{k+l}}^{2^{n+1}-2} v_i\Big). \end{aligned} \tag{B74}$$

Thus, $C^{n-k-l} Z_m$ belongs to W_X. Next we prove $\langle A_k\rangle \le \mathsf{W}_X$, or equivalently, $A_k \subseteq \mathsf{W}_X$, by induction. Denote $\mathsf{W}_X^k = \langle \Pi^\dagger C^n Z_{q2^k} \Pi C^n Z_{q2^k}^\dagger\rangle$. Suppose $A_{k-1} \subseteq \mathsf{W}_X^{k-1}$. Then, any element $g \in A_{k-1}$ has the following expression,

$$g = \phi^{-1}\Big(\begin{pmatrix} 0 & \cdots & 2^{k-1} \end{pmatrix}^T + \sum_{i=0}^{2^{n+1}-2} c_i v_i\Big). \tag{B75}$$

Note that $C^{n-k+1}Z_m$ can also be expressed in the above form as

$$C^{n-k+1}Z_m = \phi^{-1}\Big(\big(0 \ \cdots \ 2^{k-1}\big)^T + \sum_{i=(2^{n-k+2}-1)\cdot 2^{k-1}}^{2^{n+1}-2} v_i\Big). \tag{B76}$$

This implies that $C^{n-k+1}Z_m^2$ and $C^{n-k+1}Z_m g$ can be expressed as below.

$$C^{n-k+1}Z_m^2 = \phi^{-1}\Big(\big(0 \ \cdots \ 2^{k}\big)^T + \sum_{i=(2^{n-k+2}-1)\cdot 2^{k-1}}^{2^{n+1}-2} 2v_i\Big); \tag{B77}$$

$$gC^{n-k+1}Z_m = \phi^{-1}\Big(\big(0 \ \cdots \ 2^{k}\big)^T + \sum_{i=0}^{(2^{n-k+2}-1)\cdot 2^{k-1}-1} c_i v_i + \sum_{i=(2^{n-k+2}-1)\cdot 2^{k-1}}^{2^{n+1}-2} (c_i+1)v_i\Big). \tag{B78}$$

Therefore, any element $g' \in A_k$ has the expression,

$$g' = \phi^{-1}\Big(\big(0 \ \cdots \ 2^{k}\big)^T + \sum_{i=0}^{2^{n+1}-2} c_i v_i\Big). \tag{B79}$$

It means that all elements of A_k can be constructed with u and v_i, and belong to W_X^k. Now we have shown that $A_{k-1} \subseteq \mathsf{W}_X^{k-1}$ implies $A_k \subseteq \mathsf{W}_X^k$. For the induction argument to hold, we only need to check the initial condition, i.e., $A_1 \subseteq \mathsf{W}_X^1$. This reduces to Case 2. Thus, we prove that $A_k \subseteq \mathsf{W}_X$ and $\langle A_k\rangle \leq \mathsf{W}_X$. Hence, $\langle C^{n-k}Z_m, C^{n-k-1}Z_m, \cdots, Z_m\rangle \times \langle A_k\rangle \leq \mathsf{W}_X$. Combining with Eq. (B73), we obtain that $\mathsf{W}_X = \langle C^{n-k}Z_m, C^{n-k-1}Z_m, \cdots\rangle \times \langle A_k\rangle$ and proof is done. □

4) Case 3.2: $m = q2^k$ with $k > n$.
In this case, $\gcd(2^{n+1}, m) = 2^{n+1}$ and $\mathsf{W}_X = \langle A_{n+1}\rangle$ as described in the previous case.

Below we present examples of the twirling groups for target gate U to be C^nZ and CZ_m, which have been mentioned in the main text. In the next subsection, we will analyze their sample complexity and computational complexity.

Example 2. *The optimal twirling group* G *in CRU for multi-qubit controlled* Z *gate* C^nZ *is*

$$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle = \mathsf{X} \ltimes \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z\rangle. \tag{B80}$$

Example 3. *The optimal twirling group* G *in CRU for controlled phase gate* CZ_m *is*

$$\langle CZ_m, Z_m, X\rangle = \mathsf{X} \ltimes \langle CZ_m, Z_m\rangle, \tag{B81}$$

if m *is odd, and*

$$\langle CZ_{m/2}, Z_m, X\rangle = \mathsf{X} \ltimes \langle CZ_{m/2}, Z_m\rangle, \tag{B82}$$

if m *is even and* $m/2$ *is odd, and*

$$\langle CZ_{m/4}, Z_{m/2}, X, CZ_{m/2}Z_m^1, CZ_{m/2}Z_m^2\rangle = \mathsf{X} \ltimes \langle CZ_{m/4}, Z_{m/2}, CZ_{m/2}Z_m^1, CZ_{m/2}Z_m^2\rangle, \tag{B83}$$

if $m/2$ *is even. It is worth mentioning that* $\langle CZ_{m/4}, Z_{m/2}, CZ_{m/2}Z_m^1, CZ_{m/2}Z_m^2\rangle$ *is a subgroup of* $\langle CZ_{m/2}, Z_m\rangle$.

8. Complexity analysis

To enable randomized benchmarking or any other quantum information tasks with twirling groups, one needs to sample from the group and compute the multiplication of group elements. It is necessary to analyze the sample complexity and the computational complexity of the twirling group. The sample complexity is directly related to the cardinality of the corresponding twirling group. The computational complexity is related to both the group structure and the algorithm for computing the multiplication. We will provide a group multiplication algorithm and present its complexity as the upper bound of the group computational complexity. In the discussion below, we distinguish the number of qubits N and the number of controlled qubits n. In general, $n \leq N-1$.

To benchmark C^nZ gate, the twirling group is $\mathsf{G}_{C^nZ} = \langle C^{n-1}Z, C^{n-2}Z\cdots, CZ, Z, X\rangle$ shown in Example 2. Note that this group has a semi-direct product structure,

$$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle = \langle X\rangle \rtimes (\langle C^{n-1}Z\rangle \times \langle C^{n-2}Z\rangle \times \cdots \langle Z\rangle). \tag{B84}$$

This means that any element in G_{C^nZ} can be written in the form of $\Pi W_{n-1}W_{n-2}\cdots W_1W_0$ where $\Pi \in \langle X\rangle$ and $W_l \in \langle C^lZ\rangle$. To sample from this twirling group G_{C^nZ}, we just need to sample independently from $\langle C^{n-1}Z\rangle$, $\langle C^{n-2}Z\rangle$, $\cdots$, $\langle Z\rangle$, and $\langle X\rangle$. Since the orders of all generators are 2, we can use a binary string whose length equals the number of generators of the group to represent an arbitrary group element. Then, we can sample the group element by sampling the binary string. For instance, the generators of $\langle CZ\rangle$ are $\{CZ_{1,2}, CZ_{1,3}, \cdots, CZ_{1,N}, \cdots, CZ_{2,3}, \cdots, CZ_{N-1,N}\}$. So a binary string with length $\binom{N}{2}$ is enough for sampling from $\langle CZ\rangle$. The total length needed to sample G_{C^nZ} is given by

$$\log|\mathsf{G}_{C^nZ}| = \log\left|\langle X\rangle \rtimes \left(\langle C^{n-1}Z\rangle \times \cdots \times \langle CZ\rangle \times \langle Z\rangle\right)\right| = n + \sum_{i=1}^{n}\binom{N}{i} = O(N^n), \tag{B85}$$

which is indeed the sample complexity. For computing the inverse gates, we will utilize the following identity. Given a C^nZ gate acting on qubits $i_1, \cdots, i_{n+1}$ and a subset of $\{i_1, \cdots, i_{n+1}\}$, I, then

$$C^nZ_{i_1,i_2,\cdots,i_{n+1}}\Pi_{i\in I}X_i = \Pi_{i\in I}X_i\Pi_{S\subseteq I}C^{n-|S|}Z_{\{i_1,\cdots,i_{n+1}\}\setminus S}. \tag{B86}$$

So, the multiplication of two elements in the group is calculated by

$$\begin{aligned}
&\Pi^{(1)}W_{n-1}^{(1)}\cdots W_0^{(1)}\Pi^{(2)}W_{n-1}^{(2)}\cdots W_0^{(2)}\\
=&(-1)^{\Pi^{(2)}\cdot W_0^{(1)}}\Pi^{(1)}W_{n-1}^{(1)}\cdots W_1^{(1)}\Pi^{(2)}W_0^{(1)}W_{n-1}^{(2)}\cdots W_0^{(2)}\\
=&(-1)^{\Pi^{(2)}\cdot W_0^{(1)}}\Pi^{(1)}W_{n-1}^{(1)}\cdots W_2^{(1)}\Pi^{(2)}W_0'W_1^{(1)}W_0^{(1)}W_{n-1}^{(2)}\cdots W_0^{(2)}\\
=&(-1)^{\Pi^{(2)}\cdot W_0^{(1)}}\Pi^{(1)}\Pi^{(2)}W_{n-2}'W_{n-1}^{(1)}\cdots W_0'W_1^{(1)}W_0^{(1)}W_{n-1}^{(2)}W_{n-2}^{(2)}\cdots W_1^{(2)}W_0^{(2)}.
\end{aligned} \tag{B87}$$

The $W_0', \cdots, W_{n-2}'$ in the third and fourth lines are indeed the additional controlled-Z gates and multi-qubit controlled-Z gates in Eq. (B86).

Now, we give a brief analysis of the total complexity. The key point is utilizing Eq. (B86) to shift $\Pi^{(2)}$ gate to the left. We first fix an integer l and focus on C^lZ. Now we try to study the swapping between C^lZ and $\Pi^{(2)}$. By Eq. (B86), the total complexity for shifting $\Pi^{(2)}$ across a C^lZ gate is $O(2^{l+1})$ since the cardinality of subset I is at most $l+1$ and the number of subsets of I is at most 2^{l+1}. As the number of C^lZ gates on N qubits is at most $\binom{N}{l+1}$, the total complexity for shifting $\Pi^{(2)}$ to the leftmost is

$$\text{group multiplication complexity} = \sum_{l=0}^{n-1}\binom{N}{l+1}\cdot 2^{l+1}. \tag{B88}$$

The right part is bounded by N^n. Hence, the total group multiplication complexity is $O(N^n)$. Notice that this bound is quite untight. For example, if we let $n = N$, the right part of Eq. (B88) is indeed $3^n - 1$.

The inverse of an element $\Pi W_{n-1}\cdots W_0$ is $W_{n-1}\cdots W_0\Pi$, which equals the multiplication of element $W_{n-1}\cdots W_1 1_{\langle X\rangle}$ and $1_{\langle C^{n-1}Z\rangle}\cdots 1_{\langle Z\rangle}\Pi$, where 1_{G} represents the identity in group G and is indeed the all-zero string in the binary string representation of group elements. The computational complexity for computing inverse gate is then equal to the complexity for group element multiplication, i.e. $O(N^n)$.

To benchmark CZ_m gate, the twirling group by our protocol would always be a subgroup of $\mathsf{G}_{CZ_m} = \langle CZ_m, Z_m, X\rangle$. For simplicity, we only analyze the complexity of $\langle CZ_m, Z_m, X\rangle$. This complexity would be the upper bound of the complexity of the twirling group shown in Example 3. $\langle CZ_m, Z_m, X\rangle$ has a similar direct product structure like $\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, Z, X\rangle$,

$$\mathsf{G}_{CZ_m} = \langle X\rangle \rtimes (\langle CZ_m\rangle \times \langle Z_m\rangle). \tag{B89}$$

Thus, each element in G_{CZ_m} can be expressed as $W_2W_1\Pi$ where $W_2 \in \langle CZ_m\rangle, W_1 \in \langle Z_m\rangle$ and $\Pi \in \langle X\rangle$. In this case, the orders of generators CZ_m and Z_m are both m. We need to use $\lceil \log m\rceil$ bits to record their multiplicities and to uniquely identify one group element. Then, to represent an element in $\langle CZ_m\rangle$, we should use a bit string with length $\lceil \log m\rceil N(N-1)/2$. The total number of bits to express a group element is

$$\log\left|\langle X\rangle \rtimes (\langle CZ_m\rangle \times \langle Z_m\rangle)\right| = \left(\frac{N(N-1)}{2} + N\right)\lceil \log m\rceil + N = O(N^2\log m), \tag{B90}$$

which is indeed the sample complexity. The equation Eq. (B86) can be extended for CZ_m in the following way

$$CZ_m^kX_i = X_iCZ_m^{-k}(Z_m)_j^k, \tag{B91}$$

where the CZ_m here is on qubit i, j. By a similar argument like C^nZ, we can see that the group multiplication complexity, and thus inverse computation complexity, is $O(N^2 \log m)$.

The group size and computational complexity for the CNOT dihedral group have been elaborated in [16]. So, we omit the details here. Note that for CZ_m, their results only apply to $m = 2^k$, and ours admit m taking arbitrary positive integers. The scaling of the complexity results has been summarized and listed in the main text. In Fig. 8, we also show an accurate result of the size comparison between our group and the CNOT dihedral group by considering $N = n + 1$. The size of our group increases slower with respect to the qubit number than that of the CNOT dihedral group.

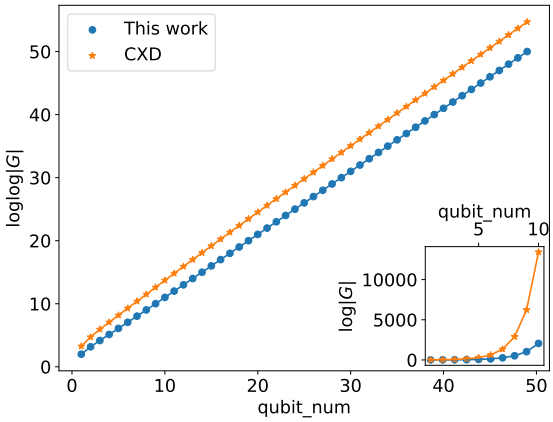


FIG. 8. Size comparison between the twirling groups for C^nZ in our method and in [16], associated with blue and orange patterns, respectively. $|\mathsf{G}|$ represents the group size. If taking a double logarithmic scale, both two kinds of groups increased nearly linearly with respect to the qubit number, but the CNOT dihedral group increased faster. The smaller figure demonstrates the results in the logarithm scale.

Below, we discuss the hierarchy of twirling groups, which may be additionally interesting. In the main text, we have mentioned that by choosing the twirled gate to be C^nZ_m with increasing n and m, the size of the twirling group and the computational complexity of the group multiplication would keep increasing. The two complexities are closely related to the classical simulability of the twirling group, and the hierarchy of the groups actually forms a computational complexity hierarchy. The results may be useful in other fields like quantum complexity theory and quantum simulation.

Here, we discuss the hierarchy of groups themselves rather than the hierarchy of their sample and computational complexities. Specifically, we focus on the group of $\langle C^nZ, C^{n-2}Z, \cdots, CZ, X, Z_{2^k}\rangle$. Starting from Pauli group $\langle X, Z\rangle$, there are two ways to build up the hierarchy. One way is adding the number of sides of dihedral groups, i.e., increasing the number k in $\langle X, Z_{2^k}\rangle$. The other way is adding multi-qubit controlled Z gates C^nZ or even adding multi-qubit controlled X gates C^nX. Interestingly, C^nZ can be contained in the CNOT-dihedral group $\langle CX, X, Z_{2^k}\rangle$ by choosing $k = n + 1$ [16]. The hierarchy is summarized below.

$$\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z_{2^k}\rangle \leq \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z_{2^{k+1}}\rangle \cdots \leq \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z(\theta)\rangle \cdots; \tag{B92}$$

$$\langle C^{n-1}X, X, Z_{2^k}\rangle \leq \langle C^{n-1}X, X, Z_{2^{k+1}}\rangle \cdots \leq \langle C^{n-1}X, X, Z(\theta)\rangle \cdots \leq \mathrm{CRU}_n; \tag{B93}$$

$$\langle X, Z_{2^k}\rangle \leq \langle CZ, X, Z_{2^k}\rangle \leq \langle CCZ, CZ, X, Z_{2^k}\rangle \cdots; \tag{B94}$$

$$\langle X, Z_{2^k}\rangle \leq \langle CX, X, Z_{2^k}\rangle \leq \langle CCX, X, Z_{2^k}\rangle \cdots \leq \mathrm{CRU}; \tag{B95}$$

$$\langle C^{k-1}Z, C^{k-2}Z, \cdots, CZ, X, Z_{2^k}\rangle \leq \langle CX, X, Z_{2^k}\rangle. \tag{B96}$$

Here, CRU_n represents the CRU on n qubits and CRU represents $\cup_n \mathrm{CRU}_n$. As shown in the above equations, the upper limit of the hierarchy is CRU. The key point is that $\langle C^{n-1}X, X\rangle$ contains all permutations over the computational bases $\{0, 1\}^n$ based on Lemma 10. Thus, $\langle C^{n-1}X, X, Z(\theta)\rangle$ contains all quantum gates like ΠW where Π is an arbitrary permutation gate, and W is an arbitrary diagonal gate. This is exactly the necessary and sufficient condition for CRU. Also, from the group hierarchy, it is clearer that the twirling group of our protocol is smaller than that in [16] for multi-qubit controlled-Z gates C^nZ. It is straightforward by investigating $\langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z\rangle \leq \langle C^{n-1}Z, C^{n-2}Z, \cdots, CZ, X, Z_{2^k}\rangle \leq \langle CX, X, Z_{2^k}\rangle$.

Appendix C: Simulation

In this part, we present additional details of the simulation, including the setting of the noise model, the benchmarking protocol using the ZX-SPAM setting, and additional benchmarking results. We denote the total number of qubits as N, which is 2 for the CS gate and $n+1$ for the C^nZ gate.

1. Noise model

In summary, to simulate a noisy quantum gate, we consider local dephasing noise, local amplitude damping noise, and unitary noise. The strength of the former two noises depends on the time to implement the gate. To calculate this time, we first decompose this gate into a series of CNOT gates, Pauli X gates, multi-qubit controlled phase gates, and single-qubit phase gates. We sum the time to implement each part to obtain the total time of implementing a gate, denoted as T_{gate}. Then, the former two types of noise are simulated as follows.

1) Local dephasing channel Λ_d. The dephasing channel on qubit i, where $1 \le i \le N$, is defined as

$$\Lambda_d^i(\rho) = K_0^i \rho K_0^{i\dagger} + K_1^i \rho K_1^{i\dagger}. \tag{C1}$$

Here,

$$K_0^i \equiv \begin{pmatrix} \sqrt{p_i} & 0 \\ 0 & \sqrt{p_i} \end{pmatrix}, K_1^i \equiv \begin{pmatrix} \sqrt{1-p_i} & 0 \\ 0 & -\sqrt{1-p_i} \end{pmatrix}, \tag{C2}$$

where p_i denotes the dephasing strength. The local dephasing channel is defined as

$$\Lambda_d = \bigotimes_{i=1}^{N} \Lambda_d^i. \tag{C3}$$

In the simulation, we set all $p_i, 1 \le i \le N$ to be the same, which equals

$$p_i = e^{-\frac{T_{gate}}{T_2}}, \tag{C4}$$

where T_{gate} denotes the time to implement a gate and depends on the concrete gate and $T_2 = 15000$ns.

2) Local amplitude damping channel Λ_a. The amplitude damping channel on qubit i, where $1 \le i \le N$, is defined as

$$\Lambda_a^i(\rho) = K_0^i \rho K_0^{i\dagger} + K_1^i \rho K_1^{i\dagger}. \tag{C5}$$

Here,

$$K_0^i \equiv \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{q_i} \end{pmatrix}, K_1^i \equiv \begin{pmatrix} 0 & \sqrt{1-q_i} \\ 0 & 0 \end{pmatrix}. \tag{C6}$$

The parameter q_i denotes the noise strength of the amplitude damping channel. The local amplitude damping channel on N qubits is defined as

$$\Lambda_a = \bigotimes_{i=1}^{N} \Lambda_a^i. \tag{C7}$$

In the simulation, we set all $q_i, 1 \le i \le N$ to be the same, which equals

$$q_i = e^{-\frac{T_{gate}}{T_1}}, \tag{C8}$$

where T_{gate} depends on the concrete gate and $T_1 = 25000$ns.

To simulate the unitary noise for each gate, we also utilize the decomposition of the gate and consider unitary errors for each decomposed part. Then, the noisy gate is simulated by sequentially acting local dephasing noise, local amplitude damping noise, and gate with unitary errors.

We introduce the calculation of T_{gate} and the unitary noise for each gate below. Recall that all of the twirling gates and the twirled gate belong to CRU. A CRU gate can be separated into a permutation part and a diagonal part. The time T_{gate} is evaluated by the sum of times of two parts. The unitary errors are also considered separately. We discuss the noise models for the two parts respectively.

1. The permutation part is $\langle CX, X\rangle$ for the CNOT dihedral group and $\langle X\rangle$ for the other simulated groups. Note that one can always implement a number of CNOT gates, and then implement one layer of Pauli X gates to realize an element in $\langle CX, X\rangle$; $\langle X\rangle$ is realized by one layer of Pauli X gates.

   For the part constructed by CNOT gates, denoted as 'CNOT part', we decompose the circuit based on the algorithm in Ref. [54] and get a synthesis of the circuit with CNOT gates and SWAP gates. We calculate the circuit depth of the 'CNOT part,' denoted as 'd_{CNOT},' by counting the minimum layer to realize all CNOT gates. If two CNOT gates can be realized parallelly, then they are in the same layer. The SWAP gates do not contribute to the counting of circuit depth since one can always virtually implement SWAP gates by manually changing the label of qubits. Then, the time to implement the 'CNOT part' is

$$t_{CNOT} = d_{CNOT} \times 2 \times 30\text{ns}. \tag{C9}$$

   If there are no gates in the 'CNOT part,' t_{CNOT} is set as 0. For each decomposed CNOT gate, we consider its unitary noise as

$$\text{Controlled}-\exp(i\delta X), \tag{C10}$$

   where $\delta = 0.005\pi$ and X is the Pauli X gate.

   Similarly, the circuit depth of the Pauli X layer is 1 as long as one qubit is acted nontrivially by a Pauli X gate and 0 otherwise, which is denoted as d_X. The time to implement the Pauli X layer is

$$t_X = d_x \times 30\text{ns}. \tag{C11}$$

   The unitary error for each decomposed Pauli X gate is

$$\exp(i\delta X), \tag{C12}$$

   where $\delta = 0.005\pi$. The total time of the permutation part is $t_{perm} = t_{CNOT} + t_X$.

2. The diagonal part of a CRU gate, W, can be generated by multi-qubit controlled phase gates and single-qubit Z rotation:

$$W = \prod_{r=1}^{N} \prod_{i \in I_r} U_i^r, \tag{C13}$$

   where $U_i^r = C^{r-1}Z_l$ is a multi-qubit controlled phase gate with $r-1$ controlled qubit number, and l denotes the phase. When $r = 1$, U_i^r is a single-qubit Z rotation. The notation I_r is an index set. When $I_{r'}$ is empty, there is no multi-qubit controlled phase gate with $r'-1$ controlled qubit number in the decomposition of W. Since U_i^r commutes with each other, we consider all these gates can be implemented at the same time in the simulation. Thus, the time to implement W is set as

$$t_{diag} = \max(\{r | I_r \neq \varnothing\}) \times 30\text{ns}. \tag{C14}$$

   This is the time to implement the multi-qubit controlled phase gate with the largest controlled qubit number. The unitary error is a multiple-Z coupling, considered for each multi-qubit controlled phase gate in the decomposition. For $C^{r-1}Z_l = e^{i\frac{2\pi}{l}|1\rangle\langle 1|^{\otimes r}}$ gate, the unitary error is

$$U_Z = e^{i\delta r \sum_{z\in\{0,1\}^r} 2^{|z|-r}|z\rangle\langle z|}, \tag{C15}$$

   where $\delta = 0.005\pi$, and $|z|$ is the number of 1 in z. The state with more 1 is influenced more.

Besides the noise of the gate, we also set errors for state preparation and measurement (SPAM). In our simulation experiments, the initial states are $U|0\rangle^{\otimes N}$ where U is a local Clifford gate. In the simulation, we suppose the local Clifford gate does not bring errors, and the error of the initial state comes from the noise when preparing $|0\rangle^{\otimes n}$. We assume that on each qubit, the state $|0\rangle$ would turn into $|1\rangle$ with probability 0.02, resulting from the thermal excitation. We take the measurement to be noiseless as there is always a measurement error correction in experiments.

2. Benchmarking procedure with the ZX-SPAM

Below, we present the benchmarking procedure with the ZX-SPAM when the twirling group contains the CNOT dihedral group CZD = $\langle CZ, X, S\rangle$. Before the introduction of the concrete benchmarking procedure, we explain why the ZX-SPAM suffices to extract all different diagonal terms when the twirling group contains the CNOT dihedral group.

We take CS gate for an example and present the Liouville representation of its twirled noise channel under several different twirling groups, including Pauli group $\mathsf{P} = \langle X, Z\rangle$, CZ Pauli group CZP = $\langle CZ, X, Z\rangle$, CZ dihedral group CZD = $\langle CZ, X, S\rangle$, and CNOT dihedral group CXD = $\langle CX, X, T\rangle$. The results are shown in Figs. 9, 10, 11, and 12.

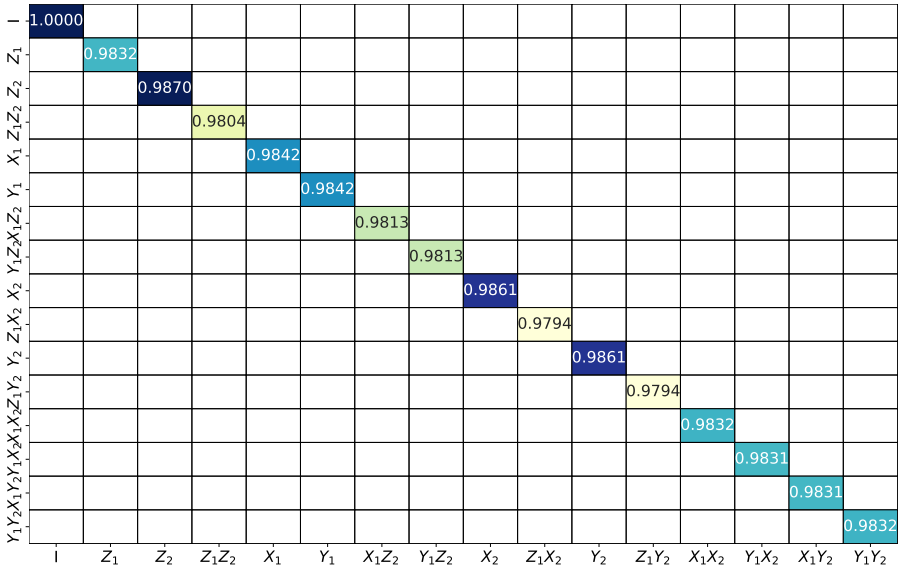


FIG. 9. Pauli Liouville representation of Λ_{P} where Λ is the noise channel of CS gate and P denotes the Pauli group. The values in vacant squares are all 0, and we omit them. The two axes record the bases of the Pauli-Liouville representation.

The matrix elements of the original noise channel Λ of CS gate in the Pauli-Liouville representation are all non-zero and contain complex numbers. After twirled by the Pauli group, there are only diagonal terms left to be nonzero, as shown in Fig. 9. If we use a larger twirling group, the CZ Pauli group, for any $X_a \in \{X_1, X_2, X_1X_2\}$, the matrix elements corresponding to X_a and $X_aZ_1Z_2$ would be averaged, and the elements of X_aZ_1 and X_aZ_2 would be also averaged. The matrix elements of Λ_{CZP} are shown in Fig. 10. Its diagonal blocks are associated with the bases set in $\{\{\mathbb{I}\}, \{Z_1\}, \{Z_2\}, \{Z_1Z_2\}, \{X_aZ_1, X_aZ_2\}, \{X_a, X_aZ_1Z_2\} | X_a \in \{X_1, X_2, X_1X_2\}\}$. Thus, the number of different diagonal terms is 7, excluding the one associated with $\{\mathbb{I}\}$. Nonetheless, via the ZX-SPAM, one can at most obtain $2(2^N - 1) = 6$ different diagonal terms,

$$\mathrm{tr}(P\Lambda(P))/2^N, P \in (\{\mathbb{I}, Z\}^{\otimes N} \bigcup \{\mathbb{I}, X\}^{\otimes N}) \backslash \mathbb{I}^{\otimes N}. \tag{C16}$$

Thus, the twirling of the CZ Pauli group does not suffice.

If we further use a bit larger twirling group, CZ dihedral group, the matrix elements corresponding to $X_a\mathsf{Z}$ would be all averaged as shown in Fig. 11. The number of different diagonal terms is just $2(2^N - 1) = 6$. In this case, using the ZX-SPAM, we can extract the diagonal terms of Λ^m_{CZD} and hence obtain the diagonal terms of Λ_{CZD} via fitting. If using the CNOT dihedral group for twirling, the twirled noise channel Λ_{CXD} would be more symmetric than Λ_{CZD}. The elements corresponding to Z_1, Z_2, and Z_1Z_2 will be averaged and the elements corresponding to $\{X_a\mathsf{Z} | X_a \in \{X_1, X_2, X_1X_2\}\}$ will also be averaged. Thus, Λ_{CXD} only has two parameters as shown in Fig. 12. Using the ZX-SPAM can also extract all different diagonal terms of Λ_{CXD}.

In Box 1, we introduce the detailed benchmarking procedure with the ZX-SPAM. The twirling group contains the CZ dihedral group.

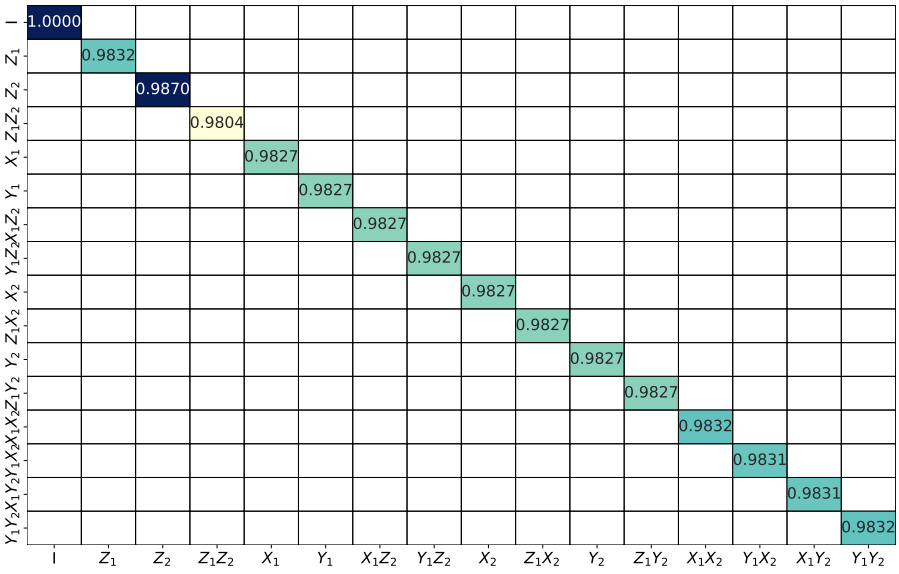


FIG. 10. Pauli Liouville representation of Λ_{CZP} where Λ is the noise channel of CS gate and CZP denotes the twirling group $\langle CZ, Z, X\rangle$. The values in vacant squares are all 0, and we omit them. Note that the elements corresponding to X_1X_2 and Y_1Y_2 are the same, and the elements corresponding to X_1Y_2 and X_1Y_2 are the same. But the two values are different. Generally, the elements corresponding to X_1 and Y_1 are different, and the elements corresponding to X_2 and Y_2 are also different. However, in this case, the noise channel is very special, and the difference can hardly be seen by accident.

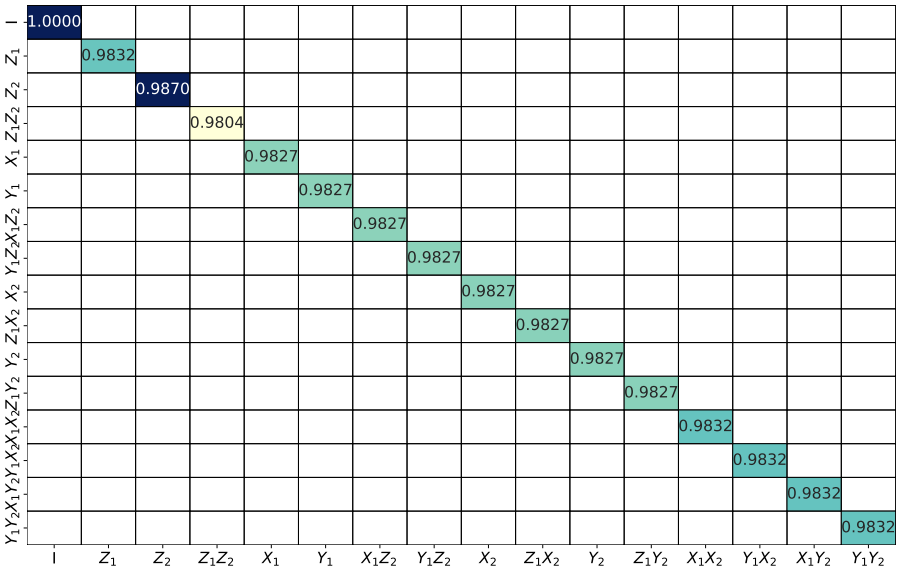


FIG. 11. Pauli Liouville representation of Λ_{CZD} where Λ is the noise channel of CS gate and CZD denotes the twirling group $\langle CZ, Z, S\rangle$. The values in vacant squares are all 0, and we omit them. The matrix elements corresponding to $X_a\mathsf{Z}$ are all averaged respectively for $X_a \in \{X_1, X_2, X_1X_2\}$.

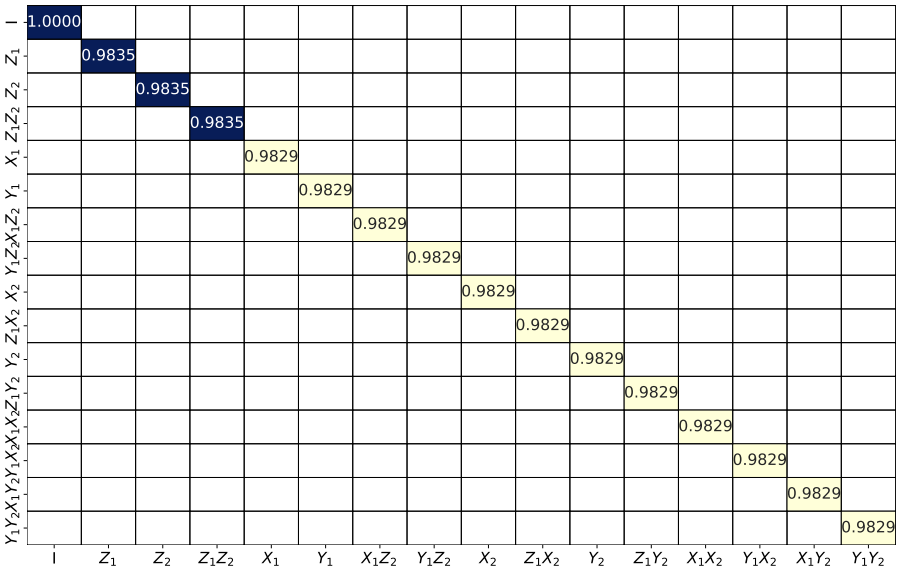


FIG. 12. Pauli Liouville representation of Λ_{CXD} where Λ is the noise channel of CS gate and CXD denotes the twirling group $\langle CX, X, T\rangle$. The values in vacant squares are all 0, and we omit them. The second to the fourth diagonal elements are averaged, and the last 12 diagonal elements are also averaged.

Box 1: Procedures for benchmarking with the ZX-SPAM

1. First initialize the state, $|\psi_0\rangle = |0\rangle^{\otimes N}$ or $|\psi_1\rangle = |+\rangle^{\otimes N}$ where N is the number of qubits.
2. Choose a positive integer, M, and choose two sets of positive integers $\{m_1, m_2, ..., m_M\}$ and $\{K_1, K_2, ..., K_M\}$. Here, $\{m_1, m_2, ..., m_M\}$ is the set of circuit depths and for $1 \le i \le m$, K_i is the number of sampled sequences when circuit depth equals m_i.
3. For each integer $1 \le i \le M$, uniformly and randomly sample $2m_i$ gates from twirling group G, where $\mathsf{G} \supseteq \langle CZ, S, X\rangle$, for K_i times, which we denote $\{G_{j,1}, G_{j,2}, \cdots, G_{j,2m_i}\}$, $1 \le j \le K_i$.
4. For each integer $1 \le i \le M$ and $1 \le j \le K_i$, implement gate sequence
$$\widetilde{S}(j, m_i) = \widetilde{U}_{inv} \prod_{t=1}^{m_i} U^\dagger G_{j,2t} U G_{j,2t-1}, \tag{C17}$$
where $\widetilde{\cdot}$ represents the noisy version of the quantum gate and $U_{inv} = (\prod_{t=1}^{m_i} U^\dagger G_{j,2t} U G_{j,2t-1})^\dagger$; U is the target gate.
5. For initial state $|\psi_0\rangle = |0\rangle^{\otimes N}$, measuring all observables from $\{\mathbb{I}, Z\}^{\otimes N}$ of the final state via Z-basis measurement. For initial state $|\psi_1\rangle = |+\rangle^{\otimes N}$, measuring all observables from $\{\mathbb{I}, X\}^{\otimes N}$ of the final state via X-basis measurement. That is, for each $Q_k \in \{\mathbb{I}, Z\}^{\otimes N}$, estimate
$$f_Z(j, m_i, k) = \mathrm{tr}\big(\widetilde{Q}_k \widetilde{S}(j, m_i)(\widetilde{\rho}_0)\big). \tag{C18}$$
For each $P_k \in \{\mathbb{I}, X\}^{\otimes N}$, estimate
$$f_X(j, m_i, k) = \mathrm{tr}\big(\widetilde{P}_k \widetilde{S}(j, m_i)(\widetilde{\rho}_1)\big). \tag{C19}$$
Here, $\widetilde{\rho}_0$ and $\widetilde{\rho}_1$ are noisy versions of $|0\rangle^{\otimes N}$ and $|+\rangle^{\otimes N}$, respectively.

6. Average the results of different gate sequences and obtain

$$f_Z(m_i,k) = \frac{1}{K_i}\sum_{j=1}^{K_i} f_Z(j,m_i,k) \tag{C20}$$

$$f_X(m_i,k) = \frac{1}{K_i}\sum_{j=1}^{K_i} f_X(j,m_i,k). \tag{C21}$$

7. For each Q_k, fit $f_Z(m_i,k)$ to function

$$f(m_i) = A\lambda_{Z,k}^{m_i} \tag{C22}$$

and obtain $\lambda_{Z,k}$. For each P_k, fit $f_X(m_i,k)$ to function

$$f(m_i) = A\lambda_{X,k}^{m_i} \tag{C23}$$

and obtain $\lambda_{X,k}$.

8. Estimate the fidelity of the target gate via

$$F = \frac{\sum_k \lambda_{Z,k} + 2^N(\sum_k \lambda_{X,k} - 1)}{4^N}. \tag{C24}$$

9. If one wants to separate further the noise of the twirling gates and the target gate, execute the following step. Replace the target gate with identity and repeat the above processes to estimate the fidelity of the twirling groups, F_G. Then, estimate the fidelity of U with F and F_G by

$$F_U = \frac{d^2F - 1}{d^2F_G - 1}(1 - \frac{1}{d^2}) + \frac{1}{d^2}. \tag{C25}$$

Note that in the above benchmarking procedure, we use the circuit structure in Ref. [15] instead of Ref. [14]. That means we implement random twirling gates interleaved with U and $U^\dagger$ instead of being just interleaved with U. For C^nZ gate, $U = U_{inv}$ and this modification does not influence anything. For more general case that U is C^nZ_m gate like CS gate, under a mild assumption that U and $U^\dagger$ has the same noise Λ, the above procedure can estimate the fidelity of $\sqrt{\mathcal{U}^\dagger\Lambda_G\mathcal{U}\Lambda_G}$, which is equal to the fidelity of Λ. In this modified procedure, the circuit depth only needs to be chosen as the multiples of 2. While within the circuit structure in Fig. 6, just like [14], the circuit depth needs to be chosen as the multiples of the order of the target gate to ensure the inverse gate belonging to the twirling group. For instance, the order of C^nZ_m is m and can be large for C^nZ_m with large m. In this case, the modified benchmarking procedure in Box 1 can offer an advantage of shorter circuit depth in circuit implementation. One can also use the circuit structure in Fig. 6 to benchmark the noise channel.

3. Fidelities of dressed and twirling gates

In this part, we show the full results of the dressed fidelity, or the fidelity of the composite noise channel of the target gate and the twirling group, the fidelity of the twirling groups, and the fidelity of the target gates. Depending on the twirling group and the SPAM setting, we have five benchmarking methods:

1) Pauli, random SPAM.

2) CZ dihedral group (for the CS gate) or $\langle C^{n-1}Z,\cdots,CZ,X,Z\rangle$ (for the C^nZ gate), random SPAM.

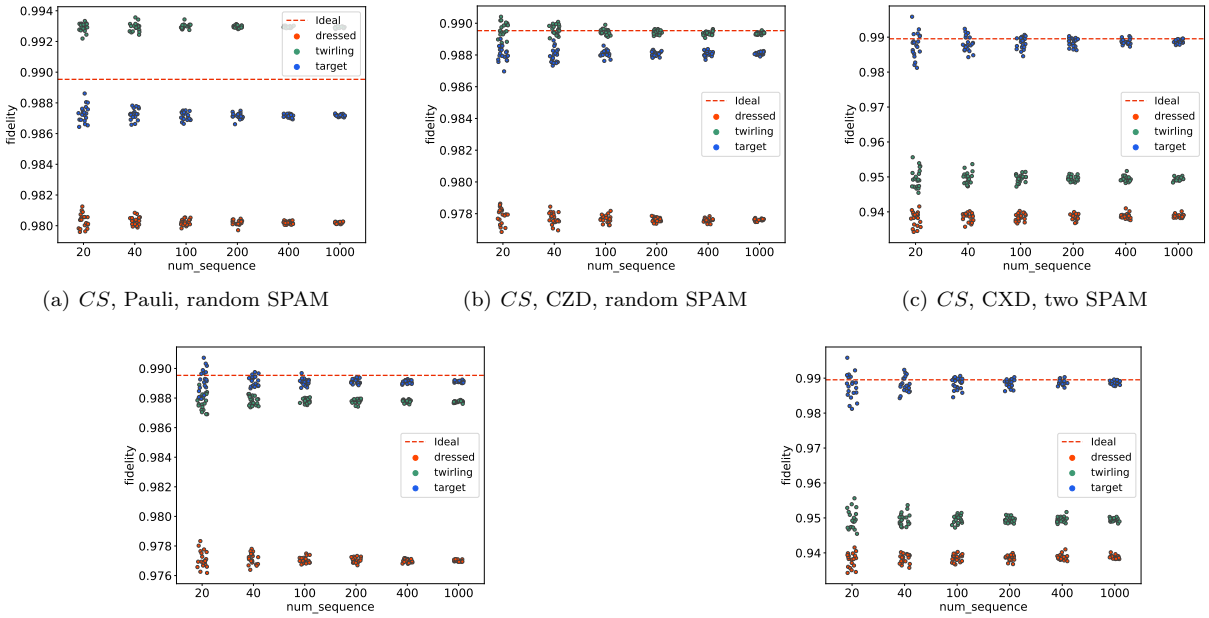
3) CNOT dihedral group, two SPAM.

4) $\langle C^{n-1}Z,\cdots,CZ,X,S\rangle$, ZX-SPAM.

5) CNOT dihedral group, ZX-SPAM.

For simplicity, the CZ dihedral group or $\langle C^{n-1}Z,\cdots,CZ,X,S\rangle$ is denoted as CZD. The group $\langle C^{n-1}Z,\cdots,CZ,X,Z\rangle$ is denoted as CZP. The CNOT dihedral group is denoted as CXD.

Recall that the benchmarking results in the main text cover the CS, CCZ, $CCCZ$, and $CCCCZ$ gates. We list the benchmarking results of the four gates with five methods in Figs. 13, 14, 15, and 16, respectively. We can see that the results of the Pauli group have the minimum fluctuation, thanks to the low error rate of the twirling group. The results of CZP or CZD are effective in a small-scale system below 5 qubits. The CNOT dihedral group fails in benchmarking when the qubit number is no less than 4. Thus, in the main text, we do not present associated results.



(a) CS, Pauli, random SPAM (b) CS, CZD, random SPAM (c) CS, CXD, two SPAM

(d) CS, CZD, ZX-SPAM (e) CS, CXD, ZX-SPAM

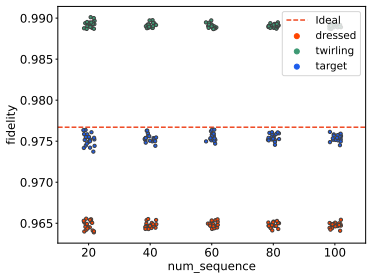
FIG. 13. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the CS gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.

4. Simulation results of C^5Z

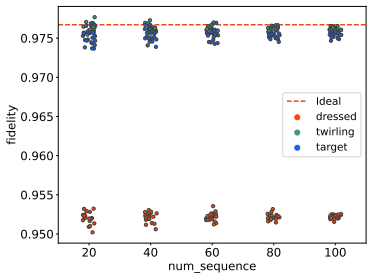
In this part, we present the benchmarking results of C^5Z, with the Pauli group, $\langle C^4Z, \cdots, CZ, X, S\rangle$, and $\langle C^4Z, \cdots, CZ, X, Z\rangle$. The latter two are labeled with CZD and CZP, respectively. The results are shown in Fig. 17. It can be seen that due to a large gate-dependent noise, the results of the latter two methods fluctuate a lot. Nonetheless, from Fig. 17(d), we can still observe that though the variance of the results associated with the Pauli group is lower, the bias associated with CZD is smaller. If one can utilize a sufficient number of random sequences, the fluctuation of results corresponding to CZD can decrease. In contrast, the bias of the results associated with the Pauli group does not change with the number of sampled sequences increase.

5. Simulation results with noiseless twirling groups

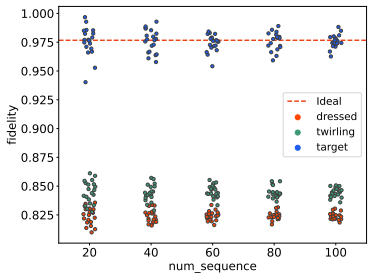
To further understand how the choice of the twirling group influences the final benchmarking results, we also simulate the benchmarking procedure with the noiseless twirling group. The noise model of the target gate does not change and is still given in Appendix C 1. In Figs. 18 and 19, we present the results of benchmarking the CS and CCZ gates with five benchmarking methods, which are introduced in Appendix C 3. It can be seen that the methods 'CXD, two SPAM,' 'CZD, ZX-SPAM,' and 'CXD, ZX-SPAM' can both estimate fidelity unbiasedly. The results of the methods 'Pauli, random SPAM' and 'CZD (CZP), random SPAM' deviate from the ideal value, and the results of 'Pauli, random SPAM' deviate more. The deviation of 'Pauli, random SPAM' and 'CZD (CZP), random SPAM' can be explained by the effect of random SPAM settings. The randomly sampled SPAM introduces a bias to the fidelity estimation. Meanwhile, the weak twirling 'ability' of the Pauli group, or the less symmetric of the twirled



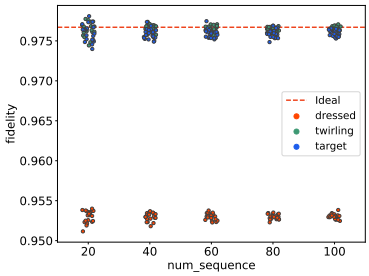
(a) CCZ, Pauli, random SPAM



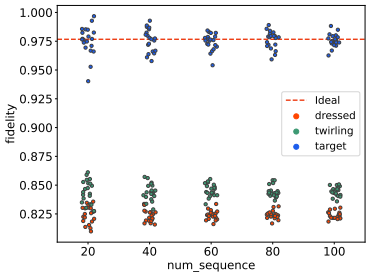
(b) CCZ, CZP, random SPAM



(c) CCZ, CXD, two SPAM

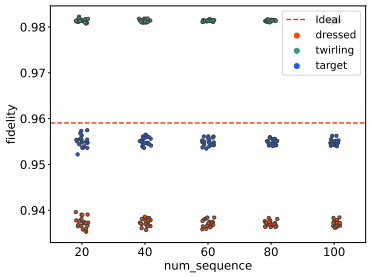


(d) CCZ, CZD, ZX-SPAM

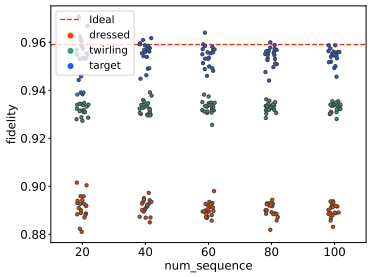


(e) CCZ, CXD, ZX-SPAM

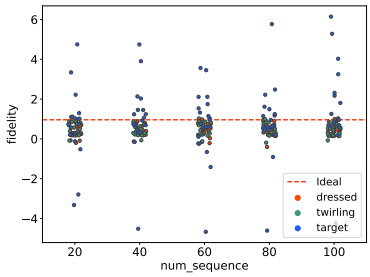
FIG. 14. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the CCZ gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.



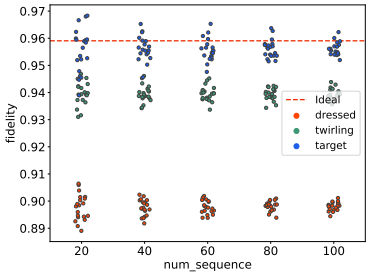
(a) $CCCZ$, Pauli, random SPAM



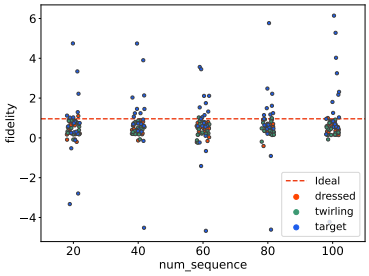
(b) $CCCZ$, CZP, random SPAM



(c) $CCCZ$, CXD, two SPAM

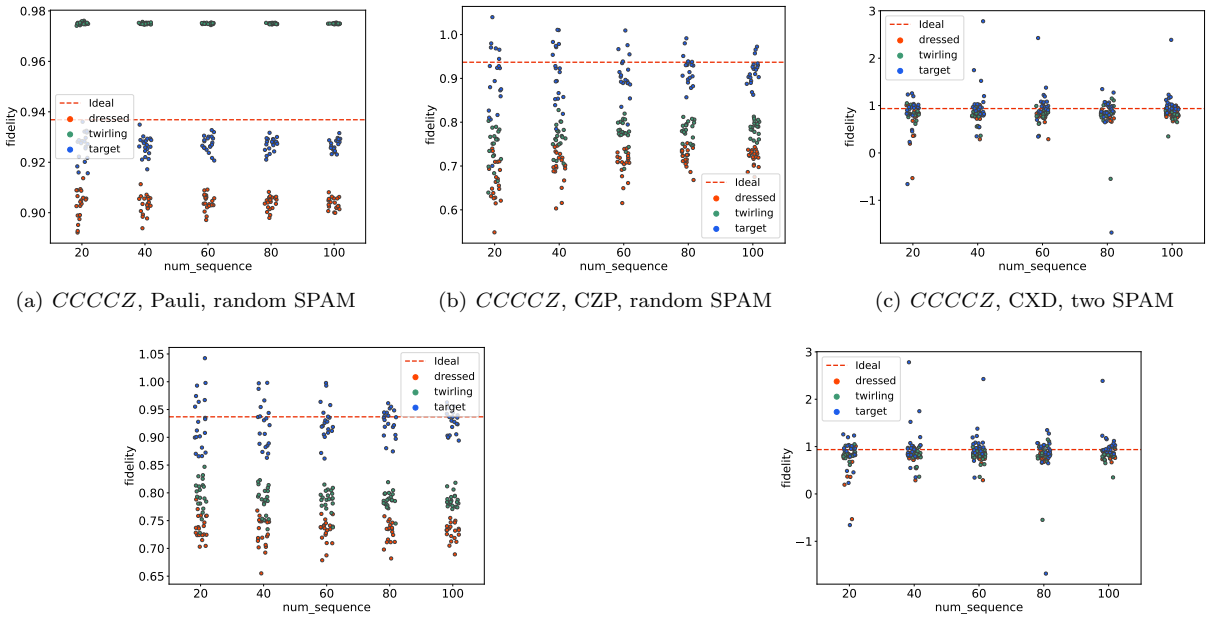


(d) $CCCZ$, CZD, ZX-SPAM



(e) $CCCZ$, CXD, ZX-SPAM

FIG. 15. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the $CCCZ$ gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.

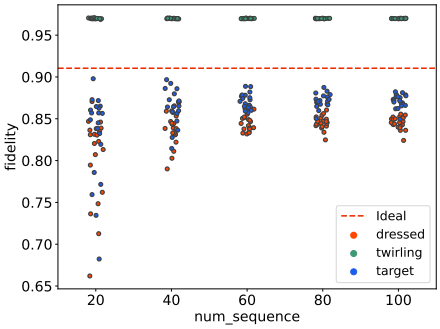


(a) $CCCCZ$, Pauli, random SPAM (b) $CCCCZ$, CZP, random SPAM (c) $CCCCZ$, CXD, two SPAM

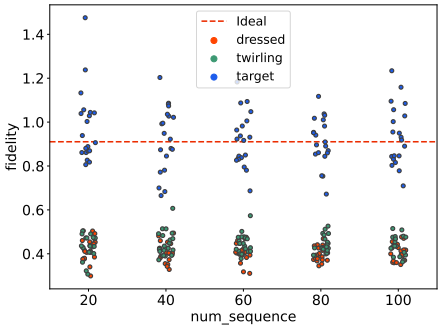
(d) $CCCCZ$, CZD, ZX-SPAM (e) $CCCCZ$, CXD, ZX-SPAM

FIG. 16. The results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the $CCCCZ$ gate with five benchmarking methods. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity.

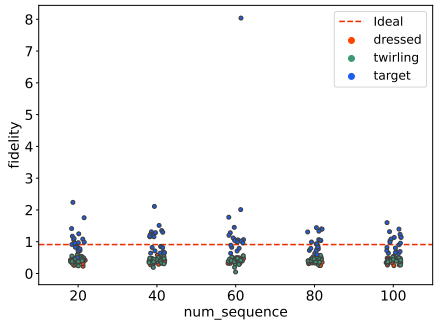
noise channel associated with the Pauli group, introduces another bias effect to the fidelity estimation. Thus, the results of the Pauli group deviate the most in the case of the noiseless twirling group. In practice, the Pauli group is the least noisy. Hence, when considering a noisy twirling group, especially in a larger system, benchmarking with the Pauli group can provide relatively more effective fidelity estimation than other groups.



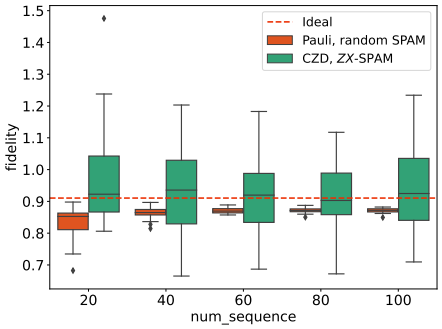
(a) $CCCCCZ$, Pauli, random SPAM



(b) $CCCCCZ$, CZD, ZX-SPAM

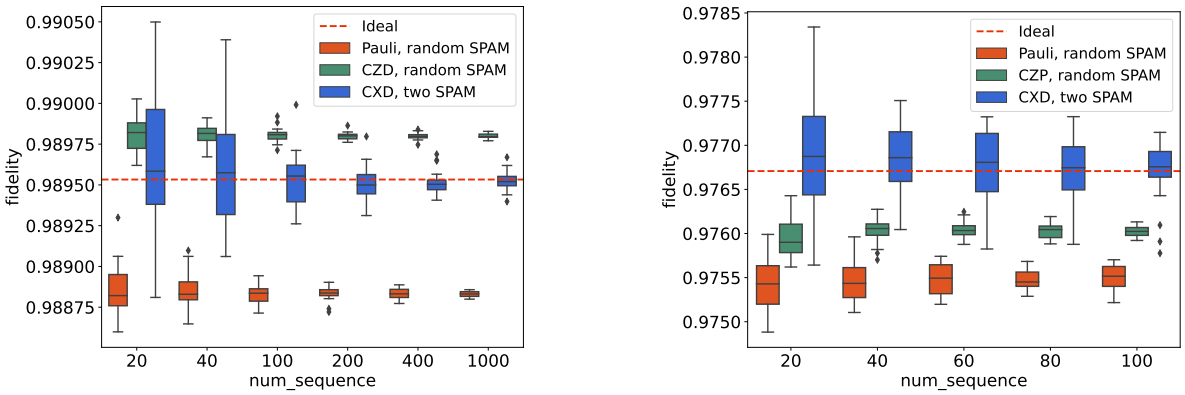


(c) $CCCCCZ$, CZP, random SPAM



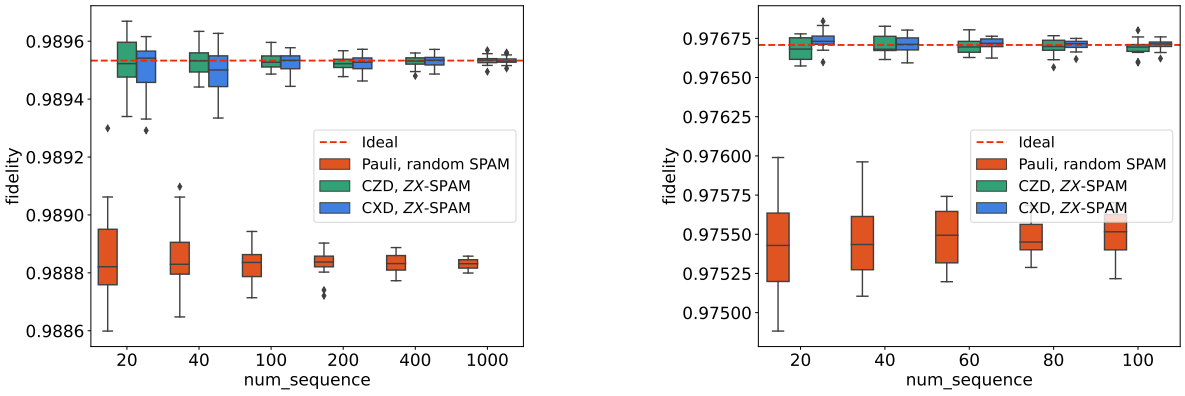
(d) $CCCCCZ$, Pauli vs CZD

FIG. 17. Figures (a), (b), and (c) show the results of the dressed fidelity, the twirling group fidelity, and the target gate fidelity for the $CCCCCZ$ gate with three benchmarking methods. The three methods are the Pauli group with a random SPAM, $\langle C^4Z, \cdots, CZ, X, S\rangle$ with the ZX-SPAM, and $\langle C^4Z, \cdots, CZ, X, Z\rangle$ with a random SPAM. The red dashed line is the theoretical value of the target gate fidelity. The horizontal axis denotes the number of sampled sequences for each depth and the vertical axis denotes the fidelity. Figure (d) compares the target gate fidelities associated with the Pauli group and $\langle C^4Z, \cdots, CZ, X, S\rangle$. It can be seen that the variance of the results corresponding to the Pauli group is lower and the bias associated with $\langle C^4Z, \cdots, CZ, X, S\rangle$ is smaller.



(a) CS (b) CCZ

FIG. 18. Benchmarking results for the CS and CCZ gates in Figures (a) and (b), respectively, with the optimal twirling group, the Pauli group, and the CNOT dihedral group in the case of noiseless twirling group. The optimal twirling group is the CZ dihedral group $\langle CZ, Z, S\rangle$ for CS, and the CZ Pauli group $\langle CZ, Z, P\rangle$ for CCZ. The red dashed line is the theoretical value of the noise channel fidelity. Each box plot contains 20 fidelities, and each fidelity is estimated with circuit depths $\{2, 4, 6, 8, 10\}$, and the total number of different gate sequences for each depth is specified by the horizontal axis. Here, for the Pauli group and the optimal group, we randomly sample and estimate 20 different diagonal terms of the twirled noise channel. We mark this setting with 'random SPAM' on the label. For the CNOT dihedral group, we estimate only two different diagonal terms of the twirled noise channel. We mark this setting with 'two SPAM' on the label. Each SPAM setting prepares an eigenstate of a Pauli observable with eigenvalue 1 and measures this Pauli observable. In 'two SPAM', the two Pauli observables are chosen as $Z^{\otimes N}$ and $X^{\otimes N}$.



(a) CS (b) CCZ

FIG. 19. Benchmarking results for the CS and CCZ gates in Figures (a) and (b), respectively, with the Pauli group, the CZ dihedral group, and the CNOT dihedral group in the case of the noiseless twirling group. The red dashed line is the theoretical fidelity value. Each box plot contains 20 fidelities. The setting of circuit depths and sampling is the same as in Figure 18. Nonetheless, for the CZ dihedral group and the CNOT dihedral group, we adopt the SPAM setting and the benchmarking procedure of Box 1. We mark this setting with 'ZX-SPAM' on the label.

Non-Markovian Quantum Exceptional Points

Jhen-Dong Lin[1 2 *] Po-Chen Kuo[1 2] Neill Lambert[4 5] Adam Miranowicz[6]
Franco Nori[4 5 7] Yueh-Nan Chen[1 2 3 †]

[1] *Department of Physics, National Cheng Kung University, 701 Tainan, Taiwan*
[2] *Center for Quantum Frontiers of Research & Technology, NCKU, 70101 Tainan, Taiwan*
[3] *Physics Division, National Center for Theoretical Sciences, Taipei 106319, Taiwan*
[4] *RIKEN Center for Quantum Computing (RQC), Wakoshi, Saitama 351-0198, Japan*
[5] *Theoretical Quantum Physics Laboratory, RIKEN Cluster for Pioneering Research, Wako-shi, Saitama 351-0198, Japan*
[6] *Institute of Spintronics and Quantum Information, Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland*
[7] *Department of Physics, The University of Michigan, Ann Arbor, 48109-1040 Michigan, USA*

Abstract. Exceptional points (EPs) are spectral singularities of non-Hermitian operators. In this work, we propose a theoretical framework based on two numerically exact descriptions of non-Markovian dynamics: the pseudomode mapping and the hierarchical equations of motion. We unveil pure non-Markovian EPs that are unobservable in the Markovian limit. Moreover, we show that structured environments can elevate EP order, thereby enhancing the system's sensitivity. These findings lay a theoretical foundation and open new avenues for non-Markovian reservoir engineering and non-Hermitian physics.

Keywords: non-Hermitian physics, exceptional points, and non-Markovian open quantum systems

1 Motivation

Spectral singularities for non-Hermitian systems, known as exceptional points (EPs), have attracted intense research attention over the past decades. Recently, investigations of EPs have extended into the full quantum regime where the temporal evolution of an open quantum system is governed by a Lindblad master equation or, equivalently, by a Liouvillian superoperator. In this context, the EPs associated with Liouvillian superoperators are termed as quantum EPs or Liouvillian EPs (LEPs) [1]. It has been demonstrated that *pure quantum EPs* exist, which are phenomena without (semi-)classical counterparts.

To date, the exploration for LEPs has largely been confined to the memoryless Markovian limit, which is only valid in cases of sufficiently weak system-environment interaction or environments without any structure. Accordingly, whether the concepts of EPs can be directly applied to the non-Markovian regime remains an open question.

In this work, we aim to address this theoretical gap. The main result lies in the development of a systematic framework for quantum EPs associated with generic non-Markovian open systems. The idea is based on applying the pseudomode equation of motion (PMEOM) [2] and the hierarchical equations of motion (HEOM) [3], which can be used to describe a large class of system-environment models. In this case, the dynamics is governed by what we call *extended Liouvillian superoperators*, enabling us to perform conventional spectral analysis, identifying the corresponding EPs, and revealing their impacts on the non-Markovian open systems.

*jhendonglin@gmail.com
†yuehnan@mail.ncku.edu.tw

2 Results

2.1 General framework

The general framework is described in Fig. 1. Specifically, we consider an open quantum system (S) coupled to a bosonic environment (E). In this case, the exact dynamics of the open system's reduced density matrix can be written as

$$\rho_{\rm S}(t) = \hat{\mathcal{T}} \exp\left\{\hat{\mathcal{F}}\Big[Q, C(t)\Big]\right\} \rho_{\rm S}(0). \tag{1}$$

Here, $\hat{\mathcal{T}}$ denotes the time-ordering operator and $\hat{\mathcal{F}}$ represents the Feynman-Vernon influence functional. An essential feature of $\hat{\mathcal{F}}$ is its exclusive dependence on the system-environment coupling operator Q and the environmental correlation function $C(t)$.

For a broad range of cases, the correlation function can be efficiently expressed as a finite weighted summation of exponential terms, i.e., $C(t) = \sum_i \alpha_i^2 \exp(-i\Omega_i t - \gamma_i|t|/2)$. With this expression, one can construct the PMEOM [2]:

$$\begin{aligned} &\frac{d}{dt}\rho_{\rm S+PM}(t) = \mathcal{L}_{\rm S+PM}[\rho_{\rm S+PM}(t)] \\ &= -i[H_{\rm S+PM}, \rho_{\rm S+PM}(t)] + \sum_i \gamma_i \mathcal{L}_{a_i}[\rho_{\rm S+PM}(t)], \\ &\text{with}\ \ H_{\rm S+PM} = H_{\rm S} + \sum_i \Omega_i a_i^\dagger a_i + \alpha_i Q(a_i^\dagger + a_i), \end{aligned} \tag{2}$$

where, $H_{\rm S}$ denotes the system Hamiltonian, $\{a_i\}$ represent the pseudomodes, and we introduce the dissipator $\mathcal{L}_{a_i}[\bullet] = a_i \bullet a_i^\dagger - \{a_i^\dagger a_i, \bullet\}/2$. The exact dynamics of S can be obtained by tracing out the pseudomodes (PM), i.e., $\rho_{\rm S}(t) = {\rm tr}_{\rm PM}[\rho_{\rm S+PM}(t)]$, after solving the PMEOM.

To describe EPs, one considers a family of parametrized extended Liouvillian superoperators

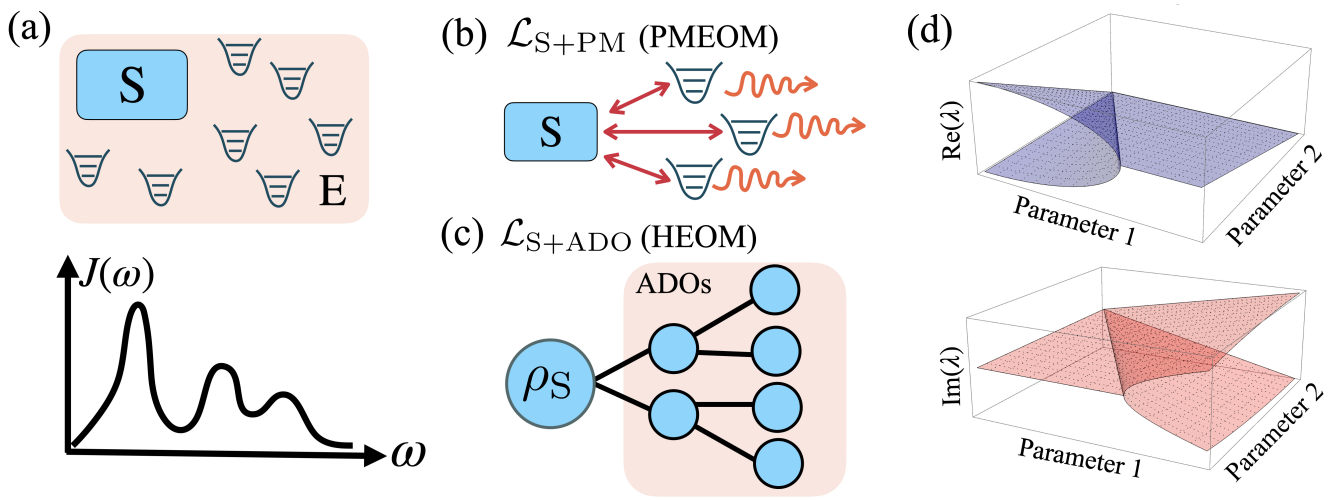


Figure 1: Schematic illustration depicting EPs for (a) a generic non-Markovian open-system model, where the structured environment is captured by the spectral density function $J(\omega)$. For a given spectral density function and the corresponding environmental correlation function, the exact non-Markovian dynamics can either be described by (b) PMEOM or (c) HEOM with the corresponding extended Liouvillian superoperators: $\mathcal{L}_{\mathrm{S+PM}}$ and $\mathcal{L}_{\mathrm{S+ADO}}$. (d) The non-Markovian EPs can then be identified by observing the complex spectrum of these extended Liouvillian superoperators.

$\mathcal{L}_{\mathrm{S+PM}}(\boldsymbol{\xi})$, bearing in mind that $\boldsymbol{\xi}$ includes the parameters related to both the system and the structured environments. Suppose that $\boldsymbol{\xi}_{\mathrm{EP}n}$ represents an nth-order EP in the parameter space, where n different eigenvalues and the corresponding eigenmatrices $\{\lambda_i, \hat{\rho}_{\mathrm{S+PM},i}\}_{i\in A}$ coalesce into $\{\lambda_{\mathrm{EP}}, \hat{\rho}_{\mathrm{S+PM},\lambda_{\mathrm{EP}}}\}$. Here, A denotes a set of indices. Due to the coalescence of the eigenmatrices, the corresponding n-dimensional eigensubspace for $\mathcal{L}_{\mathrm{S+PM}}(\boldsymbol{\xi}_{\mathrm{EP}})$ cannot be diagonalized. Nevertheless, a Jordan block for the subspace can be constructed by introducing generalized eigenmatrices $\{\hat{\rho}^{(j)}_{\mathrm{S+PM},\lambda_{\mathrm{EP}}}\}_{j=0,\cdots,n-1}$, such that the system dynamics can be expressed by

$$\rho_{\mathrm{S}}(t) = \sum_{i\notin A} c_i e^{\lambda_i t}\hat{\rho}_{\mathrm{S},i} + e^{\lambda_{\mathrm{EP}}t}\sum_{j=0}^{n-1}\sum_{m=0}^{j}\frac{t^m \tilde{c}_m}{m!}\hat{\rho}^{(j)}_{\mathrm{S},\lambda_{\mathrm{EP}}}, \quad (3)$$

where the reduced generalized eigenmatrices are introduced as $\hat{\rho}^{(j)}_{\mathrm{S},\lambda_{\mathrm{EP}}} = \mathrm{tr}_{\mathrm{PM}}(\hat{\rho}^{(j)}_{\mathrm{S+PM},\lambda_{\mathrm{EP}}})$. Equation (3) suggests that the polynomial time dependence, a common dynamical signature of EPs, could be observed in the reduced dynamics of the open system.

A similar procedure can be utilized under the framework of HEOM. In this context, a set of auxiliary density operators (ADOs) is introduced to capture the non-Markovian and non-perturbative effects. Similarly, we can define the extended quantum state $\rho_{\mathrm{S+ADO}}$ that contains both the system reduced state and the ADOs. The dynamics of the extended state is governed by the extended Liouvillian superoperator $\mathcal{L}_{\mathrm{S+ADO}}$. The system reduced state can be obtained through a linear operation, specifically $\rho_{\mathrm{S}}(t) = \mathcal{P}[\rho_{\mathrm{S+ADO}}(t)]$, where $\mathcal{P}$ is a superoperator for discarding all ADOs. Therefore, the EPs for non-Markovian open quantum systems can also be equivalently characterized under the framework of the HEOM by introducing the corresponding (generalized) reduced eigenmatrices, which are expressed as $\tilde{\rho}_{\mathrm{S},i} = \mathcal{P}(\tilde{\rho}_{\mathrm{S+ADO},i})$ and $\tilde{\rho}^{(j)}_{\mathrm{S},\lambda_{\mathrm{EP}}} = \mathcal{P}(\tilde{\rho}^{(j)}_{\mathrm{S+ADO},\lambda_{\mathrm{EP}}})$.

2.2 Example: Spin-boson model

Here, we consider a qubit representing the open system, where the system Hamiltonian and system-environment coupling operator are $H_{\mathrm{S}} = \omega_0|e\rangle\langle e|$ and $\tilde{Q} = \sigma_-$, respectively. We consider a Lorentzian spectral density that is expressed by

$$J_L(\omega) = \frac{1}{2}\frac{\Gamma\Lambda^2}{(\omega-\omega_0)^2+\Lambda^2}, \quad (4)$$

where Γ and Λ denote the coupling strength and the spectral width, respectively. In the interaction picture, the environmental correlation function can be expressed by a single exponential term, i.e., $C(t) = (\Gamma\Lambda/2)\exp(-\Lambda|t|)$. Therefore, the PMEOM can be constructed by introducing a single pseudomode with the damping rate $\gamma = 2\Lambda$ and the qubit-pseudomode coupling strength $\alpha = \sqrt{\Gamma\Lambda/2}$.

We find that $\Gamma = \Lambda/2$ corresponds to a second-order EP (EP2) and a third-order EP (EP3). Notably, these EPs are purely non-Markovian, because they are unobservable in the Markovian limit. Specifically, in such a limit, the spectral width (and thus the damping rate of the pseudomode) becomes infinite, $\Lambda \to \infty$. Therefore, the pseudomode can be adiabatically eliminated and the dynamics is governed by a qubit-only Markovian master equation, i.e., $\dot{\rho}_{\mathrm{S}}(t) = \Gamma[2\sigma_-\rho_{\mathrm{S}}(t)\sigma_+ - \{\sigma_+\sigma_-, \rho_{\mathrm{S}}(t)\}]/2$. Intuitively, there is only one qubit decay channel without internal tunneling between the qubit energy levels, thereby EP does not emerge in this scenario.

2.3 Example: Two coupled bosonic modes

Here, we examine two coupled resonant modes embadded in a Lorentzian environment. In the Markovian limit ($\Lambda \to \infty$), the resulting system-only effective non-Hermitian Hamiltonian within the rotating frame is given by:

$$\mathbf{H}_{\text{eff,S}} = \begin{pmatrix} 0 & \chi \\ \chi & i\Gamma \end{pmatrix}, \tag{5}$$

where χ denotes the coupling strength between the modes. The corresponding eigenvalues are $(i\Gamma \pm \sqrt{4\chi^2 - \Gamma^2})/2$, indicating the presence of an EP2 if $|\chi| = \Gamma/2$ with the degenerate eigenvalue $i\Gamma/2$. With a finite width Λ, the effective Hamiltonian takes the form:

$$\mathbf{H}_{\text{eff,S+PM}} = \begin{pmatrix} 0 & \chi & 0 \\ \chi & 0 & \sqrt{\frac{\Gamma\Lambda}{2}} \\ 0 & \sqrt{\frac{\Gamma\Lambda}{2}} & i\Lambda \end{pmatrix}. \tag{6}$$

By matching the coefficients of the characteristic polynomial, an EP3 is identified with the following criteria: $\{|\chi| = \Lambda/3\sqrt{3},\ \Gamma = 16\Lambda/27\}$, and the degenerate eigenvalue is $i\Lambda/3$. In other words, the EP can be transformed from second to third order with the introduction of the structured environmental characteristics.

This upgrade can lead to a further enhancement in the system sensitivity. For instance, we introduce a perturbation $\epsilon > 0$ to the coupling strength $\chi \to \chi(1+\epsilon)$. For the scenario in the exact Markovian limit, the eigenvalues in the vicinity of the EP2 are $\left\{i\Gamma/2 - \Gamma\sqrt{\epsilon}/\sqrt{2} + O\left(\epsilon^{3/2}\right), i\Gamma/2 + \Gamma\sqrt{\epsilon}/\sqrt{2} + O\left(\epsilon^{3/2}\right)\right\}$. In contrast to this case, for the scenario with a finite spectral width, the eigenvalues in the vicinity of the EP3 take the form $\{i\Lambda/3 + x_1\Lambda\epsilon^{1/3} + O(\epsilon^{2/3}), i\Lambda/3 + x_2\Lambda\epsilon^{1/3} + O(\epsilon^{2/3}), i\Lambda/3 + x_3\Lambda\epsilon^{1/3} + O(\epsilon^{2/3})\}$, where x_1, x_2, and x_3 are constants. We observe a change from a square-root bifurcation for the Markovian EP2 to a cubic-root bifurcation for the non-Markovian EP3 in response to the external perturbation, signifying the enhancement of the system sensitivity to the external perturbation.

3 Discussions

We have presented a general theory on characterizing non-Markovian EP based on pseudomode mapping and hierarchical equations of motion. We uncover the presence of purely non-Markovian EP. Additionally, the incorporation of non-Markovian effects can increase the order of the EP by effectively increasing the dimension of both the extended Liouvillian superoperator. This presents an innovative strategy for hunting higher-order EPs.

Although this work focuses exclusively on a bosonic environment, the proposed framework can be directly generalized to the scenarios with arbitrary combinations of bosonic and fermionic baths [3, 4]. Moreover, beyond the PMEOM and HEOM, our method of describing non-Markovian EPs via extended Liouvillian superoperators can also be applied to other pertinent methodologies, such as the dissipaton-embedded master equation [5] and reaction-coordinate mapping [6].

Future work involves further generalizing the theory of non-Markovian EPs. For instance, it is worthwhile to explore the potential applications emerging from the intricate interplay between the (non-)Markovian exceptional and diabolic points [7, 8] or the exotic topology and geometry of the parameter space [9, 10]. Such investigations could uncover new aspects of non-Markovian EPs, enhancing our understanding of open quantum systems embedded in environments with memory effects.

References

[1] F. Minganti, A. Miranowicz, R. W. Chhajlany, and F. Nori, "Quantum exceptional points of non-Hermitian Hamiltonians and Liouvillians: The effects of quantum jumps", Phys. Rev. A **100**, 062131 (2019).

[2] N. Lambert, S. Ahmed, M. Cirio, and F. Nori, "Modelling the ultra-strongly coupled spin-boson model with unphysical modes", Nat. Commun. **10**, 3721 (2019).

[3] N. Lambert, T. Raheja, S. Cross, P. Menczel, S. Ahmed, A. Pitchford, D. Burgarth, and F. Nori, "QuTiP-BoFiN: a bosonic and fermionic numerical hierarchical-equations-of-motion library with applications in light-harvesting, quantum control, and single-molecule electronics", Phys. Rev. Res. **5**, 013181 (2023).

[4] Y.-T. Huang, P.-C. Kuo, N. Lambert, M. Cirio, S. Cross, S.-L. Yang, F. Nori, and Y.-N. Chen, "An efficient Julia framework for hierarchical equations of motion in open quantum systems", Commun. Phys. **6**, 313 (2023).

[5] Y. Yan, "Theory of open quantum systems with bath of electrons and phonons and spins: Many-dissipaton density matrixes approach", J. Chem. Phys. **140** (2014).

[6] J. Iles-Smith, N. Lambert, and A. Nazir, "Environmental dynamics, correlations, and the emergence of noncanonical equilibrium states in open quantum systems", Phys. Rev. A **90**, 032114 (2014).

[7] I. I. Arkhipov, A. Miranowicz, F. Minganti, Ş. K. Özdemir, and F. Nori, "Dynamically crossing diabolic points while encircling exceptional curves: A programmable symmetric-asymmetric multimode switch", Nat. Commun. **14**, 2076 (2023).

[8] J. Perina Jr, A. Miranowicz, G. Chimczak, and A. Kowalewska-Kudlaszyk, "Quantum Liouvillian exceptional and diabolical points for bosonic fields with quadratic Hamiltonians: the Heisenberg-Langevin equation approach", Quantum **6**, 883 (2022).

[9] E. J. Bergholtz, J. C. Budich, and F. K. Kunst, "Exceptional topology of non-Hermitian systems", Rev. Mod. Phys. **93**, 015005 (2021).

[10] C.-Y. Ju, A. Miranowicz, F. Minganti, C.-T. Chan, G.-Y. Chen, and F. Nori, "Einstein's quantum elevator: Hermitization of non-Hermitian Hamiltonians via a generalized vielbein formalism", Phys. Rev. Res. **4**, 023070 (2022).

Optimal quantum sampling on distributed databases

Longyun Chen[1] * Jingcheng Liu[1] † Penghui Yao[1 2] ‡

[1] *State Key Laboratory for Novel Software Technology, Nanjing University*
[2] *Hefei National Laboratory*

Abstract.
Quantum sampling, a fundamental subroutine in numerous quantum algorithms, involves encoding a given probability distribution in the amplitudes of a pure state. In light of the hefty cost of large-scale quantum storage, we initiate the study of quantum sampling in a distributed setting. Specifically, we assume that the data is distributed among multiple machines, and each machine solely maintains a basic oracle that counts the multiplicity of individual elements. Given a quantum sampling task, which is to sample from the joint database, a coordinator can make oracle queries to all machines. We focus on the oblivious communication model, where communication between the coordinator and the machines is predetermined. We present both sequential and parallel algorithms: the sequential algorithm queries the machines sequentially, while the parallel algorithm allows the coordinator to query all the machines simultaneously. Furthermore, we prove that both algorithms are optimal in their respective settings.

Keywords: Quantum sampling, distributed quantum computing, quantum query complexity, adversary method

1 Introduction

Quantum sampling is a fundamental computational task in quantum computing that encodes a given distribution in the amplitudes of a quantum state. More specifically, the algorithm has access to a distribution $(p_1, \ldots, p_N)$ and is supposed to output the state $\sum_{i=1}^{N} \sqrt{p_i} \left|i\right\rangle$, where $\{\left|1\right\rangle, \ldots, \left|N\right\rangle\}$ is a set of computational bases. Quantum sampling was inspired by the famous Grover's algorithm [10] and is nowadays a key subroutine in many quantum algorithms. For example, the well-known Harrow-Hassidim-Lloyd algorithm [13], which solves a system of linear equations $Ax = b$ with an exponential speedup over the fastest classical algorithm, requires encoding the vector b to the amplitude of a pure state $\left|b\right\rangle = \sum_i b_i \left|i\right\rangle$ up to normalization. Many quantum algorithms for learning functions and distributions also require quantum sampling on a given distribution [9, 5, 6]. It is also known that the quantum advantages of certain quantum learning algorithms require quantum sampling and the advantages would vanish if quantum sampling was replaced by classical sampling [9]. Moreover, quantum sampling has also found many algorithmic applications, such as quantum walk [17, 21, 22], quantum mean estimation [8, 11, 12], and quantum coupon collector [4]. Thus, a number of works have been devoted to designing algorithms and analyzing the complexities of quantum sampling [3, 19, 15].

It is still challenging to have quantum storage for big data. Due to the limit of large-scale quantum computers, this paper initiates the study of *distributed quantum sampling*. A distributed database consists of several machines and a coordinator. The distributed database has a publicly known maximum capacity ν for each kind of element, which is an upper bound on multiplicities of the elements. With the oracle design declared later, the parameter ν is needed to bound the dimension of the register. A large dataset is distributed across these machines, and each machine implements a simple oracle that maintains how many times an element appears in its share of the dataset. The goal of the coordinator is to produce the quantum state $\sum_{i=1}^{N} \sqrt{p_i} \left|i\right\rangle$, where p_i is the probability that you get i when sampling uniformly from the distributed database. To do so, the coordinator can only communicate with the machines through oracle queries. If we only allow classical communications, then the coordinator has to send queries to each machine, asking the multiplicity of every possible element. After the coordinator has learned p_i for every i, then it has to prepare the quantum state by itself. For a dataset with a data universe of N elements containing all distinct elements distributed across n machines, the query complexity could be as large as nN. We study the quantum query complexity where we allow quantum communications between the coordinator and the machines, and we are able to show significant speed-up.

Main results

In this paper, we exhibit two distributed quantum sampling algorithms using sequential and parallel queries, respectively. Informally speaking, in the sequential model, the coordinator makes queries to each machine sequentially. In the parallel model, the coordinator queries all machines simultaneously. We further show that both algorithms achieve optimal query complexity among all oblivious algorithms.

Theorem 1 (Main result, informal) *Given a distributed database consisting of n machines with maximum capacity ν, there exists an algorithm for quantum sampling with $O(n\sqrt{\nu N/M})$ sequential queries in the oblivious model. If parallel queries are allowed, then $O(\sqrt{\nu N/M})$ queries are sufficient. Here N is the size of the data universe, and M is the total number of elements stored across the distributed database counting multiplic-*

*longyunchen@smail.nju.edu.cn
†liu@nju.edu.cn
‡phyao1985@gmail.com

ities. Moreover, both algorithms are optimal in the oblivious communication model.

These two optimal algorithms are designed by directly expanding the centralized quantum sampling algorithm. Their optimality is established by extending the method proposed by Zalka for Grover's algorithm in the centralized setting [23]. Our proof suggests that the essential barrier for distributed quantum sampling is the same as that for the centralized setting.

Related work

The problem of quantum sampling was raised after Grover's algorithm [10]. The quantum query complexity of quantum sampling has been studied in various contexts. Shi [20] introduced the problem of index erasure: given an injective function $f : [n] \to [m]$ via a black-box oracle, the task is to prepare the quantum state, which is the uniform superposition on the image of f, i.e., $\sum_{x=1}^{n} |f(x)\rangle / \sqrt{n}$. Index erasure can be viewed as a uniform quantum sampling over a subset of the universe. This problem is closely related to graph isomorphism, and the tight query complexity of the problem was later established by Ambainis, Magnin, Roetteler, and Roland in the coherent setting [3] and by Lindzey and Rosmanis in the non-coherent setting [16]. Ozols, Roetteler, and Roland [19] further introduced quantum rejection sampling, which converts a quantum state to another quantum state with a given amplitude. The quantum query complexity of quantum state conversion has been established in [15].

In addition to quantum query complexity, quantum sampling has also been studied in other models of computation. Aharonov and Ta-Shama studied the problem of preparing $\sum_{i\in\Omega} \sqrt{p_i} |i\rangle$ given the description of a classical circuit with output distribution p. A weaker quantum sampling model, where an extra register is allowed, that is, $\sum_i \sqrt{p_i} |i\rangle |c_i\rangle$, has also been considered in [12, 6].

2 Preliminaries

For integer $N > 0$, let $[N]$ represent the set $\{1, \cdots, N\}$. Given a multiset S, Supp(S) represents the support of S. For any element x, the multiplicity of x is the number of occurrences in S. The cardinality of a multiset S, denoted by $|S|$, is the sum of the multiplicities of all its elements.

Here we give a brief introduction to quantum computing and the notations used in this paper. Readers may refer to [18] for a thorough treatment. Consider a Hilbert space $\mathcal{H}$ endowed with an inner product $\langle \cdot, \cdot \rangle$. A quantum state is a positive semidefinite matrix with a trace equal to 1. Let $|\psi\rangle$ be a vector in $\mathcal{H}$. The norm of $|\psi\rangle$, denoted by $\||\psi\rangle\|$ is defined to be $\|\psi\| := \sqrt{\langle\psi|\,\psi\rangle}$. For any two vectors $|\phi\rangle$ and $|\psi\rangle$ in $\mathcal{H}$, the distance between them is $\||\phi\rangle - |\psi\rangle\|$. A quantum register A is associated with a Hilbert space $\mathcal{H}_A$. The composition of two registers A and B, denoted by AB, is associated with the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. The identity operator on $\mathcal{H}_A$, (and associated register A) is denoted I_A. The subscript A may be omitted when it is clear from the context.

3 Distributed databases

In this section, we formally introduce the model of distributed databases considered in this paper. A distributed database consists of several databases, each of which stores part of the data and is maintained by a machine. Moreover, each machine also implements some simple operations. There is a coordinator who makes queries to each machine and outputs the answer at the end of the algorithm. The coordinator is assumed to be a quantum computer. In this paper, we assume that the coordinator sends an element from the dataset to a machine, and the machine answers the multiplicity of the element, i.e., the number of occurrences of the element in the machine. A mathematical formulation is given below. We are interested in minimizing the number of quantum queries made by the coordinator.

In this paper, we only consider the *oblivious* communication model, where the order of the communication between the coordinator and the machines is predetermined (only depends on the public knowledge known to the coordinator). The oblivious communication model has been studied in [14].

We conjecture that non-oblivious communication does not help us to save the number of queries. It is worth noting that the final output of the algorithm is supposed to be a pure state. Thus, in the quantum circuits model, all intermediate measurements can be removed by the principle of deferred measurement and the gentle measurement argument [18, 1]. However, it is not clear whether they can be extended to a distributed setting. We leave it for future work.

Quantum sampling on distributed databases

Suppose that the data universe is represented by the set $[N] := \{1, \cdots, N\}$, and the dataset is distributed among n machines. The coordinator maintains a quantum state with three registers

$$|\rho\rangle = \sum_{i\in[N]} \alpha_i |i\rangle |s_i\rangle |w_i\rangle .$$

The first register is N-dimensional for element storage, the second register is $(\nu + 1)$-dimensional to store the outcome of the oracle, and the last one is the ancillary register, whose dimension remains to be determined by the algorithm design. In our algorithm, w_i should belong to $\{0, 1\}$.

In this paper, we consider two models of queries: *sequential* queries and *parallel* queries. In the sequential model, the coordinator sends queries to the machines sequentially. In a sequential model, suppose the coordinator makes a query to the j-th machine. It sends the first two registers to the j-th machine. The j-th machine implements the following operation $\mathcal{O}_j$:

$$\mathcal{O}_j |i\rangle |s\rangle = |i\rangle |(s + c_{ij}) \mod (\nu + 1)\rangle , \qquad (1)$$

Table 1: Table of Notations

Symbol	Meaning
n	the count of the machines
N	the number of the varieties of elements
T_j	the dataset (multiset) on the j-th machine
c_{ij}	the multiplicity of element i in T_j
$c_i := \sum_{j\in[n]} c_{ij}$	the total count of occurrences of i across all machines
$M := \sum_{i\in[N]} c_i$	the total count of the elements over all machines
$M_j := \lvert T_j\rvert$	the count of the elements on the j-th machine
$\mathrm{Supp}(T_j)$	the support set of distinct elements in T_j
$m_j := \lvert\mathrm{Supp}(T_j)\rvert$	the count of distinct elements on the j-th machine
ν	maximum capacity of the database
$\lvert\psi\rangle$	the quantum sampling state defined in Eq.(4)
$\lvert\pi\rangle := \frac{1}{\sqrt{N}}\sum_{i\in[N]}\lvert i\rangle$	the uniform superposition state
$\mathcal{O}_j$	the oracle of the j-th machine

where c_{ij} is the multiplicity of an element i in the j-th machine and ν is the maximum capacity of the distributed database, which is known to the coordinator. Thus, $\nu \geq \max_{i\in[N]}(\sum_{j=1}^{n} c_{ij})$ is an upper bound for the multiplicities of the elements.

It is worth noting that the oracle operation can be easily extended to a dynamic database. It is low-cost to update oracle operation $\mathcal{O}_j$ if the datasets are changed. For instance, if the multiplicity of element i in the j-th machine increases or decreases by 1, i.e., c_{ij} increases or decreases by 1, we can simply update $\mathcal{O}_j$ by left multiplying operator U or $U^\dagger$, respectively, where $U\lvert i\rangle\lvert s\rangle = \lvert i\rangle\lvert(s+1) \mod (\nu+1)\rangle$.

In the parallel model, the coordinator may send multiple queries to distinct machines simultaneously. To be more specific, in the parallel model, the state with the coordinator contains four registers

$$\lvert\rho\rangle = \sum_{\bar{i}\in[N]^n} \alpha_{\bar{i}}\lvert\bar{i}\rangle\lvert s^{\bar{i}}\rangle\lvert b^{\bar{i}}\rangle\lvert w_{\bar{i}}\rangle,$$

where $\bar{i} = (i_1,\ldots,i_n) \in [N]^n, b^{\bar{i}} \in \{0,1\}^n$. Thus, each of the first three registers contains n qudits. When the coordinator makes a query, it sends three qudits, one from each of the first three registers, to each machine. For $j \in [n]$, the j-th machine is implementing

$$\hat{\mathcal{O}}_j\lvert i_j\rangle\lvert s_j\rangle\lvert b_j\rangle = \lvert i_j\rangle\lvert s_j + c_{i_j,j}\cdot b_j \mod (\nu+1)\rangle\lvert b_j\rangle \tag{2}$$

where ν is an upper bound on the multiplicities of i_j. It is not hard to see that the operation in Equation (2) can be realized by the query operation in the sequential model defined in Equation (1). Thus a parallel query

$$\mathcal{O}\lvert\bar{i}\rangle\lvert s_1\cdots s_n\rangle\lvert b_1\cdots b_n\rangle = \bigotimes_{j=1}^{n}\hat{\mathcal{O}}_j\lvert i_j, s_j, b_j\rangle \tag{3}$$

can be implemented by n sequential queries.

To describe the problem of quantum sampling, we need to further introduce some notations. For $j \in [n]$, the dataset on the j-th machine is denoted by a multiset T_j. Then, the multiset T_j is completely determined by the values of c_{ij} defined in Equation (1).

Now we are ready to formally define the problem of quantum sampling on distributed databases. Given datasets $\{T_j\}_{j\in[n]}$, a quantum sampling algorithm is supposed to produce the state

$$\lvert\psi\rangle = \frac{1}{\sqrt{M}}\sum_{i\in[N]}\sqrt{c_i}\lvert i\rangle, \tag{4}$$

where $c_i = \sum_{j\in[n]} c_{ij}$ is the total occurrences of the element i across all machines and $M = \sum_{i\in[N]} c_i$ is the total count of all elements on the machines. Since c_i/M is the frequency of the element i appearing over all machines, measuring the state $\lvert\psi\rangle$ under the computational basis $\{\lvert i\rangle\}_{i\in[N]}$ is equivalent to sampling over the datasets.

All parameters and notations are summarized in Table 1.

4 Algorithms

By the amplitude amplification [7], we can obtain quantum sampling algorithms for sequential and parallel queries. Their details are specified in Appendix A.

Theorem 2 (Sampling with sequential queries) *Given parameters $\varepsilon \in (0,1)$, and N, M, n, ν as in Table 1 satisfying $\nu \geq \frac{M}{N\varepsilon}$, there exists an algorithm for quantum sampling which makes $O\left(n\sqrt{\nu N/M}\right)$ queries and outputs a state $\lvert\phi\rangle$ satisfying that $\lvert\langle\phi\lvert\psi\rangle\rvert \geq 1-\varepsilon$, where $\lvert\psi\rangle$ is the quantum sampling state defined in Equation (4).*

Theorem 3 (Sampling with parallel queries) *Given parameters $\varepsilon \in (0,1)$, and N, M, n, ν as in Table 1 satisfying $\nu \geq \frac{M}{N\varepsilon}$, there exists an algorithm for quantum sampling which makes $O\left(\sqrt{\nu N/M}\right)$ parallel queries and outputs a state $\lvert\phi\rangle$ satisfying that $\lvert\langle\phi\lvert\psi\rangle\rvert \geq 1-\varepsilon$.*

Both of them are optimal within their models separately with respect to the quantum query complexity, which is concluded by showing the lower bound on the queries in Appendix B.

References

[1] Scott Aaronson and Guy N. Rothblum. Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 322–333, New York, NY, USA, 2019. Association for Computing Machinery.

[2] Andris Ambainis. Quantum lower bounds by quantum arguments. *Journal of Computer and System Sciences*, 64(4):750–767, 2002.

[3] Andris Ambainis, Loïck Magnin, Martin Roetteler, and Jeremie Roland. Symmetry-assisted adversaries for quantum state generation. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 167–177, 2011.

[4] Srinivasan Arunachalam, Aleksandrs Belovs, Andrew M. Childs, Robin Kothari, Ansis Rosmanis, and Ronald de Wolf. Quantum Coupon Collector. In Steven T. Flammia, editor, *15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020)*, volume 158 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:17, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[5] Srinivasan Arunachalam, Sourav Chakraborty, Troy Lee, Manaswi Paraashar, and Ronald de Wolf. Two new results about quantum exact learning. *Quantum*, 5:587, November 2021.

[6] Srinivasan Arunachalam and Ronald De Wolf. Optimal quantum sample complexity of learning algorithms. *J. Mach. Learn. Res.*, 19(1):2879–2878, jan 2018.

[7] Gilles Brassard, Peter Hø yer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. In *Quantum computation and information (Washington, DC, 2000)*, volume 305 of *Contemp. Math.*, pages 53–74. Amer. Math. Soc., Providence, RI, 2002.

[8] Arjan Cornelissen, Yassine Hamoudi, and Sofiene Jerbi. Near-optimal quantum algorithms for multivariate mean estimation. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 33–43, New York, NY, USA, 2022. Association for Computing Machinery.

[9] András Gilyén and Tongyang Li. Distributional Property Testing in a Quantum World. In Thomas Vidick, editor, *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 25:1–25:19, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[10] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, page 212–219, New York, NY, USA, 1996. Association for Computing Machinery.

[11] Yassine Hamoudi. Quantum Sub-Gaussian Mean Estimator. In *29th Annual European Symposium on Algorithms*, Lisbon, Portugal, September 2021.

[12] Yassine Hamoudi and Frédéric Magniez. Quantum Chebyshev's Inequality and Applications. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, Patras, Greece, July 2019.

[13] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103:150502, Oct 2009.

[14] François Le Gall and Daiki Suruga. *Bounds on Oblivious Multiparty Quantum Communication Complexity*, page 641–657. Springer International Publishing, 2022.

[15] Troy Lee, Rajat Mittal, Ben W. Reichardt, Robert Špalek, and Mario Szegedy. Quantum query complexity of state conversion. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 344–353, 2011.

[16] Nathan Lindzey and Ansis Rosmanis. A Tight Lower Bound For Non-Coherent Index Erasure. In Thomas Vidick, editor, *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 59:1–59:37, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[17] Frédéric Magniez, Ashwin Nayak, Jérémie Roland, and Miklos Santha. Search via quantum walk. *SIAM Journal on Computing*, 40(1):142–164, 2011.

[18] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, USA, 10th edition, 2011.

[19] Maris Ozols, Martin Roetteler, and Jérémie Roland. Quantum rejection sampling. *ACM Trans. Comput. Theory*, 5(3), aug 2013.

[20] Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 513–519, 2002.

[21] M. Szegedy. Quantum speed-up of markov chain based algorithms. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 32–41, 2004.

[22] Pawel Wocjan and Anura Abeyesinghe. Speedup via quantum sampling. *Phys. Rev. A*, 78:042336, Oct 2008.

[23] Christof Zalka. Grover's quantum searching algorithm is optimal. *Physical Review A*, 60(4):2746–2751, oct 1999.

A Algorithm design

A.1 Sequential queries

We start by giving a quantum sampling algorithm for the sequential model. A key operator in our algorithm is the *distributing operator* D such that

$$D\left|i,0\right\rangle = \sqrt{\frac{c_i}{\nu}}\left|i,0\right\rangle + \sqrt{\frac{\nu-c_i}{\nu}}\left|i,1\right\rangle, \tag{5}$$

We claim that there exists a unitary operator satisfying the above equation, and is thus a valid quantum operation.

Lemma 4 *The operator D can be extended to a unitary operator on the whole Hilbert space.*

Proof. Eq. (5) defines the operator D on the domain of a subspace spanned by $\{|i,0\rangle\}_{i\in[N]}$. It can be checked that $\langle i,0|D^\dagger D|j,0\rangle = \delta_{ij}$ with the Kronecker notation δ_{ij}. Thus, D preserves the inner product on the subspace, which can be extended to a unitary operator on the whole space. □

Notice that the operator D depends on the input $\{T_j\}$ due to its definition in Equation (5). The next lemma shows that the operator D can be realized by $2n$ calls of $\mathcal{O}_j$'s.

Lemma 5 *The operator D can be implemented with $2n$ queries given by the oracles $\mathcal{O}_j$ defined in Eq.* (1) *and unitary operators independent of the input.*

Proof. The implementation of the operator D is given by the following three steps:

$$\begin{aligned}
|i,0,0\rangle &\xrightarrow{\mathcal{O}_1\cdots\mathcal{O}_n\otimes I} |i,c_i,0\rangle \\
&\xrightarrow{\mathcal{U}} \sqrt{\frac{c_i}{\nu}}\left|i,c_i,0\right\rangle + \sqrt{\frac{\nu-c_i}{\nu}}\left|i,c_i,1\right\rangle \\
&\xrightarrow{\mathcal{O}_1^\dagger\cdots\mathcal{O}_n^\dagger\otimes I} \sqrt{\frac{c_i}{\nu}}\left|i,0,0\right\rangle + \sqrt{\frac{\nu-c_i}{\nu}}\left|i,0,1\right\rangle.
\end{aligned}$$

The first and third steps can be realized by queries to n machines. The unitary operator $\mathcal{U}$ is defined to satisfy

$$\mathcal{U}\left|i,c,0\right\rangle = \sqrt{\frac{c}{\nu}}\left|i,c,0\right\rangle + \sqrt{\frac{\nu-c}{\nu}}\left|i,c,1\right\rangle, \tag{6}$$

which is independent of the input. It is not hard to see $\langle i,c,0|\mathcal{U}^\dagger\mathcal{U}|i',c',0\rangle = \delta_{(i,c),(i',c')}$. Thus $\mathcal{U}$ is a unitary operator. □

Proof of Theorem 2. The initial state of our algorithm is the uniform superposition state

$$|\pi\rangle = \frac{1}{\sqrt{N}}\sum_{i\in[N]}|i\rangle.$$

Recall the operator D in Equation (5). By direct calculation, we have

$$\begin{aligned}
D\left|\pi,0\right\rangle &= \frac{1}{\sqrt{N}}\sum_{i\in[N]}\left(\sqrt{\frac{c_i}{\nu}}\left|i,0\right\rangle + \sqrt{\frac{\nu-c_i}{\nu}}\left|i,1\right\rangle\right) \\
&= \sqrt{\frac{M}{\nu N}}\left|\psi,0\right\rangle + \sqrt{1-\frac{M}{\nu N}}\left|\psi^\perp,1\right\rangle, \qquad (7)
\end{aligned}$$

where $|\psi\rangle$ is the target state defined in eq. (4), and $|\psi^\perp,1\rangle$ is a pure state orthogonal to $|\psi,0\rangle$. Applying amplitude amplification [7] with $O(\sqrt{\nu N/M})$ calls of D, the final state $|\phi\rangle$ satisfies that $|\langle\phi\,|\psi\rangle\,| \geq \sqrt{(\nu N-M)/\nu N} \geq 1-\varepsilon$ by the choice of parameters.

□

A.2 Parallel queries

Modifying the algorithm for the sequential model, we then give a sampling algorithm for the parallel model. We still adopt the sampling algorithm in Theorem 2 given by amplitude amplification. The only change is the implementation of the operator

$$D : |i,0\rangle \longmapsto \sqrt{\frac{c_i}{\nu}}\left|i,0\right\rangle + \sqrt{\frac{\nu-c_i}{\nu}}\left|i,1\right\rangle$$

to reduce the query complexity.

Lemma 6 *The operator D can be implemented with four queries given by the parallel query oracle $\mathcal{O}$ defined in Eq.* (3) *and unitary operators independent of the input.*

Proof. By the proof of Theorem 5, the operator D can be implemented in three steps. The second step is a unitary operator $\mathcal{U}$ independent of the input. Here, we are going to realize the first and the third steps with the oracle $\mathcal{O}$. The first step is $|i,0\rangle \mapsto |i,c_i\rangle$. With a ancillary registers initialized as $|0^n,0^n,0^n\rangle$, this step can be completed by

$$\begin{aligned}
|i,0,0^n,0^n,0^n\rangle &\longrightarrow |i,0,i^n,0^n,1^n\rangle \\
&\xrightarrow{I\otimes I\otimes\mathcal{O}} |i,0,i^n,c_{i1}c_{i2}\cdots c_{in},1^n\rangle \\
&\longrightarrow |i,c_i,i^n,c_{i1}c_{i2}\cdots c_{in},1^n\rangle \\
&\xrightarrow{I\otimes I\otimes\mathcal{O}^\dagger} |i,c_i,i^n,0^n,1^n\rangle \\
&\longrightarrow |i,c_i,0^n,0^n,0^n\rangle.
\end{aligned}$$

This procedure only applies $\mathcal{O}$ twice. The third step is just the inverse of the first step, which can be completed similarly. □

Proof of Theorem 3. Similarly to the proof of Theorem 2, the algorithm begins with the uniform superposition state $|\pi\rangle$ and applies amplitude amplification [7] with $O(\sqrt{\nu N/M})$ calls of D, obtaining a final state $|\psi\rangle$ satisfying $|\langle\phi|\psi\rangle| \geq 1-\epsilon$. □

B Optimality

In this section, we prove that our algorithms are optimal. We consider a slightly more general setting where

each machine may have an independent maximum capacity κ_j. Thus for each $j \in [n]$, it holds $\max_{i\in[N]} c_{ij} \le \kappa_j \le \nu$. For the case where κ_j is unknown, we can just use ν for κ_j.

And the j-th machine maintains quantum oracles $\mathcal{O}_j$ and $\hat{\mathcal{O}}_j$ with the functionality

$$\mathcal{O}_j |i\rangle |s\rangle = |i\rangle |(s + c_{ij}) \mod (\nu+1)\rangle,$$
$$\hat{\mathcal{O}}_j |i,s,b\rangle = \begin{cases} (\mathcal{O}_j |i,s\rangle) \otimes |b\rangle, & b = 1, \\ |i,s,b\rangle, & b = 0. \end{cases}$$

Theorem 7 (Lower bound for the sequential model) *For any sampling algorithm in an oblivious communication model with sequential queries, if it satisfies that $F(\rho,\psi) > 9/16$, where ρ is the output state and ψ is the quantum sampling state, then its query complexity is at least $\Omega\left(\sum_{j\in[n]} \sqrt{\frac{\kappa_j N}{M}}\right)$.*

Theorem 8 (Lower bound for the parallel model) *For any sampling algorithm in an oblivious communication model with parallel queries, if it satisfies that $F(\rho,\psi) > 9/16$, where ρ is the output state and ψ is the quantum sampling state, then its query complexity is at least $\Omega\left(\max_{j\in[n]} \sqrt{\frac{\kappa_j N}{M}}\right)$.*

B.1 Oblivious queries with measurement

We start by observing that, given an oblivious algorithm with measurements for quantum sampling, one can construct an algorithm without any measurements that has the same query complexity. Therefore, we can focus on algorithms without measurements in the rest of the paper.

Lemma 9 *Let $\mathcal{A}$ be an oblivious algorithm with measurements for quantum sampling. Then there exists an algorithm $\mathcal{B}$ without measurements, which has the same query complexity and fidelity.*

It can be proved by the deferred measurement principle with minor modification. The proof is deferred in Appendix C.

B.2 Hard inputs

We prove the optimality via the quantum adversary argument [2]. To this end, we describe the hard inputs in this subsection.

Let t_k be the number of times the oracle $\hat{\mathcal{O}}_k$ and $\hat{\mathcal{O}}_k^\dagger$ is applied; then the query complexity is $\sum_{k\in[n]} t_k$. We prove a lower bound for each t_k. Notice that for an oblivious model, the order of queries is independent of the inputs. Thus, this lower bound applies to all inputs, which implies that the summation of the individual lower bounds is a lower bound for the total query complexity.

Let σ be a permutation on $[N]$ and $S \subseteq [N]$ be a subset. We say σ is *order-preserving* for S if for any $r,t \in S$, it holds that $\sigma(r) < \sigma(t)$ if and only if $r < t$.

We fix any integer k for the rest of this section. Given a sequence of multisets $T = (T_1, T_2, \cdots, T_n)$ and an order-preserving σ for $\mathrm{Supp}(T_k)$, we permute T_k by σ^{-1} to obtain a new sequence of multisets $T' = (T_1, T_2, \cdots, T'_k, \cdots, T_n)$. Specifically, define

$$c'_{ij} = \begin{cases} c_{ij}, & j \neq k, \\ c_{\sigma^{-1}(i)j}, & j = k, \end{cases}$$

where c_{ij} are the multiplicities for T_j. Notice that c'_{ij} uniquely define a sequence of $\{T'_j\}$. We write $\tilde{\sigma}^k(T) := \{T'_j\}$, and call $\tilde{\sigma}_k$ as the σ-induced permutation.

Definition 10 (Hard input condition) *Given $k \in [n]$, constants $\alpha,\beta \in (0,1]$, and a sequence of multisets $T = (T_1, T_2, \cdots, T_n)$ distributed on n machines, then T satisfies the hard input condition if*

$$M_k \ge \alpha M, \quad M_k/m_k \ge \beta\kappa_k, \quad \max_{i\in[N],j\neq k} c_{ij} + \max_{i\in[N]} c_{ik} \le \nu, \tag{8}$$

where $M_k = |T_k|$, $m_k = |\mathrm{Supp}(T_k)|$ and c_{ij} is the multiplicity of i in T_j.

Definition 11 (Hard inputs) *Given $k \in [n]$, constants $\alpha,\beta \in (0,1]$, and a sequence of multisets $T = (T_1, T_2, \cdots, T_n)$ satisfying the hard input condition in Theorem 10, the collection of hard inputs for the k-th machine is defined as*

$$\mathcal{T} := \{\tilde{\sigma}^k(T) : \sigma \text{ is order-preserving for } \mathrm{Supp}(T_k)\},$$

where $\tilde{\sigma}^k$ is the σ-induced permutation defined above.

The last condition in Theorem 10 guarantees $\tilde{\sigma}_k(T) \in \mathcal{T}$ is still a hard input with multiplicities not greater than ν.

The following lemma gives the size of the collection of hard inputs.

Lemma 12 *For any $k \in [N]$, constants $\alpha,\beta \in (0,1]$, and a sequence of multisets $T = (T_1, T_2, \cdots, T_n)$ satisfying the hard input condition, let $\mathcal{T}$ be the collection of hard inputs as given in Theorem 11. It holds that $|\mathcal{T}| = \binom{N}{m_k}$ with $m_k := |\mathrm{Supp}(T_k)|$.*

Proof. Let $S = \mathrm{Supp}(T_k)$. A claim should be concluded to calculate the size of $\mathcal{T}$.

For σ_1, σ_2 that are order-preserving for S, we claim that $\tilde{\sigma}_1^k(T) = \tilde{\sigma}_2^k(T)$ if and only if $\sigma_1(i) = \sigma_2(i)$ for every $i \in S$. The sufficiency is obvious. For the necessity, we prove by contradiction. Suppose $\tilde{\sigma}_1^k(T) = \tilde{\sigma}_2^k(T)$ and $\sigma_1(i_0) \neq \sigma_2(i_0)$ for some $i_0 \in S$. Consider the multiplicity $c'_{\sigma_1(i_0)k}$ for $T' := \tilde{\sigma}_1^k(T) = \tilde{\sigma}_2^k(T)$, it follows that $c_{i_0k} = c_{\sigma_2^{-1}(\sigma_1(i_0))k} > 0$. Denote $\sigma_2^{-1}(\sigma_1(i_0))$ by i_1. Since $\sigma_1(i_0) \neq \sigma_2(i_0)$, it holds that $i_0 \neq i_1$. If $i_0 < i_1$, then the order-preserving property of σ_1, σ_2 implies

$$\begin{aligned} |\{i > \sigma_1(i_0) | c'_{ik} > 0\}| &= |\{i > i_0 | c_{ik} > 0\}| \\ &> |\{i > i_1 | c_{ik} > 0\}| \\ &= |\{i > \sigma_2(i_1) = \sigma_1(i_0) | c'_{ik} > 0\}|, \end{aligned}$$

The inequality holds because i_1 belongs to the former set but not the latter set. The first expression and the last

expression are the same, which leads to a contradiction. It is similar for the case of $i_0 > i_1$.

With this claim, the size $|\mathcal{T}|$ equals the count of order-preserving permutations that act differently on S. Finding such a permutation is equivalent to choosing $|S|$ elements in $[N]$ as the image set $\sigma(S)$. Hence, we have $|\mathcal{T}| = \binom{N}{|S|} = \binom{N}{m_k}$, as $|S| = |\operatorname{Supp}(T_k)| =: m_k$. □

B.3 Lower bound on sequential queries

We are now ready to derive a lower bound on the query complexity. As argued above, we can bound each t_k independently. To do so, for every $k \in [n]$, we consider a collection of hard inputs $\mathcal{T}$ generated by an input T with respect to k, as given in Theorem 11.

Given an input T, let $|\psi_t^T\rangle$ be the state after t calls to the oracle. It can be expressed as

$$|\psi_t^T\rangle = U_t O_t U_{t-1} O_{t-1} \cdots U_1 O_1 U_0 |0\rangle , \tag{9}$$

where $O_1, \cdots, O_t$ are either $\hat{\mathcal{O}}_k \otimes I$ or $\hat{\mathcal{O}}_k^\dagger \otimes I$ with identity operator I on the registers that $\hat{\mathcal{O}}_k$ does not act on, and $U_0, \cdots, U_t$ are unitary operators that are independent of T_k, the datasets on the k-th machine. We consider an input $\tilde{T}$ obtained from T by removing the dataset on the k-th machine. That is, we replace T_k with an empty set, and the datasets on other machines are the same as those in T. Notice that, each of the $\{O_i\}$ is an identity operator if the dataset on the k-th machine is empty. Thus, for the dataset $\tilde{T}$, the state after t calls of the oracle is

$$|\psi_t\rangle = U_t U_{t-1} \cdots U_1 U_0 |0\rangle . \tag{10}$$

We introduce a potential function as follows:

$$D_t = \frac{1}{|\mathcal{T}|} \sum_{T \in \mathcal{T}} \left\| |\psi_t^T\rangle - |\psi_t\rangle \right\|^2 . \tag{11}$$

We note that for our collection $\mathcal{T}$, while the state $|\psi_t^T\rangle$ depends on the specific choice of $T \in \mathcal{T}$, the state $|\psi_t\rangle$ remains the same regardless of T. To see this, note that for any pair $T, T' \in \mathcal{T}$, T and T' only differs in T_k. Further, due to obliviousness of queries, the oracles not involving $\hat{\mathcal{O}}_k$ or $\hat{\mathcal{O}}_k^\dagger$ remain the same for $\tilde{T}$ (and hence for T and T'), regardless of T_k. Therefore, any input T and T' that only differs in T_k share the same state $|\psi_t\rangle$.

To elaborate our proof ideas for the lower bound, we first note that $|\psi_{t_k}^T\rangle$ is the final state of the algorithm, and it must hold that the state of the first register in $|\psi_{t_k}^T\rangle$ approximates the goal state $|\psi\rangle$. Informally speaking, since the goal state $|\psi\rangle$ also depends on T_k, which is different for different $T \in \mathcal{T}$, any sampling algorithm must bring a large enough difference between $|\psi_{t_k}^T\rangle$ and $|\psi_{t_k}\rangle$ to get $|\psi\rangle$ for every $T \in \mathcal{T}$. Through this intuition, uniformly picking a $T \in \mathcal{T}$, we consider the expectation of the variation $D_t = \mathbb{E}_{\mathcal{T}} \left[\left\| |\psi_t^T\rangle - |\psi_t\rangle \right\|^2 \right]$.

The lower bound on t_k can be obtained by the following two lemmas.

Lemma 13 *Let $\mathcal{T}$ be the collection of hard inputs for the k-th machine as in Theorem 11. Let α, β be the constants defined in Theorem 11. Suppose the fidelity between the output state ρ and the quantum sampling state ψ defined in Eq. (4) satisfies $F(\rho, \psi) \geq (1-\epsilon)^2 > 9/16$ with $\epsilon \geq 0$. If $M < \beta^2 \kappa_k N/16$ and $\alpha > 4\epsilon$, then the expectation of the variation is bounded by $D_t \geq C \frac{M_k}{M}$ for some constant C dependent to α and ϵ.*

Lemma 14 *Let $\mathcal{T}$ be the collection of hard inputs for the k-th machine as in Theorem 11. For $t \leq t_k$, it holds that $D_{t_k} \leq 4 \frac{m_k}{N} t^2$.*

With these two lemmas, we are ready to show a lower bound for the query complexity. *Proof of theorem 7.* We start by showing that for each $j \in [n], t_j = \Omega\left(\sqrt{\kappa_j N/M}\right)$. Fixing a $k \in [n]$, for a constant $\beta \in (0, 1]$, we split the proof into two cases of $M \geq \beta^2/16 \kappa_k N$ and $M < \beta^2 \kappa_k N/16$.

For $M \geq \beta^2 \kappa_k N/16$, since the model is oblivious, nothing about T_k, except κ_k, is known for the coordinator before the end of the algorithm. If $\kappa_k = 0$, clearly we have $t_k \geq 0$. If $\kappa_k > 0$, then T_k is non-empty, and we have to invoke the oracle corresponding to the k-th machine to get information about T_k. Thus, the oracle $\hat{\mathcal{O}}_k$ should be applied at least once in principle when $\kappa_k > 0$. In this way, it holds that $t_k \geq 1 \geq (\beta/4) \cdot \sqrt{\kappa_k N/M}$.

For $M < \beta^2 \kappa_k N/16$, we consider an input T' for k with constants α, β as in Theorem 11. Let $\alpha \in (4\epsilon, 1]$. Since $M < \beta^2 \kappa_k N/16$, we can put all of the elements to the k-th machine to construct an input T satisfying the hard input conditions in eq. (8). Thus, a hard input T with the same $\{\kappa_j\}$ must exist. By Lemma 13, it follows that $D_{t_k} \geq C M_k/M$ for $\mathcal{T}$ generated by T. Combining Lemma 14 with $t = t_k$, we have $4 m_k t_k^2/N \geq C M_k/M$. Recall that $M_k/m_k \geq \beta \kappa_k$ for hard input by Theorem 11, it holds that

$$t_k \geq \sqrt{\frac{C}{4} \frac{\beta \kappa_k N}{M}}. \tag{12}$$

Combining these two cases, we have $t_k \geq C' \sqrt{\kappa_k N/M}$ for some positive constant C'. Since k is arbitrary, this lower bound holds for t_j for every $j \in [n]$. By obliviousness of the queries, the value of t_j is invariant across every input with the same parameters N, M, κ_j, n, so we can directly add them together to conclude that the query complexity is

$$\sum_{j \in [n]} t_j \geq C' \sum_{j \in [n]} \sqrt{\frac{\kappa_j N}{M}} = \Omega\left(\sum_{j \in [n]} \sqrt{\frac{\kappa_j N}{M}} \right).$$

□

B.3.1 Proof of Lemma 13

To simplify the expression of the fidelity in the condition, we introduce another pure state $|\tilde{\psi}^T\rangle$.

Lemma 15 *Let $\mathcal{X}$ and $\mathcal{Y}$ be the output register and the working register of the coordinator, respectively. Let $\rho = \operatorname{Tr}_{\mathcal{Y}}[|\psi_{t_k}^T\rangle \langle \psi_{t_k}^T|]$ be the output state. If $\dim \mathcal{X} \leq \dim \mathcal{Y}$,*

then $F(\rho,\psi) = |\langle\psi_{t_k}^T|\tilde{\psi}^T\rangle|^2$ *for*

$$|\tilde{\psi}^T\rangle = \frac{1}{\sqrt{M}}\sum_{i\in[N]}\sqrt{c_i}\,|i\rangle\,|\xi_i^T\rangle \in \mathcal{X}\otimes\mathcal{Y} \tag{13}$$

with some $|\xi_i^T\rangle \in \mathcal{Y}$.

Proof. Since $|\psi_{t_k}^T\rangle$ is a purification of ρ, by Uhlmann's theorem, the fidelity $F(\rho,\psi) = \max_{|v\rangle\in\mathcal{X}\otimes\mathcal{Y}}|\langle\psi_{t_k}^T \mid v\rangle|^2$ with $|v\rangle$ satisfying $\mathrm{Tr}_{\mathcal{Y}}[|v\rangle\langle v|] = |\psi\rangle\langle\psi|$. Let $|\tilde{\psi}^T\rangle$ be the state $|v\rangle$ that makes the inner product attain the maximum. Suppose $|\tilde{\psi}^T\rangle = \sum_{i\in[N]}\kappa_i\,|i\rangle\,|\xi_i^T\rangle$. Since $\mathrm{Tr}_{\mathcal{Y}}[|\tilde{\psi}^T\rangle\langle\tilde{\psi}^T|] = |\psi\rangle\langle\psi|$, it must hold that $|\kappa_i| = \sqrt{c_i/M}$. Moving the phase of κ_i to the phase of $|\xi_i^T\rangle$, we can suppose that $\kappa_i = \sqrt{c_i/M}$, which leads to the lemma. □

Without loss of generality, we suppose that $\dim\mathcal{Y}$ is sufficiently large. Then by Lemma 15, the condition $F(\rho,\psi) \geq (1-\epsilon)^2$ of Lemma 13 implies $|\langle\psi_{t_k}^T \mid \tilde{\psi}^T\rangle| \geq 1-\epsilon$.

To obtain the lower bound for D_{t_k} given in Lemma 13, we divide D_{t_k} into two parts:

$$E_{t_k} = \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}}\left\||\psi_{t_k}^T\rangle - |\tilde{\psi}^T\rangle\right\|^2,$$
$$F_{t_k} = \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}}\left\||\psi_{t_k}\rangle - |\tilde{\psi}^T\rangle\right\|^2,$$

by a triangle inequality as follows:

$$\begin{aligned}
D_{t_k} &= \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}}\left\||\psi_{t_k}^T\rangle - |\tilde{\psi}^T\rangle + |\tilde{\psi}^T\rangle - |\psi_{t_k}\rangle\right\|^2 \\
&\geq E_{t_k} + F_{t_k} - \frac{2}{|\mathcal{T}|}\sum_{T\in\mathcal{T}}\left\||\psi_{t_k}^T\rangle - |\tilde{\psi}^T\rangle\right\|\left\||\tilde{\psi}^T\rangle - |\psi_{t_k}\rangle\right\| \\
&\geq E_{t_k} + F_{t_k} - 2\sqrt{E_{t_k}}\sqrt{F_{t_k}} \\
&= \left(\sqrt{F_{t_k}} - \sqrt{E_{t_k}}\right)^2.
\end{aligned} \tag{14}$$

Hence, we should look for a lower bound for $\sqrt{F_{t_k}} - \sqrt{E_{t_k}}$. In the following part, we upper bound E_{t_k} in Lemma 16 and lower bound F_{t_k} in Lemma 18.

Lemma 16 *Suppose* $M \leq \beta^2\kappa_k N/16$ *and* $\mathcal{T}$ *is a collection of hard inputs for* $k\in[n]$ *as in Theorem 11. If the final state* $|\psi_{t_k}^T\rangle$ *satisfies that* $|\langle\psi_{t_k}^T \mid \tilde{\psi}^T\rangle| \geq 1-\epsilon$ *for every* $T\in\mathcal{T}$, *with* $|\tilde{\psi}^T\rangle$ *given by eq.* (13), *then it holds that* $E_{t_k} \leq 2\epsilon$.

Proof. Since the change of the global phase does not affect the quantum state, without loss of generality, we can assume $\langle\psi_{t_k}^T \mid \tilde{\psi}^T\rangle = \left|\langle\psi_{t_k}^T \mid \tilde{\psi}^T\rangle\right| \geq 1-\epsilon$. Therefore,

$$\left\||\psi_{t_k}^T\rangle - |\tilde{\psi}^T\rangle\right\|^2 = 2 - 2\,\langle\psi_{t_k}^T \mid \tilde{\psi}^T\rangle \leq 2\epsilon.$$

It follows that $E_{t_k} \leq 2\epsilon$. □

To bound F_{t_k}, we need the following proposition:

Proposition 17 *Let* $\mathcal{T}$ *be a collection of hard inputs for* $k\in[n]$ *with constants* α,β, *then*

$$\sum_{T\in\mathcal{T}}\left|\langle\psi_{t_k} \mid \tilde{\psi}^T\rangle\right| \leq \sqrt{\frac{\sum_{j\neq k}M_j}{M}}|\mathcal{T}| + \sqrt{\frac{\kappa_k}{MN}}m_k|\mathcal{T}|,$$

with $|\psi_{t_k}\rangle$ *defined by Equation* (10) *and* $|\tilde{\psi}^T\rangle$ *given by Equation* (13).

Proof. Let $\mathcal{Z}$ be the space of the output register and working register. Consider an embedding map $\mathcal{Z}\hookrightarrow\mathcal{Z}\otimes\mathbb{C}^n$ defined as $|\varphi\rangle\mapsto|\varphi\rangle\,|0\rangle$, where $\mathbb{C}^n$ is the n-dimensional complex Hilbert space. Define a linear transform A on $\mathcal{Z}\otimes\mathbb{C}^n$ satisfying

$$\begin{aligned}
A\,|\tilde{\psi}^T,0\rangle &= A\left(\frac{1}{\sqrt{M}}\sum_{i\in[N]}\sqrt{c_i}\,|i,\xi_i^T,0\rangle\right) \\
&= \frac{1}{\sqrt{M}}\sum_{i\in[N]}\sum_{j\in[n]}\sqrt{c_{ij}}\,|i,\xi_i^T,j\rangle\,,
\end{aligned} \tag{15}$$

and

$$\begin{aligned}
&A\left(\frac{1}{\sqrt{M_k}}\sum_{i\in[N]}\sqrt{c_{ik}}\,|i,\xi_i^T\rangle\otimes\left(\sqrt{\frac{c_{ik}}{c_i}}\,|0\rangle + \sqrt{1-\frac{c_{ik}}{c_i}}\,|1\rangle\right)\right) \\
&= \frac{1}{\sqrt{M_k}}\sum_{i\in[N]}\sqrt{c_{ik}}\,|i,\xi_i^T,k\rangle\,.
\end{aligned} \tag{16}$$

Since $c_{ik}\leq c_i$, A is well-defined. By direct calculation, it can be verified that the definition of A preserves the inner product, so A can be supposed as a unitary [18]. By Equation (15), we have

$$\begin{aligned}
&\sum_{T\in\mathcal{T}}\left|\langle\psi_{t_k}|\tilde{\psi}^T\rangle\right| \\
&= \sum_{T\in\mathcal{T}}\left|\langle\psi_{t_k},0|\tilde{\psi}^T,0\rangle\right| \\
&= \sum_{T\in\mathcal{T}}\left|\langle\psi_{t_k},0|\,A^\dagger\frac{1}{\sqrt{M}}\sum_{i\in[N]}\sum_{j\in[n]}\sqrt{c_{ij}}\,|i,\xi_i^T,j\rangle\right| \\
&\leq \frac{1}{\sqrt{M}}\sum_{T\in\mathcal{T}}\left|\sum_{i\in[N]}\sum_{j\neq k}\sqrt{c_{ij}}\,\langle\psi_{t_k},0|\,A^\dagger\,|i,\xi_i^T,j\rangle\right| \\
&\quad + \frac{1}{\sqrt{M}}\sum_{T\in\mathcal{T}}\left|\sum_{i\in[N]}\sqrt{c_{ik}}\,\langle\psi_{t_k},0|\,A^\dagger\,|i,\xi_i^T,k\rangle\right|.
\end{aligned} \tag{17}$$

We then estimate the upper bound for these two terms in Equation (17) separately.

For the first term, note that $M_j = \sum_{i\in[N]}c_{ij}$, by Cauchy-Schwartz inequality,

$$\begin{aligned}
&\frac{1}{\sqrt{M}}\sum_{T\in\mathcal{T}}\left|\sum_{i\in[N]}\sum_{j\neq k}\sqrt{c_{ij}}\,\langle\psi_{t_k},0|\,A^\dagger\,|i,\xi_i^T,j\rangle\right| \\
&\leq \frac{1}{\sqrt{M}}\sum_{T\in\mathcal{T}}\sqrt{\sum_{j\neq k}M_j}\sqrt{\sum_{i\in[N]}\sum_{j\neq k}\left|\langle\psi_{t_k},0|\,A^\dagger\,|i,\xi_i^T,j\rangle\right|^2}
\end{aligned}$$

Since $A^\dagger$ is unitary, $\{A^\dagger |i, \xi_i^T, j\rangle\}_{i\in[N], j\in[n]}$ is an orthonormal system. Thus, the last square root is not greater than the norm of a unit vector $|\psi_{t_k}, 0\rangle$. So $\sqrt{\sum_{j\neq k} M_j/M} \cdot |\mathcal{T}|$ is an upper bound for the first term.

For the second term in Equation (17), by Equation (16) and the definition of the embedding map, we have

$$\begin{aligned}
&\frac{1}{\sqrt{M}} \sum_{T\in\mathcal{T}} \left| \sum_{i\in[N]} \sqrt{c_{ik}} \langle \psi_{t_k}, 0| A^\dagger |i, \xi_i^T, k\rangle \right| \\
=&\frac{1}{\sqrt{M}} \sum_{T\in\mathcal{T}} \left| \sum_{i\in[N]} \sqrt{\frac{c_{ik}}{c_i}} \sqrt{c_{ik}} \langle \psi_{t_k}|i, \xi_i^T\rangle \right| \\
\leq&\frac{1}{\sqrt{M}} \sum_{T\in\mathcal{T}} \sum_{i\in[N]} \sqrt{c_{ik}} \left| \langle \psi_{t_k}|i, \xi_i^T\rangle \right|.
\end{aligned}$$

Recall that $\kappa_k \geq c_{ik}$ and $c_{ik} = 0$ for $i \notin T_k$, so the last expression is not greater than $\sqrt{\frac{\kappa_k}{M}} \sum_{T\in\mathcal{T}} \sum_{i\in T_k} \left| \langle \psi_{t_k}|i, \xi_i^T\rangle \right|$. Changing the summation order, we come to

$$\begin{aligned}
&\sqrt{\frac{\kappa_k}{M}} \sum_{T\in\mathcal{T}} \sum_{i\in T_k} \left| \langle \psi_{t_k}|i, \xi_i^T\rangle \right| \\
=&\sqrt{\frac{\kappa_k}{M}} \sum_{i\in[N]} \sum_{T\in\mathcal{T}, i\in T_k} \left| \langle \psi_{t_k}|i, \xi_i^T\rangle \right| \\
\leq&\sqrt{\frac{\kappa_k}{M}} \sum_{i\in[N]} \sum_{T\in\mathcal{T}, i\in T_k} \left| \langle \psi_{t_k}|i, \iota_i\rangle \right|,
\end{aligned}$$

where "$\sum_{T\in\mathcal{T}, i\in T_k}$" means summation for all $T \in \mathcal{T}$ satisfying $i \in T_k$, and $|\iota_i\rangle$ independent to T satisfies $|\langle \psi_{t_k}|i, \iota_i\rangle| = \max_{\||\xi\rangle\|=1} |\langle \psi_{t_k}|i, \xi\rangle|$. As $|\psi_{t_k}\rangle$ is also independent of the choice of the input $T \in \mathcal{T}$, the second summation just brings a multiplier, which is the count of $T \in \mathcal{T}$ satisfying $i \in T_k$. Since the definition of $\mathcal{T}$ is to choose m_k elements in $[N]$ and assign them nonzero multiplicities, the count of the choices is $\binom{N-1}{m_k-1}$ as $i \in T_k$ means that i has to been chosen. Then by Cauchy-Schwartz inequality and Lemma 12,

$$\begin{aligned}
&\sqrt{\frac{\kappa_k}{M}} \sum_{i\in[N]} \sum_{T\in\mathcal{T}, i\in T_k} |\langle \psi_{t_k}|i, \iota_i\rangle| \\
=&\sqrt{\frac{\kappa_k}{M}} \binom{N-1}{m_k-1} \sum_{i\in[N]} |\langle \psi_{t_k}|i, \iota_i\rangle| \qquad (18) \\
\leq&\sqrt{\frac{\kappa_k}{M}} \frac{m_k}{N} \binom{N}{m_k} \sqrt{N} \sqrt{\sum_{i\in[N]} |\langle \psi_{t_k}|i, \iota_i\rangle|^2} \\
\leq&\sqrt{\frac{\kappa_k}{MN}} m_k |\mathcal{T}|.
\end{aligned}$$

Combining the above upper bounds for the two terms, we can obtain the proposition immediately. □

Lemma 18 *Let $\mathcal{T}$ be a collection generated by a hard input for $k \in [n]$ with constant β. If $M \leq \beta^2\kappa_k N/16$, then it holds that $F_{t_k} \geq M_k/2M$.*

Proof. Noting that $\left\| |\psi_{t_k}\rangle - |\tilde{\psi}^T\rangle \right\|^2 \geq 2 - 2\left|\langle \psi_{t_k}|\tilde{\psi}^T\rangle\right|$, we have

$$F_{t_k} \geq 2 - \frac{2}{|\mathcal{T}|} \sum_{T\in\mathcal{T}} \left| \langle \psi_{t_k}|\tilde{\psi}^T\rangle \right|.$$

A bound for the summation is given by Proposition 17. Through its conclusion, with $M = \sum_{j\in[n]} M_j$, we obtain a lower bound for F_{t_k}:

$$\begin{aligned}
F_{t_k} &\geq 2 - 2\sqrt{\frac{\sum_{j\neq k} M_j}{M}} - 2\sqrt{\frac{\kappa_k}{MN}} m_k \\
&= \frac{2}{1 + \sqrt{\frac{\sum_{j\neq k} M_j}{M}}} \frac{M_k}{M} - 2\sqrt{\frac{\kappa_k}{MN}} m_k \\
&\geq \frac{M_k}{M} - 2\sqrt{\frac{\kappa_k}{MN}} m_k.
\end{aligned}$$

Recalling $\frac{M_k}{m_k} \geq \beta\kappa_k$ in the hard input conditions, and the condition of $M \leq \frac{\beta^2}{16}\kappa_k N$, we can bound the second term with

$$\sqrt{\frac{\kappa_k}{MN}} \cdot m_k = \sqrt{\frac{M}{\kappa_k N}} \cdot \frac{\kappa_k m_k}{M} \leq \sqrt{\frac{\beta^2}{16}} \cdot \frac{M_k}{\beta M} = M_k/4M.$$

Thus, $F_t \geq M_k/2M$. □

To sum up, under the conditions of Lemma 13, by Lemma 16 and Lemma 18, it follows that $E_{t_k} \leq 2\epsilon, F_{t_k} \geq \frac{M_k}{2M}$. Combining them, we immediately obtain a lower bound for $\sqrt{F_{t_k}} - \sqrt{E_{t_k}}$:

$$\sqrt{F_t} - \sqrt{E_t} \geq \sqrt{\frac{M_k}{2M}} - \sqrt{2\epsilon}.$$

The definition of hard input implies $M_k \geq \alpha M$. Since we have chosen $\alpha > 4\epsilon$ as the condition of Lemma 13, it holds that $\epsilon \leq C_0 M_k/M$ for some constant $C_0 < 1/4$. With the inequality (14), the lower bound for D_t can be obtained by

$$D_t \geq \left(\sqrt{F_t} - \sqrt{E_t}\right)^2 \geq \left(\sqrt{\frac{M_k}{2M}} - \sqrt{2C_0\frac{M_k}{M}}\right)^2 = C\frac{M_k}{M}$$

with constant $C = (1/\sqrt{2} - \sqrt{2C_0})^2$. Lemma 13 follows from this.

B.3.2 Proof of Lemma 14

Since unitaries preserve norm, by the Cauchy-Schwartz inequality,

$$\begin{aligned}
D_{t+1} &= \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\| U_{t+1}\left(O_{t+1}\left|\psi_t^T\right\rangle - \left|\psi_t\right\rangle\right)\right\|^2 \\
&= \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\| O_{t+1}\left|\psi_t^T\right\rangle - \left|\psi_t\right\rangle\right\|^2 \\
&= \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\| O_{t+1}(\left|\psi_t^T\right\rangle - \left|\psi_t\right\rangle) + (O_{t+1}-I)\left|\psi_t\right\rangle\right\|^2 \\
&\le \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\| \left|\psi_t^T\right\rangle - \left|\psi_t\right\rangle\right\|^2 \\
&\quad + \frac{2}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\| \left|\psi_t^T\right\rangle - \left|\psi_t\right\rangle\right\| \left\|(O_{t+1}-I)\left|\psi_t\right\rangle\right\| \\
&\quad + \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\|(O_{t+1}-I)\left|\psi_t\right\rangle\right\|^2 \\
&\le D_t + 2\sqrt{D_t}\left[\frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\|(O_{t+1}-I)\left|\psi_t\right\rangle\right\|^2\right]^{1/2} \\
&\quad + \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\|(O_{t+1}-I)\left|\psi_t\right\rangle\right\|^2. \qquad (19)
\end{aligned}$$

It remains to prove an upper bound for $\frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\|(O_{t+1}-I)\left|\psi_t\right\rangle\right\|^2$.

Proposition 19 *For every collection $\mathcal{T}$ generated by a hard input for k, it holds that*

$$\frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \left\|(O_{t+1}-I)\left|\psi_t\right\rangle\right\|^2 \le 4\frac{m_k}{N}.$$

Proof. Since O_{t+1} is either $\hat{\mathcal{O}}_k \otimes I$ or $\hat{\mathcal{O}}_k^\dagger \otimes I$, by the definition of $\hat{\mathcal{O}}_k$, we have

$$\begin{aligned}
&(O_{t+1}-I)\left|\psi_t\right\rangle \\
&= \sum_{i\in T_k}\sum_{s=0}^{\nu}\sum_{l} \left|i,s,1,l\right\rangle\left(\left\langle i,s\oplus\pm c_{ik},1,l\right| - \left\langle i,s,1,l\right|\right)\left|\psi_t\right\rangle
\end{aligned}$$

with $x\oplus y := (x+y) \mod (\nu+1)$.

Notice that for two complex numbers a, b, it holds that $|a-b|^2 \le 2(|a|^2+|b|^2)$. Hence,

$$\begin{aligned}
&\sum_{T\in\mathcal{T}} \left\|(O_{t+1}-I)\left|\psi_t\right\rangle\right\|^2 \\
&= \sum_{T\in\mathcal{T}}\sum_{i\in T_k}\sum_{s=0}^{\nu}\sum_{l} \left|\left(\left\langle i,s\oplus\pm c_{ik},1,l\right| - \left\langle i,s,1,l\right|\right)\left|\psi_t\right\rangle\right|^2 \\
&\le 2\sum_{T\in\mathcal{T}}\sum_{i\in T_k}\sum_{s=0}^{\nu}\sum_{l} \left|\left\langle i,s\oplus\pm c_{ik},1,l|\psi_t\right\rangle\right|^2 \\
&\quad + 2\sum_{T\in\mathcal{T}}\sum_{i\in T_k}\sum_{s=0}^{\nu}\sum_{l} \left|\left\langle i,s,1,l|\psi_t\right\rangle\right|^2 \\
&= 4\sum_{T\in\mathcal{T}}\sum_{i\in T_k}\sum_{s=0}^{\nu}\sum_{l} \left|\left\langle i,s,1,l|\psi_t\right\rangle\right|^2.
\end{aligned}$$

Similarly to the deduction of Equation (18), through changing the summation order, since $\left|\psi_t\right\rangle$ is invariant across every $T\in\mathcal{T}$, and the count of $T\in\mathcal{T}$ satisfying $i\in T_k$ is $\binom{N-1}{m_k-1}$, it can be concluded that

$$\begin{aligned}
&4\sum_{T\in\mathcal{T}}\sum_{i\in T_k}\sum_{s=0}^{\nu}\sum_{l} \left|\left\langle i,s,1,l|\psi_t\right\rangle\right|^2 \\
&= 4\sum_{s=0}^{\nu}\sum_{l}\sum_{i\in[N]}\sum_{T\in\mathcal{T},i\in T_k} \left|\left\langle i,s,1,l|\psi_t\right\rangle\right|^2 \\
&= 4\binom{N-1}{m_k-1}\sum_{s=0}^{\nu}\sum_{l}\sum_{i\in[N]} \left|\left\langle i,s,1,l|\psi_t\right\rangle\right|^2 \\
&\le 4\binom{N-1}{m_k-1} = 4\frac{m_k}{N}\binom{N}{m_k}.
\end{aligned}$$

The proposition can be obtained immediately from this with Theorem 12. □

By Proposition 19 and the inequality (19), we have the relationship

$$D_{t+1} \le D_t + 4\sqrt{\frac{m_k}{N}D_t} + 4\frac{m_k}{N} = \left(\sqrt{D_t} + 2\sqrt{\frac{m_k}{N}}\right)^2.$$

This implies the upper bounds for $\sqrt{D_t}$ form an arithmetic progression. Since $D_0 = 0$, we can obtain that $D_t \le \left(2\sqrt{\frac{m_k}{N}}t\right)^2 = 4\frac{m_k}{N}t^2$.

B.4 Lower Bound on parallel queries

This subsection proves a lower bound on the query complexity for the parallel model. Similarly to the method for the sequential model, for each $k\in[n]$, we consider the number of oracle calls required for hard inputs for k, respectively. Suppose the lower bound obtained by considering the hard inputs for k is $\hat{t}_k$, then $\max_{j\in[n]}\hat{t}_j$ is a lower bound for the query complexity. Since the algorithm has no measurement, for any input T, the state after t oracles can be written as

$$\left|\psi_t^T\right\rangle = U_t O_t U_{t-1} O_{t-1}\cdots U_1 O_1 U_0\left|0\right\rangle,$$

where $O_1,\cdots,O_t$ are either $\mathcal{O}\otimes I$ or $\mathcal{O}^\dagger\otimes I$ with identity operator I on the registers that $\mathcal{O}$ doesn't act on, and $U_0,\cdots,U_t$ are unitary operators that are independent of the input. We also consider the input $\tilde{T}$ obtained from T by removing the dataset on the k-th machine, and the state with input $\tilde{T}$. Suppose the parallel oracle corresponding to $\tilde{T}$ is $\tilde{\mathcal{O}}$, then the state after t calls of this oracle is

$$\left|\psi_t\right\rangle = U_t\tilde{O}_t U_{t-1}\tilde{O}_{t-1}\cdots U_1\tilde{O}_1 U_0\left|0\right\rangle$$

where $\tilde{O}_1,\cdots,\tilde{O}_t$ are either $\tilde{\mathcal{O}}\otimes I$ or $\tilde{\mathcal{O}}^\dagger\otimes I$

With the above assumptions, we can conclude two lemmas for the parallel model similar to Lemma 13 and Lemma 14 for the sequential model.

Lemma 20 *Let $\mathcal{T}$ be the hard input for the k-th machine as in Theorem 11. Let α,β be the constants defined in*

Theorem 11. Suppose the fidelity between the output state ρ and ψ satisfies $F(\rho,\psi) \geq (1-\epsilon)^2 > 9/16$ with $\epsilon \geq 0$. If $M < \frac{\beta^2}{16}\kappa_k N$ and $\alpha > 4\epsilon$, then

$$\mathbb{E}_{\mathcal{T}}\left[\left\| |\psi_{\hat{t}_k}^T\rangle - |\psi_{\hat{t}_k}\rangle \right\|^2\right] \geq C\frac{M_k}{M}$$

for some constant C dependent to α and ϵ.

Proof. Noticing the proof of Lemma 13 is independent of the form of the oracle, it suffices to show it also holds for the parallel queries. The lemma is then obtained immediately. □

Lemma 21 *For every collection $\mathcal{T}$ generated by a hard input for k and $t \leq \hat{t}_k$, it holds that $\mathbb{E}_{\mathcal{T}}\left[\left\| |\psi_t^T\rangle - |\psi_t\rangle \right\|^2\right] \leq 4\frac{m_k}{N}t^2$ within the parallel model.*

Proof. The proof of Lemma 14 depends on the form of the oracle. But, specifically, the only part of it that depends on the oracle is the proof of Proposition 19. Hence, we only need to prove the conclusion of this proposition

$$\sum_{T\in\mathcal{T}} \left\| \left(O_{t+1} - \tilde{O}_{t+1}\right)|\psi_t\rangle \right\|^2 \leq 4\frac{m_k}{N}\binom{N}{m_k}$$

within the parallel model.

By the definition of $\mathcal{O}$ and $\tilde{\mathcal{O}}$, it follows that

$$\begin{aligned}
&(O_{t+1} - \tilde{O}_{t+1})|\psi_t\rangle \\
&= \sum_{\bar{i}:i_k\in T_k} \sum_{\bar{s}\in\{0,1,\cdots,\nu\}^n} \sum_{b\in\{0,1\}^n} \sum_{l} |\bar{i},\bar{s},b,l\rangle \cdot \\
&\qquad \left(\langle\bar{i},\bar{s},b,l| O_{t+1} - \langle\bar{i},\bar{s},b,l|\tilde{O}_{t+1}\right)|\psi_t\rangle .
\end{aligned}$$

It suffices to show

$$\begin{aligned}
&\sum_{T\in\mathcal{T}} \left\| \left(O_{t+1} - \tilde{O}_{t+1}\right)|\psi_t\rangle \right\|^2 \\
&\leq 4\sum_{T\in\mathcal{T}} \sum_{\bar{i}:i_k\in T_k} \sum_{\bar{s}\in\{0,1,\cdots,\nu\}^n} \sum_{b\in\{0,1\}^n} \sum_{l} |\langle\bar{i},\bar{s},b,l|\psi_t\rangle|^2.
\end{aligned}$$

Similarly to the proof of Proposition 19, changing the order of the summation, we can obtain the same expression as Proposition 19. This gives the lemma. □

Proof of Theorem 8. Similarly to the proof of Theorem 7, by the above two lemmas, we can conclude that $\hat{t}_k \geq C'\sqrt{\frac{\kappa_k N}{M}}$ for some constant C'. Hence, the query complexity is not less than $\max_{j\in[n]}\hat{t}_j = \Omega\left(\max_{j\in[n]}\sqrt{\frac{\kappa_j N}{M}}\right)$. □

C Proof of Lemma 9

By the assumption of the oblivious algorithm, the order of the oracles that $\mathcal{A}$ makes is predetermined and independent of the input, which, therefore, is also independent of the outcomes of the measurements. Thus, the measurement can be deferred to the end of the algorithm. Without loss of generality, it can be assumed that there is only one projective measurement [18].

Suppose this projective measurement is described by the projection operators $\{\Pi_i\}_{i=1}^N$. The algorithm now implements $\mathcal{V}$ and then follows a projective measurement. Suppose the state before this measurement is $|s_\sigma\rangle$. Then the output state is

$$\rho = \mathrm{Tr}_{\mathcal{Y}}\left[\sum_{i=1}^N \Pi_i |s_\sigma\rangle\langle s_\sigma|\Pi_i\right],$$

where $\mathcal{Y}$ is the non-output registers. The fidelity for $\mathcal{A}$ is

$$F(\rho,\psi) = \sum_{i=1}^N \sum_{\eta} |(\langle\psi|\otimes\langle\eta|)\Pi_i|s_\sigma\rangle|^2,$$

where $\{|\eta\rangle\}$ is a basis of $\mathcal{Y}$.

To give a new algorithm $\mathcal{B}$, we use an ancillary bit and choose a unitary transform U such that

$$U : |s,0\rangle \longmapsto \sum_{i=1}^N \sqrt{p_i}|s_i,i\rangle$$

with an arbitrary state $|s\rangle$, the coefficients $p_i = \langle s|\Pi_i|s\rangle$ and the normalized projections $|s_i\rangle = \Pi_i|s\rangle/\sqrt{p_i}$. It can be checked that this definition preserves the inner product, due to $\Pi_i\Pi_j = \delta_{ij}$. Hence, U is well-defined. Let the new algorithm be $\mathcal{B} = U(\mathcal{V}\otimes I)$ with initial state $|\zeta\rangle|0\rangle$, where $|\zeta\rangle$ is the initial state of the algorithm $\mathcal{A}$. Since the queries are not changed, the query complexity of $\mathcal{B}$ is the same as $\mathcal{A}$'s.

Finally, we show the fidelity for $\mathcal{B}$ is still the same. The final state of $\mathcal{B}$ is $\sum_{i=1}^N(\Pi_i|s_\sigma\rangle)\otimes|i\rangle$. Thus, if the output state is ρ', then the fidelity

$$\begin{aligned}
F(\rho',\psi) &= \sum_{\eta}\sum_{l=0}^N |\langle\psi,\eta,l|\sum_{i=1}^N(\Pi_i|s_\sigma\rangle)\otimes|i\rangle|^2 \\
&= \sum_{\eta}\sum_{l=1}^N |\langle\psi,\eta|\Pi_l|s_\sigma\rangle|^2 \\
&= F(\rho,\psi).
\end{aligned}$$

Fault-tolerant quantum computation by hybrid qubits with bosonic cat-code and single photons

Jaehak Lee[1 2] Nuri Kang[1 3] Seok-Hyung Lee[2 4] Hyunseok Jeong[2] Liang Jiang[5] Seung-Woo Lee[1 *]

[1] *Center for Quantum Information, Korea Institute of Science and Technology (KIST), Seoul 02792, Korea*
[2] *Department of Physics and Astronomy, Seoul National University, Seoul 08826, Republic of Korea*
[3] *Department of Physics, Korea University, Seoul 02841, South Korea*
[4] *Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, NSW 2006, Australia*
[5] *Pritzker School of Molecular Engineering, The University of Chicago, Chicago 60637 IL, USA*

Abstract. We introduce a fault-tolerant hybrid quantum computation by taking the advantages of both discrete variable (DV) and continuous variable (CV) systems. Particularly, we define a CV-DV hybrid qubit with bosonic cat-code and single photon and devise hybrid fusion schemes as building blocks for scalable architectures, which are implementable in current photonic platforms. The deterministic nature of the hybrid fusion enables a resource-efficient construction of cluster states, and further single photon loss can be corrected by implementing the hybrid fusion. We design fault-tolerant architectures by concatenating hybrid qubits and an outer DV quantum error correction code such as topological codes, exploring their potential merits in developing scalable quantum computation. We numerically simulate the fault-tolerance of our hybrid scheme, showing that it is at least an order of magnitude more resource-efficient over all previous proposals in photonic platforms. Moreover, it is demonstrated that our scheme allows to achieve at least 4-times higher loss thresholds compared to existing hybrid and CV approaches. We stress that our scheme is not limited to all-photonic platforms but can be implementable in other hybrid platforms including superconducting and trapped-ion systems, which allows us to find various efficient routes towards fault-tolerant quantum computing.

Keywords: Photonic hybrid quantum computation, Quantum error correction, Bosonic code

1 Hybrid quantum computation

Towards fully fault-tolerant quantum computing, various physical platforms such as photons, superconductors, trapped ions, nitrogen-vacancies in diamonds have been explored and considered as the building block for scalable quantum systems. Irrespective of the physical platforms, information is encoded into qubits defined with the basis either in DV or CV degrees of freedom, each of which has its own pros and cons. In recent years, hybrid approaches integrating different physical degrees of freedom to overcome the limitations of each platform have opened a new paradigm in quantum technologies. Hybridization may be quite a natural direction for scalability, since each platform has its own advantage depending on the circumstances and it is frequently required to convert quantum information between different platforms. Particularly, various CV-DV hybrid protocols have been recently proposed and experimentally demonstrated to combine their advantages in quantum computing and quantum communications.

Meanwhile, qubits encounter errors due to the interaction with environments and imperfect operations, which accumulate and become more severe as increasing the size of the system. Quantum error correction (QEC) provides systematic ways to protect qubits from dominant errors and allows to achieve fault-tolerance in building scalable quantum architectures. In QEC, information is typically encoded in a Hilbert space larger than a qubit space so that any error can be detected if it brings the encoded state out of the logical code space. By restoring the state back to the code space, errors can be corrected without compromising the encoding of logical information. While multiple physical qubits of finite-dimensional systems are typically used to construct single logical qubit in DV codes, a bosonic system characterized by an infinite-dimensional Hilbert space can provide a large number of degrees of freedom to encode a logical qubit in such a CV approach. Several bosonic error correction codes, in which a qubit is defined in a single oscillator, have been proposed such as GKP, binomial and cat code.

In this work, we introduce a hybrid quantum computing scheme by taking the advantages of both CV and DV systems toward fault-tolerant quantum computation [1]. In particular, we define a hybrid qubit by employing single photon and cat-code encoded state, which we call the hybrid cat-code (H-cat) qubit. The basis of H-cat qubit can then be defined as

$$\left\{|0_L\rangle = |+\rangle|\mathcal{C}_\alpha^+\rangle, |1_L\rangle = |-\rangle|\mathcal{C}_{i\alpha}^+\rangle\right\}, \tag{1}$$

where the first mode $|\pm\rangle$ represents the polarization of single photon state and the second mode represents even cat states $|\mathcal{C}_\alpha^+\rangle = \mathcal{N}_\mathcal{C}^+(|\alpha\rangle + |-\alpha\rangle)$ with the normalization factor $\mathcal{N}_\mathcal{C}^+$. Thanks to the cat-code encoded in the CV part, the effect of loss is readily correctable even without multi-qubit encoding [2, 3], while its logical basis is inherently orthogonal due to the DV part in contrast with other CV qubits. We also consider another type of hybrid qubits [4] for comparison, composed of single photon and

*swleego@gmail.com

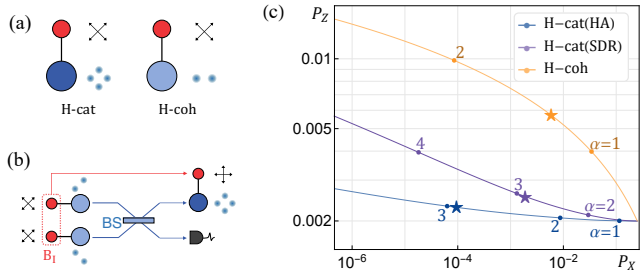


Figure 1: (a) Illustration of hybrid qubit. In our H-cat qubit, DV(red) and CV(blue) qubits are encoded, respectively, in the polarization degree of freedom and in the even cat states, while CV qubit is encoded in the coherent states(light blue) in the H-coh qubit. (b) Generation scheme of a H-cat pair. Red circles represent DV qubits and its polarization bases are represented as arrows. Blue circles represent cat-code qubits using four components of coherent states and light blue circles represent coherent-state qubits using two components of coherent states. (c) The plot represents the error rates P_X and P_Z for the fusion of hybrid qubits under loss rate $\eta = 2 \times 10^{-3}$. The orange curve is for H-coh scheme and the blue(purple) curve is for H-cat scheme employing HA (SDR) scheme. Numbers indicated at each point represent the corresponding amplitude α. Points marked as stars represent the optimal encoding amplitude α which achieves the highest loss threshold.

coherent state in the basis $\{|+\rangle|\alpha\rangle, |-\rangle|-\alpha\rangle\}$, which we call here the hybrid coherent-state (H-coh) qubit. The hybrid qubits, i.e., H-cat and H-coh, are illustrated in Fig. 1(a).

The hybrid qubits can be generated in current photonic platforms. CV-DV hybrid entangled states have been experimentally demonstrated in optical systems [5, 6] and successfully applied to quantum computing and communications in numerous experiments. In those experiments, the generated hybrid entangled states can be directly used as the H-coh qubit by simple modifications. We can also generate the H-cat qubit by using H-coh qubits by extending two-component cat states into four-component using $|\alpha\rangle$, $|i\alpha\rangle$, $|-\alpha\rangle$, and $|-i\alpha\rangle$ in CV part, using the scheme in Fig. 1(b). We stress that such CV-DV hybrid entanglement can be generated efficiently also in other platforms including superconducting and trapped-ion systems, which enables that our approach can be more generally implemented in any CV-DV hybrid platforms for quantum computing.

2 Hybrid fusion

We now introduce a *hybrid fusion*, i.e., a CV-DV hybrid entangling operation, which is performed by a joint work of Bell-state measurements on CV and DV qubits. A fusion measurement is typically applied on entangled states to generate larger size entangled states such as cluster states as prerequisites for measurement-based quantum computation (MBQC), or also enables universal gate operations via teleportation in circuit-based quantum computation. The hybrid fusion can be implemented by applying CV and DV Bell-state measurements separately, denoted here as $\mathrm{B}_{\mathcal{C}}$ and $\mathrm{B}_{\mathcal{D}}$, respectively. Its logical outcome can then be discriminated by combining the results of $\mathrm{B}_{\mathcal{C}}$ and $\mathrm{B}_{\mathcal{D}}$.

$\mathrm{B}_{\mathcal{C}}$ can be implemented by linear optics and photon-number-resolving (PNR) detectors. Two schemes were recently proposed independently in Ref. [7] and Ref. [8], which we respectively refer to HA and SDR scheme. Due to the nonorthogonality of CV basis, $\mathrm{B}_{\mathcal{C}}$ yields the X error rate p_X. In our hybrid scheme, we can remove the ambiguity by $\mathrm{B}_{\mathcal{D}}$, and thus the X error rate for hybrid fusion is reduced by half, that is, $P_X = p_X/2$. In the photon polarization encoding, $\mathrm{B}_{\mathcal{D}}$ can be chosen as a so called type II fusion that distinguishes two Bell states out of four with linear optics. Remarkably, a hybrid fusion is thus able to distinguish hybrid Bell states with certainty even if only one of $\mathrm{B}_{\mathcal{C}}$ and $\mathrm{B}_{\mathcal{D}}$ succeeds.

Moreover, in the presence of photon loss, $\mathrm{B}_{\mathcal{C}}$ can detect a single photon loss in CV part through detecting the parity change of the cat state to odd, i.e., when a total odd number of photons is detected. If two or more photons are lost, Z errors remain undetected in the fusion measurement outcome, yielding the error rate P_Z. In Fig. 1(c), we present the X and Z error rates under loss for the hybrid fusions. Specifically, P_X and P_Z are plotted for the fusions of H-cat qubits with HA and SDR, and H-coh qubit used in Ref. [4, 9] by varying the amplitude α under a fixed loss rate $\eta = 2 \times 10^{-3}$. It shows that the effect of loss can be substantially suppressed in the hybrid fusion of H-cat compared to H-coh thanks to the bosonic cat-code error correction in the CV part. As common tendencies, P_X error can be exponentially suppressed as α grows because the basis of CV part are more distinguishable, while P_Z only increases linearly with $|\alpha|^2$ because its state becomes more fragile against photon loss. The result clearly shows that P_Z is suppressed by employing the cat-code error correction. By taking larger encoding amplitude α, we can always achieve much smaller P_X and P_Z in the fusions of H-cat qubits than H-coh qubits.

3 Fault-tolerance analysis

Let us now design quantum computing architectures based on the hybrid qubits. We construct a Raussendorf-Harrington-Goyal (RHG) lattice embedding the surface code, by merging 3-qubit micro cluster states using hybrid fusions, as shown in Fig. 2(a). In the RHG lattice, errors that occur in the hybrid fusion propagate to adjacent qubits, so that a corresponding error rate is assigned to each individual qubit on the lattice. Based on assigned error rates, we can find the error pattern matching the syndrome measurement using the weighted minimum-weight perfect matching, and then count remaining error chains connecting two primal boundaries and determine whether a logical error occurs. We perform a Monte Carlo simulation to find the logical error rate p_L for different code distances d, and then investigate whether the errors are accumulated, i.e., p_L increases or

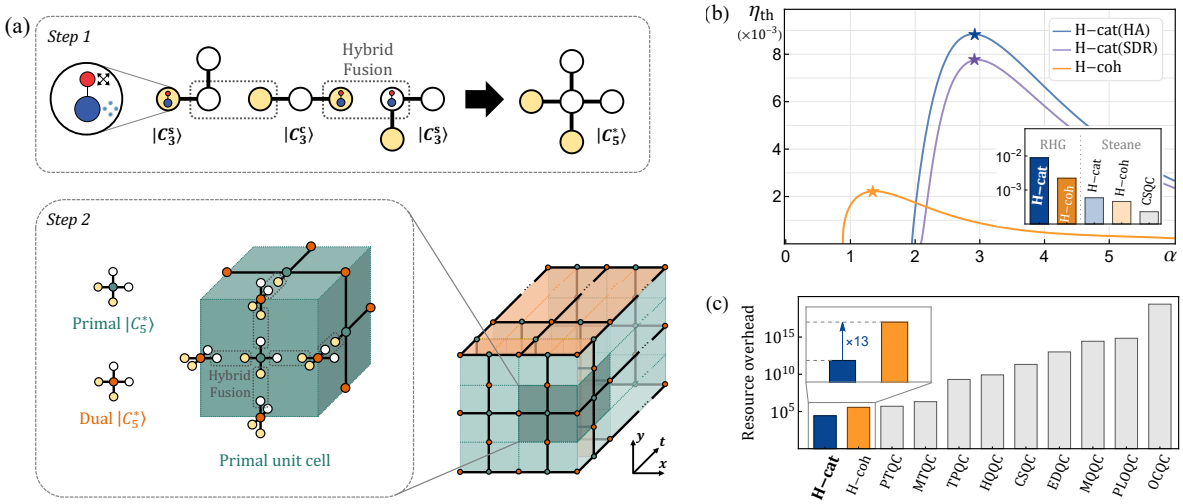


Figure 2: (a) A schematic to build a RHG lattice by hybrid qubits. 3-qubit micro H-cluster states are merged using hybrid fusions. (b) Loss thresholds obtained by simulation in RHG lattice with different schemes. Points marked as stars represent the highest loss threshold. The inset represents the comparison of optimal thresholds including previous approaches with Steane code. (c) Comparison of the resource overheads of existing photonic quantum computing proposals to achieve the logical error rate $p_L = 10^{-6}$.

not as increasing d. The fault-tolerance noise thresholds can then be determined as the maximum physical error rates by which the logical errors are not accumulated with d.

We present the loss thresholds of photonic MBQC based on H-cat and H-coh qubits in Fig. 2(b) by changing the encoding amplitude α in CV part. It shows that hybrid MBQC with H-cat qubits achieves the highest loss thresholds over other approaches including MBQC with H-coh qubits as well as all hybrid and CV approach in circuit-based model. The loss threshold of quantum computing with H-cat 0.89% is about 4-times larger than the maximum 0.22% estimated with H-coh [9].

To investigate the resource overhead, we estimate the number of unit resources referred to as $\mathcal{N}_{p_L}$ to achieves the target logical error rate p_L. In Fig. 2(c), we present the resource cost estimation of the proposed hybrid quantum computing schemes with H-cat and H-coh qubits, also comparing with existing MBQC photonic schemes. For hybrid schemes, we choose H-coh pair as a unit resource for fair comparison with the resource estimation in previous works [4, 9]. As a result, we can estimate that the resource overhead for the hybrid MBQC with H-cat qubits is $\mathcal{N}_{10^{-6}} = 2.7 \times 10^4$, which is 13 times more efficient compared to the overhead for the MBQC with H-coh qubits $\mathcal{N}_{10^{-6}} = 3.6 \times 10^5$. Remarkably, employing H-cat qubits can reduce an order of magnitude resource cost compared to the approach with H-coh qubits. It also shows that our hybrid approach is at least an order of magnitude more resource-efficient compared to all the other photonic proposals with respect to the cost of the resources, while the direct comparison is not straightforward due to the different types of resource states. The improved resource efficiency is achieved in our scheme thanks to the deterministic nature of the proposed hybrid fusion and the loss-tolerance of the resource states by inherently encoded bosonic cat-code.

References

[1] J. Lee, N. Kang, S.-H. Lee, H. Jeong, L. Jiang, S.-W. Lee, *submitted*, arXiv preprint arXiv:2401.00450.

[2] Z. Leghtas, G. Kirchmair, B. Vlastakis, R. J. Schoelkopf, M. H. Devoret, and M. Mirrahimi, Phys. Rev. Lett. **111**, 120501 (2013).

[3] M. Bergmann and P. van Loock, Phys. Rev. A **94**, 042332 (2016).

[4] S.-W. Lee and H. Jeong, Phys. Rev. A **87**, 022326 (2013).

[5] H. Jeong, A. Zavatta, M. Kang, S.-W. Lee, L. S. Costanzo, S. Grandi, T. C. Ralph, and M. Bellini, Nature Photonics **8**, 564 (2014).

[6] O. Morin, K. Huang, J. Liu, H. Le Jeannic, C. Fabre, and J. Laurat, Nature Photonics **8**, 570 (2014).

[7] J. Hastrup and U. L. Andersen, Phys. Rev. Research **4**, 043065 (2022).

[8] D. Su, I. Dhand, and T. C. Ralph, Phys. Rev. A **106**, 042614 (2022).

[9] S. Omkar, Y. S. Teo, and H. Jeong, Phys. Rev. Lett. **125**, 060501 (2020).

Entanglement witnesses and nonlocal maximum confidences in multipartite quantum state discrimination

Donghoon Ha[1] Jeong San Kim[1] *

[1] *Department of Applied Mathematics and Institute of Natural Sciences, Kyung Hee University, Yongin 17104, Republic of Korea*

Abstract. We consider multipartite quantum state discrimination and provide a specific relation between the properties of entanglement witness and quantum nonlocality inherent in the confidence of measurements. We first provide the definition of the confidence of measurements as well as its useful properties for various types of multipartite measurements. We show that globally maximum confidence that cannot be achieved by local operations and classical communication strongly depends on the existence of entanglement witness. We also provide conditions for an upper bound on maximum of locally-achievable confidences. Finally, we establish a method in terms of entanglement witness to construct quantum state ensemble with nonlocal maximum confidences.

Keywords: entanglement witness, maximum confidence, quantum nonlocality

Quantum nonlocality is an important feature of multipartite quantum systems without any classical counterpart [1–3]. In discriminating multipartite quantum states, nonlocal phenomenon occurs when a globally possible discrimination strategy cannot be realized only by *local operations and classical communication*(LOCC) [4]. The first nonlocality of quantum state discrimination was shown through orthogonal quantum states with local indistinguishability [5–7]. In general, orthogonal quantum states can be perfectly discriminated by using an appropriate measurement, whereas it is not true for nonorthogonal quantum states [8–11]. Nonlocality of quantum state discrimination can also occurs in discriminating nonorthogonal quantum states; there exist some nonorthogonal quantum states where the globally optimal discrimination cannot be realized using only LOCC measurements [12–15].

The phenomenon of nonlocality also arises in the correlation distributed in a multipartite quantum system. Quantum entanglement is a nonlocal correlation that cannot be created only by LOCC [1]. The nonlocal nature of entanglement can be used as a resource for quantum operations such as quantum teleportation and entangling measurements [16–18]. Thus, it is an important and even necessary task to detect the presence of entanglement inherent in multipartite quantum states. *Entanglement witness*(EW) is an entanglement-detecting observable providing a negative expectation value for some entangled states, whereas its expectation value is nonnegative for all separable states [19–22]. Recently, it was shown that quantum nonlocality arising in multipartite quantum state discrimination is closely related to the existence of EW [23, 24]. These results establish possible relationship between different types of nonlocality from various quantum phenomena.

Here, we consider multipartite quantum state discrimination and provide a specific relation between the properties of EW and quantum nonlocality inherent in the confidence of measurements. We first provide the definition of the confidence of measurements as well as its useful properties for various types of multipartite measurements. We show that globally maximum confidence that cannot be achieved by LOCC meausurements strongly depends on the existence of EW. We also provide conditions for an upper bound on maximum of locally-achievable confidences. Finally, we establish a method in terms of EW to construct quantum state ensemble with nonlocal maximum confidences.

For a multipartite Hilbert spaces $\mathcal{H} = \bigotimes_{k=1}^{m} \mathbb{C}^{d_k}$ with positive integers $m \geqslant 2$ and $d_1, \ldots, d_m$, let us denote by $\mathbb{H}$ the set of all Hermitian operators acting on $\mathcal{H}$. We also denote the set of all positive-semidefinite operators in $\mathbb{H}$ by

$$\mathbb{H}_+ = \{E \in \mathbb{H} \,|\, \langle v| E |v\rangle \geqslant 0 \;\; \forall |v\rangle \in \mathcal{H}\}. \tag{1}$$

A multipartite quantum state is described by $\rho \in \mathbb{H}_+$ with $\mathrm{Tr}\rho = 1$ and a measurement is expressed by $\{M_i\}_i \subseteq \mathbb{H}_+$ satisfying $\sum_i M_i = \mathbb{1}$ where $\mathbb{1}$ is the identity operator in $\mathbb{H}$. When ρ is measured in $\{M_i\}_i$, M_i is detected with the probability $\mathrm{Tr}(\rho M_i)$.

Definition 1 *$E \in \mathbb{H}_+$ is called* separable *if it can be described by a conic combination of product states, that is,*

$$E = \sum_l p_l \bigotimes_{k=1}^{m} \sigma_l^{(k)} \tag{2}$$

where $\{p_l\}_l$ is a set of nonnegative real numbers and $\{\sigma_l^{(k)}\}_l$ is a set of states acting on $\mathbb{C}^{d_k}$ for each $k = 1, \ldots, m$.

We denote the set of all separable operators in $\mathbb{H}_+$ by

$$\mathbb{SEP} = \{E \in \mathbb{H}_+ \,|\, E : \text{separable}\}. \tag{3}$$

A measurement $\{M_i\}_i$ is called a *separable measurement* if $M_i \in \mathbb{SEP}$ for all i. We also say that a measurement is a *LOCC measurement* if it can be realized by LOCC. Note that every LOCC measurement is a separable measurement [4].

*freddie1@khu.ac.kr

Definition 2 *$E \in \mathbb{H}$ is called* block positive *if*

$$\mathrm{Tr}(EF) \geqslant 0 \tag{4}$$

for all $F \in \mathbb{SEP}$.

We denote by $\mathbb{SEP}^*$ the set of all block-positive operators in $\mathbb{H}$, that is,

$$\mathbb{SEP}^* = \{E \in \mathbb{H} \,|\, E : \text{block positive}\}. \tag{5}$$

We note that

$$\mathbb{SEP} \subseteq \mathbb{H}_+ \subseteq \mathbb{SEP}^* \subseteq \mathbb{H}. \tag{6}$$

Definition 3 *$W \in \mathbb{H}$ is called an* EW *if $\mathrm{Tr}(WE) \geqslant 0$ for all $E \in \mathbb{SEP}$ but $\mathrm{Tr}(WF) < 0$ for some $F \in \mathbb{H}_+ \setminus \mathbb{SEP}$, or equivalently,*

$$W \in \mathbb{SEP}^* \setminus \mathbb{H}_+. \tag{7}$$

An EW W is called *optimal* if there is no other EW detecting more entangled states than W does; in other words, there does not exist $W' \in \mathbb{SEP}^* \setminus \mathbb{H}_+$ satisfying $\mathrm{Tr}(W'E) < 0$ for all $E \in \mathbb{H}_+ \setminus \mathbb{SEP}$ with $\mathrm{Tr}(WE) < 0$ and $\mathrm{Tr}(W'F) < 0$ for some $F \in \mathbb{H}_+ \setminus \mathbb{SEP}$ with $\mathrm{Tr}(WF) \geqslant 0$ [21]. An EW W is called *weakly optimal* if there exists a separable state σ satisfying

$$\mathrm{Tr}(\sigma W) = 0. \tag{8}$$

We note that weakly optimality is a necessary but not sufficient condition for an EW to be optimal [22, 25].

Let us consider the situation of discriminating the quantum states from the ensemble,

$$\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n, \tag{9}$$

where the state ρ_i is prepared with the *nonzero* probability η_i for each $i = 1, \ldots, n$. The *average quantum state* of $\mathcal{E}$ is denoted by ρ_0, that is,

$$\rho_0 = \sum_{i=1}^n \eta_i \rho_i. \tag{10}$$

We further consider the discrimination of the quantum state ensemble $\mathcal{E}$ in Eq. (9) using a measurement $\mathcal{M} = \{M_i\}_{i=0}^n$. For each $i = 1, \ldots, n$, we guess the prepared state to be ρ_i if the measurement result is M_i. On the other hand, the measurement result is inconclusive when we obtain M_0.

Definition 4 *For a quantum state ensemble $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ and a measurement $\mathcal{M} = \{M_i\}_{i=0}^n$, the* confidence *of ρ_j is the conditional probability $\Pr(\rho_j|M_j)$ that the prepared state is ρ_j when the measurement result is M_j, that is,*

$$\Pr(\rho_j|M_j) = \frac{\eta_j \mathrm{Tr}(\rho_j M_j)}{\mathrm{Tr}(\rho_0 M_j)}. \tag{11}$$

For each $j = 1, \ldots, n$, the confidence of ρ_j is well defined only when

$$\mathrm{Tr}(\rho_0 M_j) \neq 0. \tag{12}$$

For each $j = 1, \ldots, n$, the *maximum confidence* of ρ_j is

$$\mathcal{C}_j(\mathcal{E}) = \max_{\substack{\mathcal{M} \text{ with} \\ \mathrm{Tr}(\rho_0 M_j) \neq 0}} \Pr(\rho_j|M_j), \tag{13}$$

where the maximum is taken over all possible measurements $\mathcal{M} = \{M_i\}_{i=0}^n$ with Eq. (12). Each $\mathcal{C}_j(\mathcal{E})$ in Eq. (13) is already known as the largest eigenvalue of $\sqrt{\rho_0}^{-1} \eta_j \rho_j \sqrt{\rho_0}^{-1}$ where $\sqrt{E}$ is the positive square root of $E \in \mathbb{H}_+$, and F^{-1} is the inverse of the Hermitian operator F on its support [26, 27]. From this fact, we can easily verify that

$$\begin{aligned} \mathcal{C}_j(\mathcal{E}) &= \min\{q \in \mathbb{R} \,|\, q\mathbb{1} - \sqrt{\rho_0}^{-1} \eta_j \rho_j \sqrt{\rho_0}^{-1} \in \mathbb{H}_+\} \\ &= \min\{q \in \mathbb{R} \,|\, q\rho_0 - \eta_j \rho_j \in \mathbb{H}_+\}, \quad j = 1, \ldots, n. \end{aligned} \tag{14}$$

The *maximum-confidence discrimination* of $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ is to discriminate the states from the ensemble $\mathcal{E}$ using a measurement $\mathcal{M} = \{M_i\}_{i=0}^n$ achieving the maximum confidences in Eq. (13).

When the available measurements are restricted to separable measurements, we denote the maximum achievable confidences by

$$\mathcal{S}_j(\mathcal{E}) = \max_{\substack{\text{Separable } \mathcal{M} \\ \text{with } \mathrm{Tr}(\rho_0 M_j) \neq 0}} \Pr(\rho_j|M_j), \quad j = 1, \ldots, n. \tag{15}$$

Similarly, we denote the maximum achievable confidences by LOCC measurements as

$$\mathcal{L}_j(\mathcal{E}) = \max_{\substack{\text{LOCC } \mathcal{M} \\ \text{with } \mathrm{Tr}(\rho_0 M_j) \neq 0}} \Pr(\rho_j|M_j), \quad j = 1, \ldots, n. \tag{16}$$

For each $j = 1, \ldots, n$, it follows from the definitions of $\mathcal{C}_j(\mathcal{E})$, $\mathcal{S}_j(\mathcal{E})$ and $\mathcal{L}_j(\mathcal{E})$ that

$$0 < \mathcal{L}_j(\mathcal{E}) \leqslant \mathcal{S}_j(\mathcal{E}) \leqslant \mathcal{C}_j(\mathcal{E}) \leqslant 1. \tag{17}$$

The following theorem shows that every maximum achievable confidence by separable measurements is also achievable by LOCC measurements.

Theorem 5 *For a multipartite quantum state ensemble $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ and each $j = 1, \ldots, n$, we have*

$$\mathcal{L}_j(\mathcal{E}) = \mathcal{S}_j(\mathcal{E}) = \max_{\substack{M \in \mathbb{SEP} \\ \mathrm{Tr}(\rho_0 M) = 1}} \eta_j \mathrm{Tr}(\rho_j M), \tag{18}$$

where the maximum is taken over all possible separable operator M in $\mathbb{SEP}$ and ρ_0 is the average state of $\mathcal{E}$ in Eq. (10)*.*

Now, let us consider the minimum quantities

$$\mathcal{Q}_j(\mathcal{E}) = \min_{q \in \mathbb{R}_j(\mathcal{E})} q, \quad j = 1, \ldots, n, \tag{19}$$

where

$$\mathbb{R}_j(\mathcal{E}) = \{q \in \mathbb{R} \,|\, q\rho_0 - \eta_j \rho_j \in \mathbb{SEP}^*\}. \tag{20}$$

Each $\mathcal{Q}_j(\mathcal{E})$ in Eq. (19) is an upper bound of $\mathcal{S}_j(\mathcal{E})$ in Eq. (16).

For an ensemble $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ and each $j = 1, \ldots, n$, the following theorem shows that $\mathcal{S}_j(\mathcal{E})$ is equal to $\mathcal{Q}_j(\mathcal{E})$.

Theorem 6 *For a multipartite quantum state ensemble* $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ *and each* $j = 1, \ldots, n$, *we have*

$$\mathcal{S}_j(\mathcal{E}) = \mathcal{Q}_j(\mathcal{E}). \tag{21}$$

For a multipartite quantum state ensemble $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ and each $j = 1, \ldots, n$, the following theorem provides a necessary and sufficient condition for $q \in \mathbb{R}_j(\mathcal{E})$ to give $\mathcal{Q}_j(\mathcal{E})$.

Theorem 7 *For a multipartite quantum state ensemble* $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ *and* $q \in \mathbb{R}_j(\mathcal{E})$ *with* $j \in \{1, \ldots, n\}$,

$$q = \mathcal{Q}_j(\mathcal{E}) \tag{22}$$

if and only if there exists a separable state σ *satisfying*

$$\mathrm{Tr}[\sigma(q\rho_0 - \eta_j\rho_j)] = 0, \ \mathrm{Tr}(\sigma\rho_0) > 0. \tag{23}$$

For a multipartite quantum state ensemble $\mathcal{E}$ in Eq. (9) and each $j = 1, \ldots, n$, we say that the maximum confidence of ρ_j is *nonlocal* if it cannot be achieved by LOCC measurements, that is,

$$\mathcal{L}_j(\mathcal{E}) < \mathcal{C}_j(\mathcal{E}), \tag{24}$$

where $\mathcal{C}_j(\mathcal{E})$ and $\mathcal{L}_j(\mathcal{E})$ are defined in Eqs. (13) and (16), respectively. From Theorems 5 and 6, we note that Inequality (24) is equivalent to

$$\mathcal{Q}_j(\mathcal{E}) < \mathcal{C}_j(\mathcal{E}). \tag{25}$$

The following theorem provides a necessary and sufficient condition for Inequality (25) in terms of EW.

Theorem 8 *For a multipartite quantum state ensemble* $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$, $q \in \mathbb{R}$ *and each* $j = 1, \ldots, n$,

$$\mathcal{Q}_j(\mathcal{E}) \leqslant q < \mathcal{C}_j(\mathcal{E}) \tag{26}$$

if and only if $q\rho_0 - \eta_j\rho_j$ *is an EW.*

For a two-qubit state ensemble $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ and each $j = 1, \ldots, n$, the following corollary show that a real number q becomes $\mathcal{Q}_j(\mathcal{E})$ if $q\rho_0 - \eta_j\rho_j$ is a weakly-optimal EW.

Corollary 9 *For a two-qubit state ensemble* $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$, $q \in \mathbb{R}$ *and each* $j \in \{1, \ldots, n\}$,

$$\mathcal{Q}_j(\mathcal{E}) = q < \mathcal{C}_j(\mathcal{E}) \tag{27}$$

if and only if $q\rho_0 - \eta_j\rho_j$ *is a weakly-optimal EW. Moreover,* $\mathcal{C}_j(\mathcal{E})$ *is achievable locally when* ρ_0 *is not full rank.*

For a given set of EWs $\{W_i\}_{i=1}^n$ with $\epsilon > 0$ where ϵ is the smallest eigenvalue of

$$\mathcal{W} := \sum_{i=1}^n W_i, \tag{28}$$

Theorem 8 can used to construct quantum state ensembles where every maximum confidence is nonlocal. Let us consider the ensemble $\mathcal{E} = \{\eta_i, \rho_i\}_{i=1}^n$ with

$$\eta_i = \frac{\mathrm{Tr}(\lambda_i\mathcal{W} - \epsilon W_i)}{\mathrm{Tr}(\lambda\mathcal{W} - \epsilon\mathcal{W})}, \ \rho_i = \frac{\lambda_i\mathcal{W} - \epsilon W_i}{\mathrm{Tr}(\lambda_i\mathcal{W} - \epsilon W_i)} \tag{29}$$

where λ_i is the largest eigenvalue of W_i for each $i = 1, \ldots, n$ and λ is the sum of $\lambda_1, \ldots, \lambda_n$.

For $q \in \mathbb{R}$ and each $j = 1, \ldots, n$, a straightforward calculation leads us to

$$q\rho_0 - \eta_j\rho_j = \frac{[q(\lambda - \epsilon) - \lambda_j]\mathcal{W} + \epsilon W_j}{\mathrm{Tr}(\lambda\mathcal{W} + \epsilon\mathcal{W})}. \tag{30}$$

From Eq. (30) and Theorem 8 together with Eq. (14), we have

$$\mathcal{Q}_j(\mathcal{E}) \leqslant \frac{\lambda_j}{\lambda - \epsilon} < \mathcal{C}_j(\mathcal{E}) = \frac{\lambda_j + \epsilon\delta}{\lambda - \epsilon} \tag{31}$$

where δ is the absolute value of the smallest negative eigenvalue of $\sqrt{\mathcal{W}}^{-1} W_j \sqrt{\mathcal{W}}^{-1}$. Thus, the maximum confidence of ρ_j is nonlocal. Moreover, since ρ_0 is proportional to $\mathcal{W}$ in Eq. (28) and $\mathcal{W}$ is full rank, it follows from Theorem 7 that

$$\mathcal{Q}_j(\mathcal{E}) = \frac{\lambda_j}{\lambda - \epsilon} \tag{32}$$

when W_j is weakly optimal.

As our results provide a specific relation between EW and maximum achievable confidences by LOCC measurements, it is natural to investigate the relationip between EW and nonlocality arising in *optimal* maximum-confidence discrimination. It is also an interesting future work to investigate the relation between EW and optimal state discrimination in other state discrimination strategies.

References

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[2] A. M. Childs, D. Leung, L. Mančinska, and M. Ozols, A framework for bounding nonlocality of state discrimination, Commun. Math. Phys. **323**, 1121 (2013).

[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).

[4] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, Everything you always wanted to know about LOCC (but were afraid to ask), Commun. Math. Phys. **328**, 303 (2014).

[5] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Quantum nonlocality without entanglement, Phys. Rev. A **59**, 1070 (1999).

[6] S. Ghosh, G. Kar, A. Roy, A. Sen(De), and U. Sen, Distinguishability of Bell states, Phys. Rev. Lett. **87**, 277902 (2001).

[7] J. Walgate and L. Hardy, Nonlocality, asymmetry, and distinguishing bipartite states, Phys. Rev. Lett. **89**, 147901 (2002).

[8] A. Chefles, Quantum state discrimination, Contemp. Phys. **41**, 401 (2000).

[9] J. A. Bergou, Quantum state discrimination and selected applications, J. Phys.: Conf. Ser. **84**, 012001 (2007).

[10] S. M. Barnett and S. Croke, Quantum state discrimination, Adv. Opt. Photonics **1**, 238 (2009).

[11] J. Bae and L.-C. Kwek, Quantum state discrimination and its applications, J. Phys. A: Math. Theor. **48**, 083001 (2015).

[12] C. W. Helstrom, Quantum detection and estimation theory, J. Stat. Phys. **1**, 231 (1969).

[13] A. Peres and W. K. Wootters, Optimal detection of quantum information, Phys. Rev. Lett. **66**, 1119 (1991).

[14] R. Duan, Y. Feng, Z. Ji, and M. Ying, Distinguishing arbitrary multipartite basis unambiguously using local operations and classical communication, Phys. Rev. Lett. **98**, 230502 (2007).

[15] E. Chitambar and M.-H. Hsieh, Revisiting the optimal detection of quantum information, Phys. Rev. A **88**, 020302(R) (2013).

[16] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Phys. Rev. Lett. **70**, 1895 (1993).

[17] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein, Entangling Operations and Their Implementation Using a Small Amount of Entanglement, Phys. Rev. Lett. **86**, 544 (2001).

[18] S. Bandyopadhyay, G. Brassard, S. Kimmel, and W. K. Wootters, Entanglement cost of nonlocal measurements, Phys. Rev. A **80**, 012313 (2009).

[19] M. Horodecki, P. Horodecki, and R. Horodecki, Separability of mixed states: necessary and sufficient conditions, Phys. Lett. A **223**, 1 (1996).

[20] B. M. Terhal, Bell inequalities and the separability criterion, Phys. Lett. A **271**, 319 (2000).

[21] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki, Optimization of entanglement witnesses, Phys. Rev. A **62**, 052310 (2000).

[22] D. Chruściński and G. Sarbicki, Entanglement witnesses: construction, analysis and classification, J. Phys. A: Math. Theor. **47**, 483001 (2014).

[23] D. Ha and J. S. Kim, Entanglement witness and multipartite quantum state discrimination, J. Phys. A **56**, 205303 (2023).

[24] D. Ha and J. S. Kim, Bipartite quantum state discrimination and decomposable entanglement witness, Phys. Rev. A **107**, 052410 (2023).

[25] P. Badziąg, and P. Horodecki, R. Horodecki, and R. Augusiak, Separability in terms of a single entanglement witness, Phys. Rev. A **88**, 010301 (2013).

[26] S. Croke, E. Andersson, S. M. Barnett, C. R. Gilson, and J. Jeffers, Maximum Confidence Quantum Measurements, Phys. Rev. Lett. **96**, 070401 (2006).

[27] The support of a Hermitian operator is the subspace spanned by the all eigenvectors with nonzero eigenvalue.

Collaborative quantum sensing in an all-to-all connected sensor network

Wen-Han Png[1] * Haonan Liu[1] Travis Nicholson[1]

[1] *National University of Singapore, Centre for Quantum Technologies*

Abstract. Collaborating sensors achieve better precision than independent sensors. In this context, we proposed a novel quantum sensing protocol based on collaborating quantum sensors. We lay out generic model of all-to-all interacting collective spins sharing a common sensor bus. Then, we couple the collective spin to sensor bus such that the linear perturbation on the quantum bus can be readout through (i) collective spin operator (ii) interacting spin operator. We investigate the quantum Fisher info (QFI) for case (i) and (ii) in the entangled and separable sensor network. We find that spin-spin interaction enhances the optimum quantum Cramer Rao bound (QCRB) by $1/N$, while only requires entangling only half of the qubits. We present a numerical simulation with 1D trapped ion chain estimating two charges position and recover the consistent N scaling of QFI.

Keywords: Quantum Fisher Information, Super-Heisenberg, Trapped Ions, All-to-all Connectivity

1 Introduction

Sensing technology has been critical to the development of modern society, for it plays a strong role in healthcare, environmental monitoring, transportation, industrial automation, and security systems. Given the wide range and importance of sensing applications, innovations in sensing technology have the potential for transformative progress. For example, one possible improvement on global positioning could come in the form of collaborative sensing[11, 12], which improves position resolution via the sharing of location data between users.

Another direction that has great potential is quantum sensors. Quantum magnetometers, gravimeters, and inertial sensors have already found widespread technological applications. One exciting prospect of quantum sensors is distributed quantum sensing, whereby several sensors measuring the same parameter can be entangled, resulting in measurement resolution below the Standard Quantum Limit. In this case, the sensor precision can achieve Heisenberg scaling of $1/N$, where N is the number of sensors (as opposed to the $1/\sqrt{N}$ scaling of the Standard Quantum Limit). This effect has been proposed [8, 10, 9] and demonstrated [18] in atomic clocks.

This measurement enhancement has largely been studied for single-parameter sensing, whereby several sensors are measuring the same quantity. An interesting and less-studied case is that of multiparametric sensing, which involves many sensors measuring a distribution of values, such as thermometers attached to different places on a object with a thermal gradient. This has applications in vector field sensing [2, 6], magnetometry [13], and electrometry [21, 4, 5]. However, in the multiparametric case, it has been shown that entanglement alone does not allow for multiparametric sensing precision that scales better than $1/N$ [19].

In this Letter, we show that interacting systems can achieve super Heisenberg scaling, which surpasses the best available precision using entangled states alone. We also provide a general theory of quantum sensing with interacting states. Like collaborative GPS, we find that the interactions and entanglement form an information bus between sensor atoms, resulting in what we term *collaborative quantum sensing*. Furthermore, we provide a physical example of such a system in the form of a linear chain of trapped ions. Although these systems are often used as clocks or quantum information platforms, here we consider their use as quantum sensors of charge distributions. We find the enhanced precision scaling of our system with the ion number results in exceptional measurement resolution, which can be leveraged for a wide range of applications, from cellular biophysics to materials research to dark matter detection.

The resolution of both classical and quantum sensors can be described by the Fisher information. In sensors based on phase measurements of quantum states, the best achievable phase resolution $\Delta\theta$ for estimation of a single

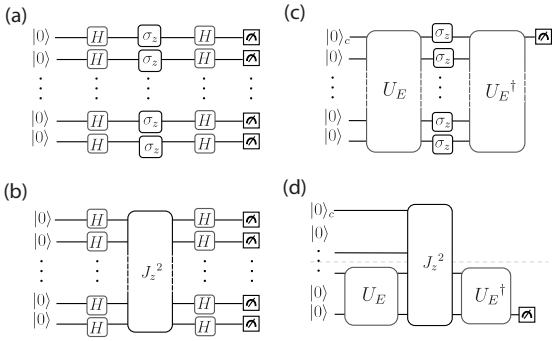


Figure 1: (a), (b), (c), (d) are the quantum circuit representation of the sensing protocol: Distributed Quantum Sensor in a Separable Network (DQSS), Collaborative Quantum Sensor in a Separable Network (CQSS), Distributed Quantum Sensor in an Entangled Network (DQSE) and Collaborative Quantum Sensor in an Entangled Network (CQSE) respectively. $|0\rangle_c$ denote the control qubit. U_E denotes the entangling operation.

*e0943469@u.nus.edu

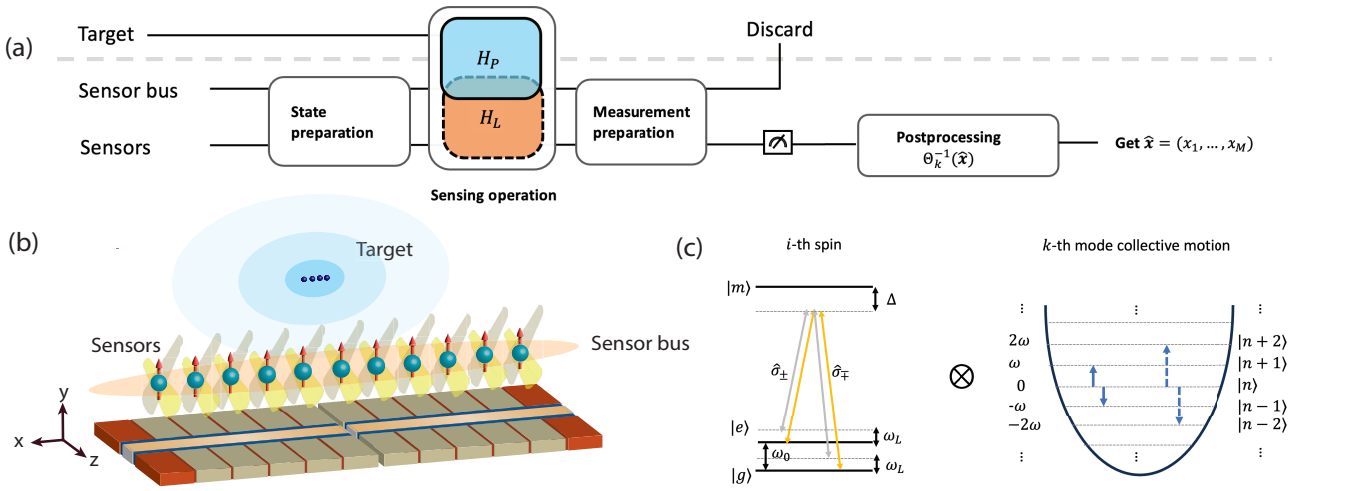


Figure 2: (a) The complete sensing protocol. The blue (red) box represents evolution with the Hamiltonian H_P (Hamiltonian H_L). After the sensing operation, we trace out the sensor bus and only measure the collective spin. In event of multi-parameter estimation, postprocessing is required to recover all estimated parameter. (b) The illustration trapped ion sensors estimating the positions of charged targets through a sensor bus (c) The Hilbert space of trapped ions comprises of the spin and motional degree of freedom. We coupled spins and collective motion through Raman beams. For DQS (CQS) protocol, we require $\omega_L \approx \omega$ ($\omega_L \approx 2\omega$) as indicated in the solid (dashed) line

parameter θ is given by the Cramer-Rao bound,

$$\Delta\theta = \frac{1}{\sqrt{F_\theta}}, \tag{1}$$

where F_θ is quantum Fisher information. For N atoms, $F_\theta = N$ when these particles are uncorrelated, whereas F_θ can be as large as N^2 for certain N-particle entangled states (e.g. squeezed states). In the case of a pth-order nonlinear interaction between particles (plus entanglement), the Fisher information can be $F_\theta = N^{2p}$.

2 Summary

We proposed the *Collaborative Quantum Sensing* (CQS) and *Distributed Quantum Sensing* (DQS) protocols, with the former encodes the perturbing signal to the collective spin operator, and the latter encodes it through the interacting-spin operator (see Fig 1). To ensure a fair comparison between the QFI of DQS and CQS, we used the same laser power and sensing duration for every instance of the number of qubits. The optimal sensing protocol is found to be CQSE, where its QCRB scales to superheisenberg limit for both collective phase estimation and multiple post-processed parameters. We demonstrated the physical implementation CQS and DQS sensing protocol, using 1D trapped ion chain to estimate the positions of two charges (see Fig. 2 for an overview). The numerical simulation successfully recovers the analytical QFI scalings of CQS and DQS, demonstrating a consistent N scaling for the collective phase estimation and multiparameter estimation on x_1, x_2. We note that the ion-phonon coupling strength is relatively weak (e.g. MHz), as the working regime is limited by $\alpha/\omega \approx 1$. Beyond this regime, trapped ion suffers from breakdown of Lamb-Dicke approximation, and leads to deviation of experimental outcome from the theoretical model. Thus, trapped ions sensor are constrained to probe perturbation with energy only at range of neV. Despite this constraint, neV perturbations remain relevant for applications in electric displacement sensing and dark matter detection [5, 1]. We note that the position of charges with small separation $\ll \mu$m is not resolvable due to this neV constraint. However, our sensor still probe a micro-scale structure at exceptional precision. Improving the precision of spin-spin coupling parameter also helps benchmarking the Ising coupling [16] . This is crucial for quantum simulation for exotic spin model such as Haldane phase, exotic frustrated magnetic states or even quantum spin liquids [14, 3]. We note that CQS and DQS protocols are based on readily available technology in trapped ion quantum computer. Entangling hundreds of ions [22, 7] and demonstrating 32 qubits GHZ state [17] has been reported. Trapped ion also feature a very mature $\sigma_z\sigma_z$ gate realization from 2 qubits [20] up to global entangling 5 qubits gate[15].

References

[1] M. Affolter, W. Ge, B. Bullock, S. C. Burd, K. A. Gilmore, J. F. Lilieholm, A. L. Carter, and J. J. Bollinger. Toward improved quantum simulations and sensing with trapped two-dimensional ion crystals via parametric amplification. *Phys. Rev. A*, 107:032425, Mar 2023.

[2] Tillmann Baumgratz and Animesh Datta. Quantum enhanced estimation of a multidimensional field. *Phys. Rev. Lett.*, 116:030801, Jan 2016.

[3] Stefan Birnkammer, Annabelle Bohrdt, Fabian Grusdt, and Michael Knap. Characterizing topological excitations of a long-range heisenberg model with trapped ions. *Physical Review B*, 105(24):L241103, 2022.

[4] M. Brownnutt, M. Kumph, P. Rabl, and R. Blatt. Ion-trap measurements of electric-field noise near surfaces. *Rev. Mod. Phys.*, 87:1419–1482, Dec 2015.

[5] Kevin A Gilmore, Matthew Affolter, Robert J Lewis-Swan, Diego Barberena, Elena Jordan, Ana Maria Rey, and John J Bollinger. Quantum-enhanced sensing of displacements and electric fields with two-dimensional trapped-ion crystals. *Science*, 373:673–678, 8 2021.

[6] Wojciech Górecki and Rafał Demkowicz-Dobrzański. Multiparameter quantum metrology in the heisenberg limit regime: Many-repetition scenario versus full optimization. *Phys. Rev. A*, 106:012424, Jul 2022.

[7] Or Katz, Marko Cetina, and Christopher Monroe. N-body interactions between trapped ion qubits via spin-dependent squeezing. *Physical Review Letters*, 129(6):063603, 2022.

[8] Eric M Kessler, Peter Komar, Michael Bishof, Liang Jiang, Anders S Sørensen, Jun Ye, and Mikhail D Lukin. Heisenberg-limited atom clocks based on entangled qubits. *Physical review letters*, 112(19):190403, 2014.

[9] Peter Komar, Eric M Kessler, Michael Bishof, Liang Jiang, Anders S Sørensen, Jun Ye, and Mikhail D Lukin. A quantum network of clocks. *Nature Physics*, 10:582–587, 2014.

[10] Péter Kómár, T Topcu, EM Kessler, Andrei Derevianko, V Vuletić, J Ye, and Mikhail D Lukin. Quantum network of atom clocks: a possible implementation with neutral atoms. *Physical review letters*, 117(6):060506, 2016.

[11] E.G. Larsson. Distributed positioning in ad hoc networks: a cramer-rao bound analysis. In *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No.03CH37484)*, volume 5, pages 2946–2950 Vol.5, 2003.

[12] E.G. Larsson. Cramer-rao bound analysis of distributed positioning in sensor networks. *IEEE Signal Processing Letters*, 11(3):334–337, 2004.

[13] David Le Sage, Koji Arai, David R Glenn, Stephen J DeVience, Linh M Pham, Lilah Rahn-Lee, Mikhail D Lukin, Amir Yacoby, Arash Komeili, and Ronald L Walsworth. Optical magnetic imaging of living cells. *Nature*, 496(7446):486–489, 2013.

[14] A Lemmer, C Cormick, D Tamascelli, T Schaetz, S F Huelga, and M B Plenio. A trapped-ion simulator for spin-boson models with structured environments. *New Journal of Physics*, 20(7):073002, jul 2018.

[15] Yao Lu, Shuaining Zhang, Kuan Zhang, Wentao Chen, Yangchao Shen, Jialiang Zhang, Jing Ning Zhang, and Kihwan Kim. Global entangling gates on arbitrary ion qubits. *Nature 2019 572:7769*, 572:363–367, 7 2019.

[16] Christopher Monroe, Wes C Campbell, L-M Duan, Z-X Gong, Alexey V Gorshkov, Paul W Hess, Rajibul Islam, Kihwan Kim, Norbert M Linke, Guido Pagano, et al. Programmable quantum simulations of spin systems with trapped ions. *Reviews of Modern Physics*, 93(2):025001, 2021.

[17] SA Moses, CH Baldwin, MS Allman, R Ancona, L Ascarrunz, C Barnes, J Bartolotta, B Bjork, P Blanchard, M Bohn, et al. A race track trapped-ion quantum processor. *arXiv preprint arXiv:2305.03828*, 2023.

[18] B.C. Nichol, R. Srinivas, D.P. Nadlinger, P. Drmota, D. Main, G. Araneda, C.J. Ballance, and D.M. Lucas. An elementary quantum network of entangled optical atomic clocks. *Nature*, 609:689, 2022.

[19] Timothy J. Proctor, Paul A. Knott, and Jacob A. Dunningham. Multiparameter estimation in networked quantum sensors. *Phys. Rev. Lett.*, 120:080501, Feb 2018.

[20] V. M. Schäfer, C. J. Ballance, K. Thirumalai, L. J. Stephenson, T. G. Ballance, A. M. Steane, and D. M. Lucas. Fast quantum logic gates with trapped-ion qubits. *Nature*, 555(7694):75–78, 2018.

[21] J. A. Sedlacek, A. Schwettmann, H. Kübler, and J. P. Shaffer. Atom-based vector microwave electrometry using rubidium rydberg atoms in a vapor cell. *Phys. Rev. Lett.*, 111:063001, Aug 2013.

[22] Yotam Shapira, Sapir Cohen, Nitzan Akerman, Ady Stern, and Roee Ozeri. Robust two-qubit gates for trapped ions using spin-dependent squeezing. *Physical Review Letters*, 130:030602, 1 2023.

Advancements in Quantum Computational Chemistry via Tensor Network-Based Algorithms for Large-Scale Execution

Shu Kanno[1 2 *]

[1] *Mitsubishi Chemical Corporation, Science & Innovation Center, Yokohama, 227-8502, Japan*
[2] *Quantum Computing Center, Keio University, Yokohama, 223-8522, Japan*

Abstract. Quantum computing holds promise for resolving complex chemical computation challenges unmanageable with classical computing methods. Despite their potential, current quantum systems are plagued by computational errors due to inherent noise. In this research, we introduce sophisticated algorithms utilizing tensor networks for the efficient preparation of quantum states and minimal quantum gate utilization during energy calculations. We have developed a novel algorithm that integrates tensor network-based system partitioning with high-fidelity quantum Monte Carlo simulations. This methodology has been validated across various chemical models, ranging from simple hydrogen plane systems to complex molecules involved in photochemistry, underscoring the precision of our approach. Further, we explore the advancements in tensor network algorithms for enhanced scalability. Our results mark a significant step towards achieving large-scale quantum chemical computations.

Keywords: Quantum chemistry, Tensor network, Quantum Monte Carlo

1 Introduction

Quantum computers have the potential to solve complex chemical computation challenges that cannot be addressed by classical computers. Current quantum computers, called noisy intermediate-scale quantum (NISQ) [1] devices, are limited in the number of qubits and quantum gates they can perform due to non-negligible physical noise. Our research focuses on the efficient preparation of quantum states using tensor networks. We proposed an algorithm combining a tensor network framework called the hybrid tensor network (HTN) [2] with a quantum version of quantum Monte Carlo (QMC) [3], where we named the algorithm HTN+QMC [4]. We have also explored advances in tensor network algorithms for large-scale computations.

2 Results

Figure 1 shows the results of energy calculations for the photochromic model molecule, MonoArylBiImidazole (MABI). Compared to the classical QMC calculation (blue), the combination of quantum calculation (light green and red) improves the energy accuracy. Furthermore, when comparing the results of quantum calculations, comparable accuracy was obtained in real device (red) and classical simulator (light green) runs. Our results demonstrate the importance of tensor network algorithms for realizing large-scale quantum chemical calculations.

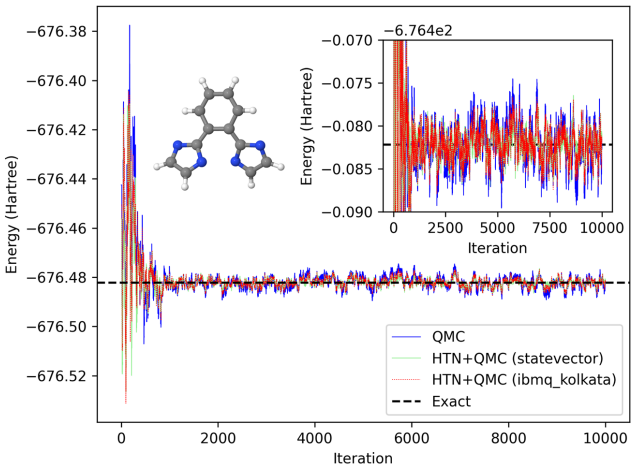


Figure 1: Results of the energy on the real device execution for MABI. The blue line represents the QMC result, whereas the light green and red lines show the HTN+QMC results of the statevector and real device procedures, respectively. The exact ground state energy is depicted by the black dashed line. The white, gray, and blue elements in the MABI structure represent the hydrogen, carbon, and nitrogen atoms, respectively. The inset in each figure presents an enlarged view along the y-axis.

References

[1] Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).

[2] Yuan, X., Sun, J., Liu, J., Zhao, Q. & Zhou, Y. Quantum simulation with hybrid tensor networks. *Phys. Rev. Lett.* **127**, 040501 (2021).

[3] Huggins, W. J. *et al.* Unbiasing fermionic quantum monte carlo with a quantum computer. *Nature* **603**, 416–420 (2022).

[4] Kanno, S. *et al.* Quantum computing quantum monte carlo with hybrid tensor network for electronic structure calculations (2023). 2303.18095.

*shu.kanno@quantum.keio.ac.jp

Simulation of Entangled States with One Bit of Communication

Peter Sidajaya[1] * Aloysius Dewen Lim[2] Baichu Yu[1 3 4] Valerio Scarani[1 2]

[1] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*
[2] *Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542*
[3] *Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Nanshan District, Shenzhen, 518055, China*
[4] *International Quantum Academy (SIQA), Shenzhen 518048, China*

Abstract. Bell's theorem shows that correlations made by entangled quantum systems cannot be replicated by Local Hidden Variables (LHV). Nevertheless, to quantify the power of a quantum correlation, it is useful to think of the additional resources that are needed to simulate it. In this work, we investigated the case of classical communication. For two qubits, the maximally entangled and some of the partially entangled states have been known to be simulatable with just one bit of communication. We used a neural network to try to close the problem for all two-qubit states completely and we presented evidence that the correlations of all partially entangled two-qubit states can be simulated with one bit of communication. On the other hand, as we go up in the dimension of the correlations, one bit of communication must necessarily fail to simulate the correlations. We give the smallest known example of quantum correlations that cannot be simulated with one bit of communication: it uses measurements on two qu5its, and thus is in the range of feasible experimental implementation.

Keywords: Bell nonlocality, entanglement, communication complexity, machine learning

1 Introduction

It is well established from Bell's theorem that Local Hidden Variables (LHVs) are inadequate to describe the behaviours of entangled quantum states [1]. Since then, some have asked how much supplementary resources, especially classical communication, do we need to simulate entangled states [2, 3, 4]. The problem was originally posed as a means to gain a more intuitive understanding of the power of entanglement. However, since Toner and Bacon's protocol to simulate a maximally entangled two-qubit state, progress in the field has been slower [4]. The latest progress came from Renner et al., who showed a protocol to simulate weakly entangled two-qubit states with just a single bit of communication [5]. On the other hand, there has been some works focusing on finding quantum behaviour that is unsimulatable with just 1-bit of communication [6, 7, 8]. The first such example came from [9], who used parallel games to find such an example. In our works [10, 11], we set out to make further progress on both fronts: to find a protocol for simulating all two-qubit states with 1-bit of communication and to find an example of an unsimulatable quantum correlation that lies in smaller scenario.

2 Simulating two-qubit states with LHV+1-bit

2.1 Neural network approach

For the first question, we used a neural network to generate numerical protocols that try to simulate their behaviours [10]. Our approach was inspired by the work of Krivachy et al. [12] where a neural network which was built with locality constraints in its architecture was used as an oracle to test whether a distribution is local. We modify the design of the network in order to use it to generate local strategies that can be done with 1-bit of supplementary communication.

The locality of the network is done by having separate networks represent the different parties and routing the different inputs according to what each party should receive. Communication is added to the model by first looking at 1-bit communication as a power for one party to choose between two options for both of them. In this way, we can actually model 1-bit of communication by having two local models (which are neural networks in themselves) and a third one that takes in Alice's inputs and outputs a number between 0 and 1 which denotes the probability of Alice choosing the first strategy, and thus sending the bit 0 to Bob. The final output is then, a convex combination of the two local models averaged over the LHVs. The architecture is illustrated in Fig. 1.

2.2 A semianalytical protocol

We first tested our model on the maximally entangled state and reobtained Toner and Bacon's model, with slight modifications. We then proceeded to train our neural network to simulate the behaviours of partially entangled states. We then wrote down the functions that approximate the behaviours of the neural network, which we will call our *semianalytical* protocols.

We benchmarked our neural network and the semianalytical protocol by comparing it with the original quantum behaviours and measuring its divergences. From the divergences, we could also consider the minimum number of sample size n needed such that we can be 95% confident that we would be able to distinguish the two behaviours in a hypothesis testing scenario. Through this method, we determined that the semianalytical protocol needs around 300 rounds of measurement to be differentiated, while the protocol of the neural network itself

*peter.sidajaya@u.nus.edu

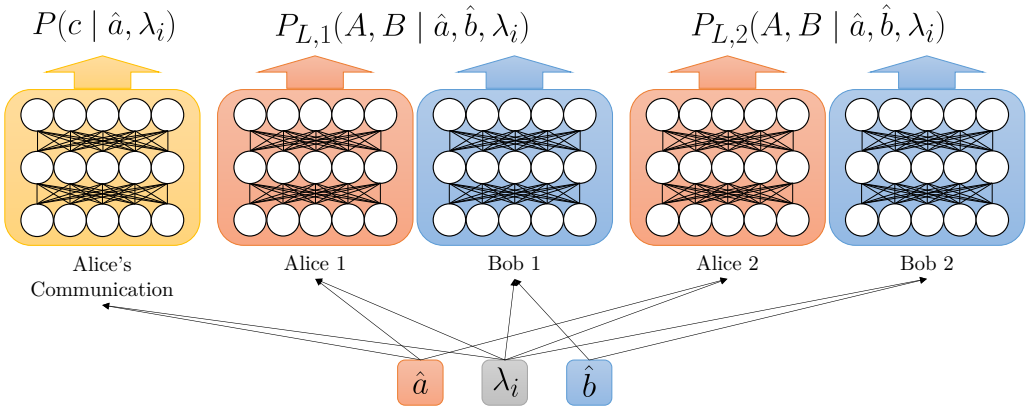


Figure 1: The architecture of the Artificial Neural Network (ANN). The model consists of two local distributions and a communication network and in each distributions the two parties are constrained by locality by routing the input accordingly. The communication network outputs a value between 0 and 1, and represents the probability of Alice sending a certain bit to Bob. The output for a particular round is then simply the convex combination of the two local distributions.

needs around upwards of 10000 rounds.

While the question of *exactly* simulating partially entangled states with 1-bit of communication remains unanswered, our works suggest that producing approximations of the quantum behaviours quite closely is possible. Our semianalytical protocols requires, on average, hundreds, and our neural networks requires tens of thousands of measurements, before it could be distinguished from the actual quantum behaviours. Taking into account that some of the two-qubit states can already been simulated by an exact protocol, these evidences suggest that all two-qubit states can be simulated with just a single bit of communication.

3 Unsimulatable correlations.

On the other front, we would also want to know what is the simplest quantum correlation that is unsimulatable by 1-bit of communication [11]. As previously mentioned, the simplest (and first) known example was given in [9]. For a fixed directional communication, the smallest given example was in the $(7, 3, 16, 16)$ scenario. This is still quite a surprisingly large scenario for just 1-bit.

The most straightforward method of finding such example is by constructing the 1-bit communication polytope and to try to find a violation of one of its facets. This method, however, only works for very small scenarios as the number of facets grow much faster for the 1-bit polytope compared to the usual local polytope. The largest known polytope is in $(3, 3, 2, 2)$, and no such violation was observed. Thus, we had to use a different approach if we would like to find a violation.

Consider these two observations: One, the maximum score of a linear inequality by a polytope is always achieved by one of its extremum points. Two, every extremum points always partition the game into two subgames where the behaviours in such subgames are local, where the partition is defined by the communication strategy of that point. Since the behaviours of these two subgames do not interact with each other, the maximum 1-bit score is always achieved by one of the extremum points whose behaviours in the local subgames are also maximum. Thus, instead of running through every extremum points, we could instead run through every partition and find the maximum scores of the subgames.

This small adjustment allowed us to calculate the 1-bit bound for scenarios in a higher dimension than previously possible and we found a Bell game which has a higher quantum score than the 1-bit bound in the $(5, 2, 5, 5)$ scenario. The inequality is given by

$$\mathcal{I}_d(\mathcal{P}) = \sum_{a,x,y=0}^{d-1} \sum_{b=0}^{1} [a - b = xy \bmod d]\, P(a, b|x, y), \quad (1)$$

where $[\cdot]$ is the Iverson bracket, which equals 1 if the statement inside is true and 0 otherwise. This inequality is a truncated XOR-5 game and has a local bound $S_{\mathcal{L}} = 6$, 1-bit bound $S_{\mathcal{L}_{+1}} = 7$, and $7.1777 \leq S_{\mathcal{Q}} \leq 7.1788$. Since $S_{\mathcal{Q}} > S_{\mathcal{L}_{+1}}$, the quantum correlation that achieves the quantum violations cannot be simulated by 1-bit of communication.

This example is smaller than the previously known example and is thus in range of feasible implementation. This method might also be used to calculate the 1-bit bidirectional communication score. However, we have yet to find a Bell game which has a higher quantum score than the 1-bit bidirectional communication that is smaller than the previously known example found in [9].

4 Conclusions

Classical communication was conceived as a way to quantify the power of a quantum correlation. However, it turns out that finding out the relationship between different quantum correlations and just a single bit of communication is a difficult task. In our works, we have furthered the work on exactly that problem, and to gain more intuition on the power of an entangled state, though still not completely. First, we have provided evidence for the simulability of all two-qubit states. On the other

hand, we also gave an example of a Bell game where quantum correlations beat 1-bit of communication, in a smaller scenario than previously known. Meanwhile, the problem of finding an exact analytical protocol for all two-qubit states and the problem of whether a smaller example of a violation of 1-bit bidirectional communication remains open.

References

[1] J. S. Bell, "On the einstein podolsky rosen paradox," *Physics Physique Fizika*, vol. 1, no. 3, p. 195, 1964.

[2] M. Steiner, "Towards quantifying non-local information transfer: finite-bit non-locality," *Physics Letters A*, vol. 270, no. 5, pp. 239–244, 2000.

[3] J. A. Csirik, "Cost of exactly simulating a bell pair using classical communication," *Phys. Rev. A*, vol. 66, p. 014302, Jul 2002.

[4] B. F. Toner and D. Bacon, "Communication cost of simulating bell correlations," *Phys. Rev. Lett.*, vol. 91, p. 187904, Oct 2003.

[5] M. J. Renner and M. T. Quintino, "The minimal communication cost for simulating entangled qubits," *arXiv preprint arXiv:2207.12457*, 2022.

[6] D. Bacon and B. F. Toner, "Bell inequalities with auxiliary communication," *Phys. Rev. Lett.*, vol. 90, p. 157904, Apr 2003.

[7] K. Maxwell and E. Chitambar, "Bell inequalities with communication assistance," *Phys. Rev. A*, vol. 89, p. 042108, Apr 2014.

[8] E. Zambrini Cruzeiro and N. Gisin, "Bell inequalities with one bit of communication," *Entropy*, vol. 21, no. 2, p. 171, 2019.

[9] I. Márton, E. Bene, P. Diviánszky, and T. Vértesi, "Beating one bit of communication with and without quantum pseudo-telepathy," *arXiv preprint arXiv:2308.10771*, 2023.

[10] P. Sidajaya, A. D. Lim, B. Yu, and V. Scarani, "Neural network approach to the simulation of entangled states with one bit of communication," *Quantum*, vol. 7, p. 1150, 2023.

[11] P. Sidajaya and V. Scarani, "Beating one bit of communication with quantum correlations in smaller dimension," *Phys. Rev. A*, 2024. (forthcoming).

[12] T. Kriváchy, Y. Cai, D. Cavalcanti, A. Tavakoli, N. Gisin, and N. Brunner, "A neural network oracle for quantum nonlocality problems in networks," *npj Quantum Information*, vol. 6, no. 1, pp. 1–7, 2020.

Encoded-fusion based quantum computation for high thresholds with linear optics

Wooyeong Song[1 2] Nuri Kang[1 3] Yong-Su Kim[1 4] Seung-Woo Lee[1 *]

[1] *Center for Quantum information, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea*

[2] *Quantum Network Research Center, Korea Institute of Science and Technology Information (KISTI), Daejeon 34141, Republic of Korea*

[3] *Department of Physics, Korea University, Seoul 02841, Republic of Korea*

[4] *Division of Nano & Information Technology, KIST school, Korea University of Science and Technology, Seoul 02792, Republic of Korea*

Abstract. We propose a fault-tolerant quantum computation scheme in a measurement-based manner with finite-sized entangled resource states and encoded-fusion scheme with linear optics. The encoded-fusion is an entangled measurement devised to enhance the fusion success probability in the presence of losses and errors based on a quantum error-correcting code. We apply an encoded-fusion scheme to construct a fault-tolerant network configuration in three-dimensional RHG lattice based on the surface code. Numerical simulations show that our scheme allows us to achieve up to ten times higher loss thresholds than non-encoded fusion approaches with limited numbers of photons used in fusion.

Keywords: Photonic quantum computation, Encoded-fusion, Fusion-based quantum computation

1 Introduction

Toward scalable and practical quantum computation, photonic systems have been considered as leading platforms thanks to high-quality sources and detectors, efficient modularity and connectivity, and long decoherence time at room temperature. In particular, extremely fast measurements on photons make them well-suited for measurement-based quantum computing (MBQC). In MBQC, universal gate operations can be realized by applying single-qubit measurements on entangled resource states. The resource state, typically a cluster state, is prepared offline so that the computation is performed easily via single-qubit measurements only. However, due to the non-deterministic fusion – a projective entangling measurement applied on entangled photons to create larger size resource states – and loss in photonic platforms, the preparation of a cluster state required to construct a fault-tolerant architecture consumes an extensive number of entangled photons.

In order to circumvent such formidable prerequisites for MBQC, fusion-based quantum computing (FBQC) was recently proposed [1]. FBQC is performed via fusion measurements applied between small constant-sized entangled resource states, so there is no need for extensive entanglement prepared and maintained stable during the process. The architecture of FBQC basically consists of resource states and fusions, which are connected to each other to create a specific network configuration called a fusion network. By constructing a fusion network, a quantum error-correcting code is implemented. The fusion thus plays a crucial role in FBQC and its efficiency directly affects the computation performance. However, the fusion success probability is limited by 50% with linear optics, and its boost with ancillary entangled photons turned out to be in a trade-off with the loss-tolerance of FBQC [2]. Therefore, non-deterministic fusions in the presence of photon loss can degrade the performance of FBQC, which becomes more crucial when increasing the system size, and, as a result, it may be still challenging to build a fault-tolerant quantum computing architecture in photonic platforms.

In this work, we propose a fault-tolerant quantum computing scheme performed in a measurement-based manner with finite-sized entangled states and encoded-fusions, called Encoded-fusion based quantum computing (EFBQC). Here, an encoded-fusion is devised to enhance the fusion success probability even under photon losses by a quantum error correcting code (QECC). We show that an encoded-fusion based on arbitrary (n, m)-generalized Shor code can be implemented with linear optics and

*swleego@gmail.com

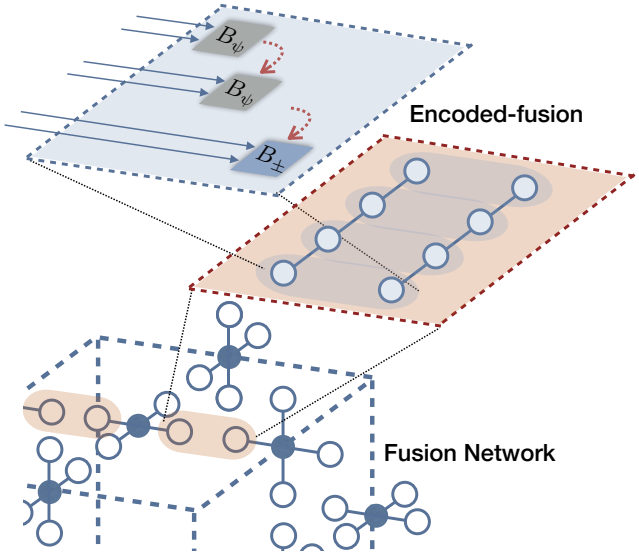


Figure 1: Schematics of encoded-fusion based quantum computing. In a fusion network, the photons participating in fusions are encoded in a QEC code. The encoded-fusion protocol can be performed by applying linear-optic Bell state measurements actively in a concatenate manner between encoded-qubits.

active feed-forwards. We then apply the encoded-fusion to construct a fusion network in a three-dimensional RHG lattice to implement the standard surface code. Numerical simulations show that our scheme allows us to achieve much higher loss threshold for individual photons than non-encoded fusion approach [3].

2 Results

Let us introduce the encoded-fusion based quantum computing (EFBQC). In EFBQC, the process to create a fusion network and logical gate operations are conceptually equivalent to FBQC in Ref. [1] except that the fusion is replaced with the encoded-fusion scheme, and the qubits that make up the resource states are encoded in the QEC code. The brief schematic of EFBQC is shown in Figure 1.

Here, we consider the (n, m)-generalized Shor code for resource state qubits. And the encoded-fusion protocol between the qubits consists of sequences of physical Bell state measurements, which are performed adaptively. We examine the 3-dimensional RHG lattice for EFBQC, and it can be composed of 4-star resource states or 6-ring resource states, as shown in Figure 2.

To demonstrate the performance of EFBQC, we compare the photon loss threshold of EFBQC in 3-

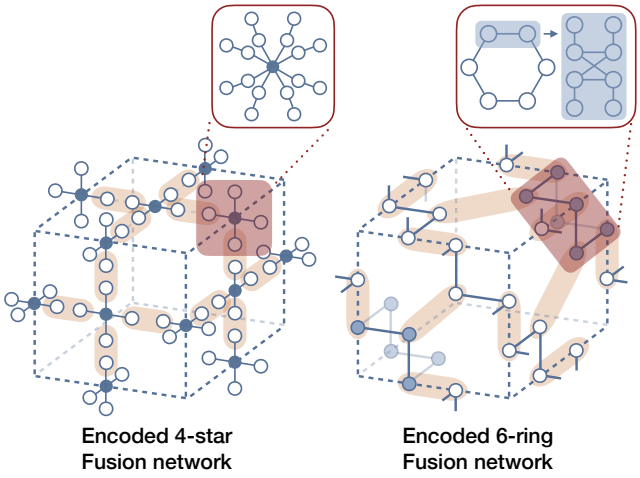


Figure 2: The fusion networks for 3-dimensional RHG lattice fabricated with the 4-star and 6-ring resource states. The insets illustrate the form of the encoded-4-star and -6-ring resource states when $(n, m) = (2, 2)$ as the simplest example.

dimensional RHG lattice with the results of FBQC. The results are shown in Figure 3 for various encoding numbers (n, m). The results for 6-ring (magenta) and 4-star (cyan) shows the FBQC results with non-encoded resource states and fusion boosting scheme proposed in Ref. [2]. And the results for $(2, 2)$-6-ring (green) and -4-star (purple) shows the FBQC results with $(2, 2)$-Shor encoded resource states without encoded fusion (with fusion boosting). It is clearly shown that EFBQC results in much higher loss thresholds, and notably, these loss thresholds can be improved as the number of photons used increases. This is not the case with FBQC using fusion boosting in Ref. [2], where there is a trade-off between fusion success probability and loss threshold. This trade-off is demonstrated by the decreasing tendency of the loss threshold as the number of photons used to boost the fusion success probability increases.

Note that both 4-star and 6-ring encoded-fusion networks can reach arbitrarily up to 14% by increasing the encoding number (n, m). Such a maximum threshold may be the characteristic of current choices of concatenated error correcting codes i.e., generalized Shor code and surface code, and thus possibly can be enhanced further with other codes [4, 5, 6].

3 Remarks

We have proposed a fault-tolerant quantum computation scheme, performed in a measurement-

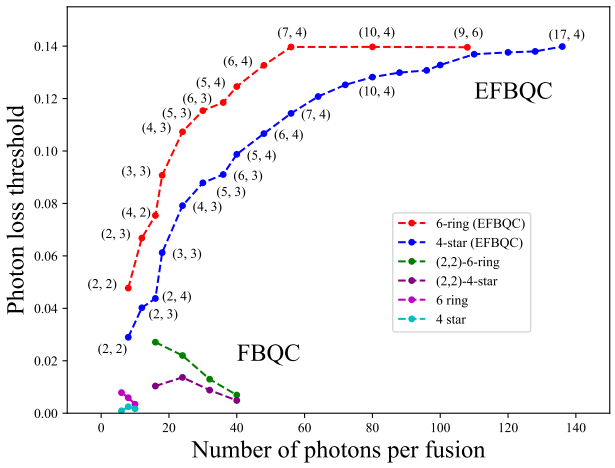


Figure 3: Photon loss thresholds for the total number of photons used per fusion. The thresholds of EFBQC are maximized by optimizing the encoded-fusion protocol for a given the encoding number (n, m). The threshold for EFBQC generally gets higher as increasing the number of photons used per fusion, while the threshold for FBQC boosted with ancillary Bell pairs decreases. EFBQCs for encoded-4-star and -6-ring resource states respectively yield 11.44% and 13.97% loss thresholds per photon when $(n, m) = (7, 4)$, and both arbitrarily reach up to 14% as increasing the encoding number (n, m).

based manner with finite-sized entangled resource states and fusion protected by quantum error correction. In contrast to previous FBQC schemes [4, 5, 6], our scheme uses a concatenation of two independent QEC protocols, one for the fusion itself and the other for the network configuration. The encoded-fusion is logically an entangling measurement, but is more aimed at correcting photon loss, fusion failure, and resource errors within the fusion process by implementing a QEC code. We have applied the encoded-fusion to construct a fusion network in RHG lattice. By numerical simulation, we have demonstrated that our scheme improves the loss thresholds up to 10 times higher than non-encoded fusion approach [1], and allows us to attain $\sim 14\%$ loss thresholds per individual photon, which is to our knowledge, a record-high threshold among recent achievements in photonic quantum computing [1, 4, 5, 6]. We have also shown that EFBQC outperforms FBQC with respect to the attainable threshold by consuming the same number of photons. Finally, we note that our approach is not limited to any specific configuration or code, but generally applicable for any structures implementing arbitrary codes and resource states, expecting further enhancements of the thresholds. Developing encoded-fusion protocols with other QECs [5, 6] would be also valuable as next step of research.

References

[1] S. Bartolucci, P. Birchall, H. Bombín, H. Cable, C. Dawson, M. Gimeno-Segovia, E. Johnston, K. Kieling, N. Nickerson, M. Pant, F. Pastowski, T. Rudolph and C. Sparrow. Fusion-based quantum computation. Nat. Commun., **14**, 912, 2023.

[2] W. P. Grice. Arbitrary complete Bell-state measurement using only linear optical elements. Phys. Rev. A, **84**, 042331, 2011.

[3] W. Song, N. Kang, Y. Kim and S.-W. Lee. Encoded-fusion based quantum computation for high thresholds with linear optics. *In preparation.*

[4] K. Sahay, J. Claes and S. Puri. Tailoring fusion-based error correction for high thresholds to biased fusion failure. Phys. Rev. Lett., **131**, 120604, 2023.

[5] S. Paesani and B. J. Brown. High-threshold quantum computing by fusing one-dimensional cluster states. Phys. Rev. Lett., **131**, 120603, 2023.

[6] T. J. Bell, L. A. Petterson and S. Paesani. Optimizing Graph Codes for Measurement-Based Loss Tolerance. PRX Quantum, **4**, 2, 2023.

Observing the quantum fault-tolerant threshold with entangled photons

Kai Sun[1] * Ze-Yan Hao[1] Jin-Shi Xu[1] † Chuan-Feng Li[1] ‡ Guang-Can Guo[1]

[1] *CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China*

Abstract. Fault-tolerant schemes, in which logical qubits are encoded by several physical qubits, enable to the output of a higher probability of correct logical qubits under the presence of errors. Here, based on an all-optical setup, we experimentally demonstrate the existence of the threshold for the fault-tolerant protocol in which four physical qubits encoded two logical qubits are realized as the spatial modes of two entangled photons. The developed high-accuracy optical system may provide a reliable platform to investigate error propagation in more complex circuits with fault-tolerant gates.

Keywords: fault-tolerance, threshold, entangled photons, spatial mode

we experimentally demonstrate the threshold of error rate for quantum circuits formed with fault-tolerant (FT) gates implemented in an all-optical setup. Based on the encoding method, we encode two logical qubits using four qubits which are mapped to the optical path information of two entangled photons. Besides the preparation stage, we experimentally implement a single-qubit Hadamard gate and a two-qubit CNOT gate in the logical space to form a complete quantum circuit in which error gates are imported based on the bit-flip error. When comparing the output probabilities of the encoded circuit and those of non-encoded circuit, we could determine the FT threshold of the error rate. Our results clearly demonstrate that when the error rate remains below the threshold, the probability to obtain correct output results in the FT circuit is higher than that of the corresponding non-encoded circuit. On the other side, if the error rate is above the threshold, no benefit is obtained from the FT implementation.

According to the FT protocol in [1], two logical qubits are encoded with four physical qubits as follows:

$$\begin{aligned} |00\rangle_l &= (|0000\rangle + |1111\rangle)/\sqrt{2}, \\ |01\rangle_l &= (|0011\rangle + |1100\rangle)/\sqrt{2}, \\ |10\rangle_l &= (|0101\rangle + |1010\rangle)/\sqrt{2}, \\ |11\rangle_l &= (|0110\rangle + |1001\rangle)/\sqrt{2}, \end{aligned} \tag{1}$$

where $\{|00\rangle_l, |01\rangle_l, |10\rangle_l, |11\rangle_l\}$ represent the logical bases, and $\{|0000\rangle, |0011\rangle, |0101\rangle, |0110\rangle, |1001\rangle, |1010\rangle, |1100\rangle, |1111\rangle\}$ represent the bases of four physical qubits (the encoded space only involves even number of $|1\rangle$ in physical qubits). The four physical qubits are mapping to coincident modes of two entangled photons. As shown in Fig. 1**a**, with optical spatial modes on each side marked as $|00\rangle, |01\rangle, |10\rangle$, and $|11\rangle$, the basis of four physical qubits is denoted as the coincidence count between two spatial modes from the side A and B, respectively. As an illustration, the coincident mode $|mnij\rangle \equiv |mn\rangle_A \otimes |ij\rangle_B$ ($m, n \in \{0,1\}_A$ and $i, j \in \{0,1\}_B$), i.e., the intensity and phase of basis $|mnij\rangle$ are related to the coincidence count between modes $|mn\rangle_A$ and $|ij\rangle_B$. This method of mapping the qubits to optical spatial modes could simulate the operation on individual qubit with the evolution of spatial modes [2].

By coherently adjusting spatial modes, single- and two-qubit gates can be conveniently realized. Logical state $|00\rangle_l = (|0000\rangle + |1111\rangle)/\sqrt{2}$ can be FT prepared with post-selection following the circuit starting from initial physical state $|0000\rangle$. In this protocol, a set of quantum gates, such as σ^x, Hadamard and CNOT gates, operated on logical qubits can be implemented in a FT manner. As a result, a circuit, only formed by these FT gates, is implemented FT and there exists a threshold of the error rate. Our main task is to experimentally demonstrate the existence of the threshold in the FT circuit.

In experiment, as shown in Fig. 1**b**, a continuous-wave diode laser with the wavelength 404 nm and a bandwidth 0.048 nm is used to pump a 20 mm-long periodically poled KTP (PPKTP) crystal with the help of a polarized Sagnac interferometer to generate polarization-entangled photons $|\Phi\rangle = (|H_A H_B\rangle + |V_A V_B\rangle)/\sqrt{2}$. Based on this entangled source, Fig. 1**c** shows $|00\rangle_l$ could be achieved with several BDs and HWPs which are adjusted along the preparation circuit. More details could be found in Ref. [3].

With the high performance of Hadamard and CNOT gates on physical qubits in experiment, experimental results show that operations in this platform are extremely accurate, allowing to observe the threshold effect in FT protocol. For the circuit implementing logical operation H_2, the threshold is $p = 0.978$ in theory. This threshold is consistent with Fig. 2**a**, in which the experimental probability of correct output, F_p, is larger than the prediction f_p of non-encoded circuit detected in the same experimental platform (see more details in SM [3]) for $p > 0.978$. On the other hand, when $p < 0.978$, we obtain $F_p < f_p$. Experimental results of logical operation $CNOT_{21} \cdot H_2$ are shown in Fig. 2**b**, in which the predicted threshold is $p = 0.968$. The experimentally obtained F_p is higher (lower) than corresponding f_p for p above (below) the threshold.

Using a concise FT protocol, we experimentally demonstrate the threshold of a complete FT circuit with a Hadamard gate and a CNOT gate on logical qubits, be-

*ksun678@ustc.edu.cn
†jsxu@ustc.edu.cn
‡cfli@ustc.edu.cn

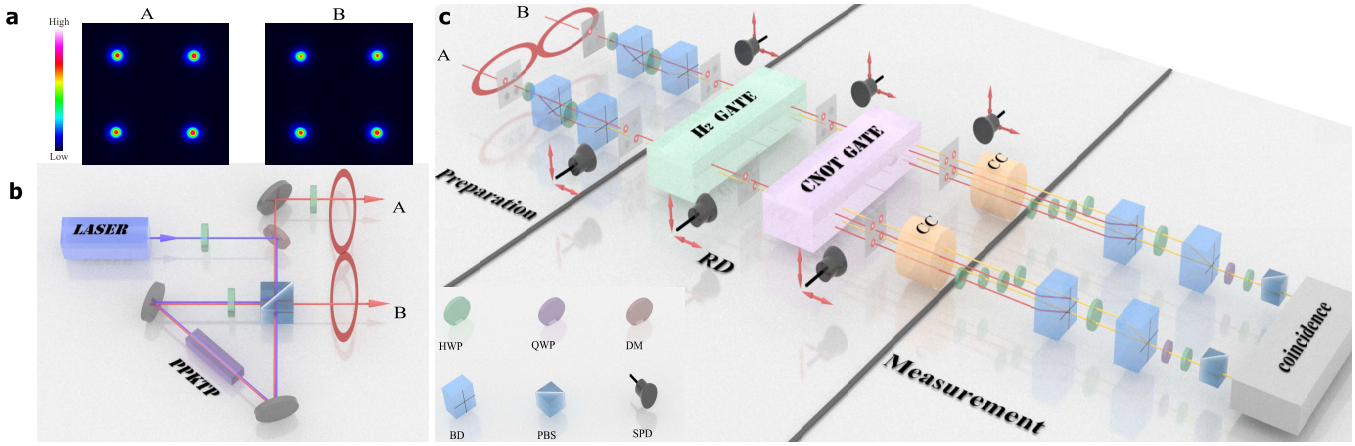


Figure 1: Experimental setup for verification of fault-tolerant threshold in quantum circuits. **a** Experimental images of optical spatial modes on sides of A and B generated by exploiting a group of several beam displacers (BDs) and half-wave plates (HWPs). **b** The unit to prepare entangled photon pairs. **c** Spatial mode evolutions of fault-tolerant circuits including the stages of preparation, logical operations (H_2 and $CNOT_{21}$), and measurement. Output spatial modes on each side of every stage are detected with removable detectors (RDs), which are built with single-photon detectors (SPD) placed on two-dimensional movable platforms, for coincidence counts to estimate the imported error rate. Final spatial modes on each side are combined together, where optical path differences among the modes on each side are offset by compensation crystals (CC), and then measured with a quarter-wave plate (QWP), a HWP, and a polarization beam splitter (PBS). The coincidence device deals with the detected signals from two sides and outputs the coincidence count.

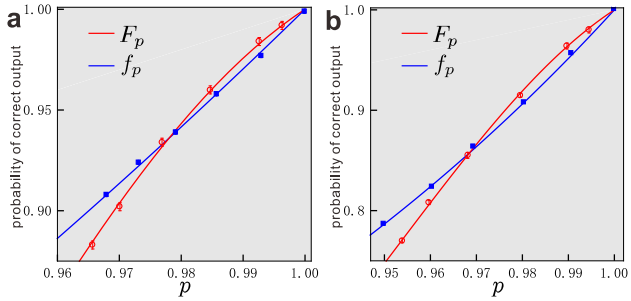


Figure 2: Experimental probabilities of correct output for different operations. Panels **a** and **b** show experimental probabilities of correct output, F_p, according to success probability $p = 1-\epsilon$ for H_2 and $CNOT_{21}\cdot H_2$, respectively. Blue and red curves represent theoretical predictions of the non-encoded circuit (f_p) and fault-tolerant circuit (F_p), respectively. Blue squares (errorbars are too small to show) and red hollow points with error bars indicate the corresponding experimental results. All error bars are estimated as standard deviations of photon counts assuming a Poisson distribution.

sides preparing and measuring processes, with the bit-flip error in each operator. Generally, to verify a FT protocol, the successful output probability of any circuit, formed with FT gates, should be higher than that of the corresponding uncoded circuit when the error rate is below a threshold. Note that rich encoding paths in the experimental setup enable different circuits for physical qubits to realize the same operations in logical space. However, only some configurations are FT. Besides the circuit realized in this work, to implement another different circuit, we just need to rotate BDs and HWPs. And for some complicated cases, we need simply to add more sets of BDs and HWPs.

To completely demonstrate a universal FT quantum computation remains a long-standing challenging. High-accuracy operations that can be achieved using optical systems establish an appropriate platform to simulate the error propagation in FT circuits, especially to investigate the behavior of coherent errors. Also, based on this experimental platform, nonlocal errors affecting the entangled property could be further investigated under this encoding framework. Moreover, despite of limitation of the scale of optical system, this work facilitates the potential investigation of FT protocols with breakthroughs of large-scale experimental implementations of quantum technology in other physical platforms.

References

[1] D. Gottesman. Quantum fault tolerance in small experiments. *arXiv*: 1610.03507 (2016).

[2] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn. Linear optical quantum computing with photonic qubits. *Reviews of Modern Physics* **79**, 135–174 (2007).

[3] K. Sun *et al.* Optical demonstration of quantum fault-tolerant threshold. *Light: Science & Applications* **11**, 203 (2022).

Symmetric Clifford twirling for cost-optimal quantum error mitigation in early FTQC regime

Kento Tsubouchi[1] * Yosuke Mitsuhashi[2] Kunal Sharma[3] Nobuyuki Yoshioka[1] †

[1] *Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[2] *Department of Basic Science, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan*
[3] *IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA*

Abstract. We propose symmetric Clifford twirling, which uses symmetric Clifford operators commuting with certain Pauli subgroups to twirl noise affecting quantum gates. We characterize the conversion of each Pauli noise through this method, showing that certain Pauli noise can be scrambled to noise exponentially close to global white noise. We further demonstrate that highly structured circuits, such as Trotterized Hamiltonian simulations, can have their effective noise scrambled to global white noise. This method allows us to mitigate errors in non-Clifford operations in early FTQC regimes with minimal sampling overhead and provides new insights into fields where randomness and symmetry are crucial.

Keywords: early FTQC, quantum error mitigation, quantum error correction, twirling

1 Introduction

Fault-tolerant quantum computing (FTQC) using quantum error correction has been a focal point of research in recent decades as a robust countermeasure for errors affecting quantum computers [1–8]. In many quantum error-correcting codes, in particular CSS codes, non-Clifford operations cannot be implemented fault-tolerantly and hence require sophisticated techniques such as magic state distillation and gate teleportation. This results in significant overhead in realizing FTQC. [9–15]. Therefore, in the early stages of FTQC, known as the early FTQC regime, non-Clifford operations are anticipated to be susceptible to a non-negligible amount of logical errors.

Recent research has revealed that such logical errors can be efficiently addressed by converting the errors to global white noise [16–18]. This is because it allows for cost-optimal quantum error mitigation (QEM) [19–21] with minimal sampling overhead [16] or the implementation of certain quantum algorithms robust to global white noise [17, 18]. While such a conversion can relax the noise requirements in noisy non-Clifford operations and significantly decrease the hardware overhead to perform magic state distillation, there is limited knowledge on how to achieve this conversion. One solution is to utilize the symmetry in the target operation; we only consider Clifford operations that commute with the non-Clifford gate. While such an idea has been proposed for the case of Pauli twirling in Ref. [22], the method does not contribute to scrambling the noise among the global system, since only local operations are considered. In order to fully exemplify the early FTQC scheme, it is an urgent task to establish a unified understanding and methodology regarding the full symmetric Clifford operations.

In this work, we address this issue by proposing symmetric Clifford twirling, a twirling method that uses symmetric Clifford operators [23] commuting with certain Pauli subgroups. By appropriately choosing the Pauli subgroup, we obtain symmetric Clifford operators that also commute with non-Clifford operations, allowing us to twirl their noise. We fully characterize how Pauli noise channels are converted through symmetric Clifford twirling, and demonstrate that some Pauli noise can be scrambled to noise exponentially close to global white noise. Furthermore, we show that the effective noise channel of some highly structured circuits is scrambled to global white noise (termed the white-noise approximation [24]), and that we can accelerate the scrambling using only a single CNOT gate. We apply our techniques to the Trotterized Hamiltonian simulation of spin models and validate the efficacy of both symmetric Clifford twirling and the white-noise approximation in the early FTQC era.

2 Problem setup

In the early FTQC regime, it is expected that we cannot supply a large number of logical qubits for magic state factories because of the limited number of physical qubits available. Hence we shall choose magic state distillation protocol with mild suppression and low space overhead [25, 26], which results in a non-negligible amount of distillation error. This naturally leads us to assume that, in the early FTQC regime, the main source of logical error is the non-Clifford operations, while Clifford operations can be implemented with a negligible amount of errors (see Sec. S1 of Technical Manuscript (TM) [27] for details).

For the sake of clarity, let us especially consider the case where we want to implement Pauli-Z rotation gate $U = R_z(\theta) \otimes I^{\otimes n-1}$ on the first qubit of n-qubit logical circuit, and Pauli noise

$$\begin{aligned}\mathcal{N} = (1-p_{\mathrm{err}})\mathcal{I} + p_x\mathcal{E}_{\mathrm{X}\otimes\mathrm{I}^{\otimes n-1}} \\ + p_y\mathcal{E}_{\mathrm{Y}\otimes\mathrm{I}^{\otimes n-1}} + p_z\mathcal{E}_{\mathrm{Z}\otimes\mathrm{I}^{\otimes n-1}}\end{aligned} \tag{1}$$

affects the non-Clifford unitary U, where $\mathcal{E}_P(\cdot) = P \cdot P^\dagger$

*tsubouchi@noneq.t.u-tokyo.ac.jp
†nyoshioka@ap.t.u-tokyo.ac.jp

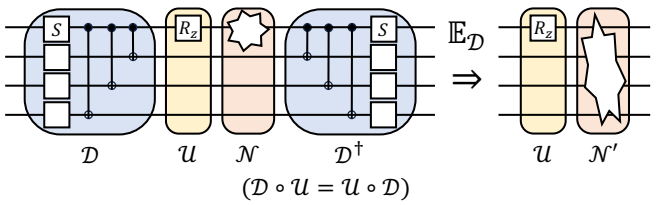


Figure 1: Conceptual diagram of symmetric Clifford twirling. By randomly sampling Clifford unitary $\mathcal{D}(\cdot) = D \cdot D^\dagger$ that commutes with non-Clifford layer $\mathcal{U}(\cdot) = U \cdot U^\dagger$, we can scramble the noise layer $\mathcal{N}$ without affecting $\mathcal{U}$.

and $p_{\rm err} = p_x + p_y + p_z$ is the total error rate. Our goal for this work is to scramble the noise layer $\mathcal{N}$ into global white noise defined as

$$\mathcal{N}_{{\rm wn},p_{\rm err}} := (1 - p_{\rm err})\mathcal{I} + p_{\rm err}\mathbb{E}_i[\mathcal{E}_{P_i}], \tag{2}$$

where $\mathbb{E}_i$ represents the uniform average over n-qubit Pauli noises $\mathcal{E}_{P_i}(\cdot) = P_i \cdot P_i^\dagger$. This is because it allows us to mitigate errors simply by rescaling the noisy expectation value with minimal sampling overhead of $e^{2p_{\rm tot}}$, which is not only a quadratic improvement from the previous method called probabilistic error cancellation [28–30], but also saturates the lower bound on the sampling overhead [16] (see Sec. S2 of TM [27] for details). Here, $p_{\rm tot}$ denotes the total logical error probability of the circuit. Furthermore, we can implement certain quantum algorithms robust to global white noise [17, 18] through this conversion.

3 Symmetric Clifford twirling

One naive way to obtain global white noise is to perform Clifford twirling: by randomly choosing a gate D from the n-qubit Clifford group $\mathcal{G}_n$ and applying the gate D and its conjugation $D^\dagger$ before and after the noise layer $\mathcal{N}$, we can scramble the noise layer $\mathcal{N}$ into the global white noise $\mathcal{N}_{{\rm wn},p_{\rm err}}$ as

$$\mathscr{T}(\mathcal{N}) := \mathbb{E}_{D\in\mathcal{G}_n}[\mathcal{D}^\dagger \circ \mathcal{N} \circ \mathcal{D}] = \mathcal{N}_{{\rm wn},p_{\rm err}}, \tag{3}$$

where $\mathcal{D}(\cdot) := D \cdot D^\dagger$ and $\mathscr{T}$ denotes the superchannel representing Clifford twirling [31]. This operation is, however, not considered as a practical option in the community, since the noise is inseparable from the target non-Clifford unitary. In order to insert D before the noise layer $\mathcal{N}$, one must insert $U^\dagger DU$ before the non-Clifford layer $\mathcal{U}$, which may introduce additional errors if $U^\dagger DU$ is a non-Clifford unitary. One feasible alternative is to consider Clifford unitaries D that commute with U, since $U^\dagger DU = D$ becomes a Clifford unitary in this case. This allows us to perform the twirling with negligible errors (see Fig. 1).

In order to realize such a twirling, let us characterize Clifford unitaries that commute with U. We denote the set of n-qubit Pauli operator as $\mathcal{P}_n := \{\mathrm{I},\mathrm{X},\mathrm{Y},\mathrm{Z}\}^{\otimes n}$ and define a Pauli subgroup

$$\mathcal{Q}_U := \langle\{P \in \mathcal{P}_n \mid {\rm tr}[PU] \neq 0\}\rangle, \tag{4}$$

Table 1: Reduction rate in the distance between the Pauli noise $\mathcal{N}$ represented as Eq. (1) and global white noise $\mathcal{N}_{{\rm wn},p_{\rm err}}$ when implementing symmetric Clifford twirling. The distance metric we use is the 2-norm v of the error probabilities p_i we define in TM, which can be used to bound the bias between the ideal and mitigated expectation value (see Sec. S3 of TM [27] for details).

Noise model	Pauli Z	depolarizing	Pauli X or Y
Reduction rate	1	$1/\sqrt{3}$	2^{-n}

where $\langle\cdot\rangle$ represents the group generated by the elements within the brackets. Additionally, let us define the $\mathcal{Q}_U$-symmetric Clifford group as:

$$\mathcal{G}_{n,\mathcal{Q}_U} := \{C \in \mathcal{G}_n \mid \forall P \in \mathcal{Q}_U,\ [C,P] = 0\}, \tag{5}$$

where its complete and unique construction method using simple quantum gates is given in Ref. [23]. From the definition, $D \in \mathcal{G}_{n,\mathcal{Q}_U}$ commutes with the non-Clifford operator U, enabling us to twirl the noise layer $\mathcal{N}$ using symmetric Clifford unitary $D \in \mathcal{G}_{n,\mathcal{Q}_U}$. We term this twirling as *symmetric Clifford twirling*, and the superchannel describing this twirling is defined as:

$$\mathscr{T}_{\mathcal{Q}_U}(\mathcal{N}) := \mathbb{E}_{D\in\mathcal{G}_{n,\mathcal{Q}_U}}[\mathcal{D}^\dagger \circ \mathcal{N} \circ \mathcal{D}]. \tag{6}$$

Especially when we consider the Pauli rotation gate $U = R_z(\theta) \otimes I^{\otimes n-1}$, $\mathcal{Q}_U$ simplifies to $\mathcal{Q}_U = \{\mathrm{I},\mathrm{Z}\} \otimes \{\mathrm{I}\}^{\otimes n-1}$. In this case, we can express the effect of symmetric Clifford twirling to the Pauli noise as presented in the following Theorem.

Theorem 1 *Let* $\mathcal{E}_{P\otimes\mathrm{I}^{\otimes n-1}}(\cdot) = (P \otimes \mathrm{I}^{\otimes n-1}) \cdot (P \otimes \mathrm{I}^{\otimes n-1})^\dagger$ *be a single-qubit Pauli channel with* $P = \mathrm{X},\mathrm{Y},\mathrm{Z}$ *and* $\mathcal{Q}_U = \{\mathrm{I},\mathrm{Z}\} \otimes \{\mathrm{I}\}^{\otimes n-1}$. *Then, by applying symmetric Clifford twirling to the Pauli channel as* $\mathscr{T}_{\mathcal{Q}_U}(\mathcal{E}_{P\otimes\mathrm{I}^{\otimes n-1}}) = \mathbb{E}_{D\in\mathcal{G}_{n,\mathcal{Q}_U}}[\mathcal{D}^\dagger \circ \mathcal{E}_{P\otimes\mathrm{I}^{\otimes n-1}} \circ \mathcal{D}]$, *we can scramble the Pauli-X and Y channels as*

$$\mathscr{T}_{\mathcal{Q}_U}(\mathcal{E}_{P\otimes\mathrm{I}^{\otimes n-1}}) = \mathop{\mathbb{E}}_{\substack{Q_1\in\{\mathrm{X},\mathrm{Y}\}\\ Q_2\in\mathcal{P}_{n-1}}}[\mathcal{E}_{Q_1\otimes Q_2}] \tag{7}$$

for $P = \mathrm{X},\mathrm{Y}$, *while the Pauli-Z channel cannot be scrambled through the symmetric Clifford twirling:*

$$\mathscr{T}_{\mathcal{Q}_U}(\mathcal{E}_{\mathrm{Z}\otimes\mathrm{I}^{\otimes n-1}}) = \mathcal{E}_{\mathrm{Z}\otimes\mathrm{I}^{\otimes n-1}}. \tag{8}$$

We generalize Theorem 1 to an arbitrarily non-Clifford gate U and Pauli noise $\mathcal{N}$ in TM [27]. Theorem 1 indicates that Pauli-Z noise remains unscrambled, while Pauli-X and Y noises are well dispersed among other qubits (albeit not exactly transformed to global white noise) through symmetric Clifford twirling, and become exponentially close to global white noise (Table 1). This encourages us to devise a method for implementing the T gate or the $R_z(\theta)$ gate such that the predominant error is Pauli-X or Y error [32]. Alternatively, we may focus on mitigating Pauli-Z noize using probabilistic error cancellation and subsequently use symmetric Clifford twirling to address the remaining Pauli-X and Y noises.

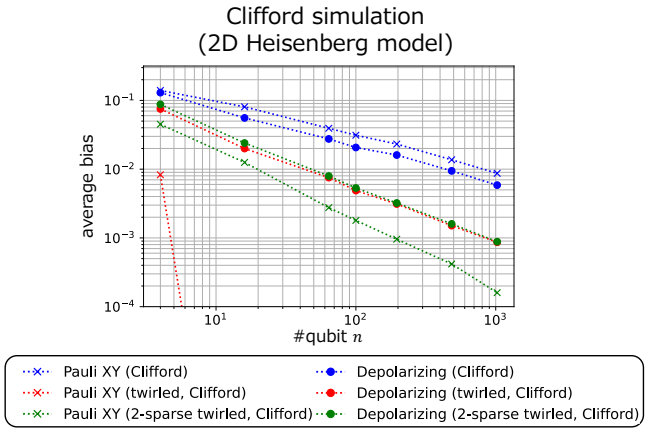


Figure 2: Performance of symmetric Clifford twirling for Trotterized Hamiltonian simulation of 2D Heisenberg model. Here we represent the average bias over random Pauli operators $P \in \mathcal{P}_n$ defined as $|R2^{-n}|\mathrm{tr}[\mathcal{N}_{\mathrm{eff}}(P)P]| - 1|$, where $\mathcal{N}_{\mathrm{eff}}$ is the effective noise channel of the noisy logical circuit. We fix the total error probability as $p_{\mathrm{tot}} := p_{\mathrm{err}}L = 1$ and set the Trotter step size as 50. The circle and the x dots represent the result for depolarizing noise and Pauli X and Y noise, while the red, blue, and green lines represent the results without symmetric Clifford twirling, with symmetric Clifford twirling, and with 2-sparse symmetric Clifford twirling.

It is worth noting that n-qubit Clifford unitary $D \in \mathcal{G}_{n,\mathcal{Q}_U}$ can be implemented with negligible errors using multi-qubit Pauli measurement (see Sec. S1 of TM [27] for details). Furthermore, even if the implementation of n-qubit Clifford unitary results in a non-negligible amount of logical errors, we can still effectively scramble the noise to some extent, by limiting the sampled symmetric Clifford unitaries to 2-qubit Clifford unitaries that always act nontrivially on the first qubit. We call such a twirling 2-*sparse symmetric Clifford twirling*, which enables us to scramble Pauli-X and Y noise to a Pauli noise with uniform distribution only among weight-2 Pauli channels, and obtain a noise that is polynomially close to the global white noise. We note that this process can be performed using only a single CNOT gate and some single-qubit Clifford gates, allowing us to scramble noise with minimal additional errors. We may also generalize such a sparse twirling to involve k-local Clifford gates to complete the symmetric Clifford twirling up to the kth order (see Sec. S4 of TM [27] for details).

4 Numerical analysis

To evaluate the efficacy of symmetric Clifford twirling in mitigating errors, we apply our techniques to the dynamics simulation circuit of first-order Suzuki-Trotter decomposition for the 2D Heisenberg model with the open boundary condition. We assume that $\mathcal{N}$ represents either Pauli noise consisting of Pauli-X and Y errors with $p_x = p_y = p_{\mathrm{err}}/2$, or depolarizing noise with $p_x = p_y = p_z = p_{\mathrm{err}}/3$. We set the total error rate to $p_{\mathrm{tot}} = 1$, which results in the constant sampling overhead of $e^2 \sim 7$. We further replace the $R_z(\theta)$ gate with $R_z(\pi/2) =$S gate to perform a large-scale Clifford simulation where the quantum advantage is expected.

We depict our results in Fig. 2. We observe that bias between ideal and error-mitigated expectation value decreases roughly as $1/\sqrt{n}$ as we increase the qubit count, even without performing symmetric Clifford twirling. This can be considered as the effect of white-noise approximation [24], where the effective noise of the quantum circuit is scrambled to the global white noise. We can view the symmetric Clifford twirling as a method to accelerate the noise scrambling in white-noise approximation. Indeed, we can see that the average bias significantly decreases when there is no Pauli-Z error. This consists with the result in Theorem 1, which states that Pauli-X and Y noise can be twirled into a noise that is exponentially close to the global white noise, while scrambling of Pauli-Z noise is prohibited. This result motivates us to devise a method for synthesizing the $R_z(\theta)$ gate with the dominant algorithmic error being Pauli-X or Y noise. Such a gate compilation is achieved in a probabilistic way under some scenarios [32], while when one can achieve such compilation is still an open question.

We note that we can accelerate white-noise approximation even for depolarizing noise where Pauli-Z noise exists. Furthermore, we can still accelerate the noise scrambling by only sparsely twirling the noise using a 2-qubit symmetric Clifford operator, and improve the scaling of the bias from $1/\sqrt{n}$ to $1/n$ when there is no Pauli-Z noise. We emphasize that the difference in the performance of the full twirling and the sparse twirling is negligible when the Pauli-Z noise remains untwirled, as we can see from the results of depolarizing noise.

It is worth noting that the validity of the white-noise approximation for such highly structured circuits is a phenomenon unique to the early FTQC regime. In fact, it is known that the approximation fails for shallow, log-depth quantum circuits in the noisy intermediate-scale quantum (NISQ) regime [33]. Furthermore, symmetric Clifford twirling is also unique to the early FTQC regime, where we can perform Clifford operations with negligible errors. Our new methodology thus paves the way for addressing logical errors in a significantly novel manner and contributes to the realization of quantum advantage.

References

[1] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory", Physical review A **52**, R2493 (1995).

[2] E. Knill, R. Laflamme, and W. Zurek, "Threshold accuracy for quantum computation", arXiv preprint quant-ph/9610011 (1996).

[3] D. Aharonov and M. Ben-Or, "Fault-tolerant quantum computation with constant error", in Proceedings of the twenty-ninth annual acm symposium on theory of computing (1997), pp. 176–188.

[4] D. A. Lidar and T. A. Brun, *Quantum error correction* (Cambridge university press, 2013).

[5] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. Girvin, L. Jiang, et al., "Extending the lifetime of a quantum bit with error correction in superconducting circuits", Nature **536**, 441 (2016).

[6] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, et al., "Realizing repeated quantum error correction in a distance-three surface code", Nature **605**, 669 (2022).

[7] G. Q. AI, "Suppressing quantum errors by scaling a surface code logical qubit", Nature **614**, 676 (2023).

[8] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, et al., "Logical quantum processor based on reconfigurable atom arrays", Nature **626**, 58 (2024).

[9] B. Eastin and E. Knill, "Restrictions on transversal encoded quantum gate sets", Physical review letters **102**, 110502 (2009).

[10] S. Bravyi and A. Kitaev, "Universal quantum computation with ideal clifford gates and noisy ancillas", Phys. Rev. A **71**, 022316 (2005).

[11] S. Bravyi and J. Haah, "Magic-state distillation with low overhead", Physical Review A **86**, 052329 (2012).

[12] A. G. Fowler, S. J. Devitt, and C. Jones, "Surface code implementation of block code state distillation", Scientific reports **3**, 1939 (2013).

[13] A. M. Meier, B. Eastin, and E. Knill, "Magic-state distillation with the four-qubit code", arXiv preprint arXiv:1204.4221 (2012).

[14] D. Litinski, "Magic state distillation: not as costly as you think", Quantum **3**, 205 (2019).

[15] J. O'Gorman and E. T. Campbell, "Quantum computation with realistic magic-state factories", Physical Review A **95**, 032338 (2017).

[16] K. Tsubouchi, T. Sagawa, and N. Yoshioka, "Universal cost bound of quantum error mitigation based on quantum estimation theory", Physical Review Letters **131**, 210601 (2023).

[17] Z. Ding, Y. Dong, Y. Tong, and L. Lin, "Robust ground-state energy estimation under depolarizing noise", arXiv preprint arXiv:2307.11257 (2023).

[18] Q. Liang, Y. Zhou, A. Dalal, and P. Johnson, "Modeling the performance of early fault-tolerant quantum algorithms", Physical Review Research **6**, 023118 (2024).

[19] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits", Physical review letters **119**, 180509 (2017).

[20] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid quantum-classical algorithms and quantum error mitigation", Journal of the Physical Society of Japan **90**, 032001 (2021).

[21] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, "Quantum error mitigation", Reviews of Modern Physics **95**, 045005 (2023).

[22] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, "Scalable error mitigation for noisy quantum circuits produces competitive expectation values", Nature Physics **19**, 752 (2023).

[23] Y. Mitsuhashi and N. Yoshioka, "Clifford group and unitary designs under symmetry", PRX Quantum **4**, 040331 (2023).

[24] A. M. Dalzell, N. Hunter-Jones, and F. G. Brandão, "Random quantum circuits transform local noise into global white noise", arXiv preprint arXiv:2111.14907 (2021).

[25] M. Vasmer and A. Kubica, "Morphing quantum codes", PRX Quantum **3**, 030319 (2022).

[26] T. Itogawa, Y. Takada, Y. Hirano, and K. Fujii, "Even more efficient magic state distillation by zero-level distillation", arXiv preprint arXiv:2403.03991 (2024).

[27] K. Tsubouchi, Y. Mitsuhashi, K. Sharma, and N. Yoshioka, "Symmetric clifford twirling for cost-optimal quantum error mitigation in early ftqc regime", arXiv preprint arXiv:2405.07720 (2024).

[28] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, "Error mitigation for universal gates on encoded qubits", Physical review letters **127**, 200505 (2021).

[29] M. Lostaglio and A. Ciani, "Error mitigation and quantum-assisted simulation in the error corrected regime", Physical review letters **127**, 200506 (2021).

[30] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, "Quantum error mitigation as a universal error reduction technique: applications from the nisq to the fault-tolerant quantum computing eras", PRX Quantum **3**, 010345 (2022).

[31] J. Emerson, R. Alicki, and K. Życzkowski, "Scalable noise estimation with random unitary operators", Journal of Optics B: Quantum and Semiclassical Optics **7**, S347 (2005).

[32] N. Yoshioka, S. Akibue, H. Morisaki, K. Tsubouchi, and Y. Suzuki, "Error crafting in probabilistic quantum gate synthesis", in preparation (2024).

[33] J. Foldager and B. Koczor, "Can shallow quantum circuits scramble local noise into global white noise?", arXiv preprint arXiv:2302.00881 (2023).

Schmidt Quantum Compressor

Israel F. Araujo[1 2 *] Hyeondo Oh[1] Nayeli A Rodríguez-Briones[3 4] Daniel K. Park [1 5 †]

[1] *Department of Statistics and Data Science, Yonsei University, Seoul 03722, Republic of Korea*
[2] *Department of Electronics and Systems, Federal University of Pernambuco, Recife 50740-550, Brazil*
[3] *Miller Institute for Basic Research in Science, University of California Berkeley, CA 94720, USA*
[4] *Atominstitut, Technische Universität Wien, Stadionallee 2, 1020 Vienna, Austria*
[5] *Department of Applied Statistics, Yonsei University, Seoul 03722, Republic of Korea*

Abstract. This work introduces the Schmidt Quantum Compressor, an innovative approach to quantum data compression that leverages the principles of Schmidt decomposition to encode quantum information efficiently. Unlike traditional variational quantum autoencoders, which rely on stochastic optimization and are prone to issues like barren plateaus, our deterministic method significantly reduces the complexity and computational overhead of quantum data compression. The compressor is rigorously evaluated through numerical experiments, demonstrating its potential to achieve high fidelity in quantum state reconstruction compared to variational methods. The applications of this technology span quantum simulation, communication, and distributed quantum computing, highlighting its versatility and the broad implications for future quantum technologies.

Keywords: quantum computing, Schmidt decomposition, quantum compression, quantum autoencoder

1 Introduction

Quantum compression is a technique in quantum computing aimed at reducing the dimensionality of quantum information while preserving its essential characteristics [1, 2]. This process involves encoding quantum states into a lower-dimensional space, allowing for more efficient storage and manipulation of quantum data. In this way, quantum compression can significantly optimize the use of quantum resources, making it a valuable tool in advancing quantum technologies.

Applications of quantum compression [3, 4] span a variety of fields, including quantum simulation [5], quantum communication [6], and distributed quantum computing [7, 8]. By minimizing the quantum resources required, quantum compression enhances the feasibility and efficiency of quantum simulations, enables more effective quantum communication protocols by reducing the quantum data transmitted over networks, and facilitates distributed computation within quantum networks. Importantly, this technology is not limited to purely quantum data; it can also efficiently transmit classical data through quantum networks by encoding it as a quantum state [9, 10, 11, 12, 13], thereby expanding the range of information that can be compressed and shared. These applications highlight the potential of quantum compression to improve the way both quantum and classical information is processed and transmitted.

Quantum autoencoders [14, 15, 16, 17, 18] are a popular approach to achieving quantum compression by employing a two-part mechanism: a compressor and a decompressor. The compressor maps the high-dimensional input quantum states into a lower-dimensional latent space, effectively compressing the information. This is followed by a decompressor that attempts to reconstruct the original quantum state from the compressed version. The success of a quantum autoencoder can be measured by the fidelity between the original and reconstructed states, which indicates how accurately the compression and subsequent decompression preserve the quantum information. Quantum autoencoders can be trained to maximize compression efficiency by optimizing the unitary transformations that define the encoding and decoding processes, demonstrating their critical contribution to the field of quantum data compression.

Built on the framework of variational quantum circuits, quantum autoencoders face various challenges that can impact their efficiency in compressing quantum data. These variational circuits depend on fine-tuning their parameters through optimization techniques to enhance performance. However, this fine-tuning process is often hindered by noise, which introduces errors complicating the process of determining gradient directions, increasing the likelihood of encountering what are known as *barren plateaus* [19].

Additionally, identifying the best circuit configuration becomes increasingly challenging [20]. The circuit configuration consists of two parts: an embedding, which maps classical data into a Hilbert space, and an ansatz, a parametrized circuit designed to approximate the solution to a problem. The embedding is unnecessary when working with quantum data.

Choosing the right embedding and ansatz for the circuit configuration demands a balance between the circuit's ability to represent complex states (expressibility [21]) and the ease with which it can be optimized (trainability [22]). The embedding must efficiently [23, 24] translate classical data into a quantum context, ensuring information is encoded without loss. On the ansatz side, the architecture of the variational circuit should be sufficiently sophisticated to represent the dataset's complexity without becoming unmanageable.

*ifa@yonsei.ac.kr
†dkd.park@yonsei.ac.kr

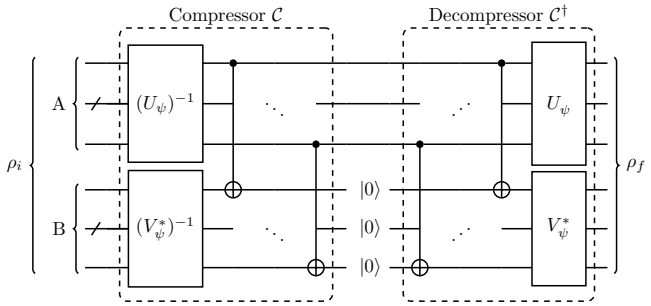


Figure 1: Schmidt Compression Protocol (SCP): quantum circuit overview. The operators U_ψ and V_ψ are derived from the typical state $|\psi\rangle$. They transform the basis of subsystems A and B to align with the computational basis. A sequence of CNOT gates then disentangles the two subsystems. When ρ_i is identical to $|\psi\rangle$, the separation is perfect, and subsystem A carries the singular values of ρ_i. Consequently, the decompressor $\mathcal{C}^\dagger$ can faithfully reproduce ρ_i without loss. For ρ_i differing from $|\psi\rangle$, the separation is approximate, with the quality of approximation ρ_f being dependent on the proximity between ρ_i and $|\psi\rangle$.

An ansatz that is too simplistic risks failing to model the data adequately, while one that is too elaborate may induce barren plateaus [19, 25] in the optimization landscape. Moreover, introducing more gates into the circuit tends to increase error propagation in NISQ-era devices, potentially corrupting the quantum state. Given these considerations, the careful selection of both embedding and ansatz is critical, as it profoundly impacts the effectiveness of variational quantum computing methods.

Furthermore, the variational nature of these circuits requires a significant amount of classical computational resources for optimization, which can be time-consuming. This is particularly challenging when dealing with complex quantum systems or attempting to scale up the quantum autoencoder for larger quantum datasets. The trade-off between the adaptability of variational circuits to different quantum compression tasks and the computational overhead introduced by the optimization process, combined with the presence of noise, requires robust error mitigation and optimization strategies in the development of quantum autoencoders.

Adopting a deterministic approach presents a strategic alternative to address the challenges faced by quantum autoencoders utilizing variational quantum circuits.

2 Schmidt compression protocol

The goal of the new protocol is to design a quantum data compression scheme based on a classical dataset or a dataset with samples from a quantum data source. The idea is to use the dataset to extract information about the source (quantum or classical), and then utilize that information to develop a quantum compression protocol.

The essence of the method is to find a typical state $|\psi\rangle$ that minimizes the distance between the entire sample set $\{|x_i\rangle\}$ and $|\psi\rangle$, thereby maximizing the fidelity of the entire dataset towards 1. The compression strategy is then built around representing states in the vicinity of $|\psi\rangle$ using fewer dimensions. We discovered that, for datasets with real and non-negative features, the optimal choice for the typical state is the average of the normalized sample vectors $|\psi\rangle = \frac{\sum_{i=1}^{M}|x_i\rangle}{M}$, where M is the number of samples.

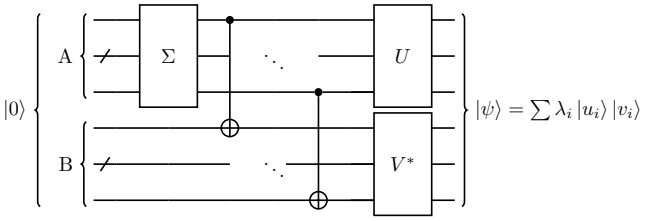


Figure 2: Schmidt Quantum State Preparation: quantum circuit overview. The operator Σ encodes the singular values of the reshaped state vector as the amplitude probabilities of subsystem A (first half of the quantum register). A sequence of CNOT gates then entangles subsystems A and B. The operators U and V are derived from the input state vector. They transform the computational basis of subsystems A and B to align with the Schmidt basis.

2.1 Protocol principle

Given a bipartite state $|\psi\rangle$ in $\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}$, its Schmidt decomposition can be written as:

$$|\psi\rangle = \sum_{i=1}^{k} \lambda_i |u_i\rangle_\mathrm{A} |v_i\rangle_\mathrm{B} . \tag{1}$$

If the typical state $|\psi\rangle$ is reshaped into a matrix form M_ψ, and then an SVD is performed on this matrix, the result is:

$$M_\psi = U_\psi \Sigma_\psi V_\psi^\dagger. \tag{2}$$

where the columns of U_ψ (left singular vectors) correspond to the Schmidt basis vectors $|u_i\rangle_\mathrm{A}$ for subsystem A, the columns of V_ψ (right singular vectors) correspond to the Schmidt basis vectors $|v_i\rangle_\mathrm{B}$ for subsystem B, and the diagonal elements of Σ_ψ are the singular values, which are equivalent to the Schmidt coefficients λ_i. It's crucial to note that the partitioning of the system into blocks do not need to be continuous or uniform in size, as illustrated in Figure 1

The compression unitary $\mathcal{C}$, depicted in Figure 1, is formulated from Equation (2) and is a derivative of the Schmidt state preparation circuit [10, 13]. Observing that the first operator in the Schmidt circuit (see Figure 2) is a state preparation mechanism that encodes the singular values into the amplitudes of subsystem A, it is feasible to reverse this circuit and eliminate the associated operator Σ to generate the state $|\lambda\rangle$ (see Figure 3). With $\mathcal{C}$ as the compression unitary, $\mathcal{C}^\dagger$ functions as the decompression unitary.

When $\mathcal{C}$ acts upon the typical state $|\psi\rangle$, it precisely produces $|\lambda\rangle$ in subsystem A and $|0\rangle$ in subsystem B,

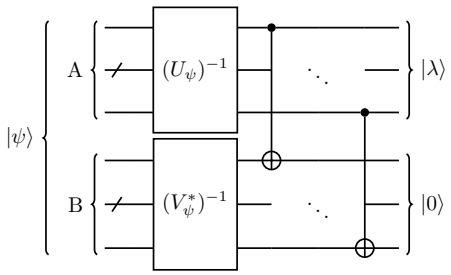


Figure 3: The figure showcases a compressor circuit adapted from the Schmidt Quantum State Preparation. In this adaptation, the original circuit is inverted, and the unitary component Σ is omitted. Upon application to a typical state $|\psi\rangle$, the circuit yields a new state $|\lambda\rangle$, which encodes the singular values formerly attributed to Σ_ψ within subsystem A. After this transformation, subsystems A and B become fully disentangled, resulting in subsystem B being characterized by the state $|0\rangle$.

as shown in Equation (4). In this configuration, $\mathcal{C}^\dagger$ is capable of flawlessly reconstructing $|\psi\rangle$ as illustrated in Figure 4a.

$$(U_\psi \otimes V_\psi^*)^{-1} |\psi\rangle = \sum_i \lambda_i |i\rangle_{\mathrm{A}} |i\rangle_{\mathrm{B}} \tag{3}$$

$$\left(\prod_{i=1}^{m} \mathrm{CNOT}_i\right) \sum_i \lambda_i |i\rangle_{\mathrm{A}} |i\rangle_{\mathrm{B}} = \sum_i \lambda_i |i\rangle_{\mathrm{A}} |0\rangle_{\mathrm{B}} \tag{4}$$

Now, consider a state $|x_i\rangle$ that is close to $|\psi\rangle$. We are interested in how $|x_i\rangle$ projects onto the known Schmidt basis vectors of $|\psi\rangle$. The projection will look like:

$$\alpha_i = \langle u_i|_{\mathrm{A}} \langle v_i|_{\mathrm{B}} |x_i\rangle \tag{5}$$

The projections translate in terms of closeness in the following way:

- **Exact match.** If $|x_i\rangle = |\psi\rangle$, the coefficients α_i will match exactly with the Schmidt coefficients λ_i.
- **Close but not exact.** If $|x_i\rangle$ is near $|\psi\rangle$ (in terms of high state fidelity between them), the coefficients α_i will be close to λ_i, but with some deviations. The magnitude and pattern of these deviations can offer insights into how $|x_i\rangle$ deviates from $|\psi\rangle$.
- **Orthogonal or unrelated.** If $|x_i\rangle$ is orthogonal or largely unrelated to $|\psi\rangle$ in terms of certain Schmidt modes, the corresponding α_i values will be small or even zero.

The entire set $\{\alpha_i\}$ can be seen as a signature or fingerprint of how $|x_i\rangle$ projects onto the Schmidt bases of $|\psi\rangle$.

3 Experiment and discussion

In this section, we contrast the performance of the Schmidt compressor algorithm with that of the Quantum Autoencoder, using fidelity as the benchmark. Fidelity measures how closely the recovered state from the complete circuit matches the initial quantum state.

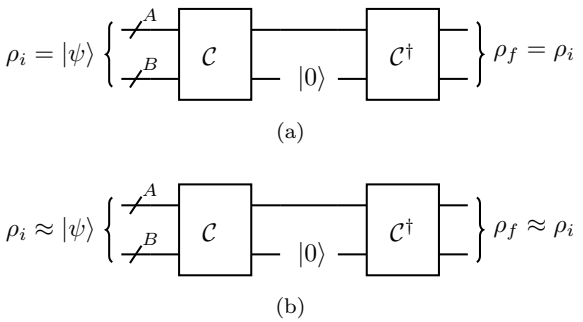


Figure 4: Operational dynamics of the complete circuit on various states. (a) Scenario where the input state precisely matches the typical state, resulting in a lossless recovery of the input state. (b) Scenario where the input state is in close proximity to the typical state, resulting in a high-fidelity approximation of the input state, although not an exact recovery.

Table 1: Comparative analysis of the Variational Quantum Autoencoder and Schmidt Compression techniques applied to the Optical Recognition of Handwritten Digits dataset [26, 27] under ideal simulation conditions. The columns labeled **Avg** display the average fidelity measured across 20 pairs of original and reconstructed states, offering insights into the effectiveness of each method in preserving quantum state information during compression.

Label	Quantum autoencoder		Schmidt compressor	
	Avg	**Std**	**Avg**	**Std**
0	0.815	0.080	**0.841**	0.073
1	0.700	0.168	0.679	0.176
2	0.715	0.115	**0.736**	0.117
3	0.699	0.117	**0.725**	0.118
4	0.694	0.093	**0.709**	0.116
5	0.705	0.091	**0.706**	0.097
6	0.744	0.082	**0.786**	0.092
7	0.703	0.098	0.699	0.117
8	0.694	0.088	**0.713**	0.093
9	0.633	0.147	**0.671**	0.147

This study employs the Optical Recognition of Handwritten Digits dataset [26, 27]. The dataset was segmented into ten unique subsets, each representing one digit class in the range from 0 to 9. Every subset comprised 180 samples. For each digit class, 20 test samples were evaluated to calculate the mean fidelity and standard deviation. The simulation findings are summarized in Table 1, which compares the performance of the Quantum Autoencoder with that of the Schmidt compressor in terms of fidelity metrics. The results show that the Schmidt compressor outperforms quantum autoencoder in 8 out of 10 data compression tasks.

The standout benefit of the Schmidt compressor is its independence from optimization, a process that can often be resource-intensive and time-consuming. Moreover, it circumvents issues commonly associated with optimization, such as barren plateaus, and eliminates the complex decision-making involved in selecting the optimal embedding and ansatz for variational circuits.

References

[1] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[2] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[3] Matej Pivoluska and Martin Plesch. Implementation of quantum compression on ibm quantum computers. *Scientific Reports*, 12(1):5841, 2022.

[4] Lee A Rozema, Dylan H Mahler, Alex Hayat, Peter S Turner, and Aephraim M Steinberg. Quantum data compression of a qubit ensemble. *Physical review letters*, 113(16):160504, 2014.

[5] Chen-Rui Fan, Bo Lu, Xue-Ting Feng, Wei-Chao Gao, and Chuan Wang. Efficient multi-qubit quantum data compression. *Quantum Engineering*, 3(2):e67, 2021.

[6] Martin Plesch and Vladimír Bužek. Efficient compression of quantum information. *Physical Review A*, 81(3):032317, 2010.

[7] Marcello Caleffi, Michele Amoretti, Davide Ferrari, Daniele Cuomo, Jessica Illiano, Antonio Manzalini, and Angela Sara Cacciapuoti. Distributed Quantum Computing: a Survey. 2022.

[8] Seng Loke. *From distributed quantum computing to quantum internet computing: an introduction*. Wiley, Hoboken, New Jersey, 2023.

[9] Mikko Möttönen, Juha J. Vartiainen, Ville Bergholm, and Martti M. Salomaa. Transformation of quantum states using uniformly controlled rotations. *Quantum Info. Comput.*, 5(6):467–473, 2005.

[10] Martin Plesch and Časlav Brukner. Quantum-state preparation with universal gate decompositions. *Phys. Rev. A*, 83:032302, 2011.

[11] Israel F Araujo, Daniel K Park, Francesco Petruccione, and Adenilton J da Silva. A divide-and-conquer algorithm for quantum state preparation. *Scientific reports*, 11(1):6329, 2021.

[12] Israel F. Araujo, Daniel K. Park, Teresa B. Ludermir, Wilson R. Oliveira, Francesco Petruccione, and Adenilton J. Da Silva. Configurable sublinear circuits for quantum state preparation. *Quantum Information Processing*, 22(2):123, 2023.

[13] Israel F. Araujo, Carsten Blank, Ismael C. S. Araújo, and Adenilton J. da Silva. Low-rank quantum state preparation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(1):161–170, 2024.

[14] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, aug 2017.

[15] Chenfeng Cao and Xin Wang. Noise-assisted quantum autoencoder. *Phys. Rev. Appl.*, 15:054012, May 2021.

[16] Chang-Jiang Huang, Hailan Ma, Qi Yin, Jun-Feng Tang, Daoyi Dong, Chunlin Chen, Guo-Yong Xiang, Chuan-Feng Li, and Guang-Can Guo. Realization of a quantum autoencoder for lossless compression of quantum data. *Physical Review A*, 102(3):032412, 2020.

[17] Hailan Ma, Chang-Jiang Huang, Chunlin Chen, Daoyi Dong, Yuanlong Wang, Re-Bing Wu, and Guo-Yong Xiang. On compression rate of quantum autoencoders: Control design, numerical and experimental realization. *Automatica*, 147:110659, 2023.

[18] L Lamata, U Alvarez-Rodriguez, J D Martín-Guerrero, M Sanz, and E Solano. Quantum autoencoders via quantum adders with genetic algorithms. *Quantum Science and Technology*, 4(1):014007, oct 2018.

[19] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, 2018.

[20] Tak Hur, Israel F. Araujo, and Daniel K. Park. Neural quantum embedding: Pushing the limits of quantum supervised learning. 2023.

[21] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021.

[22] Supanut Thanasilp, Samson Wang, Marco Cerezo, and Zoë Holmes. Exponential concentration and untrainability in quantum kernel methods. *arXiv preprint arXiv:2208.11060*, 2022.

[23] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.

[24] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

[25] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus. *PRX Quantum*, 3(1):010313, 2022.

[26] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[27] Ethem Alpaydin and Cenk Kaynak. Cascading classifiers. *Kybernetika*, 34(4):369–374, 1998.

Generalisation of Quantum Reservoir Computing with Polynomial Readout

Naomi Mona Chmielewski[1 2 *] Nina Amini[2] Joseph Mikael[1]

[1] *EDF Lab, Palaiseau, France*
[2] *CNRS, Laboratoire des Signaux et Systèmes (L2S), Université Paris-Saclay, Gif-sur-Yvette*

Abstract. We propose to analyse the ability of quantum reservoir classes with polynomial readout functions to generalise on unseen data. We find an upper bound on the Rademacher complexity as well as a risk bound that scales linearly in the maximal Lipschitz constant of the readout class. For a polynomial readout, which is often used to prove universality in quantum reservoirs, this constant scales very unfavourably in the number of qubits. Finally, we open up new avenues to find universal quantum reservoir classes with better risk bounds.

Keywords: quantum reservoir computing, quantum machine learning, time series forecasting

1 Introduction

Reservoir Computing (RC) is a machine learning paradigm that has applications in time series forecasting and other memory-based problems [1]. It represents an alternative to conventional Recurrent Neural Networks [2] that is more efficient in the computation time and data resources, as training only requires a linear regression step whereas the weights in the reservoir are left unchanged. While initially modelled as a type of neural network [3], physical implementations of RC have become popular [4]. The idea is to inject time series data collected from a source system into a physical system, to let the system react and evolve according to its natural dynamics, and to keep injecting the data at regular intervals in order to reproduce the behaviour of the source system, and to use this behaviour to make predictions.

As an example, one might collect time series data of the energy produced by a wind turbine and inject the data into the wave patterns of a water basin, similarly to [5].

Several proposals to use quantum systems as a physical RC have recently attracted attention, as the compact nature of RC lends itself well to NISQ implementations [6, 7, 8]. Furthermore, the fact that no parameter updating is needed in RC circumvents the barren plateau problem commonly encountered in Variational / Parameterised Quantum Circuits.

The two key ingredients of RC are given by the functional that describes the time evolution of the dynamical system driven by the input, and the readout function of the reservoir state that is used to make predictions. An important property that we require of a reservoir is known as the *fading memory property* (FMP), a property that states that two input sequences that were different in the distant past, but are similar in the present, should yield similar outputs from the reservoir.

Theoretical analyses of both classical RC and quantum RC have established several classes of reservoir functionals and readout functions that produce universal classes [9, 10, 11, 12, 13, 14]. A universal class is here understood to be a class of maps that can approximate any map with the FMP arbitrarily well.

Apart from universality, another important property of a machine learning model is its ability to generalise on unseen data. The *generalisation error* of a hypothesis class is often bounded through the *Rademacher complexity*, a measure that quantifies how well a hypothesis class can reproduce random noise. A high Rademacher complexity suggests that the hypothesis class might be prone to overfitting. A bound on the generalisation error is also called a risk bound. In classical RC, risk bounds have been studied for several universal reservoir classes [15].

To our knowledge, the Rademacher complexity of quantum RC has not yet been studied.

2 Contributions

We propose to bound the Rademacher complexity of a general class of quantum reservoirs. We then establish risk bounds of subclasses of the universal quantum reservoir classes introduced in [11] and [12]. Both classes employ a multivariate polynomial readout which is used to prove universality.

We find that a significant contributor to the way that the risk bound scales is given by a bound on the Lipschitz-constants of the readout functions. In particular, this leads to a very unfavourable scaling of the risk bound in the number of qubits when using a class of multivariate polynomial readouts. In other words, if one wants to use either of the quantum reservoir subclasses above, our bound suggests that the ability of these classes to perform well on unseen data quickly explodes as the number of qubits increases.

The full proofs and additional results will appear in the forthcoming work [16].

3 Main Results

3.1 An Upper Bound on the Rademacher Complexity

We consider a class $\mathcal{H}_n^{\mathrm{QRC}}$ of quantum reservoirs with n qubits with a class of Lipschitz-continuous readout func-

*naomi-mona.chmielewski@centralesupelec.fr

tions such that for all members of the class, the Lipschitz constant is bounded by some real number $\bar{L}_h$. Then, under some additional hypotheses, for k samples of Rademacher random variables, the Rademacher complexity of the class of quantum reservoirs is bounded by

$$\mathcal{R}_k(\mathcal{H}_n^{\mathrm{QRC}}) \leq \frac{\bar{L}_h}{\sqrt{k}} \ . \tag{1}$$

Now consider a class of multivariate polynomial readout functions in the measurements of the n qubits along the Z-axis with weights $\{w_i\}$ such that the maximal value of the weights is bounded by $|||W|||_\infty := \max_i |w_i|$. Furthermore, suppose that the maximal degree of the polynomial is given by $R_{\max}$. Then we find that a bound $\bar{L}_h^{\mathrm{poly}}$ on the Lipschitz-constants of the readout functions is given by

$$\bar{L}_h^{\mathrm{poly}} = |||W|||_\infty R_{\max} \cdot n\sqrt{2^n} \left(\binom{n + R_{\max}}{R_{\max}} - 1 \right) \ . \tag{2}$$

Plugging this expression into (1), we immediately see that this implies that the bound on the Rademacher complexity scales as $n\sqrt{2^n}\binom{n+R_{\max}}{R_{\max}}$ in the number of qubits.

3.2 Scaling of Risk Bounds

The risk bound that we apply here, which is an adaptation from the bound in [15], scales linearly in $\bar{L}_h$. Under certain conditions, we show that the risk bound can be written as

$$\mathcal{O}\left(\bar{L}_h \max\left\{ \frac{C_1}{m} \ , \ C_2 \frac{\log m}{m} \ , \ C_3 \sqrt{\frac{\log m}{m}} \ , \ C_4 \sqrt{\frac{\log 4/\delta}{2m}} \right\} \right)$$

where C_1, C_2, C_3, C_4 are constants that depend on the parameters of the reservoir. This bound implies that if $\bar{L}_h$ is sufficiently small, the reservoir class generalises well. On the other hand, if $\bar{L}_h$ scales badly, so does the risk bound. While this can be counteracted somewhat by the right choice of parameters (C_1, C_2, C_3, C_4), they cannot be made arbitrarily small.

Plugging in the expression of $\bar{L}_h^{\mathrm{poly}}$ from (2), it is clear that this bound scales badly in the number of qubits when using a polynomial readout. One might want to control this constant by adjusting the weights, for example by choosing $\frac{1}{|||W|||_\infty} \geq n\sqrt{2^n}\left(\binom{n+R_{\max}}{R_{\max}} - 1\right)$. This however is not advisable, as for increasing number of qubits and accounting for numerical error, the weights would quickly all be set to zero. Another possibility is to choose a linear readout, which would reduce the constant from $\bar{L}_h^{\mathrm{poly}}$ to $\bar{L}_h^{\mathrm{lin}} = |||W|||_\infty n\sqrt{2^n}$.

4 Discussion

The results from Section 3 suggest that using polynomial readouts to guarantee universality might negatively impact the reservoir class's ability to generalise. This motivates the search for universal quantum reservoir classes that do not make use of polynomial readouts to prove universality. In [17] the authors show that an input-dependent CPTP map is isomorphic to the *State Affine System* (SAS) introduced in [10], which has been shown to be universal under certain conditions, when equipped with a linear readout function. Further research could identify conditions on the CPTP map so that the universality of [10] would be applicable to the quantum reservoir.

References

[1] F. M. Bianchi, S. Scardapane, S. Lokse and R. Jenssen. Reservoir computing approaches for representation and classification of multivariate time series In *IEEE Transactions on Neural Networks and Learning Systems*, 32(5), 2021.

[2] W. Maass and T. Natschläger and H. Markram. Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations In *Neural Computation*, 14(11), 2002.

[3] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks In *GMD Report, German National Research Institute for Computer Science*, 148, 2001.

[4] K. Nakajima. Physical reservoir computing—An introductory perspective In *Japanese Journal of Applied Physics*, 59(6), 2020.

[5] C. Fernando and S. Sojakka. Pattern Recognition in a Bucket In *Advances in Artificial Life*, Chapter 63, 2003.

[6] K. Fujii and K. Nakajima. Harnessing disordered ensemble quantum dynamics for machine learning In *Phys. Rev. Applied*, 8(2), 2016.

[7] Y. Suzuki, Q. Gao, K. C. Pradel, K. Yasuoka and N. Yamamoto. Natural quantum reservoir computing for temporal information processing In *Scientific Reports*, 12(1), 2022.

[8] T. Yasuda, Y. Suzuki, T. Kubota, K. Nakajima, Q. Gao, W. Zhang, S. Shimono, H. I. Nurdin and N. Yamamoto. Quantum reservoir computing with repeated measurements on superconducting devices In *arXiv preprint*, 12(1), 2022.

[9] L. Grigoryeva and J. -P. Ortega. Echo state networks are universal In *Neural Networks*, 18, 2018.

[10] L. Grigoryeva and J. -P. Ortega. Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems In *Journal of Machine Learning Researchs*, 19, 2018.

[11] J. Chen and H. I. Nurdin. Learning Nonlinear Input-Output Maps with Dissipative Quantum Systems In *Quantum Information Processing*, 18(7), 2019.

[12] J. Chen, H. I. Nurdin and N. Yamamoto. Temporal information processing on noisy quantum computers In *Phys. Rev. Applied*, 14(2), 2020.

[13] A. Sannia, R. Martinez-Pena, M. C. Soriano G. L. Giorgi and R. Zambrini. Dissipation as a resource for Quantum Reservoir Computing In *Quantum*, 8, 2024.

[14] F. Monzani and E. Pratia. Universality conditions of unified classical and quantum reservoir computing In *arXiv preprint*, 2024.

[15] L. Gonon, L. Grigoryeva and J. -P. Ortega. Risk Bounds for Reservoir Computing In *Journal of Machine Learning Research*, 21(240), 2020.

[16] N. M. Chmielewski, N. Amini and J. Mikael. Risk Bounds for Quantum Reservoirs with Polynomial Readout *In preparation*, 2024.

[17] R. Martinez-Pena and J. -P. Ortega. Quantum reservoir computing in finite dimensions In *Physical Review E* , 107(3), 2023.

Unconditionally decoherence-free quantum error mitigation by density matrix vectorization

Zhong-Xia Shang[1 2 3 *] Zi-Han Chen[1 2 3 †] Cai-Sheng Cheng[4 ‡]

[1] *Hefei National Research Center for Physical Sciences at the Microscale and School of Physical Sciences, University of Science and Technology of China, Hefei 230026, China*

[2] *Shanghai Research Center for Quantum Science and CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, China*

[3] *Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*

[4] *Department of Chemical Physics, University of Science and Technology of China, Hefei, Anhui 230026, China*

Abstract. In this work[arXiv:2405.07592], we give a new paradigm of quantum error mitigation based on the vectorization of density matrices. Different from the ideas of existing quantum error mitigation methods that try to distill noiseless information from noisy quantum states, our proposal directly changes the way of encoding information and maps the density matrices of noisy quantum states to noiseless pure states, which is realized by a novel and NISQ-friendly measurement protocol and a classical postprocessing procedure. Our protocol requires no knowledge of the noise model, no ability to tune the noise strength, and no ancilla qubits for complicated controlled unitaries. Under our encoding, NISQ devices are always preparing pure quantum states which are highly desired resources for variational quantum algorithms to have good performance in many tasks. We show how this protocol can be well-fitted into variational quantum algorithms. We give several concrete ansatz constructions that are suitable for our proposal and do theoretical analysis on the sampling complexity, the expressibility, and the trainability. We also give a discussion on how this protocol is influenced by large noise and how it can be well combined with other quantum error mitigation protocols. The effectiveness of our proposal is demonstrated by various numerical experiments.

Keywords: Variational Quantum Algorithms,Quantum Error Mitigation,Density Matrix Vectorization

1 Basic idea

Currently, we are in the Noisy Intermediate Scale Quantum (NISQ) era [1], where high-quality qubit construction and large-scale fault-tolerant quantum computing remain elusive. Consequently, NISQ devices cannot execute complex quantum algorithms requiring deep circuits and numerous qubits [2]. To leverage the demonstrated quantum advantages of these devices [3, 4], researchers focus on developing NISQ-friendly algorithms like Variational Quantum Algorithms (VQAs) [5]. VQAs, such as Variational Quantum Eigensolver [6] and Quantum Approximate Optimization Algorithm [7], integrate classical and quantum computing by iteratively optimizing parametrized quantum circuits to minimize a cost function. Despite their resilience to noise and shallow circuit depth, VQAs suffer performance degradation due to device noise [8, 9]. This noise-induced error manifests as limitations in state preparation, circuit depth, and barren plateau problems [10]. To mitigate these issues, classical strategies like neural networks, Clifford circuits, and tensor networks are proposed [11, 12, 13]. Alternatively, Quantum Error Mitigation (QEM) methods aim to distill noiseless information from noisy states using approaches such as probabilistic error cancellation [14, 15].

In our work, we introduce a novel QEM paradigm for VQAs based on density matrix vectorization (DMV). The basic idea of our protocol is inspired by the mapping:

$$\rho \to |\rho\rangle \tag{1}$$

where $\rho = \sum_{ij} \rho_{ij}|i\rangle\langle j|$ and $|\rho\rangle$ is defined as $\frac{1}{C_\rho}\sum_{ij}\rho_{ij}|i\rangle|j\rangle$ with the normalization factor $C_\rho = ||\rho||_F = \sqrt{\sum_{ij}|\rho_{ij}|^2} = \sqrt{\mathrm{Tr}(\rho^2)}$. While DMV has been vastly used as a useful mathematical trick for simplifying many concepts in quantum information science, in this work, we see DMV in a different way. The main concept transition here is that the mapping Eq. 1 means we are treating an n-qubit density matrix ρ as a $2n$-qubit pure state $|\rho\rangle$, which shares the same idea as in Ref. [16] to demonstrate a new exponential quantum speedup. By this encoding, any quantum state either pure or mixed will always be mapped to a pure state. Note that the reverse mapping $|\rho\rangle \to \rho$ has also been adopted in several works [17, 18] for simulating open quantum systems using unitary circuits.

Currently, due to the Hermiticity and the positive semi-definiteness of ρ, we can not encode all $2n$-qubit pure states. To solve this problem, we can introduce a generalized mapping f:

$$\begin{aligned}&\{\rho_1, \rho_2, ..., \rho_K, c_1, c_2, ..., c_K\} \\ &\to |\psi\rangle = \frac{1}{C_\psi}\sum_{ij}(c_1\rho_{1,ij} + c_2\rho_{2,ij} + ... + c_K\rho_{k,ij})|i\rangle|j\rangle\end{aligned} \tag{2}$$

where we use K density matrices with K complex coefficients to form a pure state $|\psi\rangle$ with the normalization factor $C_\psi = \sqrt{\sum_{ij}|c_1\rho_{1,ij} + c_2\rho_{2,ij} + ... + c_k\rho_{k,ij}|^2}$, And it is sufficient to set $K = 4$ with 2 real and 2 imaginary coefficients to express any pure state $|\psi\rangle$. This can be understood from the matrix form $|\psi\rangle \to \psi$ which can be decomposed first into Hermitian part and $i*$Hermitian

*ustcszx@mail.ustc.edu.cn

†czh007@mail.ustc.edu.cn

‡ccs112202@mail.ustc.edu.cn

(anti-Hermitian) part, each Hermitian part is indefinite and can be further expressed as a linear combination of two positive semi-definite density matrices.

To give this mapping a real sense, we need to be able to extract information of $|\psi\rangle$ from $\{\rho_1, \rho_2, ..., \rho_K, c_1, c_2, ..., c_K\}$. Interestingly, we find that, given a Hamiltonian H_A, the expectation value with respect to $|\psi\rangle$ can be re-expressed as:

$$\langle\psi|H_A|\psi\rangle = \frac{\sum_{i,j=1}^{K} c_i^* c_j \mathrm{Tr}(H_B \rho_i \otimes \rho_j)}{\sum_{k,l=1}^{K} c_k^* c_l \mathrm{Tr}(\rho_k \rho_l)}$$

with

$$\langle il|H_B|jk\rangle = \langle ij|H_A|kl\rangle \tag{3}$$

We call H_B the substitute Hamiltonian of H_A, And it means that by measuring the values of all $\mathrm{Tr}(H_B \rho_i \otimes \rho_j)$ and $\mathrm{Tr}(\rho_i \rho_j)$, one can obtain the value of $\langle\psi|H_A|\psi\rangle$ by classical post combination.

2 Summary of our main results

Unlike the philosophy of extracting noiseless information from noisy quantum states by fighting against noise concentration in existing QEM protocols, our protocol realizes error mitigation by changing the way of encoding information. We encode pure states into linear combinations of vectorized density matrices which directly leads to unconditionally decoherence-free pure state preparations from noisy quantum circuits. By our protocol, the performance gap induced by decoherence can be directly eliminated and the gap led by low expressibility can be well mitigated without introducing additional barren plateau resources. Our protocol requires nothing(the knowledge of the noise model, the tunability for the noise strength and the controlled unitary for indirect measurements) but only needs 2-qubit collective unitaries before measurements to extract needed information. In the following, we will show four key results of our QEM protocol.

Result 1(Framework) (See fig1)the idea of classical linear combinations of quantum states has been adopted in several proposals [19, 20, 21] for variational quantum algorithms to enhance the expressibility of NISQ devices without aggravating the barren plateau problems. The key component in these protocols is the measurement strategy for values like $\langle\psi_i|O|\psi_j\rangle$ which typically require indirect hardware-challenging modified Hadamard tests [19] or complicated direct measurement strategies [22]. In contrast, an interesting and nice property of our measurement procedure is that values like $Tr(H_B \rho_i \otimes \rho_j)$ that contain the unnormalized information of $\langle\rho_i|H_A|\rho_j\rangle$ can be estimated by hardware-friendly direct measurements as shown before. Thus, our proposal inherits the advantages of classical linear combinations as in those proposals and at the same time, is easier to realize on NISQ hardware.

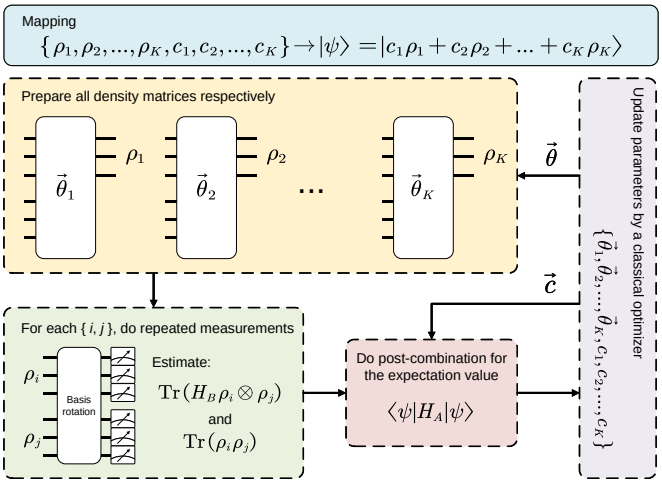


Figure 1: Running VQAs with QEM by DMV.

Result 2(Sampling complexity) The sampling complexity of estimating Eq. 3 hinges on the number K of classical combinations and the purity of $\{\rho_1, \rho_2, ..., \rho_K\}$, determined by factors such as the number of CNOT gates L and the circuit fault rate ζ, with 2^{-L} serving as a lower bound for purity in the noiseless case and $e^{-2\zeta}$ in the presence of faults modeled by a Poisson distribution.The true lower bound is thus a competition between 2^{-L} and $e^{-\zeta}$, which results in the following sampling complexity:

$$N \geq \max\left[e^{4\zeta}, 2^{2L}\right] \frac{3K^2}{\epsilon^2} \left(\frac{m||H_A||_F^2}{2^{2n}} + ||H_A||_2^2\right) \tag{4}$$

and it is also upper bounded by:

$$N \leq 2^{2n} \frac{3K^2}{\epsilon^2} \left(\frac{m||H_A||_F^2}{2^{2n}} + ||H_A||_2^2\right) \tag{5}$$

the basic idea of our QEM protocol is fundamentally different from others. However, in terms of the sampling complexity, the scaling with respect to the circuit fault rate ζ of our protocol is $e^{4\zeta}$ which interestingly coincides with the general exponential scaling behaviors of the sampling overhead in other QME protocols [23]. Indeed, our protocol solves the decoherence by directly changing the way of encoding quantum states, however, to retrieve information under this encoding, an $e^{4\zeta}$ scaling has to be respected because of the purity restriction. Compared with other protocols, our sampling complexity shares the same scaling of ζ as using the probabilistic error cancellation method [14, 15] to totally eliminate errors and as using the two-copy version of the virtual state distillation method [24]. When the task is to obtain information on pure states, our method is superior to the virtual state distillation since it needs a M-copy version with M large to capture the pure state behaviors well which, however, will result in a $e^{2M\zeta}$ scaling on the sampling complexity.

Result 3(Expressibility) To quantitatively evaluate the expressibility of ansatz we introduced above under

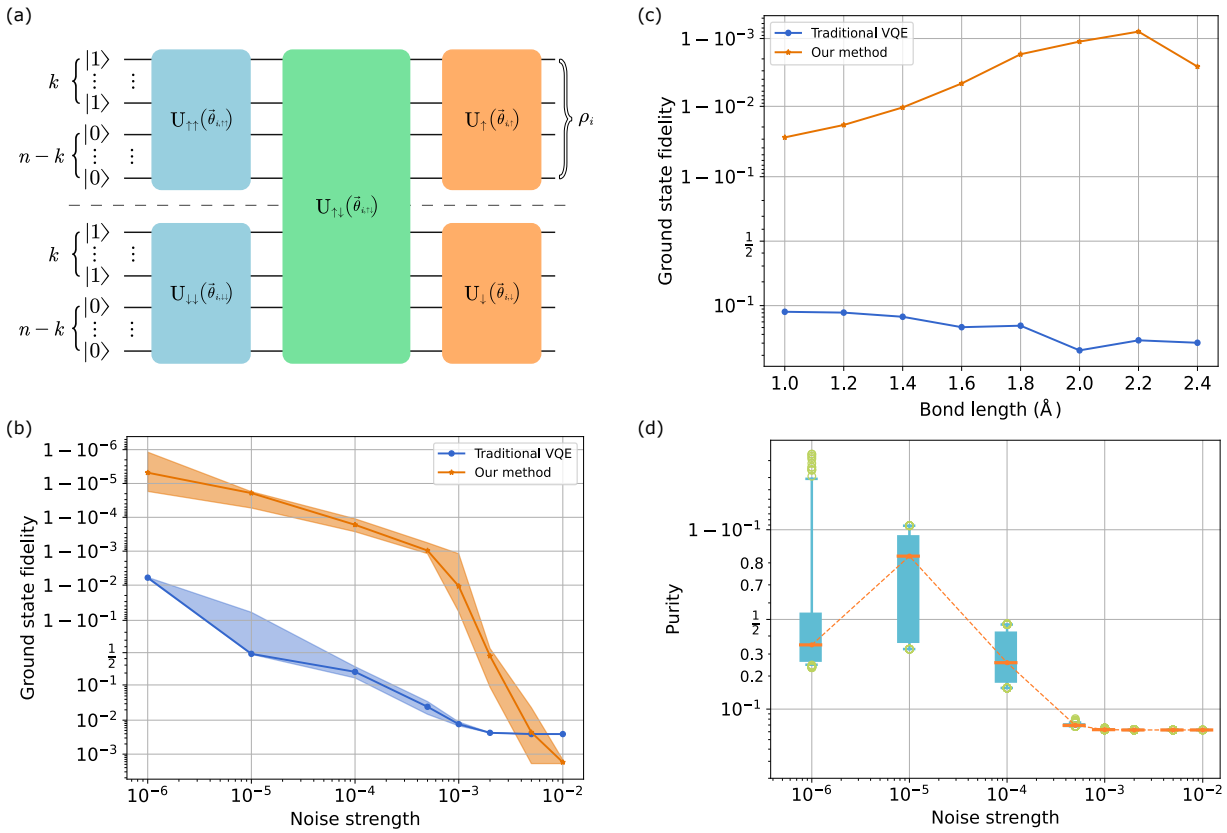


Figure 2: Solving for electronic ground states of the H_4 molecule under a spin symmetric ansatz

DMV encoding, we adopt the tools of the covering number from the statistical learning theory [25] which have been used for measuring the expressibility of ansatz in standard VQAs [26].The upper bound for the covering number of standard VQAs is given by:

$$\mathcal{N}(\mathcal{H}, \epsilon, |\cdot|) \leq \left(\frac{7N_{gt}\|H_A\|}{\epsilon}\right)^{d^{2k}N_{gt}} \tag{6}$$

For VQAs based on DMV as formulated before, the upper bound for the covering number is:

$$\mathcal{N}(\tilde{\mathcal{H}}, \epsilon, |\cdot|) \leq 2^L\mathcal{C}\left(\frac{7N_{GT}\|H_B\|}{\epsilon}\right)^{d^{2k}N_{GT}} \tag{7}$$

where $\tilde{\mathcal{H}}$ is the hypothesis space for VQAs by DMV, L is the number of CNOT gates connecting the upper and lower halves, $\mathcal{C}$ is a constant greater than 1 defined in the appendix, N_{GT} is the sum of trainable and also meaningful gates in each linear combination circuit, i.e., $N_{GT} = \sum_{i,j=1}^{K} N_{gt(ij)}$.

Taking the same observable operator, since we have $\|H_B\| = \|H_A\|$. For analytical convenience, we assume that the number of trainable gates in each combination circuit is the same as in a single standard VQA circuit, i.e., $N_{GT} = KN_{gt}$. Since this term appears in the exponential part, the upper bound for the covering number increases exponentially with the number of combinations. Additionally, the constant term $2^L\mathcal{C}$ is greater than 1, further amplifying the upper bound for the covering number. It can be seen that VQAs based on DMV have higher expressibility compared to standard VQAs.

Result 5 (numerical experiments) Additionally, we give two types of constructions of quantum circuit ansatz for preparing density matrices used for VQAs. The first is an ansatz for general purposes which can universally prepare density matrices with minimal resources .The second is how chemical-inspired ansatz specially designed for electronic structure problems of paramagnetic molecules can be well fitted into our framework. We also give numerical examples based on these ansatzes to demonstrate the performance of VQAs using our QEM protocol under various noises.For example, the numerical experiments of chemical-inspired ansatz is in fig2

References

[1] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.

[2] Ashley Montanaro. Quantum algorithms: an overview. *npj Quantum Information*, 2(1):1–8, 2016.

[3] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[4] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, et al. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020.

[5] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

[6] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.

[7] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

[8] Daniel Stilck França and Raul Garcia-Patron. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, 2021.

[9] Giacomo De Palma, Milad Marvian, Cambyse Rouzé, and Daniel Stilck França. Limitations of variational quantum algorithms: a quantum optimal transport approach. *PRX Quantum*, 4(1):010309, 2023.

[10] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature communications*, 12(1):6961, 2021.

[11] Shi-Xin Zhang, Zhou-Quan Wan, Chee-Kong Lee, Chang-Yu Hsieh, Shengyu Zhang, and Hong Yao. Variational quantum-neural hybrid eigensolver. *Physical Review Letters*, 128(12):120502, 2022.

[12] Zhong-Xia Shang, Ming-Cheng Chen, Xiao Yuan, Chao-Yang Lu, and Jian-Wei Pan. Schrödinger-heisenberg variational quantum algorithms. *Physical Review Letters*, 131(6):060406, 2023.

[13] Junxiang Huang, Wenhao He, Yukun Zhang, Yusen Wu, Bujiao Wu, and Xiao Yuan. Tensor-network-assisted variational quantum algorithm. *Physical Review A*, 108(5):052407, 2023.

[14] Kristan Temme, Sergey Bravyi, and Jay M Gambetta. Error mitigation for short-depth quantum circuits. *Physical review letters*, 119(18):180509, 2017.

[15] Suguru Endo, Simon C Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Physical Review X*, 8(3):031027, 2018.

[16] Zhong-Xia Shang, Zi-Han Chen, Ming-Cheng Chen, Chao-Yang Lu, and Jian-Wei Pan. A polynomial-time quantum algorithm for solving the ground states of a class of classically hard hamiltonians. *arXiv preprint arXiv:2401.13946*, 2024.

[17] Nobuyuki Yoshioka, Yuya O Nakagawa, Kosuke Mitarai, and Keisuke Fujii. Variational quantum algorithm for nonequilibrium steady states. *Physical Review Research*, 2(4):043289, 2020.

[18] Zhong-Xia Shang. Hermitian-preserving ansatz and variational open quantum eigensolver. *arXiv preprint arXiv:2403.03478*, 2024.

[19] William J Huggins, Joonho Lee, Unpil Baek, Bryan O'Gorman, and K Birgitta Whaley. A non-orthogonal variational quantum eigensolver. *New Journal of Physics*, 22(7):073009, 2020.

[20] Kishor Bharti and Tobias Haug. Iterative quantum-assisted eigensolver. *Physical Review A*, 104(5):L050401, 2021.

[21] Kishor Bharti and Tobias Haug. Quantum-assisted simulator. *Physical Review A*, 104(4):042418, 2021.

[22] Kosuke Mitarai and Keisuke Fujii. Methodology for replacing indirect measurements with direct measurements. *Physical Review Research*, 1(1):013006, 2019.

[23] Zhenyu Cai, Ryan Babbush, Simon C Benjamin, Suguru Endo, William J Huggins, Ying Li, Jarrod R McClean, and Thomas E O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, 95(4):045005, 2023.

[24] William J Huggins, Sam McArdle, Thomas E O'Brien, Joonho Lee, Nicholas C Rubin, Sergio Boixo, K Birgitta Whaley, Ryan Babbush, and Jarrod R McClean. Virtual distillation for quantum error mitigation. *Physical Review X*, 11(4):041036, 2021.

[25] Andrei Nikolaevich Kolmogorov and Vladimir Mikhailovich Tikhomirov. ε-entropy and ε-capacity of sets in function spaces. *Uspekhi Matematicheskikh Nauk*, 14(2):3–86, 1959.

[26] Yuxuan Du, Zhuozhuo Tu, Xiao Yuan, and Dacheng Tao. Efficient measure for the expressivity of variational quantum algorithms. *Physical Review Letters*, 128(8):080506, 2022.

Detecting Bell correlations in multipartite non-Gaussian spin states

Jiajie Guo[1] * Jordi Tura[2 3] Qiongyi He[1] Matteo Fadel[4]

[1] *State Key Laboratory for Mesoscopic Physics, School of Physics, Frontiers Science Center for Nano-optoelectronics, & Collaborative Innovation Center of Quantum Matter, Peking University, Beijing 100871, China*
[2] *$\langle aQa^L\rangle$ Applied Quantum Algorithms Leiden, The Netherlands*
[3] *Instituut-Lorentz, Universiteit Leiden, P.O. Box 9506, 2300 RA Leiden, The Netherlands*
[4] *Department of Physics, ETH Zürich, 8093 Zürich, Switzerland*

Abstract. We expand the toolbox for studying Bell correlations in multipartite systems by introducing permutationally invariant Bell inequalities (PIBIs) involving few-body correlators. First, we present around twenty families of PIBIs with up to three- or four-body correlators, that are valid for arbitrary number of particles. Compared to known inequalities, these show higher noise robustenss, or the capability to detect Bell correlations in highly non-Gaussian spin states. We then focus on finding PIBIs that are of practical experimental implementation, in the sense that the associated operators require collective spin measurements along only a few directions. To this end, we formulate this search problem as a semidefinite program that embeds the constraints required to look for PIBIs of the desired form.

Keywords: non-Gaussian spin states, Bell correlations, higher-order moments

Some correlations arising from quantum physics cannot be explained within the paradigm of local realism, and are thus called nonlocal. These are detected via the violation of a so-called Bell inequality, tested in practice through a Bell experiment. Besides their fundamental interest, nonlocal correlations are the resource enabling device-independent (DI) quantum information processing tasks, such as quantum key distribution, randomness amplification or self-testing. Although much research has focused on few-partite scenarios, mostly bipartite, nonlocal correlations also appear naturally in the multipartite regime and, in particular, in physically relevant many-body systems. With mild additional assumptions, multipartite nonlocality can be revealed in experimentally-practical ways, and take the name of Bell correlations.

Detection of Bell correlations is of great interest, as they are related to quantum critical points, metrology, open quantum systems, and bosonic systems at finite temperature, and provide an avenue to quantify DI entanglement and Bell correlation depth. However, the available inequalities are scarce, because a complete characterization is an intractable task. An approach that finds a good compromise between expressivity and complexity is to focus on Bell inequalities with particular symmetries and low-order correlators. In turn, this reduces the experimental requirements to reveal Bell correlations from them. A paradigmatic example is the use of two-body, permutationally invariant Bell inequalities (PIBIs) to detect a class of Gaussian states known as spin-squeezed states.

Despite all this progress, so far only PIBIs with up to two-body correlators are known [1]. Of particular relevance is the inequality

$$I_2 \equiv -2\mathcal{S}_0 + \frac{1}{2}\mathcal{S}_{00} - \mathcal{S}_{01} + \frac{1}{2}\mathcal{S}_{11} + 2N \geq 0, \qquad (1)$$

where $\mathcal{S}_{j_1\cdots j_k}$ are permutationally invariant (PI) observables, and $k = 1, \cdots, K$ is the order of PI correlator and $j_l = 0, 1$ is the measurement setting. I_2 enabled the experimental detection of Bell correlations in spin-squeezed BECs [2] and cold atomic ensembles [3]. But the low-order correlation pose a fundamental limit on their applicability.

*jjguo@pku.edu.cn

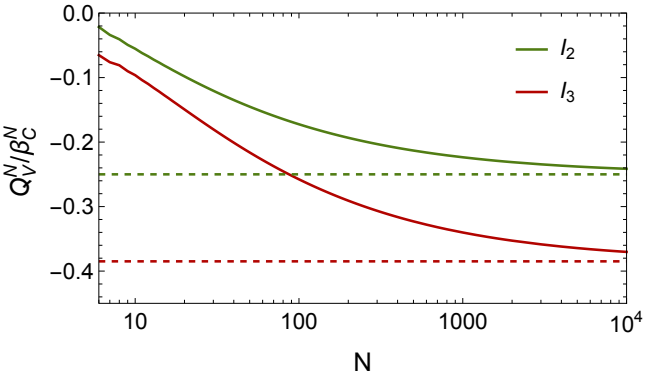


Figure 1: Maximum relative quantum violation Q_V^N/β_C^N for the 3rd-order Bell inequality I_3 and 2nd-order Bell inequality I_2, as a function of the number of parties N. The two horizontal dashed lines indicate the asymptotic violation for $N \to \infty$, which for I_2 is $-1/4$, and for I_3 is $-2\sqrt{3}/9 \approx -0.3849$.

After taking higher-order moments into consideration, in this work we present around twenty new PIBIs involving three- and four-body correlators. For example, one of the third-order PIBIs is

$$\begin{aligned} I_3 \equiv & -12(N-1)\mathcal{S}_0 - 12(N-1)\mathcal{S}_1 + 3(N-2)\mathcal{S}_{00} \\ & + 6N\mathcal{S}_{01} + 3(N-2)\mathcal{S}_{11} - 2\mathcal{S}_{000} - 3\mathcal{S}_{001} + \mathcal{S}_{111} \\ & + 12N(N-1) \geq 0, \end{aligned} \qquad (2)$$

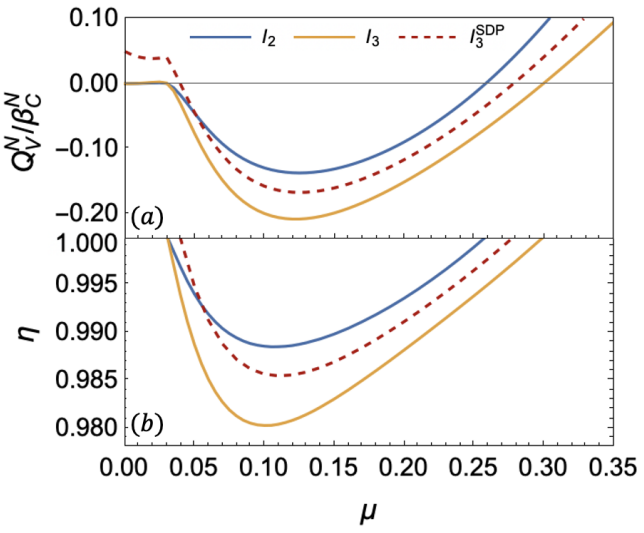


Figure 2: a): Relative quantum violation of the PIBIs I_2 (blue) and I_3 (orange) for $N = 50$ spin OAT states $|\psi(\mu)\rangle$ as a function of μ. An advantage over I_2 can also be found for $I_3^{\rm SDP}$ (red dashed), which requires to measure only one third moment of the collective spin. b): For mixed states $\rho(\eta,\mu) = \eta|\Phi(\mu)\rangle\langle\Phi(\mu)| + (1-\eta)\mathbf{I}/(N+1)$, the minimum purity η required to violate each PIBI.

and one PIBI with at most four-body correlators is

$$\begin{aligned} I_4 \equiv\ & 24(N-1)\mathcal{S}_{00} + 48(N-1)\mathcal{S}_{01} + 24(N-3)\mathcal{S}_{11} \\ & + \mathcal{S}_{0000} + 4\mathcal{S}_{0001} + 6\mathcal{S}_{0011} + 4\mathcal{S}_{0111} + \mathcal{S}_{1111} \\ & + 48N(N-1) \geq 0, \end{aligned} \tag{3}$$

which we have proven to be valid for all atom number N. Compared to known PIBIs up to second-order correlators, I_3 and I_4 provide an advantage in terms of noise robustness and sensitivity to non-Gaussian states. In Fig.1 we show Maximum relative quantum violation Q_V^N/β_C for I_2 and I_3, as a function of atom number N. It is evident the significantly better scaling for the higher-order Bell inequalities I_3 compared to I_2. In the limit $N \to \infty$, it is possible to show through a variational calculation that the relative violation of I_3 tends to $-2\sqrt{3}/9 \approx -0.3849$, which is larger than the value $-1/4$ obtained for I_2. A larger relative violation indicates a higher noise robustness, as well as the possibility to detect Bell correlations in a larger class of states.

We also show higher-order PIBIs allow us to detect Bell correlations in many-body spin states of experimental relevance. To illustrate this, let us consider the spin squeezed states $|\Phi(\mu)\rangle$ prepared through the one-axis twisting (OAT) Hamiltonian, with evolution times μ. To investigate Bell correlations in the OAT states $|\Phi(\mu)\rangle$, we compute I_2 and I_3 as a function of μ in Fig. 2 for $N = 50$, where we can observe that I_3 outperforms I_2 by reaching larger relative violation Q_V^N/β_C^N as well as detecting Bell correlations over a wider squeezing range, and thus for a larger class of states. To investigate the noise robustness, we consider OAT states mixed with white noise as $\rho(\eta,\mu) = \eta|\Phi(\mu)\rangle\langle\Phi(\mu)| + (1-\eta)\mathbf{I}/(N+1)$. In Fig. 2 we plot the minimum η for observing a PIBI violation, and show that high-order inequalities detect Bell correlations with higher noise tolerance.

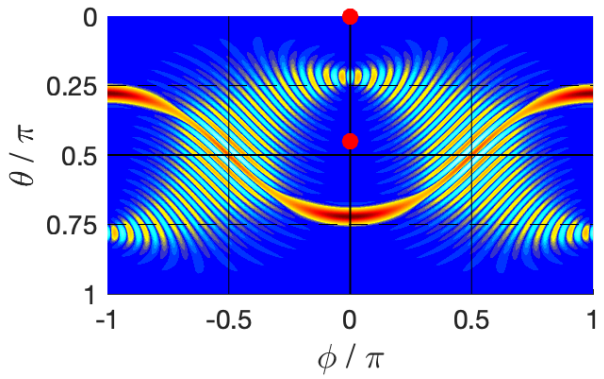


Figure 3: For $N = 50$, Wigner function of the eigenstate corresponding to the minimum eigenvalue of Bell operator $\hat{I}_4$. Red dots indicate the optimal measurement direction $\hat{S}_{\vec{n}}, \hat{S}_{\vec{m}}$, for violating I_4.

For some highly non-Gaussian states, for example, the states maximally violating inequality I_4, see Fig. 3. These states do not violate I_2, and their Bell correlations can be revealed by higher-order PIBIs.

Since higher-order moments require massive measurement statistics in practice, we want to reduce the number of higher-order measurements. For a PIBI to be experimentally practical, we note that the coefficients of the correlators must satisfy some (nonlinear) constraints. We find that these can be imposed a priori, and formulate a SDP that looks for PIBIs resulting in Bell operators of the desired form (e.g. involving only one third-moment). In Fig. 2, the red dashed line is $I_3^{\rm SDP}$ via SDP methods, it outperforms I_2. $I_3^{\rm SDP}$ has a worse noise tolerance than I_3, but is more experimentally friendly as only one direction for third-order measurements is needed. This work has been published in Physical Review Letters [4].

References

[1] J. Tura, R. Augusiak, A. B. Sainz, T. Vértesi, M. Lewenstein, A. Acín. Detecting nonlocality in many-body quantum states. *Science*, 344, 1256 (2014).

[2] R. Schmied, J. D. Bancal, B. Allard, M. Fadel, V. Scarani, P. Treutlein, and N. Sangouard. Bell correlations in a Bose-Einstein condensate. *Science*, 352, 441 (2016).

[3] N. J. Engelsen, R. Krishnakumar, O. Hosten, and M. A. Kasevich. Bell Correlations in Spin-Squeezed States of 500 000 Atoms. *Phys. Rev. Lett.* 118, 140401 (2017).

[4] J. Guo, J. Tura, Q. He, M. Fadel. Detecting Bell Correlations in Multipartite Non-Gaussian Spin States. *Phys. Rev. Lett.*, 131, 070201 (2023).

Universal readout error mitigation scheme characterized on superconducting qubits

Adrian Skasberg Aasen[1 2 *] Andras Di Giovanni[3] Hannes Rotzinger[3]
Alexey V. Ustinov[3] Martin Gärttner[2 †]

[1] *Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany*
[2] *Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany*
[3] *Institut für QuantenMaterialien und Technologien, Karlsruher Institut für Technologie, Kaiserstraße 12, 76131 Karlsruhe, Germany*

Abstract. Experiments in quantum technologies are limited by numerous sources of noise that can only be partially captured by simple analytical models and additional characterization is required. We designed an universal readout error mitigation protocol based on quantum state tomography and detector tomography. By treating readout error mitigation in the context of state tomography the method becomes largely device-, architecture-, noise source-, and quantum state-independent. We implement this method on a superconducting qubit and benchmark the protocol when varying important noise sources, such as suboptimal readout signal amplification. We observed decreases in readout infidelity by a factor of up to 30.

Keywords: Readout error mitigation, quantum state tomography, quantum detector tomography, superconducting qubit, overlapping tomography

Quantum technologies are highly dependent on accurate control and reliable readout of quantum systems. In the era of NISQ devices, there is a need for reliable operation of quantum hardware despite the noise levels present. Current experiments are limited by numerous sources of noise that can only be partially captured by simple analytical models, and additional characterization of the noise sources is required.

There are generally two approaches to dealing with noise-induced errors at the algorithmic level: *error correction* and *error mitigation* [1]. The goal of error correction is to on-the-fly correct errors that occur during a computation or simulation, by performing syndrome measurements on a logical qubit encoded onto multiple physical qubits and actively correcting them. The aim of error mitigation is more modest and tries to mitigate the effects of noise by applying post-processing onto measured data. A prominent type of noise not captured by error correction is the so-called state preparation and measurement errors, which occurs outside the standard gate formulation. Therefore, these types of noise are best handled by error mitigation methods. The focus of our work is specifically on readout errors.

Our goal is to tackle a problem that is currently not sufficiently addressed within readout error mitigation (REM), which are methods that capture a general class of readout errors, while light-weight enough to be applied to system sizes of up to 6 qubits. Such domains are very interesting when used in conjunction with methods based on quantum overlapping tomography [2, 3], which allows the construction of lower-order correlators of large quantum systems efficiently.

Our proposed method [4], illustrated in Fig. 1, consists of two stages: First, we have a calibration stage based on quantum detector tomography (QDT), which characterizes the measurement procedure in terms of generalized measurement operators, known as the positive operator-valued measure (POVM). Second, we perform the experiment of interest and apply quantum state tomography (QST) to estimate the full density matrix of the quantum system. The key insight of our procedure is that we directly integrate the information about the measurement noise into the state estimator. Thus, by treating readout error mitigation in the context of state tomography, the method becomes largely device-, architecture-, noise source-, and quantum state-independent, making it universal in the qubit domain it operates on.

Beyond the broader scope of the readout errors, the protocol comes with additional benefits which is often not enjoyed by simpler methods [5]. Most importantly we directly estimate the non-noisy state without inverting any noise channels, which often causes the resulting state to be non-physical. Furthermore, post-processing is completely separate from the experimental operation, making our method non-invasive.

The protocol relies on a notable assumption: State preparation errors must be small compared to readout errors. While this is not guaranteed for all experimental platforms, it is a reasonable assumption for superconducting qubits. It is possible to generalize our protocol by switching to more comprehensive error estimation schemes, such as gate set tomography [6], which self-consistently characterizes both readout and state preparation errors up to some gauge freedom. However, gate set tomography requires many more circuit runs, making it a prohibitively expensive calibration method.

To verify the ability of our readout error mitigation protocol, we implement it on a superconducting qubit

*aasen@kip.uni-heidelberg.de
†martin.gaerttner@uni-jena.de

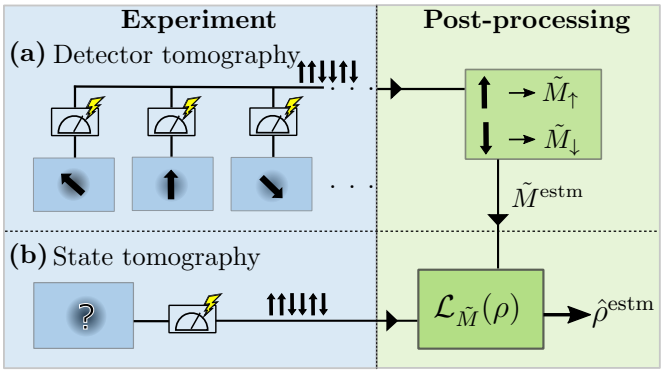


Figure 1: Protocol schematic overview. **(a)** Detector tomography: A complete set of basis states (e.g. the Pauli states) are prepared and measured repeatedly. Based on the outcomes of the measurements, a POVM is reconstructed. In a sense, it associates the measured outcome (here visualized as spin up/down measurements) with a measurement operator $\tilde{M}_i$. **(b)** State tomography: Using the reconstructed POVM, the modified likelihood function is endowed with knowledge of the operation of the measurement device. The system of interest is then prepared and measured repeatedly with the desired number of shots.

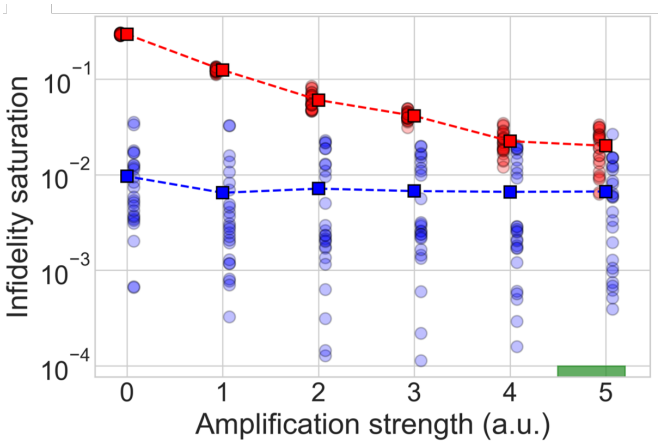


Figure 2: Benchmarking error mitigation for decreased parametric amplification. The red points are standard unmitigated QST, while the blue points are REM QST. Infidelity saturation refers to the infidelity of the reconstructed state after 240k single-shot measurements. The infidelity saturation of 20 Haar random pure states is plotted in translucent circles and shifted off center, to the left for standard QST and to the right for REM, for better visibility. The solid colored squares are the average infidelity saturation, connected by dotted lines to guide the eye. The green bar at the bottom indicates the optimal experimental parameter value.

device, and benchmark the improvement in state reconstruction fidelity due to our error mitigation scheme. We characterize the performance of the method by varying important noise sources, such as sub-optimal readout signal amplification, insufficient resonator photon population, off-resonant qubit drive, and effectively shortened T_1 and T_2 decay times. As an example, in Fig. 2, we present results for insufficient readout amplification. We see a consistent ability to mitigate any added error by lowering the amplification, even to the point of completely turning it off. Even with optimal experimental parameters we see a large improvement. Furthermore, we identified noise sources for which readout error mitigation worked well and observed decreases in readout infidelity by a factor of up to 30.

We have shown that our method greatly improves state reconstruction in the presence of readout errors, and thus reduces the quality demands on readout devices considerably. Our method adds to the toolbox of readout error mitigation schemes, and opens up new possibilities for systems with noisy readouts where accurate knowledge of the quantum state is required.

The full manuscript can be found online at arXiv:2312.04211 [4].

References

[1] Z. Cai, et.al. Quantum Error Mitigation. arXiv:2210.00921 [quant-ph], 2022.

[2] J. Cotler, F. Wilczek. Quantum Overlapping Tomography. *Phys. Rev. Lett.* 124.100401, 2020.

[3] F. Maciejewski, et.al. Modeling and mitigation of cross-talk effects in readout noise with applications to the Quantum Approximate Optimization Algorithm. *Quantum* 5, 464, 2021.

[4] A. Aasen, et.al. Universal readout error mitigation scheme characterized on superconducting qubits. arXiv:2312.04211, 2023.

[5] F. Maciejewski, et.al. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum* 4, 257, 2020.

[6] E. Nielsen, et.al. Gate Set Tomography. *Quantum* 5, 557, 2021.

An even-parity precession protocol for detecting nonclassicality and entanglement

Jinyan Chen[1] * Jackson Tiong[2] † Lin Htoo Zaw[1] ‡ Valerio Scarani[1 2] §

[1] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*
[2] *Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542*

Abstract. We introduce an even-parity precession protocol that can detect nonclassicality of some quantum states using only measurements of a uniformly-precessing variable at different points in time. Depending on the system under study, the protocol may detect the Wigner negativity of a single quantum harmonic oscillator or of a single spin $j \geq 2$; the non-Gaussian entanglement of two harmonic oscillators; or genuine multipartite entanglement of a spin ensemble, whose total spin is integer. Unlike other nonclassicality tests, simultaneous or sequential measurements are not required. Our protocol can also detect states that commute with the parity operator, which were missed by similar protocols built from Tsirelson's original precession protocol. This work also closes a long-standing gap by showing the possibility of detecting the Greenberger–Horne–Zeilinger entanglement of an even number of qubits using only collective spin measurements.

Keywords: nonclassicality, harmonic Oscillator, spin angular momentum, entanglement

1 Introduction

Uniform precession is a ubiquitous phenomenon in many physical systems, from the orbits of large celestial bodies, to the harmonic motion of mesoscopic objects in optomechanical traps, to the rotation of microscopic spins in a uniform magnetic field. As a uniformly-precessing observable has the same dynamics in both classical and quantum theory, one would not expect the simple observation of its values at different times to reveal any nonclassical signatures.

Yet, an unpublished preprint by [1] challenged this notion. He showed that if the position of a harmonic oscillator was measured at one of three random times in each round, and this was repeated over many independent rounds, the position of a quantum harmonic oscillator can be found to take a positive value more often than any state of a classical oscillator. This is an example of a mechanical task where quantum mechanics has a demonstrable advantage, in the same vein as the probability backflow of a freely-evolving particle [2] and the enhanced distance travelled by quantum projectiles [3]. Nonclassicality is therefore certified when some quantum mechanical advantage is demonstrated.

Since then, Tsirelson's original protocol has been extended to what we will call the *odd-parity precession protocol*, which certifies nonclassicality (in the form of Wigner negativity [4]) by probing a uniformly-precessing observable at one of K suitably chosen times, where K is odd [5]. In contrast to other single-system nonclassicality tests, like contextuality or Leggett–Garg inequalities [6,7], simultaneous or sequential measurements are not required. The odd-parity precession protocol has later been extended to witness entanglement. Notably, some non-Gaussian entangled states of two harmonic oscillators can be detected using only measurements of their center-of-mass position, and without the risk of false positives typical of witnesses based on uncertainty relations [8]. In spin ensembles, genuine multipartite entanglement can be detected measuring only total angular momentum. While similar witnesses of this type were known, this is the first family that detects also Greenberger–Horne–Zeilinger (GHZ) states [9].

Like any linear non-classicality witness, the odd-parity precession protocol cannot detect all states of interest (non-Gaussian, entangled). In particular, it misses all states that commute with the parity operator, as well as even-number GHZ states. In this paper, we introduce an *even*-parity precession protocol, which requires only measurements of a uniformly-precessing observable at one of K suitably chosen times, with K now even. The score is not associated to the positivity of the measurement outcome: rather, we assign an alternating positive or negative score if it falls within a region close to the origin. Nonclassicality is certified when the observed score exceeds the maximum classical score.

We show that there are states of both quantum harmonic oscillators and spin systems that violate the classical bound, some of which cannot be detected by the odd-partiy protocol. We also show that the protocol, when applied to a collective observable of a composite system, can detect entangled states of two coupled harmonic oscillators, and genuine multipartite entangled states of a spin ensemble. Our work not only contributes a new example of quantum advantage in mechanical systems, but also completely closes the problem of detecting GHZ states using collective spin measurements.

*e0944034@u.nus.edu
†tiong.jackson@gmail.com
‡zaw@l-lin.com
§valerio.scarani@gmail.com

2 The Even-parity Precession Protocol and Its Classical Bound

We shall call a physical observable $A_x(\theta)$ uniformly precessing with respect to a parameter θ if it satisfies

$$\begin{aligned} A_x(\theta) &= \cos(\theta)A_x(0) + \sin(\theta)A_y(0), \\ A_y(\theta) &= \cos(\theta)A_y(0) - \sin(\theta)A_x(0). \end{aligned} \tag{1}$$

where $A_{\hat{n}}(\theta)$ is the value of $A_{\hat{n}}$ along the angle θ in the classical case, and an operator in the Heisenberg picture in the quantum case.

The even-parity precession protocol is defined for a uniformly-precessing observable $A_x(\theta)$, a real number $\Delta \geq 0$, and an even integer $K \geq 4$, and is carried out by performing many independent rounds. In each round:

1. A value $\theta_k = \pi k/K$ for $k \in \{0, 1, \ldots, K-1\}$ is chosen;
2. $A_x(\theta_k)$ is measured at the chosen angle θ_k;
3. The score $(-1)^k$ is assigned if $|A_x(\theta_k)| \leq \Delta/2$, and the score 0 is assigned otherwise.

After many rounds, the average score is calculated as

$$s_{K,\Delta} := \frac{1}{K}\sum_{k=0}^{K-1}(-1)^k \Pr\left[|A_x(\theta_k)| \leq \frac{\Delta}{2}\right]. \tag{2}$$

In classical mechanics, the score $s^c_{K,\Delta}$ of a pure state is fully determined by its initial configuration $(A_x(0), A_y(0))$. As general classical states are probability distributions on classical pure states, this in turn implies by convexity that the classical bound is

$$\left|s^c_{K,\Delta}\right| \leq \frac{1}{K} =: \mathbf{s}^c_K, \tag{3}$$

which is independent of Δ.

3 Quantum Violations of the Classical Bound

3.1 Quantum Harmonic Oscillator

The quantum harmonic oscillator is specified by the Hamiltonian $H = P^2/(2m) + m\omega^2 X^2/2$, where the position X and momentum P are operators that satisfy the uncertainty relation $[X, P] = i\hbar\mathbb{1}$. The evolution of these observables in time is given in the Heisenberg picture as

$$\begin{aligned} X(t) &= \cos(\omega t)X + \sin(\omega t)P/(m\omega) \\ P(t) &= \cos(\omega t)P - \sin(\omega t)m\omega X. \end{aligned} \tag{4}$$

Alternatively, the observable $\sqrt{m\omega/\hbar}X(\theta/\omega)$ is the quadrature along the angle θ associated with homodyne detection in quantum optics [10]. This observable complies with Eq. (1), so the position of a harmonic oscillator, equivalently the quadrature in optical systems, is uniformly precessing with respect to θ.

Hence, we can perform the precession protocol on $X(\theta/\omega)$, and we prove the lower bound

$$K(|s^\infty_{K,\Delta}| - \mathbf{s}^c_K) \leq 2\mathcal{N}_V \tag{5}$$

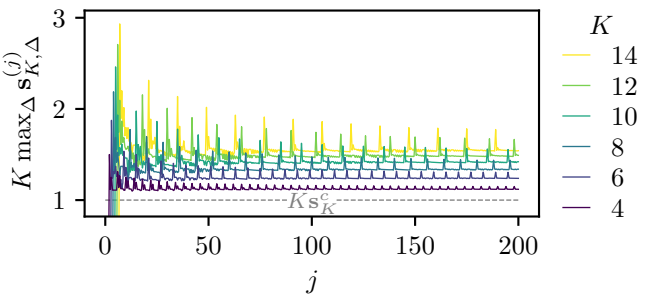


Figure 1: Maximum quantum scores $\max_\Delta \mathbf{s}^{(j)}_{K,\Delta}$ against j for $j \leq 200$. The score is zero whenever $j < K/2$, always peaks at $j = K/2$, violates the classical bound for every $j \geq K/2$, and converges to a limit as $j \to \infty$. We conjecture that the limit is $\max_\Delta \mathbf{s}^\infty_{K,\Delta}$.

on the Wigner negativity volume

$$\mathcal{N}_V := \int \mathrm{d}x \int \mathrm{d}p\, [|W(x,p)| - W(x,p)]/2\,, \tag{6}$$

which quantifies Wigner negativity as a resource [11–13].

3.2 Spin Angular Momentum

In quantum mechanics, the components of the angular momentum vector $\vec{J} = (J_x, J_y, J_z)$ satisfy the commutation relation $[J_x, J_y] = i\hbar J_z$, where $\hbar$ is the reduced Planck constant. Now, $J^{(j)}_x$ evolves under a rotation generated by the Hamiltonian $H \propto -J^{(j)}_z$, which means that we can the precession protocol on $X(J^{(j)}_x/\hbar)$. We have plotted the maximum quantum scores $\max_\Delta \mathbf{s}^{(j)}_{K,\Delta}$ against j in Fig. 1. Meanwhile, the maximum score for $j = K/2$ is analytically worked out to be

$$\mathbf{s}^{(K/2)}_{K,\Delta} = 2^{-(K-1)}\binom{K-1}{\lfloor\frac{K+\Delta}{2}\rfloor}, \tag{7}$$

The maximum value of $\mathbf{s}^{(K/2)}_{K,\Delta}$ for fixed K is given by the central binomial coefficient, which occurs when $\lfloor(K+\Delta)/2\rfloor = \lceil(K-1)/2\rceil = K/2$. And we show that $\mathbf{s}^{(j=K/2)}_{K,\Delta=0} > \mathbf{s}^c_K$ for all even $K \geq 4$. Therefore, the classical bound of the even-parity precession protocol can be violated by all quantum systems with integral spins $j \geq 2$ [5].

Finally, while we have taken the spin number j to be fixed in the above discussions, it is easily extendable the case where the protocol is performed on $\vec{J} = \oplus_{j\in\mathcal{J}}\vec{J}^{(j)}$ for some set $\mathcal{J}$ of spins.

4 Witnessing Entanglement

4.1 Witnessing Non-Gaussian Entanglement with Quadrature Measurements

A system of two identical linearly-coupled harmonic oscillators is governed by the Hamiltonian

$$H = \frac{P_+^2}{2m} + \frac{m\omega_+^2}{2}X_+^2 + \frac{P_-^2}{2m} + \frac{m\omega_-^2}{2}X_-^2, \tag{8}$$

where $X_\pm := (X_2 \pm X_1)/\sqrt{2}$, $P_\pm := (X_2 \pm X_1)/\sqrt{2}$, and $\omega_\pm := \omega \pm g/(2m)$. Here, X_+ is the center-of-mass position of the two oscillators, and its evolution in time is generated by the Hamiltonian in Eq. (8) is

$$X_+(t) = \cos(\omega_+ t)X_+ + \sin(\omega_+ t)P_+. \tag{9}$$

As such, we can perform the precession protocol on the coupled harmonic oscillators using $X_+(\theta/\omega_+)$ as the uniformly-precessing variable. We prove that $s^{\infty}_{K,\Delta} \leq \mathbf{s}^c_K + 2\mathcal{N}_V/K$, where $\mathcal{N}_V$ is the Wigner negativity volume (6) of the state $\mathrm{tr}_- \rho$.

If ρ is separable over the two oscillators, it is known that the Wigner function of $\mathrm{tr}_- \rho$ must be nonnegative by Theorem 2 of Ref. [14], which implies that $\mathcal{N}_V = 0$. However, if $\mathcal{N}_V = 0$, then $s^{\infty}_{K,\Delta} \leq \mathbf{s}^c_K$. Taking the contrapositive statement, if the even-parity precession protocol is performed on the center-of-mass position X_+, then $s^{\infty}_{K,\Delta} > \mathbf{s}^c_K$ implies that the two oscillators are entangled. Furthermore, since $2\mathcal{N}_V \geq K(s^{\infty}_{K,\Delta} - \mathbf{s}^c_K) > 0$, this criteria only detects non-Gaussian entangled states of the system with negative Wigner functions.

4.2 Witnessing Genuine Multipartite Entanglement with Collective Spin Measurements

Considering an ensemble of N particles, where the nth particle has spin j_n with angular momentum $\vec{J}^{(j_n)}$, and the total spin $\sum_{n=1}^N j_n$ is an integer, we can perform the precession protocol on the total angular momentum $J_x(\theta) = \sum_{n=1}^N J_x^{(j_n)}(\theta)$ for $K = 2\sum_{n=1}^N j_n$. And we prove that

$$|\mathrm{tr}(\rho_{\neg\mathrm{GME}} S_{K,\Delta})| \leq \frac{\mathbf{s}_{K,\Delta}}{2} =: \mathbf{s}^{K\text{-sep}}_{K,\Delta}, \tag{10}$$

where $\rho_{\neg\mathrm{GME}}$ is not *genuine multipartite entangled* state. So observing $|\mathrm{tr}(\rho S_{K,\Delta})| > \mathbf{s}^{K\text{-sep}}_{K,\Delta}$ implies that $\rho \neq \rho_{\neg\mathrm{GME}}$.

Since our witness requires only measurements of the total angular momentum, it belongs to the family of entanglement witnesses that utilize only collective observables. Past witnesses in this family have been shown to detect Dicke and many-body singlet states [15, 16], but the detection of GHZ states for more than three qubits using such a witness was not partially solved until the odd-parity precession protocol was introduced, which detects GHZ states with an odd number of qubits [9]. Our witness therefore completely closes this gap.

References

[1] Boris Tsirelson. How often is the coordinate of a harmonic oscillator positive?, 2006.

[2] A J Bracken and G F Melloy. Probability backflow and a new dimensionless quantum number. *Journal of Physics A: Mathematical and General*, 27(6):2197, mar 1994.

[3] David Trillo, Thinh P. Le, and Miguel Navascués. Quantum advantages for transportation tasks - projectiles, rockets and quantum backflow. *npj Quantum Information*, 9(1):69, Jul 2023.

[4] J. Weinbub and D. K. Ferry. Recent advances in Wigner function approaches. *Applied Physics Reviews*, 5(4):041104, 10 2018.

[5] Lin Htoo Zaw, Clive Cenxin Aw, Zakarya Lasmar, and Valerio Scarani. Detecting quantumness in uniform precessions. *Phys. Rev. A*, 106:032222, Sep 2022.

[6] Costantino Budroni, Adán Cabello, Otfried Gühne, Matthias Kleinmann, and Jan-Åke Larsson. Kochen-specker contextuality. *Rev. Mod. Phys.*, 94:045007, Dec 2022.

[7] Clive Emary, Neill Lambert, and Franco Nori. Leggett–garg inequalities. *Reports on Progress in Physics*, 77(1):016001, dec 2013.

[8] Pooja Jayachandran, Lin Htoo Zaw, and Valerio Scarani. Dynamics-based entanglement witnesses for non-gaussian states of harmonic oscillators. *Phys. Rev. Lett.*, 130:160201, Apr 2023.

[9] Khoi-Nguyen Huynh-Vu, Lin Htoo Zaw, and Valerio Scarani. Certification of genuine multipartite entanglement in spin ensembles with measurements of total angular momentum. *Phys. Rev. A*, 109:042402, Apr 2024.

[10] Leonard Mandel and Emil Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, 1995.

[11] Anatole Kenfack and Karol Życzkowski. Negativity of the Wigner function as an indicator of non-classicality. *Journal of Optics B: Quantum and Semiclassical Optics*, 6(10):396, aug 2004.

[12] Ryuji Takagi and Quntao Zhuang. Convex resource theory of non-Gaussianity. *Phys. Rev. A*, 97:062337, Jun 2018.

[13] Francesco Albarelli, Marco G. Genoni, Matteo G. A. Paris, and Alessandro Ferraro. Resource theory of quantum non-Gaussianity and Wigner negativity. *Phys. Rev. A*, 98:052350, Nov 2018.

[14] Lin Htoo Zaw. Certifiable lower bounds of wigner negativity volume and non-Gaussian entanglement with conditional displacement gates, 2024.

[15] Otfried Gühne and Géza Tóth. Entanglement detection. *Physics Reports*, 474(1):1–75, 2009.

[16] Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. Quantum metrology with nonclassical states of atomic ensembles. *Rev. Mod. Phys.*, 90:035005, Sep 2018.

Certifying entanglement dimensionality by reduction moments

Changhao Yi[1 2 3 *] Xiaodi Li[1 2 3 †] Huangjun Zhu[1 2 3 ‡]

[1] *State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai 200433, China*
[2] *Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China*
[3] *Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China*

Abstract. This paper presents a practical Schmidt number certification protocol by integrating the k-reduction map, the moment method, and the classical shadow method. Firstly, we study the spectrum of the k-reduced operators for different types of states and the corresponding k-reduction negativity, a similar notion with entanglement negativity. Secondly, we combine the moment method with the k-reduction map, and construct one series of moment criteria in the consideration of the spectrum information. Our criteria can detect more states than the fidelity-based methods, and are more practical than the correlation matrix method.

Keywords: Entanglement detection, High-dimensional entanglement, Schmidt number, k-positive map, k-reduction map, the moment method

1 Schmidt number and k-reduction map

Quantum entanglement is one of the most important and fundamental resources for many quantum technologies like quantum communication, quantum computation, quantum metrology, etc. Compared with the entanglement among qubits, high-dimensional entanglement has many advantages such as higher information capacity [1], better noise resistance [2], etc. [3]. The detection and certification of high-dimensional entanglement play the fundamental role in the applications of it to quantum technologies [4, 5].

Schmidt number or *entanglement dimensionality* characterizes the high-dimensional entanglement [6]. A pure state's Schmidt number is just its Schmidt rank; while for a mixed state $\rho_{AB} \in \mathcal{H}^A \otimes \mathcal{H}^B$, its Schmidt number is defined as

$$\mathrm{SN}(\rho_{AB}) \equiv \inf_{\mathbb{D}(\rho)} \max_{|\phi_i\rangle \in \mathbb{D}(\rho_{AB})} \mathrm{SR}(|\phi_i\rangle). \tag{1}$$

where $\mathbb{D}(\rho_{AB})$ is a pure state decomposition of ρ_{AB}. Hence, $\mathrm{SN}(\rho_{AB})$ is the minimal dimension of the subsystems of a bipartite physical system to prepare this state, which explains why it's also called entanglement dimensionality.

Currently, no method exists that is both powerful and practical for detecting high-dimensional entanglement. One popular approach, known as the *fidelity-based method* [7], relies on the fidelity condition between the state ρ_{AB} and the maximally entangled state [6]. However, this method has a significant limitation: it can only detect a restricted set of states, specifically the faithful states [8]. As a result, many common states, such as pure states with depolarized noise, may not be identified as entangled using this method.

In recent years, another method, called the *correlation matrix* or *covariance matrix method*, was proposed [9, 10]. This method, based on the condition of the 1-norm of the covariance matrix, can be conveniently realized by the randomized measurements. However, this method requires at least unitary 4-design to guarantee its performance, but there doesn't exist a simple method to realize unitary 4-design.

In this work, we propose a new method for certifying Schmidt number based on *the k-reduction map* [11]. k-reduction map is the most known example of k-positive map, defined as

$$\mathcal{R}_k(\cdot) \equiv k\mathrm{Tr}(\cdot)I - (\cdot). \tag{2}$$

There is a fundamental relationship between Schmidt number and k-reduction map [6]: given a state ρ_{AB},

$$\text{if } \mathrm{SN}(\rho_{AB}) \le k, \text{ then } (\mathcal{I}_A \otimes \mathcal{R}_k^B)(\rho_{AB}) \succeq 0. \tag{3}$$

In other words, $(\mathcal{I}_A \otimes \mathcal{R}_k^B)(\rho_{AB}) \succeq 0$ is the necessary condition for $\mathrm{SN}(\rho_{AB}) \le k$. Conversely, if the operator $(\mathcal{I}_A \otimes \mathcal{R}_k^B)(\rho_{AB}) \not\succeq 0$, then $\mathrm{SN}(\rho_{AB}) > k$. When ρ is a pure state, $(\mathcal{I}_A \otimes \mathcal{R}_k^B)(\rho) \succeq 0$ is also a sufficient condition for $\mathrm{SN}(\rho) \le k$. Condition Eq.(3), called *k-reduction condition*, is the basis of our certification protocol.

2 k-reduction negativity and k-reduction moments

The information of the spectrum of k-reduced operator $\mathcal{R}_k(\rho) \equiv (\mathcal{I}_A \otimes \mathcal{R}_k)(\rho_{AB})$ is important for our later protocol. For a pure state $|\psi\rangle$, by directly computation of the spectrum of $\mathcal{R}_k(\rho)$, we conclude that $\mathcal{R}_k(\rho)$ contains one and only one negative eigenvalue if $k < \mathrm{SN}(|\psi\rangle)$. As for a general mixed state ρ, it's easy to see that the spectrum of $\mathcal{R}_k(\rho)$ is contained in the interval $[-1, k]$ by the Weyl inequalities of matrices.

Similar with the entanglement negativity [12], we define the *k-reduction negativity* as the absolute value of the sum of negative eigenvalues of $(\mathcal{I}_A \otimes \mathcal{R}_k)(\rho_{AB})$

$$\mathcal{N}_k(\rho) \equiv \frac{1}{2}\left(\|\mathcal{R}_k(\rho)\|_1 - \mathrm{Tr}[\mathcal{R}_k(\rho)]\right) \tag{4}$$

where $(\mathcal{I}_A \otimes \mathcal{R}_k)(\rho_{AB})$ is denoted by $\mathcal{R}_k(\rho)$ for simplicity. For pure states, we can prove the following result.

*yichangh@fudan.edu.cn
†lixiaodi@fudan.edu.cn
‡zhuhuangjun@fudan.edu.cn

Theorem 1. *Suppose k is a fixed integer, $\mathcal{N}_k(\psi)$ is Schur concave for pure states $|\psi\rangle \in \mathcal{H}_{AB}$.*

When $\mathrm{SN}(|\psi\rangle) = r$, $\mathcal{N}_k(\psi) \leq 1 - k/r$, the equality is saturated when $|\psi\rangle$ is a maximally entangled state with Schmidt number r. Then we can get a further conclusion that the spectrum of $\mathcal{R}_k(\rho)$ is contained in $[\frac{k}{d} - 1, k]$ with $d = \min\{d_A, d_B\}$.

Next, we turn to the construction of the certification protocol. From the previous k-reduction condition, we understand that the key challenge is determining the positivity of $\mathcal{R}_k(\rho)$, which is generally a difficult task. Fortunately, this problem has been partially addressed in the context of entanglement detection, where some moment criteria for partial transposed states have been proposed [13, 14, 15].

Before introducing the moment criteria, we first define the n-th *k-reduction moment*

$$q_n := \mathrm{Tr}[(\mathcal{R}_k(\rho))^n], \tag{5}$$

for n being a non-negative integer. Obviously, q_n is just the sum of the n-th power of the eigenvalues of $\mathcal{R}_k(\rho)$. Given the k-reduced operator, we get a moment sequence $S_N(\rho) = (q_0, \cdots, q_N)$ for any positive integer N. Such a moment sequence $S_N(\rho)$ is called the truncated $[-1, k]$-moment sequence [16], as the spectrum of $\mathcal{R}_k(\rho)$ is contained in $[-1, k]$.[1] As a result, the problem of determining the positivity of $\mathcal{R}_k(\rho)$ is equivalent to determine if $S_N(\rho)$ is further a truncated $[0, k]$-moment sequence. The letter problem can be easily solved by consulting the theorems in [16], then we get the k-reduction moment criteria.

Theorem 2. *If $\rho \in \mathcal{H}_{AB}$ satisfies $\mathrm{SN}(\rho) \leq k$, then $B_N[\rho, k] \succeq 0$ for all positive integer N. Here $B_N[\rho, k]$ are the Hankel matrices defined as*

$$B_N[\rho, k] \equiv (q_{i+j+1})_{i,j=0}^{N}, \quad \textit{for } N \textit{ is odd}, \tag{6}$$

$$B_N[\rho, k] \equiv (kq_{i+j+1} - q_{i+j+2})_{i,j=0}^{N}, \quad \textit{for } N \textit{ is even}. \tag{7}$$

As the k-reduction condition, $B_N[\rho, k] \succeq 0$'s are necessary conditions for $\mathrm{SN}(\rho) \leq k$. Conversely, if there exists an order N such that $B_N \not\succeq 0$, then $\mathcal{R}_k(\rho) \not\succeq 0$, so we can get the conclusion that $\mathrm{SN}(\rho) > k$. For every N, $B_N[\rho, k] \succeq 0$ produces a moment condition for k-reduction moments, and the first two moment conditions are trivial, so we only consider the moment conditions for $N \geq 3$.

3 Moment estimation

Let's take the third moment condition of k-reduction moments as an example,

$$B_3 = \begin{pmatrix} q_1 & q_2 \\ q_2 & q_3 \end{pmatrix} \succeq 0, \tag{8}$$

[1] We don't use the tighter bound, because we assume the dimensions of the subsystems are unknown.

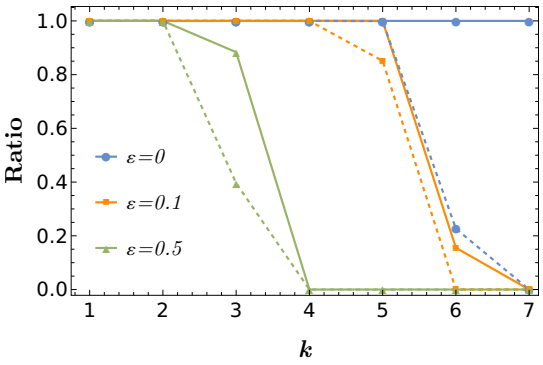


Figure 1: The depolarized Haar random pure ensembles with noise strengths $\epsilon = 0, 0.1, 0.5$ separately and $d_A = d_B = 8$. The dashed (solid) lines correspond to the correlation matrix criterion (k-reduction criterion). k is the parameter used in two criteria. Ratio is just the proportion of states in the ensemble whose Schmidt number can be certified by two criteria with parameter k.

where the explicit expressions of these three moments are

$$\begin{aligned} q_1 =& kd_B - 1 \\ q_2 =& k(kd_B - 2)\mathrm{Tr}(\rho_A^2) + \mathrm{Tr}(\rho_{AB}^2) \\ q_3 =& k^2(kd_B - 3)\mathrm{Tr}(\rho_A^3) + 3k\mathrm{Tr}[(\rho_A \otimes I_B)\rho_{AB}^2] - \mathrm{Tr}(\rho_{AB}^3). \end{aligned}$$

The above equations show that we need to estimate these terms

$$\begin{aligned} &\mathrm{Tr}(\rho_A^2), \mathrm{Tr}(\rho_{AB}^2), \\ &\mathrm{Tr}(\rho_{AB}^3), \mathrm{Tr}(\rho_A^3), \mathrm{Tr}[(\rho_A \otimes I_B)\rho_{AB}^2]. \end{aligned} \tag{9}$$

Now the problem is reduced to how to estimate these terms in Eq.(9). A popular approach is the *classical shadow* method, which can efficiently estimate the nonlinear functions of ρ and ρ_A [17]. The classical shadow method utilizes the randomized measurement outcomes to construct the classical description of a state, called the classical shadow. Taking ρ as an example, we can get these classical shadows of it, $\{\hat{\rho}_1, \cdots, \hat{\rho}_M\}$, then use these shadows to construct the estimator of $\mathrm{Tr}(\rho^2), \mathrm{Tr}(\rho^3)$ as

$$\frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \mathrm{Tr}[\hat{\rho}_i \hat{\rho}_j], \tag{10}$$

$$\frac{3!}{M(M-1)(M-2)} \sum_{1 \leq i < j < k \leq M} \mathrm{Tr}[\hat{\rho}_i \hat{\rho}_j \hat{\rho}_k]. \tag{11}$$

The estimation errors are controlled by these estimators' variances. Requiring the upper bounds of the estimation errors to be small, we can compute the sample complexity of these estimators, i.e., the order of the number of samples M. Given the unitary ensemble used in the classical shadow method is unitary 3-design, letting $M = \Omega(d_A d_B)$ is enough for estimating all these terms in Eq.(9).

4 Numerical results

k-reduction criterion Firstly, we explore the theoretical capability of the k-reduction condition Eq.(3) and

Table 1: Certification ratios by the k-reduction criterion with $\mathcal{E}_{D,K}$, $D = 16^2$.

SN= \ K=	2	3	4	5	6
2	1	1	1	1	1
3	1	1	1	1	0.9976
4	1	1	1	0.0176	0
5	1	1	0.0006	0	0
6	1	0.1596	0	0	0
7	1	0	0	0	0
8	1	0	0	0	0
9	0.0022	0	0	0	0

Table 2: Certification ratios by the covariance matrix criterion with $\mathcal{E}_{D,K}$, $D = 16^2$.

SN= \ K=	2	3	4	5	6
5	1	1	1	1	1
6	1	1	1	0	0
7	1	0	0	0	0
8	1	0	0	0	0
9	0.0034	0	0	0	0

compare it with the covariance matrix criterion given in [9, 10]. As for the fidelity-based method, the previous two criteria both performs better than it.

In the case of pure states, the sufficient and necessary condition says that the k-reduction condition can certify any pure state with Schmidt number larger than k, while there is no such guarantee that the covariance matrix criterion can certify any pure state's Schmidt number. As shown in Fig.1, we consider the ensembles of Haar random pure states with different depolarized noises, and compare the certification ratios of these two criteria. We can find that the performance of k-reduction condition is always better as long as the noise strength ε is not very large.

As for the general mixed states, k-reduction criterion doesn't have better performance than the covariance matrix criterion. Here we consider the ensembles of mixed states with induced measure $\mathcal{E}_{D,K}$, where $D = d_A d_B$ and K is the dimension of the auxiliary system. As shown in Table 1 and 2, the behaviors of k-reduction criterion are even worse for K becoming larger.

k-reduction moment criteria Secondly, we analyze the behavior of the moment criteria of k-reduction moments. From Fig.2, two key points require explanation. First, the certification ratios increase as the order N of the moment criteria increases, approaching 1 when N is sufficiently large. This is expected because the moment criterion B_N approximates the k-reduction condition more closely as N increases. Second, the certification ratios decrease as k becomes larger. This is because, for larger k, the leading term of B_N with respect to k contributes more significantly, and this leading term is

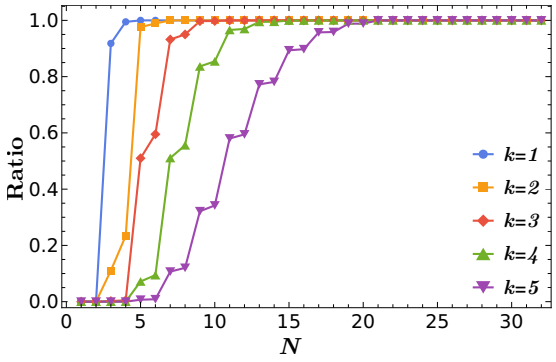


Figure 2: Certification ratios of different k-reduction moment conditions for the pure state ensemble with Schmidt rank $r = 6$ and $d = 16$.

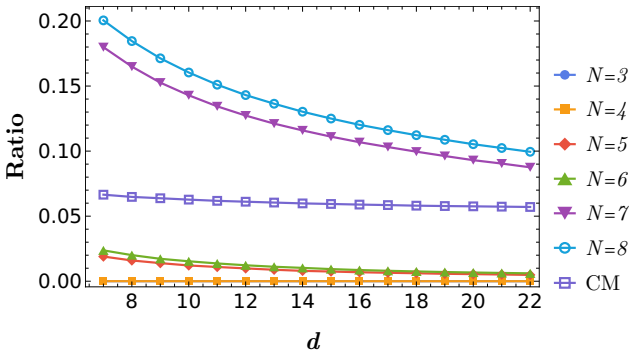


Figure 3: Comparison between the k-reduction moment criteria and the moment-based covariance matrix criterion for the pure state ensemble with $r = 6$ and $k = 5$.

approximately a weaker criterion.

Next, we compare the performances of the k-reduction moment criteria and the moment-based covariance matrix criterion. As shown in Fig.3, the certification ratios of the k-reduction moment criteria can exceed those of the moment-based covariance matrix criterion when $N \geq 7$. However, the detection ratio of the k-reduction moment criteria decreases rapidly as the local dimension d increases, whereas the detection ratio of the covariance matrix criterion varies more slowly with d.

5 Summary

Now we summarize the advantages and disadvantages of our method.

- The k-reduction condition performs significantly better than the fidelity-based method and is superior to the covariance matrix criterion for pure states or pure states with small noise.
- The moment-based covariance matrix criterion outperforms the k-reduction moment criteria at lower orders.
- The k-reduction moment criteria at lower orders can be estimated efficiently and are much more practical than the moments used in the covariance matrix method.

References

[1] P Ben Dixon, Gregory A Howland, James Schneeloch, and John C Howell. Quantum mutual information capacity for high-dimensional entangled states. *Physical review letters*, 108(14):143603, 2012.

[2] Sebastian Ecker, Frédéric Bouchard, Lukas Bulla, Florian Brandt, Oskar Kohout, Fabian Steinlechner, Robert Fickler, Mehul Malik, Yelena Guryanova, Rupert Ursin, et al. Overcoming noise in entanglement distribution. *Physical Review X*, 9(4):041042, 2019.

[3] Manuel Erhard, Mario Krenn, and Anton Zeilinger. Advances in high-dimensional quantum entanglement. *Nature Reviews Physics*, 2(7):365–381, 2020.

[4] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of modern physics*, 81(2):865, 2009.

[5] Otfried Gühne and Géza Tóth. Entanglement detection. *Physics Reports*, 474(1-6):1–75, 2009.

[6] Barbara M Terhal and Paweł Horodecki. Schmidt number for density matrices. *Physical Review A*, 61(4):040301, 2000.

[7] Jessica Bavaresco, Natalia Herrera Valencia, Claude Klöckl, Matej Pivoluska, Paul Erker, Nicolai Friis, Mehul Malik, and Marcus Huber. Measurements in two bases are sufficient for certifying high-dimensional entanglement. *Nature Physics*, 14(10):1032–1037, 2018.

[8] Mirjam Weilenmann, Benjamin Dive, David Trillo, Edgar A Aguilar, and Miguel Navascués. Entanglement detection beyond measuring fidelities. *Physical Review Letters*, 124(20):200502, 2020.

[9] Nikolai Wyderka and Andreas Ketterer. Probing the geometry of correlation matrices with randomized measurements. *PRX Quantum*, 4(2):020325, 2023.

[10] Shuheng Liu, Qiongyi He, Marcus Huber, Otfried Gühne, and Giuseppe Vitagliano. Characterizing entanglement dimensionality from randomized measurements. *PRX Quantum*, 4(2):020324, 2023.

[11] Jun Tomiyama. On the geometry of positive maps in matrix algebras. ii. *Linear algebra and its applications*, 69:169–177, 1985.

[12] G. Vidal and R. F. Werner. Computable measure of entanglement. *Phys. Rev. A*, 65:032314, Feb 2002.

[13] Andreas Elben, Richard Kueng, Hsin-Yuan Robert Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, et al. Mixed-state entanglement from local randomized measurements. *Physical Review Letters*, 125(20):200501, 2020.

[14] Antoine Neven, Jose Carrasco, Vittorio Vitale, Christian Kokail, Andreas Elben, Marcello Dalmonte, Pasquale Calabrese, Peter Zoller, Benoit Vermersch, Richard Kueng, et al. Symmetry-resolved entanglement detection using partial transpose moments. *npj Quantum Information*, 7(1):152, 2021.

[15] Xiao-Dong Yu, Satoya Imai, and Otfried Gühne. Optimal entanglement certification from moments of the partial transpose. *Phys. Rev. Lett.*, 127:060504, Aug 2021.

[16] Konrad Schmüdgen et al. *The moment problem*, volume 9. Springer, 2017.

[17] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[18] Nicolai Friis, Giuseppe Vitagliano, Mehul Malik, and Marcus Huber. Entanglement certification from theory to experiment. *Nature Reviews Physics*, 1(1):72–87, 2019.

[19] You Zhou, Pei Zeng, and Zhenhuan Liu. Single-copies estimation of entanglement negativity. *Physical Review Letters*, 125(20):200502, 2020.

[20] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P Lanyon, Peter Zoller, Rainer Blatt, and Christian F Roos. Probing rényi entanglement entropy via randomized measurements. *Science*, 364(6437):260–263, 2019.

[21] Zhenhuan Liu, Yifan Tang, Hao Dai, Pengyu Liu, Shu Chen, and Xiongfeng Ma. Detecting entanglement in quantum many-body systems via permutation moments. *Physical Review Letters*, 129(26):260501, 2022.

[22] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

[23] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 913–925, 2016.

[24] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 899–912, 2016.

[25] Andreas Elben, Steven T Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1):9–24, 2023.

[26] Steven J van Enk and Carlo WJ Beenakker. Measuring tr ρ n on single copies of ρ using random measurements. *Physical review letters*, 108(11):110503, 2012.

[27] Satoya Imai, Nikolai Wyderka, Andreas Ketterer, and Otfried Gühne. Bound entanglement from randomized measurements. *Physical Review Letters*, 126(15):150501, 2021.

[28] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 325–338, 2018.

[29] Scott Aaronson and Guy N. Rothblum. Gentle measurement of quantum states and differential privacy, 2019.

[30] Dax Enshan Koh and Sabee Grewal. Classical Shadows With Noise. *Quantum*, 6:776, August 2022.

[31] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. *PRX Quantum*, 2:030348, Sep 2021.

[32] Atithi Acharya, Siddhartha Saha, and Anirvan M. Sengupta. Shadow tomography based on informationally complete positive operator-valued measure. *Phys. Rev. A*, 104:052418, Nov 2021.

[33] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical shadow tomography with locally scrambled quantum dynamics. *Phys. Rev. Res.*, 5:023027, Apr 2023.

[34] Jonathan Kunjummen, Minh C. Tran, Daniel Carney, and Jacob M. Taylor. Shadow process tomography of quantum channels. *Phys. Rev. A*, 107:042403, Apr 2023.

[35] Christian Bertoni, Jonas Haferkamp, Marcel Hinsche, Marios Ioannou, Jens Eisert, and Hakop Pashayan. Shallow shadows: Expectation estimation using low-depth random clifford circuits, 2023.

[36] Ahmed A. Akhtar, Hong-Ye Hu, and Yi-Zhuang You. Scalable and flexible classical shadow tomography with tensor networks. *Quantum*, 7:1026, June 2023.

[37] Nathaniel Johnston, Benjamin Lovitz, and Aravindan Vijayaraghavan. Complete hierarchy of linear systems for certifying quantum entanglement of subspaces. *Physical Review A*, 106(6):062443, 2022.

[38] Anna Sanpera, Dagmar Bruß, and Maciej Lewenstein. Schmidt-number witnesses and bound entanglement. *Physical Review A*, 63(5):050301, 2001.

[39] Paul Erker, Mario Krenn, and Marcus Huber. Quantifying high dimensional entanglement with two mutually unbiased bases. *Quantum*, 1:22, 2017.

[40] Shuheng Liu, Matteo Fadel, Qiongyi He, Marcus Huber, and Giuseppe Vitagliano. Bounding entanglement dimensionality from the covariance matrix. *arXiv preprint arXiv:2208.04909*, 2022.

[41] Huangjun Zhu. Multiqubit clifford groups are unitary 3-designs. *Physical Review A*, 96(6):062336, 2017.

[42] Guillaume Aubrun, Stanisław J Szarek, and Deping Ye. Phase transitions for random states and a semicircle law for the partial transpose. *Physical Review A*, 85(3):030302, 2012.

[43] David Gross, Koenraad Audenaert, and Jens Eisert. Evenly distributed unitaries: On the structure of unitary designs. *Journal of mathematical physics*, 48(5), 2007.

[44] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

[45] Zak Webb. The clifford group forms a unitary 3-design. *arXiv preprint arXiv:1510.02769*, 2015.

[46] Man-Duen Choi. Positive linear maps on c*-algebras. *Canadian Journal of Mathematics*, 24(3):520–529, 1972.

[47] Toshiyuki Takasaki and Jun Tomiyama. On the geometry of positive maps in matrix algebras. *Mathematische Zeitschrift*, 184:101–108, 1983.

[48] Dariusz Chruściński and Gniewomir Sarbicki. Entanglement witnesses: construction, analysis and classification. *Journal of Physics A: Mathematical and Theoretical*, 47(48):483001, 2014.

[49] Jinchuan Hou, Chi-Kwong Li, Yiu-Tung Poon, Xiaofei Qi, and Nung-Sing Sze. A new criterion and a special class of k-positive maps. *Linear Algebra and its Applications*, 470:51–69, 2015.

[50] David P. DiVincenzo, Peter W. Shor, John A. Smolin, Barbara M. Terhal, and Ashish V. Thapliyal. Evidence for bound entangled states with negative partial transpose. *Phys. Rev. A*, 61:062312, May 2000.

[51] Arthur Wayne Roberts. Convex functions. In *Handbook of convex geometry*, pages 1081–1104. Elsevier, 1993.

[52] Michael A Nielsen. Conditions for a class of entanglement transformations. *Physical Review Letters*, 83(2):436, 1999.

[53] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[54] Michał Horodecki and Paweł Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Physical Review A*, 59(6):4206, 1999.

[55] https://en.wikipedia.org/wiki/Dirichlet_distribution.

[56] Pengyu Liu, Zhenhuan Liu, Shu Chen, and Xiongfeng Ma. Fundamental limitation on the detectability of entanglement. *Physical Review Letters*, 129(23):230503, 2022.

[57] Mark G. Krein, David Louvish, and A. A. Nudel'man. The markov moment problem and extremal problems. 1977.

[58] Konrad Schmüdgen. Ten lectures on the moment problem, 2020.

[59] Daniel Grier, Hakop Pashayan, and Luke Schaeffer. Sample-optimal classical shadows for pure states. *arXiv preprint arXiv:2211.11810*, 2022.

Certifying entanglement dimensionality by reduction moments

Contents

A Introduction

With the development of noisy intermediate-scale quantum (NISQ) devices, quantum systems have become increasingly controllable in laboratories. Entanglement phenomena reveal one of the most fundamental differences between quantum mechanics and classical mechanics. Thus, to demonstrate the power of quantum information technologies, the certification of entanglement is one vital step [5, 18, 13, 15, 14, 19, 20, 21].

Extracting information from a quantum state is conventionally achieved through quantum tomography of the entire state. However, the formidable challenge lies in the substantial resources demanded by quantum tomography [22, 23, 24]. Recently, *randomized measurement* [25, 26] has emerged as an alternative resource-efficient method for various tasks in quantum information, which have been used to estimate different physical quantities, such as purity [20], entanglement negativity [19] and von Neumann entropy [17, 27], etc. The technique of *classical shadow tomography* [17] effectively produces a classical description of a quantum state by integrating the ideas of randomized measurement and shadow tomography [28, 29], known as the classical shadow. This technique has many applications and inspires much research [30, 31, 32, 33, 34, 35, 36]. One particular application is the entanglement certification, where the classical shadow is used to construct the moments of the partial transposed operators to determine whether a state satisfies the *positive partial transpose* (PPT) criterion [13, 15, 14].

Entanglement dimensionality, or *Schmidt number* [6, 37], is a specific quantification of entanglement. Usually, the criteria for Schmidt number certification can find a correspondence criterion for entanglement certification. For instance, similar to the entanglement witness, *Schmidt number witness* [38] is an operator that divides the density operator space into two parts, where all states with a specific Schmidt number upper bound belong to one side. The most common Schmidt number witness is a linear combination of the identity and the maximally entangled state projector. Essentially, this witness is equivalent to the fidelity-based criterion, stating that the fidelity between a target state and the maximally entangled state cannot exceed its Schmidt number divided by the Hilbert space dimension [6, 39, 7]. However, the fidelity-based method fails to detect a large class of states, called unfaithful [8]. Even pure states with a certain depolarizing noise can be *unfaithful*. Recently, the *correlation matrix method* proposed in [27, 40, 10, 9] can resolve this problem. In these works, the authors proved that for a mixed state with certain Schmidt number upper bound, the 1-norm of its correlation matrix with re-

spect to an operator basis of the bipartite Hilbert space has an upper bound determined by the upper bound and the Hilbert space's dimension. They further use the constraint to carve out the regions of density operators with varying Schmidt numbers in the space of the second and the fourth order orthogonal moments [10, 9]. This method can detect a much broader range of quantum states compared with the fidelity-based method, but suffers from the problem of lacking of a convenient and simple unitary 4-design, for the well-known fact the Clifford group is only a unitary 3-design [41].

In this paper, we explore the capability of *k-positive map*, more specifically, the *k-reduction map*, to certify the Schmidt number of an unknown quantum state. In the problem of entanglement certification, the PPT criterion states that if a partial transposed quantum state is not positive, then it must be entangled. The k-positive map, just like partial transpose, is a special type of positive but not completely positive map, which preserves the positivity of all density operators with Schmidt number no larger than k. Similarly, if the operator obtained by applying a k-positive map to the target state is not positive, then the Schmidt number of the unknown state is larger than k. So the key step is to determine whether the new operator is positive or not. Analogous to [13, 15, 14], we will use the moment method to accomplish the task.

Our discoveries are summarized here. We provide analytical results about the spectrum of operators obtained by applying the k-reduction map on different types of states, and define the sum of the negative parts of the spectrum as the k-reduction negativity. We prove that this quantity is Schur concave for Schmidt spectrum of pure states. In terms of the application of the moment method, by taking the bounded spectrum into consideration, we complete the existing method in [15] with a new series of even-order moment conditions. We quantify the performances of our protocol and the correlation matrix protocol [9, 10] by exploring the situations where the protocol can certify the actual Schmidt number of the target state, or by computing the optimal Schmidt number lower bound certifiable by different methods. We conclude that our protocol works better than the correlation matrix protocol with sufficiently high moments for pure states and isotropic states. As for the random mixed states with induced metric [42], the performance of our protocol is also comparable to the correlation matrix protocol. Our protocol is also more competitive in consideration of unitary design [43].

The main context is organized as follows. In Section B, we review the definition of Schmidt number for mixed states and give an overview of the correlation matrix method. In Sec. C, we summarize some results about the k-positive map, and then study the spectrum of the k-reduced operators. In Sec. D, we elaborate on the moment method systematically, and then present our protocol for Schmidt number certification. In Sec.E, we describe how to estimate the reduction moments. In order to compare the performance of our protocol with the known ones, in Sec. F we analyze the detection ratios and the sample complexity of our methods. Finally, we make the conclusions and end this paper in Sec. G. A summary of main results can be found in Table 3.

B Background on Schmidt number certification criterion

B.1 Schmidt number

Consider a pure state $|\psi\rangle$ defined in a bipartite Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ with dimension $D = d_A d_B, d_A \equiv \dim(\mathcal{H}_A), d_B \equiv \dim(\mathcal{H}_B)$. Its Schmidt number is defined as

$$\mathrm{SN}(|\psi\rangle) \equiv \mathrm{rank}\left[\mathrm{Tr}_B(|\psi\rangle\langle\psi|)\right]. \tag{12}$$

We also define the *Schmidt spectrum* of $|\psi\rangle$ as the set of eigenvalues of $\mathrm{Tr}_B(|\psi\rangle\langle\psi|)$. As an instance, if the pure state $|\psi\rangle$ has Schmidt decomposition

$$|\psi\rangle = \sum_{i=0}^{r-1} \sqrt{\lambda_i}|i\rangle_A \otimes |i\rangle_B, \tag{13}$$

then its Schmidt spectrum is $\{\lambda_i\}_{i=0}^{r-1} \cup \{0\}^{d-r}$.

The convex combinations of pure states with Schmidt number no larger than r constitute a convex set, denoted as $\mathcal{S}_r$. There is an inherent hierarchy structure from the definition:

$$\mathcal{S}_1 \subset \mathcal{S}_2 \subset \cdots \subset \mathcal{S}_d = \mathcal{S}(\mathcal{H}_{AB}), \tag{14}$$

where $d = \min\{d_A, d_B\}$ and $\mathcal{S}(\mathcal{H}_{AB})$ represents the set of density operators in Hilbert space $\mathcal{H}_{AB}$. If a state ρ belongs to $\mathcal{S}_r$ but does not belong to $\mathcal{S}_{r-1}$, then it has Schmidt number r:

$$\mathrm{SN}(\rho) \equiv r, \quad \text{if } \rho \in \mathcal{S}_r/\mathcal{S}_{r-1}. \tag{15}$$

The Schmidt number is more commonly defined by generalizing the Schmidt rank for pure states through the *convex-roof construction*:

$$\mathrm{SN}(\rho) \equiv \inf_{\mathbb{D}(\rho)} \max_{\phi_i \in \mathbb{D}(\rho)} \mathrm{SN}(|\phi_i\rangle). \tag{16}$$

where $\mathbb{D}(\rho)$ is a pure state decomposition of ρ. A quantum state with Schmidt number r can be prepared from rank-r maximal entangled states, but not from states of lower rank. Hence, the Schmidt number of a mixed state ρ quantifies the minimal dimension of the Hilbert space needed for state preparation.

B.2 Method of correlation matrix

In recent years, a new method has been developed to detect the Schmidt number [27, 10, 9], which utilizes the *correlation matrix* of ρ on an operator basis of $\mathcal{H}_{AB}$. Assuming $d_A = d_B = d$, the correlation matrix of ρ is defined as

$$T_{jk} \equiv \frac{1}{d}\mathrm{Tr}\left(\rho\sigma_j^{(A)} \otimes \sigma_k^{(B)}\right), \tag{17}$$

where $\{\sigma_j^{(A)}\}$ (or $\{\sigma_j^{(B)}\}$) together with the identity forms the operator basis for the Hermitian operators on $\mathcal{H}_A$ (or $\mathcal{H}^{(B)}$), i.e.,

$$\mathrm{Tr}\left(\sigma_j^{(A)}\sigma_{j'}^{(A)}\right) = \mathrm{Tr}\left(\sigma_j^{(B)}\sigma_{j'}^{(B)}\right) = d\delta_{j,j'}. \tag{18}$$

Table 3: Comparison between different Schmidt number certification protocols for quantum states with total dimension D.

Method	Limited to faithful states	Order of unitary design	Sample complexity
Fidelity-based	Yes	/	$\Omega(1)$[7]
Correlation matrix	No	4	Unknown
Full state tomography of rank-r states	No	4	$\Omega(D^2 r)$[44]
Moments of k-reduced operator (This work)	No	3	$\Omega(D^{1/2})$ Appendix L

Suppose T has singular values $\{v_n\}_{n=0}^{d^2-2}$, its p-th Schatten norm is

$$\|T\|_p \equiv \left(\sum_{n=0}^{d^2-2} v_n^p\right)^{\frac{1}{p}} = \mathrm{Tr}\left[(TT^\top)^{\frac{p}{2}}\right]^{\frac{1}{p}}. \tag{19}$$

The correlation matrix criterion is [10, 9]:

$$\text{if } \mathrm{SN}(\rho) \le r, \text{ then } \|T\|_1 \le r - \frac{1}{d}. \tag{20}$$

It has been proved that the correlation matrices obtain by different local operator bases differ by orthogonal transformations, hence the singular values of T are invariant under local unitary transformations.

If condition Eq. (20) is violated, then we conclude that $\mathrm{SN}(\rho)$ is larger than r. However, we cannot directly estimate this 1-norm or the singular values in experiments. Instead, the second and the fourth moments of the singular values ($\|T\|_2^2, \|T\|_4^4$) can be obtained by the second and the fourth order Haar-randomized correlators $\mathcal{C}^{(n)}$ with special observables P,

$$\mathcal{C}^{(n)} \equiv \int dU_A dU_B \,\mathrm{Tr}\left[\rho U_A P U_A^\dagger \otimes U_B P U_B^\dagger\right]^n. \tag{21}$$

We can determine if the point $(\mathcal{C}^{(2)}, \mathcal{C}^{(4)})$ lies in the region

$$\mathcal{T}_r \equiv \left\{(\mathcal{C}^{(2)}, \mathcal{C}^{(4)}) : \|T\|_1 \le r - d^{-1}\right\}, \tag{22}$$

whose boundaries can be computed analytically through an optimization problem with the constraint of the correlation matrix condition Eq.(20).

The moment-based correlation matrix criterion is thus

$$\text{if } \mathrm{SN}(\rho) \le r, \text{ then } (\mathcal{C}^{(2)}, \mathcal{C}^{(4)}) \in \mathcal{T}_r. \tag{23}$$

To estimate the fourth order correlator $\mathcal{C}^{(4)}$ experimentally, we need at least unitary 4-design. However, the Clifford group fails to be unitary 4-design [41, 45].

C Method of k-reduction map

C.1 k-positive map and k-reduction map

In this paper, we will focus on the application of the *k-reduction map* to Schmidt number certification. We first introduce the notions of *k-positive map*. Let $\mathcal{L}(\mathcal{H})$ be the set of operators defined on Hilbert space $\mathcal{H}$. An operator $X \in \mathcal{L}(\mathcal{H})$ is *positive* if $\langle\psi|X|\psi\rangle \ge 0$ for all $|\psi\rangle \in \mathcal{H}$, denoted as $X \succeq 0$, while a linear map $\mathcal{M} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ is *positive* if and only if it satisfies

$$\mathcal{M}(X) \succeq 0, \quad \forall X \succeq 0. \tag{24}$$

A linear map $\mathcal{M} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H})$ is called *k-positive* if and only if the map $\mathcal{I}_k \otimes \mathcal{M} : \mathcal{L}(\mathcal{H}^k \otimes \mathcal{H}) \to \mathcal{L}(\mathcal{H}^k \otimes \mathcal{H})$ is positive, where $\mathcal{I}_k$ is the identity map for the auxiliary Hilbert space $\mathcal{H}^k$ with dimension k [46, 47, 11, 48]. The most important and common example of k-positive map is the k-reduction map [47, 11], defined as:

$$\mathcal{R}_k(X) \equiv k\,\mathrm{Tr}(X) I - X. \tag{25}$$

There are other types of k-positive map, but they are more complicated to use [49].

A linear map $\mathcal{M} : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_B)$ is k-positive map if and only if, for any pure state $|\psi\rangle \in \mathcal{H}_{AB}$ with Schmidt number at most k, $\mathcal{I}_A \otimes \mathcal{M}(|\psi\rangle\langle\psi|)$ is positive semi-definite [6, 50]. A mixed state ρ with Schmidt number no larger than k must also satisfy $\mathcal{I}_A \otimes \mathcal{M}(\rho) \succeq 0$ for any k-positive map $\mathcal{M}$, but the converse is not necessarily true. When we choose the k-positive map to be the explicit k-reduction map $\mathcal{R}_k$ (Hereafter we will use $\mathcal{R}_k(\rho)$ to denote $\mathcal{I}_A \otimes \mathcal{R}_k(\rho)$ for simplicity), we get the following results [6]:

Proposition 3. *If $\mathrm{SN}(\rho) \le k$, then*

$$\mathcal{R}_k(\rho) = k\rho_A \otimes I_B - \rho \succeq 0. \tag{26}$$

When ρ is a pure state, the condition is also sufficient.

Analogous to the violation of PPT condition, if ρ violates the above condition, we can conclude that $\mathrm{SN}(\rho) > k$. The above necessary condition is the basis of our Schmidt number certification method, so we call it the *k-reduction condition*. We also call the operators obtained by acting the k-reduced map on quantum states as the *k-reduced operators*.

C.2 Properties of the k-reduction operator

In this section, we will study basic properties of the k-reduction map, particularly the spectrum of the k-reduced operators.

We first study the explicit structure of the spectrum of $\mathcal{R}_k(|\psi\rangle\langle\psi|)$ with $\mathrm{SN}(|\psi\rangle) = r$ and $k < r$. The results are summarized in the following theorem, and its proof is in Appendix I.1:

Theorem 4. *Given a pure state $\rho = |\psi\rangle\langle\psi|$ in $\mathcal{H}_{AB}$ with*

$$|\psi\rangle = \sum_{i=0}^{r-1} \sqrt{\lambda_i}|i\rangle_A \otimes |i\rangle_B, \ \lambda_i > 0, \tag{27}$$

then the k-reduced operator $\mathcal{R}_k(\rho)$ has spectrum

$$\{x_i\} \cup \{k\lambda_i\}^{d_B-1} \cup \{0\}^{d_B(d_A-r)}$$

where $i = 0, \cdots, r-1$ and $\{x_i\}$ are the eigenvalues of

$$k\sum_{i=0}^{r-1} \lambda_i |i\rangle_A\langle i| \otimes |i\rangle_B\langle i| - |\psi\rangle\langle\psi|. \tag{28}$$

If $k \geq r$, all eigenvalues are non-negative. If $1 \leq k < r$, there exists exactly one negative eigenvalue in $\{x_i\}$.

The smallest eigenvalue of $\mathcal{R}_k(\rho)$ measures the degree to which the k-reduction criterion is violated. Thus, similar to the concept of entanglement negativity [12], we can also define the negativity of k-reduction map as

$$\begin{aligned} \mathcal{N}_k(\rho) &\equiv \frac{1}{2}\left(\|\mathcal{R}_k(\rho)\|_1 - \mathrm{Tr}[\mathcal{R}_k(\rho)]\right) \\ &= \frac{1}{2}\left(\|\mathcal{R}_k(\rho)\|_1 - kd_B + 1\right), \end{aligned} \tag{29}$$

which is the absolute value of the sum of the negative eigenvalues of $\mathcal{R}_k(\rho)$, and we call it the *k-reduction negativity*. The k-reduction criterion can thus be reinterpreted as:

$$\text{if } \mathrm{SN}(\rho) \leq k, \quad \text{then } \mathcal{N}_k(\rho) = 0. \tag{30}$$

As an instance, define the maximally entangled state with Schmidt number r as

$$|+_r\rangle \equiv \frac{1}{\sqrt{r}} \sum_{i=0}^{r-1} |i\rangle_A \otimes |i\rangle_B. \tag{31}$$

If $k < r$, $\mathcal{R}_k(|+_r\rangle\langle+_r|)$ has spectrum

$$\{k/r\}^{d_B r-1} \cup \{k/r - 1\} \cup \{0\}^{d_B(d_A-r)}, \tag{32}$$

and its k-reduction negativity is $1 - \frac{k}{r}$.

We proceed by proving an important property of the k-negativity. Given two vectors of real numbers $\mathbf{x} = (x_1, x_2, \cdots, x_r), \mathbf{y} = (y_1, y_2, \cdots, y_r)$, we rearrange the entries in the non-increasing order to obtain $\overleftarrow{\mathbf{x}} = (x'_1, x'_2, \cdots, x'_r)$ and $\overleftarrow{\mathbf{y}} = (y'_1, y'_2, \cdots, y'_r)$. If

$$\sum_{k=1}^{r'} x'_k \geq \sum_{k=1}^{r'} y'_k, \quad \forall r' \leq r, \tag{33}$$

and the equation holds when $r' = r$, then we say $\mathbf{x}$ *majorizes* $\mathbf{y}$, denoted by $\mathbf{x} \succ \mathbf{y}$. A function F is *Schur concave* [51] if the condition $\mathbf{x} \succ \mathbf{y}$ implies $F(\mathbf{x}) \leq F(\mathbf{y})$.

Considering a pure state $|\psi\rangle \in \mathcal{H}_{AB}$ with Schmidt spectrum $\boldsymbol{\lambda} = \{\lambda_i\}_{i=0}^{d-1}$ $(0 \leq \lambda_i \leq 1)$, according to Theorem 4, the k-negativity $\mathcal{N}_k(\psi)$ is sorely determined by $\boldsymbol{\lambda}$. We define a new function θ_k of the Schmidt spectrum $\boldsymbol{\lambda}$, whose value satisfies

$$\theta_k(\boldsymbol{\lambda}) \equiv \mathcal{N}_k\left(\psi\right), \tag{34}$$

where $|\psi\rangle$ is any pure state with $\boldsymbol{\lambda}$ as the Schmidt spectrum. For the benefit of proof, we define $\theta_k(\boldsymbol{\lambda})$ on region $[0, +\infty)^{\times d}/\{0\}^{\times d}$ instead. Then we give the following result, whose proof is in Appendix I.1.

Theorem 5. *For every pure state $|\psi\rangle \in \mathcal{H}_{AB}$ with Schmidt spectrum $\boldsymbol{\lambda} = \{\lambda_i\}_{i=0}^{d-1}$, we have the following results: (a) $\theta_k(\boldsymbol{\lambda})$ is Schur concave with respect to $\boldsymbol{\lambda}$ with k being fixed; (b) when $\mathrm{SN}(|\psi\rangle) = r$, $\mathcal{N}_k(\psi) \leq 1 - k/r$, the equality is saturated when $|\psi\rangle$ is a maximally entangled state with Schmidt number r; (c) the k-reduction negativity $\mathcal{N}_k(\psi)$ is non-increasing with k increasing.*

According to the Nielsen's criterion [52], if the Schmidt spectrum of $|\psi\rangle$ is majorized by that of $|\psi'\rangle$, $|\psi'\rangle$ can be produced from $|\psi\rangle$ through local operations and classical communication (LOCC). According to Theorem 5, k-negativity can be used to determine if one pure state can be transformed from another by LOCC. As a result, it is possible that the k-negativity is an entanglement monotone.

Using Theorem 5, we can also prove the following result about the eigenvalues of $\mathcal{R}_k(\rho)$, which will be useful in our protocol:

Corollary 6. *Suppose ρ is a quantum state defined on $\mathcal{H}_{AB}$, then the spectrum of $\mathcal{R}_k(\rho)$ is a subset of*

$$\left[\frac{k}{d} - 1, k\right]. \tag{35}$$

For mixed states, the spectrum of $\mathcal{R}_k(\rho)$ is difficult to analyze in general. Therefore, we will consider the pure states with depolarized noise as our next instance:

$$\rho = (1-\epsilon)|\psi\rangle\langle\psi| + \epsilon\frac{I}{d_A d_B} \tag{36}$$

with $\mathrm{SN}(|\psi\rangle) = r$. Because the k-reduction map is linear, we have

$$\mathcal{R}_k(\rho) = (1-\epsilon)\mathcal{R}_k(|\psi\rangle\langle\psi|) + \epsilon\mathcal{R}_k\left(\frac{I}{d_A d_B}\right). \tag{37}$$

The second term is merely a constant. Thus, there exists a simple linear relation between the spectrum of $\mathcal{R}_k(\rho)$ and the spectrum of $\mathcal{R}_k(|\psi\rangle\langle\psi|)$. Using this property, we can quantify the k-negativity of depolarized states as well.

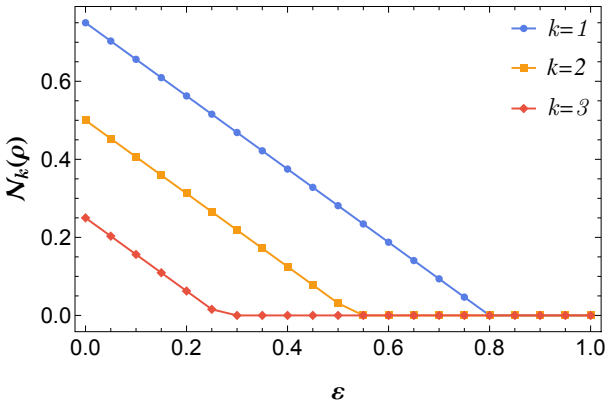


Figure 4: The k-reduction negativity $\mathcal{N}_k(\rho_{\epsilon,r})$ as a function of ϵ for $k = 1, 2, 3$. Here we set $d_A = d_B = 4, r = 4$. As ϵ approaches 1, the k-reduction negativity decreases with it. When the k-reduction negativity equals to 0, the k-reduction map cannot certify the corresponding Schmidt number lower bound any more.

Theorem 7. *Suppose ρ is a depolarized state in the form of Eq. (36), then it has k-reduction negativity*

$$\mathcal{N}_k(\rho) = \begin{cases} 0, & \epsilon \geq \epsilon^* \\ (1-\epsilon)\mathcal{N}_k(\psi) - \frac{\epsilon(d_B k - 1)}{d_A d_B}, & \epsilon < \epsilon^*, \end{cases} \tag{38}$$

with

$$\epsilon^* \equiv \frac{d_A d_B \mathcal{N}_k(\psi)}{k d_B - 1 + d_A d_B \mathcal{N}_k(\psi)}. \tag{39}$$

See Appendix I.2 for its proof. If we choose $|\psi\rangle = |+_r\rangle$ in Eq. (36), then we obtain

$$\rho_{\epsilon,r} \equiv (1-\epsilon)|+_r\rangle\langle+_r| + \epsilon \frac{I}{d_A d_B}. \tag{40}$$

We have the following result about the range of its Schmidt number

Theorem 8. *Suppose $\rho_{\epsilon,r}$ is a quantum state defined on $\mathcal{H}_{AB}$ in the form of Eq. (40). Define constant*

$$u \equiv \frac{\varepsilon}{(1-\varepsilon) d_A d_B} \tag{41}$$

for simplicity. Then its Schmidt number has lower and upper bounds:

$$\left\lceil r \frac{1+u}{1+d_B r u} \right\rceil \leq \mathrm{SN}(\rho_{\varepsilon,r}) \leq \left\lceil r \frac{1+u}{1+r^2 u} \right\rceil. \tag{42}$$

Specifically, if $\varepsilon < 0.5, r \leq \sqrt{d_A}$, then $\mathrm{SN}(\rho_{\varepsilon,r}) = r$.

See Appendix I.2 for its proof. We define $\epsilon_c^{(\mathrm{RM})}$ as the maximal value of ε that keeps $\mathcal{R}_{r-1}(\rho_{\varepsilon,r}) \not\geq 0$. According to Theorem 7, we have:

$$\epsilon_c^{(\mathrm{RM})} = \left(1 + \frac{r^2 - r}{d_A} - \frac{r}{d_A d_B}\right)^{-1}. \tag{43}$$

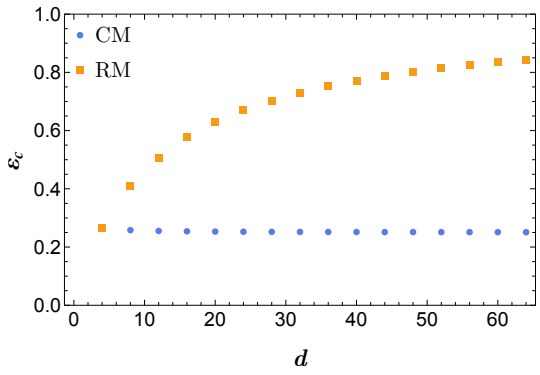


Figure 5: Comparison between $\varepsilon_c^{(\mathrm{RM})}$ and $\varepsilon_c^{(\mathrm{CM})}$ with different local dimension d. Here we choose $r = 4$, thus $\epsilon_c^{(\mathrm{CM})} = d/(4d-1)$ (see Appendix M for its proof), and $\epsilon_c^{(\mathrm{RM})}$ is given in Eq. (43). As shown by the figure, the k-reduction criterion can certify states with noise strength ϵ in a larger region than the correlation matrix criterion.

As the local dimensions d_A, d_B increase, this maximal value approaches 1, so the k-reduction criterion can certify states $\rho_{\epsilon,r}$ with any noise strength asymptotically. We also compute $\mathcal{N}_k(\rho_{\epsilon,r})$ numerically in Fig. 4.

In the last part of this section, we compare the k-reduction criterion Eq. (26) with the correlation matrix criterion Eq. (20) under different situations. From Proposition 3, we know that the Schmidt number of any pure state can be detected by the k-reduction map. However, the correlation matrix criterion Eq. (20) cannot certify all pure states. Consider the following rank-4 state

$$\sqrt{\frac{4}{5}}|00\rangle + \sqrt{\frac{1}{15}}|11\rangle + \sqrt{\frac{1}{15}}|22\rangle + \sqrt{\frac{1}{15}}|33\rangle. \tag{44}$$

with $d_A = d_B = 16$. The 1-norm of its correlation matrix is approximately 2.7231, which is smaller than $r - 1 - d^{-1} = 2.9375$. Thus, from the correlation matrix criterion, we can only conclude that its Schmidt number is at least 3 instead of 4.

Then we consider the pure state with depolarized noise $\rho_{\epsilon,r}$.

Proposition 9. *Suppose $\rho_{\epsilon,r}$ is the state in Eq. (40) with $d_A = d_B = d$, T is its correlation matrix under any operator basis, and*

$$\epsilon < \frac{1 + d^{-1}}{r} + \frac{r(r + \sqrt{r-1} + 2)}{d}. \tag{45}$$

Then we have $\|T\|_1 > r - 1 - d^{-1}$.

The proof can be found in Appendix M. Similar to $\varepsilon_c^{(\mathrm{RM})}$, we can also consider the maximal value $\epsilon_c^{(\mathrm{CM})}$ of ϵ that keeps $\|T\|_1 \geq r - 1 - d^{-1}$. When d is very large, we have $\epsilon_c^{(\mathrm{CM})} \approx r^{-1}$, which is much smaller than $\epsilon_c^{(\mathrm{RM})}$ in Eq. (43). Therefore, the correlation matrix criterion can only certify $\rho_{\epsilon,r}$ with ϵ in a narrower region than that of k-reduction criterion. An illustration of this comparison can be found in Fig. 5.

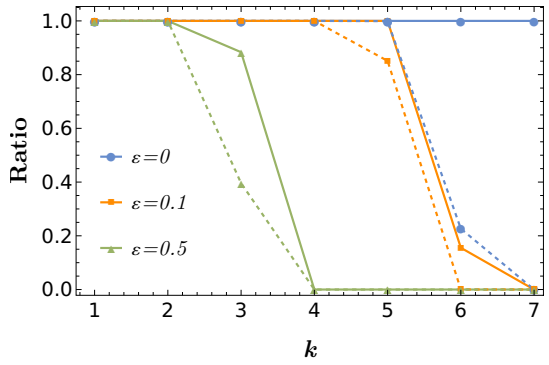


Figure 6: Comparison between the certification ratios of the two criteria for the ensembles of the depolarized Haar random pure states with $\epsilon = 0, 0.1, 0.5$ separately. Here we set $d_A = d_B = 8$. The horizontal axis represents the parameter used in the criteria. Every point is the ratio of states whose Schmidt number can be certified to be at least k by the criteria in the sampled ensemble. The dashed lines correspond to the correlation matrix criterion. The solid lines correspond to the k-reduction criterion. As shown by the figure, the performances of the k-reduction criterion are better than the correlation matrix criterion in all three situations.

Finally, we consider the Haar random pure states with depolarized noise. In Fig. 6 we numerically demonstrate that the k-reduction criterion has a higher certification ratio than the correlation matrix criterion. The certification ratio means the percentage of states certifiable by certain Schmidt number certification criterion in the ensemble of sampled states. All above discussions show that the k-reduction criterion is more powerful for the cases of pure states and pure states with depolarized noise.

D Schmidt number certification by k-reduction moments

In this section we systematically develop *the moment method*, which is designed to characterize the subset of positive operators with respect to a given set of Hermitian operators, and apply this method to the problem of certifying the lower bound of the Schmidt number of an unknown state.

D.1 PT moments and k-reduction moments

Firstly, we provide an overview of how moment methods are utilized in entanglement certification [13]. The well-known PPT criterion states that if a bipartite state ρ becomes a non-positive operator under a partial transpose operation, then ρ is necessarily entangled. We denote the partial transposed operator as $\rho^{\top_B}$.

The PPT criterion is proved to be useful for a large class of states. However, non-complete positive maps, including partial transpose and k-positive maps, are non-physical. We cannot directly implement them in laboratories. One solution to this problem is quantum tomography. After reconstructing the entire state from the outcomes of experimental measurements, we can directly test the PPT criterion by a classical computer. Unfortunately, full-state tomography suffers from the exponential dimension of the Hilbert space [22, 23, 24]. So we need to avoid using full-state quantum tomography.

Recently, researchers have found a more resource-efficient protocol [28, 29, 17]. To exploit the PPT criterion, we need to determine whether the spectrum of $\rho^{\top_B}$ does not contain negative eigenvalues. Instead of directly constructing $\rho^{\top_B}$, the positivity of $\rho^{\top_B}$ can be reflected by the partial transposed (PT) moments:

$$p_n^{\top} \equiv \mathrm{Tr}\left[(\rho^{\top_B})^n\right],\ n = 1, 2, \cdots. \tag{46}$$

The condition that employs the first n PT moments is referred to as the $p_n^{\top}$-PPT condition [13, 14, 15]. If the PT moments violate these conditions, for example, the $p_3^{\top}$-PPT condition,

$$p_3^{\top} \geq [p_2^{\top}]^2, \tag{47}$$

then we can claim that ρ is entangled. In experiments, efficient estimations of $p_n^{\top}$ can be realized using randomized measurement toolbox.

Likewise, we can design Schmidt number certification protocols using the sequence of *k-reduction moments*:

$$\begin{aligned} S_k(\rho) &\equiv (q_0, q_1, \cdots), \\ q_n &\equiv \mathrm{Tr}[(\mathcal{R}_k(\rho))^n]. \end{aligned} \tag{48}$$

According to Theorem 4, when ρ is a pure state with Schmidt spectrum $\{\lambda_i\}_{i=0}^{r-1}$,

$$q_n = \sum_{i=0}^{r-1} x_i^n + (d_B - 1)\sum_{i=0}^{r-1}(k\lambda_i)^n. \tag{49}$$

If the first few orders of $\{q_n\}$ can be estimated accurately, we can determine whether $\mathcal{R}_k(\rho) \succeq 0$ or not as well. In the following sections, we will develop our argument in detail.

D.2 The moment method

In this section, we present a few standard results of the moment method [16]. Readers with sufficient related background can directly go to Sec. D.3.

The moment problem concerns the following question: given a real sequence $S = (s_n)_{n\in\mathbb{N}_0}$ and a closed subset K, when does there exist a Radon measure μ such that $s_n = \int_K x^n d\mu(x)$ for all non-negative integer $n \in \mathbb{N}_0$? The moment methods refer to the systematic approaches for solving the moment problems. A sequence S is called a *Hamburger moment sequence* when $K = \mathbb{R}$. Likewise, if $K = [0, +\infty)$ (or $[a, b]$), the corresponding sequence is referred to as a *Stieltjes moment sequence* (or an $[a, b]$ *sequence*). Accordingly, the moment problems can be specified as the Hamburger moment problem, the Stieltjes moment problem, and the $[a, b]$-moment problem. When the sequence is a finite, the corresponding moment problems are called *truncated moment problems*. We use the following notation to represent a finite sequence generated from a truncation of an infinite sequence S:

$$\begin{aligned} S_{N_1,N_2} &\equiv (s_{N_1}, s_{N_1+1}, \cdots, s_{N_2}), \\ S_N &\equiv S_{0,N}. \end{aligned} \tag{50}$$

The arithmetic operations between sequences are defined by the corresponding operations on each entry:

$$(S \pm S')_i \equiv S_i \pm S'_i. \tag{51}$$

An important tool of the moment problem is the *Hankel matrix*. Given an integer n and a finite sequence S_{2n}, the Hankel matrix is defined as,

$$(H(S_{2n}))_{ij} \equiv s_{i+j}, \quad i,j = 0,1,\cdots,n, \tag{52}$$

which is a real symmetric matrix with dimension $(n+1)\times(n+1)$.

The Hankel matrix generated by any truncated Hamburger moment sequence must satisfy the following condition:

Lemma 10. *A real sequence S_{2n} is a truncated Hamburger moment sequence if and only if $H(S_{2n}) \succeq 0$, and*

$$(s_{n+1}, s_{n+2}, \cdots, s_{2n})^\top \in \text{range}(H(S_{2n})). \tag{53}$$

That is to say, if we know S_{2n} comes from the truncation of a Hamburger moment sequence, then we always have $H(S_{2n}) \succeq 0$.

We are particularly interested in the truncated $[a,b]$-moment problem, because in practice we can only have access to finite number of moments from experimental data, and the k-reduced operator has bounded spectrum. The standard result from the theory of moment problems provides us with the *bounded moment conditions* as a solution to this scenario.

Lemma 11 (Bounded Moment Conditions)**.** *A real sequence S_N with even N is a truncated $[a,b]$-moment sequence if and only if*

$$H(S_N) \succeq 0, \quad H(\overline{S}_{N-2}) \succeq 0, \tag{54}$$

where $\overline{S} = (\overline{s}_i)$,

$$\overline{s}_i \equiv (a+b)s_{i+1} - s_{i+2} - abs_i. \tag{55}$$

When N is odd, the conditions become

$$\begin{aligned} H(S_{1,N} - aS_{N-1}) &\succeq 0, \\ H(bS_{N-1} - S_{1,N}) &\succeq 0. \end{aligned} \tag{56}$$

Here are a few remarks:

1. S_N can be a truncated $[a,b]$-moment sequence, even when S is not a $[a,b]$-moment sequence;
2. when N is even and S is known to be a Hamburger sequence, the condition $H(S_N) \succeq 0$ is trivial according to Lemma 10. Thus, we only need $H(\overline{S}_{N-2}) \succeq 0$.

D.3 k-reduction moment conditions

We now focus on the moment sequences generated by the moments of $\mathcal{R}_k(\rho)$. To interpret the certification question as a moment problem, we first write the moments with the notation of an atomic measure. Given a k-reduced operator $\mathcal{R}_k(\rho)$ with spectrum $\{\lambda_0,\cdots,\lambda_{D-1}\}$, we can construct the following atomic measure

$$\mu_{\rho,k}(x) \equiv \sum_{i=0}^{D-1} \delta(x-\lambda_i), \tag{57}$$

Since $\text{Spec}(\mathcal{R}_k(\rho)) \subset [-1,k]$ (see Corollary 6), the sequence $S_k(\rho)$, through the atomic measure, can be considered as a $[-1,k]$-moment sequence:

$$q_n = \int_{-1}^{k} x^n d\mu_{\rho,k}(x). \tag{58}$$

Define the range of $\mathcal{R}_k$ in domain $\mathcal{S}(\mathcal{H}_{AB})$, i.e., the set of k-reduced operators, as

$$\text{Range}(\mathcal{R}_k) \equiv \left\{\mathcal{R}_k(\rho)\middle|\rho \in \mathcal{S}(\mathcal{H}_{AB})\right\},$$

and $\text{Range}^+(\mathcal{R}_k)$ is the set of positive operators in $\text{Range}(\mathcal{R}_k)$. According to the k-reduction condition,

$$\mathcal{R}_k(\mathcal{S}_k) \subset \text{Range}^+(\mathcal{R}_k) \tag{59}$$

with $\mathcal{S}_k$ defined in Eq. (14). To certify the Schmidt number of a quantum state ρ, we need to determine whether $\mathcal{R}_k(\rho) \in \text{Range}^+(\mathcal{R}_k)$ or not. So the problem is transformed into the characterization of $\text{Range}^+(\mathcal{R}_k)$ with respect to $\text{Range}(\mathcal{R}_k)$. That is, given a truncated $[-1,k]$-moment sequence $S_k(\rho)_N$, when can we conclude that it is not a truncated $[0,k]$-moment sequence? A sequence of criteria can be obtained using Lemma 11.

Introduce the Hankel matrices $\{B_N[\rho,k]\}$:

$$\begin{aligned} (B_N[\rho,k])_{ij} &\equiv q_{i+j+1}, \quad N \text{ is odd}, \\ (B_N[\rho,k])_{ij} &\equiv kq_{i+j+1} - q_{i+j+2}, \quad N \text{ is even}, \end{aligned} \tag{60}$$

where $\{q_n\}$ is defined in Eq. (48). Sometimes we also simplify $B_N[\rho,k]$ as B_N. Combining the k-reduction criterion with Lemma 10 and Lemma 11, we obtain

Theorem 12. *Suppose ρ is a state defined on $\mathcal{H}_{AB}$ and $\text{SN}(\rho) \le k$. Then for all N, $B_N[\rho,k] \succeq 0$.*

The simplest odd order condition is

$$B_3 = \begin{pmatrix} q_1 & q_2 \\ q_2 & q_3 \end{pmatrix} \succeq 0, \tag{61}$$

and the simplest even order condition is

$$B_4 = \begin{pmatrix} kq_1 - q_2 & kq_2 - q_3 \\ kq_2 - q_3 & kq_3 - q_4 \end{pmatrix} \succeq 0. \tag{62}$$

Here are a few remarks:

1. The conditions $B_1 \succeq 0, B_2 \succeq 0$ are trivial because B_1, B_2 only contain one entry, so we usually start with $N = 3$.

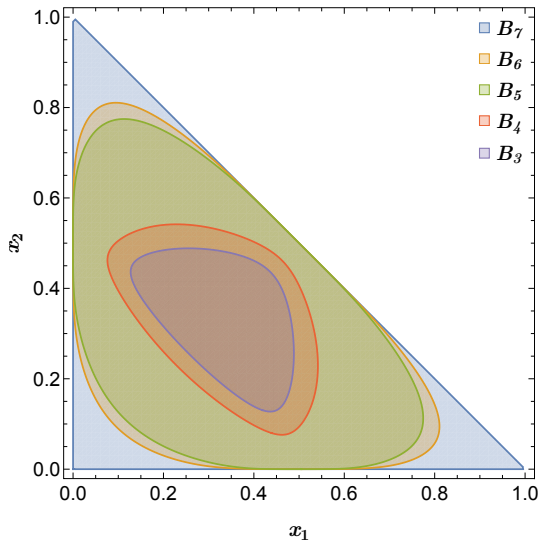


Figure 7: Plot of the detectable regions. Each point represents a state with Schmidt number 3 in the form of Eq. (64). The region of states whose Schmidt number can be detected by the N-th order moment criterion (that is, the set of states satisfying $B_N[\psi, 2] \not\succeq 0$) is illustrated in the plot. The higher the moment is, the larger the detectable region becomes. When $N = 7$, all states can be detected.

2. B_N is a submatrix of B_{N+2}, which means the N-th order moment condition $B_N \succeq 0$ is strictly not stronger than $B_{N+2} \succeq 0$.

3. If the condition is violated, i.e., $\exists N, B_N \not\succeq 0$, then we can conclude that $S_k(\rho)_N$ is not a truncated $[0, k]$-moment sequence, thus $\mathcal{R}_k(\rho) \not\succeq 0$, $\mathrm{SN}(\rho) > k$. Therefore, the N-th order moment-based criterion is

$$\text{if } B_N[\rho, k] \not\succeq 0, \quad \text{then } \mathrm{SN}(\rho) > k. \tag{63}$$

4. $B_N \succeq 0$ itself does not imply anything.

Theorem 12 provides us a series of moment-based criteria for Schmidt number certification. In Fig. 7, we use the following state to reflect the detectable regions of these moment-based conditions:

$$\sqrt{x_1}|00\rangle + \sqrt{x_2}|11\rangle + \sqrt{1 - x_1 - x_2}|22\rangle. \tag{64}$$

The set of such states whose Schmidt number can be detected by B_N is shown in the figure.

We also want to know given the dimension of the operator as D, how large N should be to guarantee that we can characterize its positivity completely. We have the following result as an answer to this question:

Corollary 13. *Suppose ρ is a quantum state with dimension D and $\mathcal{R}_k(\rho) \not\succeq 0$, then the finite moment sequence $S_k(\rho)_N$ cannot be a truncated $[0, k]$-moment sequence when $N \geq 2D$.*

The proof of the corollary is in Appendix J.3.

D.4 Certification by the third moment criterion

In the section, we formally introduce our protocol, and use the $N = 3$ moment criterion as an instance. Using the results in Sec. D, we know that if a quantum state ρ has Schmidt number at most k, then it must remain positive under the k-reduction map; thus, $\{B_N\}$, the Hankel matrices composed by the first few moments of the k-reduced operator, must all be positive semi-definite.

In linear algebra, $B_N \succeq 0$ is equivalent to the condition that the determinants of all *principal minors* of B_N are non-negative [53]. A principal minor of a matrix is the determinant of a submatrix obtained by deleting pairs of column and row whose joint entry is diagonal. The determinant of the entire matrix is also included in the principal minors. However, for the case of $N = 3$, because $q_1 = kd_B - 1 > 0$, the principle minors of the Hankel matrix other than the full determinant are trivial (if $q_3 < 0$, then the full determinant must be negative, thus it is included in the case where the full determinant is negative). We only need to test the determinant of the entire Hankel matrix. Therefore, if $\mathrm{SN}(\rho) \leq k$, then $\det(B_3[\rho, k]) \geq 0$. Conversely,

$$\text{if } \det(B_3[\rho, k]) < 0, \quad \text{then } \mathrm{SN}(\rho) > k. \tag{65}$$

In our protocol, we certify the lower bound of $\mathrm{SN}(\rho)$ based on this criterion, which can be simplified by introducing a function:

$$F_{\mathcal{R}}(\rho, k) \equiv \det(B_3[\rho, k]), \tag{66}$$

such that if $F_{\mathcal{R}}(\rho, k) < 0$, then $\mathrm{SN}(\rho) > k$.

Notice that $F_{\mathcal{R}}(\rho, k)$ is a polynomial of k:

$$F_{\mathcal{R}}(\rho, k) = \alpha_4 k^4 + \alpha_3 k^3 + \alpha_2 k^2 + \alpha_1 k + \alpha_0. \tag{67}$$

Define $p_n \equiv \mathrm{Tr}(\rho^n), a_n \equiv \mathrm{Tr}(\rho_A^n)$, and further define $p_{1,2} \equiv \mathrm{Tr}\big(\rho_A \otimes I_B \rho^2\big)$. Then the coefficients of $F_{\mathcal{R}}(\rho, k)$ are as follows:

$$\begin{aligned}
\alpha_4 &= d_B^2(a_3 - a_2^2), \\
\alpha_3 &= -4d_B(a_3 - a_2^2), \\
\alpha_2 &= d_B(3p_{1,2} - 2a_2 p_2) - 4a_2^2 + 3a_3, \\
\alpha_1 &= 4a_2 p_2 - d_B p_3 - 3p_{1,2}, \\
\alpha_0 &= p_3 - p_2^2.
\end{aligned} \tag{68}$$

Using this condition, we obtain an algorithm that can certify whether the Schmidt number of the target is smaller than a specific value or not. See Algorithm 1. The complete algorithm, which uses a general moment order N, is detailed in Algorithm 2 and includes Algorithm 1 as a subroutine.

Here are a few comments for Algorithm 1. First, we have to set truncation values for k, N because we cannot apply the k-reduction moment condition with infinite order. The common truncation value for k is d, because the largest Schmidt number is d. Second, the output is always a lower bound of the Schmidt number.

Algorithm 1 Certification of Schmidt number lower bound

Require: Target state ρ, maximal order N^*, expected Schmidt number r_{est}.
Ensure: Return whether $\text{SN}(\rho) \geq r_{\text{est}}$ or not.
1: **for** $N = 3, 4, \cdots, N^*$ **do**
2: Set $k = r_{\text{est}} - 1$. Estimate the moments $q_1, q_2, \cdots, q_N$ of $\mathcal{R}_k(\rho)$.
3: Construct the Hankel matrix B_N.
4: If B_N is not positive semi-definite, return yes (1).
5: **end for**
6: Return no (0).

Algorithm 2 Optimal certification of Schmidt number

Require: Taget state ρ, maximal order N^*, expected Schmidt number upper bound r_{upper}.
Ensure: If $\text{SN}(\rho) < r_{\text{upper}}$, then the output is $\text{SN}(\rho)$.
1: $s_{\text{temp}} = 1$.
2: **for** $k = 1, 2, \cdots, r_{\text{upper}}$ **do**
3: Input (ρ, N^*, k) to Algorithm 2.
4: **if** ouput equals 0, **then**
5: Return s_{temp}.
6: **else** $s_{\text{temp}} = k$.
7: **end if**
8: **end for**
9: Return s_{temp}.

E Moments estimation

E.1 Moment estimation by randomized measurements

For the purpose of Schmidt number certification, the entries of the Hankel matrices to be estimated are q_n or $kq_n - q_{n+1}$ with q_n introduced in Eq. (48). After we expand $\mathcal{R}_k(\rho)^n$, the polynomial is in the form of

$$\text{Tr}\left[\text{poly}(\rho, \rho_A \otimes \mathbb{I}_B)\right]. \tag{69}$$

Quantities of this form can be estimated in experiments by randomized measurement protocols, such as the *classical shadow method* [17]. There are two steps in classical shadow tomography. The first step is data acquisition. Given a unitary ensemble $\mathcal{U}$, we randomly sample an element U from it and apply it on the state ρ, then measure the rotated state in the standard basis to obtain $|\hat{s}\rangle$. Based on U and $|\hat{s}\rangle$, we can create the classical shadow of the target state:

$$\hat{\rho} = \mathcal{M}_{cs}^{-1}\left(U^\dagger|\hat{s}\rangle\langle\hat{s}|U\right). \tag{70}$$

Here $\mathcal{M}_{cs}$ is the measurement channel, whose expression depends on the specific ensemble $\mathcal{U}$. For global Clifford group ensemble, the classical shadow is

$$\hat{\rho} = (D+1)U|\hat{s}\rangle\langle\hat{s}|U^\dagger - I, \tag{71}$$

where D is the dimension of ρ. The second step involves predicting the expectation values $\text{Tr}(O_a\rho)$ of the observables by constructing estimators based on the empirical mean

$$\hat{o}_a = \frac{1}{M}\sum_{m=1}^{M}\text{Tr}(\hat{\rho}_m O_a), \tag{72}$$

or the median-of-means. The variance of these estimators is controlled by the *shadow norm*:

$$\left\|O_a - \frac{\text{Tr}(O_a)}{D}\right\|_{\text{shadow}}^2. \tag{73}$$

The formal definition can be found in [17]. Additionally, classical shadows can be employed to estimate quadratic functions of ρ, such as $o_a = \text{Tr}(O_a\rho\otimes\rho)$, using formula

$$\hat{o}_a = \frac{2}{M(M-1)}\sum_{m<n}\text{Tr}(O_a\cdot\hat{\rho}_m\otimes\hat{\rho}_n). \tag{74}$$

As an instance, given a sequence of state estimator $\{\hat{\rho}_m\}$, the purity $\text{Tr}\left(\rho^2\right)$ can be estimated by

$$\begin{aligned}&\frac{1}{M(M-1)}\sum_{m\neq n}\text{Tr}\left[\mathbb{W}_{(12)}\cdot\hat{\rho}_m\otimes\hat{\rho}_n\right]\\ =&\frac{1}{M(M-1)}\sum_{m\neq n}^{N}\text{Tr}(\hat{\rho}_m\hat{\rho}_n),\end{aligned} \tag{75}$$

where $\mathbb{W}_{(12)}$ is the swap gate between two Hilbert spaces. Other non-linear functions can be estimated in a similar formalism.

Statistical correlation method is another randomized measurement protocol that can estimate moments of a quantum state [20, 19]. The advantage of statistical correlation is we do not need to record the information of random unitaries for each measurement, while its applicable scope is much narrower than the classical shadow methods.

The performance of a randomized measurement protocol is assessed based on two primary factors: the order of unitary design and the sample complexity. Ideally, we want to sample unitaries from a Haar random ensemble, which is not realistic for large systems. Instead, we sample unitaries from a t-design to approximate a uniformly random sample. A set of unitary operators $\mathcal{U}$ is called a *unitary t-design* [43] if for all operators O,

$$\frac{1}{|\mathcal{U}|}\sum_{U\in\mathcal{U}}U^{\otimes t}O(U^\dagger)^{\otimes t} = \int d\mu U U^{\otimes t}O(U^\dagger)^{\otimes t}. \tag{76}$$

The ensemble $\mathcal{U}$ can thus be used to replace the Haar random ensemble if we only need a random unitary twirling channel with order no larger than t. High order unitary designs with $t > 4$ are hard to construct for qubit systems. Thus, it is advantageous for a randomized measurement to require a low order of unitary design.

Randomized measurement protocols are essentially based on concentration laws of statistics. The sample complexity means how many samples we need to guarantee the convergence of the statistical outcomes. It is quantified by the variance of the target estimator. Certainly, the smaller the sample complexity (variance of the

estimator) is, the better the protocol is. Because the estimators of the moments are non-linear functions of ρ, the sample complexity of classical shadow method is usually related to the dimension of the system [13].

We prove the following results about the sample complexity of Algorithm 2 with $N^* = 3$ in Appendix L.

Theorem 14. *Suppose ρ is a quantum state with dimension D, and we are given access to a unitary-3 design. Then the sample complexity of estimating p_2, p_3 and $\mathrm{Tr}(\rho_A \otimes I_B \cdot \rho^2)$ up to accuracy $\mathcal{O}(1)$ using randomized measurements is $\Omega(D^{1/2})$.*

E.2 Moment estimation by permutation tests

A simple quantum computer that enables permutation test can reduce the sample complexity of moment estimation to $\Omega(1)$. Given a sequence of states $\rho_1, \rho_2, \cdots, \rho_N$, add an ancillary qubit $|+\rangle$ and perform a controlled-full-permutation $\mathbb{W}_\pi$:

$$\begin{aligned}&\frac{|0\rangle\langle 0|}{2} \otimes \bigotimes_{n=1}^{N} \rho_n + \frac{|1\rangle\langle 1|}{2} \otimes \mathbb{W}_\pi \bigotimes_{n=1}^{N} \rho_n \mathbb{W}_\pi^{-1}, \\ +&\frac{|0\rangle\langle 1|}{2} \otimes \bigotimes_{n=1}^{N} \rho_n \mathbb{W}_\pi^{-1} + \frac{|1\rangle\langle 0|}{2} \otimes \mathbb{W}_\pi \bigotimes_{n=1}^{N} \rho_n.\end{aligned} \tag{77}$$

The ancilla qubit becomes:

$$\frac{I}{2} + \frac{1}{2}\mathrm{Tr}\left(\prod_{n=1}^{N} \rho_n\right) X. \tag{78}$$

Measure it in the Pauli X basis, the probability of measuring $|+\rangle$ state is

$$P_+ = \frac{1}{2} + \frac{1}{2}\mathrm{Tr}\left(\prod_{n=1}^{N} \rho_n\right), \tag{79}$$

and the probability of measuring $|-\rangle$ state is

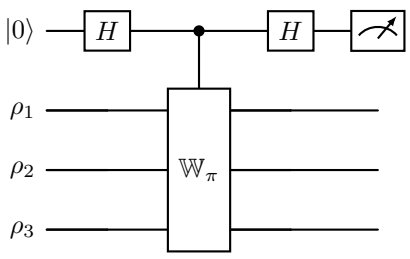
$$P_- = \frac{1}{2} - \frac{1}{2}\mathrm{Tr}\left(\prod_{n=1}^{N} \rho_n\right). \tag{80}$$

Thus, if we associate the $|+\rangle$ state with $+1$, and $|-\rangle$ with -1, the expectation value of the measurement outcome is

$$P_+ - P_- = \mathrm{Tr}\left(\prod_{n=1}^{N} \rho_n\right). \tag{81}$$

All the necessary parameters in the moment-based criteria can be written in this form, and we only need $\Omega(\epsilon^{-2})$ runs of tests to obtain error bound ϵ.

An example with $N = 3$ is demonstrated here:



F Performance

In this section, we analyze the detectability of various types of quantum states and their performance under different certification protocols. The protocols considered are Eq. (63) and Eq. (23).

Denote the certification protocol by a function $F(\rho, k)$. One example is represented in Eq.(67). Let $\mathcal{E}$ be an ensemble of quantum states, then the probability of certifying Schmidt number with lower bound k in this ensemble is,

$$\Pr\left\{F(\rho, k) < 0 \;\middle|\; \rho \in \mathcal{E}\right\}, \tag{82}$$

which is referred to as the *detection ratio* in the following paragraphs. If a state with Schmidt number r satisfies $F(\rho, r-1) < 0$, then we say the criterion can detect the state for simplicity. The higher the detection ratio is, the better the performance is for the ensemble $\mathcal{E}$.

Certainly, depending on the choice of $\mathcal{E}$, the detection ratios of the same criterion can be different. In the next section, we will look into three concrete examples: the *isotropic states*, the pure state ensemble and the induced metric state ensemble.

F.1 Isotropic states

The isotropic state is defined as

$$\rho_F \equiv \frac{1-F}{d^2-1} I + \frac{d^2F-1}{d^2-1} |+_d\rangle\langle +_d|. \tag{83}$$

An isotropic state can be constructed by averaging over the set of Haar random unitaries. For example, given integer r, the isotropic state with $F = r/d$ can be written as

$$\int dU U \otimes U^* |+_r\rangle\langle +_r| U^\dagger \otimes (U^*)^\dagger, \tag{84}$$

which is the convex combinations of infinite numbers of maximally entangled states with Schmidt number r [54]. It is natural to infer that this mixed state has Schmidt number r. In fact, for a general F, it has been proved that $\mathrm{SN}(\rho_F) = \lceil dF \rceil$, and the k-reduction map can detect the Schmidt number of all isotropic states [6]. We find that the lowest order moment criterion is enough to certify all isotropic states as well:

Proposition 15. *Suppose ρ_F is an isotropic state with $F > d^{-1}$ and $1 \le k \le \lceil dF \rceil - 1$. Then $B_3[\rho_F, k] \not\succeq 0$.*

The proof is in Appendix K. The correlation matrix criterion Eq. (20) can detect the Schmidt number of an isotropic state as well.

Proposition 16. *Suppose ρ_F is an isotropic state defined on $\mathcal{H}_{AB}$ with $d_A = d_B = d$, and T is its correlation matrix under any operator basis. Then the singular values of T are*

$$\frac{d^2F-1}{d(d^2-1)} \tag{85}$$

with multiplicity d^2-1. Therefore,

$$\|T\|_1 = dF - d^{-1}. \tag{86}$$

See Appendix M for its proof. However, for a sequence with completely even distribution, it is impossible to properly bound the 1-norm based on low order of moments. Suppose the singular values of T are all s with multiplicity D, thus

$$\|T\|_2^2 = Ds^2, \quad \|T\|_4^4 = Ds^4. \tag{87}$$

We can construct another rank-2 correlation matrix T' with singular values $\{\sigma_0, \sigma_1\}$ that shares the same moments:

$$\sigma_0^2 + \sigma_1^2 = Ds^2, \quad \sigma_0^4 + \sigma_1^4 = Ds^4. \tag{88}$$

The 1-norm of this matrix is

$$\|T'\|_1 = \sigma_0 + \sigma_1 = \sqrt{D + \sqrt{2(D^2 - D)}}s. \tag{89}$$

Based on the second and the fourth moment, one cannot rule out this possibility. To conclude that $\mathrm{SN}(\rho_F) > \lceil dF \rceil - 1$, we need at least

$$\|T\|_1 > dF - d^{-1} - 1 = Ds - 1, \tag{90}$$

which is much larger than $\|T'\|_1$. Hence, the second and the fourth moments of correlation matrix method are not enough to accurately certify the Schmidt number of isotropic states.

F.2 Ensemble of pure states, analytical results

Define the target ensemble as

$$\mathcal{E}_0 \equiv \left\{ |\psi\rangle = \sum_{i=0}^{r-1} \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_B \right\}, \tag{91}$$

where $\{|i\rangle_A\}, \{|i\rangle_B\}$ are elements of two orthogonal bases of $\mathcal{H}_A$ and $\mathcal{H}_B$ separately, and $\{\lambda_i\}$ is sampled from the uniform distribution of $(r-1)$-simplex [55]:

$$\left\{ (\lambda_0, \lambda_1, \cdots, \lambda_{r-1}) : \sum_{i=0}^{r-1} \lambda_i = 1, \lambda_i \geq 0 \right\}, \tag{92}$$

which is also called the *Dirichlet distribution*. We do not need to apply local random unitaries to $|i\rangle_A$ and $|i\rangle_B$ because the k-reduction moment criterion and the correlation matrix criterion are both invariant under local unitaries.

We first demonstrate the efficiency of our algorithm by showing that the k-reduction moment criteria can always detect $|+_r\rangle$:

Proposition 17. *Given $|+_r\rangle$ defined in Eq. (31) and $1 \leq k < r$, we have $B_N[+_r, k] \not\succeq 0$.*

See Appendix K for its proof. However, for a general pure state, the local dimension d_B plays an important role in the detectability performances of k-reduction moment conditions. This is because the N-th order moments q_n can be written as a polynomial of d_B. Thus, the performance of the criterion depends strongly on the local dimension. In Appendix K, we prove the following result about the asymptotic behavior of the odd order k-reduction moment criteria. The proof for the even order situation is similar.

Proposition 18. *Suppose $|\psi\rangle$ is a pure state define on $\mathcal{H}_{AB}$ with non-degenerate Schmidt spectrum $\{\lambda_i\}_{i=0}^{r-1}$ and N is an odd number. If $\frac{N+1}{2} \geq 2r$, then the N-th order moment condition is able to detect the Schmidt number r as*

$$B_N[\psi, r-1] \not\succeq 0. \tag{93}$$

If $\frac{N+1}{2} \leq r$ and the local dimension d_B is large enough in the sense that

$$d_B > 1 + \frac{(N+1)\mathcal{N}_{r-1}(\psi)}{2\lambda_{\min}(H_1)}, \tag{94}$$

where $H_1 \equiv H(S_{1,N}^{(1)})$, $S^{(1)} = (s_n^{(1)})$,

$$s_n^{(1)} \equiv \sum_{i=0}^{r-1} [(r-1)\lambda_i]^n, \tag{95}$$

then the N-th order moment condition fails to detect the Schmidt number r as:

$$B_N[\psi, r-1] \succeq 0. \tag{96}$$

In another word, if the target state has a large dimension and the Schmidt spectrum of $|\psi\rangle$ is non-degenerate, then the criteria almost always fail when $N \leq 2r-1$, and always succeed when $N \geq 4r-1$. Therefore, because the samples from the Dirichlet distribution satisfies $\forall i \neq j, \lambda_i \neq \lambda_j$ almost surely, the detect ratio of the k-reduced moment criterion with a small value of N tends to 0 as the system dimension gets large.

Notice that maximally entangled states with Schmidt number r have degenerate Schmidt spectrum, thus do not satisfy the precondition in Proposition 18.

As a comparison, although the correlation matrix criterion (Eq. (20)) cannot detect all pure states, it exhibits a weaker dependence on dimension. More specifically,

Proposition 19. *Suppose $|\psi\rangle$ is a pure state defined on $\mathcal{H}_{AB}$ with $d_A = d_B = d$ and its Schmidt spectrum is $\{\lambda_i\}_{i=0}^{r-1}$, T is its correlation matrix under any operator basis. Then the non-zero singular values of T are*

$$\{\sqrt{\lambda_i \lambda_j} + \eta_{ij}\}_{i,j=0}^{r-1}, \tag{97}$$

with $|\eta_{ij}| < (r + 2\sqrt{r-1} + 2)/d$.

The proof is in Appendix M. Therefore, for a pure state with Schmidt spectrum $\{\lambda_i\}_{i=0}^{r-1}$, when the local dimension is very larger, $\|T\|_1 > r - 1 - d^{-1}$ is close to the condition of

$$\sum_{i=0}^{r-1} \sqrt{\lambda_i} \geq \sqrt{r-1}. \tag{98}$$

Hence, the probability of certifying $\mathrm{SN}(\psi) \geq r$ with correlation matrix criterion almost only depends on the Schmidt spectrum. The moment-based protocol has this property as well. For large systems, we have approximations:

$$\begin{aligned} \|T\|_2^2 &\approx \sum_{i=0}^{r-1} \lambda_i = 1, \\ \|T\|_4^4 &\approx \sum_{i=0}^{r-1} \lambda_i^2. \end{aligned} \tag{99}$$

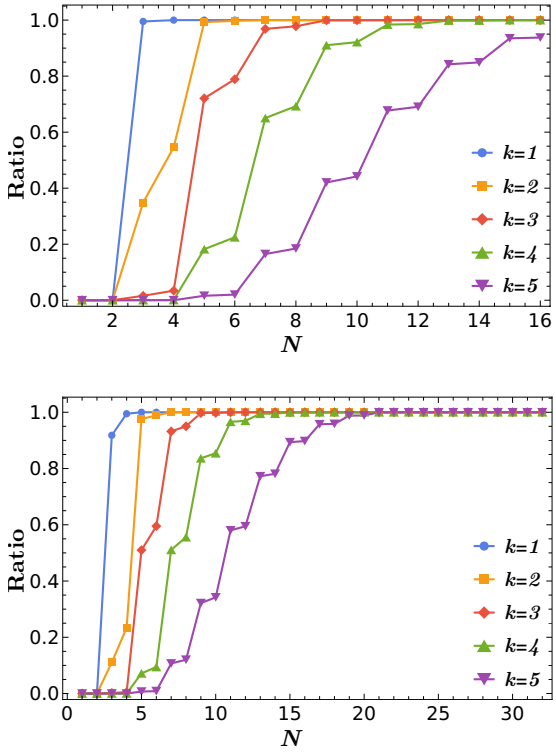


Figure 8: Detection ratios of different k-reduction moment conditions (Eq. (63)) for pure state ensemble $\mathcal{E}_0$ (Eq. (91)). The horizontal line is the moment order N. The pure state ensemble $\mathcal{E}_0$ has Schmidt rank $r = 6$. The first figure has local dimension $d = 8$, the second figure has local dimension $d = 16$. As shown by the figure, the detection ratio increases monotonically with the order of the criterion and decreases with k. When the order is high enough, eventually the detection ratios approach 1.

Therefore, the forth moment of the correlation matrix can detect the Schmidt number of a pure state when

$$\sum_{i=0}^{r-1} \lambda_i^2 \leq \frac{1}{r-1} + \mathcal{O}(d^{-1}). \tag{100}$$

The condition depends weakly on d.

F.3 Ensemble of pure states, numerical results

Here are a few numerical results about the detection ratios of pure state ensembles $\mathcal{E}_0$ mentioned in Eq. (91). In Fig. 8, we test the k-certification ratios of pure state ensemble for different orders of bounded k-reduction moment criteria (see Corollary 11). The horizontal line represents the orders of the criteria, the vertical line represents the corresponding detection ratios for different k. As demonstrated by the figure, the detection ratio increases with the criterion order and decreases with k, and eventually the detection ratios all approach 1. This phenomenon matches with our understanding:

In Fig. 9, we compare the detection ratio of our protocol with that of the correlation matrix method, and demonstrate how both detection ratios evolve with the dimension of the local system. One prominent feature of the correlation matrix method is that its detection ratio

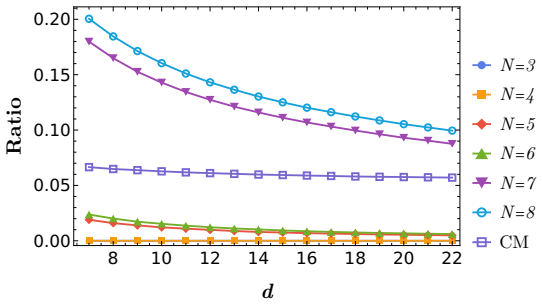


Figure 9: Comparison between the k-reduction moment criteria (Eq. (63)) and the moment-based correlation matrix criterion (Eq. (23)). The horizontal line is the local dimension d. In this test, we sample states from $\mathcal{E}_0$ (Eq. (91)) with $r = 6$ and set $k = 5$. The detection ratios are obtained by averaging over 10^6 samples. The detection ratio of the bounded moment criteria can be larger than that of the correlation matrix criterion for $N \geq 7$. However, the detection ratio of the bounded moment criteria decays fast with the local dimension, while the detection ratio of the correlation matrix criterion varies slowly with d.

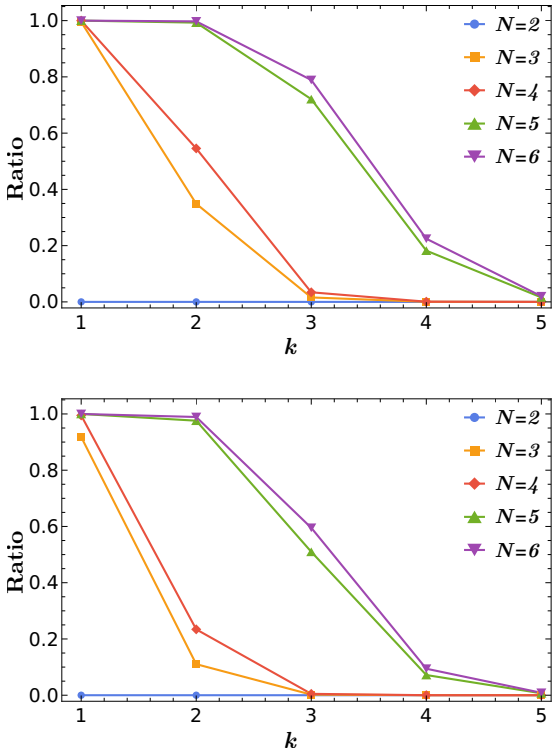


Figure 10: Detection ratios of different k-reduction moment conditions (Eq. (63)) for pure state ensemble $\mathcal{E}_0$ (Eq. (91)). The horizontal line is k. The pure state ensemble $\mathcal{E}_0$ has Schmidt rank $r = 6$. The first figure has local dimension $d = 8$, the second figure has local dimension $d = 16$.

Table 4: Detection ratios for the optimal Schmidt number lower bound obtained by Eq. (63) in induced metric ensemble $\mathcal{E}_{D,K}$ (Eq. (101)) with $D = 16^2$.

SN= \ K=	2	3	4	5	6
3	0	0	0	0	0.1506
4	0	0	0	0.9986	0.8494
5	0	0	1.0000	0.0014	0
6	0	0.9984	0	0	0
7	0.0610	0.0016	0	0	0
8	0.9386	0	0	0	0
9	0.0004	0	0	0	0

Table 5: Detection ratios for the optimal Schmidt number lower bound obtained by Eq. (23) in induced metric ensemble $\mathcal{E}_{D,K}$ (Eq. (101)) with $D = 16^2$.

SN= \ K=	2	3	4	5	6
5	0	0	0.0678	1.0000	1.0000
6	0	0.9988	0.9322	0	0
7	0.9958	0.0012	0	0	0
8	0.0042	0	0	0	0

almost does not decay with the dimension. This is because for a pure state $|\psi\rangle$, the singular values of its correlation matrix T are sorely determined by the eigenvalues of $\mathrm{Tr}_B(|\psi\rangle\langle\psi|)$, thus the criterion function $k - d^{-1} - \|T\|_1$ depends very weakly on d. On the other hand, each q_n is a polynomial function of d_B, which makes the detection ratio of our protocol highly susceptible to the size of dimension.

F.4 Ensemble of induced metric states

Let $|\psi\rangle$ be a sample of Haar random pure state on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. We set $\dim(\mathcal{H}_A) = \dim(\mathcal{H}_B) = \sqrt{D}$ and $\dim(\mathcal{H}_C) = K$. Then the ensemble of induced metric states is generated by

$$\mathcal{E}_{D,K} = \{\mathrm{Tr}_C(|\psi\rangle\langle\psi|)\}. \tag{101}$$

We do not choose the Hilbert-Schmidt ensemble with $K = D$ because the states rarely have negative eigenvalues under the action of the k-reduction map, although these states are known to be entangled [42, 56].

In the numerical tests we set $D = d^2 = 16^2$ and $K = 2, 3, 4, 5, 6$, then use Algorithm 1 to detect the Schmidt numbers of sampled states. As shown by the results (see Table 4 and Table 5), the moment method of k-reduction map provides the lower bound estimation of the Schmidt number of an unknown state ρ, which means if the estimation is k_0 by the above algorithm, then $\mathrm{SN}(\rho) \geq k_0$. From the numerical results, we cannot tell which method is better, since there are always some cases where one of the methods provides higher lower bound for Schmidt number.

Table 6: Detection ratios of Schmidt number lower bounds certifiable by criterion Eq. (26) in the induced metric ensemble $\mathcal{E}_{D,K}$ (Eq. (101)) with $D = 16^2$.

SN= \ K=	2	3	4	5	6
2	1	1	1	1	1
3	1	1	1	1	0.9976
4	1	1	1	0.0176	0
5	1	1	0.0006	0	0
6	1	0.1596	0	0	0
7	1	0	0	0	0
8	1	0	0	0	0
9	0.0022	0	0	0	0

Table 7: Detection ratios of Schmidt number lower bounds certifiable by criterion Eq. (20) in the induced metric ensemble $\mathcal{E}_{D,K}$ (Eq. (101)) with $D = 16^2$.

SN= \ K=	2	3	4	5	6
5	1	1	1	1	1
6	1	1	1	0	0
7	1	0	0	0	0
8	1	0	0	0	0
9	0.0034	0	0	0	0

G Conclusion

In this paper, we design a practical and efficient protocol for entanglement dimensionality characterization that is applicable to a wide class of quantum states. The protocol verifies whether $\mathcal{R}_k(\rho)$ is positive or not through the Hankel matrix. Each entry of the matrix is a non-linear function of ρ and ρ_A. The key procedure is thus to use the randomized measurement method to estimate several non-linear functions of the target state.

The comparison between the k-reduction map method and the correlation matrix method is analyzed from two aspects. In terms of the original criteria, the k-reduction map is able to detect the Schmidt number of all pure states, while the correlation matrix method cannot (see Eq. (44) for a counterexample). For noisy maximally entangled state with Schmidt number r and the ensemble of Haar random pure states with depolarized noise, the k-reduction map method can also detect a larger range of states.

In terms of the moment-based protocols (those can actually be implemented by experiments), our protocol is more feasible for two reasons. First, our lowest order protocol with $N^* = 3$ only needs unitary-3 design to perform, while in correlation matrix method, the estimation of $\mathcal{C}^{(4)}$ needs unitary-4 design. Second, by combining the classical shadow and the statistical correlation, we only need $\Omega(D^{1/2})$ samples to estimate all necessary parameters in the third moment k-reduction criterion. While the sample complexity of the correlation matrix

method is unknown yet. Furthermore, the third order k-reduction moment criterion is enough to certify isotropic states, which is more powerful than the moment-based correlation matrix criterion. However, the detection ratios of the low order moment-based protocols decay to zero with the dimension of the system, while the correlation matrix method depends weakly on the dimension.

A few unsolved questions in this paper also deserve further exploration. In our protocol, we only use the simplest type of k-positive map. Other extent k-positive maps [49] are very complicated, and not suitable for designing practical protocols at the first glance. Thus, can we design new k-positive maps that are more efficient than the k-reduction map? Besides, we do not take the statistical errors into consideration when analyzing the performances of our protocol. The analysis requires knowledge about the perturbation theory of Hankel matrices, which we will leave for future work. Finally, it is unclear whether the k-reduction negativity is an entanglement measure or not yet. The purpose is to show that the k-reduction negativity, as a function of ρ, does not increase under LOCC operations on ρ. A rigorous proof is still absent.

H Acknowledgement

We thank Xiaodong Yu, Zihao Li, Datong Chen and Zhenhuan Liu for helpful discussions. C.Y. and X.L. acknowledge support from the National Natural Science Foundation of China (Grant No. 92165109), National Key Research and Development Program of China (Grant No. 2022YFA1404204), and Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01).

Supplemental Materials

I Spectrum of k-reduced operators

In this section, we discuss the details about the spectrum of the k-reduced operators. We start with the simplest case of the pure states and provide a complete characterization of the spectrum of the corresponding k-reduced operators, and then consider a more involved situation, where several types of noises are added to the pure states. Although the spectrum of the k-reduced operator of a general density operator is very hard to calculate, we still succeed in extracting information about the spectrum.

I.1 Pure state

We start with the proof of Theorem 4. Part of the result has been proved in previous works [11, 6]. Here we prove it with a different method.

Proof of Theorem 4. Given the explicit Schmidt decomposition of $|\psi\rangle$, we have

$$\mathcal{R}_k(\rho) = k\sum_{i=0}^{r-1}\lambda_i|i\rangle_A\langle i|\otimes I - \sum_{i,j=0}^{r-1}\sqrt{\lambda_i\lambda_j}|i\rangle_A\langle j|\otimes|i\rangle_B\langle j|. \tag{102}$$

Notice that $\{|j\rangle_B\}_{j=0}^{r-1}$ is part of a vector basis of $\mathcal{H}_B$. Denote the full basis as $\{|j\rangle_B\}_{j=0}^{d_B-1}$. Define $\mathcal{W}$ as the space $\mathrm{span}\{|i\rangle_A\}_{i=0}^{r-1}\otimes\mathcal{H}_B$, then the bipartite Hilbert space can be decomposed as $\mathcal{H}_A\otimes\mathcal{H}_B = \mathcal{W}\oplus\mathcal{W}^\perp$, such that $\dim(\mathcal{W}) = d_Br, \dim(\mathcal{W}^\perp) = d_B(d_A-r)$, and $\mathcal{R}_k(\rho)$ is supported on $\mathcal{W}$. In space $\mathcal{W}^\perp$, there are $d_B(d_A-r)$ eigenvectors of $\mathcal{R}_k(\rho)$ with eigenvalue 0.

Operator $\mathcal{R}_k(\rho)$ can be divided into X_1+X_2, where

$$X_1 \equiv k\sum_{i=0}^{r-1}\lambda_i|ii\rangle\langle ii| - \sum_{i,j=0}^{r-1}\sqrt{\lambda_i\lambda_j}|ii\rangle\langle jj|, \tag{103}$$

$$X_2 \equiv k\sum_{i=0}^{r-1}\sum_{j=0,j\neq i}^{d_B-1}\lambda_i|ij\rangle\langle ij|. \tag{104}$$

We further define $\mathcal{W}_1 \equiv \mathrm{span}\{|ii\rangle\}_{i=0}^{r-1}$, and $\mathcal{W} = \mathcal{W}_1\oplus\mathcal{W}_2$. Notice that X_1 is supported on $\mathcal{W}_1$ and X_2 is supported on $\mathcal{W}_2$. In $\mathcal{W}_1$, X_1 has r eigenvalues denoted by $\{x_i\}_{i=0}^{r-1}$; In $\mathcal{W}_2$, X_2 has eigenvalues $\{k\lambda_i\}_{i=0}^{r-1}$ and each has multiplicity d_B-1. From here we can see that $\mathcal{R}_k(\rho)\succeq 0$ if and only if $X_1\succeq 0$.

Rewrite X_1 as $\Lambda^{\frac{1}{2}}O\Lambda^{\frac{1}{2}}$, where

$$\Lambda^{\frac{1}{2}} \equiv \sum_{i=0}^{r-1}\lambda_i^{\frac{1}{2}}|ii\rangle\langle ii|, \quad O \equiv k\sum_{i=0}^{r-1}|ii\rangle\langle ii| - \sum_{i,j=0}^{r-1}|ii\rangle\langle jj|. \tag{105}$$

The eigenvalues of O are $\{k, k-r\}$, where the multiplicity of k is $r-1$. Both Λ and O live in $\mathcal{W}_1$ as X_1 does, hence $X_1\succeq 0$ if and only if $O\succeq 0$, which further implies $\mathcal{R}_k(\rho)\succeq 0$ if and only if $k\geq r$.

When $k<r$, O has exactly one negative eigenvalue $k-r$. Because $\Lambda^{1/2}$ is reversible in $\mathcal{W}_1$, X_1 also only contains one negative eigenvalue. □

Having finished the proof of Theorem 4, we want to know more details about the eigenvalues of X_1. Assuming $|\phi\rangle = \sum_{i=0}^{r-1}\phi_i|ii\rangle$ is one eigenvector of X_1 with eigenvalue x, we substitute it into the equation $X_1|\phi\rangle = x|\phi\rangle$ and get

$$k\lambda_i\phi_i - \sum_{j=0}^{r-1}\sqrt{\lambda_i\lambda_j}\phi_j = x\phi_i \quad \text{for } i=0,\cdots,r-1. \tag{106}$$

Write the above equation as

$$(k\lambda_i - x)\phi_i = \sqrt{\lambda_i}\left(\sum_{j=0}^{r-1}\sqrt{\lambda_j}\phi_j\right). \tag{107}$$

According to the factor $k\lambda_i - x = 0$ or not, the solution of the above equation can be classified into two situations.

(a) $\exists 0 \le i \le r-1, x = k\lambda_i$. We set $i = 0$ without loss of generality. Since $k\lambda_0$ is an eigenvalue of X_1, it is a root of the characteristic function:

$$\det(k\lambda_0 - X_1) = \det(\Lambda)\times$$
$$\det\left(\sum_{i=0}^{r-1}\frac{k\lambda_0 - k\lambda_i}{\lambda_i}|ii\rangle\langle ii| + \sum_{i,j=0}^{r-1}|ii\rangle\langle jj|\right) = 0. \quad (108)$$

Because

$$\begin{vmatrix} 1 & 1 & 1 & \cdots \\ 1 & 1+\lambda_1' & 1 & \cdots \\ 1 & 1 & 1+\lambda_2' & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & \cdots \\ 0 & \lambda_1' & 0 & \cdots \\ 0 & 0 & \lambda_2' & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} = \lambda_1'\lambda_2'\cdots, \quad (109)$$

we have

$$\begin{aligned}\det(k\lambda_0 - X_1) &= \det(\Lambda)\cdot\prod_{j=1}^{r-1}\frac{k\lambda_0 - k\lambda_j}{\lambda_j} \\ &= k^{r-1}\lambda_0\prod_{j=1}^{r-1}(\lambda_0 - \lambda_j). \quad (110)\end{aligned}$$

So $x = k\lambda_0$ if and only if $\exists 0 < j \le r-1, \lambda_0 = \lambda_j$. Equivalently, for every pair $0 \le i \ne j \le r-1, \lambda_i \ne \lambda_j$ if and only if $\forall 0 \le i \le r-1, x \ne k\lambda_i$, which is our second situation.

(b) $\forall 0 \le i \le r-1, k\lambda_i - x \ne 0$. Dividing the non-zero factor on the two sides of Eq.(107), then we obtain

$$\phi_i = \frac{\sqrt{\lambda_i}}{k\lambda_i - x}\left(\sum_{j=0}^{r-1}\sqrt{\lambda_j}\phi_j\right). \quad (111)$$

Multiplying it by $\sqrt{\lambda_i}$ on the two sides and summing it over i, we get the following equation

$$\left(\sum_{i=0}^{r-1}\frac{\lambda_i}{k\lambda_i - x} - 1\right)\left(\sum_{j=0}^{r-1}\sqrt{\lambda_j}\phi_j\right) = 0. \quad (112)$$

If $\sum_{i=0}^{r-1}\sqrt{\lambda_i}\phi_i = 0$, then from Eq.(107) we have $\phi_i = 0$ for all i, which is a trivial solution. Hence, every eigenvalue $x \ne k\lambda_i$ must satisfy

$$\sum_{i=0}^{r-1}\frac{\lambda_i}{k\lambda_i - x} = 1. \quad (113)$$

If we consider the complete Schmidt spectrum of $|\psi\rangle$ including eigenvalue 0: $\boldsymbol{\lambda} = \{\lambda_i\}_{i=0}^{r-1}\cup\{0\}^{d-r}$, the eigenvalue x satisfies

$$\sum_{i=0}^{d-1}\frac{\lambda_i}{k\lambda_i - x} = 1, \quad (114)$$

where $\lambda_i = 0$ for $i \ge r$. When $k < r$, the k-reduction negativity $\mathcal{N}_k(\psi)$ is defined as the opposite of the unique negative eigenvalue of $\mathcal{R}_k(\rho)$, hence it satisfies

$$\sum_{i=0}^{d-1}\frac{\lambda_i}{k\lambda_i + \mathcal{N}_k(\psi)} = 1. \quad (115)$$

Recall that we define $\theta_k(\boldsymbol{\lambda})$ in Eq. (34) as a function of $\boldsymbol{\lambda} \in [0,1]^{\times d}$ whose value equals $\mathcal{N}_k(\psi)$ when $|\psi\rangle$ has Schmidt spectrum $\boldsymbol{\lambda}$:

$$\begin{aligned}&\text{if } k \ge r, \quad \theta_k(\boldsymbol{\lambda}) = 0, \\ &\text{if } k < r, \quad \theta_k(\boldsymbol{\lambda}) > 0, \quad s.t. \quad \sum_{i=0}^{d-1}\frac{\lambda_i}{k\lambda_i + \theta_k(\boldsymbol{\lambda})} = 1.\end{aligned} \quad (116)$$

Given a nonzero vector $\boldsymbol{x} = (x_0, x_1, \cdots, x_{d-1})$ defined in $[0,\infty)^{\times d}$, we introduce the following function

$$G(\theta;\boldsymbol{x}) \equiv \sum_{i=0}^{d-1}\frac{x_i}{kx_i + \theta}. \quad (117)$$

We will prove that for $k < r$, the equation $G(\theta;\boldsymbol{\lambda}) = 1$ has a unique positive root, so we can define $\theta_k(\boldsymbol{\lambda})$ as the positive solution of the equation. Notice that $G(\theta;\boldsymbol{\lambda})$ is continuous and decreases monotonically with θ in $(0,\infty)$. Because $G(0;\boldsymbol{x}) = r/k > 1$ and $\lim_{\theta\to\infty}G(\theta;\boldsymbol{x}) = 0 < 1$, there must exist a unique solution for $G(\theta;\boldsymbol{x}) = 1$ in $(0,+\infty)$.

Proof of Theorem 5. (a) We first prove that $\theta_k(\boldsymbol{\lambda})$ is Schur concave in $(0,1)^{\times d}$. We know that a smooth symmetric function $f(\lambda_1,\cdots,\lambda_d)$ is Schur concave if $\frac{\partial f}{\partial\lambda_\ell} - \frac{\partial f}{\partial\lambda_j}$ always has the opposite sign with $\lambda_\ell - \lambda_j$ [51]. To use this theorem, we compute the partial derivative $\frac{\partial\theta_k}{\partial\lambda_\ell}$ by

$$\left[\sum_{i=0}^{d-1}\frac{k\lambda_i}{(k\lambda_i + \theta_k)^2}\right]\cdot\frac{\partial\theta_k}{\partial\lambda_\ell} = \frac{k\theta_k}{(k\lambda_\ell + \theta_k)^2}. \quad (118)$$

Because θ_k is positive in $(0,1)^{\times d}$, when $\lambda_\ell > \lambda_j$, we have $\frac{\partial\theta_k}{\partial\lambda_\ell} < \frac{\partial\theta_k}{\partial\lambda_j}$. Therefore, θ_k is Schur concave in the open set $(0,1)^{\times d}$.

Let's consider two nonzero vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ defined in $[0,1)^{\times d}$, and assume $\boldsymbol{\lambda} \succ \boldsymbol{\eta}$, where $\boldsymbol{\lambda}$ has $r > k$ non-zero entries, $\boldsymbol{\eta}$ has $t > k$ non-zero entries. The majorization relation ensures that $r \le t$. We introduce a parameter $\delta \in (0,1)$ such that

$$\begin{aligned}\boldsymbol{\lambda}_\delta &\equiv \left((1-\delta)\lambda_0,\cdots,(1-\delta)\lambda_{r-1},\frac{\delta}{d-r},\cdots,\frac{\delta}{d-r}\right), \\ \boldsymbol{\eta}_\delta &\equiv \left((1-\delta)\eta_0,\cdots,(1-\delta)\eta_{t-1},\frac{\delta}{d-t},\cdots,\frac{\delta}{d-t}\right),\end{aligned} \quad (119)$$

then $\boldsymbol{\lambda}_\delta \succ \boldsymbol{\eta}_\delta$. Since $\boldsymbol{\lambda}_\delta, \boldsymbol{\eta}_\delta \in (0,1)^{\times d}$, according to the previous conclusion, we get

$$\theta_k(\boldsymbol{\eta}_\delta) \ge \theta_k(\boldsymbol{\lambda}_\delta). \quad (120)$$

As functions of δ, both $\theta_k(\boldsymbol{\eta}_\delta)$ and $\theta_k(\boldsymbol{\lambda}_\delta)$ are continuous in δ. Thus, taking the limit $\delta \to 0^+$, we obtain

$$\theta_k(\boldsymbol{\eta}) \ge \theta_k(\boldsymbol{\lambda}). \quad (121)$$

Let's consider another case where $\boldsymbol{\lambda} \succ \boldsymbol{\eta}$, $r \le k$ and $t \le k$ or $k \le t$, then

$$\theta_k(\boldsymbol{\eta}) \ge 0 = \theta_k(\boldsymbol{\lambda}). \quad (122)$$

Combining all the situations, we can conclude that $\theta_k(\boldsymbol{\lambda})$ is Schur concave in $[0,1)^{\times d}$.

(b) For any Schmidt spectrum $\boldsymbol{\lambda}$ with r non-zero entries, the vector that is majorized by the vector

$$\left(r^{-1},\cdots,r^{-1},0,\cdots,0\right), \tag{123}$$

with r non-zero entries. For $\left(r^{-1},\cdots,r^{-1},0,\cdots,0\right)$, $\theta_k = 1-k/r$, so we get $\theta_k(\boldsymbol{\lambda}) \leq 1-k/r$, since θ_k is Schur concave.

(c) Given $\boldsymbol{\lambda}$, as k increases from 1 to $r-1$, $\theta_k(\boldsymbol{\lambda})$ has to become smaller according to Eq. (116). If $k \geq r$, $\theta_k = 0$. So we can conclude that $\theta_k(\boldsymbol{\lambda})$ remains non-increasing as k increases. □

I.2 Pure state with noise

Since we cannot prepare an ideal pure state in laboratory, we have to consider the pure states with noises. Firstly, we consider a pure state with depolarizing noise:

$$\rho = (1-\varepsilon)|\psi\rangle\langle\psi| + \varepsilon\frac{I}{d_A d_B}, \tag{124}$$

where $\varepsilon \in [0,1]$. Applying the k-reduction map to it, we get

$$\mathcal{R}_k(\rho) = (1-\varepsilon)\mathcal{R}_k(|\psi\rangle\langle\psi|) + \frac{\varepsilon}{d_A d_B}\mathcal{R}_k(I_{AB}). \tag{125}$$

Hereafter, we denote the eigenvalues of an operator O by $\sigma_i(O)$ and arrange them in increasing order as $\sigma_0(O) \leq \sigma_1(O) \leq \cdots \leq \sigma_{d_A d_B-1}(O)$. For simplicity, we also define $\sigma_\ell \equiv \sigma_\ell(\mathcal{R}_k(|\psi\rangle\langle\psi|)), \ell = 0,1,\cdots,d_A d_B - 1$.

Proof of Theorem 7. The smallest eigenvalue of $\mathcal{R}_k(\rho)$ is

$$\sigma_0(\mathcal{R}_k(\rho)) = (1-\varepsilon)\sigma_0 + \varepsilon\frac{kd_B-1}{d_A d_B}. \tag{126}$$

If $k < r$, then $\sigma_0 < 0$ and equals $-\mathcal{N}_k(\psi)$. When ε is small enough, we have $\sigma_0(\mathcal{R}_k(\rho)) < 0$ and it equals $-\mathcal{N}_k(\rho)$. When ε is large enough, we have $\sigma_0(\mathcal{R}_k(\rho)) \geq 0$ and $\mathcal{N}_k(\rho) = 0$. The critical value of ε that separates the two situations can be obtained by solving

$$-(1-\varepsilon^*)\mathcal{N}_k(\psi) + \varepsilon^*\frac{kd_B-1}{d_A d_B} = 0. \tag{127}$$

□

If we assume $d_A, d_B \gg 1$, in order to have $\sigma_0(\mathcal{R}_k(\rho)) < 0$, we need

$$\varepsilon < \frac{\sigma_0}{\sigma_0 - \frac{kd_B-1}{d_A d_B}} \approx \frac{\sigma_0}{\sigma_0 - k/d_A} = \frac{\mathcal{N}_k(\psi)}{\mathcal{N}_k(\psi) + k/d_A}. \tag{128}$$

As illustrated in Fig. 11, when ε becomes closer and closer to 1, k/d_A must be smaller and smaller. We can only certify a small lower bound of the Schmidt number when p is close to 1.

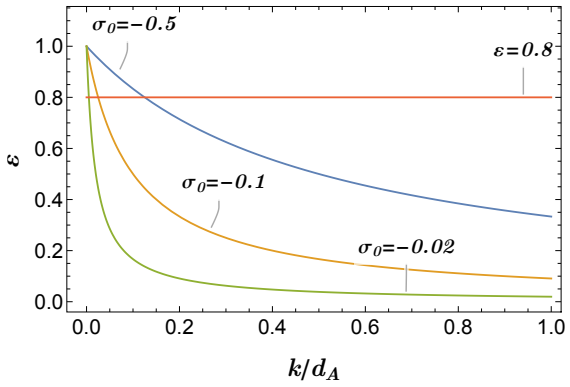


Figure 11: With fixed values of σ_0, the relation between ε and k/d_A is shown in the figure. The three curves are $\sigma_0/(\sigma_0 - k/d_A)$ with $\sigma_0 = 0.01, 0.2, 0.5$ separately. To order to have a positive k-reduction negativity, ε must be below the three curves the for three cases. The $\varepsilon = 0.8$ shows when the noise strength is large, how small k/d_A should be to certify the Schmidt number. When σ_0 approaches 0, the region of possible values for ε shrinks quickly.

The isotropic state can be viewed as a maximally state with depolarizing noise:

$$\begin{aligned}\rho_F =& \frac{1-F}{d^2-1}(I - |+_d\rangle\langle+_d|) + F|+_d\rangle\langle+_d| \\ =& \frac{1-F}{d^2-1}I + \frac{d^2F-1}{d^2-1}|+_d\rangle\langle+_d|,\end{aligned} \tag{129}$$

where we have assumed $d_A = d_B = d$. The smallest eigenvalue of the corresponding k-reduced operator is

$$\sigma_0(\mathcal{R}_k(\rho_F)) = \frac{(1-F)(kd-1)}{d^2-1} + \frac{d^2F-1}{d^2-1}\sigma_0(\mathcal{R}_k(|+_d\rangle\langle+_d|)). \tag{130}$$

The spectrum of $\mathcal{R}_k(|+_d\rangle\langle+_d|)$ contains $k/d-1$ with multiplicity 1 and k/d with multiplicity d^2-1. So

$$\sigma_0(\mathcal{R}_k(\rho_F)) = k/d - F. \tag{131}$$

We can conclude $\mathrm{SN}(\rho_F) > k$ when $F > k/d$, which is consistent with a previous result [6].

Proof of Theorem 8. To certify a lower bound of the Schmidt number of a quantum state, we can first construct a k-positive map, then prove that the state becomes non-positive after applying the map. To certify an upper bound, we can construct a pure state decomposition of that state, and find which $\mathcal{S}_r$ it belongs to.

Given the state

$$\rho_{\epsilon,r} = (1-\epsilon)|+_r\rangle\langle+_r| + \epsilon\frac{I}{d_A d_B}, \quad |+_r\rangle = \frac{1}{\sqrt{r}}\sum_{j=0}^{r-1}|j_A\rangle\otimes|j_B\rangle, \tag{132}$$

we define $\mathcal{W}_0 \equiv \mathrm{span}\{|i_A\rangle\otimes|j_B\rangle, i,j = 0,1,\cdots,r-1\}$, and let I_0 be the projector onto this subspace and $I_1 =$

$I - I_0$, then we obtain

$$\begin{aligned}\rho_{\epsilon,r} &= \left[(1-\epsilon)|+_r\rangle\langle+_r| + \epsilon\frac{I_0}{d_A d_B}\right] \oplus \epsilon\frac{I_1}{d_A d_B} \\ &= (1-\epsilon')\rho^{\text{ef}}_{\epsilon,r} \oplus \epsilon'\frac{I_1}{d_A d_B - r^2},\end{aligned} \tag{133}$$

where

$$\rho^{\text{ef}}_{\epsilon,r} \equiv \frac{1-\epsilon}{1-\epsilon'}|+_r\rangle\langle+_r| + \frac{\epsilon}{1-\epsilon'}\frac{I_0}{d_A d_B}, \quad \epsilon' \equiv \epsilon\frac{d_A d_B - r^2}{d_A d_B}. \tag{134}$$

Notice that $\rho^{\text{ef}}_{\epsilon,r}$ lives in $\mathcal{W}_0$. Every pure state decomposition of $\rho^{\text{ef}}_{\epsilon,r}$ corresponds to a pure state decomposition of $\rho_{\epsilon,r}$ with the same maximal Schmidt number. Thus,

$$\text{SN}(\rho_{\epsilon,r}) \leq \text{SN}\left(\rho^{\text{ef}}_{\epsilon,r}\right). \tag{135}$$

The state in the RHS is an isotropic state. Previous result [6] shows that the Schmidt number of

$$\frac{1-F}{r^2-1}I_0 + \frac{r^2F-1}{r^2-1}|+_r\rangle\langle+_r| \tag{136}$$

equals $\lceil rF \rceil$. Therefore, by solving the equation of F:

$$\frac{r^2F-1}{r^2-1} = \frac{1-\epsilon}{1-\epsilon'}, \tag{137}$$

and introduce $u \equiv \frac{\epsilon}{(1-\epsilon)d_A d_B}$ for simplicity, we obtain

$$\begin{aligned}\text{SN}(\rho_{\epsilon,r}) &\leq \left\lceil \frac{r^2-1}{r} \cdot \frac{d_A d_B(1-\epsilon)}{d_A d_B(1-\epsilon)+\epsilon r^2} + \frac{1}{r} \right\rceil \\ &= \left\lceil r\frac{1+u}{1+r^2u} \right\rceil.\end{aligned} \tag{138}$$

On the other hand, according to Theorem 7, the smallest eigenvalue of $\mathcal{R}_k(\rho_{\epsilon,r})$ is

$$(1-\varepsilon)(k/r-1) + \varepsilon\frac{kd_B - 1}{d_A d_B}. \tag{139}$$

When the value is smaller than 0, we can conclude that $\text{SN}(\rho_{\epsilon,r}) > k$. The critical value of k that satisfies $\sigma_0(\mathcal{R}_k(\rho_{\epsilon,r})) = 0$ is

$$r \cdot \frac{1-\varepsilon+\frac{\varepsilon}{d_A d_B}}{1-\varepsilon+\varepsilon\frac{r}{d_A}} = r\frac{1+u}{1+d_B ru}. \tag{140}$$

Thus, the optimal Schmidt number lower bound certifiable by the k-reduction map is

$$\text{SN}(\rho_{\epsilon,r}) \geq \left\lceil r\frac{1+u}{1+d_B ru} \right\rceil. \tag{141}$$

One can verify that when $\varepsilon < 1/2, r \leq \sqrt{d_A}$, we have $u < (d_A d_B)^{-1}$ and

$$r\frac{1+u}{1+d_B ru} > r-1. \tag{142}$$

So $\text{SN}(\rho_{\varepsilon,r}) = r$ in this case. □

Secondly, we consider the pure state with dephasing noise. Assume $d_A = d_B = d$ and the pure state $|\psi\rangle$ has Schmidt number r. The full state reads

$$\rho = (1-p)|\psi\rangle\langle\psi| + p|+_d\rangle\langle+_d|, \quad p \in [0,1]. \tag{143}$$

Its k-reduced operator is

$$\mathcal{R}_k(\rho) = (1-p)\mathcal{R}_k(|\psi\rangle\langle\psi|) + p\mathcal{R}_k(|+_d\rangle\langle+_d|) \tag{144}$$

which is just the convex combination of the k-reduced operators of $|\psi\rangle$ and $|+_d\rangle$. Using Weyl inequalities [53], we obtain

$$\sigma_0(\mathcal{R}_k(\rho)) \geq (1-p)\sigma_0 + p(k/d-1), \tag{145}$$

$$\begin{aligned}\sigma_0(\mathcal{R}_k(\rho)) \leq \min\{&(1-p)\sigma_0 + pk/d, \\ &(1-p)\sigma_{d^2-1} + p(k/d-1)\}.\end{aligned} \tag{146}$$

Define $\Delta \equiv \sigma_{d^2-1} - \sigma_0$. When $p < \frac{\Delta}{\Delta+1}$, the first term on right-hand side of Eq. (146) is smaller, the condition for $\sigma_0(\mathcal{R}_k(\rho)) < 0$ is

$$p < \frac{\sigma_0}{\sigma_0 - k/d_A}. \tag{147}$$

When $p > \frac{\Delta}{\Delta+1}$, the second term on right-hand side of Eq. (146) is smaller, the condition for $\sigma_0(\mathcal{R}_k(\rho)) < 0$ is

$$p > \frac{\sigma_{d^2-1}}{\sigma_{d^2-1} + 1 - k/d}. \tag{148}$$

J The moment method and moment criteria

In this section of the appendix, we aim to constructing the moment method more rigorously, and providing the proof of Corollary 13. Here we will not provide the theorems about the solutions of the Hamburger, Stieltjes and $[a,b]$-moment problems, but recommend the books [16, 57] and the lecture notes [58] to the readers.

J.1 The relative moment problems

We want to ask such questions: what are the conditions for a Hamburger moment sequence S further being a Stieltjes moment sequence, or a $[a,b]$-moment sequence ($a < 0 < b$) further being a $[0,b]$-moment sequence? We call such problems as *the relative moment problems*. Although there are no direct answers for these relative moment problems in the [16], it is easy to get the answer by comparing the known results of the moment problems. So we directly list the results in the following.

Corollary 20 (The relative infinite Stieltjes moment problem)**.** *Suppose S is a Hamburger moment sequence, and $H(S_{1,2n+1}) \succeq 0$ for all $n \in \mathbb{N}^+$, then S is also a Stieltjes moment sequence.*

Corollary 21 (The relative truncated Stieltjes moment problem)**.** *Suppose S_N is a truncated Hamburger moment sequence, and*
(a) even case $N = 2n$: $H(S_{1,2n-1}) \succeq 0$, $(s_{n+1}, \cdots, s_{2n})^T \in \text{range}(H(S_{1,2n-1}))$;
(b) odd case $N = 2n+1$: $H(S_{1,2n+1}) \succeq 0$.
Then S is also a truncated Stieltjes moment sequence:

Corollary 22 (The relative infinite $[0, b]$-moment problem). *Suppose S is an $[a, b]$-moment sequence with $a < 0 < b$, and $H(bS_{1,2n+1} - S_{2,2n+2}) \succeq 0$ (or $H(S_{1,2n+1}) \succeq 0$) for $n \in \mathbb{N}^+$, then S is also a $[0, b]$-moment sequence.*

Corollary 23 (The relative truncated $[0, b]$-moment problem). *Suppose S_N is a truncated $[a, b]$-moment sequence with $a < 0 < b$, and*
(a) even case $N = 2n$, $H(bS_{1,2n-1} - S_{2,2n}) \succeq 0$;
(b) odd case $N = 2n + 1$, $H(S_{1,2n+1}) \succeq 0$;
Then S_N is also a truncated $[0, b]$-moment sequence.

J.2 The moment criteria

Suppose $\mathcal{O}$ is a set of Hermitian operators on Hilbert space $\mathcal{H}$ with dimension D, and $\mathcal{O}^+$ is the subset of positive semidefinite operators in $\mathcal{O}$, our goal is to characterize the positive operators of $\mathcal{O}^+$ with respect to $\mathcal{O}$ by a series of moment criteria. Given an operator $X \in \mathcal{O}$, if the spectrum of X is $(\lambda_1, \cdots, \lambda_D)$, there is a corresponding atomic measure $\mu_X = \sum_{i=1}^{D} \delta(x - \lambda_i)$ with $x \in \mathbb{R}$ being a real variable, and the n-th moment is related to the spectrum of X by

$$s_n(X) \equiv \mathrm{Tr}(X^n) = \sum_{i=0}^{D-1} \lambda_i^n = \int_{-\infty}^{\infty} x^n d\mu_X. \tag{149}$$

Through the atomic measure, each operator in $\mathcal{O}$ is mapped into an infinite real sequence $S(X) = (s_0(X), s_1(X), \cdots)$, the set of all such infinite real sequences is denoted as

$$\mathcal{X}_{\mathcal{O}} = \{(s_0(X), s_1(X), \cdots) : X \in \mathcal{O}, s_n(X) = \mathrm{Tr}(X^n)\}. \tag{150}$$

Likewise, the set of infinite real sequences corresponding to $\mathcal{O}^+$ is

$$\mathcal{X}_{\mathcal{O}^+} = \{(s_0(X), s_1(X), \cdots) : Y \in \mathcal{O}^+, s_n(X) = \mathrm{Tr}(X^n)\}. \tag{151}$$

By definition, every sequence $S(X) \in \mathcal{X}_{\mathcal{O}}$ is a Hamburger moment sequence, every sequence $S(O) \in \mathcal{X}_{\mathcal{O}^+}$ is further a Stieltjes moment sequence and $\mathcal{X}_{\mathcal{O}^+} \subset \mathcal{X}_{\mathcal{O}}$.

Given the assumption that the infinite moment sequence $S(X) = (s_0(X), s_1(X), \cdots)$ of an unknown operator $X \in \mathcal{O}$ is known, the sequence $S(X)$ is naturally a Hamburger moment sequence, then we want to know if $S(X)$ is a Stieltjes moment sequence. First, we consider if the sequence $S(X)$ has a unique representing measure,

$$\int_{-\infty}^{+\infty} e^{\epsilon|x|} d\mu_X(x) = \sum_{i=1}^{D} \int_{-\infty}^{+\infty} e^{\epsilon|x|} \delta(x - \lambda_i) dx \leq D e^{\epsilon\tau} \tag{152}$$

where τ represents the upper bound of $|\lambda_i|$, which means the spectrum of the operator X is assumed to be bounded. The above result tells us that $\int_{-\infty}^{+\infty} e^{\epsilon|x|} d\mu_X(x) < \infty$, then according to the Carleman's condition [16], the Hamburger sequence $S(X)$ corresponds to a unique measure, and it's the same for operators in $\mathcal{O}^+$. The previous results mean that there is a one-to-one map between the operator set $\mathcal{O}$ and the sequence set $\mathcal{X}_{\mathcal{O}}$ (up to unitary transformation). In other words, a positive operator $X \in \mathcal{O}^+$ must correspond to a Stieltjes moment sequence and a non-positive operator corresponds to a Hamburger moment sequence.

Given an infinite moment sequence $S(X)$ of an unknown operator X, if the moment sequence is further a Stieltjes moment sequence, then we can conclude that X is a positive operator. So by Corollary 20, we get the following conditions, called *the plain moment conditions*, to determine if an unknown operator $X \in \mathcal{O}$ is positive or not:

Theorem 24 (Plain moment conditions). *An operator $X \in \mathcal{O}$ is positive semidefinite if and only if the infinite moment sequence $S(X)$ is a Stieltjes moment sequence. In other words, $S(X)$ must satisfy*

$$H(S_{1,2n+1}(X)) \succeq 0, \ \textit{for all } n \in \mathbb{N}_0. \tag{153}$$

In practice, we only have the first few moments of an operator X, so we must consider the truncated moment sequence rather than the infinite moment sequence. According to Theorem 21, we get the result.

Theorem 25 (Truncated plain moment conditions). *An operator $X \in \mathcal{O}$ is positive semidefinite if and only if the truncated moment sequence $S_N(X)$ is a Stieltjes moment sequence. In other words, $S_N(X)$ has to satisfy the following conditions:*
(1) When $N = 2n$: $H(S_{1,2n-1}(X)) \succeq 0$, and $(s_{n+1}, \cdots, s_{2n})^\top \in \mathrm{range}(H(S_{1,2n-1}(X)))$;
(2) When $N = 2n + 1$: $H(S_{1,2n+1}(X)) \succeq 0$.

We note that the condition of Hankel matrix $H_{n-1}(ES) \succeq 0$ for $N = 2n$ is the same as the one for $N = 2n - 1$, but the case of $N = 2n$ has an extra condition $(s_{n+1}, \cdots, s_{2n})^T \in \mathrm{range}(H_{n-1}(ES))$. Let $\tilde{H}_{n-1}$ denotes the extended matrix of $H_{n-1}(ES)$ by appending the column vector $(s_{n+1}, \cdots, s_{2n})^T$, so $(s_{n+1}, \cdots, s_{2n})^T \in \mathrm{range}(H_{n-1}(ES))$ is equivalent to $\mathrm{rank}\tilde{H}_{n-1} = \mathrm{rank}H_{n-1}(ES)$. However, in practice, the extra condition for $N = 2n$ does not make any difference, since rare random moment sequences satisfy this extra condition. So the truncated plain moment conditions are equivalent to the plain moment conditions effectively.

The truncated plain moment conditions for $N = 2n+1$ has been found in [15] and are used to solve the problem of entanglement certification by combining with PPT. However, we should emphasize that the truncated moment conditions used in [15] are only the necessary conditions for X to be positive, while the infinite series of plain moment conditions are necessary and sufficient conditions.

Since we know the spectrum of any k-reduced operator $\mathcal{R}_k(\rho)$ is bounded, we can use the $[a, b]$-moment method to improve the previous conditions. Generally, we assume the spectrum of an operator $X \in \mathcal{O}$ is bounded by an interval $[a, b]$, then X generates an $[a, b]$-moment sequence $S(X) = (s_0(X), s_1(X), \cdots)$. All infinite $[a, b]$-moment

sequences correspond to unique representing measures, then there is a one-to-one map between the operator set $\mathcal{O}$ and the sequence set $\mathcal{X}_{\mathcal{X}}[a,b]$. Hence, the problem of determining an operator X is positive or negative is equivalent to the one of determining if the moment sequence $S(X)$ is a $[0,b]$-moment sequence, given $S(X)$ is an $[a,b]$-moment sequence? According to Corollary 22, we get the following result.

Theorem 26 (Bounded moment conditions)**.** *An operator $X \in \mathcal{X}$ with spectrum contained in $[a,b]$ is positive semi-definite if and only if the infinite moment sequence $S(X)$ is a $[0,b]$-moment sequence. In other words, $S(X)$ must satisfy*

$$H(S_{1,2n+1}(X)) \succeq 0, \text{ for all } n \in \mathbb{N}^+, \tag{154}$$

or

$$H(bS_{1,2n+1}(X) - S_{2,2n+2}(X)) \succeq 0, \text{ for all } n \in \mathbb{N}^+. \tag{155}$$

In consideration of the truncated moment problems and their solutions 23, we get the following similar result. [Truncated bounded moment conditions] An operator $X \in \mathcal{X}$ with spectrum contained in $[a,b]$ is positive semidefinite if and only if the finite moment sequence $S_N(X)$ is a $[0,b]$-moment sequence. In other words, $S_N(X)$ must satisfy:
(1) when $N = 2n$: $H(bS_{1,2n-1}(X) - S_{2,2n}(X)) \succeq 0$.
(2) when $N = 2n+1$: $H(S_{1,2n+1}(X)) \succeq 0$.

We can see that the plain moment conditions only corresponds to the second series of bounded moment conditions, the truncated bounded moment conditions have one more series of criteria corresponding to the even orders, compared with the conditions in [15].

J.3 Boundary point of the moment cone

Naively, we think the above Theorem J.2 is already a complete characterization of $\mathcal{O}^+$ with respect to $\mathcal{O}$. However, since the truncated moment sequence is only an approximate truncation of the infinite moment sequence, $S_N \in \mathcal{X}_{\mathcal{O}}^N[a,b]$ may have more than one representing measures. So it is possible that there is another representing measure $\nu(S_N)$ of S_N letting S_N be a $[0,b]$-moment sequence, even if the corresponding operator X is not a positive operator, then we will classify X into the class of positive operators by mistake. It is exactly the reason why truncated moment conditions are only necessary conditions but not sufficient conditions to determine if an operator is positive.

We want to know the upper bound of the moment order to determine an operator X being positive definitely. In order to do this, we first review some notions of the truncated $[a,b]$-moment problem. The set of all truncated-N moment sequences is actually a convex cone, called *moment cone* and denoted by $\mathcal{X}^N[a,b]$. Each moment sequence $S_N \in \mathcal{X}^N[a,b]$ has at least one k-atomic representing measure,

$$\mu = \sum_{i=1}^{k} m_i \delta(x - t_i), \tag{156}$$

with $k \leq N+1$. The numbers $t_i \in [a,b]$ are pairwise different, called *roots* of μ, and $m_i > 0$ are the *weights*. The *index* of the representing measure μ for S_N with respect to the interval $[a,b]$ is defined as

$$\mathrm{ind}_{[a,b]}(\mu) \equiv \sum_{i=1}^{k} \epsilon(t_i), \tag{157}$$

where $\epsilon(t) = 2$ if $t \in (a,b)$, while $\epsilon(t) = 1$ if $t = a$ or b. The *index* of sequence S_N is defined as the minimal index of all representing measures for S_N, i.e.,

$$\mathrm{ind}_{[a,b]}(S_N) = \min_{\mu} \mathrm{ind}_{[a,b]}(\mu(S_N)). \tag{158}$$

A representing measure μ of S_N is called *principal* if $\mathrm{ind}_{[a,b]}(\mu) = N+1$, and further called *upper (lower) principal* if it is principal and b is (not) a root of μ. A representing measure μ of S_N is called *canonical* if $\mathrm{ind}_{[a,b]}(\mu) \leq N+2$.

The points of moment cone $\mathcal{X}^N[a,b]$ are classified into two types, boundary points and interior points, the following theorem provides a characterization of them.

Theorem 27 (Characterization of boundary and interior points of moment cone)**.** *Suppose $\mathcal{X}^N[a,b]$ is a closed convex cone in $\mathbb{R}^{N+1}$ with nonempty interior. For $S_N \in \mathcal{X}^N[a,b]$ and N can be even or odd, the following statements are equivalent:*
(1) S_N is a boundary point of $\mathcal{X}^N[a,b]$.
(2) $\det(\underline{H}_N(S)) = 0$ or $\det(\bar{H}_N(S)) = 0$.
(3) S_N is $[a,b]$-determinate.
(4) $\mathrm{ind}_{[a,b]}(S_N) \leq N$.
The following statements are equivalent:
(5) S_N is an interior point of $\mathcal{X}^N[a,b]$.
(6) The Hankel matrices $\underline{H}_N(S_N), \bar{H}_N(S_N)$ are positive definite.
(7) $\det(\underline{H}_n(S_N)) = 0$ and $\det(\bar{H}_n(S_N)) = 0$ for all $n \leq N$.
(8) $\mathrm{ind}_{[a,b]}(S_N) \geq N+1$.

Now we focus on the k-reduced operators. Let's assume $X \in \mathrm{Range}(\mathcal{R}_k)$ is a non-positive operator, then discuss the consequences of S_N being a boundary point of $\mathcal{X}^N[-1,k]$. Theorem 27 tells us that there is only one representing measure of it, which is exactly the atomic measure corresponding to the spectrum of X, there cannot exist another new representing measure letting $S_N \in \mathcal{X}^N[0,k]$. So we can conclude that S_N *which is both $[-1,k]$-moment sequence and $[0,k]$-moment sequence must be an interior point of $\mathcal{X}^N[-1,k]$.*

Suppose the spectrum of X is $\mu_X = (\lambda_1, \cdots, \lambda_D) \subset [-1,k]$, then the index of representing measure μ_X with respect to $[-1,k]$ is $\mathrm{ind}_{[-1,k]}(\mu_X) \leq 2D$, so the index of $S_N(X)$ satisfy $\mathrm{ind}_{[-1,k]}(S_N) \leq 2D$. According to theorem 27, at least when $N = 2D$, $S_N(X)$ must be a boundary point of $\mathcal{X}^N[-1,k]$. Combined with the previous discussion, we can get the conclusion that S_N *generated by X cannot be $[0,k]$-moment sequence when $N \geq 2D$.* So we can formalize the above discussions into the following important theorem.

Theorem 28 (Proof of Corollary 13)**.** *For any non-positive operator $X \in Range(\mathcal{R}_k)$, the moment sequence S_N generated by X cannot be a $[0,k]$-moment sequence when $N \geq 2D$.*

Although the above theorem is correct, but the upper bound of $\text{ind}_{[-1,k]}(S_N)$ is not tight for special cases. Consider the case of k-reduced operator $\mathcal{R}_k(\rho)$ with ρ being a pure state. We have known the structure of the spectrum of the k-reduced operator $\mathcal{R}_k(\rho)$, i.e., the spectrum totally contains at most $2r+1$ different eigenvalues $\{k\lambda_1, \cdots, k\lambda_r, 0, x_1, \cdots, x_r\} \subset [-1, k]$. Then the index of μ_X is bounded as, $\text{ind}_{[-1,k]}(\mu_X) \leq 2(2r+1) = 4r+2$, so we get a much tighter upper bound of the index of S_N as $\text{ind}_{[-1,k]}(S_N) \leq 4r+2$. This fact tells us that *we can always determine if the k-reduced operator of a pure state is a positive operator or not by the moment conditions with order larger than $4r+2$.*

K Detectable regions of Algorithm 2

Proof of Proposition 15. The isotropic state under the k-reduction map writes

$$\mathcal{R}_k(\rho_F) = \frac{k(d-d^{-1})-1+F}{d^2-1} I - \frac{d^2F-1}{d^2-1}|+_d\rangle\langle+_d|. \tag{159}$$

Its non-zero spectrum is

$$\{\lambda_\Delta\}^{d^2-1} \cup \{\lambda_\delta\}, \quad \lambda_\Delta \equiv \frac{F-1}{d^2-1} + \frac{k}{d}, \ \lambda_\delta \equiv \frac{k}{d} - F. \tag{160}$$

Therefore, the moments of $\mathcal{R}_k(\rho_F)$ are

$$q_n = (d^2-1)\lambda_\Delta^n + (-\lambda_\delta)^n. \tag{161}$$

Notice that $\text{SN}(\rho_F) = \lceil dF \rceil$. So if $k \leq \lceil dF \rceil - 1$, then $\lambda_\delta \geq (1 + dF - \lceil dF \rceil)/d > 0$.

Because $k \geq 1$, we have $\lambda_\Delta > 0$, and

$$q_1 q_3 - q_2^2 = -(d^2-1)\lambda_\Delta\lambda_\delta(\lambda_\Delta + \lambda_\delta)^2 < 0, \tag{162}$$

which means the third moment condition always works. □

For the pure state situation, we start with a few properties of the Hankel matrix. Consider the following moment sequence $S = (s_n)_{n\in\mathbb{N}}$ with

$$s_n = \sum_{i=0}^{L-1} m_i \gamma_i^n, \quad m_i \in \mathbb{N}^+, \quad \gamma_i \in \mathbb{R}/\{0\}, \quad \gamma_i \neq \gamma_j. \tag{163}$$

Here $\{\gamma_i\}$ can be regarded as the set of eigenvalues, $\{m_i\}$ are the multiplicities, and L is the total number of distinct eigenvalues. For an odd number N, the Hankel matrix $H(S_{1,N})$ writes

$$[H(S_{1,N})]_{ij} = s_{i+j+1}, \quad i,j = 0, 1, \cdots, \frac{N-1}{2}. \tag{164}$$

Define $d_N \equiv \dim(H(S_{1,N})) = \frac{N+1}{2}$ for simplicity. Then we have

Lemma 29. *Given $H(S_{1,N})$ defined as described before. Then we have (a) rank $(H(S_{1,N})) = \min\{d_N, L\}$; (b) when $d_N \geq L$, $H(S_{1,N})$ is positive semi-definite if and only if $\forall i, \gamma_i > 0$.*

Proof. (a) Let

$$e(\gamma) \equiv (1, \gamma, \gamma^2, \cdots, \gamma^{d_N-1})^\top, \tag{165}$$

then $H(S_{1,N})$ has decomposition:

$$H(S_{1,N}) = \sum_{i=0}^{L-1} m_i\gamma_i e(\gamma_i)e(\gamma_i)^\top. \tag{166}$$

Introduce matrix

$$V \equiv (e(\gamma_1)\ e(\gamma_2)\ \cdots\ e(\gamma_L)). \tag{167}$$

Because all $\{m_i\lambda_i\}$ are non-zero, we have $\text{rank}(H(S_{1,N})) = \text{rank}(V)$.

According to the Vandermonde determinant, if $d_N = L$, then

$$\det(V) = \prod_{i\neq j} |\gamma_i - \gamma_j| \neq 0. \tag{168}$$

Thus, the set of $\{e(\gamma_i)\}$ are linearly independent, and $\text{rank}(V) = L$. The situation is the same for $d_N > L$. So for $d_N \geq L$,

$$\text{rank}(H(S_{1,N})) = \text{rank}(V) = L. \tag{169}$$

If $d_N < L$, then the rank of V is the same with that of

$$(e(\gamma_1)\ e(\gamma_2)\ \cdots\ e(\gamma_{d_N})). \tag{170}$$

The determinant of this matrix is $\prod_{i\neq j\leq d_N} |\gamma_i - \gamma_j| \neq 0$, so V still has full rank. Hence

$$\text{rank}(H(S_{1,N})) = \text{rank}(V) = d_N. \tag{171}$$

(b) If $\forall i, \gamma_i > 0$, then for any real vector c,

$$c^\top H(S_{1,N})c = \sum_{i=1}^{L} m_i\gamma_i |c^\top e(\gamma_i)|^2 \geq 0, \tag{172}$$

so $H(S_{1,N})$ is positive semi-definite.

If there exists a $\gamma_{i^*} < 0$, then in the space $\text{span}\{e(\gamma_i), i = 0, 1, \cdots, L-1\}$, there must exist a v that is orthogonal to $\text{span}\{e(\gamma_i), i \neq i^*\}$ and has a non-zero overlap with $e(\gamma_{i^*})$. Therefore,

$$v^\top H(S_{1,N})v = m_{i^*}\gamma_{i^*}|v^\top e(\gamma_{i^*})|^2 < 0, \tag{173}$$

so $H(S_{1,N})$ is not positive semi-definite. □

Using this lemma, we can prove:

Proof of Proposition 18. Recall that the non-zero spectrum of $\mathcal{R}_k(|\psi\rangle\langle\psi|)$ includes

$$\{k\lambda_j\}^{d_B-1} \cup \{x_j\}, \tag{174}$$

where $\{x_j\}$ are the solutions of

$$\sum_{j=0}^{r-1} \frac{\lambda_j}{k\lambda_j - x} = 1. \tag{175}$$

Because $\{\lambda_j\}$ are distinct to each other and $k < \mathrm{SN}(\psi)$, from Theorem 4 we obtain

$$x_{j_1} \neq x_{j_2}, \quad x_j \notin \{k\lambda_j\}, \quad -1 < \min_j x_j < 0. \tag{176}$$

Therefore,

$$q_n = (d_B - 1)\sum_{j=0}^{r-1}(k\lambda_j)^n + \sum_{j=0}^{r-1} x_j^n. \tag{177}$$

In the notation of Lemma 29, it means

$$\begin{aligned} &L = 2r, \quad \gamma_i = k\lambda_i, \quad \gamma_{i+r} = x_i, \\ &m_i = d_B - 1, \quad m_{i+r} = 1, \ i = 1, 2, \cdots, r. \end{aligned} \tag{178}$$

Since there exists a negative γ_i, when $d_N \geq L = 2r$, the Hankel matrix is not positive semi-definite.

When $d_N \leq r$, let H_1 be the Hankel matrix generated by $s_n^{(1)} \equiv \sum_{j=0}^{r-1}(k\lambda_j)^n$, let H_2 be the Hankel matrix generated by $s_n^{(2)} \equiv \sum_{j=0}^{r-1} x_j^n$. Thus, $B_N[\psi, k] = (d_B - 1)H_1 + H_2$. Because it is required that $\{\lambda_j\}$ are distinct and positive, we have $H_1 \succeq 0$ and $\mathrm{rank}(B_N) = \mathrm{rank}(H_1) = d_N$. Thus, all the eigenvalues of H_1 are non-zero.

The smallest eigenvalue of $B_N[\psi, k]$ has lower bound

$$(d_B - 1)\lambda_{\min}(H_1) + \lambda_{\min}(H_2). \tag{179}$$

Furthermore, suppose x_{r-1} is the only negative eigenvalue of $\mathcal{R}_k(|\psi\rangle\langle\psi|)$, then we have lower bound:

$$\begin{aligned} \lambda_{\min}(H_2) &> \min_{v \in \mathbb{R}^{d_N}, \|v\|_2 = 1} x_{r-1}|v^\top e(x_{r-1})|^2 > d_N \cdot x_{r-1} \\ &= -\frac{N+1}{2}\mathcal{N}_k(\psi). \end{aligned} \tag{180}$$

Hence, when

$$d_B > 1 + \frac{(N+1)\mathcal{N}_k(\psi)}{2\lambda_{\min}(H_1)}, \tag{181}$$

the matrix $B_N[\psi, k]$ must be positive.

□

Proof of Proposition 17. The non-zero spectrum of $\mathcal{R}_k(|+_r\rangle\langle+_r|)$ is

$$\left\{\frac{k}{r}\right\}^{d_B r - 1} \cup \left\{\frac{k}{r} - 1\right\}. \tag{182}$$

Therefore, in the notation of Lemma 29

$$\begin{aligned} &L = 2, \quad \gamma_1 = \frac{k}{r}, \quad \gamma_2 = \frac{k}{r} - 1, \\ &m_1 = d_B r - 1, \quad m_2 = 1. \end{aligned} \tag{183}$$

When $1 \leq k \leq r - 1$, we have $\gamma_1 > 0, \gamma_2 < 0$, so the Hankel matrix is not positive semi-definite. □

L Sample complexity of Algorithm 2 with $N^* = 3$

Here are a few notations for asymptotic analysis in complexity theory. If a real function $f(x)$ satisfies

$$\lim_{x \to \infty} \frac{|f(x)|}{x^n} < \infty \tag{184}$$

then $f(x) = \mathcal{O}(x^n)$. If the above limit is not zero, then $f(x) = \Omega(x^n)$. If $f(x) = \mathcal{O}(x^n)$ and $f(x) = \Omega(x^n)$, then $f(x) = \Theta(x^n)$. On the other hand, if

$$\lim_{x \to \infty} \frac{|f(x)|}{x^n} = 0, \tag{185}$$

then $f(x) = o(x^n)$. If $f(x) = \mathcal{O}(x^n), g(x) = \mathcal{O}(x^m)$, then $f(x) + g(x) = \mathcal{O}(x^{\max\{m,n\}})$, and $f(x)g(x) = \mathcal{O}(x^{m+n})$; if $f(x) = \mathcal{O}(x^n), g(x) = o(x^n)$, then $f(x)+g(x) = \mathcal{O}(x^n)$.

In order to computer the variance of estimators, we need to estimate the average over k-fold random unitaries. Using Weingarten calculus, we obtain

$$\Phi^{(k)}(\cdot) \equiv \int dU U^{\otimes k}(\cdot)(U^\dagger)^{\otimes k} = \sum_{\pi,\sigma \in S_k} C_{\pi,\sigma}\,\mathrm{Tr}(\mathbb{W}_\pi \cdot)\mathbb{W}_\sigma, \tag{186}$$

where the matrix C is

$$Q_{\pi,\sigma} \equiv d^{\#\mathrm{cycles}(\pi\sigma)}, \quad C \equiv Q^{-1} \tag{187}$$

and d is the dimension of each U. The exponential index $\#\mathrm{cycles}(\pi)$ is how many cycles the permutation π can be decomposed into. For example, in S_3 group, the identity can be written as $(1)(2)(3)$, so it has 3 cycles; (123) itself is a cycle, so it has 1 cycle.

For example,

$$\begin{aligned} \Phi^{(2)}(X) =& \frac{1}{D^2 - 1}[I\,\mathrm{Tr}(X) + \mathbb{W}\,\mathrm{Tr}(\mathbb{W}X) \\ &- D^{-1}\mathbb{W}\,\mathrm{Tr}(X) - D^{-1}I\,\mathrm{Tr}(\mathbb{W}X)]. \end{aligned} \tag{188}$$

Here we use $\mathbb{W}$ to represent $\mathbb{W}_{(12)}$ for simplicity.

The following lemma is useful in our analysis.

Lemma 30. *Given $d, k \in \mathbb{N}^+$, the entries of C defined in Eq. (187) have bounds:*

$$|C_{\pi,\pi^{-1}} - d^{-k}| \leq 2d^{-k-1}; \quad |C_{\pi,\sigma}| \leq 2d^{-k-1}, \ \sigma \neq \pi^{-1}. \tag{189}$$

Proof. Let

$$Q_0 \equiv d^k \sum_{\pi \in S_k} \delta_{\pi,\pi^{-1}}, \quad Q_1 \equiv Q - Q_0. \tag{190}$$

For all $\pi\sigma \neq 1$, $d^{\#\mathrm{cycles}(\pi\sigma)} \leq d^{k-1}$. Thus $\|Q_1\|_\infty \leq d^{k-1}$, and

$$\|Q_0^{-1/2} Q_1 Q_0^{-1/2}\|_\infty \leq \|Q_0^{-1}\|_\infty \|Q_1\|_\infty \leq d^k \cdot d^{k-1} = d^{-1}. \tag{191}$$

Because Q_0 is a real symmetric reversible, Q_1 is real symmetric, and $\|Q_0^{-1/2}Q_1Q_0^{-1/2}\|_\infty < 1$, we have

$$\begin{aligned} C &= (Q_0+Q_1)^{-1} \\ &= Q_0^{-1/2}(I+Q_0^{-1/2}Q_1Q_0^{-1/2})^{-1}Q_0^{-1/2} \\ &= Q_0^{-1/2}\sum_{n=0}^{\infty}(-Q_0^{-1/2}Q_1Q_0^{-1/2})^n Q_0^{-1/2} \\ &= Q_0^{-1/2}\sum_{n=0}^{\infty}(-Q_0^{-1/2}Q_1Q_0^{-1/2})^n Q_0^{-1/2} \\ &= Q_0^{-1} - Q_0^{-1}Q_1Q_0^{-1} + \cdots . \end{aligned} \tag{192}$$

Finally,

$$\begin{aligned} \|C - Q_0^{-1}\|_\infty &\le \sum_{n=1}^{\infty}\|Q_1\|_\infty^n \|Q_0^{-1}\|_\infty^{n+1} \le \sum_{n=1}^{\infty}(d^{k-1})^n (d^{-k})^{n+1} \\ &= d^{-k}\sum_{n=1}^{\infty} d^{-n} \le 2d^{-k-1}. \end{aligned} \tag{193}$$

It is equivalent to

$$|C_{\pi,\pi^{-1}} - d^{-k}| \le 2d^{-k-1}; \quad |C_{\pi,\sigma}| \le 2d^{-k-1},\ \sigma \ne \pi^{-1}. \tag{194}$$

□

In the following paragraphs, we will show that given access to unitary-3 designs, the sample complexity of estimating $\mathrm{Tr}(\rho^2), \mathrm{Tr}(\rho^3), \mathrm{Tr}(\rho^2\cdot\rho_A\otimes I_B)$ using the classical shadow methods are $\Omega(D), \Omega(D^{1/3}), \Omega(D^{1/2})$ separately. Here D is the dimension of ρ. Thus, if we use the statistical correlation method to estimate $\mathrm{Tr}(\rho^2)$ [20], and the classical shadow method to estimate the other two quantities, then the sample complexity of the third moment k-reduction criterion by randomized measurement is $\Omega(D^{1/2})$.

Proof of Theorem 14. In the classical shadow protocol, we first sample a random unitary U from ensemble $\mathcal{U}$, then apply U on the target state ρ and measure the rotated state in the standard basis to obtain $|b\rangle$. Let $\hat{\Psi}$ be the random variable whose value is $U|b\rangle\langle b|U^\dagger$, then

$$P[\hat{\Psi} = U|b\rangle\langle b|U^\dagger] = \frac{1}{|\mathcal{U}|}\langle b|U^\dagger\rho U|b\rangle. \tag{195}$$

After computation, the expectation value is

$$\begin{aligned} \mathbb{E}[\hat{\Psi}] &= \frac{1}{|\mathcal{U}|}\sum_{U\in\mathcal{U}}\sum_{|b\rangle} U|b\rangle\langle b|U^\dagger\cdot\langle b|U\rho U^\dagger|b\rangle \\ &= \sum_{|b\rangle}\mathrm{Tr}_2\left[\frac{1}{|\mathcal{U}|}\sum_{U\in\mathcal{U}} U^{\otimes 2}|b\rangle\langle b|^{\otimes 2}(U^\dagger)^{\otimes 2}\cdot I\otimes\rho\right] \\ &= \sum_{|b\rangle}\mathrm{Tr}_2\left[\Phi^{(2)}\left(|b\rangle\langle b|^{\otimes 2}\right)\cdot I\otimes\rho\right] \\ &= \sum_{|b\rangle}\mathrm{Tr}_2\left[\frac{1}{D(D+1)}(I+\mathbb{W})\cdot I\otimes\rho\right] \\ &= \frac{\rho+I}{D+1}. \end{aligned} \tag{196}$$

where $\mathcal{U}$ forms a unitary-2 design. Therefore, the estimator for the state is

$$\hat{\rho} \equiv (D+1)\hat{\Psi} - I, \tag{197}$$

and

$$\begin{aligned} \mathbb{E}[\hat{\rho}\otimes\hat{\rho}] =& (D+1)^2\mathbb{E}[\hat{\Psi}\otimes\hat{\Psi}] - (D+1)\mathbb{E}[\hat{\Psi}]\otimes I \\ &- (D+1)I\otimes\mathbb{E}[\hat{\Psi}] + I\otimes I. \end{aligned} \tag{198}$$

To compute the variances, we also need $\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]$, which can be computed from

$$\begin{aligned} \mathbb{E}[\hat{\Psi}\otimes\hat{\Psi}] &= \frac{1}{|\mathcal{U}|}\sum_{U\in\mathcal{U}}\sum_{|b\rangle} U|b\rangle\langle b|U^\dagger\otimes U|b\rangle\langle b|U^\dagger\cdot\langle b|U\rho U^\dagger|b\rangle, \\ &= \sum_{|b\rangle}\mathrm{Tr}_3\left[\frac{1}{|\mathcal{U}|}\sum_{U\in\mathcal{U}} U^{\otimes 3}|b\rangle\langle b|^{\otimes 3}(U^\dagger)^{\otimes 3}\cdot I\otimes I\otimes\rho\right] \\ &= \sum_{|b\rangle}\mathrm{Tr}_3\left[\Phi^{(3)}\left(|b\rangle\langle b|^{\otimes 3}\right)\cdot I\otimes I\otimes\rho\right] \\ &= \frac{2}{(D+1)(D+2)}[I\otimes I + \rho\otimes I + I\otimes\rho]\Pi_s. \end{aligned} \tag{199}$$

when$\mathcal{U}$ forms a unitary-3 design. Here $\Pi_s = (I+\mathbb{W})/2$. The derivation of the last step can be found in Lemma 14 of [59]. Accordingly,

$$\mathbb{E}[\hat{\rho}\otimes\hat{\rho}] = \frac{D+1}{D+2}(I\otimes I + I\otimes\rho + \rho\otimes I)\mathbb{W} - \frac{1}{D+2}(\rho\otimes I + I\otimes\rho + I\otimes I).$$

Let $R_\rho \equiv I\otimes I + I\otimes\rho + \rho\otimes I$. Then the expectation value can be simplified to

$$\mathbb{E}[\hat{\rho}\otimes\hat{\rho}] = \frac{D+1}{D+2}R_\rho\mathbb{W} - \frac{1}{D+2}R_\rho. \tag{200}$$

(a). *Variance of* $\mathrm{Tr}(\rho^2)$ *by the classical shadow.* Label the state estimator of each run as $\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_M$ separately, where M is the total number of samples. Then the estimator for the purity is (recall that $p_n \equiv \mathrm{Tr}(\rho^n)$):

$$\hat{p}_2 = \binom{M}{2}^{-1}\sum_{j<k}\mathrm{Tr}(\hat{\rho}_j\hat{\rho}_k). \tag{201}$$

Its variance $\mathrm{Var}[\hat{p}_2]$ can be computed by

$$\begin{aligned} \mathrm{Var}[\hat{p}_2] &= \mathbb{E}[\hat{p}_2^2] - p_2^2 \\ &= \binom{M}{2}^{-2}\sum_{j_1<k_1, j_2<k_2}\mathbb{E}\left[\mathrm{Tr}(\hat{\rho}_{j_1}\otimes\hat{\rho}_{j_2}\cdot\hat{\rho}_{k_1}\otimes\hat{\rho}_{k_2}) - p_2^2\right] \\ &= \binom{M}{2}^{-1}\left\{2(M-2)\left[\mathrm{Tr}(\rho\otimes\rho\cdot\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]) - p_2^2\right]\right. \\ &\quad \left. + \left[\mathrm{Tr}\left(\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]^2\right) - p_2^2\right]\right. \\ &\le \frac{4(M-2)}{M(M-1)}\mathrm{Tr}(\rho\otimes\rho\cdot\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]) \\ &\quad + \frac{2}{M(M-1)}\mathrm{Tr}\left(\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]^2\right). \end{aligned}$$

Using

$$\begin{aligned}\mathrm{Tr}(R_\rho \mathbb{W} R_\rho \mathbb{W}) &= \mathrm{Tr}\big(R_\rho^2\big) = D^2+4D+2Dp_2+2,\\ \mathrm{Tr}\big(R_\rho^2 \mathbb{W}\big) &= D+4+4p_2,\end{aligned} \tag{202}$$

we obtain

$$\begin{aligned}\mathrm{Tr}(\rho\otimes\rho\cdot\mathbb{E}[\hat\rho\otimes\hat\rho]) &= \frac{D+1}{D+2}(p_2+2p_3)-\frac{1}{D+2}(1+2p_2)\\ &\le p_2+2p_3,\\ \mathrm{Tr}\big(\mathbb{E}[\hat\rho\otimes\hat\rho]^2\big) &= (D+1)^2-\frac{D^2+2(D+1)(D+4)}{(D+2)^2}\\ &\quad+2\frac{D^2-2}{D+2}p_2\\ &\le (D+1)^2+2Dp_2.\end{aligned}$$

Hence,

$$\mathrm{Var}[\hat p_2]\le \frac{4(M-2)}{M(M-1)}\cdot(p_2+2p_3)+\frac{2}{M(M-1)}((D+1)^2+2Dp_2).$$

To guarantee that $\mathrm{Var}[\hat p_2]=\mathcal{O}(1)$, $M=\Omega(D)$ number of samples suffices.

(b). *Variance of* $\mathrm{Tr}\big(\rho^3\big)$ *by classical shadow.* The third moment p_3 has estimator

$$\hat p_3 = \binom{M}{3}^{-1}\sum_{i<j<k}\mathrm{Tr}\left(\hat\rho_i\hat\rho_j\hat\rho_k\right). \tag{203}$$

Its variance is

$$\begin{aligned}\mathrm{Var}[\hat p_3^2] =& \binom{M}{3}^{-2}\sum_{i_1<j_1<k_1,i_2<j_2<k_2}\mathbb{E}[\mathrm{Tr}(\hat\rho_{i_1}\hat\rho_{j_1}\hat\rho_{k_1})\\ &\mathrm{Tr}(\hat\rho_{i_2}\hat\rho_{j_2}\hat\rho_{k_2})-p_3^2]\\ =& \binom{M}{3}^{-2}\sum_{i_1<j_1<k_1,i_2<j_2<k_2}\mathbb{E}[\mathrm{Tr}\left((\hat\rho_{i_1}\otimes\hat\rho_{i_2})\cdot\right.\\ &\left.(\hat\rho_{j_1}\otimes\hat\rho_{j_2})\cdot(\hat\rho_{k_1}\otimes\hat\rho_{k_2})-p_3^2\right]\end{aligned}$$

Now we analyze the expectation value of each summand with different indices. Based on how many pairs of $\hat\rho$ sharing the same index, we can classify the summand into 4 types:

1. all indices are distinct;
2. there exists exactly a pair of identical indices;
3. there exist exactly two pairs of identical indices;
4. there exist three pairs of identical indices.

Type 1, the summand equals 0.

Type 2, the summand equals

$$S_2 = \mathrm{Tr}\big(\mathbb{E}[\hat\rho\otimes\hat\rho]\cdot\rho^2\otimes\rho^2\big)-p_3^2. \tag{204}$$

For fixed $i_1=i_2=m$, then there are totally $\binom{M-m}{2}$ different pairs of (j,k). So there are totally

$$\sum_{m=1}^{M-2}\binom{M-m}{2}\left[\binom{M-m}{2}-1\right]=\mathcal{O}(M^5) \tag{205}$$

number of summands of this type.

Type 3, the summand equals

$$S_3 = \mathrm{Tr}\big(\mathbb{E}[\hat\rho\otimes\hat\rho]^2\cdot\rho\otimes\rho\big)-p_3^2, \tag{206}$$

and there are totally

$$\sum_{m=1}^{M}\sum_{m'=m+1}^{M}(M-m'+1)(M-m')=\mathcal{O}(M^4) \tag{207}$$

number of summands of this type.

Type 4, the summand equals

$$S_4 = \mathrm{Tr}\big(\mathbb{E}[\hat\rho\otimes\hat\rho]^3\big)-p_3^2, \tag{208}$$

and there are totally $\binom{M}{3}=\mathcal{O}(M^3)$ number of summands of this type.

Therefore,

$$\mathrm{Var}[\hat p_3^2]=\mathcal{O}(M^{-1})S_2+\mathcal{O}(M^{-2})S_3+\mathcal{O}(M^{-3})S_4. \tag{209}$$

After computation, these summands have magnitudes:

$$\begin{aligned}S_2 \le& \frac{D+1}{D+2}(p_4+2p_5)-\frac{1}{D+2}(p_2^2+2p_2p_3)-p_3^2\\ =&\mathcal{O}(1),\\ S_3 \le& \frac{1+(D+1)^2}{(D+2)^2}\,\mathrm{Tr}\big[R_\rho^2\cdot\rho\otimes\rho\big]\\ =&\frac{1+(D+1)^2}{(D+2)^2}(1+4p_2+2p_3+2p_2^2)\\ =&\mathcal{O}(1),\\ S_4 \le& \frac{1}{(D+2)^3}[(D+1)^3\,\mathrm{Tr}(R_\rho\mathbb{W}R_\rho\mathbb{W}R_\rho\mathbb{W})\\ &+3(D+1)\,\mathrm{Tr}\big(R_\rho^3\mathbb{W}\big)]\\ =&\frac{(D+1)^3+3(D+1)}{(D+2)^3}(D+6+12p_2+8p_3)\\ =&\mathcal{O}(D).\end{aligned} \tag{210}$$

Hence,

$$\mathrm{Var}[\hat p_3]=\mathcal{O}(M^{-1})+\mathcal{O}(M^{-2})+\mathcal{O}(M^{-3}D). \tag{211}$$

To guarantee that $\mathrm{Var}[\hat p_3]=\mathcal{O}(1)$, $M=\Omega(D^{1/3})$ number of samples suffices.

(c). *Variance of* $p_{1,2}$ *by classical shadow.* In order to estimate $p_{1,2}$, we need to construct classical shadow for $\rho_A\otimes I_B$ as well. Denote the estimator of $\rho_A\otimes I_B$ as $\hat a$, and suppose there are totally L of samples, then the estimator of $p_{1,2}$ is

$$\hat p_{1,2} = L^{-1}\binom{M}{2}^{-1}\sum_{n=0}^{L-1}\sum_{i<j}^{M-1}\mathrm{Tr}\left(\hat a_n\hat\rho_i\hat\rho_j\right) \tag{212}$$

Similarly, its variance is

$$\begin{aligned}\mathrm{Var}[\hat p_{1,2}] =& L^{-2}\binom{M}{2}^{-2}\sum_{n_1,n_2}^{L-1}\sum_{i_1<j_1,i_2<j_2}^{M-1}\\ &\mathbb{E}[\mathrm{Tr}\left(\hat a_{n_1}\hat\rho_{i_1}\hat\rho_{j_1}\right)\mathrm{Tr}\left(\hat a_{n_2}\hat\rho_{i_2}\hat\rho_{j_2}\right)-p_{1,2}^2].\end{aligned} \tag{213}$$

The summand in RHS can be rewritten as

$$\mathrm{Tr}\left(\mathbb{E}\left[\hat{a}_{n_1}\otimes\hat{a}_{n_2}\right]\cdot\mathbb{E}[\hat{\rho}_{i_1}\hat{\rho}_{j_1}\otimes\hat{\rho}_{i_2}\hat{\rho}_{j_2}]\right)-p_{1,2}^2. \tag{214}$$

It has been proved that

$$\begin{aligned}
&\sum_{n_1,n_2}^{L-1}\mathbb{E}\left[\hat{a}_{n_1}\otimes\hat{a}_{n_2}\right]\\
=&L(L-1)\rho_A\otimes I_B\otimes\rho_A\otimes I_B+L\mathbb{E}[\hat{a}\otimes\hat{a}]\\
=&L(L-1)\rho_A\otimes I_B\otimes\rho_A\otimes I_B\\
&+L\left(\frac{D_A+1}{D_A+2}R_{\rho_A}\mathbb{W}_A-\frac{1}{D_A+2}R_{\rho_A}\right)\otimes I_B\otimes I_B,\\
&\sum_{i_1<j_1,i_2<j_2}^{M-1}\mathbb{E}[\hat{\rho}_{i_1}\hat{\rho}_{j_1}\otimes\hat{\rho}_{i_2}\hat{\rho}_{j_2}]\\
=&\binom{M}{2}^2\rho^2\otimes\rho^2+\binom{M}{2}\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]^2\\
&+\binom{M}{2}(M-2)\{\rho\otimes\rho\cdot\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]+\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]\cdot\rho\otimes\rho\},
\end{aligned}$$

Here $\mathcal{R}_{\rho_A}:=I_A\otimes I_A+I_A\otimes\rho_A+\rho_A\otimes I_A$ and $\mathbb{W}_A$ is the swap gate between two Hilbert spaces $\mathcal{H}_A$.

After computation, we obtain

$$\begin{aligned}
&\mathrm{Var}[\hat{p}_{1,2}]\\
=&\mathcal{O}(L^{-1})\,\mathrm{Tr}\big(R_{\rho_A}\mathbb{W}_A\otimes I_B^{\otimes 2}\cdot\rho^2\otimes\rho^2\big)\\
&+\mathcal{O}(M^{-2})\,\mathrm{Tr}\big(\rho_A^{\otimes 2}\otimes I_B^{\otimes 2}\cdot\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]^2\big)\\
&+\mathcal{O}(L^{-1}M^{-2})\,\mathrm{Tr}\big(R_{\rho_A}\mathbb{W}_A\otimes I_B^{\otimes 2}\cdot\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]^2\big)\\
&+\mathcal{O}(L^{-1})\,\mathrm{Tr}\big(\rho_A^{\otimes 2}\otimes I_B^{\otimes 2}\cdot\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]\cdot\rho\otimes\rho\big)\\
&+\mathcal{O}(M^{-1}L^{-1})\,\mathrm{Tr}\big(R_{\rho_A}\mathbb{W}_A\otimes I_B^{\otimes 2}\cdot\mathbb{E}[\hat{\rho}\otimes\hat{\rho}]\cdot\rho\otimes\rho\big)\\
=&\mathcal{O}(L^{-1})+\mathcal{O}(M^{-2}D_B^2)+\mathcal{O}(L^{-1}M^{-2}D_B^2D_A)\\
&+\mathcal{O}(L^{-1})+\mathcal{O}(M^{-1}L^{-1}).
\end{aligned}\tag{215}$$

Therefore,

$$\mathrm{Var}[\hat{p}_{12}]=\mathcal{O}(M^{-2}D_B^2)+\mathcal{O}(L^{-1}M^{-2}D_B^2D_A). \tag{216}$$

To guarantee that $\mathrm{Var}[\hat{p}_{12}]=\mathcal{O}(1)$, we need

$$M=\Omega(D_B),\quad LM^2=\Omega(D_B^2D_A). \tag{217}$$

Thus, when we choose $L=\Omega(D_A), M=\Omega(D_B)$, $L+M=\Omega(D_A+D_B)$ number of samples suffices. If $D_A=D_B=\sqrt{D}$, then the sample complexity is $\Omega(D^{\frac{1}{2}})$.

(d) *Variance of* $\mathrm{Tr}\big(\rho^2\big)$ *by statistical correlation.* We give a self-consistent analysis for the sample complexity of estimating $\mathrm{Tr}\big(\rho^2\big)$ with statistical correlation method as well. Here is the protocol for estimating $\mathrm{Tr}\big(\rho^2\big)$ using statistical correlation:

- sample a sequence of random unitaries $\{U_1,U_2,\cdots,U_N\}$;
- for each U_n, apply it to ρ then measure state $U_n\rho U_n^\dagger$ in the standard basis;
- repeat the measurement for M times to obtain $\{|s_{n,1}\rangle,|s_{n,2}\rangle,\cdots,|s_{n,M}\rangle\}$;
- output

$$\frac{1}{N}\sum_{n=0}^{N-1}\frac{2}{M(M-1)}\sum_{m_1<m_2}^{M}\Xi(s_{n,m_1},s_{n,m_2}) \tag{218}$$

where

$$\Xi(s_1,s_2)\equiv(D+1)\delta_{s_1,s_2}-1. \tag{219}$$

Define

$$\hat{\Xi}\equiv\sum_{s_1,s_2}\Xi(s_1,s_2)|s_1,s_2\rangle\langle s_1,s_2|. \tag{220}$$

For each U, the probability of obtaining measurement outcome s is

$$P(s|U)=\langle s|U\rho U^\dagger|s\rangle. \tag{221}$$

Notice that the average over measurement outcomes gives

$$\begin{aligned}
\mathbb{E}_M[\Xi(s_1,s_2)]&=\sum_{s_1,s_2}\Xi(s_1,s_2)P(s_1|U)P(s_2|U)\\
&=\mathrm{Tr}\Big[(U\rho U^\dagger)^{\otimes 2}\hat{\Xi}\Big].
\end{aligned}\tag{222}$$

Denote the estimator for $\mathrm{Tr}\Big[(U_n\rho U_n^\dagger)^{\otimes 2}\hat{\Xi}\Big]$ as T_n (the summand of Eq. (218) with respect to n), thus the variance of $\mathrm{Tr}\big(\rho^2\big)$ is

$$\frac{1}{N}\mathbb{E}_U[\mathbb{E}_M[T_n^2]-\mathbb{E}_M[T_n]^2]. \tag{223}$$

Now we expand T_n^2 as

$$\begin{aligned}
T_n^2=&\frac{1}{M^2(M-1)^2}\sum_{m_1\neq m_2,m_3\neq m_4}\Xi(s_{m_1},s_{m_2})\Xi(s_{m_3},s_{m_4})\\
=&(M^{-4})\sum_{m_1\neq m_2\neq m_3\neq m_4}\Xi(s_{m_1},s_{m_2})\Xi(s_{m_3},s_{m_4})\\
&+(M^{-4})\sum_{m_1\neq m_2\neq m_3}\Xi(s_{m_1},s_{m_2})\Xi(s_{m_1},s_{m_3})\\
&+(M^{-4})\sum_{m_1\neq m_2}\Xi^2(s_{m_1},s_{m_2}),
\end{aligned}\tag{224}$$

$$\begin{aligned}
\mathbb{E}_M[T_n^2]=&(1)\,\mathrm{Tr}\Big((U_n\rho U_n^\dagger)^{\otimes 2}\hat{\Xi}\Big)^2+(M^{-1})\,\mathrm{Tr}\Big((U_n\rho U_n^\dagger)^{\otimes 3}\hat{\Xi}'\Big)\\
&+(M^{-2})\,\mathrm{Tr}\Big((U_n\rho U_n^\dagger)^{\otimes 2}\hat{\Xi}^2\Big)
\end{aligned}\tag{225}$$

where

$$\langle s_1,s_2,s_3|\hat{\Xi}'|s_1,s_2,s_3\rangle=\Xi(s_1,s_2)\Xi(s_1,s_3). \tag{226}$$

When we compute $\mathbb{E}_M[T_n^2]]-\mathbb{E}_M[T_n]^2$, we can ignore the (1) term because it always cancels with $\mathbb{E}_M[T_n]^2=\mathrm{Tr}\Big((U_n\rho U_n^\dagger)^{\otimes 2}\hat{\Xi}\Big)^2$. Hence,

$$\begin{aligned}
&\mathbb{E}_M[T_n^2]]-\mathbb{E}_M[T_n]^2\\
=&(M^{-1})\left[\mathrm{Tr}\Big((U\rho U^\dagger)^{\otimes 3}\hat{\Xi}'\Big)-\mathrm{Tr}\Big((U_n\rho U_n^\dagger)^{\otimes 2}\hat{\Xi}\Big)^2\right]\\
&+(M^{-2})[\mathrm{Tr}\Big((U\rho U^\dagger)^{\otimes 2}\hat{\Xi}^2\Big)\\
&-\mathrm{Tr}\Big((U_n\rho U_n^\dagger)^{\otimes 2}\hat{\Xi}\Big)^2]
\end{aligned}\tag{227}$$

$$
\begin{aligned}
&\mathbb{E}_U[\mathbb{E}_M[T_n^2]] - \mathbb{E}_U[\mathbb{E}_M[T_n]^2] \\
=&(M^{-1})\operatorname{Tr}\left(\mathbb{E}_U(U\rho U^\dagger)^{\otimes 3}\hat{\Xi}'\right) + (M^{-2})\operatorname{Tr}\left(\mathbb{E}_U(U\rho U^\dagger)^{\otimes 2}\hat{\Xi}^2\right) \\
=&(M^{-1})\operatorname{Tr}\left[\Phi^{(3)}(\hat{\Xi}')\rho^{\otimes 3}\right] + \mathcal{O}(M^{-2})\operatorname{Tr}\left[\Phi^{(2)}(\hat{\Xi}^2)\rho^{\otimes 2}\right]
\end{aligned}
$$

The last equality holds as long as the unitary ensemble forms a unitary-3 design.

Using Lemma 30, we obtain

$$
\operatorname{Tr}\left[\Phi^{(2)}(\hat{\Xi}^2)\rho^{\otimes 2}\right] = \left(D^{-2}\max_{\pi\in S_2}\operatorname{Tr}\left(\hat{\Xi}^2\mathbb{W}_\pi\right)\right), \tag{228}
$$

$$
\operatorname{Tr}\left[\Phi^{(3)}(\hat{\Xi}')\rho^{\otimes 3}\right] = \left(D^{-3}\max_{\pi\in S_3}\operatorname{Tr}\left(\hat{\Xi}'\mathbb{W}_\pi\right)\right). \tag{229}
$$

The values of the traces are:

$$
\operatorname{Tr}\left(\hat{\Xi}^2\right) = \sum_{s_1,s_2}\Xi^2(s_1,s_2) = D(D^2+D-1), \tag{230}
$$

$$
\operatorname{Tr}\left(\hat{\Xi}^2\mathbb{W}_{(12)}\right) = \sum_{s}\Xi^2(s,s) = D^3, \tag{231}
$$

$$
\operatorname{Tr}\left(\hat{\Xi}'\right) = \sum_{s_1,s_2,s_3}\Xi(s_1,s_2)\Xi(s_1,s_3) = D, \tag{232}
$$

$$
\operatorname{Tr}\left(\hat{\Xi}'\mathbb{W}_{(12)}\right) = \sum_{s_1,s_2}\Xi(s_1,s_1)\Xi(s_1,s_2) = D, \tag{233}
$$

$$
\operatorname{Tr}\left(\hat{\Xi}'\mathbb{W}_{(123)}\right) = \sum_{s}\Xi(s,s)\Xi(s,s) = D^3. \tag{234}
$$

Thus,

$$
\operatorname{Tr}\left[\Phi^{(2)}(\hat{\Xi}^2)\rho^{\otimes 2}\right] = (D), \quad \operatorname{Tr}\left[\Phi^{(3)}(\hat{\Xi}')\rho^{\otimes 3}\right] = (1). \tag{235}
$$

The variance has upper bound

$$
\operatorname{Var}[\hat{p}_2] = \operatorname{Var}\left[\frac{1}{N}\sum_{n=0}^{N-1}T_n\right] = \frac{1}{NM^2}(D) + \frac{1}{NM}(1). \tag{236}
$$

If we choose $N = \Omega(1), M = \Omega(D^{1/2})$, then $NM = \Omega(D^{1/2})$ number of samples suffices to guarantee that $\operatorname{Var}[\hat{p}_2] = \mathcal{O}(1)$. □

M Detectable regions of the correlation matrix method

In this section we study what type states with Schmidt number r can be certified by criterion Eq. (20). When $d_A = d_B = d$, we can choose $\{\sigma_j^{(A)}\}$ and $\{\sigma_k^{(B)}\}$ to be the same set of operator bases. Therefore, we will drop the upper indices and only use $\{\sigma_j\}$ to denote the basis operators for simplicity.

Proof of Proposition 16. We start with a simpler situation where the state is the maximally entangled state $|+_d\rangle$, then

$$
\begin{aligned}
T_{jk} =& \frac{1}{d}\langle +_d|\sigma_j\otimes\sigma_k|+_d\rangle \\
=& \frac{1}{d^2}\sum_{m,n=0}^{d-1}\langle mm|\sigma_j\otimes\sigma_k|nn\rangle \\
=& \frac{1}{d^2}\operatorname{Tr}\left(\sigma_j\sigma_k^\top\right).
\end{aligned} \tag{237}
$$

Clearly, T is diagonal with diagonal entries belonging to $\{\pm 1/d, \pm i/d\}$. Thus, the singular values of T are all d^{-1}, and

$$
\|T\|_1 = (d^2-1)\cdot d^{-1} = d - d^{-1}. \tag{238}
$$

Now we use this result to analyze the isotropic state. Because I does mot contribute the correlation matrices, the correlation matrix of the isotropic state is proportional to that of $|+_d\rangle$, which is

$$
T_{jk} = \frac{d^2F-1}{d(d^2-1)}\langle +_d|\sigma_j\otimes\sigma_k|+_d\rangle. \tag{239}
$$

The singular values of T are all $\frac{d^2F-1}{d(d^2-1)}$. □

Pure state. Recall that for a system with local dimensions $d_A = d_B = d$, the correlation matrix of $|\psi\rangle = \sum_{n=0}^{r-1}\sqrt{\lambda_n}|n\rangle_A\otimes|n\rangle_B$ writes

$$
T_{jk} = \frac{1}{d}\sum_{m,n=0}^{r-1}\sqrt{\lambda_m}\sqrt{\lambda_n}\operatorname{Tr}(|n\rangle_A\langle m|\otimes|n\rangle_B\langle m|\cdot\sigma_j\otimes\sigma_k). \tag{240}
$$

Rewrite it as

$$
T_{jk} = \sum_{m,n=1}^{r}\sqrt{\lambda_m\lambda_n}\frac{\langle m|\sigma_j|n\rangle}{\sqrt{d}}\frac{\langle m|\sigma_k|n\rangle}{\sqrt{d}}. \tag{241}
$$

Using the fact that $|m\rangle\langle n| = \frac{1}{d}\sum_\sigma\langle m|\sigma|n\rangle\sigma$, we obtain

$$
\frac{1}{d}\sum_\sigma\langle m|\sigma|n\rangle\langle m'|\sigma|n'\rangle = \delta_{mm'}\delta_{n,n'}, \tag{242}
$$

$$
\delta_{mn}\delta_{m'n'} + \sum_{\sigma\neq I}\langle m|\sigma|n\rangle\langle m'|\sigma|n'\rangle = d\delta_{mm'}\delta_{nn'}. \tag{243}
$$

Define

$$
|B_{m,n}\rangle \equiv \sum_{j=1}^{d^2-1}\frac{\langle m|\sigma_j|n\rangle^*}{\sqrt{d}}|e_j\rangle, \tag{244}
$$

$$
\hat{J}c \equiv c^*, \quad \hat{J}|B_{m,n}\rangle = \sum_{j=1}^{d^2-1}\frac{\langle m|\sigma_j|n\rangle}{\sqrt{d}}|e_j\rangle. \tag{245}
$$

then

$$
\begin{aligned}
T =& \sum_{m,n=1}^{r}\sqrt{\lambda_m\lambda_n}\hat{J}|B_{m,n}\rangle\langle B_{m,n}|, \\
\langle B_{m,n}|B_{m',n'}\rangle =& \delta_{mm'}\delta_{nn'} - \frac{1}{d}\delta_{mn}\delta_{m'n'}, \\
\hat{J}T|B_{m,n}\rangle =& \sqrt{\lambda_m\lambda_n}|B_{m,n}\rangle - \frac{1}{d}\delta_{mn}\sum_n\lambda_n|B_{n,n}\rangle.
\end{aligned} \tag{246}
$$

Now we can see that T has rank at most r^2. We can further write $\hat{J}T$ as $T_\alpha - \frac{r}{d}T_\beta$, where

$$T_\alpha \equiv \sum_{m,n=1}^{r} \sqrt{\lambda_m\lambda_n}|B_{m,n}\rangle\langle B_{m,n}|, notag \tag{247}$$

$$T_\beta \equiv \sum_{n=1}^{r} \lambda_n|B_{n,n}\rangle \cdot \sum_{m=1}^{r}\langle B_{m,m}|\frac{1}{\sqrt{r}}. \tag{248}$$

From here we can prove Theorem 19.

Proof of Theorem 19. The singular values of T are the same with those of $\hat{J}T$, so we directly analyze $\hat{J}T$ here. Besides, because $\|T_\beta\|_\infty \leq 1$, in the limit $d \to \infty$, the singular values of $\hat{J}T$ are the same with those of T_1. Therefore, we focus on the singular values of T_α.

Notice that the set of vectors $\{|B_{m,n}\rangle, m \neq n\}$ satisfies

$$\langle B_{m,n}|B_{m',n'}\rangle = \delta_{m,m'}\delta_{n,n'}, \tag{249}$$

so they are part of a vector basis and there are totally $r(r-1)$ of them. The other r vectors $\{|B_{m,m}\rangle\}$ are orthogonal to the previous $r(r-1)$ vectors, and

$$\langle B_{m,m}|B_{n,n}\rangle = \delta_{m,n} - \frac{1}{d}. \tag{250}$$

Construct the following mutually orthogonal vectors using Schmidt decomposition:

$$\begin{aligned}
|\tilde{B}_1\rangle &\equiv |B_{1,1}\rangle, \\
|\tilde{B}_2\rangle &\equiv |B_{2,2}\rangle - \langle B_{1,1}|B_{2,2}\rangle|B_{1,1}\rangle \\
&= |B_{2,2}\rangle + \frac{1}{d}|B_{1,1}\rangle \\
|\tilde{B}_3\rangle &\equiv |B_{3,3}\rangle - \langle B_{1,1}|B_{3,3}\rangle|B_{1,1}\rangle - \langle B_{2,2}|B_{3,3}\rangle|B_{2,2}\rangle \\
&= |B_{3,3}\rangle + \frac{1}{d}|B_{1,1}\rangle + \frac{1}{d}|B_{2,2}\rangle.
\end{aligned} \tag{251}$$

By mathematical induction, eventually we obtain

$$\begin{aligned}
|\tilde{B}_n\rangle &= |B_{n,n}\rangle + \frac{1}{d}\sum_{n'<n}|B_{n',n'}\rangle, \\
\langle\tilde{B}_n|\tilde{B}_n\rangle &= 1 - \frac{1}{d} - \frac{n-1}{d^2} - \frac{2n^2-5n+3}{d^3}.
\end{aligned} \tag{252}$$

For all $n = 1, 2, \cdots, r$, we have

$$\||\tilde{B}_n\rangle\langle\tilde{B}_n| - |B_{n,n}\rangle\langle B_{n,n}|\|_\infty \leq \||\tilde{B}_n\rangle - |B_{n,n}\rangle\|_2 \leq \frac{\sqrt{r-1}}{d}. \tag{253}$$

Define

$$\begin{aligned}
\tilde{T}_\alpha \equiv &\sum_{m\neq n}\sqrt{\lambda_m\lambda_n}|B_{m,n}\rangle\langle B_{m,n}| \\
&+ \sum_{n=1}^{r}\lambda_n\frac{|\tilde{B}_n\rangle\langle\tilde{B}_n|}{1-\frac{1}{d}-\frac{n-1}{d^2}-\frac{2n^2-5n+3}{d^3}},
\end{aligned} \tag{254}$$

which has eigenvalues

$$\{\sqrt{\lambda_m\lambda_n}, \lambda_m(1+O(d^{-1})), \quad m \neq n = 1, 2, \cdots, r\}. \tag{255}$$

Its norm distance to $\hat{J}T$ is bounded by

$$\begin{aligned}
\|\hat{J}T - \tilde{T}_\alpha\|_\infty &\leq \|T_\alpha - \tilde{T}_\alpha\|_\infty + \frac{r}{d}\|T_\beta\|_\infty \\
&\leq \frac{2}{d} + \frac{\sqrt{r-1}}{d} + \frac{r}{d}.
\end{aligned} \tag{256}$$

The theorem is proved using Weyl's inequality from here. □

As a special example, we consider the rank-4 maximally entangled state defined on $2L$ qubits:

$$|+_4\rangle = \frac{1}{2}\left(|0\rangle\otimes|0\rangle + |1\rangle\otimes|1\rangle + |2\rangle\otimes|2\rangle + |3\rangle\otimes|3\rangle\right). \tag{257}$$

Its correlation matrix satisfies:

Proposition 31 (The computation of $\epsilon_c^{(\mathrm{CM})}$ in Fig. 5)**.** *Suppose $|+_4\rangle$ is a multi-qubit state is defined on $\mathcal{H}_{AB}$ with $d_A = d_B = d = 2^L, L \in \mathbb{N}^+$, then*

$$\|T\|_1 = 4 - d^{-1}. \tag{258}$$

Proof. Because the singular values of correlation matrix are invariant under local unitaries, we can assume $|+_4\rangle = (|00\rangle+|11\rangle+|22\rangle+|33\rangle)/2$, where the states $|0\rangle, |1\rangle, |2\rangle, |3\rangle$ are

$$|0\rangle^{L-2}|00\rangle, \quad |0\rangle^{L-2}|01\rangle, \quad |0\rangle^{L-2}|10\rangle, \quad |0\rangle^{L-2}|11\rangle, \tag{259}$$

and the operator basis to be the standard Pauli basis. We can classify the Pauli operators into the following groups. Let O_X be a Pauli-X operator, then define

$$\langle O_X\rangle \equiv \{cO_XO_Z : c = \pm1, \pm i,\ O_Z \text{ is a Pauli-}Z \text{ operator}\}. \tag{260}$$

Therefore, $T_{jk} \neq 0$ if and only if σ_j, σ_k belongs to one of the groups $\langle I\rangle, \langle X_1\rangle, \langle X_2\rangle, \langle X_1X_2\rangle$ simultaneously. We can thus decompose T as the direct sum of the four correlation matrices accordingly:

$$T = T_{\langle I\rangle} \oplus T_{\langle X_1\rangle} \oplus T_{\langle X_2\rangle} \oplus T_{\langle X_1X_2\rangle}. \tag{261}$$

If $\sigma_j, \sigma_k \in \langle I\rangle$, then $\langle m|\sigma_j|n\rangle = \langle m|\sigma_k|n\rangle = \delta_{m,n}$, thus

$$(T_{\langle I\rangle})_{jk} = \frac{1}{d}, \quad \sigma_j, \sigma_k \in \langle I\rangle. \tag{262}$$

Also notice that $|\langle I\rangle| = 2^L - 1 = d - 1$.

If $\sigma_j, \sigma_k \in \langle X_1\rangle$, write them as

$$\sigma_j = (-i)^{s_1}Z_1^{s_1}Z_2^{s_2}\cdots Z_L^{s_L}X_1, \quad s_n = 0, 1, \tag{263}$$

$$\sigma_k = (-i)^{r_1}Z_1^{r_1}Z_2^{r_2}\cdots Z_L^{r_L}X_1, \quad r_n = 0, 1. \tag{264}$$

Then $\langle 0|\sigma_j|1\rangle = (-i)^{s_1}, \langle 1|\sigma_j|0\rangle = i^{s_1}, \langle 2|\sigma_j|3\rangle = (-1)^{s_2}(-i)^{s_1}, \langle 3|\sigma_j|2\rangle = (-i)^{s_1} = (-1)^{s_2}i^{s_1}$, other entries are 0, we further have

$$\begin{aligned}
(T_{\langle X_1\rangle})_{jk} =& \frac{1}{4d}[(-i)^{s_1+r_1} + i^{s_1+r_1} + (-1)^{s_2+r_2}(-i)^{s_1+r_1} \\
&+ (-1)^{s_2+r_2}i^{s_1+r_1}] \\
=& \frac{i^{s_1+r_1}}{4d}[1+(-1)^{s_1+r_1}][1+(-1)^{s_2+r_2}].
\end{aligned} \tag{265}$$

In the matrix form

$$T_{\langle X_1\rangle} = \frac{1}{d}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \otimes I_{2^{L-2}}, \tag{266}$$

which has singular value $\{d^{-1}\}$ with multiplicity $2^L = d$. The computation for $T_{\langle X_2\rangle}$ and $T_{\langle X_1X_2\rangle}$ is similar. Thus,

$$\|T\|_1 = \|T_{\langle I\rangle}\|_1 + \|T_{\langle X_1\rangle}\|_1 + \|T_{\langle X_2\rangle}\|_1 + \|T_{\langle X_1X_2\rangle}\|_1 = 4 - d^{-1}. \tag{267}$$

□

Proof of Proposition 9. Given state $\rho = (1-\epsilon)|+_r\rangle\langle+_r| + \frac{\epsilon}{d^2}I$, the identity part does not contribute to the correlation matrix, thus

$$T_{jk} = (1-\epsilon)\frac{1}{r}\sum_{m,n=0}^{r-1} \frac{\langle m|\sigma_j|n\rangle}{\sqrt{d}} \frac{\langle m|\sigma_k|n\rangle}{\sqrt{d}} \tag{268}$$

According to Proposition 19, if $d > (r + \sqrt{r-1} + 2)/\eta$, each singular value of T belongs to region

$$\left[\frac{1-\epsilon}{r} - \eta, \frac{1-\epsilon}{r} + \eta\right], \tag{269}$$

Thus,

$$\|T\|_1 \geq (1-\epsilon)r - r^2\eta. \tag{270}$$

In order to guarantee that $\|T\|_1 > r - 1 - d^{-1}$, we need at least

$$\epsilon < \frac{1+d^{-1}}{r} + r\eta < \frac{1+d^{-1}}{r} + \frac{r(r+\sqrt{r-1}+2)}{d}. \tag{271}$$

□

Memory Effects in Quantum State Verification

Siyuan Chen[1 2 *] Wei Xie[3] Kun Wang[1 †]

[1] *Institute for Quantum Computing, Baidu Research, Beijing 100193, China*

[2] *College of Physics, Jilin University*

[3] *School of Computer Science and Technology, University of Science and Technology of China*

1 Overview of the results

We investigate the memory effects in quantum state verification [1, 2, 3] and show that quantum memories can substantially improve QSV efficiency. Specifically,

1. We establish an analytic formula for optimizing two-copy state verification in Theorem 1.

2. Building on Theorem 1, we construct a two-copy optimal verification protocol for graph states, utilizing only Clifford gates. This strategy is transversal, requires no magic resource [4] in fault-tolerant quantum computation, and reveals the latent structure inherent in graph states.

3. For multi-copy availability, we present a general dimension expansion technique become increasingly advantageous with growing memory resources, ultimately approaching the theoretical limit of efficiency.

Our findings demonstrate that quantum memories dramatically enhance state verification tasks, shedding light on error-resistant strategies and practical applications of large-scale quantum memory-assisted verification. We believe the results are beneficial to the broader audience of AQIS, *especially to those who are working in benchmarking and improving the qualities of near-term quantum computers.*

A full technical version can be found in the attached technical PDF and is accessible in arXiv:2312.11066.

2 Quantum memory assisted state verification

In this verification strategy, n spatially disparate verifiers conduct a test as follows: First, they store k copies of d-dimensional qudits in their local quantum memories; Then, they measure their local copies in $\mathcal{H}^k \equiv \mathcal{H}^{\otimes k}$ using (possibly entangled) measurements and make a decision based on the outcomes. This "store-and-measure" strategy is vividly illustrated in Fig. 1 for $k = 2$. The test will be repeated M times and the total number of consumed states is Mk. We designate this quantum memory-assisted strategy as an (n, k, d) *verification strategy*. The standard verification strategies fall under the category of $(n, 1, d)$ strategies.

In the good case, the overall state stored in the quantum memories admits a tensor product structure: $|\Psi\rangle := \bigotimes_{r=1}^{k} |\psi\rangle^{(r)}$, where the superscript r represents the r-th copy in the quantum memory. The verifiers perform a local binary measurement $\{T_\ell, \mathbb{1} - T_\ell\}$ such that state $|\Psi\rangle$ passes the test with certainty. In the bad case, we assume that the k states produced by the quantum device are independent, indicating that the fake state in the composite space $\mathcal{H}^{nk}$ has the form

$$\xi = \bigotimes_{r=1}^{k} \sigma^{(r)}, \tag{1}$$

where each $\sigma^{(r)}$ satisfies $\langle\psi|\sigma^{(r)}|\psi\rangle \leq 1 - \varepsilon$. Correspondingly, the maximal probability that the fake state ξ in the bad case can pass the test is

$$p(\Omega) := \max_{\langle\psi|\sigma^{(r)}|\psi\rangle \leq 1-\varepsilon} \mathrm{Tr}\left[\Omega\left(\bigotimes_{r=1}^{k} \sigma^{(r)}\right)\right]. \tag{2}$$

The minimum required number of measurements to saturate the worst-case failure probability, denoted as $M_m(\Omega)$, is given by $M_m(\Omega) = \ln\delta / \ln p(\Omega)$. Thus, the total number of copies consumed by the verification strategy Ω satisfies

$$N_m(\Omega) = kM_m(\Omega) = \frac{k \ln \delta}{\ln p(\Omega)}. \tag{3}$$

The verifiers' objective is to design efficient memory-assisted strategies Ω that minimize the number of copies consumed.

3 Two-copy verification strategy

We analytically solve the maximization problem in Eq. (2) for the case of $k = 2$, yielding an exact analytic formula for optimizing two-copy state verification. First of all, we simplify the form of the optimisation in Eq. (2). Regarding the permutation invariant nature of the verifiers, we show that

*chance.siyuan@gmail.com

†nju.wangkun@gmail.com

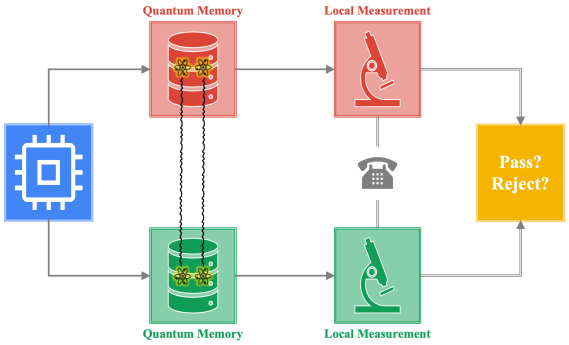


Figure 1: Schematic view of quantum memory assisted state verification. In this $(2,2,d)$ strategy, the verifiers store two copies of quantum states (represented by atoms) in their local quantum memories. They then agree on local measurements via classical communication and perform these measurements on their respective qudits. Finally, they make a "pass/reject" decision from the measurement outcomes.

it is best to consider verification strategies that are symmetric with respect to the two state copies; i.e., $\mathbb{F}_{1\leftrightarrow 2}\Omega\mathbb{F}_{1\leftrightarrow 2} = \Omega$, where $\mathbb{F}_{1\leftrightarrow 2}$ is the swap operator between the first and second copy. Regarding the restriction conditions in Eq. (2), we make the following useful observations: (a) it suffices to consider fake product states without classical correlation; (b) it suffices to optimize over pure fake states; and (c) If the quantum device is not too bad, i.e., there exists an *insurance infidelity* $\varepsilon_{\max} \geq \varepsilon$ such that $\langle\psi|\sigma|\psi\rangle \geq 1-\varepsilon_{\max}$ for all σ, it is then suffices to consider fake states σ for which $\langle\psi|\sigma|\psi\rangle = 1-\varepsilon$. We introduce the following two projectors

$$\mathbb{P}_s := \frac{\mathbb{F}_{1\leftrightarrow 2} + \mathbb{I}_{12}}{2}, \tag{4}$$

$$\mathbb{P}_\psi := |\psi\rangle\langle\psi| \otimes (\mathbb{I} - |\psi\rangle\langle\psi|) \tag{5}$$

which are useful in deriving the analytic formula. Note that $\mathbb{P}_s$ is the projector onto the symmetric subspace of $\mathcal{H}^n \otimes \mathcal{H}^n$. For any symmetric two-copy verification strategy Ω, define the doubly projected operator $\Omega_\star := 2\mathbb{P}_\psi\mathbb{P}_s\Omega\mathbb{P}_s\mathbb{P}_\psi$. Let $\lambda_\star(\Omega)$ be the maximal eigenvalue of the projected operator $\Omega_\star$. We show that, $\lambda_\star$ is the intrinsic property of Ω which underpins Ω's verification efficiency, as elucidated in the ensuing theorem.

Theorem 1 *When $\lambda_\star(\Omega) < 1$ and the existence of insurance fidelity $\varepsilon_{\max}$ is guaranteed, it holds that*

$$p(\Omega) = 1 - 2(1-\lambda_\star(\Omega))\varepsilon + \mathcal{O}(\varepsilon^{1.5}). \tag{6}$$

Correspondingly, the sample complexity of Ω is given by

$$N_m(\Omega) = \frac{2\ln\delta}{\ln p(\Omega)} \approx \frac{1}{(1-\lambda_\star(\Omega))\varepsilon}\ln\frac{1}{\delta}. \tag{7}$$

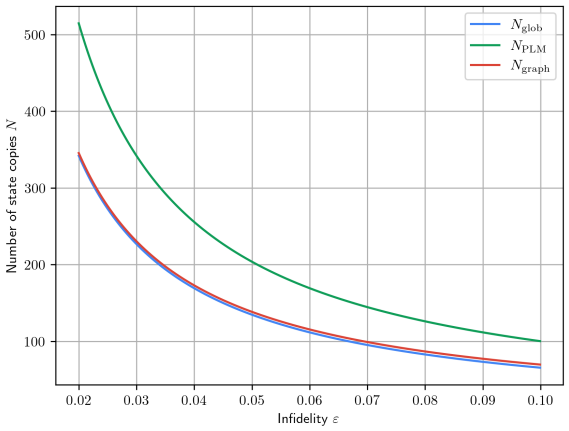
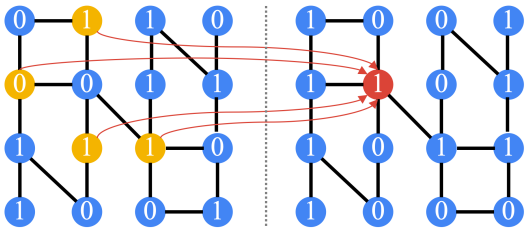


Figure 2: Comparison of the total number of state copies required to verify the Bell state for different strategies as a function of the infidelity ε, where $\delta = 0.001$. Here, N_{graph} is the sample complexity of our proposed two-copy graph verification strategy, N_{PLM} is the sample complexity of the optimal strategy by Pallister *et al.* [2], and N_{glob} is the sample complexity of the globally optimal strategy.



(i) A graph code $\boldsymbol{b}$. (ii) Induced parity code $c(\boldsymbol{b})$.

Figure 3: Schematic view of a graph code $\boldsymbol{b}$ of a graph and its induced parity code $c(\boldsymbol{b})$. The binary value of a vertex (red vertex) in the induced parity code is given by the summation modulus 2 of the values of its adjacent vertices (yellow vertices) in the graph code $\boldsymbol{b}$.

4 Demonstrative exmaple: Graph states

Stabilizer operations have been shown to be efficiently classically simulatable [5]. Under local Clifford transformation, any stablizer states can be reduced into a graph state [6]. We leverage Theorem 1 and one-bit teleportation construction [7] to construct a two-copy Clifford verification strategy for arbitrary multi-qubit graph state $|G\rangle$ associated with a graph $G = (V, E)$, demonstrating that for those state devoid of magic, moderate quantum memory usage, rather than magic resource, can boost the QSV efficiency to global optimality.

To formally describe our two-copy verification strategy for graph states, we begin by introducing the

concept of graph codes of a graph $G = (V, E)$. Let $n = |V|$ be the number of vertices. A *graph code* $\boldsymbol{b} \in \{0,1\}^n$ is an n-bit binary string that assigns the binary value $\boldsymbol{b}_v \in \{0,1\}$ to vertex $v \in V$. Each graph code $\boldsymbol{b}$ uniquely induces a *parity code* $c(\boldsymbol{b}) \in \{0,1\}^n$, where the binary string map $c : \{0,1\}^n \to \{0,1\}^n$ is defined as $c_u(\boldsymbol{b}) := \sum_{v\in V, u\sim v} \boldsymbol{b}_v \pmod 2$, c_u is the value of vertex u, and $u \sim v$ means that u is adjacent to v. Fig. 3 visualizes an example. Let $|\Phi_{00}\rangle := (|00\rangle + |11\rangle)/\sqrt{2}$ be the standard two-qubit Bell state. A binary code pair (m, n) induces a locally transformed Bell state via $|\Phi_{mn}\rangle := (\mathbb{I} \otimes X^m Z^n)|\Phi_{00}\rangle$, where X and Z are the Pauli operators. Our two-copy strategy for $|G\rangle$ involves binary measurement $\{\Omega_g, \mathbb{I} - \Omega_g\}$, where Ω_g corresponding to passing the test is

$$\Omega_g = \sum_{\boldsymbol{b}\in\{0,1\}^n} \bigotimes_{j=1}^{n} |\Phi_{c_j(\boldsymbol{b})\boldsymbol{b}_j}\rangle\langle\Phi_{c_j(\boldsymbol{b})\boldsymbol{b}_j}|_{O_j O_j'}, \tag{8}$$

where O_j, O_j' represent two qubits held by the j-th verifier. The verification strategy carries out as follows. In each test, the verifiers first store two copies of the states. Then, the j-th verifier measures his qubits $O_j O_j'$ with the Bell measurement $\{|\Phi_{mn}\rangle\langle\Phi_{mn}|\}_{m,n\in\{0,1\}}$ and records the outcome as $\boldsymbol{b}_j = m$ and $\boldsymbol{b}_j' = n$. Finally, they classically communicate the outcomes and obtain two graph codes $\boldsymbol{b}, \boldsymbol{b}'$ of the graph G. The states pass the test if and only if $\boldsymbol{b} = c(\boldsymbol{b}')$.

Regarding the performance of our two-copy verification strategy Ω_g, we can prove that $\lambda_\star(\Omega_g) = 0$ and $\varepsilon_{\max} > 1 - \varepsilon$. Thus its optimal efficiency is achieved with a sample complexity of $N_{\text{graph}}(\Omega_g) \approx 1/\varepsilon \ln 1/\delta$ using Eq. (7), indicating that Ω_g achieves globally optimal efficiency. As a showcase, we compare its efficiency with the optimal single-copy verification strategy [2] on verifying the canonical Bell state $|\Phi_{00}\rangle$. As shown in Fig. 2, our two-copy strategy rapidly converges towards globally optimal when $\varepsilon \to 0$, reducing the sample complexity by 50% compared to the optimal single-copy strategy. Note that our two-copy verification strategy for the Bell state bears similarities with the celebrated entanglement-swapping protocol [8, 9], an important component of quantum networks.

5 Dimension expansion

It is demanding to generalize Theorem 1 to $k > 2$. We present a general technique for constructing *efficient* verification strategies for arbitrary k, inspired by the observation that every k-tensor state $|\Psi\rangle$ can be equivalently viewed as a single n-partite state with local dimension d^k. This "dimension expansion" from d to d^k leverages quantum memory to establish an equivalence between an $(n, 1, d^k)$ verification strategy and an (n, k, d) strategy. Concretely, we relax the maximization problem in Eq. (2) by considering any quantum state ξ in $\mathcal{H}^{nk}$ satisfying the fidelity constraint $\langle\Psi|\xi|\Psi\rangle \le (1-\varepsilon)^k$, thus providing an upper bound for $p(\Omega)$:

$$p(\Omega) \le \max_{\langle\Psi|\xi|\Psi\rangle\le(1-\varepsilon)^k} \mathrm{Tr}[\Omega\xi] = 1 - \nu(\Omega)\varepsilon', \tag{9}$$

where $\varepsilon' := 1 - (1-\varepsilon)^k$. Note that $p(\Omega)$ is completely determined by $\nu(\Omega)$, analogous to the single-copy state verification case.

References

[1] Masahito Hayashi and Tomoyuki Morimae. Verifiable measurement-only blind quantum computing with stabilizer testing. *Physical Review Letters*, 115:220502, 2015.

[2] Sam Pallister, Noah Linden, and Ashley Montanaro. Optimal verification of entangled states with local measurements. *Physical Review Letters*, 120:170502, 2018.

[3] Huangjun Zhu and Masahito Hayashi. Efficient verification of pure quantum states in the adversarial scenario. *Physical Review Letters*, 123:260504, 2019.

[4] Victor Veitch, SA Hamed Mousavian, Daniel Gottesman, and Joseph Emerson. The resource theory of stabilizer quantum computation. *New Journal of Physics*, 16(1):013009, 2014.

[5] Daniel Gottesman. The Heisenberg representation of quantum computers. *arXiv preprint quant-ph/9807006*, 1998.

[6] Maarten Van den Nest, Jeroen Dehaene, and Bart De Moor. Graphical description of the action of local clifford transformations on graph states. *Physical Review A*, 69(2):022316, 2004.

[7] Xinlan Zhou, Debbie W Leung, and Isaac L Chuang. Methodology for quantum logic gate construction. *Physical Review A*, 62(5):052316, 2000.

[8] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert. "event-ready-detectors" bell experiment via entanglement swapping. *Phys. Rev. Lett.*, 71:4287–4290, 1993.

[9] Matthäus Halder, Alexios Beveratos, Nicolas Gisin, Valerio Scarani, Christoph Simon, and Hugo Zbinden. Entangling independent pho-

tons by time measurement. *Nature Physics*, 3(10):692–695, August 2007.

Efficient Concatenated Bosonic Code for Additive Gaussian Noise

Kosuke Fukui,[1] Takaya Matsuura,[2] and Nicolas C. Menicucci[2]

[1]*Department of Applied Physics, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[2]*Centre for Quantum Computation and Communication Technology, School of Science, RMIT University, Melbourne, Victoria 3000, Australia*

Bosonic codes offer noise resilience for quantum information processing. Good performance often comes at a price of complex decoding schemes, limiting their practicality. Here, we propose using a Gottesman-Kitaev-Preskill (GKP) code to detect and discard error-prone qubits, concatenated with a quantum parity code to handle the residual errors. Our method employs a simple, linear-time decoder that nevertheless offers significant performance improvements over the standard decoder. Our work may have applications in a wide range of quantum computation and communication scenarios.

Bosonic codes protect discrete quantum information encoded in bosonic mode(s). The infinite-dimensional nature of the bosonic Hilbert space allows more sophisticated encoding than the conventional single-photon encoding [1] or matter-based qubits [2]. The Gottesman–Kitaev–Preskill (GKP) qubit [3] has emerged as a promising bosonic qubit for fault-tolerant quantum computation due to its excellent performance against common types of noise [4]. Experiments involving trapped ions [5] and superconducting circuits [6] have demonstrated a GKP qubit, with the latter boasting a squeezing level close to 10 dB. This level is sufficient for fault tolerance in some proposed architectures [7, 8] and is approaching what is required by others [9, 10]. Recently, optical systems have demonstrated in proof-of-principle experiments of a GKP qubit [11].

The ultimate goal of a large-scale, fault-tolerant quantum computation will require additional innovations, and its ultimate architecture remains an open question. Such requirements on the device can be roughly classified into scalability (many qubits) and fault tolerance (of good quality) [12], and architectures designed to use bosonic qubits as the information carriers have recently demonstrated prominent advances in both areas. Analog quantum error correction (QEC) [13] makes strides toward achieving the goal of Ref. [14] by using the real-valued syndrome of a GKP qubit to improve error recovery in a concatenated code by selecting the most likely error pattern for a given syndrome in a continuous variable (CV)-level decoder. In fact, when used with a suitable qubit code, analog QEC can achieve the hashing bound of additive Gaussian noise [7, 13]. This would seem to be the end of the story except for one major drawback: The decoder for analog QEC employs a type of belief propagation [15] that may become unwieldy in real-world implementations. This is especially true in optical architectures, where fast processing of the outcomes is vital [9, 10], i.e. good performance often comes at a price of complex decoding schemes, limiting their practicality. Thus, in such cases—and especially when hardware-level control is used—reducing the number of bits required to represent outcomes may be critical to fast decoding.

In this work, we make significant progress toward achieving this goal. What we would like is a simple CV-level decoder that generates discrete outcomes that can be fed directly into a qubit-level code at the next level of concatenation. We propose using the GKP code [3] to detect and discard error-prone qubits, concatenated with a quantum parity code (QPC) [16] to handle the residual errors [17]. This is the key innovation that makes further improvements feasible since more complicated codes or additional layers of concatenation do not require modifying the CV-level decoding scheme, thus keeping the decoder simple and efficient. In the following, we will briefly summarize our main results, and we refer to the full paper (Ref. [17]) for the detailed explanation of our method.

The highly reliable measurement (HRM) [7] is the key to improving the performance of the code without the computational overhead required for conventional analog QEC. In the measurement of the GKP qubit, the measurement outcomes—each of the form $s_{\mathrm{m}} = n\sqrt{\pi} + \Delta_{\mathrm{m}}$ with integer n and $|\Delta_{\mathrm{m}}| \leq \sqrt{\pi}/2$, where even and odd n correspond to 0 and 1 logical bit values, respectively—together form the syndrome, as shown in Fig. 1(a). The bit-or phase-flip errors occur when the GKP syndrome value s_{m}, which is wrapped $\mathrm{mod}\sqrt{\pi}$, misidentifies a definite displacement $u \in [-\sqrt{\pi}, \sqrt{\pi})$ as $u \pm \sqrt{\pi}$ [3]. The HRM buffers against this possibility by introducing a danger zone of outcomes $0 \leq \sqrt{\pi}/2 - |\Delta_{\mathrm{m}}| < \delta$ for some $\delta > 0$, as shown in Fig. 1(b). Outcomes in this zone are flagged as unreliable, with $\delta \to 0$ recovering the usual case [3]. This corresponds to flagging as unreliable any displacement $u \pmod{2\sqrt{\pi}}$ that falls within δ of a crossover point $\pm\sqrt{\pi}/2$. Thus, the HRM is a ternary (three-outcome) decoder for GKP qubits that maps each raw CV outcome s_{m} from $\mathbb{R} \to \{\pm 1, E\}$, where E represents an untrustworthy value. Specifically, when the HRM flags a result s_{m} as unreliable, the corresponding qubit is discarded and treated as a located erasure error ($s_{\mathrm{m}} \to E$), while otherwise the result is kept and binned as usual ($s_{\mathrm{m}} \to \pm 1$) depending on which of an even or odd multiple of $\sqrt{\pi}$ s_{m} is close to. For error probabilities of the HRM, we define three cases: the measurement result is correct, $P^{(\mathrm{c})} = \Pr(|u| < \sqrt{\pi}/2 - \delta)$; the result is incorrect, $P^{(\mathrm{i})} = \Pr(|u| > \sqrt{\pi}/2 + \delta)$; or the result is unreliable and the qubit discarded, $P^{(\mathrm{d})} = \Pr(-\delta < |u| - \sqrt{\pi}/2 < \delta)$. We further define the "success probability", $1 - P^{(\mathrm{d})}$, as the probability the qubit was not discarded and the "postselected error probability", $P^{(\mathrm{i})}_{\mathrm{post}} = P^{(\mathrm{i})}/(1 - P^{(\mathrm{d})})$, as the probability of getting an incorrect outcome within the sample of qubits that are not discarded. Figures 1(c) and 1(d) show that decreasing the postselected error probability (by increasing δ) reduces the

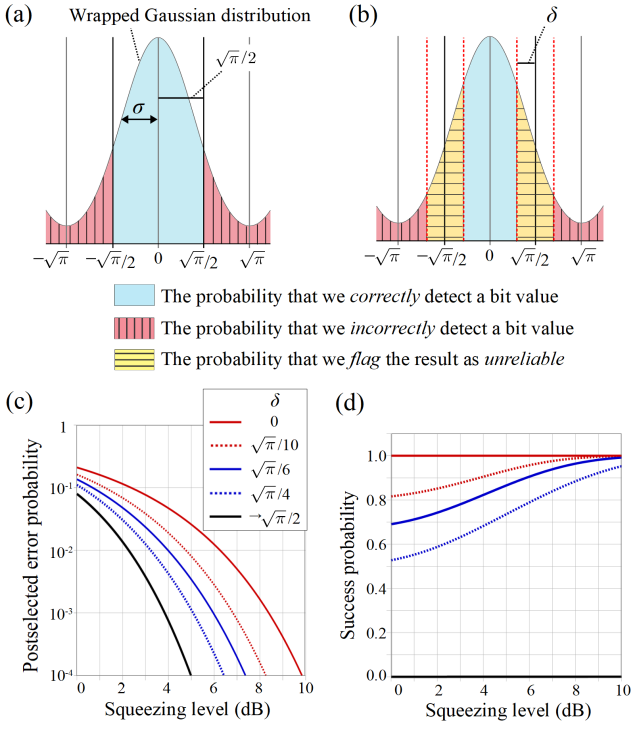


FIG. 1. The highly reliable measurement (HRM). (a) Effect of a displacement $u \bmod 2\sqrt{\pi}$, distributed according tothe wrapped distribution $p(u) = \frac{1}{2\sqrt{\pi}}\vartheta\left(-\frac{u}{2\sqrt{\pi}}, \frac{i\sigma^2}{2}\right)$ with variance σ^2, where $\vartheta(z,\tau) = \sum_{m\in\mathbb{Z}} \exp\left[2\pi i\left(\frac{1}{2}m^2\tau + mz\right)\right]$ is a Jacobi theta function of the third kind. (b) The HRM flags outcomes in the 2δ-wide "danger zone" (yellow) as unreliable. (c) Postselected error probability of the HRM for several values of δ. (d) Corresponding success probability. Note: (Squeezing level in dB) $= -10\log_{10}(\sigma^2/\sigma^2_{\text{vac}})$, where the vacuum variance $\sigma^2_{\text{vac}} = \frac{1}{2}$.

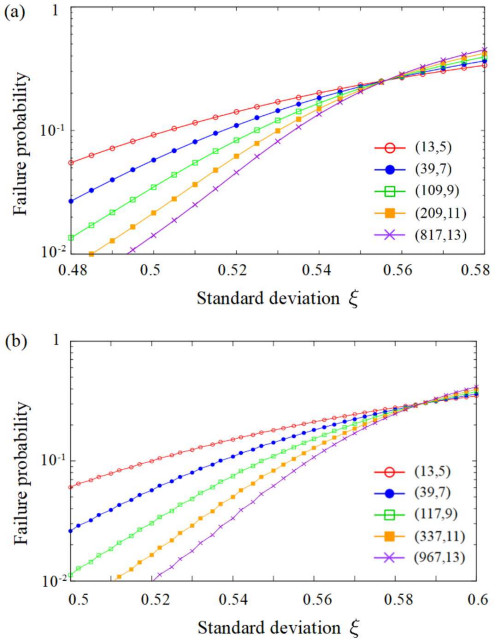


FIG. 2. Failure probabilities using the (n,m) QPC for (a) $\delta_X = \delta_Z = 0$ (conventional GKP error correction [3]) and (b) optimized values δ_X and δ_Z, where $\delta_{Z(X)}$ is the parameter for the HRM in the $Z(X)$ basis.

success probability.

Our decoder uses the CV-level measurement outcome from GKP error correction merely in the CV-level decoder to decide whether to keep the qubit or discard it entirely and treat it as a located erasure error. Loss-tolerant QEC codes such as the QPC [16] are well suited to dealing with the discarded qubits [18–20]. The (n,m) QPC is an nm-qubit code built from n blocks of m qubits. Logical basis states are $|\pm\rangle_{\text{L}} = 2^{-n/2}\left(|0\rangle^{\otimes m} \pm |1\rangle^{\otimes m}\right)^{\otimes n}$. In our code, the physical qubit states are GKP qubits. Here, concatenating GKP qubits with one of the loss-tolerant codes compensates for the discarded ("lost") qubits due to using the HRM. The key insight of our work is that coarse graining the real-valued outcomes to a single ternary outcome, $\mathbb{R} \to \{\pm 1, E\}$, decreasing the error probability of the postselected GKP qubit. This trade of unlocated errors for located erasures makes the logical qubit more robust. This is a simple, local decoding step and does not require complicated modeling of CV-level errors since the HRM maps *locally detected* unreliable results to lost qubits at known locations, while analog QEC requires modeling the joint likelihood of real-valued outcomes over multimode code words, which will be intractable when the code size gets larger. For low-latency applications such as hardware-level decoding for high-throughput optical architectures [9], processing these values is computationally expensive compared to using one- or two-bit values. We refer to the full paper [17] for the detailed explanation of our CV-decoder.

Specifically, figure 2 shows the performance of (a) the QPC without HRM and (b) that with HRM, as a function of the standard deviation of the GKP qubit for several sizes (n,m) of the QPC. In Fig. 2(a), we optimized the value of n for the given m so that the failure probability is minimized; in Fig. 2(b), we optimized δ_Z and δ_X, as well as n, so that the failure probability is minimized, where $\delta_{Z(X)}$ is the parameter for the HRM in the $Z(X)$ basis. The conventional method (a) gives a threshold of $\xi \approx 0.555$, matching previous work with concatenated codes and simple decoding [3, 14]. Our improved method—-discarding unreliable outcomes—-greatly surpasses this, achieving a threshold $\xi \approx 0.585$, as shown in Fig. 1(c). Thus, our method employs a simple, linear-time decoder that nevertheless offers significant performance improvements over the standard decoder.

In conclusion, we have shown that concatenating the GKP code with a QPC considerably improves its performances with a small code and straightforward decoding, linear in the number of modes. Respectively, (1) decoding happens in linear time since the CV-level decoding is entirely local; and (2) the HRM wraps each GKP qubit in a simple "error-detecting" code, so concatenating with any qubit-level code designed to handle erasures [21–23] can benefit from this type of outcome mapping. Further applications and extensions include improved decoding in GKP-based architectures (e.g., [9, 10, 24]) and in codes that exploit biased noise (e.g., [25–27]). In ad-

dition, our method can be used directly with other qubit-level decoders e.g., topological ones [9]. There is room for further improvement of the threshold to achieve the hashing bound of additive Gaussian noise by using more complicated codes, additional layers of concatenation, or analog quantum error correction. But our innovation however is rooted in the simple and efficient decoder, where our method considerably improves its performances with a small code and straightforward decoding in linear time.

This work was partly supported by JST PRESTO Grant No. JPMJPR23FA, JST Moonshot R&D Grant No. JPMJMS2064 and No. JPMJMS2061, UTokyo Foundation, and donations from Nichia Corporation. T.M. acknowledges JSPS Overseas Research Fellowships. This work is supported by the Australian Research Council (ARC) Centre of Excellence for Quantum Computation and Communication Technology (Project No. CE170100012).

[1] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, Reviews of modern physics **79**, 135 (2007).
[2] J. Schreier, A. A. Houck, J. Koch, D. I. Schuster, B. Johnson, J. Chow, J. M. Gambetta, J. Majer, L. Frunzio, M. H. Devoret, *et al.*, Physical Review B **77**, 180502 (2008).
[3] D. Gottesman, A. Kitaev, and J. Preskill, Physical Review A **64**, 012310 (2001).
[4] V. V. Albert, K. Noh, K. Duivenvoorden, D. J. Young, R. Brierley, P. Reinhold, C. Vuillot, L. Li, C. Shen, S. Girvin, *et al.*, Physical Review A **97**, 032346 (2018).
[5] C. Flühmann, T. L. Nguyen, M. Marinelli, V. Negnevitsky, K. Mehta, and J. P. Home, Nature **566**, 513 (2019).
[6] P. Campagne-Ibarcq, A. Eickbusch, S. Touzard, E. Zalys-Geller, N. E. Frattini, V. V. Sivak, P. Reinhold, S. Puri, S. Shankar, R. J. Schoelkopf, *et al.*, Nature **584**, 368 (2020).
[7] K. Fukui, A. Tomita, A. Okamoto, and K. Fujii, Physical review X **8**, 021054 (2018).
[8] K. Fukui, Physical Review A **107**, 052414 (2023).
[9] J. E. Bourassa, R. N. Alexander, M. Vasmer, A. Patil, I. Tzitrin, T. Matsuura, D. Su, B. Q. Baragiola, S. Guha, G. Dauphinais, K. K. Sabapathy, N. C. Menicucci, and I. Dhand, Quantum **5**, 392 (2021).
[10] I. Tzitrin, T. Matsuura, R. N. Alexander, G. Dauphinais, J. E. Bourassa, K. K. Sabapathy, N. C. Menicucci, and I. Dhand, PRX Quantum **2**, 040353 (2021).
[11] S. Konno, W. Asavanant, F. Hanamura, H. Nagayoshi, K. Fukui, A. Sakaguchi, R. Ide, F. China, M. Yabuno, S. Miki, *et al.*, Science **383**, 289 (2024).
[12] D. P. DiVincenzo, Fortschritte der Physik: Progress of Physics **48**, 771 (2000).
[13] K. Fukui, A. Tomita, and A. Okamoto, Physical review letters **119**, 180507 (2017).
[14] J. Harrington and J. Preskill, Physical Review A **64**, 062301 (2001).
[15] D. Poulin, Physical Review A **74**, 052333 (2006).
[16] T. C. Ralph, A. Hayes, and A. Gilchrist, Physical review letters **95**, 100501 (2005).
[17] K. Fukui, T. Matsuura, and N. C. Menicucci, Physical Review Letters **131**, 170603 (2023).
[18] S. Muralidharan, J. Kim, N. Lütkenhaus, M. D. Lukin, and L. Jiang, Physical review letters **112**, 250501 (2014).
[19] W. J. Munro, A. M. Stephens, S. J. Devitt, K. A. Harrison, and K. Nemoto, Nature Photonics **6**, 777 (2012).
[20] F. Ewert, M. Bergmann, and P. van Loock, Physical review letters **117**, 210501 (2016).
[21] M. Varnava, D. E. Browne, and T. Rudolph, Physical review letters **97**, 120501 (2006).
[22] S. D. Barrett and T. M. Stace, Physical review letters **105**, 200502 (2010).
[23] J.-P. Tillich and G. Zémor, IEEE Transactions on Information Theory **60**, 1193 (2013).
[24] M. V. Larsen, C. Chamberland, K. Noh, J. S. Neergaard-Nielsen, and U. L. Andersen, PRX Quantum **2**, 030325 (2021).
[25] D. K. Tuckett, *Tailoring surface codes: Improvements in quantum error correction with biased noise*, Ph.D. thesis, The University of Sydney (2020-01-01).
[26] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, Nature Communications **12**, 2172 (2021).
[27] L. Hänggli, M. Heinze, and R. König, Physical Review A **102**, 052408 (2020).

Randomness expansion from self-tests of contextuality secure against quantum adversaries

Jaskaran Singh[1] * Cameron Foreman[2 3] † Kishor Bharti[4 5] ‡ Adán Cabello[6 7] §

[1] *Department of Physics and Center for Quantum Frontiers of Research & Technology (QFort), National Cheng Kung University, Tainan 701, Taiwan*

[2] *Quantinuum, Partnership House, Carlisle Place, London SW1P 1BX, United Kingdom*

[3] *Department of Computer Science, University College London, London, United Kingdom*

[4] *A*STAR Quantum Innovation Centre (Q.InC), Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore, 138632, Republic of Singapore.*

[5] *Centre for Quantum Engineering, Research and Education, TCG CREST, Sector V, Salt Lake, Kolkata 700091, India.*

[6] *Departamento de Física Aplicada II, Universidad de Sevilla, 41012 Sevilla, Spain*

[7] *Instituto Carlos I de Física Teórica y Computacional, Universidad de Sevilla, 41012 Sevilla, Spain*

Abstract. Randomness expansion secure against quantum adversaries requires either a violation of a Bell inequality at the cost of ensuring no-communication and high detection efficiencies or trusted characterization of some or all of the devices. In this work, we establish that local contextuality-based self-tests are sufficient to provide random numbers that are secure against quantum adversaries without fully characterizing and trusting the devices nor the strict experimental conditions of Bell scenarios. Our scheme is semi device-independent, in the sense that it inherits the assumptions required for the soundness of the underlying contextuality test. We leverage the recent results on self-testing of contextual correlations to show that our scheme provides random numbers which are $\mathcal{O}(\sqrt{\epsilon})$-close to being uniformly distributed and uncorrelated from an unbounded quantum adversary, where ϵ is the robustness paramter of the self-test. For example, we show that a recent experiment on the violation of the 5-cycle noncontextuality inequality [X.-M. Hu *et al.*, npj Quantum Inf. 9, 103 (2023)] guarantees the generation of 0.9878 secure random bits per round.

Keywords: quantum randomness, self-testing, contextuality, semi device-independent, randomness expansion

1 Introduction

Randomness expansion (RE) is the task of using an initial random string to generate a longer one and serves as the underlying protocol for many commercial random number generators. Quantum theory allows for device-independent randomness expansion (DI-RE), for performing the task under minimal assumptions and without comprehensive information of the inner workings of the devices. This is accomplished by observing a violation of a Bell inequality. These protocols can be considered as the pinnacle of security, as they offer security against unbounded quantum adversaries (by which we mean computationally unbounded adversaries who may have access to additional quantum side-information) with only minimal assumptions on the physical devices. However, while they offer maximum security, they also impose extremely strict experimental requirements which are very difficult to implement in the lab. In order to mitigate these issues, a significant effort has been put forth into the development of alternate quantum randomness expansion (QRE) schemes which may offer a similar notion of security as DI-RE but sometimes against classical adversaries and with less stringent experimental requirements [3, 4, 5]. The downside is that slightly stronger assumptions are required, like having a trusted or fully characterized source or measurement. Such schemes are categorized as semi device-independent QREs (SDI-QREs).

The approach of SDI-QRE has generated significant interest in the community leading to several schemes with varying levels of trust and characterization of devices.

In this paper, we propose an alternate SDI-QRE protocol which is based on self-testing of contextual correlations and offers universally composable security against quantum adversaries. Particularly, our protocol does not require complete characterization of the preparation and measurement devices, while the security is determined via the violation of a non-contextuality (NC) inequality. Essentially, we show that whenever the contextual correlations can be robustly self-tested with robustness parameter $\mathcal{O}(\sqrt{\epsilon})$, where ϵ is the deviation from the maximum quantum value, it implies that the QRE scheme can produce uniformly random bits which are $\mathcal{O}(\sqrt{\epsilon})$-close to being uncorrelated from a quantum adversary. Our result re-enforces the idea that contextuality should be considered as a strong candidate for locally generating secure random bit strings much like Bell non-locality, albeit with slightly stronger assumptions (which can be experimentally enforced).

Notably, while our proposal is not fully DI, and therefore requires some characterization of the device to test

*jaskaran@gs.ncku.edu.tw

†cameron.foreman@quantinuum.com

‡kishor.bharti1@gmail.com

§adan@us.es

security, it offers several advantages that make it worth investigating independently. Foremost, our proposal is based on self-testing of contextual correlations under an alternate set of assumptions and can be generalized to a huge class of contextuality-based self-tests [8, 9]. Loophole-free experimental implementations of the same have also been performed recently [6, 7] which enable our scheme to be experimentally accessible. Secondly, we propose an information-theoretically secure QRE scheme in a localized manner (i.e. on a single device). This sets our scheme apart from DI-RE protocols which rely on at least two non-communicating devices.

2 Self-testing of quantum contextual correlations

We adopt the framework of Ref. [8] for odd N-cycle NC inequalities, which shows that localized quantum systems can be self-tested via NC inequalities using the graph theoretic framework of Ref. [10]. For brevity, more details on NC inequalities and the underlying assumptions that are required in the self-testing can be found in our accompanying technical manuscript.

A canonical version of an NC inequality is defined as a linear sum of all the probability assignments $p(1|i)$ to the vertices v_i of an exclusivity graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ such that $p(1|i) + p(1|j) \leq 1$ for all $(v_i, v_j) \in E$. Mathematically, it is written as

$$\beta = \sum_{i=1}^{N} p(1|i) \leq \beta_{nc}(\mathcal{G}) \leq \beta_{qc}(\mathcal{G}), \tag{1}$$

where $\beta_{nc}(\mathcal{G})$ is maximum value of the sum of probabilities attained under the NC assumption and $\beta_{qc}(\mathcal{G})$ is the maximum value attained by quantum theory for a particular choice of projective measurements and quantum state. Now, we can state the result of robust local self-testing of a NC correlations [8].

Definition 1 (Local self-testing of relations) *Consider the NC inequality in Eq.* (1) *and its corresponding exclusivity graph $\mathcal{G}$ for which the optimal quantum strategy is $\mathcal{S}_1 : \left(|v_0\rangle\langle v_0|, \{\Pi_i\}_{i=1}^N\right)$ for $i = 0, 1, \ldots, N$. The NC inequality provides a robust local self-test for this strategy, in the sense, that for any other quantum strategy $\mathcal{S}_2 : \left(|\tilde{v}_0\rangle\langle\tilde{v}_0|, \{\tilde{\Pi}_i\}_{i=1}^N\right)$ that achieves $\beta = \beta_{qc}(\mathcal{G}) - \epsilon$, there exists a global isometry V such that $\left\||\tilde{v}_i\rangle\langle\tilde{v}_i| - V|v_i\rangle\langle v_i|V^\dagger\right\| \leq \sqrt{\epsilon}$.*

3 Protocol and security

We consider a party Alice who wants to securely expand her private randomness locally under a specific set of assumptions (detailed in the accompanying technical manuscript). We describe our necessary notation in Fig. 1 and our protocol to achieve this task in Fig. 2. We consider a quantum adversary Eve, who may have some knowledge of the string of bits generated by Alice. Therefore, the joint state of Alice's m registers storing the produced random bits (the *key*) and the state of Eve

Parameters and notation:

$n \in \mathbb{N}$ - Total number of rounds.

$\mathcal{H}$ - Hilbert space of dimension $d \in \mathbb{N}$.

$N \geq 5 \in \mathbb{N}/2\mathbb{N}$ - Odd total number of measurements greater than or equal to 5.

$\mathcal{M}_i = \{\Pi_i, \mathbb{1} - \Pi_i\}$, $i \in \{1, \ldots, N\}$ - Dichotomic projective measurements with outcomes labelled by $a \in \{0, 1\}$ corresponding to $\mathbb{1} - \Pi_i$ and Π_i, respectively.

$\rho \in \mathcal{H}$ - Initial quantum state on which $\mathcal{M}_i$ are implemented.

β - N-cycle NC inequality as defined in Eq. (1).

$\left(\rho, \{\Pi_i\}_{i=1}^N\right)$ - Quantum realization that obtains $\beta = \beta_{qc}(\mathcal{G})$.

$\epsilon \in (0, 1)$ - Parameter to quantify the deviation of β from $\beta_{qc}(\mathcal{G})$.

$\omega_0 = 1$ and $\omega_1 = \cos\frac{\pi}{N}$.

Figure 1: Parameters and notations used in the protocol.

can be written as a classical-quantum (cq) state which is given as

$$\rho_{KE} = \sum_{\boldsymbol{k} \in \{0,1\}^m} p(\boldsymbol{k}) |\boldsymbol{k}\rangle\langle\boldsymbol{k}| \otimes \rho_{\boldsymbol{k}}^E, \tag{2}$$

where $\rho_{\boldsymbol{k}}^E$ is the reduced state of Eve conditioned on the string $\boldsymbol{k}$.

The string $\boldsymbol{k}$, conditioned on not aborting, is said to be ϵ_{sec}-secure if it can be distinguished (by Eve) from a string generated by a uniform distribution with probability at most $1/2 + \epsilon_{\text{sec}}/2$. This security parameter ϵ_{sec} is quantified by a bound on the trace distance,

$$\left\|\rho_{KE} - 2^{-m}\mathbb{1}_K \otimes \rho_E\right\| \leq \epsilon_{\text{sec}} , \tag{3}$$

where $2^{-m}\mathbb{1}_K$ is the maximally mixed state [2]. It should be noted that this security condition is composable.

In order to show that the random numbers generated from our protocol are ϵ_{sec}-secure, we proceed as follows (for proofs we refer to the accompanying technical manuscript).

Theorem 1 *Let $\mathcal{S}$ be a quantum strategy which achieves $\beta = \beta_{qc}(\mathcal{G})$ for the NC inequality in Eq.* (1)*. This inequality provides a local self-test of relations for all quantum strategies $\tilde{\mathcal{S}} : (\tilde{\rho}_A, \{|\tilde{a}_i\rangle\langle\tilde{a}_i|\}_{a,i})$ that achieve $\beta = \beta_{qc}(\mathcal{G}) - \epsilon$ and there exists an isometry $U = V_A^\dagger \otimes \mathbb{1}_E$ and an ancillary state $|\xi\rangle_E$ such that for the purification $|\psi\rangle_{AE}$ of the state $\tilde{\rho}_A$ and the measurement outcomes*

Procedure

1. Alice chooses a quantum realization $\left(\rho, \{\Pi_i\}_{i=1}^N\right)$ of an odd N-cycle NC inequality.

2. **While** $j \leq n$:

   For $q \in [0,1]$, choose $T_j = 0$ with probability $1-q$ and $T_j = 1$ otherwise.

   If $T_j = 0$ (Key round):

   Implement $\mathcal{M}_1$ on state ρ to obtain a.

   If $a = 0$:

   Record a as k_j with probability ω_1 and set $j = j+1$.

   Else $a = 1$:

   Record a as k_j with probability ω_0 and set $j = j+1$.

   Else $T_j = 1$ (Spot-check round):

   Randomly choose $i \in \{1, \ldots, N\}$.

   Implement $\mathcal{M}_i$ on state ρ to obtain a. Record a, i and set $j = j+1$.

3. Using the statistics from all spot-check rounds evaluate β.

   If $\beta_{qc}(\mathcal{G}) - \beta < \epsilon$:

   Abort the protocol.

   Else

   Obtain the bit string $\mathbf{k}$ as a concatenation of all bit values k_j.

Figure 2: Protocol for randomness expansion.

$\{|a_i\rangle\langle a_i|\}$,

$$\left\| U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i| \otimes \mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE} U^\dagger - \left(|a_i\rangle\langle a_i| \otimes \mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|_A \otimes |\xi\rangle\langle\xi|_E\right) \right\| \leq 2\epsilon' \quad \forall a, i, \tag{4}$$

where $\epsilon' = 2\sqrt{\epsilon}$. This shows that self-testing quantum strategies are $2\epsilon'$-close to uncorrelated from Eve. Following Theorem 1, we can finally show that

Theorem 2 *All self-testing quantum strategies that achieve $\beta = \beta_{qc}(\mathcal{G}) - \epsilon$ satisfy*

$$\left\|\rho_{AE}^{\otimes m} - 2^{-m}\mathbb{1}_K \otimes \rho_E\right\| \leq 8\sqrt{\epsilon}, \tag{5}$$

where $\rho_E = \bigotimes_{j=1}^m |\xi_j\rangle\langle\xi_j|$ is the reduced state of Eve for m key rounds, and $|\xi_j\rangle$ is her state corresponding to the jth round.

This proves that self-tests of contextuality can offer information-theoretically secure randomness expansion, against a quantum adversary, with security parameter $\epsilon_{sec} = 8\sqrt{\epsilon}$.

Next, we calculate the net randomness expansion of our scheme. The amount of randomness produced is quantified by the smooth min-entropy $H_{\min}^{\delta}(K|E)$. In our protocol, after n total number of rounds the produced randomness is $H_{\min}^{\delta}(K|E) = m$, for $\delta = 8\sqrt{\epsilon}$. Our protocol also consumes a certain amount of randomness which should be kept private and independent of the settings chosen. The amount of initial private randomness can be quantified by noting that it is required for (a) choosing whether the round will be a spot-check with probability q, (b) randomly choosing which measurement to implement, (c) post-selecting the outcome 1 with probability ω_1 (note that the outcome 1 occurs with probability $1/[1+\cos(\pi/N)]$) and (d) randomly choosing the sequence of measurement for the self-test (for more details, see Ref. [6]). In total, the amount of initial private randomness required is

$$l_{\text{in}} = nh(q) + q\log N + h\left(\frac{\omega_1}{1+\cos\frac{\pi}{N}}\right) + 1\ , \tag{6}$$

where $h(\cdot)$ is the binary entropy function. Therefore, the net randomness expansion per round can be evaluated as $r = \frac{m - l_{\text{in}}}{n}$.

As an example we consider the data of the KCBS self-test performed in Ref. [6] and find that our protocol could generate $r = 0.9878$ bits per round when $n = 10^4$ out of which 10^2 rounds can be used for spot checking with $\epsilon_{sec} = 10^{-2}$. We note that the value of ϵ_{sec} can be improved by considering higher number of rounds with more precise results.

4 Conclusion

We have proposed a scheme for locally performing SDI-QRE secure against an unbounded quantum adversary. We leverage the results of Ref. [8] to show that self-testing contextual correlations can certify information-theoretic security of randomness expansion. Unlike many of the SDI schemes developed so far, our scheme does not require trusted measurements (but assumes that the measurements satisfy certain constraints. Instead, a different set of assumptions are required which can be (and have been) experimentally enforced [6, 7, 11, 12].

Additionally, the security of our scheme does not rely on entanglement, no-communication or even on the requirement of at least two parties. On the downside, it is necessary to provide some characterization of the devices for the self-test such that the orthogonality and repeatability conditions are satisfied. This increases the amount of resources required for the certification of security. However, there is always a compromise between the level of trust and security and the amount of resources required to achieve it, and our scheme proposes a different trade-off as compared to other works. It may be argued that the requirement of correlations that offer close-to-maximal violation of NC inequalities may be a downside of our scheme, but unlike Bell experiments and as evidenced by the experiment in Ref. [6], it is already possible to achieve such correlations in the lab which makes our scheme much more relevant.

References

[1] S. Fehr and R. Gelles and C. Schaffner. Security and composability of randomness expansion from Bell inequalities. *Phys. Rev. A* 87, 012335.

[2] C. Portmann and R. Renner. Security in quantum cryptography. *Rev. Mod. Phys.* 94, 025008 (2022).

[3] Z. Cao and H. Zhou and X. Yuan and X. Ma. Source-independent quantum random number generation. *Phys. Rev. X* 6, 011020 (2016).

[4] J. B. Brask and A. Martin and W. Esposito *et al.*. Megahertz-rate semi-device-independent quantum random number generators based on unambiguous state discrimination. *Phys. Rev. Appl.* 7, 054018 (2017).

[5] D. Drahi and N. Walk and M. J. Hoban and A. K. Fedorov *et al.*. Certified quantum random numbers from untrusted light. *Phys. Rev. X* 10, 041048 (2020).

[6] X. Hu and Y. Xie and A. S. Arora *et al.*. Self-testing of a single quantum system from theory to experiment. *npj Quantum Inf.* 9, 103 (2023).

[7] P. Wang and J. Zhang and C.-Y. Luan *et al.*. Significant loophole-free test of Kochen-Specker contextuality using two species of atomic ions. *Sci. Adv.* 8, eabk1660 (2022).

[8] K. Bharti and M. Ray, Maharshi and A. Varvitsiotis *et al.*. Robust Self-Testing of Quantum Systems via Noncontextuality Inequalities. *Phys. Rev. Lett.* 122, 250403 (2019).

[9] K. Bharti and M. Ray and Z.-P. Xu *et al.*. Graph-Theoretic approach for self-testing in Bell scenarios. *PRX Quantum* 3, 030344 (2022).

[10] A. Cabello and S. Severini and A. Winter. Graph-Theoretic Approach to Quantum Correlations. *Phys. Rev. Lett.* 112, 040401 (2014).

[11] A. Zhang and H. Xu and J. Xie *et al.*. Experimental Test of Contextuality in Quantum and Classical Systems. *Phys. Rev. Lett.* 122, 080401 (2019).

[12] M. Um and Q. Zhao and J. Zhang *et al.*. Randomness Expansion Secured by Quantum Contextuality. *Phys. Rev. Appl.* 13, 034077 (2020).

Randomness expansion from self-tests of contextuality secure against quantum adversaries

Jaskaran Singh,[1,*] Cameron Foreman,[2,3,†] Kishor Bharti,[4,5,‡] and Adán Cabello[6,7,§]

[1] *Department of Physics and Center for Quantum Frontiers of Research & Technology (QFort), National Cheng Kung University, Tainan 701, Taiwan*
[2] *Quantinuum, Partnership House, Carlisle Place, London SW1P 1BX, United Kingdom*
[3] *Department of Computer Science, University College London, London, United Kingdom*
[4] *A*STAR Quantum Innovation Centre (Q.InC), Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore, 138632, Republic of Singapore.*
[5] *Centre for Quantum Engineering, Research and Education, TCG CREST, Sector V, Salt Lake, Kolkata 700091, India.*
[6] *Departamento de Física Aplicada II, Universidad de Sevilla, 41012 Sevilla, Spain*
[7] *Instituto Carlos I de Física Teórica y Computacional, Universidad de Sevilla, 41012 Sevilla, Spain*

Achieving universally composable randomness expansion secure against quantum side-information requires a loophole-free violation of a Bell inequality at the cost of ensuring no-communication and high detection efficiencies or the characterization of some or all of the devices by expending more resources. In this work, we establish that local contextuality-based self-tests are sufficient to provide universally composable random numbers which are secure against quantum side-information without fully characterizing and trusting the devices and the strict experimental conditions of Bell scenarios. Our scheme is semi device-independent and inherits the assumptions required for the quantum soundness of the underlying contextuality test. We leverage the recent results on self-testing of contextual correlations to show that our scheme provides uniform random numbers which are $\mathcal{O}(\sqrt{\epsilon})$-close (in trace norm) to being uncorrelated from a quantum adversary where ϵ is the deviation from the maximum quantum value of the non-contextuality inequality. As an example, we show that a recent experiment on the violation of the 5-cycle non-contextuality inequality [X.-M. Hu *et al.*, npj Quantum Inf. 9, 103 (2023)] can already guarantee the generation of 0.9878 secure random bits per round

Introduction.— Random numbers are an indispensable resource in several information processing tasks including gaming, simulations, computing and many tasks in cryptography. Quantum theory has proven useful in generating "true randomness" [1–6] by utilizing quantum processes which are inherently non-deterministic [7–22] with several commercial applications [23, 24].

As an added advantage, quantum theory also allows for device-independent randomness expansion (DI-RE) [25–31] which can be implemented under minimal assumptions and without comprehensive information of the inner workings of the devices. This offshoot of quantum cryptography can allow the certification of secure randomness against computationally unbounded adversaries who may posses quantum side-information (which we call unbounded quantum adversaries) by exploiting the properties of observing the violation of a Bell inequality [32–34]. A considerable amount of research has been devoted to the theoretical and experimental development of such protocols [4, 35–39]. These protocols can be (at least theoretically) considered as the pinnacle of security as they offer composable security [40] against unbounded quantum adversaries under minimal assumptions on the physical devices.

However, while DI-RE offers maximum security, such schemes also impose extremely strict experimental requirements which are very difficult to implement. These requirements generally include no-communication and loophole-free Bell violation. In order to mitigate these issues a significant effort has been put forth into the development of alternate quantum randomness expansion (QRE) schemes which may offer a similar notion of security as DI-RE but sometimes against classical adversaries or with less stringent experimental requirements [41–51]. The downside is that slightly stronger assumptions are required, like having a trusted or fully characterized source or measurements. Such schemes are categorized as semi device-independent QREs (SDI-QREs).

Recently, SDI-QREs based on the violation of a non-contextuality (NC) inequality [52–55] have also been proposed [56, 57]. This approach, in principle, sounds more advantageous than the aforementioned QRE protocols as it does not require strict experimental conditions as DI-RE and the full characterization of either the preparation or measurement devices. While their approach sounds favorable, even in their work contextual correlations by themselves (without any additional assumptions apart from those needed in tests of contextuality) cannot guarantee composable security against a quantum adversary with minimal trust on the devices.

Consequently it is natural to ask whether it is possible to generate universally secure quantum random numbers without fully characterizing the devices and the strong requirements of DI-RE.

In this paper, we conclusively answer this question and propose a SDI-QRE protocol based on self-testing of contextual correlations which offers universally composable security. Particularly, our protocol does not require

complete characterization of the preparation and measurement devices, while its security is derived via the violation of a NC inequality. Essentially, we show that whenever the contextual correlations can be robustly self-tested with robustness parameter ϵ, it implies that the QRE scheme can produce uniformly random bits which are $\mathcal{O}(\sqrt{\epsilon})$-close (in trace norm) to being uncorrelated from a quantum adversary. Our result re-enforces the idea that contextuality should be considered as a strong candidate for locally generating secure random bit strings much like Bell non-locality, albeit with slightly stronger assumptions (which can be experimentally enforced)

Notably, while our proposal is not fully DI, and therefore requires some characterization of the device to test security, it offers several advantages that make it worth investigating independently. Foremost, our proposal is based on self-testing of contextual correlations under an alternate set of assumptions. Significantly loophole-free experimental implementations of the same have been performed recently [58, 59] which can already achieve the desired correlations. This enables our scheme to be experimentally accessible and will act as a catalyst for further research into local certification of security of QRE schemes. Secondly, our scheme certifies universally composable security of a QRE scheme in its natural setting: a single local device.

Self-testing of quantum contextual correlations.—We adopt the framework of Ref. [60] which shows that localized quantum systems can be self-tested via NC inequalities using the graph theoretic framework of Ref. [53]. This notion of self-test is markedly distinct than the self-tests performed in Bell scenarios which have two spatially separated parties. As an example, in the Bell-CHSH scenario it is possible to show that the maximum violation of the Bell-CHSH inequality implies that both the parties are implementing Pauli qubit observables on a two-qubit maximally entangled state. However, a self-test of NC inequalities only implies that the set of projectors involved in the NC inequality follow the same exclusivity relations as the set of projectors which maximally violate the inequality and the overlap with the handle (underlying preparation) is the same in both cases. Moreover, the underlying assumptions in both the frameworks are different. For more details on NC inequalities and the underlying assumptions that are required in the self-testing see Appendix .

We start by defining an exclusivity graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ which is defined as a pair of set of vertices $\mathcal{V} = \{v_i\}_{i=1}^N$ and a set of edges $\mathcal{E} = \{v_i, v_j\}_{i,j}$. If the pair of vertices $\{v_i, v_j\} \in \mathcal{E}$, then the corresponding vertices are termed as exclusive and are denoted by $v_i \sim v_j$.

A canonical version of an NC inequality is then written as a linear sum of all the probability assignments $p(1|i)$ to the vertices of a graph $\mathcal{G}$ such that $p(1|i) + p(1|j) \leq 1$ for $v_i \sim v_j \ \forall i, j$. Mathematically, it is written as

$$\beta = \sum_{i=1}^{N} p(1|i) \leq \beta_{nc}(\mathcal{G}), \tag{1}$$

where $\beta_{nc}(\mathcal{G})$ is maximum value of the sum of probabilities which is attained under the NC assumption that the probability assigned to a vertex i is independent of the assignment to vertex $j \neq i$. A quantum realizations of this scenario is the one in which projectors Π_i are assigned to each vertex i, such that $\mathrm{tr}\,(\Pi_i \Pi_j) = 0$ for $v_i \sim v_j$. These projectors are considered to act on a quantum state ρ and the required probabilities are evaluated according to the Born rule $p(1|i) = \mathrm{tr}\,(\rho \Pi_i)$. In this case, the inequality (1) can attain a maximum value $\beta_{qc}(\mathcal{G}) \geq \beta_{nc}(\mathcal{G})$, which is known as the quantum bound of the inequality.

Now, we can state the result of robust local self-testing of a NC correlations [60].

Definition 1 (Local self-testing of relations)**.** *Consider the NC inequality in Eq. (1) and its corresponding exclusivity graph $\mathcal{G}$ for which the optimal quantum strategy be $\mathcal{S}_1 : \left(|v_0\rangle\langle v_0|, \{\Pi_i\}_{i=1}^N\right)$. This strategy achieves $\beta = \beta_{qc}(\mathcal{G})$ and satisfies the relations $tr\,(\Pi_i \Pi_j) = 0$ for all $v_i \sim v_j$. The NC inequality provides a robust local self-test for these relations, in the sense, that for any other quantum strategy $\mathcal{S}_2 : \left(|\tilde{v}_0\rangle\langle\tilde{v}_0|, \{\tilde{\Pi}_i\}_{i=1}^N\right)$ that achieves $\beta = \beta_{qc}(\mathcal{G}) - \epsilon$, there exists an isometry V such that $\left\| |\tilde{v}_i\rangle\langle\tilde{v}_i| - V|v_i\rangle\langle v_i| V^\dagger \right\| \leq \sqrt{\epsilon}$ for $i = 0, 1, \ldots, N$.*

Here, $\|A\| = \mathrm{tr}\left(\sqrt{A^\dagger A}\right)$ denotes the trace norm of a matrix A. Next, we will use this definition of local self-testing of relations to show that it is possible to locally certify the security of a random number generator without entanglement.

Protocol.—We consider a party Alice who is interested in expanding her private randomness in a way that she can certify the security of her randomness locally under a set of assumptions (as detailed in Appendix). We describe a protocol to achieve this task in Fig. 1. Here, it should be noted that while we present our findings for odd N-cycle NC inequalities it generalizes for arbitrary NC inequalities which can provide a self-test following Ref. [60].

In our protocol it should be noted that in the key round, for the measurements $\mathcal{M}_i$ Alice randomly post-selects the outcomes $\mathbb{1} - \Pi_i$ with probability $\omega_1 = \cos\pi/N$, while the outcomes corresponding to Π_i are selected with probability $\omega_0 = 1$. This is done in order to make the resultant bit string unbiased. After post-selection the resultant probability distribution of the outcomes $\hat{p}(a|i)$ is simply

$$\hat{p}(a|i) = \frac{\omega_a p(a|i)}{\sum_{a=0}^{1} \omega_a p(a|i)} = \frac{1}{2} \quad \forall a \in \{0, 1\}. \tag{2}$$

Protocol for randomness expansion

Parameters and notation:

$n \in \mathbb{N}$ - Total number of rounds.

$\mathcal{H}$ - Hilbert space of dimension $d \in \mathbb{N}$.

$N \geq 5 \in \mathbb{N}/2\mathbb{N}$ - Total number of measurements which are taken to be odd and greater than or equal to 5.

$\mathcal{M}_i = \{\Pi_i, \mathbb{1} - \Pi_i\}$, $i \in \{1, \ldots, N\}$ - Two outcome projective measurements with outcomes labelled by $a \in \{0, 1\}$ corresponding to $\mathbb{1} - \Pi_i$ and Π_i respectively.

$\rho \in \mathcal{H}$ - Initial quantum state on which the measurements are performed.

β - N-cycle NC inequality as defined in Eq. (1) with maximum quantum value $\beta_{qc}(\mathcal{G})$.

$\left(\rho, \{\Pi_i\}_{i=1}^N\right)$ - Quantum realization that obtains $\beta = \beta_{qc}(\mathcal{G})$.

$\epsilon \in (0, 1)$ - Parameter to quantify the deviation of β from $\beta_{qc}(\mathcal{G})$ under trace norm.

$\omega_0 = 1$ and $\omega_1 = \cos\frac{\pi}{N}$.

Procedure

1. Alice chooses a quantum realization $\left(\rho, \{\Pi_i\}_{i=1}^N\right)$ of an odd N-cycle NC inequality.
2. **While** $j \leq n$:

   For $q \in [0, 1]$, choose $T_j = 0$ with probability $1 - q$ and $T_j = 1$ otherwise.

   If $T_j = 0$ (Key round):

   Perform the measurement $\mathcal{M}_1$ on state ρ to obtain outcome a.

   If $a = 0$:

   Record a as k_j with probability ω_1 and set $j = j + 1$.

   Else $a = 1$:

   Record a as k_j with probability ω_0 and set $j = j + 1$.

   Else $T_j = 1$ (Spot-check round):

   Randomly choose $i \in \{1, \ldots, N\}$ with uniform probability.

   Perform the measurement $\mathcal{M}_i$ on state ρ to obtain outcome a. Record the outcome a and the measurement i and set $j = j + 1$.
3. Using the statistics from all spot-check rounds evaluate β.

   If $\beta_{qc}(\mathcal{G}) - \beta < \epsilon$:

   Abort the protocol.

   Else

   Obtain the bit string $\mathbf{k}$ as a concatenation of all bit values k_j.

FIG. 1. The template protocol for randomness expansion using self-testing of contextual correlations.

Here, it should be noted that the post-selection is performed on the key rounds only and as such does not introduce any loopholes in the test of the NC inequality which is performed on a separate set of rounds.

In the case when Alice is able to successfully pass the self-test using the statistics of spot-check rounds, she can certify that the results obtained in the key rounds (after post-selection) can be used to generate a secure and random key.

Security.—In order to show security we consider that a computationally unbounded quantum adversary, Eve, may share some correlations with Alice. Therefore, the joint state of Alice's m registers storing the bits of the key and the state of Eve can be written as a classical-quantum (cq) state which is given as

$$\rho_{KE} = \sum_{\boldsymbol{k} \in \{0,1\}^m} p(\boldsymbol{k}) \left|\boldsymbol{k}\right\rangle \left\langle \boldsymbol{k}\right| \otimes \rho_{\boldsymbol{k}}^E, \tag{3}$$

where $\rho_{\boldsymbol{k}}^E$ is the reduced state of Eve conditioned on the string $\boldsymbol{k}$.

The string $\boldsymbol{k}$ is said to be ϵ_{sec}-secure if it can be distinguished (by Eve) from a string generated by a uniform distribution with probability at most $1/2 + \epsilon_{sec}/2$. For cq-states, this distinguishability is quantified by a bound on the trace norm (due to the maximum guessing probability in quantum state discrimination).

$$\left\|\rho_{KE} - 2^{-m}\mathbb{1}_K \otimes \rho_E\right\| \leq \epsilon_{sec}, \tag{4}$$

where it should be noted that the reduced state of Eve is now uncorrelated with the bit string of Alice.

In order to show that the random numbers generated from our protocol are ϵ_{sec}-secure (for some value of ϵ_{sec}), we proceed as follows. First, we show that the strategy of Alice is ϵ'-close to being uncorrelated with Eve where $\epsilon' = 2\sqrt{\epsilon}$ if the N-cycle NC inequality satisfies $\beta \leq \beta_{qc} - \epsilon$. In this step we use the self-test proposed in Ref. [60] and derive an extraction map for the case when Alice observes the maximal violation of the NC inequality. An extraction map is defined as a map that can transform a quantum system into another quantum system having some desirable properties. Afterwards we show that in case of close-to-maximal violation, our extraction map extracts a strategy which is $4\sqrt{\epsilon}$-close to the target (desired) strategy if Alice observes $\beta \leq \beta_{qc} - \epsilon$. Then, we show that after performing post-selection in the key rounds by Alice, the bit string with her can be made uniform. Applying both the two results together, we are then able to prove Eq. (4) with $\epsilon_{sec} = 8\sqrt{\epsilon}$ for our case.

Here, a quantum strategy is defined as the quantum realization of a NC scenario given by the tuple $\mathcal{S} = (\left|v_0\right\rangle \left\langle v_0\right|, \{\left|a_i\right\rangle \left\langle a_i\right|\}_{a,i})$ such that for this realization $\beta = \beta_{qc}(\mathcal{G})$. In our work the strategy to be self-tested is denoted by a tilde.

Theorem 1. *Let $\mathcal{S}$ be the target quantum strategy which achieves $\beta = \beta_{qc}(\mathcal{G})$ for the NC inequality in Eq. (1).*

This inequality provides a local self-test for all quantum strategies $\tilde{\mathcal{S}} : (\tilde{\rho}_A, \{|\tilde{a}_i\rangle\langle\tilde{a}_i|\}_{a,i})$ *that also achieve* $\beta = \beta_{qc}(\mathcal{G})$ *and there exists an isometry* $U = V_A^\dagger \otimes \mathbb{1}_E$ *and an ancillary state* $|\xi\rangle_E$ *such that for the purification* $|\psi\rangle_{AE}$ *of the state* $\tilde{\rho}_A$*,*

$$U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i| \otimes \mathbb{1}_E |\psi\rangle_{AE}\right) = \left(|a_i\rangle\langle a_i|v_0\rangle_A\right) \otimes |\xi\rangle_E \quad \forall a, i. \tag{5}$$

Proof. See appendix. □

The specific form of U in this case forms our extraction map. Using this extraction map we can show that the post-measurement state of Alice is given by (see Appendix for the proof)

$$\rho_{AE} = \sum_{a=0}^{1} p(a|i) |a_i\rangle\langle a_i| \otimes |\xi\rangle\langle\xi|_E \quad \forall i, \tag{6}$$

where as can be seen the state of Eve, denoted by $|\xi\rangle_E$ is uncorrelated with Alice's.

Next, we look at the case where Alice can only achieve a violation of the NC inequality which is ϵ-close to the maximal one.

Theorem 2. *Let* $\mathcal{S}$ *be a quantum strategy which achieves* $\beta = \beta_{qc}(\mathcal{G})$ *for the NC inequality in Eq. (1). This inequality provides a local self-test of relations for all quantum strategies* $\tilde{\mathcal{S}} : (\tilde{\rho}_A, \{|\tilde{a}_i\rangle\langle\tilde{a}_i|\}_{a,i})$ *that achieve* $\beta = \beta_{qc}(\mathcal{G}) - \epsilon$ *and there exists an isometry* $U = V_A^\dagger \otimes \mathbb{1}_E$ *and an ancillary state* $|\xi\rangle_E$ *such that for the purification* $|\psi\rangle_{AE}$ *of the state* $\tilde{\rho}_A$ *and the measurement outcomes* $\{|a_i\rangle\langle a_i|\}$*,*

$$\left\| U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i| \otimes \mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE} U^\dagger - \left(|a_i\rangle\langle a_i| \otimes \mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|_A \otimes |\xi\rangle\langle\xi|_E\right) \right\| \leq 2\epsilon' \quad \forall a, i. \tag{7}$$

Proof. See appendix. □

As can be seen the same extraction map as before is sufficient to showcase our results. Similarly as before, we can use these results to show that the joint state of Alice's outcomes and Eve in the case of close-to-optimal violation satisfies

$$\left\| \rho_{AE} - \sum_{a=0}^{1} p(a|i) |a_i\rangle\langle a_i| \otimes |\xi\rangle\langle\xi|_E \right\| \leq 4\epsilon' \quad \forall i. \tag{8}$$

So far we have shown that the post-measurement states of Alice $4\epsilon'$-close to being uncorrelated from Eve whenever the rounds are spot-check rounds. Since Alice randomly selects each round as either a spot-check round or a key round without Eve knowing, it is expected that the post-measurements states of Alice in the key round will also follow a similar behavior. In this case, the distribution $p(a|i)$ will be replaced with post-selection probability distribution $\hat{p}(a|i)$.

Theorem 3. *For all quantum strategies* $\tilde{S} : (\tilde{\rho}_A, \{|\tilde{a}_i\rangle\langle\tilde{a}_i|\}_{a,i})$ *that can be self-tested following Theorem 5, Alice's* m *key registers satisfy*

$$\left\| \rho_{AE}^{\otimes m} - 2^{-m} \sum_{\boldsymbol{k}\in\{0,1\}^m} |\boldsymbol{k}\rangle\langle\boldsymbol{k}| \otimes \rho_E \right\| \leq 4\epsilon', \tag{9}$$

where $\rho_E = \bigotimes_{j=1}^{m} |\xi_j\rangle\langle\xi_j|$ *is the reduced state of Eve for the* m *rounds, and* $|\xi_j\rangle$ *is her state corresponding to the* j*th key round.*

Proof. See appendix. □

We can further simplify the expression by noting that $\sum_{\boldsymbol{k}\in\{0,1\}^m} |\boldsymbol{k}\rangle\langle\boldsymbol{k}| = \mathbb{1}$ and putting $\rho_{AE}^{\otimes m} = \rho_{KE}$ to obtain $\|\rho_{KE} - 2^{-m}\mathbb{1} \otimes \rho_E\| \leq 4\epsilon'$, which is exactly the expression we set out to prove. The parameter $4\epsilon'$ can then be identified with ϵ_{sec} which quantifies the security of the protocol. For a higher level of security we would like it be as small as possible.

Next, we can bound the length of the secure random bit string that can be generated using our approach. Note that the amount of randomness produced is quantified by the smooth min-entropy $H_{\min}^{\delta}(K|E)$ (for more details see Appendix). In our case, since we already show that after self-testing the key registers of Alice are uniform and uncorrelated with Eve, we can safely set the parameter $\delta = 4\epsilon' = 8\sqrt{\epsilon}$ and obtain $H_{\min}^{\delta}(K|E) = m$. However, our protocol also consumes a certain amount of randomness which should be kept private and independent of the settings chosen. The amount of initial private randomness can be quantified by noting that it is required for (a) choosing whether the round will be a spot-check with probability q, (b) randomly choosing which measurement to implement and (c) post-selecting the outcome 1 with probability ω_1 (note that the outcome 1 occurs with probability $1/[1+\cos(\pi/N)]$) and (d) randomly choosing the sequence of measurement for the self-test (e.g. to measure $p(1|i=2)$ for the self-test, one could choose to perform $\mathcal{M}_2$ followed by $\mathcal{M}_3$ or vice-versa. For more details see Ref. [59].). In total, the amount of initial private randomness required, for a single round, is

$$l_{\text{in}} = nh(q) + q\log N + h\left(\frac{\omega_1}{1+\cos\frac{\pi}{N}}\right) + 1\ , \tag{10}$$

where $h(\cdot)$ is the binary entropy function.

Therefore, the net randomness expansion per round can be evaluated as $r = \frac{m - l_{\text{in}}}{n}$ Moreover, it has been shown that for a large number of rounds n, an effective value of probability for spot-checking is $q = \frac{1}{\sqrt{n}}$. We plot the amount of randomness generated per round as a function of the total number of experimental rounds in Fig. 2.

As an example for randomness expansion using our scheme, we can consider the data of the KCBS self-test

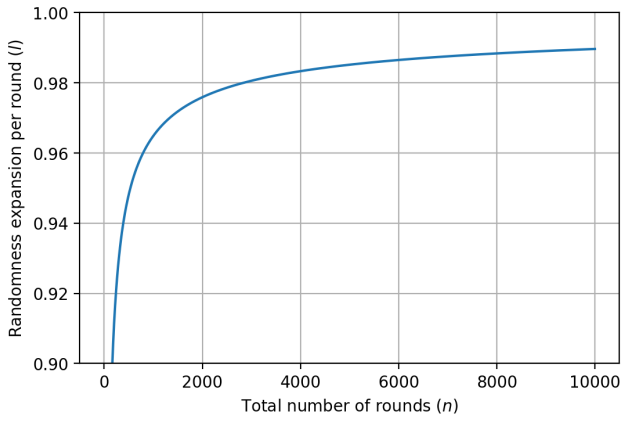


FIG. 2. The amount of secure randomness expansion per round as a function of total number of experimental rounds. We take $N = 5$ corresponding to the KCBS scenario and set $q = \frac{1}{\sqrt{n}}$ as the probability for spot-checking.

performed in Ref. [59] to estimate the randomness generated per key generation round. The maximum violation of the KCBS inequality experimentally observed in Table 2 of Ref. [59] is $\beta = 2.236$ which yields $\epsilon \approx 10^{-5}$. As a result $\epsilon_{\text{sec}} = 4\epsilon' \approx 10^{-2}$ is the security parameter. Secondly, the experiment performed 10^4 rounds of self-testing; We can consider that 10^2 rounds can be used for spot checking (corresponding to $q = \frac{1}{\sqrt{n}}$) out of a total of 10^4 rounds. In this case we find that the randomness expansion is $r = 0.9878$ bits per round (assuming that the accuracy of the KCBS self-test is still maintained with lower number of spot-check rounds).

Discussion.—We have put forth a scheme for locally certifying universal composability of a QRNG against a quantum adversary in a semi-device independent manner. We leverage the results of Ref. [60] to show that contextual correlations that offer close-to-maximal violation of a NC inequality are sufficient to certify secure randomness expansion. Unlike many of the semi-device independent schemes presented in Ref. [56, 57, 61], our scheme does not require trusted measurements or preparation sources and is secure against a quantum adversary (as opposed to a classical one). Instead, a different set of assumptions are required which can be (and have been) experimentally enforced [56, 57, 59, 62–66].

The security of our scheme is guaranteed by the self-test of the NC inequality and does not rely on entanglement, no-communication or even on the requirement of at least two parties. More importantly, it does not require the use of fully trusted components. On the downside, it is necessary to provide some characterization the QRNG for the self-test such that it obeys the required orthogonality and repeatability conditions. This increases the amount of resources that are required for the certification of security. However, this is expected as there is always a compromise between the level of trust and security and the amount of resources required to achieve it. It may be argued that achieving correlations that offer close-to-maximal violation of NC inequalities may be a downside of our scheme, but as is evidenced by the experiment in Ref. [59], it is already possible to achieve such correlations in the lab which makes our scheme much more relevant.

Similar to other works on randomness expansion based on self-testing in Bell scenarios [67], our results only hold for close-to-maximum violation of the NC inequality. By using the data of KCBS self-test in Ref. [59] we are able to show that the experiment is capable of generating 0.9878 bits per round of secure randomness with $\epsilon_{sec} \approx 10^{-2}$. Here, it can be noted that the value ϵ_{sec} may not be impressive from a cryptographic point of view. We attribute this to a limitation of Ref. [59] in which they only report violation of KCBS inequality upto 3 significant digits. It should also be noted that the order of our security parameter is similar to some of the robust self-tests in Bell scenarios [68–70] which can also be leveraged for randomness expansion. However, while so far no experiment has even come close to achieving such correlations for Bell scenarios (in a loophole-free manner), self-testing contextual correlations that can even achieve maximum quantum violation have already been experimentally realized. This makes our scheme experiment-ready.

Following our work several future directions of research can be identified. As a next step, it now remains to develop techniques for estimating the min-entropy conditioned on a quantum adversary for contextual correlations achieving arbitrary violation of the NC inequalities. Even though it is possible to experimentally achieve maximum quantum violation of NC inequalities in a loophole-free manner, the min-entropy approach would offer tighter bounds on randomness expansion rates. We have reasons to believe that device dependent techniques similar to the ones developed in Ref. [71] can be appropriately modified to fit our scheme and provide tighter bounds on the net randomness expansion in a SDI manner as shown in Ref. [72]. Secondly, deriving tighter bounds on the robustness for self-tests of contextual correlations would also be an interesting problem that will have significant impact on the security parameter in our scheme.

* jaskaran@gs.ncku.edu.tw
† cameron.foreman@quantinuum.com
‡ kishor.bharti1@gmail.com
§ adan@us.es

[1] A. Acín and L. Masanes, Certified randomness in quantum physics, Nature (London) **540**, 213 (2016).
[2] M. Herrero-Collantes and J. C. Garcia-Escartin, Quantum random number generators, Rev. Mod. Phys. **89**,

015004 (2017).
[3] X. Ma, X. Yuan, Z. Cao, B. Qi, and Z. Zhang, Quantum random number generation, npj Quantum Inf. **2**, 1 (2016).
[4] P. Bierhorst, E. Knill, S. Glancy, Y. Zhang, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, M. J. Stevens, and L. K. Shalm, Experimentally generated randomness certified by the impossibility of superluminal signals, Nature (London) **556**, 223 (2018).
[5] P. O. J.G. Rarity and P. Tapster, Quantum random-number generation and key sharing, J. Mod. Opt. **41**, 2435 (1994).
[6] J. F. Dynes, Z. L. Yuan, A. W. Sharpe, and A. J. Shields, A high speed, postprocessing free, quantum random number generator, Appl. Phys. Lett. **93**, 031109 (2008).
[7] H. Schmidt, Quantum-Mechanical Random-Number Generator, J. Appl. Phys. **41**, 462 (1970).
[8] T. Tsurumaru, T. Sasaki, and I. Tsutsui, Secure random number generation from parity symmetric radiations, Commun. Phys. **5**, 1 (2022).
[9] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, A fast and compact quantum random number generator, Rev. Sci. Instrum. **71**, 1675 (2000).
[10] P. X. Wang, G. L. Long, and Y. S. Li, Scheme for a quantum random number generator, J. of Appl. Phys. **100**, 056107 (2006).
[11] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, Optical quantum random number generator, J. Mod. Opt. **47**, 595 (2000).
[12] P. Bronner, A. Strunz, C. Silberhorn, and J.-P. Meyn, Demonstrating quantum random with single photons, Eur. J. Phys. **30**, 1189 (2009).
[13] M. Gräfe, R. Heilmann, A. Perez-Leija, R. Keil, F. Dreisow, M. Heinrich, H. Moya-Cessa, S. Nolte, D. N. Christodoulides, and A. Szameit, On-chip generation of high-order single-photon W-states, Nat. Photonics **8**, 791 (2014).
[14] Y.-Q. Nie, H.-F. Zhang, Z. Zhang, J. Wang, X. Ma, J. Zhang, and J.-W. Pan, Practical and fast quantum random number generation based on photon arrival time relative to external reference, Appl. Phys. Lett. **104**, 051110 (2014).
[15] A. Khanmohammadi, R. Enne, M. Hofbauer, and H. Zimmermanna, A monolithic silicon quantum random number generator based on measurement of photon detection time, IEEE Photonics J. **7**, 1 (2015).
[16] Q. Yan, B. Zhao, Q. Liao, and N. Zhou, Multi-bit quantum random number generation by measuring positions of arrival photons, Rev. Sci. Instrum. **85**, 103116 (2014).
[17] Y. Shen, L. Tian, and H. Zou, Practical quantum random number generator based on measuring the shot noise of vacuum states, Phys. Rev. A **81**, 063814 (2010).
[18] C. Gabriel, C. Wittmann, D. Sych, R. Dong, W. Mauerer, U. L. Andersen, C. Marquardt, and G. Leuchs, A generator for unique quantum random numbers based on vacuum states, Nat. Photonics **4**, 711 (2010).
[19] M. A. Wayne, E. R. Jeffrey, G. M. Akselrod, and P. G. Kwiat, Photon arrival time quantum random number generation, J. Mod. Opt. **56**, 516 (2009).
[20] M. Wahl, M. Leifgen, M. Berlin, T. Röhlicke, H.-J. Rahn, and O. Benson, An ultrafast quantum random number generator with provably bounded output bias based on photon arrival time measurements, Appl. Phys. Lett. **98**, 171105 (2011).
[21] M. Ren, E. Wu, Y. Liang, Y. Jian, G. Wu, and H. Zeng, Quantum random-number generator based on a photon-number-resolving detector, Phys. Rev. A **83**, 023820 (2011).
[22] M. J. Applegate, O. Thomas, J. F. Dynes, Z. L. Yuan, D. A. Ritchie, and A. J. Shields, Efficient and robust quantum random number generation by photon number detection, Appl. Phys. Lett. **107**, 071106 (2015).
[23] Quantum random number generation (QRNG) - ID Quantique (2023).
[24] Quantum random number generator (Tropos-QRNG) | QNu labs (2023).
[25] R. Colbeck, Quantum and relativistic protocols for secure multi-party computation, arXiv **arXiv:0911.3814** (2011).
[26] R. Colbeck and A. Kent, Private randomness expansion with untrusted devices, J. Phys. A: Math. Theor. **44**, 095305 (2011).
[27] P. J. Brown, S. Ragy, and R. Colbeck, A framework for quantum-secure device-independent randomness expansion, IEEE Trans. Inf. Theory **66**, 2964 (2020).
[28] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Random numbers certified by Bell's theorem, Nature (London) **464**, 1021 (2010).
[29] S. Pironio and S. Massar, Security of practical private randomness generation, Phys. Rev. A **87**, 012336 (2013).
[30] S. Fehr, R. Gelles, and C. Schaffner, Security and composability of randomness expansion from Bell inequalities, Phys. Rev. A **87**, 012335 (2013).
[31] O. Nieto-Silleras, C. Bamps, J. Silman, and S. Pironio, Device-independent randomness generation from several Bell estimators, New J. Phys. **20**, 023049 (2018).
[32] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics Physique Fizika **1**, 195 (1964).
[33] J. S. Bell, On the problem of hidden variables in quantum mechanics, Rev. Mod. Phys. **38**, 447 (1966).
[34] Z.-P. Xu, J. Steinberg, J. Singh, A. J. López-Tarrida, J. R. Portillo, and A. Cabello, Graph-theoretic approach to Bell experiments with low detection efficiency, Quantum **7**, 922 (2023).
[35] Y. Liu, Q. Zhao, M.-H. Li, J.-Y. Guan, Y. Zhang, B. Bai, W. Zhang, W.-Z. Liu, C. Wu, X. Yuan, H. Li, W. J. Munro, Z. Wang, L. You, J. Zhang, X. Ma, J. Fan, Q. Zhang, and J.-W. Pan, Device-independent quantum random-number generation, Nature (London) **562**, 548 (2018).
[36] M.-H. Li, X. Zhang, W.-Z. Liu, S.-R. Zhao, B. Bai, Y. Liu, Q. Zhao, Y. Peng, J. Zhang, Y. Zhang, W. J. Munro, X. Ma, Q. Zhang, J. Fan, and J.-W. Pan, Experimental realization of device-independent quantum randomness expansion, Phys. Rev. Lett. **126**, 050503 (2021).
[37] Y. Liu, X. Yuan, M.-H. Li, W. Zhang, Q. Zhao, J. Zhong, Y. Cao, Y.-H. Li, L.-K. Chen, H. Li, T. Peng, Y.-A. Chen, C.-Z. Peng, S.-C. Shi, Z. Wang, L. You, X. Ma, J. Fan, Q. Zhang, and J.-W. Pan, High-speed device-independent quantum random number generation without a detection loophole, Phys. Rev. Lett. **120**, 010503 (2018).
[38] L. K. Shalm, Y. Zhang, J. C. Bienfang, C. Schlager, M. J. Stevens, M. D. Mazurek, C. Abellán, W. Amaya, M. W.

Mitchell, M. A. Alhejji, H. Fu, J. Ornstein, R. P. Mirin, S. W. Nam, and E. Knill, Device-independent randomness expansion with entangled photons, Nat. Phys. **17**, 452 (2021).
[39] W.-Z. Liu, M.-H. Li, S. Ragy, S.-R. Zhao, B. Bai, Y. Liu, P. J. Brown, J. Zhang, R. Colbeck, J. Fan, Q. Zhang, and J.-W. Pan, Device-independent randomness expansion against quantum side information, Nat. Phys. **17**, 448 (2021).
[40] C. Portmann and R. Renner, Security in quantum cryptography, Rev. Mod. Phys. **94**, 025008 (2022).
[41] Z. Cao, H. Zhou, and X. Ma, Loss-tolerant measurement-device-independent quantum random number generation, New J. Phys. **17**, 125011 (2015).
[42] E. Passaro, D. Cavalcanti, P. Skrzypczyk, and A. Acín, Optimal randomness certification in the quantum steering and prepare-and-measure scenarios, New J. Phys. **17**, 113010 (2015).
[43] Y.-Q. Nie, J.-Y. Guan, H. Zhou, Q. Zhang, X. Ma, J. Zhang, and J.-W. Pan, Experimental measurement-device-independent quantum random-number generation, Phys. Rev. A **94**, 060301 (2016).
[44] F. Xu, J. H. Shapiro, and F. N. C. Wong, Experimental fast quantum random number generation using high-dimensional entanglement with entropy monitoring, Optica **3**, 1266 (2016).
[45] Z. Cao, H. Zhou, X. Yuan, and X. Ma, Source-independent quantum random number generation, Phys. Rev. X **6**, 011020 (2016).
[46] F. Bischof, H. Kampermann, and D. Bruß, Measurement-device-independent randomness generation with arbitrary quantum states, Phys. Rev. A **95**, 062305 (2017).
[47] I. Šupić, P. Skrzypczyk, and D. Cavalcanti, Measurement-device-independent entanglement and randomness estimation in quantum networks, Phys. Rev. A **95**, 042340 (2017).
[48] J. B. Brask, A. Martin, W. Esposito, R. Houlmann, J. Bowles, H. Zbinden, and N. Brunner, Megahertz-rate semi-device-independent quantum random number generators based on unambiguous state discrimination, Phys. Rev. Appl. **7**, 054018 (2017).
[49] M. Avesani, D. G. Marangon, G. Vallone, and P. Villoresi, Source-device-independent heterodyne-based quantum random number generator at 17 Gbps, Nat. Commun. **9**, 1 (2018).
[50] D. Drahi, N. Walk, M. J. Hoban, A. K. Fedorov, R. Shakhovoy, A. Feimov, Y. Kurochkin, W. S. Kolthammer, J. Nunn, J. Barrett, and I. A. Walmsley, Certified quantum random numbers from untrusted light, Phys. Rev. X **10**, 041048 (2020).
[51] C. Foreman, S. Wright, A. Edgington, M. Berta, and F. J. Curchod, Practical randomness amplification and privatisation with implementations on quantum computers, Quantum **7**, 969 (2023).
[52] A. Cabello, JoséM. Estebaranz, and G. García-Alcaine, Bell-Kochen-Specker theorem: A proof with 18 vectors, Phys. Lett. A **212**, 183 (1996).
[53] A. Cabello, S. Severini, and A. Winter, Graph-theoretic approach to quantum correlations, Phys. Rev. Lett. **112**, 040401 (2014).
[54] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, Phys. Rev. A **71**, 052108 (2005).
[55] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, J. Appl. Math. Mech. **17**, 59 (1967).
[56] M. Um, X. Zhang, J. Zhang, Y. Wang, S. Yangchao, D.-L. Deng, L.-M. Duan, and K. Kim, Experimental certification of random numbers via quantum contextuality, Sci. Rep. **3**, 1 (2013).
[57] M. Um, Q. Zhao, J. Zhang, P. Wang, Y. Wang, M. Qiao, H. Zhou, X. Ma, and K. Kim, Randomness expansion secured by quantum contextuality, Phys. Rev. Appl. **13**, 034077 (2020).
[58] P. Wang, J. Zhang, C.-Y. Luan, M. Um, Y. Wang, M. Qiao, T. Xie, J.-N. Zhang, A. Cabello, and K. Kim, Significant loophole-free test of Kochen-Specker contextuality using two species of atomic ions, Sci. Adv. **8**, 10.1126/sciadv.abk1660 (2022).
[59] X.-M. Hu, Y. Xie, A. S. Arora, M.-Z. Ai, K. Bharti, J. Zhang, W. Wu, P.-X. Chen, J.-M. Cui, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, J. Roland, A. Cabello, and L.-C. Kwek, Self-testing of a single quantum system from theory to experiment, npj Quantum Inf. **9**, 1 (2023).
[60] K. Bharti, M. Ray, A. Varvitsiotis, N. A. Warsi, A. Cabello, and L.-C. Kwek, Robust self-testing of quantum systems via noncontextuality inequalities, Phys. Rev. Lett. **122**, 250403 (2019).
[61] M. Pivoluska, M. Plesch, M. Farkas, N. Ružičková, C. Flegel, N. H. Valencia, W. McCutcheon, M. Malik, and E. A. Aguilar, Semi-device-independent random number generation with flexible assumptions, npj Quantum Inf. **7**, 1 (2021).
[62] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, Experimental non-classicality of an indivisible quantum system, Nature (London) **474**, 490 (2011).
[63] M. Jerger, Y. Reshitnyk, M. Oppliger, A. Potočnik, M. Mondal, A. Wallraff, K. Goodenough, S. Wehner, K. Juliusson, N. K. Langford, and A. Fedorov, Contextuality without nonlocality in a superconducting quantum system, Nat. Commun. **7**, 1 (2016).
[64] X. Zhan, P. Kurzyński, D. Kaszlikowski, K. Wang, Z. Bian, Y. Zhang, and P. Xue, Experimental detection of information deficit in a photonic contextuality scenario, Phys. Rev. Lett. **119**, 220403 (2017).
[65] M. Malinowski, C. Zhang, F. M. Leupold, A. Cabello, J. Alonso, and J. P. Home, Probing the limits of correlations in an indivisible quantum system, Phys. Rev. A **98**, 050102 (2018).
[66] A. Zhang, H. Xu, J. Xie, H. Zhang, B. J. Smith, M. S. Kim, and L. Zhang, Experimental test of contextuality in quantum and classical systems, Phys. Rev. Lett. **122**, 080401 (2019).
[67] L. Wooltorton, P. Brown, and R. Colbeck, Tight analytic bound on the trade-off between device-independent randomness and nonlocality, Phys. Rev. Lett. **129**, 150403 (2022).
[68] M. McKague, T. H. Yang, and V. Scarani, Robust self-testing of the singlet, J. Phys. A: Math. Theor. **45**, 455304 (2012).
[69] C. Bamps and S. Pironio, Sum-of-squares decompositions for a family of Clauser-Horne-Shimony-Holt-like inequalities and their application to self-testing, Phys. Rev. A **91**, 052111 (2015).
[70] I. Šupić, R. Augusiak, A. Salavrakos, and A. Acín, Self-testing protocols based on the chained Bell inequalities, New J. Phys. **18**, 035013 (2016).

[71] M. Araújo, M. Huber, M. Navascués, M. Pivoluska, and A. Tavakoli, Quantum key distribution rates from semidefinite programming, Quantum **7**, 1019 (2023).
[72] C. R. i. Carceller, L. N. Faria, Z.-H. Liu, N. Sguerso, U. L. Andersen, J. S. Neergaard-Nielsen, and J. B. Brask, Improving semi-device-independent randomness certification by entropy accumulation, arXiv 10.48550/arXiv.2405.04244 (2024).
[73] M. McKague, Device independent quantum key distribution secure against coherent attacks with memoryless measurement devices, New J. Phys. **11**, 103037 (2009).
[74] T. Scheidl, R. Ursin, J. Kofler, S. Ramelow, X.-S. Ma, T. Herbst, L. Ratschbacher, A. Fedrizzi, N. K. Langford, T. Jennewein, and A. Zeilinger, Violation of local realism with freedom of choice, Proc. Natl. Acad. Sci. U.S.A. **107**, 19708 (2010).
[75] F. Dupuis and O. Fawzi, Entropy accumulation with improved second-order term, IEEE Trans. Inf. Theory **65**, 7596 (2019).
[76] Y. Zhang, E. Knill, and P. Bierhorst, Certifying quantum randomness by probability estimation, Phys. Rev. A **98**, 040304 (2018).
[77] A. E. Rastegin, Relations for certain symmetric norms and anti-norms before and after partial trace, J. Stat. Phys. **148**, 1040 (2012).
[78] R. Renner and S. Wolf, Smooth Renyi entropy and applications, in *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.* (2004) pp. 233–.
[79] R. Renner, Security of quantum key distribution, arXiv 10.48550/arXiv.quant-ph/0512258 (2005).
[80] M. Tomamichel, R. Colbeck, and R. Renner, Duality between smooth min- and max-entropies, IEEE Trans. Inf. Theory **56**, 4674 (2010).
[81] M. Tomamichel, *Quantum Information Processing with Finite Resources* (Springer International Publishing, Cham, Switzerland).

Self-testing of NC inequalities

We start by defining a graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ which is defined as a pair of set of vertices $\mathcal{V} = \{v_i\}_i$ and a set of edges $\mathcal{E} = \{v_i, v_j\}_{i,j}$. If the pair of vertices $\{v_i, v_j\} \in \mathcal{E}$, then the corresponding vertices are termed as adjacent and will be hereafter denoted by $v_i \sim v_j$. We can consider different theories that assign probabilities to measurement events to characterize the different classes of correlations. Without loss of generality, let $p : \mathcal{V} \to [0, 1]$ be a probability assignment, such that a vertex v_i is mapped to a probability $p_i \in [0, 1]$. Moreover, since we only focus on exclusivity graphs, by definition of exclusive events it is required that $p_i + p_j \leq 1$ if $v_i \sim v_j$.

We can now define a canonical version of an NC inequality as a linear sum of all the probability assignments to the vertices of a graph $\mathcal{G}$. Mathematically, it is written as

$$\beta = \sum_{i=1}^{N} p(1|i) \leq \beta_{nc}(\mathcal{G}), \tag{11}$$

where $\beta_{nc}(\mathcal{G})$ maximum value of the sum of probabilities which is attained under the NC assumption that the probability assigned to a vertex i is independent of the assignment to vertex $j \neq i$. This value can be identified with the independence number of the graph $\mathcal{G}$.

In the approach of Ref. [53], the authors consider a set of N measurement events $\{e_i\}_{i=1}^{N}$ which are associated with the vertices of a graph $\mathcal{G}$, such that two mutually exclusive events e_i and e_j, which can be perfectly distinguished by two jointly measurable observables, are associated with adjacent vertices $v_i \sim v_j$. The corresponding graph $\mathcal{G}$ is termed as an exclusivity graph.

We consider different theories that assign probabilities to measurement events to characterize the different classes of correlations. Without loss of generality, let $p : \mathcal{V} \to [0, 1]$ be a probability assignment, such that a vertex v_i is mapped to a probability $p_i \in [0, 1]$. Moreover, since we only focus on exclusivity graphs, by definition of exclusive events it is required that $p_i + p_j \leq 1$ if $v_i \sim v_j$. Different theories can provide different probability assignments to all the events. For our purposes we focus on two of those, namely, deterministic NC and quantum theory.

A deterministic NC theory makes the assignment $p : \mathcal{V} \to \{0, 1\}$ such that an assignment to a particular event e_k is independent of the probability assignments for the events $\{e_j\}_{j\neq k}$. The polytope of deterministic non-contextual assignments corresponding to a graph $\mathcal{G}$, denoted by $\mathcal{P}_{nc}(\mathcal{G})$ forms a convex hull of all non-contextual assignments. Any probability assignment that does not lie in $\mathcal{P}_{nc}(\mathcal{G})$ is termed as contextual.

A quantum realization, $(\rho, \{\Pi_i\})$, of the graph $\mathcal{G}$ can be made by considering an underlying Hilbert space $\mathcal{H}$, a quantum state $\rho = |v_0\rangle \langle v_0|$ and a set of measurements $\mathcal{M}_i = \{\Pi_i, \mathbb{1} - \Pi_i\}$, $i \in \{1, \ldots, N\}$, such that each $\Pi_i = |v_i\rangle \langle v_i|$ corresponds to a vertex $v_i \in \mathcal{V}$ and for every $v_i \sim v_j$, $\mathrm{tr}\,(\Pi_i \Pi_j) = 0$. Let us denote the outcomes of each of the measurements $\mathcal{M}_i$ by $a \in \{0, 1\}$, where the outcome 1 corresponds to Π_i. Then the probability assigned to every vertex v_i can be written as a conditional probability $p(1|i) = \mathrm{tr}\,(\rho \Pi_i)$.

For certain quantum realizations, the inequality (11) can attain a maximum value $\beta_{qc}(\mathcal{G}) \geq \beta_{nc}(\mathcal{G})$, which is known as the quantum bound of the inequality. Using the graph theoretic formalism, the quantum bound can be identified with the Lovász number ϑ of the exclusivity graph $\mathcal{G}$. For more details see Refs. [53, 60].

In our work, for simplification, we only consider odd N-cycle NC scenarios in which $v_i \sim v_{i+1}$, such that $\mathrm{tr}\left(\Pi_i \Pi_{i+1}\right) = 0$ for $i \in \{1, \ldots, N\}$ and $N + 1 \equiv 1$. The NC bound of an odd N-cycle NC inequality is

$$\alpha = \frac{N-1}{2}, \tag{12}$$

while the quantum bound is found to be

$$\vartheta = \frac{N \cos \pi/N}{1 + \cos \pi/N}, \tag{13}$$

which can be achieved by choosing

$$\begin{aligned} |v_0\rangle &= (1, 0, 0)^T \\ |v_j\rangle &= (\cos\theta, \sin\theta \sin\phi_j, \sin\theta \cos\phi_j)^T \quad \forall j, \end{aligned} \tag{14}$$

where $\cos^2\theta = \frac{\cos \pi/N}{1+\cos \pi/N}$ and $\phi_j = \frac{j\pi(N-1)}{N}$ for $j \in \{1, \ldots, N\}$. Additionally, in this case, the probabilities $p(1|i)$ are found to be

$$p(1|i) = \frac{\cos \pi/N}{1 + \cos \pi/N} \quad \forall i, \tag{15}$$

which shows that each of the projectors $\Pi_i = |v_i\rangle \langle v_i|$ have an equal overlap with the state $\rho = |v_0\rangle \langle v_0|$.

Required assumptions

In this section we provide a detailed discussion on the assumptions that have been made while deriving our results. Many of them follow from the set of assumptions which are necessary in self-testing of contextual correlations, while some are technical assumptions which are standard in several randomness expansion schemes. Tests of the former have already been implemented in Ref. [59].

1. All measurements involved in our result are assumed to be repeatable. This means that consecutively and sequentially performing the same measurement yields the same result. A test of whether this assumption is satisfied in the experiments can be made by estimating the quantity $R = p(00|ii) + p(11|ii)$, where $p(aa|ii)$ is the probability of obtaining the outcome a twice in sequence when the measurement corresponding to setting i is implemented sequentially.

2. All measurements involved in our result are assumed to follow the required orthogonality relationships. A test of whether this assumption is satisfied in the experiments can be made by testing the quantities for $i \sim j$: $p(11|ij) = p(11|ji) = 0$, $p(10|ij) = p(01|ji)$, $p(01|ij) = p(10|ji)$ and $p(00|ij) = p(00|ji)$.

3. The measurement devices are memoryless. This assumption is required for the self-test and the derivation of our main result. While this assumption may sound too restrictive, it is rather often an implicit assumption in several cryptographic applications including QKD [73], randomness expansion [48] and tests of Bell inequality [74] to name a few. There also exist some methods which can allow to bypass this assumption [75, 76]. Specifically, it has been shown that the amount of private entropy produced is almost the same as in the case of no memory.

4. It is also assumed that Alice has access to some (short) private randomness to implement some operations in the protocol that require random selection. These operations include selecting whether a round will be a spot-check, randomly selecting the measurements or randomly selecting a key generation for post-selection. However, as we show in the main text, after a certain number of key generation rounds, the protocol can be used to expand the amount of this randomness (hence the name randomness expansion).

5. The lab of Alice is shielded from an adversary such that no information about her measurement outcomes can leak out.

6. We *do not* need to assume that the state prepared by Alice is a pure state. However, it is known that only (close-to) pure states can exhibit a violation of the n-cycle NC inequalities that are used in our result. Note, however that the same cannot be said for the measurements. We already assume that the measurements be repeatable and satisfy the orthogonality relationships.

Proofs of main theorems

In this section, we first prove our results for the case when Alice observes maximum violation of the NC inequality, i.e. $\beta = \beta_{qc}(\mathcal{G})$. Afterwards, we use similar techniques to prove our results in the case when $\beta \leq \beta_{qc}(\mathcal{G}) - \epsilon$, where $\epsilon \ll 1$.

We start by showing a particular extraction map that will be used throughout all our proofs.

Theorem 4. *Let $\mathcal{S}_1$ be the target quantum strategy which achieves $\beta = \beta_{qc}(\mathcal{G})$ for the NC inequality in Eq. (1). This inequality provides a local self-test of relations for all quantum strategies $\mathcal{S} : \left(\tilde{\rho}_A, \{\tilde{\Pi}_i\}_{i=1}^N\right)$ that also achieve $\beta = \beta_{qc}(\mathcal{G})$ and there exists an isometry $U = V_A^\dagger \otimes \mathbb{1}_E$ and an ancillary state $|\xi\rangle_E$ such that for the purification $|\psi\rangle_{AE}$ of the state ρ_A,*

$$U\left[\left(|\tilde{a}_i\rangle\langle\tilde{a}_i| \otimes \mathbb{1}_E\right)|\psi\rangle_{AE}\right] = \left(|a_i\rangle\langle a_i|v_0\rangle_A\right) \otimes |\xi\rangle_E \quad \forall a, i. \tag{16}$$

Proof. The purification $|\psi\rangle_{AE}$ of the state $\tilde{\rho}_A$ can be written as

$$|\psi\rangle_{AE} = \sum_l \sqrt{c_l}\,|\phi_l\rangle_A \otimes |l\rangle_E\,, \tag{17}$$

where $c_l \geq 0\ \forall l$ and the state $\tilde{\rho}_A = \mathrm{tr}_E\left(|\psi\rangle_{AE}\langle\psi|\right) = \sum_l c_l\,|\phi_l\rangle\langle\phi_l|$. Let us also define the projectors $\tilde{\Pi}_i^A = |\tilde{v}_i\rangle\langle\tilde{v}_i|_A \otimes \mathbb{1}_E$ (any other projector can also be taken instead of $\mathbb{1}_E$ without changing the proof). Following Definition 1, if Alice observes maximum violation of the NC inequality then the following relations hold: $|\phi_l\rangle_A = V_A\,|v_0\rangle_A$ and $|\tilde{v}_i\rangle\langle\tilde{v}_i|_A = V_A\,|v_i\rangle\langle v_i|\,V_A^\dagger\ \forall i$.

Next, we define our extraction map by considering the unitary operator

$$U = V_A^\dagger \otimes \mathbb{1}_E, \tag{18}$$

using which we can show that

$$\begin{aligned} U\left(\tilde{\Pi}_i^A\,|\psi\rangle_{AE}\right) &= U\left(|\tilde{v}_i\rangle\langle\tilde{v}_i|_A \otimes \mathbb{1}_E\right)U^\dagger U\,|\psi\rangle_{AE} \\ &= \sum_l \sqrt{c_l}\left(V_A^\dagger\,|\tilde{v}_i\rangle\langle\tilde{v}_i|_A\,V_A \otimes \mathbb{1}_E\right)\left(V_A^\dagger\,|\phi_l\rangle_A \otimes |l\rangle_E\right) \\ &= \left(|v_i\rangle\langle v_i|_A \otimes \mathbb{1}_E\right)\left(\sum_l \sqrt{c_l}\,|v_0\rangle \otimes |l\rangle_E\right) \\ &= \left(\Pi_i^A\,|v_0\rangle_A\right) \otimes |\xi\rangle_E \quad \forall i, \end{aligned} \tag{19}$$

where we have taken $|\xi\rangle_E = \sum_l \sqrt{c_l}\,|l\rangle_E$.

In a similar fashion as above we can show that

$$U\left[\left(\mathbb{1} - \tilde{\Pi}_i^A\right)|\psi\rangle_{AE}\right] = \left[\left(\mathbb{1} - \Pi_i^A\right)|v_0\rangle_A\right] \otimes |\xi\rangle_E\,. \tag{20}$$

Using Eqs. (19) and (20) together we can conclude the proof. □

Next, we can evaluate the post-measurement cq-state of Alice and Eve, which is given by

$$\rho_{AE} = \sum_{a=0}^{1} |a_i\rangle\langle a_i| \otimes \rho_E^{a_i} \quad \forall i, \tag{21}$$

where $\rho_E^{a_i} = \mathrm{tr}_A\left[\left(|\tilde{a}_i\rangle\langle\tilde{a}_i| \otimes \mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\right]$.

For the moment, let us focus on the term $\mathrm{tr}_A\left[\left(\tilde{\Pi}_i \otimes \mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\right]$ and using the fact that the isometry $U^\dagger U = \mathbb{1}$, we have

$$\begin{aligned} &\mathrm{tr}_A\left[\left(\tilde{\Pi}_i^A \otimes \mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\right] \\ &= \mathrm{tr}_A\left[U^\dagger U\left(\tilde{\Pi}_i^A \otimes \mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\,U^\dagger U\right] \\ &= \mathrm{tr}_A\left[U^\dagger\left(\Pi_i^A\,|v_0\rangle\langle v_0|_A \otimes |\xi\rangle\langle\xi|_E\,U\right)\right] \\ &= \mathrm{tr}\left(\Pi_i^A\,|v_0\rangle\langle v_0|\right)|\xi\rangle\langle\xi|_E \\ &= p(1|i)\,|\xi\rangle\langle\xi|_E\,, \end{aligned} \tag{22}$$

which, after performing a similar analysis for the term $\mathrm{tr}_A\left[\left(\mathbb{1}-\tilde{\Pi}_i\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\right]$ results in the post-measurement state being

$$\rho_{AE}=\sum_{a=0}^{1}p(a|i)\,|a_i\rangle\langle a_i|\otimes|\xi\rangle\langle\xi|_E\,, \tag{23}$$

where it can be seen that Eve is uncorrelated with the results of Alice's measurements. Next, we move on to prove our results in the case when Alice observes close-to-maximum violation of the N-cycle NC inequality.

Theorem 5. *Let $\mathcal{S}$ be a quantum strategy which achieves $\beta=\beta_{qc}(\mathcal{G})$ for the NC inequality in Eq. (1). This inequality provides a local self-test of relations for all quantum strategies $\tilde{\mathcal{S}}:(\tilde{\rho}_A,\{|\tilde{a}_i\rangle\langle\tilde{a}_i|\}_{a,i})$ that achieve $\beta=\beta_{qc}(\mathcal{G})-\epsilon$ ($\epsilon\ll 1$) and there exists an isometry $U=V_A^\dagger\otimes\mathbb{1}_E$ and an ancillary state $|\xi\rangle_E$ such that for the purification $|\psi\rangle_{AE}$ of the state $\tilde{\rho}_A$,*

$$\left\|\begin{aligned}&U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\,U^\dagger\\&\quad-\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|_A\otimes|\xi\rangle\langle\xi|_E\right)\end{aligned}\right\|\leq 2\epsilon'\quad\forall a,i. \tag{24}$$

Proof. The proof can be broken down into two steps. In the first step we bound the distance

$$\left\|\begin{aligned}&U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\,U^\dagger\\&\quad-\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\end{aligned}\right\|\leq\epsilon'\quad\forall a,i, \tag{25}$$

which can be done by noting that

$$\begin{aligned}&\left\|\begin{aligned}&U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\,U^\dagger\\&\quad-\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\end{aligned}\right\|\\&=\left\|U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)U^\dagger-|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right\|\\&=\left\|V_A^\dagger|\tilde{a}_i\rangle\langle\tilde{a}_i|\,V_A-|a_i\rangle\langle a_i|\right\|\\&\leq\epsilon',\end{aligned} \tag{26}$$

where in the last line we used the definition of the robust self-test of NC inequalities.

In the second step we bound the distance

$$\|\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}-\left(|a_i\rangle\langle a_i|\,\mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|_A\otimes|\xi\rangle\langle\xi|_E\right)\|\leq\epsilon' \tag{27}$$

which can be accomplished by noting that for any state $\tilde{\rho}_A=\sum_{l,k}c_{lk}\,|\phi_l\rangle\langle\phi_k|_A$ which is part of the strategy used to achieve the local self-test, must be ϵ'-close to the pure state that achieves the maximum value of the NC inequality for a fixed set of projectors. Therefore, we have

$$\left\|V_A^\dagger|\phi\rangle\langle\phi|_A\,V_A-|v_0\rangle\langle v_0|\right\|\leq\epsilon', \tag{28}$$

which can be used to show that

$$\left\|\begin{aligned}&V_A^\dagger|\phi\rangle\langle\phi|_A\,V_A\otimes|\xi\rangle\langle\xi|\\&\quad-|v_0\rangle\langle v_0|\otimes|\xi\rangle\langle\xi|\end{aligned}\right\|\leq\epsilon', \tag{29}$$

which implies

$$\begin{aligned}&\|\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}-\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|_A\otimes|\xi\rangle\langle\xi|_E\right)\|\\&=\left\|U|\phi\rangle\langle\phi|_A\otimes|\xi\rangle\langle\xi|\,U^\dagger-|v_0\rangle\langle v_0|_A\otimes|\xi\rangle\langle\xi|_E\right\|\\&=\||\psi\rangle\langle\psi|_{AE}-|v_0\rangle\langle v_0|_A\otimes|\xi\rangle\langle\xi|_E\|\\&\leq\epsilon'\end{aligned} \tag{30}$$

Next, using Eqs. (25) and (30) we can write

$$\begin{aligned}
&\left\| \begin{array}{l} U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}U^\dagger \\ \quad -\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|_A\otimes|\xi\rangle\langle\xi|_E\right) \end{array} \right\| \\
\leq &\left\| \begin{array}{l} U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}U^\dagger \\ \quad -\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE} \end{array} \right\| + \left\| \begin{array}{l} \left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE} \\ \quad -\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|_A\otimes|\xi\rangle\langle\xi|_E\right) \end{array} \right\| \\
= &\ 2\epsilon'
\end{aligned} \tag{31}$$

where we have the triangle property of the trace distance which states that for any three operators ρ, τ and σ, we have $\|\rho-\sigma\| \leq \|\rho-\tau\| + \|\tau-\sigma\|$.

□

Next, following the same technique as before we can show that the post-measurement cq-state of Alice and Eve is uncorrelated except with a probability $4\epsilon'$. Specifically, we wish to show that

$$\left\| \rho_{AE} - \sum_{a=0}^{1} p(a|i)\,|a_i\rangle\langle a_i|_A \otimes |\xi\rangle\langle\xi|_E \right\| \leq 4\epsilon', \tag{32}$$

where $\rho_{AE} = \sum_{a=0}^{1} |a_i\rangle\langle a_i|_A \otimes \rho_E^{a_i}$ and $\rho_E^{a_i} = \mathrm{tr}_A\left[\left(|a_i\rangle\langle a_i|_A\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_{AE}\right]$.

We can see this result by noting that from Eq. (24) and using a result from Ref. [77] which states that for an operator $X \in \mathcal{H}_1\otimes\mathcal{H}_2$, where $\mathcal{H}_i$ denotes the i-th Hilbert space, the norm satisfies $\|\mathrm{tr}_1(X)\| \leq \|X\|$, where $\mathrm{tr}_1(\cdot)$ is the partial trace over the system 1. Therefore, we have (for all permissible values of a and i)

$$\left\| \begin{array}{l} \mathrm{tr}_A\left[U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_E U^\dagger\right] \\ \quad -\mathrm{tr}_A\left[\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|\otimes|\xi\rangle\langle\xi|_E\right)\right] \end{array} \right\| \leq 2\epsilon'. \tag{33}$$

However, the term on the left hand side can be re-written as

$$\begin{aligned}
&\left\| \begin{array}{l} \mathrm{tr}_A\left[U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_E U^\dagger\right] \\ \quad -\mathrm{tr}_A\left[\left(|a_i\rangle\langle a_i|\otimes\mathbb{1}_E\right)\left(|v_0\rangle\langle v_0|\otimes|\xi\rangle\langle\xi|_E\right)\right] \end{array} \right\| \\
= &\left\| \mathrm{tr}_A\left[U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_E U^\dagger\right] - p(a|i)\,|\xi\rangle\langle\xi|_E \right\| \\
= &\left\| \begin{array}{l} |a_i\rangle\langle a_i|\otimes\mathrm{tr}_A\left[U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_E U^\dagger\right] \\ \quad -p(a|i)\,|a_i\rangle\langle a_i|\otimes|\xi\rangle\langle\xi|_E \end{array} \right\|,
\end{aligned} \tag{34}$$

which, for all permissible values of a and i, gives us

$$\left\| \begin{array}{l} |a_i\rangle\langle a_i|\otimes\mathrm{tr}_A\left[U\left(|\tilde{a}_i\rangle\langle\tilde{a}_i|\otimes\mathbb{1}_E\right)|\psi\rangle\langle\psi|_E U^\dagger\right] \\ \quad -p(a|i)\,|a_i\rangle\langle a_i|\otimes|\xi\rangle\langle\xi|_E \end{array} \right\| \leq 2\epsilon'. \tag{35}$$

Next, by using the property of sub-additivity of the trace distance we obtain

$$\left\| \rho_{AE} - \sum_{a=0}^{1} p(a|i)\,|a_i\rangle\langle a_i|\otimes|\xi\rangle\langle\xi|_E \right\| \leq 4\epsilon', \tag{36}$$

which proves our result.

Entropic quantities

In this section we describe some of the entropic quantities that we employ in our work. Our work mainly revolves around the operational formulation of min-entropy and its smooth version [78, 79]. We first define the guessing probability for an adversary Eve. For a given cq state $\rho_{KE} = \sum_k p(k)\,|k\rangle\langle k|\otimes\rho_k^E$, the guessing probability is defined

as the maximum probability with which Eve can guess the outcome of a measurement on the system K if she has access to system E. Mathematically, it is written as

$$p_{\text{guess}}(K|E) := \max_{\{M_k\}} \sum_k p(k) \text{tr}\left(M_k \rho_k^E\right), \tag{37}$$

where $\{M_k\}$ are the positive-operator-valued-measures (POVM) elements of a measurement that Eve implements on her system E. Following this definition we can now define min-entropy as

$$H_{\min}(K|E) := -\log\left(p_{\text{guess}}(K|E)\right), \tag{38}$$

where log is taken base 2. Here it should be noted that the quantity $H_{\min}(K|E)$ is evaluated for the state ρ_{KE}. However, in our work we are concerned with states that may only be δ-close (in trace norm) to the state ρ_{KE} such that $\delta \geq 0$. In such a case, we cannot directly employ the min-entropy and therefore need to use its smoothed version, $H_{\min}^{\delta}(K|E)$ which is defined over the set of states that are δ-close to the state ρ_{KE} as

$$H_{\min}^{\delta}(K|E) := \sup_{\rho'_{KE} \in \mathcal{B}_\delta(\rho_{KE})} H_{\min}(K|E)_{\rho'_{KE}}, \tag{39}$$

where $\mathcal{B}_\delta(\rho_{KE})$ denotes a δ-ball centered at ρ_{KE} with respect to the purified distance [80] and the min-entropy is defined over the set of states ρ'_{KE}. For a detailed analysis on min-entropy and smooth min-entropy we refer to Ref. [81]

Understanding Generalization in Quantum Machine Learning with Margins

Tak Hur[1] * Daniel K. Park[1 2] †

[1] *Department of Statistics and Data Science, Yonsei University, Seoul 03722, Republic of Korea*
[2] *Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea*

Keywords: Quantum Machine Learning, Generalization, Statistical Learning Theory, Margin

1 Introduction

Quantum machine learning (QML) stands out as an innovative application of quantum computation. The success of QML algorithm does not solely depend on how well the model fits the training data but, more importantly, on their ability to accurately predict the outcomes of previously unseen data. This crucial capability, known as generalization, has been extensively explored and analyzed through the lens of statistical learning theory. However, recent studies have highlighted the limitations of current understandings of generalization based on uniform bounds in both classical and quantum machine learning frameworks [1, 2]. In this work, we propose a *complexity measures* based on margin distribution, which can accurately capture the generalization performance of QML models.

2 Rethinking Generalization

Suppose there is an unknown joint probability distribution $\mathcal{D}$ governing the quantum state ρ and its corresponding label y. In this section, for simplicity, we will consider $\rho \in \mathbb{C}^{2^n \times 2^n}$ and $y \in \{-1, +1\}$, namely n-qubit binary classification task. With m independent and identically distributed (i.i.d) samples $S = \{\rho_i, y_i\}_{i=1}^m$, the goal is to find a hypothesis h^* with small true error $R(h^*) = \mathbb{E}_{\rho,y\sim\mathcal{D}}[\mathbb{1}_{\mathrm{sgn}(h^*(\rho))\neq y}]$. Since the true distribution $\mathcal{D}$ is unknown, we alternatively find h (from a hypothesis class $\mathcal{H}$) with small empirical risk, $\hat{R}_S(h) = 1/|S| \sum_{\rho,y\in S} \mathbb{1}_{\mathrm{sgn}(h(\rho))\neq y}$. For a hypothesis h, we define a generalization gap as a difference between true and empirical risk, $g(h) = R(h) - \hat{R}_S(h)$. A common way to understand generalization is to upper bound $g(h)$ by a *complexity measure* of the hypothesis class $\mathcal{H}$. For example, the hypothesis class of Quantum Neural Networks (QNN) with parameterized quantum circuit $U(\theta)$ and observable O can be expressed as $\mathcal{H}_{\mathrm{QNN}} = \{\rho \mapsto \mathrm{Tr}(OU(\theta)\rho U^\dagger(\theta)) : \theta \in \Theta\}$.

Theorem 1 (Rademacher Complexity Bound) *For any $\delta > 0$, with probability at least $1 - \delta$ over a sample S of size m drawn according to $\mathcal{D}$, following holds for any $h \in \mathcal{H}_{\mathrm{QNN}}$,*

$$R(h) \leq \hat{R}_S(h) + \hat{\mathfrak{R}}_S(\mathrm{sgn} \circ \mathcal{H}_{\mathrm{QNN}}) + 3\sqrt{\frac{\log(2/\delta)}{2m}}. \quad (1)$$

Here, $\mathrm{sgn} \circ \mathcal{H}_{\mathrm{QNN}} = \{\rho \mapsto \mathrm{sgn}(h(\rho)) : h \in \mathcal{H}_{\mathrm{QNN}}\}$, and $\hat{\mathfrak{R}}_S(\mathcal{H}) = \mathbb{E}_\sigma[\sup_{h\in\mathcal{H}} \frac{1}{m} \sum_i \sigma_i h(\rho_i)]$, where σ_i are i.i.d Rademacher random variables that takes value ± 1 with equal probability $1/2$.

Although Theorem 1 provides rigorous theoretical guarantee for generalization, it can result in vacuous upper bound, especially when $\mathcal{H}_{\mathrm{QNN}}$ is extensive enough to overfit random labels. For example, consider a corrupted sample $\tilde{S} = \{\rho_i, \tilde{y}_i\}_{i=1}^m$, where each $\tilde{y}_i$ are independently assigned ± 1 with a probability $1/2$, irrespective of the data ρ_i. Suppose $\mathcal{H}_{\mathrm{QNN}}$ can overfit the corrupted sample $\tilde{S}$, i.e. $\exists h \in \mathcal{H}_{\mathrm{QNN}}$ s.t. $\hat{R}_{\tilde{S}}(h) \approx 0$. Since the true error with respect to the corrupted distribution is 0.5 for all h, the analysis indicates that $0.5 \lesssim \hat{\mathfrak{R}}_S(\mathrm{sgn} \circ \mathcal{H}) + 3\sqrt{\log(2/\delta)/2m}$. Consequently, the Rademacher complexity bound $g(h) \lesssim 0.5$ is uninformative for binary classification.

Ref [1] highlighted that modern (classical) machine learning models, due to their large size and extensive numbers of parameters, can overfit random labels, suggesting our understanding of generalization is incomplete. Similarly, Ref [2] demonstrated that Quantum Convolutional Neural Networks (QCNNs) can also overfit random labels in the Quantum Phase Recognition problem, indicating this issue extends to quantum machine learning. It is important to note that this problem is not restricted to Rademacher Complexity bound, but any uniform generalization bounds, including the results from Ref [3, 4, 5, 6, 7, 8].

3 Margin based Generalization in Quantum Machine Learning

The concept of margin has been extensively explored since the early days of machine learning, offering theoretical foundations for Support Vector Machines [9]. The margin quantifies the difference between the output for correct labels and incorrect labels. More specifically, in k-class classification, for a data point (x, y), where $x \in \mathcal{X}$ and $y \in [k]$, and a classifier $f : \mathcal{X} \mapsto \mathbb{R}^k$, margin is defined as $f(x)_y - \max_{j\neq y} f(x)_j$. Here, the k-dimensional vector output of the classifier corresponds to the probability of

*takh0404@yonsei.ac.kr
†dkd.park@yonsei.ac.kr

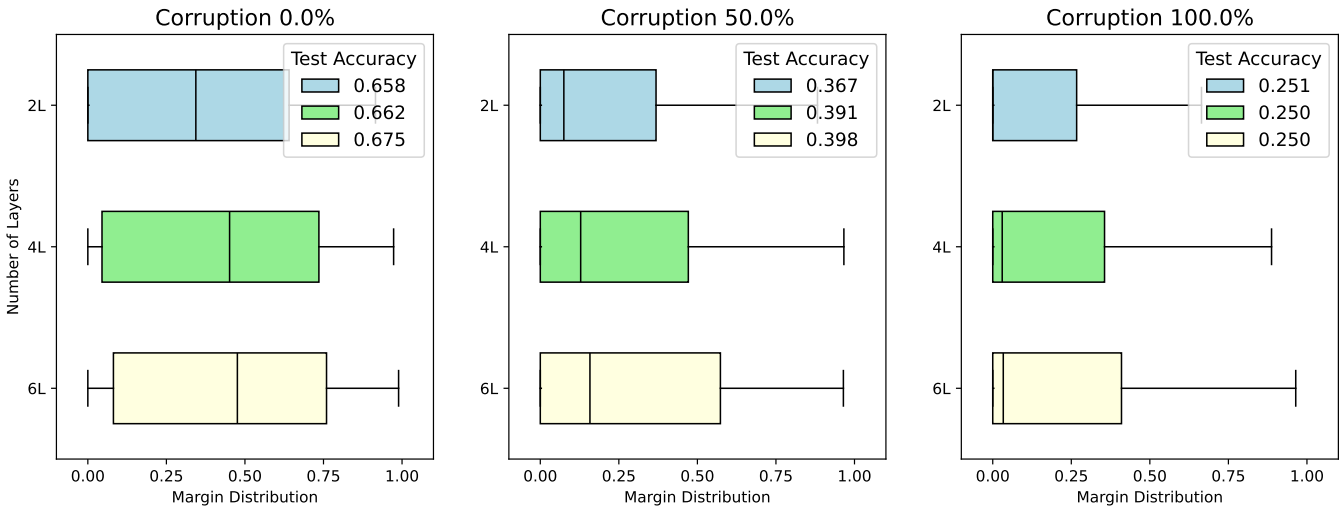


Figure 1: A (Tukey) box-and-whisker plot depicting the margin distributions of optimized Quantum Convolutional Neural Networks (QCNNs). The results for QCNNs with two, four, and six layers are displayed, along with their corresponding test accuracies. QCNNs were trained for 4-class classification task aimed at quantum phase recognition (QPR). The experiment was performed with varying degrees of label noise: QPR dataset with pure labels (left), half randomly labelled dataset (middle), and full randomly labelled datasets (right). As the noise (corruption) level increases, the margin distributions tend to exhibit a more pronounced skew towards the left, indicating that a greater proportion of samples are classified with smaller margins. Notably, the margin distribution exhibits a strong positive correlation with test accuracy across all scenarios.

assigning x to each class. Recently, Ref [10] proposed a generalization bound based on margins, normalized by a spectral norm of the weights, in the context of deep neural networks. It illustrated that complexity measures based on margin can address the shortcomings of uniform generalization bounds, as will be explained in more detail later in this section. Furthermore, it empirically demonstrated a significant correlation between margin-based measures and generalization error. Since then, margin is extensively used as a tool to understand generalization in (classical) machine learning [11, 12, 13, 14, 15].

The notion of margin can be extended to understand generalization performances of quantum machine learning models. Consider a k-class classification employing quantum neural networks, where the hypothesis class is defined as $\mathcal{H}_{\text{QNN}} = \{\rho \mapsto [\text{Tr}(\mathcal{M}_i U(\theta)\rho U^\dagger(\theta))]_{i=1}^k : \theta \in \Theta\}$. Here, measurement outcome of $\mathcal{M}_i$ represents the probability of assigning ρ to label i.

Theorem 2 (Margin Bound for QNN) *For any $\delta > 0$ and $\gamma > 0$, with probability at least $1-\delta$ over a sample S of size m drawn according to $\mathcal{D}$, following holds for any $h \in \mathcal{H}_{\text{QNN}}$,*

$$R(h) \leq \hat{R}_\gamma(h) + \frac{2}{\gamma}\hat{\mathfrak{R}}_S(\mathcal{H}_{\text{QNN}}) + 3\sqrt{\frac{\log(2/\delta)}{2m}}. \qquad (2)$$

Here, $\hat{R}_\gamma(h)$ represents the empirical margin error, quantifying the number of samples whose classification margin falls below the threshold γ. Formally, it is defined as $\hat{R}_\gamma(h) = 1/|S| \sum_{\rho,y\in S} \mathbb{1}_{h(\rho)_y \leq \max_{j\neq y} h(\rho)_j + \gamma}$.

The upper bound described in Equation 2 comprises of two competing terms: selecting a larger γ increases $\hat{R}_\gamma(h)$, while simultaneously decreasing $2\hat{\mathfrak{R}}_S(\mathcal{H}_{\text{QNN}})/\gamma$. According to Theorem 2, a hypothesis that classifies S with large margins results in a tighter upper bound, as opting for a larger γ does not significantly increase $\hat{R}_\gamma(h)$. Thus the *margin distribution*, which is the distribution of margins of sample S, plays a crucial role in comprehending the generalization performance of QML models.

Unlike uniform generalization bounds, margin bound provides distinct results depending on the distribution of the data. For instance, if we corrupt the sample from S to $\tilde{S}$ (and the data distribution from $\mathcal{D}$ to $\tilde{\mathcal{D}}$) as outlined in Section 2, the margin distribution will also vary, leading to a different generalization upper bound. If the margin distribution skews toward left as the data are corrupted, the margin bounds correctly explains the increasing generalization gap, a subtlety that uniform generalization bounds fail to capture.

Remark 1 *It is noteworthy that we can further upper bound the $\hat{\mathfrak{R}}_S(\mathcal{H}_{\text{QNN}})$ and achieve more interpretable results. For instance, Ref [6, 7, 8] analyze the Rademacher complexity of QNN through the lens of quantum resource theory. Additionally, Ref [4] quantifies the covering number of QNN based on the numbers of parameters. This result, combined with Dudley's entropy integral, can be utilized to establish the upper bound of Rademacher complexity [16]. However, in this study, our primary focus lies on exploring the margin distributions of quantum machine learning models and how margin-based complexity measures strongly correlate with generalization gap.*

4 Experimental Results

This section experimentally demonstrate strong correlation between margins and generalization performances

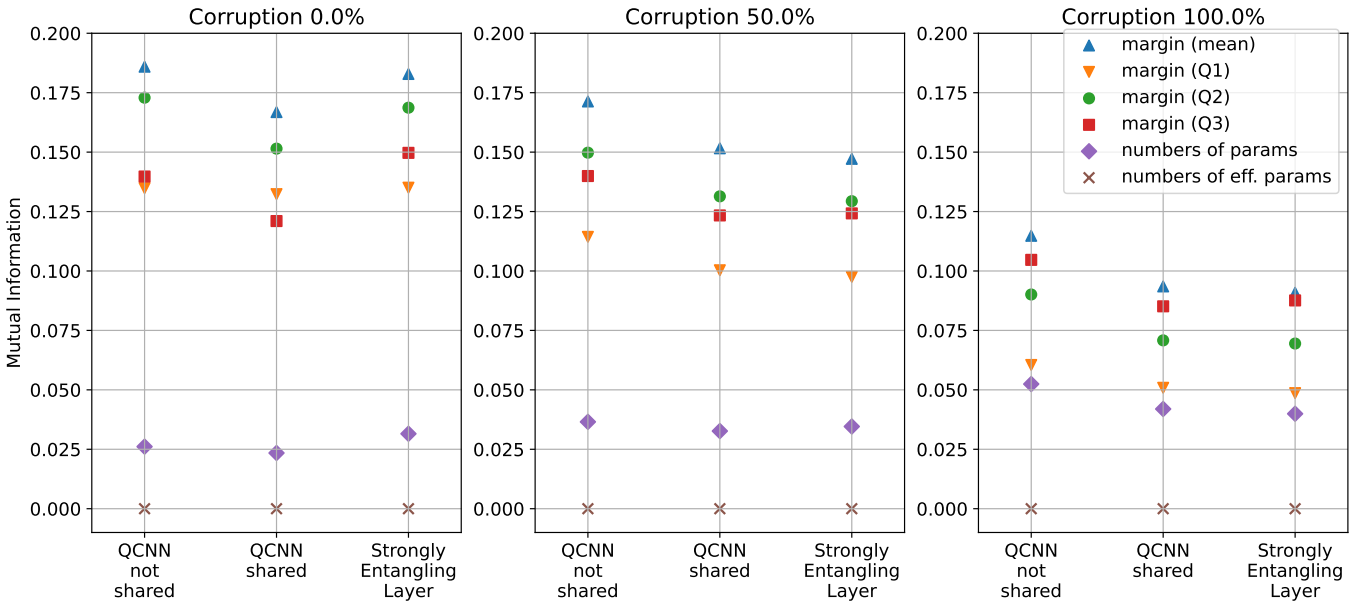


Figure 2: A comparative analysis demonstrating mutual information between generalization gap and various complexity measures. This includes four margin-based complexity measures: mean, lower quartile (Q1), median (Q2), and upper quartile (Q3), along with two parameter-based complexity measures: number of parameters and number of effective parameters. The experiments were conducted using three distinct variational ansatz: 1) QCNN without parameter sharing, 2) QCNN with parameter sharing 3) `StronglyEntanglingLayer` (see Ref [17] for details). Furthermore, the experiments were repeated under label corruption as outlined in Section 1. In all scenarios, margin-based complexity measure exhibited more mutual information about the generalization gap compared to parameter-based complexity measures. Notably, the mutual information tends to decrease with higher levels of corruption.

of QML models. We conducted extensive tests on Quantum Neural Networks (QNNs) with various hyperparameters, including circuit architecture, variational ansatz, number of layers, number of training samples, training batch size, maximum training iteration. The models were trained to perform the Quantum Phase Recognition (QPR) task, which involves classifying the phases of the ground state of the generalized cluster Hamiltonian, defined as $H = \sum_{j=1}^{n}(Z_j - J_1 X_j X_{j+1} - J_2 X_{j-1} Z_j X_{j+1})$ (see Ref [4, 2] for details). Additionally, the experiments were conducted under different levels of label noise: pure labels (r=0.0), half random labels (r=0.5), and fully random labels (r=1.0).

Figure 1 illustrates margin distributions of the optimized QCNNs in a box-and-whisker plot, alongside their respective test accuracies, conducted with varying numbers of QCNN layers. Across all layer configurations, the test accuracy decreases (and consequently, the generalization gap increases) as the labels are randomly corrupted with increasing levels of noise. The margin distributions exhibit significant leftward skew as the labels are corrupted. Thus, the margin bounds (Equation 2) correctly captures the generalization behavior under label corruption. Moreover, QCNNs with a larger number of layers tend to have higher test accuracy and exhibit right-skewed margin distributions, which further validates that margin distribution effectively captures the generalization performance in QML.

In Figure 2, we compare four margin-based complexity measures—mean, lower quartile, median, and upper quartile of the margin distribution—against parameter-based complexity measures. The latter includes 1) the number of parameters and 2) the number of effective parameters, which underwent significant changes during the optimization process. We evaluated mutual information between generalization gap and various complexity measures, treating them as random variables depending on sample S and hyperparameters of the models. Intuitively, a larger mutual information value indicates that the complexity measure contains more information about the generalization gap. Consequently, there is less uncertainty about generalization given the complexity measure. The experiments were conducted with three distinct variational ansatz: 1) QCNN without parameter sharing [18], 2) QCNN with parameter sharing [19, 20], and 3) `StronglyEntanglingLayers` [17]. Across all models, the mutual information values with margin-based complexity measures are significantly larger than those with parameters-based counterparts, indicating that margin distribution is more effective tool for understanding the generalization performance of QML models.

References

[1] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

[2] Elies Gil-Fuster, Jens Eisert, and Carlos Bravo-Prieto. Understanding quantum machine learning also requires rethinking generalization. *Nature Communications*, 15(1):1–12, 2024.

[3] Matthias C Caro, Elies Gil-Fuster, Johannes Jakob Meyer, Jens Eisert, and Ryan Sweke. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum*, 5:582, 2021.

[4] Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles. Generalization in quantum machine learning from few training data. *Nature communications*, 13(1):4919, 2022.

[5] Matthias C Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T Sornborger, Lukasz Cincio, Patrick J Coles, and Zoë Holmes. Out-of-distribution generalization for learning quantum dynamics. *Nature Communications*, 14(1):3751, 2023.

[6] Kaifeng Bu, Dax Enshan Koh, Lu Li, Qingxian Luo, and Yaobo Zhang. Rademacher complexity of noisy quantum circuits. *arXiv preprint arXiv:2103.03139*, 2021.

[7] Kaifeng Bu, Dax Enshan Koh, Lu Li, Qingxian Luo, and Yaobo Zhang. Statistical complexity of quantum circuits. *Physical Review A*, 105(6):062431, 2022.

[8] Kaifeng Bu, Dax Enshan Koh, Lu Li, Qingxian Luo, and Yaobo Zhang. Effects of quantum resources and noise on the statistical complexity of quantum circuits. *Quantum Science and Technology*, 8(2):025013, 2023.

[9] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.

[10] Peter L Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally-normalized margin bounds for neural networks. *Advances in neural information processing systems*, 30, 2017.

[11] Behnam Neyshabur, Srinadh Bhojanapalli, and Nathan Srebro. A pac-bayesian approach to spectrally-normalized margin bounds for neural networks. *arXiv preprint arXiv:1707.09564*, 2017.

[12] Yiding Jiang, Dilip Krishnan, Hossein Mobahi, and Samy Bengio. Predicting the generalization gap in deep networks with margin distributions. *arXiv preprint arXiv:1810.00113*, 2018.

[13] Behnam Neyshabur, Zhiyuan Li, Srinadh Bhojanapalli, Yann LeCun, and Nathan Srebro. Towards understanding the role of over-parametrization in generalization of neural networks. *arXiv preprint arXiv:1805.12076*, 2018.

[14] Alexander R Farhang, Jeremy D Bernstein, Kushal Tirumala, Yang Liu, and Yisong Yue. Investigating generalization by controlling normalized margin. In *International Conference on Machine Learning*, pages 6324–6336. PMLR, 2022.

[15] Yiding Jiang, Pierre Foret, Scott Yak, Daniel M Roy, Hossein Mobahi, Gintare Karolina Dziugaite, Samy Bengio, Suriya Gunasekar, Isabelle Guyon, and Behnam Neyshabur. Neurips 2020 competition: Predicting generalization in deep learning. *arXiv preprint arXiv:2012.07976*, 2020.

[16] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[17] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, M. Sohaib Alam, Shahnawaz Ahmed, Juan Miguel Arrazola, Carsten Blank, Alain Delgado, Soran Jahangiri, Keri McKiernan, Johannes Jakob Meyer, Zeyue Niu, Antal Száva, and Nathan Killoran. Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2020.

[18] Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vid Stojevic, Andrew G. Green, and Simone Severini. Hierarchical quantum classifiers. *npj Quantum Information*, 4(1):65, December 2018.

[19] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, December 2019.

[20] Tak Hur, Leeseok Kim, and Daniel K Park. Quantum convolutional neural network for classical data classification. *Quantum Machine Intelligence*, 4(1):3, 2022.

Quantum Pattern Engine

Ruo Cheng Huang[1] * Paul M. Riechers[1 2] † Mile Gu[1 3 4] ‡ Varun Narasimhachar[1 5] §

[1] *School of Physical and Mathematical Sciences, Nanyang Technological University, 637371 Singapore, Singapore*
[2] *Beyond Institute for Theoretical Science (BITS), San Francisco, CA, USA*
[3] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore*
[4] *MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, 117543, Singapore*
[5] *Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632*

Abstract. Quantum information-processing techniques enable work extraction from a system's inherently quantum features, in addition to its classical free energy. Meanwhile, the science of computational mechanics affords tools for the predictive modelling of non-Markovian stochastic processes. We combine tools from these two sciences to develop a technique for predictive work extraction from non-Markovian stochastic processes with quantum outputs. We demonstrate that this technique can extract more work than non-predictive quantum work extraction protocols, on one hand, and predictive work extraction without quantum information processing, on the other. We discover a phase transition in the efficacy of memory for work extraction from quantum processes, which has no classical precedent. Our work opens the prospect of machines that harness environmental free energy in an essentially quantum time-varying form.

Keywords: Quantum Thermodynamics, Stochastic Processes, Computational Mechanics, Temporal Correlations

1 Introduction

In the earliest heat engines, a combustible fuel was burned to maintain a temperature gradient between hot and cold heat reservoirs. The second law of thermodynamics holds that no engine can sustainably function with a single reservoir [1, 2, 3, 4]. While thought experiments such as Maxwell's demon and Szilard's engine initially appear to defy this law [5], a more complete understanding of thermodynamics resolved the apparent paradox: the resource powering the engine need not be a temperature gradient, but may be *any* form of free energy—even information [6, 7, 8, 9, 10]. The emerging field of quantum thermodynamics has continued to expand the scope of "fuel" to increasingly general forms of free energy. There has been both theoretical and experimental advancement in constructing engines that can harness the free energy locked up in quantum coherence, over and above classical free energy [11, 12, 13, 14, 15].

The story does not stop there—in addition to *static* fuel, there is also a *dynamical* fuel-like resource embodied by complex thermodynamic processes. The framework of *computational mechanics* in complexity science offers powerful techniques for the characterization and manipulation of stochastic processes. The future behaviour of such a process in general cannot be known perfectly using data from its past. Nevertheless, temporal correlations, i.e., patterns in a process's behaviour over time, enable prediction. These correlations may even be *non-Markovian*, whereby the future of a process depend not only on its present, but also on its distant past. Epsilon machines and their quantum extensions [16, 17, 18] perform memory-optimal predictive modelling of stochastic processes. Pattern extractors [14, 19] leverage prediction to extract useful work from the classical free energy present in temporal patterns, exchanging heat with a single bath. However, these predictive engines are not equipped to harness the free energy locked up in quantum degrees of freedom.

1.1 Our contributions

Here, we develop the theoretical prototype for a *predictive quantum engine*: a machine that charges a battery by feeding on a multipartite quantum system whose parts are temporally correlated via a classical stochastic process [**?**]. It can extract free energy beyond what is accessible to current quantum engines or classical predictive engines. We present a systematic construction of such an engine for arbitrary classical processes and quantum output states. We illustrate its application on example stochastic processes of correlated non-orthogonal qubits. (Fig. 1). We also use this test case to benchmark the performance of our engine against various alternatives, including one without coherent quantum information processing, and one without predictive functionality. Our predictive quantum engine outperforms these alternatives in terms of work output. We show that parametrized processes of correlated non-orthogonal quantum outputs exhibit phase boundaries between parametric regions where memory of past observations can and cannot enhance the work yield—despite the apparently smooth change of memoryful correlations in the process across this boundary. The sudden lack of memory advantage is thus fundamentally thermodynamic (since prediction per se has more freedom than during work extraction) and fundamentally quantum (since classical engines can exploit all the process's inherent memory). Finally, we generalize the Information Processing Sec-

*ruocheng001@e.ntu.edu.sg
†pmriechers@gmail.com
‡gumile@ntu.edu.sg
§varun.achar@gmail.com

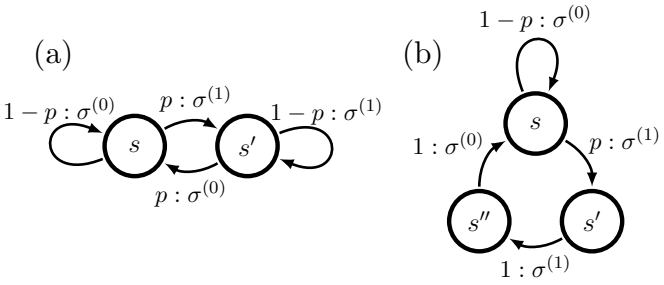


Figure 1: Latent-state sources of correlated quantum processes. Each arrow represents a transition between latent states; the label $p:\sigma^{(x)}$ indicates that the transition happens with probability p and produces a quantum state $\sigma^{(x)}$. (a) Perturbed-coin process. (b) 2-1 golden-mean process.

ond Law (IPSL) to the quantum regime and derive the fundamental bounds on a quantum pattern engine's performance.

2 Framework

Rather than using memoryless quantum sources which produce independent and identically distributed (IID) quantum states at each discrete time [20, 21], we consider general *finite-state sources of quantum states* which can create nontrivial correlation across time. Some simple examples are depicted in Fig. 1. these memoryful sources of states can be represented by a hidden Markov Model (HMM). The states generated are separable but can have non-classical correlations in the form of discord [22]. These sources generalize the kindred 'classically controlled qubit sources' of Ref. [23].

We restrict the the engine to possess no quantum memory and limited classical memory. Each of the quantum states generated also has an immediate expiration date, hence the quantum states generated must be fed into the engine one at a time rather than storing everything for later time. We allow the HMM to be arbitrary in its alphabet, size, statistics as well as the quantum outputs' dimensionality and purity. Lastly, we assume that the source of the fuel tape is known exactly, which entails the complete knowledge of the HMM

The engine will operate relying on its internal memory which keeps track of "belief states", $\boldsymbol{\eta}_t$. The memory will allow the engine to predict what the next expected state, ξ_t, will be. The engine then attempts to extract work from the quantum states based on the identity of ξ_t. The battery storing the work will eventually be measured and the memory will be updated. The process proceeds cyclically as shown in Fig. 2.

3 Results

Here we provide a summary of our results.

1. We generalised the so-called "mixed-state presentation" into the quantum regime where the states are non-orthogonal [24, 25, 26, 27, 28].

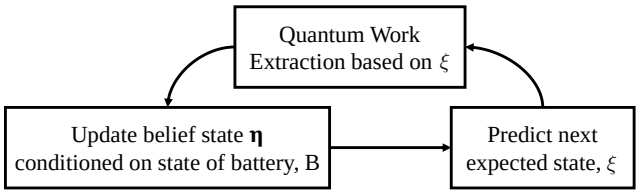


Figure 2: The protocol proceeds cyclically to fine-tune the belief state.

2. We demonstrated the superior performance of our engine by comparing it against other engines that lacked either memory or the ability to operate coherently on quantum states.

3. We discovered a phase transition in efficacy of knowledge in work extraction with respect to process parametrization. This along with the general performance of our engine can be found in Fig. 3

4. Finally, we provide a fundamental limit of this quantum pattern engine by invoking the quantum information processing second law to act as an upper bound.

4 Discussion

We developed the theoretical prototype for a *quantum pattern engine*: a machine that can adaptively extract useful work from quantum stochastic processes by exploiting knowledge of the temporal patterns they contain. We witnessed that, in the presence of coherence, the memory-assisted quantum approach will always outperform the memory-assisted classical approach. We also demonstrated its advantage over engines that can only harness static quantum resources. We found a phase transition marking the onset of memory advantage. It is an open question whether this phase transition coincides with the onset of quantum discord.

We found how to update the state of knowledge about any latent-state generator of a quantum process, given any POVM on the current quantum output. Furthermore, the fundamental thermodynamic limits of work extraction from correlated multi-partite quantum systems was found, setting the ultimate benchmark.

Despite the advances presented here, many open questions remain for future work.Although designing the protocol guarantees maximal work extraction locally in time, it remains an interesting open question whether there is a superior steady-state approach that sacrifices short-term work extraction for greater knowledge and long-term returns. It may be possible to extend our method to more complex quantum processes, e.g., to those with entangled temporal correlations. This would, however, likely require a quantum memory. On the other hand, our method can immediately be adapted to applications where the pattern is spatial instead of temporal (e.g., states of many-body systems), and where

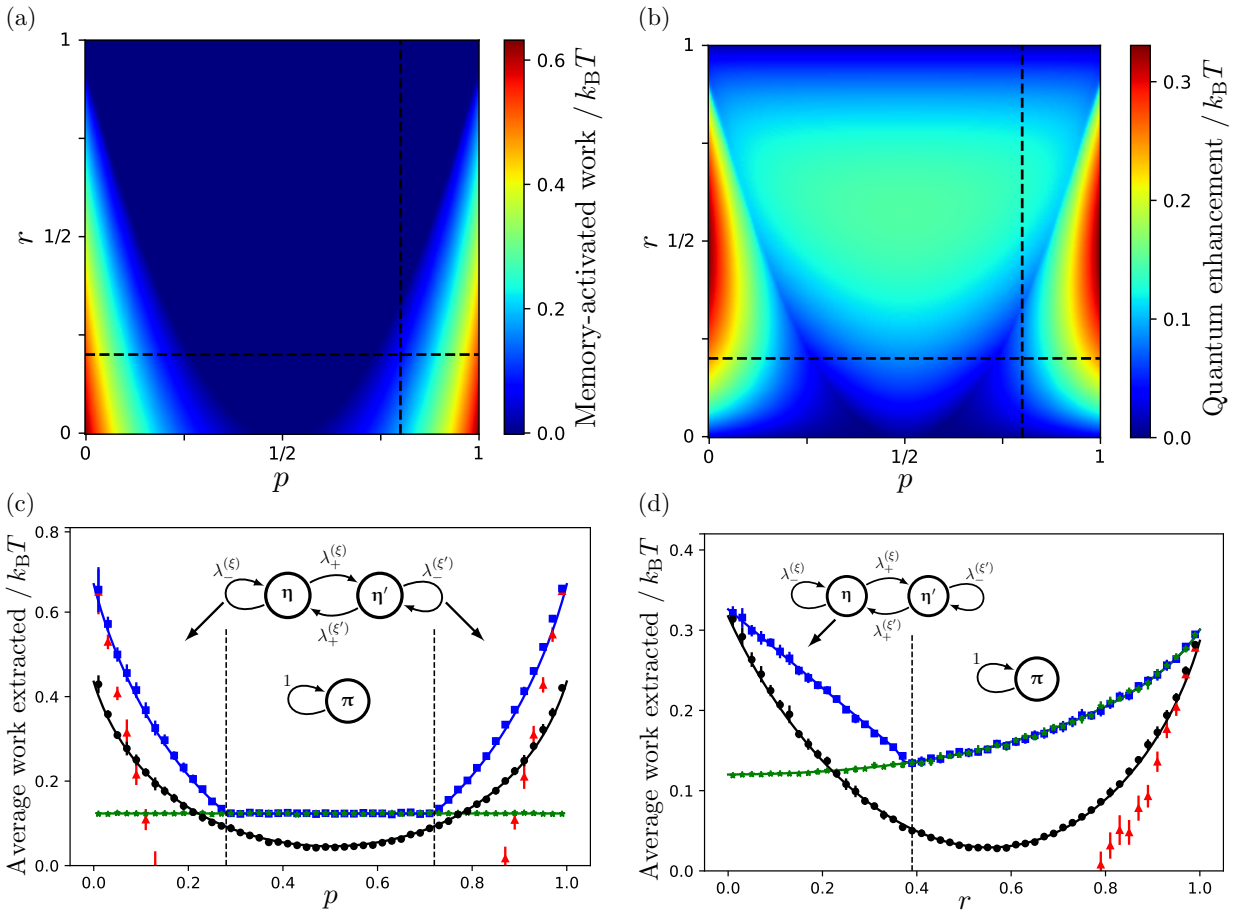


Figure 3: Comparison between average work-extraction rates of various approaches. (a) Memory enhancement of work extracted, compared to memoryless quantum approach. (b) Quantum enhancement of work extraction, compared to memoryful classical approach. Panels (c) and (d) reveal phase transitions in memory enhancement through cross-sections of parameter space. Analytic results (solid lines) and simulations (markers) are shown. Blue (squares) represents memory-assisted quantum approach; black (circles) represents memory-assisted classical approach; red (triangles) represents overcommitment approach; green (stars) represents memoryless quantum approach.

the engine is constrained to operate locally on small regions at a time.

For full detail of the paper, please refer to [29]. If time permits, we shall present our most recent work of utilizing techniques from reinforcement learning such as dynamic programming to achieve the maximal work.

References

[1] Daniel V Schroeder. An introduction to thermal physics, 1999.

[2] Stephen J Blundell and Katherine M Blundell. *Concepts in thermal physics*. Oup Oxford, 2009.

[3] James Sethna. *Statistical mechanics: entropy, order parameters, and complexity*, volume 14. Oxford University Press, USA, 2021.

[4] Herbert B Callen. Thermodynamics and an introduction to thermostatistics, 1998.

[5] Leo Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioral Science*, 9(4):301–310, 1964.

[6] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM journal of research and development*, 5(3):183–191, 1961.

[7] Charles H Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

[8] Dibyendu Mandal and Christopher Jarzynski. Work and information processing in a solvable model of Maxwell's demon. *Proceedings of the National Academy of Sciences*, 109(29):11641–11645, 2012.

[9] Dibyendu Mandal, HT Quan, and Christopher Jarzynski. Maxwell's refrigerator: an exactly solvable model. *Physical review letters*, 111(3):030602, 2013.

[10] Juan MR Parrondo, Jordan M Horowitz, and Takahiro Sagawa. Thermodynamics of information. *Nature physics*, 11(2):131–139, 2015.

[11] Paul Skrzypczyk, Anthony J Short, and Sandu Popescu. Work extraction and thermodynamics for individual quantum systems. *Nature communications*, 5, 2014.

[12] Johan Åberg. Catalytic coherence. *Physical review letters*, 113(15):150402, 2014.

[13] Kamil Korzekwa, Matteo Lostaglio, Jonathan Oppenheim, and David Jennings. The extraction of work from quantum coherence. *New Journal of Physics*, 18(2):023045, 2016.

[14] Andrew JP Garner, Jayne Thompson, Vlatko Vedral, and Mile Gu. Thermodynamics of complexity and pattern manipulation. *Physical Review E*, 95(4):042140, 2017.

[15] Matteo Lostaglio. Thermodynamic laws for populations and quantum coherence: A self-contained introduction to the resource theory approach to thermodynamics. *arXiv preprint arXiv:1807.11549*, 2018.

[16] James P Crutchfield and Karl Young. Inferring statistical complexity. *Physical review letters*, 63(2):105, 1989.

[17] Cosma Rohilla Shalizi and James P Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of statistical physics*, 104(3):817–879, 2001.

[18] James P Crutchfield. Between order and chaos. *Nature Physics*, 8(1):17–24, 2012.

[19] Alexander B. Boyd, Dibyendu Mandal, and James P. Crutchfield. Correlation-powered information engines and the thermodynamics of self-correction. *Phys. Rev. E*, 95:012152, Jan 2017.

[20] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2010.

[21] Philipp Strasberg, Gernot Schaller, Tobias Brandes, and Massimiliano Esposito. Quantum and information thermodynamics: A unifying framework based on repeated interactions. *Phys. Rev. X*, 7:021003, Apr 2017.

[22] Kavan Modi, Tomasz Paterek, Wonmin Son, Vlatko Vedral, and Mark Williamson. Unified view of quantum and classical correlations. *Phys. Rev. Lett.*, 104:080501, Feb 2010.

[23] Ariadna E Venegas-Li, Alexandra M Jurgens, and James P Crutchfield. Measurement-induced randomness and structure in controlled qubit processes. *Physical Review E*, 102(4):040102, 2020.

[24] James P Crutchfield. The calculi of emergence: computation, dynamics and induction. *Physica D: Nonlinear Phenomena*, 75(1-3):11–54, 1994.

[25] Christopher J Ellison, John R Mahoney, and James P Crutchfield. Prediction, retrodiction, and the amount of information stored in the present. *Journal of Statistical Physics*, 136(6):1005–1034, 2009.

[26] Sarah E Marzen and James P Crutchfield. Nearly maximally predictive features and their dimensions. *Physical Review E*, 95(5):051301, 2017.

[27] Paul M Riechers and James P Crutchfield. Spectral simplicity of apparent complexity. I. The nondiagonalizable metadynamics of prediction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(3):033115, 2018.

[28] Alexandra M Jurgens and James P Crutchfield. Shannon entropy rate of hidden Markov processes. *Journal of Statistical Physics*, 183(2):1–18, 2021.

[29] Ruo Cheng Huang, Paul M Riechers, Mile Gu, and Varun Narasimhachar. Engines for predictive work extraction from memoryful quantum stochastic processes. *Quantum*, 7:1203, 2023.

Realization of algorithmic identification of cause and effect in quantum correlations

Zhao-An Wang[1 2 *]

[1] *CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei, 230026, China*

[2] *CAS Center For Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, 230026, China*

Abstract. Causal inference revealing causal dependencies between variables from empirical data has found applications in multiple sub-fields of scientific research. Here, we have devised a photonic setup and experimentally realized an algorithm capable of identifying any two-qubit statistical correlations generated by the two basic causal structures under an observational scenario, thus revealing a universal quantum advantage in causal inference over its classical counterpart. We further demonstrate the explainability and stability of our causal discovery method, which is widely sought in data processing algorithms. Employing a fully observational approach, our result paves the way for studying quantum causality in general settings.

Keywords: quantum algorithm, causal structures, quantum correlations

1 Introduction

According to the Reichenbach's common cause principle (RCCP) [1], a latent variable that jointly influences two events can always produce the same correlation as a direct causal link between the two events. The consequence of RCCP for causal identification is the inability to differentiate between a common cause (CC) and a direct cause (DC) using observational data generated by classical entities. Given that quantum correlations have drastically different behaviors from their classical counterparts, it is natural to consider the extent to which a quantum perspective on correlations and causations can overcome the limitations imposed by RCCP. The question has motivated a number of works exploring the quantum advantages in identifying causal relationships [3, 4, 5]. In this study, we tackle the quantum counterpart of the two-point causal identification problem. Our aim is to determine whether the causal structure of a two-point qubit correlation is induced by the same particle going through a unitary quantum channel which amounts to a DC, or two (possibly entangled) particles being successively measured which constitutes a CC.

2 Quantum causality with geometry

In classical causal theory, a DC implies that an earlier variable directly influences a subsequent variable, while a CC signifies a confounding factor co-influencing the two variables. In quantum theory, the notion of variables is replaced by quantum states. As shown in Fig. 1a and b, a quantum channel connecting input and output states can serve as a quantum analog of DC, whereas a bipartite system consisting of two subsystems gives rise to a quantum CC. The concept can be formulated with a quantum comb representation where a quantum gate switches to an either identity or SWAP gate determining the latent causal structure (Fig. 1c). We need to note that in the classical regime, it is insufficient to distinguish the causal structures by merely observing the temporal correlation of the two states and intervention such as placebos in clinical examinations is needed. On the contrary, one can take advantage of quantum correlations to distinguish them [2].

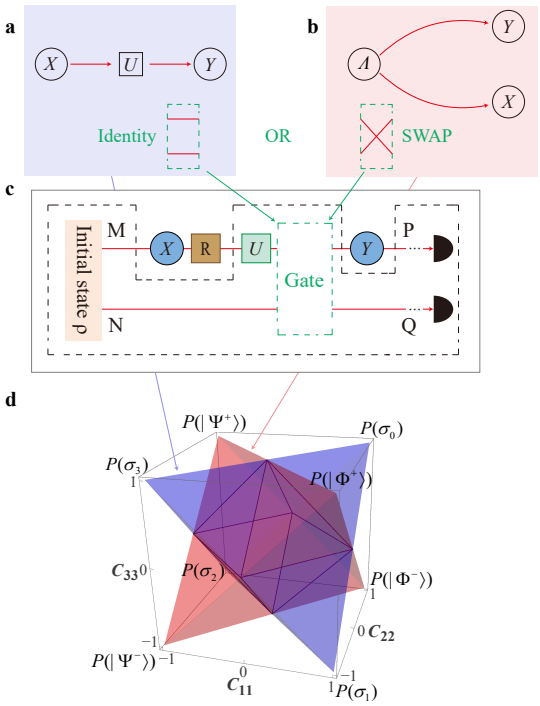


Figure 1: **Perspectives of two-point quantum causality. a, b**: Directed acyclic graph of direct cause (DC) and common cause (CC). **c**: Quantum comb representation of causal structures. **d**: Geometric description of two-qubit causal structures.

Moreover, one can map the two-point quantum correlations derived from canonical Pauli measurements to points in a three-dimensional space formulating a geo-

*zawang@mail.ustc.edu.cn

Algorithm 1 The algorithm for discrimination of the two-point qubit causal structures. $\mathcal{T}_{\mathrm{DC}}$ ($\mathcal{T}_{\mathrm{CC}}$) is the DC- (CC-) tetrahedron in the main text, whose four vertices are $(+1,+1,+1)^{\mathrm{T}}, (+1,-1,-1)^{\mathrm{T}}, (-1,+1,-1)^{\mathrm{T}}, (-1,-1,+1)^{\mathrm{T}}$ $((-1,-1,-1)^{\mathrm{T}}, (-1,+1,+1)^{\mathrm{T}}, (+1,-1,+1)^{\mathrm{T}}, (+1,+1,-1)^{\mathrm{T}})$. The quantities δ, ε and ε' are cutoff values and can be conveniently chosen. The function $\mathcal{D}$ means the Euclidean distance of its two arguments.

Require: A two-point Pauli correlation $\mathbf{P} = (C_{11}, C_{22}, C_{33})^T, C_{ii} := p(x = y|\sigma_i, \sigma_i) - p(x \neq y|\sigma_i, \sigma_i), \mathbf{P} \in \mathcal{T}_{\mathrm{DC}} \cup \mathcal{T}_{\mathrm{CC}}$. The sum of all entries of $\mathbf{P}$ is b
Ensure: The latent causal structure of the correlation, DC or CC
if $\mathbf{P} \notin \mathcal{T}_{\mathrm{DC}} \cap \mathcal{T}_{\mathrm{CC}}$ **then** ▷ Distinguishable by Pauli correlations
  if $\mathbf{P}$ is in the DC tetrahedron **then return** DC
  else return CC
  end if
else if $1 - b < \delta$ **then** ▷ Distinguishable by symmetrically modified Pauli correlations
  calculate V
  measure the new correlation C_{33}^V
  if $1 - C_{33}^V < \varepsilon$ **then return** DC
  else return CC
  end if
else ▷ Distinguishable by asymmetrically modified Pauli correlations
  calculate $V = V_2 V_1$, $V' = V_2 \sigma_1 V_1$ ▷ V_1 and V_2 are obtained in two consecutive rounds
  measure the new correlation $\mathbf{P} = (C_{11}^{V'}, C_{22}^{V'}, C_{33}^{V'})^T$
  if $\mathcal{D}(\mathbf{P}_{V'}, \mathbf{P}(\sigma_3)) < \varepsilon'$ **then return** DC
  else return CC
  end if
end if

metric representation [6]. In Fig. 1d, the blue tetrahedron encompasses the set of DC cases while the red one encompasses the set of CC cases [7].

3 Experimental demonstrating Algorithmic causal identification

As the two tetrahedra are non-identical, cases observed in the disjoint areas will imply causation. However, in the overlapping area, both DC and CC causal explanations are possible and extra measurements are needed. We perform a quantum algorithm in Algorithm 1 to eliminate this ambiguity. Generally, by exerting a pair of delicately designed operators, the algorithm moves the points out of the overlapping area thus the causal structures become distinguishable.

We implement the quantum causal identification algorithm by realizing the comb representation in Fig. 1c in an optical platform as depicted in Fig. 2. The quantum gate that controls the latent causal structure is realized by a PBS based Sagnac ring. In Fig. 3, We test two families of temporally-ordered quantum systems, of which causality is not unveiled by their Pauli correlations and the initial state ρ is pure. The result shows that in the second measurement, when the latent causal structure is DC (blue), we move the ambiguous points (yellow) to the top of the cube, leaving the overlapping area; while the red points representing CC structures can not reach the top, thus the causal structure is successfully distinguished. Similarly, we also extend to cases where the initial states are mixed (not shown limited by the length). The distinguishability is quantified by two distance measures $1 - C_{33}^V$ and $\mathcal{D}(\mathbf{P}_{V'}, \mathbf{P}(\sigma_3))$ and the distinguishing criteria are set with cutoff values ε and ε'.

4 Discussion

The algorithm is built upon quantum correlation and unitary operations within an observational scheme, eliminating the need for interventions, which sets it apart from its classical counterpart and results in reduced resource requirements. Besides the framework we have used, physicists have also researched other formalisms such as pseudo-density matrix formalism and process matrix formalism, to properly describe causal structures and seek less intervening and go on less resource consuming inference methods. Digging out more causal inference criteria that take the advantages of the quantum world could be a natural trend. Therefore, we hope that our comprehensive exploration of two-point quantum correlations from a causal perspective will contribute to the further advancement of quantum causal inference and pave the way for the construction of causal networks involving various quantum resources.

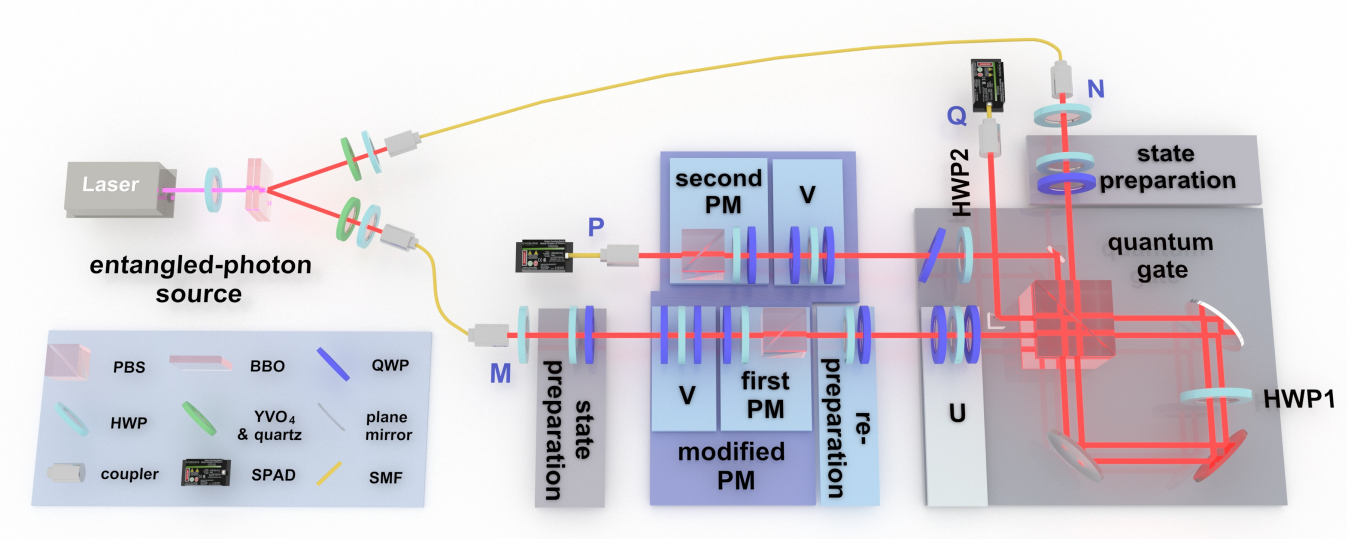


Figure 2: **Experimental setup.** The checkpoints M, N, P, and Q match the notations in Fig. 1. The monochromatic panels were inside the quantum comb and the blue panels were accessible to the causal discovery algorithm.

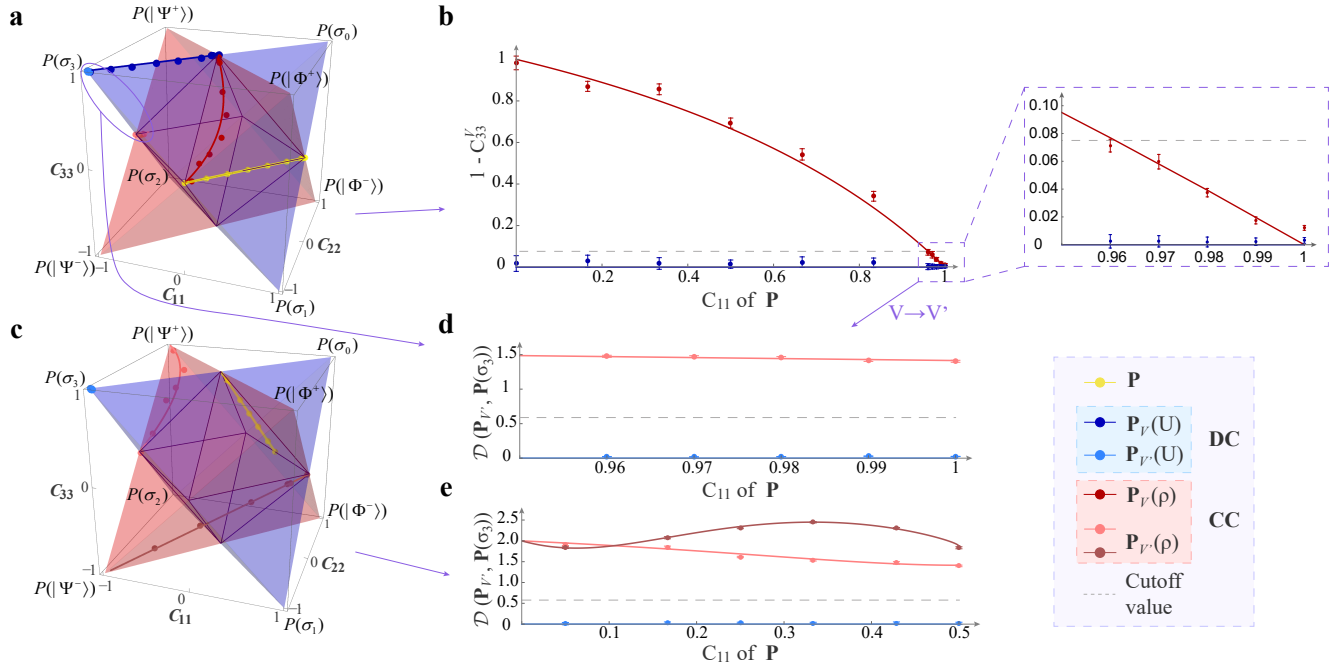


Figure 3: **Experimental results for pure initial states.** **a**: Visualization of $\mathbf{P}_V$ when $\mathbf{P}$ moves on one edge of the causally indistinguishable octahedron. The $\mathbf{P}_{V'}$ are evaluated only when the $\mathbf{P}$ are close to the exceptional plane. **b**: Predicted (line) and measured (data points) DC-criterion correspond to $\mathbf{P}_V$. **c**: Visualization of $\mathbf{P}_{V'}$ when $\mathbf{P}$ moves on the exceptional plane. **d, e**: Predicted and measured distance criterion for DC correspond to the points $\mathbf{P}_{V'}$ in **a** and **c**, respectively. The pink and brown lines show the effect of different phases of Bell states in state preparation.

References

[1] Reichenbach, H. and Reichenbach, M. *The Direction of Time*, University of California Press, 1991.

[2] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch. A quantum advantage for inferring causal structure. Nat. Phys. 11, 414, 2015.

[3] J. F. Fitzsimons, J. A. Jones, and V. Vedral Quantum correlations which imply causation, Sci. Rep. 5, 18281, 2015.

[4] F. Costa and S. Shrapnel Quantum causal modelling, New J. Phys. 18, 063032, 2016

[5] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens Quantum common cause and quantum causal model Phys. Rev. X 7, 031021, 2017

[6] C. Zhang, Y. Hou, and D. Song. Quantum observation scheme universally identifying causalities from correlations Phys. Rev. A 101, 062103, 2020.

[7] M. Hu and Y. Hou. Discrimination between quantum common causes and quantum causality. Phys. Rev. A 97, 062125, 2018.

[8] For more details, please refer to the full version Phys. Rev. A 109, 012406, 2024.

Correlation-Pattern-Based Continuous Variable Entanglement Detection through Neural Networks

Xiaoting Gao[1] *

[1] *State Key Laboratory for Mesoscopic Physics, School of Physics, Frontiers Science Center for Nano-optoelectronics, Peking University, Beijing 100871, China*

Abstract. Entanglement in continuous-variable non-Gaussian states provides irreplaceable advantages in many quantum information tasks. However, the sheer amount of information in such states grows exponentially and makes a full characterization impossible. Here, we develop a neural network that allows us to use correlation patterns to effectively detect continuous-variable entanglement through homodyne detection. Our algorithm works not only on any kind of Gaussian state but also on a whole class of experimentally achievable non-Gaussian states. The findings provide a new approach for experimental detection of continuous-variable quantum correlations without resorting to a full tomography of the state and confirm the exciting potential of neural networks in quantum information processing.

Keywords: Non-Gaussian entanglement, Neural networks, Homodyne detection

The study of quantum entanglement is experiencing a thorough theoretical development and impressive experimental progress, leading to important applications in quantum cryptography, quantum metrology, and quantum computation. It is, therefore, crucial to find reliable and practical methods to detect entanglement. However, once quantum objects are sufficiently large, it becomes practically impossible to decide whether an arbitrary state is entangled or not. In quantum physics, "large" is measured in what we call the dimension of the Hilbert space that describes the object. In our work, we focus on one of the most popular quantum systems out there: light. The different modes of light are mathematically described by infinite-dimensional Hilbert spaces, i.e. the continuous variable (CV) regime, which makes the study of quantum correlations between these modes, in general, impossible. By focusing on specific, yet experimentally relevant classes of states, we managed to circumvent these problems. We leverage the power of artificial neural networks to use measurements of the electric field -known as *homodyne measurements*- to decide whether a state is likely to be entangled or not.

The conventional entanglement criteria which rely on the knowledge of reconstructed density matrix, such as the positive partial transpose (PPT) criterion [1] or the quantum Fisher information (QFI) criterion proposed in Ref. [2], are experimentally infeasible for general non-Gaussian states, which possess complex non-Gaussian Wigner functions. An innovative approach to overcome this issue is provided by deep neural networks, which can work with limited amounts of data from actual measurements. Recently, neural networks have found extensive applications in quantum physics and quantum information science, including detecting quantum features, characterizing quantum states and quantum channels, and solving many-body problems. A key step thus lies in selecting an appropriate training data set to ensure that the networks can effectively and universally learn the features of the quantum system. Keeping our focus on the homodyne measurements, which are the primary means of revealing CV entanglement in experiments, we seek to answer the following question in this paper: Can neural networks be used to detect entanglement for general non-Gaussian states?

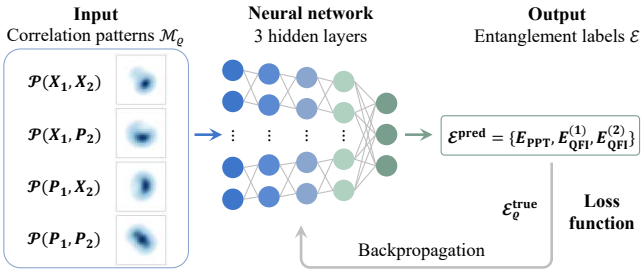


Figure 1: Scheme of the training data processing. The generation of the training data set begins with a series of random density matrices ϱ. Then for each density matrix one generates $24\times24\times4$-dimensional correlation patterns as input data of the neural network. At the output, 3 entanglement labels are computed from ϱ and fed into the neural network for training. The loss function, a binary cross-entropy loss, is evaluated between the true entanglement labels and the predicted labels output from the neural network.

The first step in achieving this goal is to use a numerical method to efficiently simulate a large number of quantum states of light that require only linear optics and a small number of single-photon operations to be produced. While this class of states is very far from arbitrary, it does contain the states that can be produced with state-of-the-art experiments. For our simulated states, we can easily check whether or not they are entangled.

On top of this, we can also reproduce the typical measurement statistics for homodyne measurements. This information is then used to train a machine learning algorithm that can use homodyne measurements as input to decide whether or not the state from which these measurements originate is entangled (Shown in Fig. 1).

*gaoxiaoting@stu.pku.edu.cn

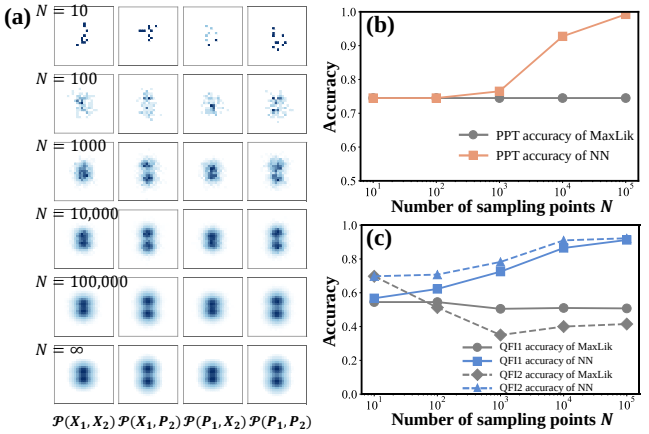


Figure 2: (a) Binned correlation patterns of a state $\hat{\rho} \in \varrho_{\text{test}}$ when the number of Monte Carlo sampling points $N = 10, 100, 1000, 10,000$ and $100,000$. The plot with $N = \infty$ shows the correlation patterns directly discretized from the theoretical joint probability distributions. (b) Accuracy of PPT-type entanglement prediction from the neural network (orange line) and MaxLik algorithm (gray line) against the same value of N in (a). (c) Accuracy of QFI-type entanglement prediction from the neural network (blue lines) and MaxLik (gray lines) algorithm against N. Solid and dashed lines represent the first and second-order QFI, respectively.

After $3,000$ epochs of training the loss function has converged and the network has captured the features mapping the correlation patterns to the entanglement labels, without providing the full density matrices ϱ. This is a crucial element since experiments generally do not have access to the full density matrix of a produced state, and what can be acquired is partial correlation information from measurements.

To test the network with experimental-like data, we simulate the homodyne measurement outcomes via a Monte Carlo sampling method. The test data are obtained from previously unseen quantum states, denoted as ϱ_{test}. For each pattern of the state in ϱ_{test}, we perform N repetitions of sampling to simulate the joint measurement events for each mode, forming a $2 \times N$-dimensional outcomes and used to recover the joint probability distributions. However, directly feeding the raw sampling results into the neural network is infeasible, as the input layer of our trained network requires $24 \times 24 \times 4$-dimensional data. Thus, we also bin each $2 \times N$ sampling points into a 24×24-dimensional matrix. Figure 2(a) shows the discretized correlation patterns with different numbers of sampling points N. The plot with $N = \infty$ is directly obtained from discretizing the theoretical joint probability distributions. As the number of samples N increases from 10 to $100,000$, the Monte Carlo sampling result converges towards the $N = \infty$ case.

Finally, we rely on another type of machine learning tool to visualize what is happening during the training process (Shown in Fig. 3). A clustering algorithm called t-SNE will group quantum states depending on how similar their homodyne measurement statistics are. This way, we can see many clusters of entangled states appearing, which form a useful guideline for future research. Furthermore, this visualization provides a useful sanity check. Quantum states that are too different from anything that the machine learning algorithm was trained on -and thus states for which the algorithm cannot be trusted- will not belong to the previously identified clusters. In such a case, one would have to enrich the set quantum states used for the training.

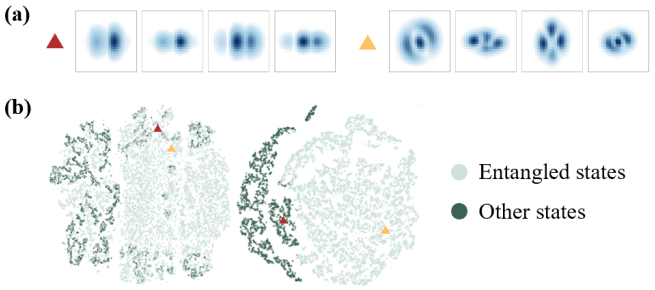


Figure 3: Two-dimensional clusters of two-mode CV states. (a) Examples of $24 \times 24 \times 4$-dimensional correlation patterns for two input states. (b) Left: The 2-dimensional clustering of states before being fed into the network, where t-SNE preserves the pairwise similarities between data points. Right: The same dimension reduction process is conducted on the 64-dimensional array from the last hidden layer of the neural network. Points representing PPT-type entangled are colored in light green, while others are colored in deep green.

This work builds upon the idea that artificial intelligence is a crucial tool to help us recognize patterns in the intricate measurement data that are obtained in quantum optics experiments. In the case of our work, the impressive feature is that the artificial neural network really manages to achieve a goal that, at present, cannot be attained in any other way. This work has been accepted by Physical Review Letters and available on [3].

References

[1] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. **77**, 1413 (1996).

[2] M. Gessner, L. Pezzè, and A. Smerzi, Efficient entanglement criteria for discrete, continuous, and hybrid variables, Phys. Rev. A **94**, 020101 (2016).

[3] X. Gao, M. Isoard, F. Sun, C. E. Lopetegui, Y. Xiang, V. Parigi, Q. He, and M. Walschaers, Correlation-pattern-based continuous-variable entanglement detection through neural networks, arXiv:2310.20570 (2023).

Arbitrary Amplification of Quantum Coherence in Asymptotic and Catalytic Transformation

Naoto Shiraishi[1] * Ryuji Takagi[1] †

[1] *Department of Basic Science, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-0041, Japan*

Abstract. Quantum coherence is one of the fundamental aspects distinguishing classical and quantum theories. Coherence between different energy eigenstates is particularly important, as it serves as a valuable resource under the law of energy conservation. A fundamental question in this setting is how well one can prepare good coherent states from low coherent states and whether a given coherent state is convertible to another one. Here, we show that any low coherent state is convertible to any high coherent state arbitrarily well in two operational settings: asymptotic and catalytic transformations. For a variant of asymptotic coherence manipulation where one aims to prepare desired states in local subsystems, the rate of transformation becomes unbounded regardless of how weak the initial coherence is. In a non-asymptotic transformation with a catalyst, a helper state that locally remains in the original form after the transformation, we show that an arbitrary state can be obtained from any low coherent state. Applying this to the standard asymptotic setting, we find that a catalyst can increase the coherence distillation rate significantly—from zero to infinite rate. We also prove that such anomalous transformation requires small but non-zero coherence in relevant modes, establishing the condition under which a sharp transition of the operational capability occurs. Our results provide a general characterization of the coherence transformability in these operational settings and showcase their peculiar properties compared to other common resource theories such as entanglement and quantum thermodynamics.

Keywords: Quantum coherence, resource theories, quantum thermodynamics, catalyst, information theory

1 Background

In this presentation [1], we investigate the ability of manipulation of quantum systems under the law of energy conservation. In this setting, the quantum coherence among different energy eigenstates is important from both applicational and theoretical perspectives. In applications, quantum coherence is a valuable resource in constructing quantum clocks [2] and performing quantum metrology [3]. Under the law of energy conservation, coherence in the above sense is easily lost due to decoherence, while it is impossible to create and inflate coherence without any help. In this regard, coherence is a precious quantum resource that should be utilized as efficiently as possible. From the theoretical perspective, this problem has been investigated in the light of the *resource theory of asymmetry (unspeakable coherence)* [4, 5], and is also strongly related to quantum thermodynamics [6] since quantum thermodynamics also respects the energy conservation.

To investigate the manipulation power with coherence, two frameworks of state transformations— catalytic and asymptotic transformations—have been actively investigated. In the *catalytic transformation*, one is allowed to borrow the help of an auxiliary system called catalyst—an ancillary system that should return to its own state at the end of the process. In particular, *correlated catalyst*, which could have a correlation with the main system after the transformation, has shown to be effective in enhancing the resource manipulability for several physical settings [7–19]. Another standard setting is *asymptotic transformation*, which concerns the conversion of many copies of an initial quantum state to many copies of another target state [20]. The key performance quantifier of the asymptotic transformation is its transformation rate, which is the ratio of the number of copies of the final state to those of the initial one. To respect the law of energy conservation, we restrict a class of possible operations to the set of covariant operations, which is a standard setting in this research field [4, 5, 21].

Previous studies have revealed that both transformations have severe limitations on the manipulation of coherence. In particular, in the catalytic transformation, the coherence no-broadcasting theorem [22, 23] shows that no coherence could be created even with a correlated catalyst if the input state is exactly incoherent. From these observations, coherence manipulation is considered to be much harder compared to other resource manipulations.

2 Summary of results

Contrarily to these previous implications, in this presentation we demonstrate that we can accomplish arbitrary state conversions in both frameworks, as long as the initial state has (even a tiny amount of) nonzero coherence on relevant energy modes. This result includes a striking case that an almost incoherent initial state is converted into a maximally coherent state, which means that quantum coherence can be arbitrarily amplified in the operational settings we describe below. These protocols solve several open problems on single-shot [24] and asymptotic [18] coherence transformation with correlated catalysts.

On the opposite side, we derive no-go theorems that an exactly zero coherent mode in the initial state should result in zero coherence on this mode in the final state. Combining these two, our results elucidate a sharp threshold in the resource theory of asymmetry, which distinguishes one regime where we can perform arbitrary state conversions and another regime where we can gain no coherence. Our results provide the first general characterization of the feasible manipulation of quantum coherence in asymptotic and catalytic settings, offering potential applications to quantum metrology and quantum thermodynamics.

3 Correlated-catalytic transformation

We consider state transformation with *covariant operations*, the standard set of physically available operations

* shiraishi@phys.c.u-tokyo.ac.jp
† ryujitakagi.pat@gmail.com

implementable by an energy-conserving unitary [5]. We denote by ρ and ρ' the initial and the target states, respectively. In the correlated-catalytic transformation, we employ an auxiliary system C called catalyst that retains the same form after the transformation but helps state conversion in the system S. We say that ρ is convertible to ρ' with a correlated-catalytic transformation if there exists a finite-dimensional catalytic system C with a catalyst state c, and a covariant operation Λ on SC such that $\tau = \Lambda(\rho \otimes c)$ with $\mathrm{Tr}_S[\tau] = c$ and $\mathrm{Tr}_C[\tau] = \rho'$. Our final state may have a correlation between the system and the catalyst, which reflects the name "correlated catalyst". The correlated catalyst is employed in various resource theories [7–19], which usually accompanies insightful and physically robust results. In most studies, the final state is allowed to have an arbitrarily small but finite error. If the final state has no error, we say that this transformation is *exact*. The central question in this setting is whether a given state ρ can be converted into another given state ρ' with a correlated catalyst.

As explained, the coherence no-broadcasting theorem [22, 23] states that a fully incoherent initial state is convertible only to an incoherent state through a covariant operation even with the help of a correlated catalyst. This may give an impression that a correlated catalyst offers little advantage in state convertibility with quantum coherence. However, we show exactly the opposite—correlated catalysts allow enormous operational power, and the only exception is the case with no coherence in the initial state.

Theorem 1 (Fig. 1 (Left, b))**.** *For arbitrary states ρ and ρ', ρ is convertible to ρ' with a correlated catalyst with an arbitrarily small error as long as ρ has nonzero coherence on relevant modes. Moreover, the transformation can be made exact if ρ' is full rank. In addition, the correlation between the system and the catalyst can be made arbitrarily small.*

This theorem claims that an almost incoherent state can be transformed into an almost maximally coherent state with a correlated catalyst, which solves the conjecture in Ref. [24] in the affirmative. Notably, the correlation between the system and the catalyst can be made arbitrarily small, implying that the final state τ is extremely close to a product state of $\rho' \otimes c$.

We remark that the above correlated-catalytic transformation can be *exact* for almost all final states without measure-zero exceptions. This is an advantage of our result compared to other results on correlated-catalytic transformations since most of the previous results apply only to the case that the final state has a small but nonzero error.

As its direct consequence, we also solve in the affirmative an open problem on catalytic asymptotic transformations raised in Ref. [18], asking whether a correlated catalyst improves the rate of (standard) asymptotic transformations. Theorem 1 suggests the solution of the above conjecture in a drastic manner that the transformation rate is improved from zero without a catalyst to infinity with a catalyst.

As for the converse, we provide a no-go result which we call *mode no-broadcasting theorem* prohibiting a correlated-catalytic transformation from incoherent mode to coherent mode. This is an extension of the coherence no-broadcasting theorem [22, 23]—the coherence no-broadcasting theorem applies only when the initial state is completely incoherent (i.e., all modes have no coherence), while our mode no-broadcasting theorem applies to coherent initial states if the coherent modes are relatively irrational to the mode of interest. Intuitively speaking, coherence on relatively irrational modes provides no help to create coherence on a certain mode. Our no-go theorem formalizes this intuition.

Theorem 2 (Fig. 1 (Right))**.** *Consider two states ρ and ρ' such that ρ' has a coherence on a mode and all coherent modes in ρ is relatively irrational to this mode. Then, there is no correlated-catalytic covariant transformation from ρ to ρ'.*

We stress that previously nothing was known about the capability of correlated-catalytic covariant operations other than the coherence no-broadcasting theorem. Theorems 1 and 2 provide the first general characterization of the feasible coherence manipulation with the help of correlated catalysts.

4 Asymptotic transformation

We next consider the asymptotic transformation. In the standard framework of asymptotic transformation, one considers a series $\{\Lambda_n\}_n$ of covariant operations that transforms $\rho^{\otimes n}$ to $\rho'^{\otimes \lfloor rn \rfloor}$ with taking the limit of $n \to \infty$. The key quantity in this framework is the asymptotic transformation rate $R(\rho \to \rho')$, which is the supremum over all achievable rates.

We in particular employ the framework of the *asymptotic marginal transformations* [19, 25], where the reduced state of $\Lambda_n\left(\rho^{\otimes n}\right)$ on every subsystem approaches ρ' with an arbitrarily small error. If the reduced state of any single subsystem exactly coincides with the target state ρ' for some finite n, we say that this asymptotic marginal transformation is *exact*. This setting is especially appealing in the scenario where multiple parties are separated from each other and would like to consume a good coherent state locally. In such a setting, the quality of the resource state is determined by how close the local marginal state is to the desired final state ρ'.

We remark that the above definition is different from the standard definition of asymptotic transformations where we measure the error on the entire system, not on each single subsystem. To distinguish these two, we denote by $\tilde{R}(\rho \to \rho')$ the asymptotic marginal transformation rate and by $R(\rho \to \rho')$ the (standard) asymptotic transformation rate, respectively. The asymptotic marginal transformation rate is larger than or equal to the asymptotic transformation rate; $\tilde{R}(\rho \to \rho') \geq R(\rho \to \rho')$, as the former only focuses on the accuracy of preparing good marginal states. This, for instance, allows for correlation among different subsystems in the final state. Nevertheless, $\tilde{R}(\rho \to \rho')$ is known to come with a fundamental upper bound for a wide class of resource theories [25], and these two rates indeed coincide for entanglement distillation [19]. This suggests that the above relaxation of the asymptotic setting may not realize a significant change in the ability of transformation. On the basis of the above suggestion, together with the limitations on coherence distillation [26] showing that $R(\rho \to \rho') = 0$ for a generic mixed state ρ and a pure coherent state ρ', the marginal asymptotic transformation on coherence seems to have severely limited power of state conversions.

Nevertheless, we prove that any low coherent state can actually be transformed to any high coherent state with an arbitrarily high transformation rate. Namely, there is no restriction on coherence transformation in this operational

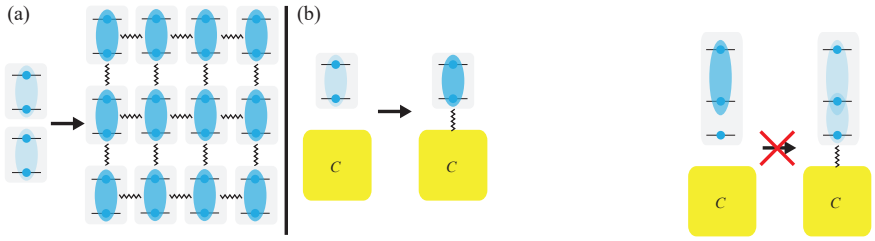


Figure 1: (Left, a): An asymptotic marginal transformation maps n copies of ρ into m copies of ρ' with correlation among copies. Theorem 3 states that for almost all ρ and ρ', the rate of asymptotic marginal transformation of $\rho \to \rho'$ defined as $\lim_{n\to\infty} \max_m \frac{m}{n}$ is unbounded. (Left, b): A correlated-catalytic transformation maps a product state of system S and catalyst C written as $\rho \otimes c$ to τ such that $\mathrm{Tr}_S[\tau] = c$ and $\mathrm{Tr}_C[\tau] = \rho'$. Theorem 1 states that for almost all ρ and ρ', ρ is convertible to ρ' with a correlated catalyst. In addition, the strength of the correlation between the main and catalytic systems can be made arbitrarily small. (Right): Suppose that a mode has no coherence in the initial state. Then even if the initial state has coherence on other modes irrationally related to the mode in interest, a covariant operation with a correlated catalyst cannot provide coherence on this mode. This restriction is stronger than the coherence no-broadcasting theorem [22, 23].

setting, and all states without measure-zero exceptions admit infinite asymptotic marginal distillation rates.

Theorem 3 (Fig. 1 (Left, a))**.** *For arbitrary states ρ and ρ', $\tilde{R}(\rho \to \rho')$ diverges as long as ρ has non-zero coherence on relevant modes. Moreover, the asymptotic marginal transformation can be made exact if ρ' is full rank. In both cases, the correlation between one subsystem and the others can be made arbitrarily small.*

On the other hand, if ρ does not have coherence on relevant modes, even a single copy of ρ' with arbitrarily small errors cannot be prepared from any number of copies of ρ.

We emphasize that we put no requirement on the amount of coherence in the initial state ρ, meaning that a negligibly small but nonzero coherence suffices to obtain a maximally coherent state with an arbitrary rate. The only meaningful distinction lies in whether the state has (maybe extremely small but) non-zero coherence or has exactly zero coherence, and if a system has non-zero coherence, the amount of coherence does not matter in this setting.

It is also remarkable that the diverging rate is obtained for the *exact* transformation. In fact, results in asymptotic resource transformations typically break down when no errors are allowed, and therefore much less is known for exact transformation compared to transformation with an arbitrarily small but finite error. Our theorem presents an exceptionally rare setting in which exact transformations have an enormous ability of state conversions.

5 Proof techniques

We briefly discuss several novel proof techniques we introduce in our work. The full proofs can be found in the technical manuscript. Theorem 1 follows from Theorem 3, and to show Theorem 3 we employ another type of catalytic transformation called *marginal catalytic transformation* as a subroutine in the construction. The marginal catalytic transformation uses multiple initially uncorrelated catalysts. At the end of the protocol, catalysts may have correlations among them, but their reduced state remains invariant. We show that by exploiting the property of marginal catalysts, we can turn a marginal-catalytic transformation into a correlated catalytic one.

To this end, we first turn marginal catalytic transformation into asymptotic marginal transformation, from which Theorem 3 naturally follows. We prove this by utilizing the fact that marginal catalysts are not infinitely reusable but *partially* reusable. We combine this observation with the recent result [24] showing that an arbitrary coherence transformation is possible with marginal catalysts. This allows us to prepare an arbitrary number of copies of the target state in the sense of asymptotic marginal transformation, resulting in the diverging transformation rate. We then turn the asymptotic marginal transformations into the correlated-catalytic transformations. We accomplish this by extending the construction introduced by one of the authors [10] to convert asymptotic *marginal* transformations to correlated-catalytic transformations, resulting in Theorem 1.

The core idea of showing Theorem 2 is to prove that correlated catalysts cannot exploit coherence on a certain mode to create coherence on another mode that is only irrationally related. To prove this, we present the approach using *ladder systems*, allowing one to separately deal with coherence that is rationally related. We then introduce the technique of *complete degeneration*, which virtually changes energy levels so that the modes except the one under study become incoherent. This allows us to reduce the mode broadcasting setting to the one for coherence broadcasting, resulting in Theorem 2. We expect that these techniques will find much use in showing further properties of coherence manipulation and other related problems.

6 Discussion

We showed the anomalous potential of the manipulation of quantum coherence in the asymptotic and catalytic coherence distillation. Our results shed light on the power of correlation in resource manipulation. The importance of correlation has already been discussed intensively in the context of quantum thermodynamics [27–30]. Our results confirm that the unbounded power of coherence transformation is also present in the setting with more operational motivation—asymptotic and correlated-catalytic coherence transformation—lifting quantum coherence as a tangible operational resource.

References

[1] Naoto Shiraishi and Ryuji Takagi. Arbitrary amplification of quantum coherence in asymptotic and catalytic transformation. *Phys. Rev. Lett.*, 132:180202, May 2024.

[2] D. Janzing and T. Beth. Quasi-order of clocks and their synchronism and quantum bounds for copying timing information. *IEEE Trans. Inf. Theory*, 49(1):230–240, 2003.

[3] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, Jan 2006.

[4] Gilad Gour and Robert W Spekkens. The resource theory of quantum reference frames: manipulations and monotones. *New J. Phys.*, 10(3):033023, mar 2008.

[5] Iman Marvian Mashhad. *Symmetry, asymmetry and quantum information*. PhD thesis, 2012.

[6] Matteo Lostaglio, David Jennings, and Terry Rudolph. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. *Nat. Commun.*, 6:6383, March 2015.

[7] Henrik Wilming, Rodrigo Gallego, and Jens Eisert. Axiomatic characterization of the quantum relative entropy and free energy. *Entropy*, 19(6):241, 2017.

[8] Markus P. Müller. Correlating thermal machines and the second law at the nanoscale. *Phys. Rev. X*, 8:041051, Dec 2018.

[9] Paul Boes, Jens Eisert, Rodrigo Gallego, Markus P. Müller, and Henrik Wilming. Von neumann entropy from unitarity. *Phys. Rev. Lett.*, 122:210402, May 2019.

[10] Naoto Shiraishi and Takahiro Sagawa. Quantum thermodynamics of correlated-catalytic state conversion at small scale. *Phys. Rev. Lett.*, 126:150502, Apr 2021.

[11] Seok Hyung Lie and Hyunseok Jeong. Catalytic quantum randomness as a correlational resource. *Phys. Rev. Res.*, 3:043089, Oct 2021.

[12] H. Wilming. Entropy and reversible catalysis. *Phys. Rev. Lett.*, 127:260402, Dec 2021.

[13] Patryk Lipka-Bartosik and Paul Skrzypczyk. Catalytic quantum teleportation. *Phys. Rev. Lett.*, 127:080502, Aug 2021.

[14] Tulja Varun Kondra, Chandan Datta, and Alexander Streltsov. Catalytic transformations of pure entangled states. *Phys. Rev. Lett.*, 127:150503, Oct 2021.

[15] Roberto Rubboli and Marco Tomamichel. Fundamental limits on correlated catalytic state transformations. *Phys. Rev. Lett.*, 129:120506, Sep 2022.

[16] Henrik Wilming. Correlations in typicality and an affirmative solution to the exact catalytic entropy conjecture. *Quantum*, 6:858, November 2022.

[17] Benjamin Yadin, Hyejung H Jee, Carlo Sparaciari, Gerardo Adesso, and Alessio Serafini. Catalytic gaussian thermal operations. *J. Phys. A: Math. Theor.*, 55(32):325301, jul 2022.

[18] Ludovico Lami, Bartosz Regula, and Alexander Streltsov. Catalysis cannot overcome bound entanglement. May 2023.

[19] Ray Ganardi, Tulja Varun Kondra, and Alexander Streltsov. Catalytic and asymptotic equivalence for quantum entanglement. May 2023.

[20] Mark M Wilde. *Quantum information theory*. Cambridge university press, 2013.

[21] Eric Chitambar and Gilad Gour. Quantum resource theories. *Rev. Mod. Phys.*, 91:025001, Apr 2019.

[22] Matteo Lostaglio and Markus P. Müller. Coherence and asymmetry cannot be broadcast. *Phys. Rev. Lett.*, 123:020403, Jul 2019.

[23] Iman Marvian and Robert W. Spekkens. No-broadcasting theorem for quantum asymmetry and coherence and a trade-off relation for approximate broadcasting. *Phys. Rev. Lett.*, 123:020404, Jul 2019.

[24] Ryuji Takagi and Naoto Shiraishi. Correlation in catalysts enables arbitrary manipulation of quantum coherence. *Phys. Rev. Lett.*, 128:240501, Jun 2022.

[25] Giovanni Ferrari, Ludovico Lami, Thomas Theurer, and Martin B. Plenio. Asymptotic State Transformations of Continuous Variable Resources. *Commun. Math. Phys.*, 398(1):291, 2023.

[26] Iman Marvian. Coherence distillation machines are impossible in quantum thermodynamics. *Nat. Commun.*, 11(1):25, January 2020.

[27] Matteo Lostaglio, Markus P. Müller, and Michele Pastena. Stochastic independence as a resource in small-scale thermodynamics. *Phys. Rev. Lett.*, 115:150402, Oct 2015.

[28] Markus P. Müller and Michele Pastena. A generalization of majorization that characterizes shannon entropy. *IEEE Trans. Inf. Theory*, 62(4):1711–1720, 2016.

[29] Facundo Sapienza, Federico Cerisola, and Augusto J. Roncaglia. Correlations as a resource in quantum thermodynamics. *Nat. Commun.*, 10:2492, June 2019.

[30] Seok Hyung Lie and Nelly H. Y. Ng. Catalysis always degrades external quantum correlations. *Phys. Rev. A*, 108:012417, Jul 2023.

Arbitrary Amplification of Quantum Coherence in Asymptotic and Catalytic Transformation

Naoto Shiraishi* and Ryuji Takagi†
Department of Basic Science, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

(Received 9 January 2024; accepted 25 March 2024; published 2 May 2024)

Quantum coherence is one of the fundamental aspects distinguishing classical and quantum theories. Coherence between different energy eigenstates is particularly important, as it serves as a valuable resource under the law of energy conservation. A fundamental question in this setting is how well one can prepare good coherent states from low coherent states and whether a given coherent state is convertible to another one. Here, we show that any low coherent state is convertible to any high coherent state arbitrarily well in two operational settings: asymptotic and catalytic transformations. For a variant of asymptotic coherence manipulation where one aims to prepare desired states in local subsystems, the rate of transformation becomes unbounded regardless of how weak the initial coherence is. In a non-asymptotic transformation with a catalyst, a helper state that locally remains in the original form after the transformation, we show that an arbitrary state can be obtained from any low coherent state. Applying this to the standard asymptotic setting, we find that a catalyst can increase the coherence distillation rate significantly—from zero to infinite rate. We also prove that such anomalous transformation requires small but nonzero coherence in relevant modes, establishing the condition under which a sharp transition of the operational capability occurs. Our results provide a general characterization of the coherence transformability in these operational settings and showcase their peculiar properties compared to other common resource theories such as entanglement and quantum thermodynamics.

DOI: 10.1103/PhysRevLett.132.180202

Quantum coherence between different energy eigenstates is a valuable resource inevitable for quantum clocks [1], metrology [2], and work extraction [3]. Under the law of energy conservation, coherence in the above sense is easily lost due to decoherence, while it is impossible to create and inflate coherence without any help. In this regard, coherence is a precious quantum resource that should be utilized as efficiently as possible.

A central problem concerning quantum coherence as an operational resource is to characterize its manipulability with energy-conserving unitary [4,5]. This physical setting comes with a fundamental constraint that the total amount of quantum coherence cannot be increased by energy-conserving operations. To understand its manipulation power, two formalisms of state transformations—asymptotic and catalytic transformations—have been actively investigated.

One standard setting for resource manipulation is the *asymptotic transformation*, where one aims to convert many copies of the initial quantum state to many copies of another target state [6]. The key performance quantifier of the asymptotic manipulation is its transformation rate, the ratio of the number of copies of the final state to those of the initial one. On the asymptotic coherence manipulation, it has been shown that there is a strong limitation that the transformation rate from generic mixed states to pure coherent states is zero [7].

Another standard setting for resource manipulation is *catalytic transformation*, where one is allowed to borrow the help of another auxiliary system called catalyst—an ancillary system that should return to its own state at the end of the process. In particular, *correlated catalyst*, which could have a correlation with the main system after the transformation, has been shown to be effective in enhancing the resource manipulability for several physical settings [8–20]. However, similarly to the asymptotic transformation, fundamental limitations on catalytic enhancement have been observed. A notable result is the coherence no-broadcasting theorem [21,22], showing that no coherence could be created with a correlated catalyst if the input state is exactly incoherent. These previous studies, both on asymptotic and catalytic transformations, indicate the potential difficulty of manipulating quantum coherence.

Contrarily to these suggestions, we here show that an arbitrary coherence manipulation is enabled in asymptotic and catalytic coherence transformation. We consider a variant of the asymptotic transformation where one aims to prepare a target state on each subsystem [23] and show that the transformation rate becomes unbounded if the initial state has nonzero coherence. In the correlated-catalytic transformation, we prove that arbitrary state transformation becomes possible as long as the initial state has nonzero coherence. This shows that the observation from the coherence no-broadcasting theorem is unstable

about the perturbation of the initial state in the following sense: As long as the initial state contains even a tiny amount of coherence, every coherent state suddenly becomes reachable. In addition, for target states besides measure-zero exceptions, *exact* transformation is possible, which is a much stronger claim than the conventional resource-theoretic results allowing a small error in the final state.

As a direct consequence of our result, we show that the standard asymptotic transformation rate becomes infinite with the help of correlated catalysts. This resolves the open problem proposed in Ref. [11], asking whether correlated catalysts could improve the asymptotic rate at all, in the most drastic manner—catalysts can make undistillable coherent states infinitely distillable.

Our protocols require a nonzero amount of coherence—even if extremely small—in the initial state to implement arbitrary state conversions. To fully characterize this requirement, we formalize no-go theorems on state conversions by introducing the notion of *resonant coherent modes*. These no-go theorems reveal that initial coherence, even if it is negligibly small, is inevitable for arbitrary state conversions, and exactly zero coherence must result in zero coherence. Together with the feasible transformations described above, these characterize state transformability both in asymptotic and catalytic settings, revealing that the distinction between zero and nonzero coherence is an extremely sharp threshold.

We remark that these "amplification" effects do not contradict the physical requirements that the total amount of coherence should not increase. Our results rest on the fact that coherence can *locally* increase, as observed in several settings previously [24,25]. Our results extend these observations in the context of asymptotic and catalytic coherence manipulation and provide general characterizations of the anomalous coherence amplification phenomena observed in each operational setting.

Coherence transformation.—Superposition between energy eigenstates is manifested in time evolution. For a system with Hamiltonian H, a state ρ is called *coherent* if $\mathcal{U}_t(\rho) \neq \rho$ for some time t, where $\mathcal{U}_t(\rho) \coloneqq e^{-iHt}\rho e^{iHt}$ is the unitary time evolution. A state is called *incoherent* if it is not coherent. We remark that the coherence we consider in this work is what is so-called *unspeakable coherence* [26]. (Not to be confused with another type known as *speakable coherence* [27].)

Available operations in manipulating quantum coherence should not create coherence from incoherent states, as respecting the law of energy conservation. In reflecting this restriction, a natural set of available operations for the coherence manipulation is the *covariant operations* with time translation [28], i.e., the action of a channel $\Lambda\colon S \to S'$, where S and S' are input and output systems, commutes with the unitary time evolution as $\Lambda \circ \mathcal{U}_t^S = \mathcal{U}_t^{S'} \circ \Lambda$ for all t [4,5,29]. From the operational perspective, any covariant operation Λ can be equivalently written by an energy-conserving unitary U and an incoherent state η as $\Lambda(\rho) = \mathrm{Tr}_A[U(\rho \otimes \eta)U^\dagger]$, where A is some auxiliary system [5,30]. In other words, covariant operations are operations which can be implemented by an energy-conserving unitary with incoherent states.

Coherent modes.—Our findings clarify that whether relevant modes have (maybe tiny but) nonzero coherence leads to a drastic change. To formalize this, we introduce the notion of resonant coherent modes. A mode for Δ is a pair of two energy levels with energy difference Δ, and a state ρ has a coherent mode Δ if $\rho_{ij} \neq 0$ with $E_i - E_j = \Delta$ is satisfied for some i, j, where $\rho_{ij} \coloneqq \langle i|\rho|j\rangle$ and $|i\rangle$ is an energy eigenstate with energy E_i for the given Hamiltonian H. We then define the set $\mathcal{C}(\rho)$ of resonant coherent modes of state ρ as all linear combinations of nonzero coherent modes with integer coefficients, i.e.,

$$\mathcal{C}(\rho) \coloneqq \left\{ x | x = \sum_{i,j(\rho_{ij}\neq 0)} n_{ij}\Delta_{ij}, n_{ij} \in \mathbb{Z} \right\} \tag{1}$$

for an energy interval $\Delta_{ij} = E_i - E_j$ and a nonzero off-diagonal entry ρ_{ij} of a density matrix ρ for energies E_i and E_j. Notably, in the asymptotic and catalytic coherence manipulation, one can create a coherence on mode $\Delta = \Delta_1 + \Delta_2$ if the initial state has coherence on modes Δ_1 and Δ_2 [31].

Asymptotic manipulation.—We first consider the asymptotic manipulation. Suppose ρ is an initial state and ρ' is a target state. In the standard framework of asymptotic transformation, one considers a series $\{\Lambda_n\}_n$ of available operations that transforms $\rho^{\otimes n}$ to $\rho'^{\otimes \lfloor rn \rfloor}$ with vanishing error at the limit of $n \to \infty$. The asymptotic transformation rate $R(\rho \to \rho')$ is the supremum over all achievable rates r. When ρ' is a pure state ϕ, it is particularly called asymptotic distillation. For coherence distillation with covariant operations, the distillation rate $R(\rho \to \phi)$ is known to be zero for an arbitrary full-rank state ρ and an arbitrary coherent pure state ϕ [7], which puts a fundamental restriction on the tangibility of coherence as an operational resource.

However, the necessity of obtaining the state close to $\phi^{\otimes \lfloor rn \rfloor}$ can be reasonably relaxed for many operational settings. For instance, consider the scenario where multiple parties are separated from each other and would like to consume a good coherent state locally. In such a setting, the quality of the resource state is determined by how close the local marginal state is to the maximally coherent state. The framework that fits this operational setting was considered previously and called *asymptotic marginal transformation* [9,23]. Suppose ρ and ρ' are the states on the systems S and S' respectively. The state ρ can be converted to ρ' with an asymptotic marginal transformation with rate r if there exists a series of available operations $\{\Lambda_n\}_n$ from $S^{\otimes n}$ to $S'^{\otimes \lfloor rn \rfloor}$ such that the reduced state of

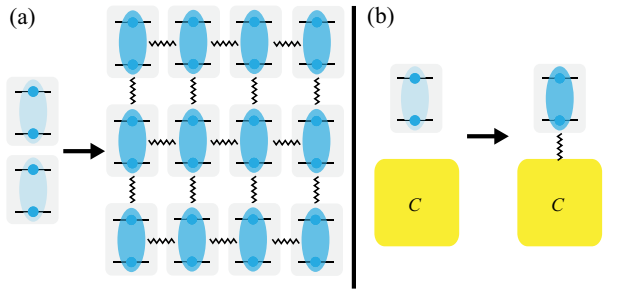


FIG. 1. (a): An asymptotic marginal transformation maps n copies of ρ into m copies of ρ' with correlation among copies. Theorem 1 states that for almost all ρ and ρ', the rate of asymptotic marginal transformation of $\rho \to \rho'$ defined as $\lim_{n\to\infty} \max_m (m/n)$ is unbounded. (b): A correlated-catalytic transformation maps a product state of system S and catalyst C written as $\rho \otimes c$ to τ such that $\mathrm{Tr}_S[\tau] = c$ and $\mathrm{Tr}_C[\tau] = \rho'$. Theorem 2 states that for almost all ρ and ρ', ρ is convertible to ρ' with a correlated catalyst. In addition, the strength of the correlation between the main and catalytic systems can be made arbitrarily small.

$\Lambda_n(\rho^{\otimes n})$ on every subsystem approaches ρ' with a vanishing error at the $n \to \infty$ limit. If the reduced state of any single subsystem exactly coincides with the target state ρ' for some finite n, we say that this asymptotic marginal transformation is *exact*. Although the asymptotic marginal transformation rate $\tilde{R}(\rho \to \rho')$, which is defined as the highest achievable rate in the marginal asymptotic conversion, serves as an upper bound of the standard asymptotic transformation $\tilde{R}(\rho \to \rho') \geq R(\rho \to \rho')$, these two rates coincide in many settings such as entanglement, quantum thermodynamics, and nonclassicality [23], suggesting that this relaxation may not realize a significant improvement in the ability of transformation (see also Sec. II A in the Supplemental Material [32]).

Despite these previous observations, we prove that any low coherent state can be transformed to any high coherent state with an arbitrarily high transformation rate. Namely, there is no restriction on coherence transformation, and all states without measure-zero exceptions admit infinite asymptotic marginal distillation rates [Fig. 1(a)].

Theorem 1.—For arbitrary states ρ and ρ', $\tilde{R}(\rho \to \rho')$ diverges if ρ has nonzero coherence in the sense of $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$. Moreover, the asymptotic marginal transformation can be made exact if ρ' is full rank. In both cases, the correlation between one subsystem and the others can be made arbitrarily small. On the other hand, if $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$, even a single copy of ρ' cannot be prepared from any number of copies of ρ with arbitrarily small error.

We remark that $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$ is quite a mild condition since any state ρ with extremely small but nonzero coherence on all modes automatically passes this requirement regardless of ρ'.

Theorem 1 provides the complete characterization of the general asymptotic marginal coherence transformation, including the case of distillation when ρ' is pure. Intuitively speaking, if the initial state contains nonzero coherence on modes that are coherent in the target state, then an arbitrary rate can be realized. The condition $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$, whether the state has (maybe extremely small but) nonzero coherence or has exactly zero coherence, serves as an extremely sharp and the only threshold separating infinite and zero asymptotic transformation rates.

The diverging rate for the exact transformation shown in Theorem 1 is also remarkable. In fact, asymptotic resource transformation typically comes with a severe restriction when no errors are allowed, and therefore much less is known for exact transformation compared to transformation with a vanishing error. Our result presents a rare scenario in which exact transformation realizes an outstanding performance that coincides with the performance of nonzero error transformation.

Correlated-catalytic transformation.—We now consider the correlated-catalytic transformation, where we employ an auxiliary system C called catalyst which does not change its own state between the initial and the final state but helps state conversion in system S. We say that ρ is convertible to ρ' through correlated-catalytic transformation if there exists a finite-dimensional catalytic system C with a catalyst state c, and a covariant operation Λ on SC such that $\tau = \Lambda(\rho \otimes c)$ with $\mathrm{Tr}_S[\tau] = c$ and $\mathrm{Tr}_C[\tau] = \rho'$. Our final state may have a correlation between the system and the catalyst, which reflects the name "correlated catalyst."

We investigate covariant operations with a correlated catalyst. Recent studies have revealed a severe limitation for correlated-catalytic covariant operations, called the coherence no-broadcasting theorem [21,22]. This theorem states that a fully incoherent initial state is convertible only to an incoherent state through a covariant operation even with the help of a correlated catalyst. This may suggest that a correlated catalyst offers little advantage in state convertibility with quantum coherence. However, we show exactly the opposite—correlated catalysts allow enormous operational power to most covariant state conversions, and the only exception is the case with no coherence in the initial state.

Theorem 2.—For arbitrary states ρ and ρ', ρ is convertible to ρ' with a correlated catalyst with an arbitrarily small error if $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$, and the transformation can be made exact if ρ' is full rank. In addition, the correlation between the system and catalyst can be made arbitrarily small.

This shows that a correlated catalyst enables an arbitrary coherence amplification—an almost incoherent state can be transformed to an almost maximally coherent state with a correlated catalyst [Fig. 1(b)], solving the conjecture in Ref. [31] in the affirmative. Similarly to the case of asymptotic transformation, the only meaningful distinction lies in whether the state has nonzero coherent modes or not.

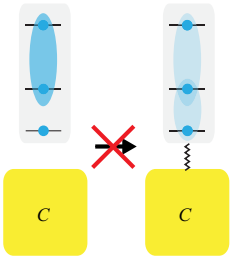


FIG. 2. Suppose that a mode has no coherence in the initial state. Then even if the initial state has coherence on other modes irrationally related to the mode in interest, a covariant operation with a correlated catalyst cannot provide coherence on this mode. This restriction is stronger than the coherence no-broadcasting theorem [21,22].

Notably, the correlation between the system and the catalyst can be made arbitrarily small, implying that the final state τ is extremely close to a product state of $\rho' \otimes c$.

By choosing the initial state as $\rho^{\otimes n}$ and the target state as $\phi^{\otimes rn}$ for a coherent state ρ and a pure coherent state ϕ, there exists a correlated catalyst that enables the transformation from $\rho^{\otimes n}$ to $\phi^{\otimes rn}$ with an arbitrarily small error for every n and r. This setup corresponds to the standard (not marginal) asymptotic distillation, i.e., the error in the final state is measured for the entire state $\phi^{\otimes rn}$, assisted by correlated catalysts. Noting that the standard asymptotic distillation rate $R(\rho \to \phi)$ without a catalyst is zero for every full-rank state ρ [7], our result gives the first example for which the catalyst improves the asymptotic transformation rate, resolving the open problem raised in Ref. [11].

As for the converse, we expect that $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$ also gives the necessary condition for the transformation to exist. Here, we give a partial result toward the full solution to this problem. As naturally guessed, coherence in the initial state would not be helpful in creating coherence on the mode that is only irrationally related to the resonant coherent modes. To formalize this, we introduce $\mathcal{C}'$, which is an extension of $\mathcal{C}$ to rational coefficients: $\mathcal{C}'(\rho) := \{x | x = \sum_{i,j(\rho_{ij}\neq 0)} n_{ij}\Delta_{ij}, n_{ij} \in \mathbb{Q}\}$. We then obtain the necessary condition for the approximate correlated-catalytic covariant transformation (Fig. 2).

Theorem 3.—For two states ρ and ρ' such that $\mathcal{C}'(\rho') \not\subseteq \mathcal{C}'(\rho)$, there does not exist a correlated-catalytic covariant transformation from ρ to ρ'.

This result can be understood as *mode no-broadcasting*—new coherent modes cannot be created by a covariant operation with a correlated catalyst. This contains the coherence no-broadcasting theorem as a special case with $\mathcal{C}'(\rho) = \{0\}$ and extends it to the case of coherent initial states. We conjecture that the above condition is strengthened to $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$, which would provide the exact characterization of the feasible coherence transformation with a correlated catalyst together with Theorem 2.

Proof sketch.—Here, we provide a proof sketch of our main results. The complete proofs are presented in the Supplemental Material [32].

We first outline the proof of the achievable part of Theorem 1. Our protocol employs another operational framework known as marginal-catalytic transformations [31] (see also Ref. [24]). In particular, it was shown that for any full-rank state ρ' there exists a set $C_1, \ldots, C_N$ of catalytic systems with states $c_1, \ldots, c_N$ and a covariant operation Λ: $S \otimes C_1 \otimes \cdots \otimes C_N \to S' \otimes C_1 \otimes \cdots \otimes C_N$ such that $\Lambda(c_1 \otimes \cdots \otimes c_N) = \tau$ with $\mathrm{Tr}_{C_1,\ldots,C_N}[\tau] = \rho'$ and $\mathrm{Tr}_{\setminus C_i}[\tau] = c_i$ for all $i = 1, \ldots, N$. Furthermore, these catalysts are partially reusable: If we have N^k sets of catalysts $c_1 \otimes \cdots \otimes c_N$, an appropriate recombination of them allows one to prepare $(k+1)N^k$ copies of ρ' with these marginal catalysts.

We now construct our protocol, which is inspired by Ref. [9]. We first show that a set $c_1 \otimes \cdots \otimes c_N$ of catalysts can be prepared exactly from $\rho^{\otimes\mu}$ by a covariant operation for some integer μ. This allows us to transform μN^k copies of ρ into N^k sets of catalysts. Using these catalysts, we obtain $(k+1)N^k$ copies of ρ' by a marginal catalytic covariant transformation, after which we discard the catalytic systems. The transformation rate is $(k+1)/\mu$, which can be made arbitrarily large by setting sufficiently large k. This transformation can be made exact for a full-rank target state ρ' by employing the result in Ref. [18].

The converse part of Theorem 1 can be shown by utilizing the properties of the modes of asymmetry [83].

Theorem 2 can be obtained by applying the well-known technique to derive correlated-catalytic convertibility from asymptotic convertibility with vanishing error. This type of result was first shown in Ref. [16] in the context of quantum thermodynamics (cf. Refs. [84,85] for exact asymptotic transformation), and a general form of statement is explicitly shown and proven in Ref. [31]. In particular, this construction was recently used to convert asymptotic marginal transformation to correlated-catalytic transformation [18].

We finally outline the proof of Theorem 3. We suppose contrarily that the final state ρ' has coherence on a mode $\Delta E \notin \mathcal{C}'(\rho)$ and derive contradiction with the coherence no-broadcasting theorem [21,22]. Let $L(\Delta)$ be an infinite-dimensional system whose energy levels form a ladder with energy interval Δ. We embed the main system S and catalytic system C into a product of ladder systems $L(\Delta_0) \otimes L(\Delta_1) \otimes L(\Delta_2) \otimes \cdots$ such that Δ_0 is an integer multiple of ΔE ($\Delta E = m\Delta_0$ for some integer m), and the set $\{\Delta_0, \Delta_1, \Delta_2, \ldots\}$ are rational-linearly independent. By assumption, ρ' has coherence in $L(\Delta_0)$, for which ρ is incoherent. For brevity, we abbreviate the set $\Delta_1, \Delta_2, \ldots$ as $\tilde{\boldsymbol{\Delta}}$.

Our key observation is that since a covariant operation acts on each rational-linearly independent mode separately, if $\rho \otimes c$ on $L(\Delta_0) \otimes L(\tilde{\boldsymbol{\Delta}})$ can be transformed to τ by a covariant operation, the same transformation is also possible

on systems with another arbitrary set $\tilde{\boldsymbol{\Delta}}' = (\Delta_1', \Delta_2', \ldots)$. Namely, $\rho \otimes c \to \tau$ on $L(\Delta_0) \otimes L(\tilde{\boldsymbol{\Delta}})$ (the same density matrix on ladders with different energy spacings) is possible by a covariant operation. Setting $\tilde{\boldsymbol{\Delta}}' = \mathbf{0}$ in the above modification, where all states outside $L(\Delta_0)$ are degenerate, we find that ρ on $L(\Delta_0) \otimes L(\mathbf{0})$ is completely incoherent. On the other hand, the final state ρ' has coherence in $L(\Delta_0)$, which contradicts the coherence no-broadcasting theorem.

Discussion.—We showed the anomalous potential of the manipulation of quantum coherence in the asymptotic and catalytic coherence distillation. These results are highly special to quantum coherence that cannot be seen in other resource theories such as entanglement [86,87], quantum thermodynamics [19], and speakable coherence [26,27,88] (see Sec. V in the Supplemental Material [32]). Related to this, we stress that our result is different from the well-known embezzlement phenomena observed in several resource theories [89,90], admitting arbitrary state conversions by allowing a small error in a catalyst. Our framework allows no errors in the catalyst, and thus the operational capability comes from an entirely different mechanism.

Our results shed light on the power of correlation in resource manipulation. In fact, without correlation, amplification of coherence is impossible in both asymptotic and catalytic settings. The importance of correlation has already been discussed intensively in the context of quantum thermodynamics [91–94]. Quantum thermodynamics with an uncorrelated catalyst has many restrictions with Rényi entropies in state convertibility [84,89,95,96], while most of the restrictions are lifted by proper use of correlations, and only the second law of thermodynamics with the relative entropy remains [14,16]. For the coherence transformation, previous studies [24,31] showed an astonishing operational power enabled by correlations between multiple catalysts. Our results confirm that the unbounded power of coherence transformation is also present in the setting with much more operational motivation—asymptotic and correlated-catalytic coherence transformation—lifting quantum coherence as an even more tangible operational resource.

Note added.—During the completion of our manuscript, we became aware of an independent related work by Kondra *et al.* [97], which was concurrently posted to arXiv with ours. Also, an anonymous referee of the QIP conference notified us that when ρ' is pure and the period (the minimum time after which the state returns to the original one) for ρ and ρ' coincide, one can also obtain the diverging asymptotic marginal transformation rate (with an arbitrary small error) by generalizing the construction for sublinear coherence distillation in Ref. [7] [Supplementary Note 7] to the case of marginal asymptotic conversion. This approach, which is different from ours, in fact admits a larger target state, up to the size sublinear in the number of copies of ρ. Our Theorem 1, on the other hand, applies to the fully general setting and contains further insights into the possibility of exact transformation and fundamental limitations imposed by the resonant coherence modes. We thank the referee for their insightful comments.

We thank Eunwoo Lee for discussions on group representations, and Kohdai Kuroiwa for the asymptotic continuity. N. S. was supported by JSPS Grants-in-Aid for Scientific Research Grant No. JP19K14615. R. T. is supported by JSPS KAKENHI Grant No. JP23K19028.

*shiraishi@phys.c.u-tokyo.ac.jp
†ryujitakagi.pat@gmail.com

[1] D. Janzing and T. Beth, IEEE Trans. Inf. Theory **49**, 230 (2003).
[2] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).
[3] M. Lostaglio, D. Jennings, and T. Rudolph, Nat. Commun. **6**, 6383 (2015).
[4] G. Gour and R. W. Spekkens, New J. Phys. **10**, 033023 (2008).
[5] I. Marvian Mashhad, Symmetry, asymmetry and quantum information, Ph.D. thesis, 2012.
[6] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, England, 2013).
[7] I. Marvian, Nat. Commun. **11**, 25 (2020).
[8] P. Boes, J. Eisert, R. Gallego, M. P. Müller, and H. Wilming, Phys. Rev. Lett. **122**, 210402 (2019).
[9] R. Ganardi, T. Varun Kondra, and A. Streltsov, arXiv:2305.03488.
[10] T. V. Kondra, C. Datta, and A. Streltsov, Phys. Rev. Lett. **127**, 150503 (2021).
[11] Lami, L., B. Regula, and A. Streltsov, arXiv:2305.03489.
[12] S. H. Lie and H. Jeong, Phys. Rev. Res. **3**, 043089 (2021).
[13] P. Lipka-Bartosik and P. Skrzypczyk, Phys. Rev. Lett. **127**, 080502 (2021).
[14] M. P. Müller, Phys. Rev. X **8**, 041051 (2018).
[15] R. Rubboli and M. Tomamichel, Phys. Rev. Lett. **129**, 120506 (2022).
[16] N. Shiraishi and T. Sagawa, Phys. Rev. Lett. **126**, 150502 (2021).
[17] H. Wilming, Phys. Rev. Lett. **127**, 260402 (2021).
[18] H. Wilming, Quantum **6**, 858 (2022).
[19] H. Wilming, R. Gallego, and J. Eisert, Entropy **19**, 241 (2017).
[20] B. Yadin, H. H. Jee, C. Sparaciari, G. Adesso, and A. Serafini, J. Phys. A **55**, 325301 (2022).
[21] M. Lostaglio and M. P. Müller, Phys. Rev. Lett. **123**, 020403 (2019).
[22] I. Marvian and R. W. Spekkens, Phys. Rev. Lett. **123**, 020404 (2019).
[23] G. Ferrari, L. Lami, T. Theurer, and M. B. Plenio, Commun. Math. Phys. **398**, 291 (2023).
[24] F. Ding, X. Hu, and H. Fan, Phys. Rev. A **103**, 022403 (2021).
[25] G. Manzano, R. Silva, and J. M. R. Parrondo, Phys. Rev. E **99**, 042135 (2019).

[26] I. Marvian and R. W. Spekkens, Phys. Rev. A **94**, 052324 (2016).
[27] T. Baumgratz, M. Cramer, and M. B. Plenio, Phys. Rev. Lett. **113**, 140401 (2014).
[28] In the context of resource theories, covariant operations correspond to the class of completely resource nongenerating operations (see also Proposition S.4 in the Supplemental Material), which constitutes the standard set of free operations considered for the resource theory of unspeakable coherence.
[29] E. Chitambar and G. Gour, Rev. Mod. Phys. **91**, 025001 (2019).
[30] M. Keyl and R. F. Werner, J. Math. Phys. (N.Y.) **40**, 3283 (1999).
[31] R. Takagi and N. Shiraishi, Phys. Rev. Lett. **128**, 240501 (2022).
[32] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.180202 for detailed proofs and discussions of our main results, which includes Refs. [33–82].
[33] F. Albarelli, M. G. Genoni, M. G. A. Paris, and A. Ferraro, Phys. Rev. A **98**, 052350 (2018).
[34] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).
[35] A. Anshu, M.-H. Hsieh, and R. Jain, Phys. Rev. Lett. **121**, 190504 (2018).
[36] K. Audenaert, M. B. Plenio, and J. Eisert, Phys. Rev. Lett. **90**, 027901 (2003).
[37] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Phys. Rev. Lett. **111**, 250404 (2013).
[38] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).
[39] K. Bu, U. Singh, and J. Wu, Phys. Rev. A **93**, 042326 (2016).
[40] E. T Campbell, Phys. Rev. A **83**, 032317 (2011).
[41] Coladangelo, A., and D. Leung, arXiv:1910.11354.
[42] K. Fang and Z.-W. Liu, Phys. Rev. Lett. **125**, 060405 (2020).
[43] K. Fang and Z.-W. Liu, PRX Quantum **3**, 010337 (2022).
[44] M. G. Genoni and M. G. A. Paris, Phys. Rev. A **82**, 052341 (2010).
[45] T. Gonda and R. W. Spekkens, Compositionality **5**, 7 (2023).
[46] G. Gour, Phys. Rev. A **95**, 062314 (2017).
[47] G. Gour and C. M. Scandolo, arXiv:2101.01552.
[48] G. Gour and A. Winter, Phys. Rev. Lett. **123**, 150401 (2019).
[49] F. Hansen, Proc. Natl. Acad. Sci. U.S.A. **105**, 9909 (2008).
[50] A. Hickey and G. Gour, J. Phys. A **51**, 414009 (2018).
[51] M. Horodecki and J. Oppenheim, Nat. Commun. **4**, 2059 (2013).
[52] M. Horodecki and J. Oppenheim, Int. J. Mod. Phys. B **27**, 1345019 (2013).
[53] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[54] M. Howard and E. Campbell, Phys. Rev. Lett. **118**, 090501 (2017).
[55] D. Jonathan and M. B. Plenio, Phys. Rev. Lett. **83**, 3566 (1999).
[56] K. Kuroiwa and H. Yamasaki, Quantum **4**, 355 (2020).
[57] H. Kwon, K. C. Tan, T. Volkoff, and H. Jeong, Phys. Rev. Lett. **122**, 040503 (2019).
[58] Y. Liu and X. Yuan, Phys. Rev. Res. **2**, 012035 (2020).
[59] Z.-W. Liu, K. Bu, and R. Takagi, Phys. Rev. Lett. **123**, 020401 (2019).
[60] Z.-W. Liu and A. Winter, arXiv:1904.04201.
[61] I. Marvian, Phys. Rev. Lett. **129**, 190502 (2022).
[62] I. Marvian and R. W. Spekkens, New J. Phys. **15**, 033001 (2013).
[63] I. Marvian and R. W. Spekkens, Nat. Commun. **5**, 3821 (2014).
[64] B. Regula, J. Phys. A **51**, 045303 (2018).
[65] B. Regula, Phys. Rev. Lett. **128**, 110505 (2022).
[66] B. Regula, Quantum **6**, 817 (2022).
[67] B. Regula, K. Bu, R. Takagi, and Z.-W. Liu, Phys. Rev. A **101**, 062315 (2020).
[68] B. Regula, K. Fang, X. Wang, and M. Gu, New J. Phys. **21**, 103017 (2019).
[69] B. Regula and L. Lami, arXiv:2211.15678.
[70] B. Regula, L. Lami, G. Ferrari, and R. Takagi, Phys. Rev. Lett. **126**, 110403 (2021).
[71] B. Regula and R. Takagi, Nat. Commun. **12**, 4411 (2021).
[72] B. Regula and R. Takagi, Phys. Rev. Lett. **127**, 060402 (2021).
[73] B. Synak-Radtke and M. Horodecki, J. Phys. A **39**, L423 (2006).
[74] R. Takagi, Sci. Rep. **9**, 14562 (2019).
[75] R. Takagi and B. Regula, Phys. Rev. X **9**, 031053 (2019).
[76] R. Takagi, B. Regula, K. Bu, Z.-W. Liu, and G. Adesso, Phys. Rev. Lett. **122**, 140402 (2019).
[77] R. Takagi, B. Regula, and M. M. Wilde, PRX Quantum **3**, 010348 (2022).
[78] R. Takagi and Q. Zhuang, Phys. Rev. A **97**, 062337 (2018).
[79] R. Uola, T. Kraft, J. Shang, X.-D. Yu, and O. Gühne, Phys. Rev. Lett. **122**, 130404 (2019).
[80] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, New J. Phys. **16**, 013009 (2014).
[81] B. Yadin, F. C. Binder, J. Thompson, V. Narasimhachar, M. Gu, and M. S. Kim, Phys. Rev. X **8**, 041038 (2018).
[82] C. Zhang, B. Yadin, Z.-B. Hou, H. Cao, B.-H. Liu, Y.-F. Huang, R. Maity, V. Vedral, C.-F. Li, G.-C. Guo, and D. Girolami, Phys. Rev. A **96**, 042327 (2017).
[83] I. Marvian and R. W. Spekkens, Phys. Rev. A **90**, 062110 (2014).
[84] G. Aubrun and I. Nechita, Commun. Math. Phys. **278**, 133 (2008).
[85] R. Duan, Y. Feng, X. Li, and M. Ying, Phys. Rev. A **71**, 042319 (2005).
[86] R. Alicki and M. Fannes, J. Phys. A **37**, L55 (2004).
[87] M. Christandl and A. Winter, J. Math. Phys. (N.Y.) **45**, 829 (2004).
[88] Z. Xi, Y. Li, and H. Fan, Sci. Rep. **5**, 10922 (2015).
[89] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, Proc. Natl. Acad. Sci. U.S.A. **112**, 3275 (2015).
[90] W. van Dam and P. Hayden, Phys. Rev. A **67**, 060302 (2003).
[91] S. H. Lie and N. H. Y. Ng, Phys. Rev. A **108**, 012417 (2023).
[92] M. Lostaglio, M. P. Müller, and M. Pastena, Phys. Rev. Lett. **115**, 150402 (2015).

[93] M. P. Müller and M. Pastena, IEEE Trans. Inf. Theory **62**, 1711 (2016).
[94] F. Sapienza, F. Cerisola, and A. J. Roncaglia, Nat. Commun. **10**, 2492 (2019).
[95] M. Klimesh, arXiv:0709.3680.
[96] S. Turgut, J. Phys. A **40**, 12185 (2007).
[97] T.Varun Kondra, R. Ganardi, and A. Streltsov, arXiv:2308.12814.
[98] J. I. Cirac, A. K. Ekert, and C. Macchiavello, Phys. Rev. Lett. **82**, 4344 (1999).
[99] X. Wang and M. M. Wilde, Phys. Rev. Lett. **125**, 040502 (2020).

Supplemental Material
"Arbitrary Amplification of Quantum Coherence in Asymptotic and Catalytic Transformation"

Naoto Shiraishi and Ryuji Takagi
Department of Basic Science, The University of Tokyo

Contents

Appendix I: General setups

I.A Preliminaries

1. *Quantum resource theories*

The coherence transformation discussed in this work can be understood in a broader context of quantum resource theories [1], which is a general framework that describes the quantification and feasible manipulation of quantities that are considered "precious" under given physical settings. The physical restriction can be formalized by specifying a set $\mathbb{F}$ of *free states* and a set $\mathbb{O}$ of *free operations* that are assumed to be freely accessible in the setting of interest. The requirement imposed on free operations is that they should not be able to create resourceful (i.e., non-free) states from free states, i.e., every free operation $\Lambda \in \mathbb{O}$ satisfies $\Lambda(\sigma) \in \mathbb{F}$ for all $\sigma \in \mathbb{F}$, justifying the notion of "free" operations.

These concepts not only allow us to characterize the resourceful states but also motivate us to *quantify* the amount of resourcefulness contained in a given resourceful state $\rho \notin \mathbb{F}$. The resource quantification can be formalized by *resource measures*, also known as *resource monotones*, which are

functions from quantum states to real numbers that return the "amount" of resourcefulness. For a function R to be a valid resource measure, it is required that (1) it takes the minimum value for free states: $R(\sigma) = c$ for all $\sigma \in \mathbb{F}$ with some constant c and $R(\rho) \geq c$ for every state ρ, (2) it does not increase under free operations (monotonicity): $R(\rho) \geq R(\Lambda(\rho))$ is satisfied for all free operations $\Lambda \in \mathbb{O}$ and all states ρ.

The fundamental problem in any operational setting is to characterize the feasible state transformation by a quantum process accessible in a given setting. The resource theory framework allows us to formulate this question as follows: for given states ρ and ρ', does there exist a free operation $\Lambda \in \mathbb{O}$ such that $\Lambda(\rho) = \rho'$? We call this type of state transformation enabled by a free operation *free transformation*. The problem of feasible state transformation is closely related to the notion of resource quantification since the monotonicity of a resource measure R ensures $R(\rho) \geq R(\Lambda(\rho))$. Therefore, $R(\rho) \geq R(\rho')$ serves as a necessary condition for the free transformation from ρ to ρ' to be possible. This also reflects the intuition that "free operations should not increase the amount of precious resources". On the other hand, proving the sufficient part requires one to construct a specific free operation Λ that actually transforms ρ to ρ'. The ultimate goal then is to obtain the necessary and sufficient conditions for feasible free transformation for arbitrary given states ρ and ρ'.

The performance of free transformation can be studied in various settings, each of which comes with an operational motivation. In this work, we mainly focus on two standard settings known as *asymptotic* and *catalytic* transformation, which we discuss in Sec. II and Sec. III respectively.

The characterization of feasible free transformation admits not only advances in the fundamental understanding of operational aspects of quantum mechanics but also practical consequences. Quantum information processing protocols typically employ a pure state prepared in the standard form as its "fuel". For instance, quantum teleportation utilizes the maximally entangled state (called "e-bit") to realize the optimal performance. However, available quantum states are typically not provided in the desired form, mainly due to inevitable noise. This requires one to prepare such standard pure states from given noisy states. This procedure is called *resource distillation* and serves as a key routine in numerous settings, such as entanglement manipulation [2], fault-tolerant quantum computation [3], and work extraction [4, 5]. The generality of resource theories allows us to study the performance of the distillation process as a problem of free transformation. Indeed, distillation corresponds to a specific setting where we choose ρ' as the desired pure state ϕ.

The resource-theoretic framework can be applied to numerous physical settings to describe key quantities such as entanglement transformation under local operations and classical communication [6], quantum non-classicality [7, 8] and non-Gaussianity [9–11] in continuous-variable systems, quantum thermodynamics [4, 5, 12], and non-Cliffordness (magicness) in fault-tolerant quantum computation [13, 14]. In recent years, the unified understanding of the operational characterization of the individual quantum resources has been developed under the framework of *general resource theories*, which keeps the generality of the choice of free objects and seeks the properties shared by all such quantum resources in the context of, e.g., advantages in discrimination tasks [15–18], resource quantification [15–24], and resource distillation and dilution [25–38]. The main focus of this work is a specific instance of resource theory, which we introduce in the next subsection. Nevertheless, having a background understanding and context in general quantum resource theories is helpful in interpreting and appreciating the results of individual theories. We also discuss extensions of our main results to general resource theories in Sec. V.

2. *Quantum coherence and covariant operation*

Throughout this study, we consider the situation under the law of energy conservation. Under this constraint, we cannot create coherence between energy eigenstates with different energies without any help, while it decoheres very easily. We first clarify an *incoherent state*, which has no coherence in it:

Definition S.1 (Incoherent state)**.** Consider a system with Hamiltonian $H = \sum_i E_i \Pi_i$ where $\Pi_i := \sum_\alpha |E_i, \alpha\rangle\langle E_i, \alpha|$ is the projector onto the subspace of energy eigenstates with energy E_i. Here, $|E_i, \alpha\rangle$ is an energy eigenstate with energy E_i, and α distinguishes the degeneracy in the same energy. A state ρ is incoherent if this state is block-diagonal with respect to the energy eigenbasis: $\rho = \sum_i p_i \sigma_i$ with some probability distribution $\{p_i\}_i$ and state σ_i such that $\Pi_i \sigma_i \Pi_i = \sigma_i$ for all i. Equivalently, a state ρ is incoherent if $\rho = e^{-iHt} \rho e^{iHt}$ holds for any $t \in \mathbb{R}$.

The equivalence can be confirmed as follows: Let $\rho_{ij} := \langle i|\rho|j\rangle$ be an off-diagonal element in the sense of block diagonalization, i.e., $|i\rangle$ and $|j\rangle$ are energy eigenstates with the energies E_i and E_j such that $E_i \neq E_j$. Then, by expressing $\rho(t) := e^{-iHt}\rho e^{iHt}$, we have $\rho_{ij}(t) = e^{i(E_j-E_i)t}\rho_{ij}$. Hence, $\rho_{ij}(t) = \rho_{ij}$ is equivalent to $\rho_{ij} = 0$ for $E_i \neq E_j$.

Since a coherent state cannot be prepared from an incoherent state under energy conservation, coherence among energy eigenstates with different energies is considered as a precious resource in this setting, while incoherent states can be regarded as free states which we can freely use and waste.

The class of possible operations under the law of energy conservation is characterized by *covariant operations*. We consider a state transformation from system S to system S', whose Hamiltonians are respectively H_S and $H_{S'}$. Let ρ and ρ' be states in S and S'. We first present one definition of covariant operations whose physical picture is most transparent and then provide equivalent definitions that are more axiomatic.

Definition S.2 (Covariant operation)**.** An operation $\Lambda : S \to S'$ is a covariant operation if the following relation is satisfied: There exist auxiliary systems A and A' with Hamiltonians H_A and $H_{A'}$ satisfying $S \otimes A = S' \otimes A'$ such that the operation Λ can be expressed as

$$\Lambda(\rho) = \mathrm{Tr}_{A'}[U(\rho \otimes \eta)U^\dagger], \tag{1}$$

where U is an energy-conserving unitary satisfying

$$U(H_S \otimes I_A + I_S \otimes H_A)U^\dagger = H_{S'} \otimes I_{A'} + I_{S'} \otimes H_{A'}, \tag{2}$$

with the identity operator I, and η is an incoherent state in A.

The above definition manifests the fact that a covariant operation is implementable under the law of energy conservation with an incoherent state.

We remark that there are several equivalent characterizations of covariant operations: The equivalence of these definitions is proven in, e.g., Refs. [39, 40].

Proposition S.3. *An operation Λ: $S \to S'$ is a covariant operation if and only if*

$$\Lambda(e^{-iH_S t}\rho e^{iH_S t}) = e^{-iH_{S'}t}\Lambda(\rho)e^{iH_{S'}t} \tag{3}$$

holds for any ρ and any $t \in \mathbb{R}$.

Proposition S.4. *An operation Λ: $S \to S'$ is a covariant operation if and only if the following condition is satisfied: For any auxiliary system B with Hamiltonian H_B, if a state τ on SB is an incoherent state with respect to $H_S \otimes I_B + I_S \otimes H_B$, then the final state $\Lambda \otimes \mathcal{I}_B(\tau)$ is also an incoherent state with respect to $H_{S'} \otimes I_B + I_{S'} \otimes H_B$. Here, $\Lambda \otimes \mathcal{I}_B$ is an operation that applies Λ on system S and leave system B as it is.*

The above operational setting provides a resource theory with the set $\mathbb{F}$ of free states being incoherent states and the set $\mathbb{O}$ of free operations being covariant operations. This was introduced under the name of the resource theory of asymmetry [41], as the coherence is manifested by the resource that causes the asymmetry under time translation $\{e^{-iHt}\}_{t\in\mathbb{R}}$, which constructs a unitary representation of the group $\mathbb{R}$. Therefore, we also use the word "asymmetry" interchangeably to refer to coherence. Although the framework of resource theory of asymmetry encompasses the general group G, in this manuscript we refer to the case $G = \mathbb{R}$ with time translation as its unitary representation unless otherwise stated. For instance, we say "asymmetry measure" to refer to a resource measure defined for the resource theory of asymmetry in the above sense.

We also remark on the potential confusion regarding the use of the word "coherence". In the context of resource theories, another standard framework to describe quantum superposition is to look at off-diagonal terms with respect to a given orthonormal basis $\{|i\rangle\}_i$ [42], where there is no concept of energy or time translation. The coherence characterized in this framework is known as *speakable* coherence, while coherence relevant in this work is known as *unspeakable* coherence [43]. The former quantity is rooted in the computational notion where we are given a fixed computational basis, while the latter is relevant to the physical system equipped with a certain Hamiltonian, making it a relevant quantity for quantum clock and work extraction in the quantum thermodynamic setting. The properties of these two different notions of coherence are significantly different — we indeed see that our main results regarding unbounded amplification of coherence never apply to the speakable coherence, as we discuss in Sec. V B.

I.B Coherent modes

To formally state our result, let us introduce the notion of *coherent modes*. The idea of coherent modes stems from the observation that the difference in the energy levels to which coherence is attributed plays a crucial role in characterizing the coherence transformation. Classifying coherence in terms of energy levels can be formalized by the *modes of asymmetry* introduced in Ref. [44].

Definition S.5 (Modes of asymmetry [44])**.** The modes of asymmetry is defined as the set of a mode with non-zero coherence:

$$\mathcal{D}(\rho) := \left\{ \Delta_{ij} \;\middle|\; \Delta_{ij} = E_i - E_j,\ (i,j) \in \mathcal{M}(\rho) \right\}, \tag{4}$$

where $\mathcal{M}(\rho) := \{(i,j) | \langle E_i, \alpha | \rho | E_j, \beta \rangle \neq 0\}$ is the set of integer pairs (i,j) such that the off-diagonal element of ρ with respect to energy eigenstates with energies E_i and E_j is non-zero.

As we shall show in the remainder, extremely small but non-zero coherence can be amplified arbitrarily, while exactly zero coherence never becomes non-zero coherence. Therefore, we need to determine whether state ρ has coherence on the modes where ρ' also has coherence. To describe the presence and the absence of coherence on a mode, we define a set of *resonant coherent modes* in state ρ on a system equipped with energy levels $\{E_i\}_i$, which is an extension of the mode of asymmetry [44].

Definition S.6 (Set of resonant coherent modes)**.** Denoting by $\Delta_{ij} := E_i - E_j$, we define a set of resonant coherent modes $\mathcal{C}(\rho)$ as a linear combination of non-zero coherent mode of ρ with integer coefficients:

$$\mathcal{C}(\rho) := \left\{ x \;\middle|\; x = \sum_{(i,j)\in\mathcal{M}} n_{ij}\Delta_{ij} : n_{ij} \in \mathbb{Z} \right\}. \tag{5}$$

Here we consider a linear combination of coherent mode because if we have coherent modes with energy difference 1 and $\sqrt{2}$, then we can create a coherent mode with energy difference $1+\sqrt{2}$ through a covariant operation in both the asymptotic-marginal and correlated-catalytic transformation [45]. The relation $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$ means that all resonant coherent modes of ρ' are also those of ρ.

A simple but important fact on $\mathcal{C}(\rho)$ is that ρ and its tensor-product state $\rho^{\otimes n}$ has the same set of resonant coherent modes:

$$\mathcal{C}(\rho^{\otimes n}) = \mathcal{C}(\rho). \tag{6}$$

Appendix II: Asymptotic transformation

II.A Preliminaries

Before proceeding to our main arguments, we first present general frameworks of resource theories not restricted to quantum coherence, and then proceed to quantum coherence. One of the effective approaches to studying free transformation is to leverage the tools and ideas from Shannon theory, aiming to transform many copies of the initial state into many copies of the target state. We then study the rate of the number of copies of the initial and final states. This framework is generally called *asymptotic transformation*.

There are several approaches to quantifying the performance of asymptotic transformation. One of the ways is to consider the maximum transformation rate at which a free operation can transform many copies of ρ into a state that approaches many copies of the target state ρ'. Specifically, we call r an achievable rate if for any $\varepsilon > 0$ there exists a series $\{\Lambda_n\}_n$ of free operations (i.e., $\Lambda_n \in \mathbb{O}$ for all n) such that $\lim_{n\to\infty} \|\Lambda_n(\rho^{\otimes n}) - \rho'^{\otimes \lfloor rn \rfloor}\|_1 < \varepsilon$. The *asymptotic transformation rate* $R(\rho \to \rho')$ is the supremum of the achievable rates given by

$$R(\rho \to \rho') := \sup \left\{ r \;\middle|\; \lim_{n\to\infty} \|\Lambda_n(\rho^{\otimes n}) - \rho'^{\otimes \lfloor rn \rfloor}\|_1 = 0,\ \Lambda_n \in \mathbb{O} : S^{\otimes n} \to S'^{\otimes \lfloor rn \rfloor} \right\}. \tag{7}$$

In particular, when the target state ρ' is a pure state ϕ, the rate $R(\rho \to \phi)$ is customarily called *asymptotic distillation rate* with the target state ϕ. We note that there may exist a finite error in the final state for finite n, while it should vanish in the asymptotic limit.

The asymptotic transformation rate is relevant when one would like to prepare uncorrelated copies of the target state ρ'. Another operational setting is when multiparties are separate apart and each of them would like to obtain a state close to the target state ρ'. A reasonable goal in this setting is to obtain a state whose local marginal states are close to the target state while distributing the good local state to as many parties as possible. The transformation rate suitable for characterizing such a setting was previously studied [46, 47]. Here, we call it *asymptotic marginal transformation rate* and formally define it as follows.

Definition S.7 (Asymptotic marginal transformation rate)**.** Let ρ and ρ' be states on systems S and S'. We say that asymptotic marginal transformation rate r is achievable if there is a series $\{\Lambda_n\}_n$ of free operations with $\Lambda_n : S^{\otimes n} \to S'^{\otimes \lfloor rn \rfloor}$ such that $\lim_{n\to\infty} \max_i \|\mathrm{Tr}_{\setminus i}\Lambda_n(\rho^{\otimes n}) - \rho'\|_1 = 0$. Here, $\mathrm{Tr}_{\setminus i}$ represents a partial trace over subsystems except for the ith one. The *asymptotic marginal transformation rate* $\tilde{R}(\rho \to \rho')$ is the supremum over the achievable rates, i.e.,

$$\tilde{R}(\rho \to \rho') := \sup\left\{ r \;\middle|\; \lim_{n\to\infty} \max_i \|\mathrm{Tr}_{\setminus i}\Lambda_n(\rho^{\otimes n}) - \rho'\|_1 = 0,\ \Lambda_n : S^{\otimes n} \to S'^{\otimes \lfloor rn \rfloor} \in \mathbb{O} \right\}. \tag{8}$$

When ρ' is a pure state ϕ, we particularly call $\tilde{R}(\rho \to \phi)$ *asymptotic marginal distillation rate* with the target state ϕ. Unlike the standard asymptotic transformation, the asymptotic marginal transformation does not require the final state to approach a product of the target state at the infinite-copy limit, allowing for some correlation among the subsystem in the final state. However, if $\lfloor rn \rfloor$ copies are used separately and do not interact with each other, the final state is indistinguishable from the case of an exact transformation.

The contractivity of the trace distance under the partial trace ensures that $\|\Lambda_n(\rho^{\otimes n}) - \rho'^{\otimes \lfloor rn \rfloor}\|_1 < \varepsilon$ implies $\|\mathrm{Tr}_{\setminus i}\Lambda_n(\rho^{\otimes n}) - \rho'\|_1 < \varepsilon$, leading to $R(\rho \to \rho') \le \tilde{R}(\rho \to \rho')$. The previous study [46] found that the asymptotic marginal transformation is explicitly upper bounded as

$$R(\rho \to \rho') \le \tilde{R}(\rho \to \rho') \le \frac{G(\rho)}{G(\rho')}, \tag{9}$$

where G is a resource measure that is superadditive, tensor-product additive, and lower semi-continuous. In the case of the theory of entanglement, a much stronger claim was shown. Let $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ be the maximally entangled state. If ρ is distillable (i.e., $R(\rho \to \Phi) > 0$), the above two rates coincide; $R(\rho \to \Phi) = \tilde{R}(\rho \to \Phi)$ [47].

Our result also involves a stronger notion than the asymptotic marginal transformation, which we call *asymptotic exact marginal transformation*. In the asymptotic transformation, one typically allows non-zero errors that vanish asymptotically at the infinite-copy limit. This can be understood via the non-asymptotic transformation with non-zero error. Let ρ and ρ' be states on system S and S'. Then we define the asymptotic marginal transformation rate with the target state ρ' by

$$\tilde{R}(\rho \to \rho') := \lim_{\varepsilon\to 0} \lim_{n\to\infty} \sup\left\{ r \;\middle|\; \max_i \|\mathrm{Tr}_{\setminus i}[\Lambda_n(\rho^{\otimes n})] - \rho'\|_1 < \varepsilon,\ \Lambda_n : S^{\otimes n} \to S'^{\otimes \lfloor rn \rfloor} \in \mathbb{O} \right\}. \tag{10}$$

As is explicit in the above formulation, the asymptotic marginal transformation admits non-zero errors for every n as long as it asymptotically vanishes at the limit $n \to \infty$. On the other hand, we can also define the *asymptotic exact marginal transformation* and its rate as follows:

$$\tilde{R}^0(\rho \to \rho') := \lim_{n\to\infty} \sup\left\{ r \;\middle|\; \forall i\ \mathrm{Tr}_{\setminus i}[\Lambda_n(\rho^{\otimes n})] = \rho',\ \Lambda_n : S^{\otimes n} \to S'^{\otimes \lfloor rn \rfloor} \in \mathbb{O} \right\}. \tag{11}$$

As seen from the definition, the asymptotic exact marginal transformation does not allow any error, which is a much more stringent requirement than the asymptotic marginal transformation. In fact, since every sequence of operation $\{\Lambda_n\}_n$ that achieves the asymptotic marginal exact transformation clearly realizes the transformation whose error vanishes at the limit of $n \to \infty$, we always have $\tilde{R}^0(\rho \to \rho') \le \tilde{R}(\rho \to \rho')$.

We note that despite the clear relation between two asymptotic rates $R(\rho \to \rho') \le \tilde{R}(\rho \to \rho')$, there is no simple relation between the asymptotic rate R and asymptotic exact marginal rate $\tilde{R}^0$. The

asymptotic exact marginal transformation rate can similarly be introduced for the standard asymptotic transformation, and it was extensively studied in the context of zero-error distillable entanglement and entanglement cost [48–50].

In the theory of coherence, the performance of the standard asymptotic transformation rate was studied in the context of coherence distillation. Notably, Ref. [40] showed a fundamental restriction in the coherence distillation — every full-rank state has a zero asymptotic distillation rate.

Theorem S.8 (Coherence distillation is impossible [40])**.** *Consider the asymptotic transformation with covariant operations. For an arbitrary full-rank state ρ and a target pure coherent state ϕ, $R(\rho \to \phi) = 0$ holds.*

This result appears to suggest that obtaining high-quality coherent bits would come with a fundamental difficulty. However, we show that the distillation rate behaves in a dramatically different manner if we consider the asymptotic marginal transformation. Namely, we prove that these two rates take two opposite extremes, and the difference between these two becomes unbounded, realizing $R(\rho \to \phi) = 0$ and $\tilde{R}(\rho \to \phi) = \infty$ for almost all coherent states ρ. To the best of our knowledge, this is the first example of the resource theory for which these two rates do not coincide. We also show that the coherence manipulation is free from the severe restriction of the exact transformation when it comes to the asymptotic marginal transformation. The asymptotic exact marginal rate $\tilde{R}^0$ also diverges for most coherent states.

II.B Arbitrary coherence amplification: asymptotic marginal transformation (Theorem 1)

Now we are in a position to prove our first main result Theorem 1 in the main text. We break down the claim of Theorem 1 on the asymptotic marginal transformation into two parts: the diverging rate for $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$ (discussed in this subsection) and vanishing rate for $\mathcal{C}(\rho') \nsubseteq \mathcal{C}(\rho)$ (discussed in Sec. II E). We will see that the resonant coherent modes introduced above play an essential role in dividing these two regimes and completely characterizes the power of asymptotic marginal transformation.

We begin with the case when the infinite rate can be realized. We also show that the strength of correlation can be made arbitrarily small, which makes the asymptotic marginal transformation even more operationally relevant.

Theorem S.9 (The first part of Theorem 1 in the main text)**.** *For arbitrary states ρ and ρ', $\tilde{R}(\rho \to \rho')$ diverges if $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$. Moreover, the correlation between one subsystem and the others can be made arbitrarily small, i.e., for any $\varepsilon > 0$ the final state τ satisfies $\|\tau - \rho' \otimes \mathrm{Tr}_i \tau\|_1 < \varepsilon$ for any i-th copy.*

In fact, we show an even stronger claim, which shows that asymptotic exact marginal transformation rates also diverge for almost all transformations.

Theorem S.10. *Suppose ρ and ρ' are arbitrary states such that $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$ and ρ' is full rank. Then, the asymptotic exact marginal transformation rate $\tilde{R}^0(\rho \to \rho')$ diverges. Moreover, the correlation between one subsystem and the others can be made arbitrarily small, i.e., for any $\varepsilon > 0$ the final state τ satisfies $\|\tau - \rho' \otimes \mathrm{Tr}_i \tau\|_1 < \varepsilon$ for any i-th copy.*

We remark that Theorem S.9 is a direct consequence of Theorem S.10 as shown in the following.

Proof of Theorem S.9. Theorem S.9 can be derived from Theorem S.10 by noting that full-rank states are dense in the state space and thus every state has a full-rank state in its arbitrary neighborhood. Let ρ' be an arbitrary state in S' satisfying $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$. For a given $\delta > 0$, there exists a full-rank state $\tilde{\rho}$ in S' such that $\|\rho' - \tilde{\rho}\|_1 < \delta$. Theorem S.10 ensures that for every $r > 0$, there exists an integer n and a covariant operation $\Lambda : S^{\otimes n} \to S'^{\otimes \lfloor rn \rfloor}$ such that $\mathrm{Tr}_{\setminus i} \Lambda(\rho^{\otimes n}) = \tilde{\rho}$ for all i. This ensures that

$$\begin{aligned} \|\mathrm{Tr}_{\setminus i} \Lambda(\rho^{\otimes n}) - \rho'\|_1 &\leq \|\mathrm{Tr}_{\setminus i} \Lambda(\rho^{\otimes n}) - \tilde{\rho}\|_1 + \|\rho' - \tilde{\rho}\|_1 \\ &< \delta \end{aligned} \tag{12}$$

for every i, which shows that $R(\rho \to \rho')$ diverges. It is also guaranteed by Theorem S.10 that the final state $\tau = \Lambda(\rho^{\otimes n})$ satisfies $\|\tau - \rho' \otimes \mathrm{Tr}_{\setminus i}\tau\|_1 < \varepsilon$ for all i. □

Therefore, we focus on proving Theorem S.10. To this end, we introduce several results needed for the proof.

One of the key observations in proving our theorems is that any state conversion is possible by a covariant operation with the help of a class of auxiliary states called *marginal catalysts*.

Definition S.11 (Marginal-catalytic transformation)**.** We say that a state ξ in system S is convertible to a state ρ' in the system S' by a marginal-catalytic free transformation if there exists finite-dimensional catalytic systems $C_1, \ldots, C_M$ with states $c_1, \ldots, c_M$ and a free operation $\Lambda : S \otimes C_1 \otimes \cdots \otimes C_M \to S' \otimes C_1 \otimes \cdots \otimes C_M \in \mathbb{O}$ such that

$$\tau = \Lambda(\xi \otimes c_1 \otimes \cdots \otimes c_M), \quad \mathrm{Tr}_{\setminus S}\tau = \rho', \quad \mathrm{Tr}_{\setminus C_i}\tau = c_i \ \forall i. \tag{13}$$

The marginal-catalytic transformation was first introduced in the context of quantum thermodynamics [51], and it was recently shown that an arbitrary state transformation is possible by a covariant operation with a marginal catalyst [45]. The precise statement is as follows:

Lemma S.12 (Theorem 2 in Ref. [45])**.** *For any two quantum states ξ and ρ', and for any accuracy $\varepsilon > 0$, there exists a set of two-level catalytic systems $C_1, \ldots, C_N$ with full-rank qubit mixed states $c_1, \ldots, c_N$ and a covariant operation Λ: $S \otimes C_1 \otimes \cdots \otimes C_N \to S' \otimes C_1 \otimes \cdots \otimes C_N$ such that $\Lambda(\xi \otimes c_1 \otimes \cdots \otimes c_N) = \tau$ with $\|\mathrm{Tr}_{C_1,\ldots,C_N}[\tau] - \rho'\|_1 < \varepsilon$ and $\mathrm{Tr}_{\setminus C_i}[\tau] = c_i$ for all $i = 1, \ldots, N$.*

Note that although the above statement is on the marginal-catalytic transformations with vanishing error, as demonstrated soon after (Proposition. S.15) this can be strengthened into the form of the exact marginal-catalytic transformations.

The framework of marginal-catalytic transformations comes with a less clear operational meaning compared to the correlated catalyst, as the final catalyst cannot be reused indefinitely due to the correlation among catalytic subsystems. Nevertheless, we show that the marginal-catalytic covariant transformation constructed in Ref. [45] serves as an effective subroutine in our protocols.

The key property of marginal catalysts that lends themselves to useful subroutines is that, although the marginal catalysts cannot be reused indefinitely for state transformation, they can still give some operational advantage. It was shown that they are *partially* reusable, meaning that the catalyst allows a larger number of transformations than the number of sets of catalysts provided. For brevity, we denote a set $C_1, \ldots, C_N$ of catalytic systems and their initial states $c_1, \ldots, c_N$ by $\boldsymbol{C}$ and $\boldsymbol{c}$, respectively.

Lemma S.13 (Supplemental Material of Ref. [45])**.** *Let ξ and ρ' be arbitrary states in systems S and S'. For any accuracy $\varepsilon > 0$, let $\boldsymbol{C}$ and $\boldsymbol{c}$ be the sets of catalytic systems and states ensured by Lemma S.12. Then, N^k sets of catalysts $\boldsymbol{c}^{\otimes N^k}$ in $\boldsymbol{C}^{\otimes N^k}$ can realize $(k+1)N^k$ marginal-catalytic transformation from ξ to ρ', i.e., there is a covariant operation Λ: $S^{\otimes(k+1)N^k} \otimes \boldsymbol{C}^{\otimes N^k} \to S'^{\otimes(k+1)N^k} \otimes \boldsymbol{C}^{\otimes N^k}$ such that*

$$\Lambda(\xi^{\otimes(k+1)N^k} \otimes \boldsymbol{c}^{\otimes N^k}) = \tau, \quad \|\mathrm{Tr}_{\setminus S_j}[\tau] - \rho'\|_1 < \varepsilon, \quad \mathrm{Tr}_{\setminus C_{ij}}[\tau] = c_i \ \forall i \tag{14}$$

for every $1 \le j \le (k+1)N^k$. Here, S_j denotes the j th main system, and C_{ij} refers to the i th catalyst in the j th copy.

The idea is to reuse the marginal catalysts by recombining them in a way that the correlation present in the final state does not affect the next round of transformation. Below we present a way of such a recombination. We label N^k copies of catalysts C_j by a tuple of k integers $(n_1, n_2, \ldots, n_k)$ with $n_i \in \{1, 2, \ldots, N\}$. In the first step, we coordinate catalysts into N^k groups of $C_1, \ldots, C_N$ such that catalysts in the same group have the same label $(n_1, n_2, \ldots, n_k)$. Using these N^k sets of catalysts, we have N^k conversions in this step. In the l-th step $(2 \le l \le k+1)$, we coordinate catalysts into N^k groups of $C_1, \ldots, C_N$ in the following manner: N catalysts in the same group has the same labels $(n_1, n_2, \ldots, n_k)$ except for n_l, and there exists an integer g such that the label n_l with catalyst C_j is written as $n_l = g + j \mod N$ for all C_j. Using these N^k sets of catalysts, in each step we have N^k conversions. It is easy to confirm that in any group all the catalysts have no correlation before the catalytic transformation. Through the whole procedure, we have $(k+1)N^k$ conversions. We refer interested readers to the first section of the Supplemental Material of Ref. [45] for further details.

We also introduce a useful lemma, which connects approximate transformations to exact transformations. The essential idea was shown by Wilming [52] for the catalytic entropy conjecture, and following this we present a statement in a general form.

Lemma S.14 (Ref. [52]). *We consider a quantum system equipped with the trace distance $d(\rho,\sigma) := \frac{1}{2}\|\rho-\sigma\|_1$. Given a sequence of convex closed sets of quantum states $\{S_n\}_{n=1}^{\infty}$ satisfying $S_n \subseteq S_{n+1}$, let V be a set of states such that for any $\varepsilon > 0$ and any $\sigma \in V$ there exist a sufficiently large N and a state $\eta \in S_N$ such that $d(\sigma,\eta) < \varepsilon$. If κ is an interior state of V in terms of distance d, then there exists an integer n such that $\kappa \in S_n$.*

Let S_n be the set of exactly convertible states with parameter n (e.g., the number of copies and the size of a catalyst), and let V be the set of approximately convertible states from ρ. Lemma S.14 suggests that if we can convert ρ to κ approximately and κ is an interior state of V, then we can convert ρ to κ exactly. The condition of convexity is fulfilled if a classical mixture is a free operation in the resource theory in consideration. This fact enables us to interpret results on approximate transformations to exact transformations.

Proof of Lemma S.14. We prove it by contradiction. Suppose contrarily that $\kappa \in V$ is an interior state while it does not belong to $S_\infty := \lim_{n\to\infty} S_n$.

Consider an open ball $B_\delta = \left\{ \rho \;\middle|\; \|\rho-\kappa\|_1 \le \delta,\ \rho \ge 0,\ \mathrm{Tr}(\rho) = 1 \right\}$ with its center κ and radius δ such that $B_\delta \subseteq V$. By definition, there exists n such that for any $\eta \in B_\delta$, S_n has a state $\xi(\eta) \in S_n$ satisfying $\|\xi(\eta)-\eta\|_1 < \delta/2$. Since S_n is convex and $\kappa \notin S_n$, there uniquely exists a state $a \in S_\infty$ closest to κ. Let $\rho(t) := -(t-1)a + t\kappa,\ t \ge 1$ be the set of states on the ray from κ along the line going through a and κ. Since $\mathrm{Tr}(\rho(t)) = 1$ for every t, $\rho(t)$ is a quantum state if and only if $\rho(t) \ge 0$. We also have $\|\rho(t)-\kappa\|_1 = (t-1)\|a-\kappa\|_1$, which is continuous and increasing with t. Therefore, for every $\delta' < \delta$, there exists $t' \ge 1$ such that $\|\rho(t')-\kappa\|_1 = \delta'$ and $\rho(t') \in B_\delta$. Taking $\delta' = 2\delta/3$, we get

$$\|\rho(t')-a\| = t'\|a-\kappa\|_1 \ge (t'-1)\|a-\kappa\|_1 = \|\rho(t')-\kappa\|_1 = 2\delta/3. \tag{15}$$

We also note that

$$\min_{b\in S_n}\|\rho(t')-b\| = \min_{b\in S_n}\|-(t'-1)a+t'\kappa-b\|_1 = t'\min_{b\in S_n}\left\|\kappa-\left[(1-t'^{-1})a+t'^{-1}b\right]\right\|_1 = t'\|\kappa-a\|_1 \tag{16}$$

where in the last equality, we used the fact that $(1-t'^{-1})a + t'^{-1}b \in S_n$ due to the convexity of S_n and the assumption that a is the closest state to κ and thus $b = a$ achieves the minimum. Together with the assumption of S_n that $\min_{b\in S_n}\|\rho(t')-b\| \le \delta/2$, we obtain

$$\delta/2 \ge \min_{b\in S_n}\|\rho(t')-b\| = \|\rho(t')-a\| \ge \|\rho(t')-\kappa\|_1 = 2\delta/3 \tag{17}$$

which is a contradiction.

□

Applying Lemma S.14 to Lemma S.12, we have the following proposition, showing that, if the target state is full rank, the marginal transformation can be made exact.

Proposition S.15. *For any quantum state ξ and a full-rank state ρ', there exists a set of two-level catalytic systems $C_1,\dots,C_N$ with full-rank mixed states $c_1,\dots,c_N$ and a covariant operation Λ such that $\Lambda(\xi\otimes c_1\otimes\cdots\otimes c_N) = \tau$ with $\mathrm{Tr}_{C_1,\dots,C_N}[\tau] = \rho'$ and $\mathrm{Tr}_{\setminus C_i}[\tau] = c_i$ for all $i = 1,\dots,N$.*

We also show that an arbitrarily good qubit coherent state can be prepared from finite copies of a general state ρ as long as it contains the mode of asymmetry for the qubit coherent state (recall Definition S.5).

Lemma S.16. *Let ρ be an arbitrary state and σ be an arbitrary two-level state such that $\mathcal{D}(\sigma) \subseteq \mathcal{D}(\rho)$. Then, for every $\varepsilon > 0$, there exists a positive integer n and a covariant operation Λ such that $\|\Lambda(\rho^{\otimes n}) - \sigma\|_1 < \varepsilon$.*

Proof. For every $\Delta \in \mathcal{D}(\rho)$, one can transform ρ to a weakly coherent qubit state η with $\mathcal{D}(\eta) = \{0, \pm\Delta\}$ by a covariant operation using the protocol in Ref. [45, Lemma 11 in Supplemental Material]. Let Λ_1 be this covariant operation such that $\Lambda_1(\rho) = \eta$. As pointed out in Ref. [40], the protocol introduced in Ref. [53] allows one to transform many weakly coherent qubit states to one copy of a good coherent state by a covariant operation. That is, for every $\varepsilon' > 0$, there is a positive integer m and a covariant operation Λ_2 such that $\|\Lambda_2(\eta^{\otimes m}) - |+\rangle\langle+|\|_1 < \varepsilon'$ where $|+\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ is the maximally coherent state with $\Delta \in \mathcal{D}(|+\rangle\langle+|)$.

We now employ the fact that every full-rank qubit state σ can be prepared exactly by a covariant operation from finite copies of $|+\rangle\langle+|$ defined in the same system. This is because the coherence cost $1/R(|+\rangle\langle+| \to \sigma)$ is proportional to the ratio of the quantum Fisher information of σ to that of $|+\rangle$ [54], and $|+\rangle$ has the largest quantum Fisher information in a two-level system, particularly ensuring that the coherence cost is upper bounded by 1. Therefore, for every $\delta > 0$ there is an integer k' and a covariant operation $\tilde{\Lambda}_3$ such that $\|\tilde{\Lambda}_3(|+\rangle\langle+|^{\otimes k'}) - \sigma^{\otimes k'}\|_1 < \delta$. Taking the partial trace other than the first subsystem and noting the data-processing inequality of the trace norm, we get $\|\Lambda'_3(|+\rangle\langle+|^{\otimes k'}) - \sigma\|_1 < \delta$ where $\Lambda'_3 = \mathrm{Tr}_{\setminus 1} \circ \tilde{\Lambda}_3$ is also a covariant operation. We then note that this applies to an arbitrary qubit state σ, which ensures the existence of the exact transformation that prepares σ because of Lemma S.14. We let Λ_3 be such a covariant operation satisfying $\Lambda_3(|+\rangle\langle+|^{\otimes k}) = \sigma$ for some integer k.

Then,

$$\begin{aligned} \|\Lambda_3 \circ \Lambda_2^{\otimes k} \circ \Lambda_1^{\otimes km}(\rho^{\otimes km}) - \sigma\|_1 &\leq \|\Lambda_3 \circ \Lambda_2^{\otimes k}(\eta^{\otimes km}) - \Lambda_3(|+\rangle\langle+|^{\otimes k})\|_1 \\ &\leq \| \left[\Lambda_2(\eta^{\otimes m})\right]^{\otimes k} - |+\rangle\langle+|^{\otimes k}\|_1 \\ &< k\varepsilon' \end{aligned} \tag{18}$$

where in the last inequality we used $\|\Lambda_2(\eta^{\otimes m}) - |+\rangle\langle+|\|_1 \leq \varepsilon'$ and the following general inequality satisfying for all states ρ_1 and ρ_2 and every integer n:

$$\begin{aligned} \|\rho_1^{\otimes n} - \rho_2^{\otimes n}\|_1 &\leq \|\rho_1^{\otimes n} - \rho_2 \otimes \rho_1^{\otimes n-1}\|_1 + \|\rho_2 \otimes \rho_1^{\otimes n-1} - \rho_2^{\otimes 2} \otimes \rho_1^{\otimes n-2}\|_1 + \cdots + \|\rho_2^{\otimes n-1} \otimes \rho_1 - \rho_2^{\otimes n}\|_1 \\ &\leq n\|\rho_1 - \rho_2\|_1. \end{aligned} \tag{19}$$

Since ε' in (18) can be taken as small as one wishes, the target error ε can be realized by choosing $\varepsilon' = \varepsilon/k$. Noting that $\Lambda := \Lambda_3 \circ \Lambda_2^{\otimes k} \circ \Lambda_1^{\otimes km}$ is also a covariant operation concludes the proof. □

Using Lemma S.14, we can improve Lemma S.16 as the exact preparation of arbitrary full-rank final states.

Lemma S.17. *Let ρ be an arbitrary state and σ be an arbitrary two-level full-rank state such that $\mathcal{D}(\sigma) \subseteq \mathcal{D}(\rho)$. Then, there exists a positive integer n and a covariant operation Λ such that $\Lambda(\rho^{\otimes n}) = \sigma$.*

We are now in a position to present the proof of Theorem S.10.

Proof of Theorem S.10. We first construct an asymptotic conversion protocol whose transformation rate is larger than any integer R without taking into account the small correlation condition, and then demonstrate how to suppress correlation.

Let ρ be a state in S and ρ' be a full-rank state in S' such that $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$. We aim to construct the catalysts $c_1, \ldots, c_N$ by a covariant operation from multiple (but finite) copies of the initial state ρ. Lemma S.17 ensures that there exists $\{\mu_i\}_{i=1}^N$ such that $\rho^{\otimes \mu_i}$ is convertible to c_i exactly by a covariant operation. We write $\mu := \sum_{i=1}^N \mu_i$. We remark that the exact preparation of the catalysts here is essential to avoid the errors from accumulating during the multiple uses of the prepared catalysts $c_1, \ldots, c_N$.

We now construct the desired covariant operation as follows: We start with μN^k copies of ρ. We first convert μN^k copies of ρ into N^k copies of $c_1 \otimes \cdots \otimes c_N$ by the above protocol. Then, using the partial reusability of the marginal catalysts shown in Lemma S.13 and the exact marginal-catalytic covariant transformation ensured by Proposition S.15, a set of free states $\xi^{\otimes (k+1)N^k}$ can be converted into state Σ on $S'^{\otimes (k+1)N^k}$ such that its reduced state to any single copy $i = 1, \ldots, (k+1)N^k$ is equal to ρ' exactly ($\mathrm{Tr}_{\setminus i}[\Sigma] = \rho'$ for all $1 \leq i \leq (k+1)N^k$). Overall, this protocol realizes the covariant

operation that transforms $\rho^{\otimes n}$ to m copies of ρ' with $n := \mu N^k$ and $m := (k+1)N^k$ in the sense of the marginal asymptotic conversion.

Since

$$\frac{m}{n} = \frac{(k+1)N^k}{\mu N^k} = \frac{k+1}{\mu} \tag{20}$$

can become arbitrarily large by taking a sufficiently large k, we get that any transformation rate R is an achievable asymptotic exact marginal transformation rate.

We now demonstrate how to ensure that the correlation between a copy and the remainder of copies is less than ε in the sense of trace distance. Let κ be a state such that (i) κ is ε' close to a pure state (i.e., there exists a pure state ψ such that $\|\kappa - \psi\|_1 \leq \varepsilon'$), (ii) κ is convertible to ρ' by a covariant operation, (iii) $\mathcal{C}(\kappa) = \mathcal{C}(\rho')$. The existence of such κ is guaranteed by Lemma S.17 as follows. First, if $\tilde{\kappa}$ satisfies (i) and (iii), then $\kappa = \tilde{\kappa}^{\otimes r}$ with sufficiently large r also satisfies (i) with $\varepsilon' \to r\varepsilon'$ and (iii), as well as (ii) due to Lemma S.17. It therefore suffices to ensure the existence of such $\tilde{\kappa}$ satisfying (i) and (iii) with replacing ε' by ε'/r. Such a state for an arbitrary $\varepsilon' > 0$ can be constructed by, e.g., purifying ρ' using the auxiliary system with a trivial Hamiltonian and applying the depolarizing channel with sufficiently small noise strength.

We replace ρ' in the above protocol by κ, and let Λ be the covariant operation converting $\rho^{\otimes n}$ to $\kappa^{\otimes m}$. Due to the condition (i), the correlation between state κ and other copies is bounded as $\|\Lambda(\rho^{\otimes n}) - \kappa \otimes \mathrm{Tr}_i[\Lambda(\rho^{\otimes n})]\|_1 < 2\varepsilon' + \sqrt{\varepsilon'/2}$ [47] for all i. Take small enough ε' such that $2\varepsilon' + \sqrt{\varepsilon'/2} < \varepsilon$ and let κ be a state satisfying (i)–(iii) for this ε'. Let $\mathcal{E}$ be a covariant operation $\mathcal{E}(\kappa) = \rho'$ ensured by the condition (ii). Then, $\tau := \mathcal{E}^{\otimes m} \circ \Lambda(\rho^{\otimes n})$ is the desired final state because

$$\mathrm{Tr}_{\setminus i}\tau = \mathcal{E}(\mathrm{Tr}_{\setminus i}\Lambda(\rho^{\otimes n})) = \mathcal{E}(\kappa) = \rho' \tag{21}$$

for all i, and

$$\begin{aligned} \|\tau - \rho' \otimes \mathrm{Tr}_i\tau\|_1 &= \|\mathcal{E}^{\otimes m} \circ \Lambda(\rho^{\otimes n}) - \mathcal{E}^{\otimes m}(\kappa \otimes \mathrm{Tr}_i\Lambda(\rho^{\otimes n})\|_1 \\ &\leq \|\Lambda(\rho^{\otimes n}) - \kappa \otimes \mathrm{Tr}_i\Lambda(\rho^{\otimes n})\|_1 \\ &< \varepsilon \end{aligned} \tag{22}$$

for all i, where we used the contractivity of the trace norm under $\mathcal{E}^{\otimes m}$. □

We remark that our protocol requires exact catalysts, not approximated ones, unlike the previous result on entanglement [47], where the catalyst can be an approximated one. This difference comes from the difference between a correlated single catalyst and multiple marginal catalysts. If we employ a single correlated catalyst with some error, this error does not increase through the above process. On the other hand, if we employ multiple marginal catalysts with some errors, these errors may increase through the above process. To avoid this trouble, we should prepare catalysts without any small errors.

II.C Consistency with zero coherence distillation rate

Theorem S.9 ensures that one can produce a state with an arbitrary size whose marginal is arbitrarily close to a pure state—which includes the maximally coherent state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$—with an arbitrarily small correlation between one and the other subsystems. One may wonder how this would be consistent with the vanishing asymptotic distillation rate $R(\rho \to \phi) = 0$ for every full-rank state ρ and pure coherent state ϕ [40], as it might appear that the final state with arbitrarily small correlation should also be arbitrarily close to the tensor product of ϕ, leading to the contradiction with the zero asymptotic distillation rate. Here, we show that these two results are consistent.

The key observation is that even if a state τ_m satisfies $S'^{\otimes m}$ with $\|\mathrm{Tr}_{\setminus i}\tau_m - \phi\|_1 < \varepsilon$ for all i and $\|\tau_m - \phi \otimes \mathrm{Tr}_i\tau\|_1 < \varepsilon$ with small $\varepsilon > 0$, τ_m can generally be far from $\phi^{\otimes m}$ for large m. Here we bound both the error in each subsystem and the amount of correlation by the same variable ε for notational simplicity. We denote the distance between τ_m and $\phi^{\otimes m}$ by

$$\|\tau_m - \phi^{\otimes m}\|_1 := f(m, \varepsilon). \tag{23}$$

Then, in order for $\{\tau_m\}_m$ to be a valid family of states for the (standard) asymptotic transformation by taking small ε, it is required that

$$\lim_{\varepsilon\to 0}\lim_{m\to\infty} f(m,\varepsilon) = 0. \tag{24}$$

However, this requirement does not hold in general. For instance, let $\phi = |+\rangle\langle+|$ and consider a class of states defined by

$$\tau_m = (1-\delta)\left[(1-\varepsilon/2)|+\rangle\langle+| + \frac{\varepsilon}{2}|-\rangle\langle-|\right]^{\otimes m} + \frac{\delta}{2}(|+\rangle\langle+|^{\otimes m} + |-\rangle\langle-|^{\otimes m}). \tag{25}$$

This satisfies $\|\mathrm{Tr}_{\setminus i}\tau_m - |+\rangle\langle+|\|_1 \le \varepsilon$ and $\|\tau_m - |+\rangle\langle+| \otimes \mathrm{Tr}_i\tau\|_1 \le \varepsilon$ for every i for sufficiently small δ. On the other hand,

$$f(m,\varepsilon) = 2[1-(1-\delta)(1-\varepsilon/2)^m - \delta/2]. \tag{26}$$

which satisfies $\lim_{\varepsilon\to 0}\lim_{m\to\infty} f(m,\varepsilon) = 2(1-\delta/2) > 0$.

This example shows that the combination of "good local states" and "small correlation" does not necessarily result in a "good global state" and ensures that Theorem S.9 is not in contradiction with the zero asymptotic distillation rate. We remark that the above example may not be a general form that can be obtained by our protocol. Nevertheless, this example is already sufficient to argue that there is no definite inconsistency between our result and the zero asymptotic distillation rate.

II.D Consistency with asymmetry measure

Besides the zero asymptotic distillation rate, one might also find it strange that an arbitrary number of highly coherent states can be prepared in every subsystem from the viewpoint of resource measures. Theorem S.9 shows that from a state with an arbitrarily small amount of coherence, one could prepare a state whose local state is arbitrarily close to a highly coherent state with an arbitrarily small correlation. If we naively think that the small correlation would ensure that the total coherence is approximately the sum of local asymmetry, this appears to lead to the contradiction.

To formalize this concern, consider a tensor-product additive asymmetry measure, i.e., an asymmetry measure R satisfying $R(\otimes_i \rho_i) = \sum_i R(\rho_i)$. The standard tensor-product additive asymmetry measure includes quantum Fisher information [55] and Wigner-Yanase skew information [56, 57], which are in a family of additive asymmetry measures known as metric-adjusted skew informations [57, 58].

As in the previous subsection, let $\tau_m = \Lambda(\rho^{\otimes n})$ be a state in $S'^{\otimes m}$ such that $\|\mathrm{Tr}_{\setminus i}\tau_m - \phi\|_1 < \varepsilon$ for all i and $\|\tau_m - \phi\otimes\mathrm{Tr}_i\tau\|_1 < \varepsilon$ for some pure coherent state ϕ. Using the tensor-product additivity of R and the monotonicity of R under covariant operations, we get

$$R(\rho) = \frac{1}{n}R(\rho^{\otimes n}) \ge \frac{1}{n}R(\Lambda(\rho^{\otimes n})) = \frac{1}{n}R(\tau_m). \tag{27}$$

If R is also continuous with respect to the trace distance, Theorem S.9 claims that even for a state ρ for which $R(\rho)$ is arbitrarily close to 0, (27) holds for an arbitrarily large m and arbitrarily small ε for a sufficiently large n. This appears a counterintuitive claim, and indeed, if τ_m had no correlation at all, this would immediately meet the contradiction. This is because if $\tau_m = \otimes_i\rho'_i$ for some coherent states ρ'_i close to ϕ, then tensor-product additivity of R would imply $\frac{1}{n}R(\tau_m) = \frac{1}{n}\sum_{i=1}^m R(\rho'_i) \sim \frac{m}{n}R(\phi)$, which could be made arbitrarily large by taking large m.

The above concern is based on the naive observation that (1) τ_m should be close to $\phi^{\otimes n}$, and (2) $R(\phi^{\otimes m}) = mR(\phi)$, therefore (3) $R(\tau_m)$ should be close to $mR(\phi)$, which might cause an inconsistency. To see where this argument breaks down, let us consider the continuity of R. Let ρ_1 and ρ_2 be two d-dimensional states and $t := \|\rho_1 - \rho_2\|_1$. The continuity can be formalized as

$$|R(\rho_1) - R(\rho_2)| \le g(d,t) \tag{28}$$

where g is a continuous bounded function for $t \in [0,1]$ such that $\lim_{t\to 0} g(d,t) = 0$. Let us take $m = rn$ and $t_{r,n,\varepsilon} := \|\tau_{rn} - \phi^{\otimes rn}\|_1$. Then, the tensor-product additivity and the continuity of R

gives

$$\begin{aligned} R(\rho) &\geq \frac{1}{n} R(\tau_m) \\ &\geq rR(\phi) - \frac{g(2^{rn}, t_{r,n,\varepsilon})}{n}. \end{aligned} \tag{29}$$

This inequality prohibits the case of $\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{g(2^{rn},t_{r,n,\varepsilon})}{n} = 0$, because by taking sufficiently small ε and sufficiently large n (that depends on the chosen ε), the right-hand side of (29) could become arbitrarily close to $rR(\phi)$, which would violate (29) by taking large enough r.

To avoid $\lim_{n\to\infty}\frac{g(2^{rn},t_{r,n,\varepsilon})}{n} = 0$, the function g must scale with d at least as $\log d$ (i.e., linear in n). If g grows faster than $\log d$, then we get $\lim_{n\to\infty} g(2^{rn}, t_{r,n,\varepsilon})/n = \infty$, making (29) consistent. Therefore, the only case that the consistency with (29) might be in question is when g grows with $\log d$ and gives $g(2^{rn}, t_{r,n,\varepsilon})/n =: h(t_{r,n,\varepsilon})$ satisfying $\lim_{x\to 0} h(x) = 0$. When this holds, the function R is said to be *asymptotically continuous* [59]. In this case, the issue may arise if $\lim_{\varepsilon\to 0}\lim_{n\to\infty} t_{r,n,\varepsilon} = 0$ for an arbitrary τ_{rn} such that $\|\mathrm{Tr}_{\setminus i}\tau_{rn} - \phi\|_1 \leq \varepsilon$ for all i and $\|\tau_{rn} - \phi\otimes\mathrm{Tr}_i\tau_{rn}\|_1 \leq \varepsilon$. However, this is not the case in general as we show in the previous section with the example in (25). This confirms that the apparent inconsistency with additive asymmetry measures actually does not arise.

In fact, the above argument can be employed to obtain a general continuity property of weakly tensor-product additive functions, which may be of independent interest. Let f be a map from quantum states to real numbers satisfying $|f(\rho_1) - f(\rho_2)| \leq g(d, t)$ for arbitrary d-dimensional quantum states ρ_1 and ρ_2, where $t = \|\rho_1 - \rho_2\|_1$ and g is a continuous bounded function for $t \in [0, 1]$ such that $\lim_{t\to 0} g(d, t) = 0$. We say that f is *more than asymptotically continuous* [60] if $\lim_{d\to\infty}\frac{g(d,t_d)}{\log d} = 0$ for an arbitrary sequence $\{t_d\}_d$ such that $t_d \in [0, 1]$.

Proposition S.18. *Let f be an arbitrary non-constant map from quantum states to real numbers that is weakly tensor-product additive, i.e., $f(\rho^{\otimes n}) = nf(\rho)$ for an arbitrary state ρ and a positive integer n. Then, f cannot be "more than asymptotically continuous".*

Proof. Since f is not a constant map, there exist states ρ and σ such that $f(\rho) < f(\sigma)$. Let d be the dimension of ρ and σ, and let $t_n := \|\rho^{\otimes n} - \sigma^{\otimes n}\|_1$. Then, we get

$$\begin{aligned} f(\rho) &= \frac{1}{n} f(\rho^{\otimes n}) \\ &\geq \frac{1}{n} f(\sigma^{\otimes n}) - \frac{g(d^n, t_n)}{n} \\ &= f(\sigma) - \frac{g(d^n, t_n)}{n} \end{aligned} \tag{30}$$

where the first and the last equalities are due to the weak tensor-product additivity. If f is more than asymptotically continuous, we have $\lim_{n\to\infty} g(d^n, t_n)/n = 0$. This then implies $f(\rho) \geq f(\sigma)$ by taking the $n \to \infty$ limit on both sides, which is a contradiction. □

We remark that Ref. [60] showed—with a different technique—that tensor-product additive, permutationally invariant non-constant functions cannot be more than asymptotically continuous. Proposition S.18 extends it to all *weakly* tensor-product additive non-constant functions, which may not necessarily be permutationally invariant.

II.E No-go theorem for asymptotic marginal transformation (Theorem 1)

Theorem S.9 states that if $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$ is satisfied, the asymptotic marginal transformation rate $\tilde{R}(\rho \to \rho')$ becomes unbounded. Here, we show its opposite: $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$ implies $\tilde{R}(\rho \to \rho') = 0$, showing that the condition in terms of the set of resonant coherent modes serves as a sharp threshold between the infinite and zero transformation rates.

Theorem S.19 (The second part of Theorem 1 in the main text)**.** *If $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$, even a single copy of ρ' cannot be prepared from any number of copies of ρ with arbitrarily small error, which particularly implies that the asymptotic marginal transformation rate becomes zero: $\tilde{R}(\rho \to \rho') = 0$.*

This result can be formalized as follows.

Theorem S.20. *Suppose two states ρ and ρ' satisfy $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$. Then, there exists $\varepsilon > 0$ such that for every sequence $\{\Lambda_n\}_n$ of covariant operations, $\|\Lambda_n(\rho^{\otimes n}) - \rho'\|_1 > \varepsilon$ holds for all n.*

The main idea behind the proof is that the modes of asymmetry defined in Definition S.5 cannot be created by a covariant operation [44].

Proof of Theorem S.20. Suppose, to the contrary, that ρ and ρ' satisfy $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$ but there exists a series $\{\Lambda_n\}_n$ of covariant operations such that for every $\varepsilon > 0$, there is a sufficiently large n satisfying $\|\Lambda_n(\rho^{\otimes n}) - \rho'\|_1 < \varepsilon$.

Take sufficiently small $\varepsilon > 0$ such that $\mathcal{C}(\rho') \subseteq \mathcal{C}\left(\Lambda_n(\rho^{\otimes n})\right)$. There always exists such ε because each entry of the density matrix is continuous with respect to the change in the density matrix with trace distance, and thus every non-zero mode remains as a non-zero mode with a sufficiently small perturbation (while zero modes could turn into non-zero modes under arbitrarily small perturbation.)

This means that, together with the assumption $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$, there exists n and a covariant operation Λ_n such that $\mathcal{C}\left(\Lambda_n(\rho^{\otimes n})\right) \not\subseteq \mathcal{C}(\rho)$. By Definitions S.5 and S.6 of modes of asymmetry and sets of resonant coherent modes, $\mathcal{D}(\rho_1) \subseteq \mathcal{C}(\rho_2)$ implies $\mathcal{C}(\rho_1) \subseteq \mathcal{C}(\rho_2)$ for arbitrary states ρ_1 and ρ_2. This observation leads to $\mathcal{D}\left(\Lambda_n(\rho^{\otimes n})\right) \not\subseteq \mathcal{C}(\rho)$, because if $\mathcal{D}\left(\Lambda_n(\rho^{\otimes n})\right) \subseteq \mathcal{C}(\rho)$ held contrarily to our claim, then $\mathcal{C}\left(\Lambda_n(\rho^{\otimes n})\right) \subseteq \mathcal{C}(\rho)$ would also hold, contradicting our previous finding $\mathcal{C}\left(\Lambda_n(\rho^{\otimes n})\right) \not\subseteq \mathcal{C}(\rho)$.

Notice that

$$\mathcal{D}(\rho^{\otimes m}) = \left\{ x \;\middle|\; x = \sum_{(i,j)\in\mathcal{M}} n_{ij} \Delta_{ij} : n_{ij} \in \mathbb{Z},\ \sum_{(i,j)} |n_{ij}| \leq m \right\} \subseteq \mathcal{C}(\rho) \tag{31}$$

for any m. This, together with $\mathcal{D}(\Lambda_n(\rho^{\otimes n})) \not\subseteq \mathcal{C}(\rho)$, implies $\mathcal{D}\left(\Lambda_n(\rho^{\otimes n})\right) \not\subseteq \mathcal{D}(\rho^{\otimes n})$. However, this contradicts the fact that the modes of asymmetry cannot be created by a covariant operation [44], i.e., $\mathcal{D}(\Lambda(\rho)) \subseteq \mathcal{D}(\rho)$ for every state ρ and covariant operation Λ. □

We also obtain the no-go theorem for asymptotic exact marginal transformation. The case when $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$ can be shown as a direct consequence of Theorem S.20. We can also add another constraint when ρ is pure.

Theorem S.21. *If either (1) $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$ or (2) ρ' is pure coherent state and ρ is full rank, there is no covariant operation Λ such that $\Lambda(\rho^{\otimes n}) = \rho'$ for any integer n. As a result, we get $\tilde{R}^0(\rho \to \rho') = 0$*

Proof. Theorem S.20 implies the part for the case when $\mathcal{C}(\rho') \not\subseteq \mathcal{C}(\rho)$. The case when ρ' is a pure coherent state and ρ is full rank is prohibited by the fact that pure coherent states have the diverging purity of coherence [40] while full rank states have the finite purity of coherence. □

Appendix III: Catalytic transformation

III.A Preliminaries

Another approach to enhance the desired transformation is to use an auxiliary state that aids the transformation. In particular, when the auxiliary state is returned to the original form, it is called a *catalyst*. The most direct framework for such a transformation is to transform a state $\rho \otimes c$ to the final state $\rho' \otimes c$, where the catalyst c can be reused for another transformation. We particularly call this *product catalyst*, as the transformation keeps the product structure in the final state. It turned out that the product catalyst is able to enhance the feasible transformation in various settings, such as entanglement transformation [61–63], quantum thermodynamics [12], speakable coherence [64], and magic state transformation in the context of fault-tolerant quantum computation [65].

If one focuses on the infinite reusability of the catalyst, the above transformation framework can be extended. After one round of the transformation, if it is promised that a fresh initial state (uncorrelated with the previous final state) is prepared for the next transformation, what only matters for the next transformation is the reduced state in the catalytic system — as long as the reduced state in the catalytic system remains intact from the initial form, it can still be reused indefinitely. This is called *correlated catalyst*, which we formally define as follows.

Definition S.22 (Correlated-catalytic transformation)**.** We say that state ρ in system S is convertible to state ρ' in system S' by a correlated-catalytic transformation if there exists a finite-dimensional catalytic system C with a state c and a free operation $\Lambda \in \mathbb{O} : S \otimes C \to S' \otimes C$ such that

$$\tau = \Lambda(\rho \otimes c), \quad \mathrm{Tr}_C[\tau] = \rho', \quad \mathrm{Tr}_{S'}[\tau] = c. \tag{32}$$

The correlated-catalytic transformation has been proven to be effective in several physical settings. A prominent example is quantum thermodynamics, where a correlated catalyst enables us to recover the second law of thermodynamics in a conventional form [66, 67]. It has also been discovered that similar observations can be made to the entanglement transformation [68, 69] and a general class of resource theories [45].

However, when it comes to the covariant operations, several previous works indicated that the use of a catalyst may not help much for the coherence transformation. First, it is known that a pure product catalyst does not enhance the transformation at all in the covariant operations [70, 71]. In addition, as for the correlated catalyst, it was shown that one could not create any coherence from an incoherent state with a correlated-catalytic covariant operation, i.e., if ρ is incoherent, then so is ρ' [72, 73] (see Theorem. S.25 for its precise statement).

III.B Arbitrary coherence transformation with correlated catalysts (Theorem 2)

Contrarily to the aforementioned implications, we here show that correlated catalysts can provide a dramatic operational power, and the limitations imposed on the correlated-catalytic covariant transformation shown in Refs. [72, 73] (coherence no-broadcasting theorem) is exceptional for the incoherent input states.

There exists a standard technique to reduce asymptotic transformations to correlated-catalytic transformations. This technique for non-exact asymptotic transformations was first discussed by Shiraishi and Sagawa [67]. This paper has commented its general applicability, and a number of papers have applied this technique to various resource theories including entropy conjecture [74], entanglement [47, 69], and teleportation [68]. The statement in a general form is presented by the authors (Proposition 4 in Ref. [45]).

For our purpose, we here express a slightly generalized version of the statement. The proof is the same as that in Ref. [45] and was employed in Ref. [47, 52].

Proposition S.23 (Slight generalization of Ref. [45])**.** *Consider a resource theory such that a set of free operations includes the relabeling of a classical register and the conditioning of free operations by a classical register. Then, if there is a free transformation $\Lambda : S^{\otimes n} \to S^{\otimes n}$ which maps $\rho^{\otimes n}$ to τ with $\sigma := \frac{1}{n}\sum_{i=1}^{n} \mathrm{Tr}_{\setminus i}[\tau]$, then there exists a correlated-catalytic free transformation mapping ρ to σ.*

We here assume, without loss of generality, that the initial state ρ and the final state σ are states in the same system S. This is because if the system S' for the final state was different from the input system S, one could consider an enlarged system $S \oplus S'$ as its input and output system and simply call it S. This modification is harmless since an embedding process is a covariant operation and we can make possible errors in the final state inside the S' part.

For completeness, we here present the construction of the desired correlated-catalytic transformations. Suppose that a covariant operation $\mathcal{E}$: $S^{\otimes n} \to S^{\otimes n}$ converts $\rho^{\otimes n}$ to τ with $\sigma := \frac{1}{n}\sum_{i=1}^{n} \mathrm{Tr}_{\setminus i}[\tau]$. Then, defining the reduced state of copies from the first copy to the i-th copy as

$$\tau_i := \mathrm{Tr}_{C_{i+1},\ldots,C_n}[\tau] \in S^{\otimes i}, \tag{33}$$

we construct the catalytic system $C = S^{\otimes(n-1)} \otimes R$ with state c as

$$c := \frac{1}{n}\sum_{k=1}^{n} \rho^{\otimes(k-1)} \otimes \tau_{n-k} \otimes |k\rangle\langle k|_R \,. \tag{34}$$

Here R is a classical register system spanned by $\{|k\rangle\}_{k=1}^{n}$. The initial state of the composite system is written as

$$\rho \otimes c = \frac{1}{n}\sum_{k=1}^{n} \rho^{\otimes k} \otimes \tau_{n-k} \otimes |k\rangle\langle k|_R \,. \tag{35}$$

For later discussion, we say that the first copy corresponds to system S, and the latter $k-1$ copies correspond to catalyst C.

Our free operation consists of two steps. In the first step, we apply Λ if the classical register is $|n\rangle\langle n|$ and leave the system otherwise. The resulting state is

$$\frac{1}{n}\left(\sum_{k=1}^{n-1}\rho^{\otimes k}\otimes\tau_{n-k}\otimes|k\rangle\langle k|_R+\tau\otimes|n\rangle\langle n|_R\right). \tag{36}$$

In the second step, we relabel the classical register as $k\to k+1$ for $k\neq n$ and $n\to 1$, which results in

$$\frac{1}{n}\sum_{k=1}^{n}\rho^{\otimes(k-1)}\otimes\tau_{n-k+1}\otimes|k\rangle\langle k|_R. \tag{37}$$

Finally, by regarding the first $n-1$ copies as catalyst C and the last copy as system S, the state of the catalyst returns to its own state:

$$\begin{aligned}\mathrm{Tr}_n\left[\frac{1}{n}\sum_{k=1}^{n}\rho^{\otimes(k-1)}\otimes\tau_{n-k+1}\otimes|k\rangle\langle k|_R\right]&=\frac{1}{n}\sum_{k=1}^{n}\rho^{\otimes(k-1)}\otimes\mathrm{Tr}_n[\tau_{n-k+1}]\otimes|k\rangle\langle k|_R\\&=\frac{1}{n}\sum_{k=1}^{n}\rho^{\otimes(k-1)}\otimes\tau_{n-k}\otimes|k\rangle\langle k|_R\\&=c,\end{aligned} \tag{38}$$

and the state of the system is σ:

$$\mathrm{Tr}_{\backslash n,R}\left[\frac{1}{n}\sum_{k=1}^{n}\rho^{\otimes(k-1)}\otimes\tau_{n-k+1}\otimes|k\rangle\langle k|_R\right]=\frac{1}{n}\sum_{k=1}^{n}\mathrm{Tr}_{\backslash n}[\tau_{n-k+1}]=\frac{1}{n}\sum_{k=1}^{n}\mathrm{Tr}_{\backslash k}[\tau]=\sigma. \tag{39}$$

Now we state our second main result on correlated-catalytic transformations, stating that any coherent state is convertible to any full-rank state with a correlated catalyst.

Theorem S.24 (Theorem 2 in the main text)**.** *Let ρ and ρ' be arbitrary states such that $\mathcal{C}(\rho')\subseteq\mathcal{C}(\rho)$. Then, for any accuracy $\delta>0$ and correlation strength $\varepsilon>0$, there exists a finite-dimensional catalyst c and a covariant operation Λ such that*

$$\Lambda(\rho\otimes c)=\tau,\quad \|\mathrm{Tr}_C[\tau]-\rho'\|_1<\delta,\quad \mathrm{Tr}_S[\tau]=c,\quad \|\tau-\rho'\otimes c\|_1<\varepsilon. \tag{40}$$

Moreover, if ρ' is full rank, the transformation can be made exact, i.e., $\mathrm{Tr}_C[\tau]=\rho'$.

Proof. If we do not require the smallness of correlation, this is a direct consequence of Theorem S.10 and Proposition S.23. Since one can let the system for the classical register come with the trivial Hamiltonian, in which case relabeling of the classical register is a covariant operation, Proposition S.23 can be applied to the case of covariant operations. Theorem S.10 then implies that the exact catalytic transformation is possible for a full-rank target state ρ'. Since the set of full-rank states is dense in the state space, for every non-full rank state ρ' and every $\delta>0$, there exists a full-rank state that is δ-close to ρ', which shows the statement in the theorem.

To make the correlation arbitrarily small, we employ the trick that has already been used in the proof of Theorem S.10. Let κ be a state such that (i) κ is ε' close to a pure state (i.e., there exists a pure state ψ such that $\|\kappa-\psi\|_1\leq\varepsilon'$), (ii) κ is convertible to ρ' by a covariant operation, (iii) $\mathcal{C}(\kappa)=\mathcal{C}(\rho')$. Applying Proposition S.23 to the marginal-asymptotic transformation from ρ to κ with transformation rate $R=1$ ensured by Theorem S.10, we obtain a correlated-catalytic transformation from ρ to κ. By taking sufficiently small ε' in the condition (i), the correlation between state κ and the catalyst is bounded from above by ε, which means that $\tau'=\Lambda(\rho\otimes c)$ satisfies $\|\tau'-\kappa\otimes c\|_1<\varepsilon$. Finally, applying a covariant operation that transforms κ to ρ', we arrive at the desired correlated-catalytic transformation with the correlation less than ε. Here, we used the fact that an operation acting on only the system (not on the catalyst) does not increase the correlation between the system and the catalyst. □

The anomalous power of correlated-catalytic transformations for a two-level system was first argued by Ref. [71], while as pointed by Ref. [45], there was a gap in their proof. The same paper [45] conjectured that the condition $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$ would be the necessary and sufficient condition for arbitrary amplification of coherence with a correlated catalyst. Theorem S.24 solves the sufficient part of this conjecture, together with the unproved claim in Ref. [71], in the affirmative. In the next section, we discuss the necessary part of this conjecture and present the significant step toward the full resolution of this problem.

III.C Mode no-broadcasting (Theorem 3)

We here investigate the limitation of correlated-catalytic transformations with covariant operations. We first describe an important result known as the coherence no-broadcasting theorem found by Lostaglio and Müller [72] and Marvian and Spekkens [73] independently.

Theorem S.25 (Coherence no-broadcasting [72, 73])**.** *Let ρ be an incoherent state on system S. Consider a correlated-catalytic transformation from ρ to ρ' by a covariant operation Λ on SC with catalyst C: $\mathrm{Tr}_S[\Lambda(\rho \otimes c)] = c$. Then, the final state of the system $\rho' = \mathrm{Tr}_C[\Lambda(\rho \otimes c)]$ is still an incoherent state.*

This theorem applies only when an initial state of the system is completely incoherent. However, it is highly plausible that even if an initial state has some coherence on some modes, this coherence provides no advantage to create coherence on a mode that is only irrationally related to theirs. This intuition is indeed true, and we can prove the following theorem, which is an extension of the above coherence no-broadcasting theorem to the level of each mode. To state our finding, we slightly extend the definition of a set of resonant coherent modes.

For a given set $\mathcal{S}$ of real numbers, we say that a real number x can be written by a *rational-linear combination* of the elements in $\mathcal{S}$ if there exists a set $\{a_j\}_j$ of rational numbers such that $x = \sum_j a_j s_j$ for elements $s_j \in \mathcal{S}$. We then define $\mathcal{C}'(\rho)$ as the set of all real numbers that can be written by a rational linear combination of the set $\mathcal{D}(\rho)$ of modes of asymmetry defined in Definition S.5.

Definition S.26 (Set of rational coherent modes)**.** We define a set of rational coherence modes denoted by $\mathcal{C}'(\rho)$ as rational-linear combinations of non-zero coherent modes of ρ:

$$\mathcal{C}'(\rho) := \left\{ x \;\middle|\; x = \sum_{\Delta \in \mathcal{D}(\rho)} a_\Delta \Delta : a_\Delta \in \mathbb{Q} \right\}. \tag{41}$$

We then show the following no-go theorem.

Theorem S.27 (Theorem 3 in the main text: Mode no-broadcasting (weak version))**.** *Consider a correlated-catalytic transformation from ρ to ρ' through a covariant operation Λ on SC with catalyst C: $\mathrm{Tr}_S[\Lambda(\rho \otimes c)] = c$. Then, the final state of the system $\rho' = \mathrm{Tr}_C[\Lambda(\rho \otimes c)]$ has no coherence on a mode that is only irrationally related to coherent modes of ρ, i.e., $\mathcal{C}'(\rho') \subseteq \mathcal{C}'(\rho)$.*

Namely, if $\mathcal{C}'(\rho') \nsubseteq \mathcal{C}'(\rho)$, then no correlated-catalytic transformation maps ρ to ρ'.

To show this, we first remark that the input system S and the output system S' of a covariant operation can generally be different. However, as pointed out in Ref. [75], one can model such a transformation with another covariant operation with the same input and output systems by considering an extended space $\mathcal{H}_{\mathrm{in}} \oplus \mathcal{H}_{\mathrm{out}}$, where $\mathcal{H}_{\mathrm{in,out}}$ are the Hilbert spaces underlying the systems S and S' respectively. Therefore, we consider S as the system on the space that is already extended in this way and focus on a covariant operation such that SC is its input and output space.

Our proof of Theorem S.27 is based on the proof by contradiction. We suppose contrarily that there exists a correlated-catalytic transformation from ρ to ρ' despite $\mathcal{C}'(\rho') \nsubseteq \mathcal{C}'(\rho)$. Our goal is to construct a protocol violating the coherence no-broadcasting theorem.

To deal with any system in a unified framework, we embed the main and the catalytic systems in a product of *ladder systems* whose energy levels form an infinite ladder.

Definition S.28 (Ladder system)**.** A ladder system with energy interval Δ denoted by $L(\Delta)$ has states labeled by two integers (n, a) with $n \in \mathbb{Z}$ and $a \in \mathbb{N}$. The state $|n, a\rangle$ is an energy eigenstate with energy $n\Delta$. The label a distinguishes degenerate energy eigenstates.

Suppose that the main system S and catalytic system C can be embedded into the collection of the ladder systems. We consider this extended ladder system as our system S and write it as $S = L_S(\Delta_1) \otimes L_S(\Delta_2) \otimes \cdots =: L_S(\boldsymbol{\Delta})$ with abbreviation $\boldsymbol{\Delta} = (\Delta_1, \Delta_2, \ldots)$. Similarly, we set the catalytic system as $C = L_C(\boldsymbol{\Delta})$. For later convenience, we write a subsystem whose energy interval is multiples of Δ_i as $X(\Delta_i) := L_S(\Delta_i) \otimes L_C(\Delta_i)$. We also denote by $X(\boldsymbol{\Delta}) := L_S(\boldsymbol{\Delta}) \otimes L_C(\boldsymbol{\Delta})$.

We remark that this embedding is always possible when all energies for S and C can be written as integer-linear combinations of $\boldsymbol{\Delta}$. For instance, suppose that an energy E for (either the main or catalytic) system is written as $E = \sum_j n_j \Delta_j$ for some integers $\{n_j\}_j$ and elements $\{\Delta_j\}_j$ of $\boldsymbol{\Delta}$. Then, the energy eigenstates $|E, \alpha\rangle$ (where α distinguishes the degeneracy) can be mapped as $|E, \alpha\rangle \to \otimes_j |n_j, \alpha\rangle_{\Delta_j}$ where $|n_j, \alpha\rangle_{\Delta_j}$ is an energy eigenstate in $L(\Delta_j)$. By construction, $\otimes_j |n_j, \alpha\rangle_{\Delta_j}$ is an energy eigenstate of the extended ladder system with energy $\sum_j n_j \Delta_j$.

In the following, we say that a set $\mathcal{S}$ is *rational-linearly independent* if, for all elements $x_i \in \mathcal{S}$, there is no set $\{a_j\}_j$ of rational numbers such that $x_i = \sum_{j \neq i} a_j x_j$. Our key observation is that if a covariant operation on $X(\boldsymbol{\Delta})$ exists and all $\boldsymbol{\Delta}$ is rational-linearly independent, then the value of $\boldsymbol{\Delta}$ is in fact irrelevant. To describe this, we denote by $\rho[\boldsymbol{\Delta}]$ a quantum state on $S = L_S(\boldsymbol{\Delta})$ whose density matrix is ρ. Similarly, we denote by $c[\boldsymbol{\Delta}]$ and $\tau[\boldsymbol{\Delta}]$ quantum states on $C = L_C(\boldsymbol{\Delta})$ and $SC = X(\boldsymbol{\Delta})$, respectively.

Lemma S.29. *Let $\Delta_1, \Delta_2, \ldots$ be energy intervals that are rational-linearly independent. Suppose that a state $\tau[\boldsymbol{\Delta}]$ on $SC = X(\boldsymbol{\Delta})$ is convertible to $\tau'[\boldsymbol{\Delta}]$ on SC by a covariant operation. Then, for any $\boldsymbol{\Delta}'$ state $\tau[\boldsymbol{\Delta}']$ on $X(\boldsymbol{\Delta}')$ is also convertible to $\tau'[\boldsymbol{\Delta}']$ on $X(\boldsymbol{\Delta}')$ by a covariant operation.*

Proof. By definition, a covariant operation Λ on SC can be expressed by using an auxiliary system $A = L_A(\boldsymbol{\Delta})$ with its incoherent state η as

$$\Lambda(\tau) = \mathrm{Tr}_{A'}[U(\tau \otimes \eta)U^\dagger], \tag{42}$$

where U is an energy-conserving unitary on SCA. Owing to the energy conservation and rational-linear independence of energy intervals, energy change in S or C occurs only in each $L_S(\Delta_i) \otimes L_C(\Delta_i) \otimes L_A(\Delta_i)$, since the energy change with a multiple of Δ_i cannot be compensated by any rational-linear combination of Δ_j's with $j \neq i$. In other words, using a set of operators $\{K_i^j\}_j$ acting on $L_S(\Delta_i) \otimes L_C(\Delta_i) \otimes L_A(\Delta_i)$, any energy-conserving unitary U can be expressed in the form of

$$U = \sum_j \bigotimes_i K_i^j \tag{43}$$

with

$$[K_i^j, H_{SCA}(\Delta_i)] = 0 \tag{44}$$

for any i and j. Here, $H_{SCA}(\Delta_i)$ is a Hamiltonian on $L_S(\Delta_i) \otimes L_C(\Delta_i) \otimes L_A(\Delta_i)$.

As seen from Eq. (43), U conserves energy regardless of the value of Δ_i's. This directly implies that the same U for $\boldsymbol{\Delta}'$ with the initial state of A as $\eta[\boldsymbol{\Delta}']$ serves as the desired covariant operation. □

We are now in a position to prove Theorem S.27.

Proof of Theorem S.27. Let ρ and ρ' be states on the system S, and suppose $\mathcal{C}'(\rho') \not\subseteq \mathcal{C}'(\rho)$. Then, there exists a mode in ρ' that cannot be written by a rational-linear combination of the modes in ρ. Let Δ_0 be the energy interval for such a mode. We show that one can construct a set $\boldsymbol{\Delta}$ of intervals that embeds SC and would lead to the contradiction with coherence no-broadcasting theorem.

To construct the desired $\boldsymbol{\Delta}$, we start with $\boldsymbol{\Delta} = \{\Delta_0\}$ and add elements to $\boldsymbol{\Delta}$ step by step. Let $\overline{\mathcal{D}}(\rho)$ be a set of modes where ρ does not have coherence, which is a complement of a set $\mathcal{D}(\rho)$ of coherent modes. Let $\Delta_i(\rho)$ be the ith element in $\mathcal{D}(\rho)$ (with an arbitrary order). We run the following procedure.

(i) If $\Delta_i(\rho)$ cannot be written by a rational-linear combination of the elements already in $\boldsymbol{\Delta}$, add $\Delta_i(\rho)$ to $\boldsymbol{\Delta}$. On the other hand, if $\Delta_i(\rho) = \sum_j \frac{m_j}{n_j} \Delta_j$ for $\Delta_j \in \boldsymbol{\Delta}$ and some integers m_j and n_j, we redefine all elements $\Delta \in \boldsymbol{\Delta}$ as $\Delta \to \Delta / \prod_j n_j$, while not adding $\Delta_i(\rho)$ to $\boldsymbol{\Delta}$. We sequentially apply this procedure for $i = 1, 2, \ldots |\mathcal{D}(\rho)|$.

(ii) We then apply the same procedure for all incoherent modes in $\overline{\mathcal{D}}(\rho)$. Namely, if an incoherent mode cannot be written as a rational-linear combination of the elements already in $\boldsymbol{\Delta}$, we add the incoherent mode into $\boldsymbol{\Delta}$. Otherwise, we redefine the elements in $\boldsymbol{\Delta}$ by dividing them by some integer.

(iii) We apply the same procedure for energy intervals of the catalytic system C.

By construction, the resulting set $\boldsymbol{\Delta}$ satisfies that (1) all energies in S and C can be written as an integer-linear combination of the elements in $\boldsymbol{\Delta}$, and thus SC can be embedded in $L_S(\boldsymbol{\Delta}) \otimes L_C(\boldsymbol{\Delta})$ (2) $\boldsymbol{\Delta}$ is rational-linearly independent (3) no combination of Δ_0 and other modes in $\boldsymbol{\Delta}$ results in a coherent mode of ρ. The condition (3) is confirmed as follows: If a combination of Δ_0 and other modes in $\boldsymbol{\Delta}$ results in a coherent mode of ρ, one could write Δ_0 as a rational-linear combination of modes in $\mathcal{D}(\rho)$, which would contradict the assumption that $\Delta_0 \notin \mathcal{C}'(\rho)$. A remarkable point of this construction lies in the fact that defining $\tilde{\boldsymbol{\Delta}} := \boldsymbol{\Delta} \setminus \{\Delta_0\}$ so that $\boldsymbol{\Delta} = \{\Delta_0, \tilde{\boldsymbol{\Delta}}\}$, any state ρ on $L_S(\Delta_0) \otimes L_S(\tilde{\boldsymbol{\Delta}})$ which is incoherent on $L_S(\Delta_0)$ (i.e., $\Delta_0 \notin \mathcal{C}'(\rho)$) can be expressed as $\rho = \sum_{i,\alpha,\alpha'} p_i |i,\alpha\rangle\langle i,\alpha'|_{\Delta_0} \otimes \sigma_i$, since the coefficient of $|i,\alpha\rangle\langle j,\alpha'|_{\Delta_0}$ terms with $i \neq j$ should be zero due to absence of coherence on $L_S(\Delta_0)$. Here, σ_i are states on $L_S(\tilde{\boldsymbol{\Delta}})$ and p_i are nonnegative coefficients satisfying $\sum_i p_i = 1$.

Then, suppose contrarily that a covariant operation Λ on SC converts $\Lambda(\rho \otimes c) = \tau$ with $\mathrm{Tr}_S[\tau] = c$ and $\mathrm{Tr}_C[\tau]$ has non-zero coherence for the mode with energy interval Δ_0. This implies that the reduced state of $\mathrm{Tr}_C[\tau]$ on $L_S(\Delta_0)$ has non-zero coherence. We aim to show that this contradicts the coherence no-broadcasting theorem.

To this end, we introduce a method of *complete degeneration*. By setting $\boldsymbol{\Delta}' = \{\Delta_0, \mathbf{0}\}$ in Lemma S.29, we have a correlated-catalytic covariant transformation from $\rho[\{\Delta_0, \mathbf{0}\}]$ to $\rho'[\{\Delta_0, \mathbf{0}\}]$ with a catalyst $c[\{\Delta_0, \mathbf{0}\}]$. Remarkably, the form $\rho = \sum_{i,\alpha,\alpha'} p_i |i,\alpha\rangle\langle i,\alpha'|_{\Delta_0} \otimes \sigma_i$ and the fact that $L_S(\mathbf{0})$ has a single energy with fully degenerate energy eigenstates—hence no coherence in any state—the initial state of the system $\rho[\{\Delta_0, \mathbf{0}\}]$ is an incoherent state. On the other hand, the final state of the system, $\rho'[\{\Delta_0, \mathbf{0}\}]$ has non-zero coherence in $L_S(\Delta_0)$. In summary, an incoherent state $\rho[\{\Delta_0, \mathbf{0}\}]$ is converted into a coherent state $\rho'[\{\Delta_0, \mathbf{0}\}]$ by a covariant operation with a correlated catalyst, which contradicts the coherence no-broadcasting theorem. □

Although we do not have a proof at present, we expect that the condition $\mathcal{C}'(\rho') \nsubseteq \mathcal{C}'(\rho)$ in our mode no–broadcasting theorem can be lifted to $\mathcal{C}(\rho') \nsubseteq \mathcal{C}(\rho)$.

Conjecture S.30 (Mode no-broadcasting (strong version))**.** Consider a correlated-catalytic transformation from ρ to ρ' by a covariant operation Λ with a catalyst c: $\mathrm{Tr}_S[\Lambda(\rho \otimes c)] = c$. Then, the final state of the system $\rho' = \mathrm{Tr}_C[\Lambda(\rho \otimes c)]$ satisfies $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$.

The nontrivial part of showing the conjecture is to rule out the possibility of creating non-zero coherence from the coherence on another mode that is rationally related. We leave a thorough investigation for future work.

Appendix IV: Asymptotic coherence manipulation with correlated catalyst

The power of correlated catalyst has mainly been considered in the context of enhancing single-shot transformations, i.e., whether a catalyst could perform the transformation from a single copy of ρ to a single copy of ρ' that is not realizable without the help of catalysts.

An interesting—yet still much unexplored—question is whether correlated catalysts could enhance the *asymptotic* transformation rate by using correlated catalysts alongside the asymptotic transformation. This question was recently raised and studied in the context of entanglement distillation [47, 76]. Here, let us formally introduce relevant quantities.

Definition S.31 (Asymptotic correlated-catalytic transformation rate)**.** Let ρ and ρ' be states on systems S and S'. We say that the rate r is achievable in asymptotic correlated-catalytic transformation if there is a series $\{c_n\}_n$ of finite-dimensional states in some systems $\{C_n\}_n$ and a series $\{\Lambda_n\}_n$ of free operations with $\Lambda_n : S^{\otimes n} \otimes C_n \to S'^{\otimes \lfloor rn \rfloor} \otimes C_n$ such that for any $\varepsilon > 0$ there exists sufficiently large N and for $n > N$

$$\|\mathrm{Tr}_{C_n} \Lambda_n(\rho^{\otimes n} \otimes c_n) - \rho'^{\otimes \lfloor rn \rfloor}\|_1 < \varepsilon, \quad \mathrm{Tr}_{\setminus C_n} \Lambda_n(\rho^{\otimes n} \otimes c_n) = c_n \tag{45}$$

is satisfied. The *asymptotic correlated-catalytic transformation rate* $R_{\mathrm{cc}}(\rho \to \rho')$ is the supremum over the achievable rates.

Analogously to the case of other asymptotic transformations, we can also introduce the asymptotic exact rate.

Definition S.32 (Asymptotic exact correlated-catalytic transformation rate)**.** Let ρ and ρ' be states on systems S and S'. We say that the rate r is achievable in asymptotic exact correlated-catalytic transformation if there is a series $\{c_n\}_n$ of finite-dimensional states in some systems $\{C_n\}_n$ and a series $\{\Lambda_n\}_n$ of free operations with $\Lambda_n : S^{\otimes n} \otimes C_n \to S'^{\otimes \lfloor rn \rfloor} \otimes C_n$ such that there exists sufficiently large N and for $n > N$

$$\begin{aligned} \mathrm{Tr}_{C_n} \Lambda_n(\rho^{\otimes n} \otimes c_n) &= \rho'^{\otimes \lfloor rn \rfloor} \\ \mathrm{Tr}_{\setminus C_n} \Lambda_n(\rho^{\otimes n} \otimes c_n) &= c_n. \end{aligned} \tag{46}$$

The *asymptotic exact correlated-catalytic transformation rate* $R^0_{\mathrm{cc}}(\rho \to \rho')$ is the supremum over the achievable rates.

In the context of entanglement distillation, it was found that correlated catalysts cannot enable non-zero distillation rate for positive-partial-transpose (PPT) entangled states [47] or cannot increase the distillation rates for distillable entangled states [76]. Ref. [76] also showed that, in the setting of speakable coherence [42]—related but different framework from that for superposition of energy eigenstates, which we discussed in this article—distillable coherence or coherence cost does not change with the help of correlated catalysts. It was then proposed as an open problem whether correlated catalysts could ever improve asymptotic transformation rates in any physical setting.

Let us now consider our setting of coherence distillation with covariant operations. Recall that the standard asymptotic rate $R(\rho \to \phi)$ of coherence distillation by covariant operations is zero for all full-rank state ρ and pure state ϕ (Theorem S.8). This shows that all full-rank states are "bound coherent" states analogous to bound entanglement in the resource theory of entanglement, and the corresponding question is whether correlated catalysts could improve this rate, i.e., whether it is possible to obtain $R_{\mathrm{cc}}(\rho \to \phi) > 0$. Our results answer this question in the most drastic way.

Corollary S.33. *Let ρ be a state in S and ρ' be a state in S' such that $\mathcal{C}(\rho') \subseteq \mathcal{C}(\rho)$. Then, $R_{\mathrm{cc}}(\rho \to \rho')$ diverges. Moreover, if ρ' is full rank, $R^0_{\mathrm{cc}}(\rho \to \rho')$ also diverges. In both cases, the correlation between the main and catalytic systems can be made arbitrarily small.*

Proof. This is a direct consequence of Theorem S.24 by taking $\rho^{\otimes n}$ as the initial state and $\rho'^{\otimes Rn}$ as the target state for an arbitrary R. The condition in Theorem S.24 is satisfied because $\mathcal{C}(\sigma^{\otimes m}) = \mathcal{C}(\sigma)$ for every state σ and integer m. □

Appendix V: Extension to general resource theories

V.A Asymptotic-marginal and correlated-catalytic free transformation

The arguments to prove Theorems S.10 and S.24 provide a systematic way of constructing asymptotic-marginal and correlated-catalytic transformations from a marginal catalytic transformation (recall its definition in Definition S.11). Notably, what we have employed is only the aforementioned general properties, and other specific properties of quantum coherence are not utilized. Therefore, these results can directly be extended to general resource theories. We omit the proofs because they are essentially the same as those for Theorems S.10 and S.24.

Theorem S.34. *Consider a resource theory with set $\mathbb{O}$ of free operations. Let ρ and ρ' be arbitrary states on S and S'. Suppose that ρ can be transformed to ρ' by marginal-catalytic free transformation, i.e., there exist catalytic systems $C_1, \ldots, C_N$ with state $c_1, \ldots, c_N$ and a free operation $\Lambda \in \mathbb{O} : S \otimes C_1 \otimes \cdots \otimes C_N \to S' \otimes C_1 \otimes \cdots \otimes C_N$ such that $\tau = \Lambda(\rho \otimes c_1 \otimes \cdots \otimes c_N)$ satisfies $\mathrm{Tr}_{C_1,\ldots,C_N}[\tau] = \rho'$ and $\mathrm{Tr}_{\setminus C_i}[\tau] = c_i$ for any $1 \le i \le N$. Suppose also that there exists a free operation $\mathcal{E} \in \mathbb{O} : S^{\otimes m} \to C_1 \otimes \cdots \otimes C_N$ such that $\mathcal{E}(\rho^{\otimes m}) = c_1 \otimes \cdots \otimes c_N$ for some integer m.*

Then, for any $\delta > 0$, there exist sufficiently large integers n and m with $\frac{m}{n} > 1 - \delta$ and a free operation $\mathcal{K} \in \mathbb{O}$ on $S^{\otimes n} \to S'^{\otimes m}$ such that

$$\mathrm{Tr}_{\setminus i}[\mathcal{K}(\rho^{\otimes n})] = \rho' \tag{47}$$

for any $1 \leq i \leq m$.

Theorem S.35. *Consider a resource theory whose free operations include the relabeling of classical registers and free operations conditioned by classical labels. Let ρ and ρ' be arbitrary states on S and S' such that ρ can be transformed by marginal-catalytic free transformation. Suppose also that there exists a free operation $\mathcal{E} : S^{\otimes m} \to C_1 \otimes \cdots \otimes C_N$ such that $\mathcal{E}(\rho^{\otimes m}) = c_1 \otimes \cdots \otimes c_N$ for some integer m. Then, there exists a finite-dimensional catalytic system C, its state c, and a free operation $\mathcal{K}$ on SC such that*

$$\mathcal{K}(\rho \otimes c) = \tau, \quad \mathrm{Tr}_C[\tau] = \rho', \quad \mathrm{Tr}_S[\tau] = c. \tag{48}$$

V.B Restrictions imposed by resource measures

Theorems S.9 and S.24 appear highly anomalous compared to results in other resource theories. One may wonder why such apparent amplification enabled by correlation is not seen in other resource theories. To elucidate the specialty of (unspeakable) quantum coherence, we see these phenomena from the viewpoint of resource measures.

To this end, we recall the limitations imposed on correlated and marginal catalytic transformations.

Proposition S.36 (Proposition 3 of Ref. [45])**.** *Let $\mathfrak{R}$ be a resource measure that is tensor-product additive and superadditive. Then, $\mathfrak{R}(\rho) \geq \mathfrak{R}(\rho')$ holds if ρ is convertible to ρ' by a correlated-catalytic or a marginal-catalytic free transformation.*

Here, a resource measure $\mathfrak{R}$ is tensor-product additive if $\mathfrak{R}(\rho\otimes\sigma) = \mathfrak{R}(\rho)+\mathfrak{R}(\sigma)$, and is superadditive if a state τ on a composite system AB satisfies $\mathfrak{R}(\tau) \geq \mathfrak{R}(\mathrm{Tr}_A[\tau]) + \mathfrak{R}(\mathrm{Tr}_B[\tau])$. This theorem directly implies that the existence of even a single resource measure satisfying the above conditions and the nontriviality, i.e., there exist two states ρ and ρ' such that $0 < \mathfrak{R}(\rho) < \mathfrak{R}(\rho')$, prohibits arbitrary state conversions by a correlated-catalytic free transformation, since conversion $\rho \to \rho'$ with $\mathfrak{R}(\rho) < \mathfrak{R}(\rho')$ is impossible.

The restriction on the asymptotic marginal transformation is obtained in a similar manner.

Proposition S.37. *Let $\mathfrak{R}$ be a resource measure that is tensor-product additive and superadditive. Then, $\mathfrak{R}(\rho) \geq \tilde{R}^0(\rho \to \rho')\,\mathfrak{R}(\rho')$ holds.*

Proof. By definition of the asymptotic exact marginal transformation rate, for every $\delta > 0$, there exists a sufficiently large n and a free operation $\Lambda \in \mathbb{O} : S \to S'^{\otimes \lfloor (\tilde{R}^0(\rho\to\rho') - \delta)n \rfloor}$ such that $\mathrm{Tr}_{\setminus i}\Lambda(\rho^{\otimes n}) = \rho'$ holds for all i. Using such n, we get

$$\begin{aligned}
\mathfrak{R}(\rho) &= \frac{1}{n}\mathfrak{R}(\rho^{\otimes n}) \\
&\geq \frac{1}{n}\mathfrak{R}(\Lambda(\rho^{\otimes n})) \\
&\geq \frac{\left\lfloor n\left(\tilde{R}^0(\rho\to\rho') - \delta\right)\right\rfloor}{n}\mathfrak{R}(\rho') \\
&\geq \frac{\left[n\left(\tilde{R}^0(\rho\to\rho') - \delta\right) - 1\right]}{n}\mathfrak{R}(\rho') \\
&= \left[\left(\tilde{R}^0(\rho\to\rho') - \delta\right) - 1/n\right]\mathfrak{R}(\rho')
\end{aligned} \tag{49}$$

where the first line is due to the tensor-product additivity of $\mathfrak{R}$, the second due to the monotonicity, and the third line due to the superadditivity of $\mathfrak{R}$. The statement follows by noting that $\delta > 0$ and $1/n$ can be made arbitrarily small by taking a sufficiently large n. □

Most resource theories including entanglement [77, 78], quantum thermodynamics [79], speakable coherence [42, 80] have such a nontrivial measure. (See, e.g., [29, 81] for a couple of exceptions.) In contrast, any nontrivial faithful measure in the resource theory of (unspeakable) coherence is shown not to be superadditive [73]. This distinguishes the resource theory of quantum coherence from other resource theories.

[1] E. Chitambar and G. Gour, *Quantum resource theories*, Rev. Mod. Phys. **91**, 025001 (2019).
[2] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, *Concentrating partial entanglement by local operations*, Phys. Rev. A **53**, 2046–2052 (1996).
[3] S. Bravyi and A. Kitaev, *Universal quantum computation with ideal Clifford gates and noisy ancillas*, Phys. Rev. A **71**, 022316 (2005).
[4] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, *Resource Theory of Quantum States Out of Thermal Equilibrium*, Phys. Rev. Lett. **111**, 250404 (2013).
[5] M. Horodecki and J. Oppenheim, *Fundamental limitations for quantum and nanoscale thermodynamics*, Nat. Commun. **4**, 2059 (2013).
[6] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Quantum entanglement*, Rev. Mod. Phys. **81**, 865–942 (2009).
[7] B. Yadin, F. C. Binder, J. Thompson, V. Narasimhachar, M. Gu, and M. S. Kim, *Operational Resource Theory of Continuous-Variable Nonclassicality*, Phys. Rev. X **8**, 041038 (2018).
[8] H. Kwon, K. C. Tan, T. Volkoff, and H. Jeong, *Nonclassicality as a Quantifiable Resource for Quantum Metrology*, Phys. Rev. Lett. **122**, 040503 (2019).
[9] M. G. Genoni and M. G. A. Paris, *Quantifying non-Gaussianity for quantum information*, Phys. Rev. A **82**, 052341 (2010).
[10] R. Takagi and Q. Zhuang, *Convex resource theory of non-Gaussianity*, Phys. Rev. A **97**, 062337 (2018).
[11] F. Albarelli, M. G. Genoni, M. G. A. Paris, and A. Ferraro, *Resource theory of quantum non-Gaussianity and Wigner negativity*, Phys. Rev. A **98**, 052350 (2018).
[12] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, *The second laws of quantum thermodynamics*, Proc. Nat. Acad. Sci. **112**, 3275 (2015).
[13] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, *The resource theory of stabilizer quantum computation*, New J. Phys. **16**, 013009 (2014).
[14] M. Howard and E. Campbell, *Application of a Resource Theory for Magic States to Fault-Tolerant Quantum Computing*, Phys. Rev. Lett. **118**, 090501 (2017).
[15] R. Takagi, B. Regula, K. Bu, Z.-W. Liu, and G. Adesso, *Operational Advantage of Quantum Resources in Subchannel Discrimination*, Phys. Rev. Lett. **122**, 140402 (2019).
[16] R. Takagi and B. Regula, *General Resource Theories in Quantum Mechanics and Beyond: Operational Characterization via Discrimination Tasks*, Phys. Rev. X **9**, 031053 (2019).
[17] R. Uola, T. Kraft, J. Shang, X.-D. Yu, and O. Gühne, *Quantifying Quantum Resources with Conic Programming*, Phys. Rev. Lett. **122**, 130404 (2019).
[18] B. Regula, L. Lami, G. Ferrari, and R. Takagi, *Operational Quantification of Continuous-Variable Quantum Resources*, Phys. Rev. Lett. **126**, 110403 (2021).
[19] A. Anshu, M.-H. Hsieh, and R. Jain, *Quantifying Resources in General Resource Theory with Catalysts*, Phys. Rev. Lett. **121**, 190504 (2018).
[20] B. Regula, *Convex Geometry of Quantum Resource Quantification*, J. Phys. A: Math. Theor. **51**, 045303 (2018).
[21] Z.-W. Liu and A. Winter, *Resource theories of quantum channels and the universal role of resource erasure*, (2019), arXiv:1904.04201.
[22] G. Gour and A. Winter, *How to Quantify a Dynamical Quantum Resource*, Phys. Rev. Lett. **123**, 150401 (2019).
[23] G. Gour and C. M. Scandolo, *Dynamical Resources*, (2020), arXiv:2101.01552.
[24] T. Gonda and R. W. Spekkens, *Monotones in General Resource Theories*, Compositionality **5**, 7 (2023).
[25] M. Horodecki and J. Oppenheim, *(Quantumness in the context of) Resource Theories*, Int. J. Mod. Phys. B **27**, 1345019 (2013).
[26] G. Gour, *Quantum Resource Theories in the Single-Shot Regime*, Phys. Rev. A **95**, 062314 (2017).
[27] Z.-W. Liu, K. Bu, and R. Takagi, *One-Shot Operational Quantum Resource Theory*, Phys. Rev. Lett. **123**, 020401 (2019).
[28] B. Regula, K. Bu, R. Takagi, and Z.-W. Liu, *Benchmarking one-shot distillation in general quantum resource theories*, Phys. Rev. A **101**, 062315 (2020).
[29] K. Kuroiwa and H. Yamasaki, *General Quantum Resource Theories: Distillation, Formation and Consistent Resource Measures*, Quantum **4**, 355 (2020).
[30] K. Fang and Z.-W. Liu, *No-Go Theorems for Quantum Resource Purification*, Phys. Rev. Lett. **125**,

060405 (2020).
[31] Y. Liu and X. Yuan, *Operational resource theory of quantum channels*, Phys. Rev. Research **2**, 012035 (2020).
[32] K. Fang and Z.-W. Liu, *No-Go Theorems for Quantum Resource Purification: New Approach and Channel Theory*, PRX Quantum **3**, 010337 (2022).
[33] B. Regula and R. Takagi, *One-Shot Manipulation of Dynamical Quantum Resources*, Phys. Rev. Lett. **127**, 060402 (2021).
[34] B. Regula and R. Takagi, *Fundamental Limitations on Distillation of Quantum Channel Resources*, Nat. Commun. **12**, 4411 (2021).
[35] B. Regula and L. Lami, *Functional analytic insights into irreversibility of quantum resources*, (2022), arXiv:2211.15678.
[36] B. Regula, *Tight constraints on probabilistic convertibility of quantum states*, Quantum **6**, 817 (2022).
[37] R. Takagi, B. Regula, and M. M. Wilde, *One-Shot Yield-Cost Relations in General Quantum Resource Theories*, PRX Quantum **3**, 010348 (2022).
[38] B. Regula, *Probabilistic Transformations of Quantum Resources*, Phys. Rev. Lett. **128**, 110505 (2022).
[39] M. Keyl and R. F. Werner, *Optimal cloning of pure states, testing single clones*, J. Math. Phys. **40**, 3283 (1999).
[40] I. Marvian, *Coherence distillation machines are impossible in quantum thermodynamics*, Nat. Commun. **11**, 25 (2020).
[41] G. Gour and R. W. Spekkens, *The resource theory of quantum reference frames: manipulations and monotones*, New J. Phys. **10**, 033023 (2008).
[42] T. Baumgratz, M. Cramer, and M. B. Plenio, *Quantifying Coherence*, Phys. Rev. Lett. **113**, 140401 (2014).
[43] I. Marvian and R. W. Spekkens, *How to quantify coherence: Distinguishing speakable and unspeakable notions*, Phys. Rev. A **94**, 052324 (2016).
[44] I. Marvian and R. W. Spekkens, *Modes of asymmetry: The application of harmonic analysis to symmetric quantum dynamics and quantum reference frames*, Phys. Rev. A **90**, 062110 (2014).
[45] R. Takagi and N. Shiraishi, *Correlation in Catalysts Enables Arbitrary Manipulation of Quantum Coherence*, Phys. Rev. Lett. **128**, 240501 (2022).
[46] G. Ferrari, L. Lami, T. Theurer, and M. B. Plenio, *Asymptotic State Transformations of Continuous Variable Resources*, Commun. Math. Phys. **398**, 291 (2023).
[47] R. Ganardi, T. Varun Kondra, and A. Streltsov, *Catalytic and asymptotic equivalence for quantum entanglement*, (2023), arXiv:2305.03488.
[48] K. Audenaert, M. B. Plenio, and J. Eisert, *Entanglement Cost under Positive-Partial-Transpose-Preserving Operations*, Phys. Rev. Lett. **90**, 027901 (2003).
[49] B. Regula, K. Fang, X. Wang, and M. Gu, *One-shot entanglement distillation beyond local operations and classical communication*, New J. Phys. **21**, 103017 (2019).
[50] X. Wang and M. M. Wilde, *Cost of Quantum Entanglement Simplified*, Phys. Rev. Lett. **125**, 040502 (2020).
[51] M. Lostaglio, M. P. Müller, and M. Pastena, *Stochastic Independence as a Resource in Small-Scale Thermodynamics*, Phys. Rev. Lett. **115**, 150402 (2015).
[52] H. Wilming, *Correlations in typicality and an affirmative solution to the exact catalytic entropy conjecture*, Quantum **6**, 858 (2022).
[53] J. I. Cirac, A. K. Ekert, and C. Macchiavello, *Optimal Purification of Single Qubits*, Phys. Rev. Lett. **82**, 4344–4347 (1999).
[54] I. Marvian, *Operational Interpretation of Quantum Fisher Information in Quantum Thermodynamics*, Phys. Rev. Lett. **129**, 190502 (2022).
[55] C. Zhang, B. Yadin, Z.-B. Hou, H. Cao, B.-H. Liu, Y.-F. Huang, R. Maity, V. Vedral, C.-F. Li, G.-C. Guo, and D. Girolami, *Detecting metrologically useful asymmetry and entanglement by a few local measurements*, Phys. Rev. A **96**, 042327 (2017).
[56] I. Marvian and R. W. Spekkens, *Extending Noether's theorem by quantifying the asymmetry of quantum states*, Nat. Commun. **5**, 3821 (2014).
[57] R. Takagi, *Skew informations from an operational view via resource theory of asymmetry*, Sci. Rep. **9**, 14562 (2019).
[58] F. Hansen, *Metric adjusted skew information*, Proc. Natl. Acad. Sci. **105**, 9909 (2008).
[59] B. Synak-Radtke and M. Horodecki, *On asymptotic continuity of functions of quantum states*, J. Phys. A: Math. Gen. **39**, L423 (2006).
[60] A. Coladangelo and D. Leung, *Additive entanglemement measures cannot be more than asymptotically continuous*, (2019), arXiv:1910.11354.
[61] D. Jonathan and M. B. Plenio, *Entanglement-Assisted Local Manipulation of Pure Quantum States*, Phys. Rev. Lett. **83**, 3566–3569 (1999).
[62] M. Klimesh, *Inequalities that Collectively Completely Characterize the Catalytic Majorization Relation*, (2007), arXiv:0709.3680.
[63] S. Turgut, *Catalytic transformations for bipartite pure states*, J. Phys. A: Math. Theor. **40**, 12185 (2007).

[64] K. Bu, U. Singh, and J. Wu, *Catalytic coherence transformations*, Phys. Rev. A **93**, 042326 (2016).
[65] E. T. Campbell, *Catalysis and activation of magic states in fault-tolerant architectures*, Phys. Rev. A **83**, 032317 (2011).
[66] M. P. Müller, *Correlating Thermal Machines and the Second Law at the Nanoscale*, Phys. Rev. X **8**, 041051 (2018).
[67] N. Shiraishi and T. Sagawa, *Quantum Thermodynamics of Correlated-Catalytic State Conversion at Small Scale*, Phys. Rev. Lett. **126**, 150502 (2021).
[68] P. Lipka-Bartosik and P. Skrzypczyk, *Catalytic Quantum Teleportation*, Phys. Rev. Lett. **127**, 080502 (2021).
[69] T. V. Kondra, C. Datta, and A. Streltsov, *Catalytic Transformations of Pure Entangled States*, Phys. Rev. Lett. **127**, 150503 (2021).
[70] I. Marvian and R. W. Spekkens, *The theory of manipulations of pure state asymmetry: I. Basic tools, equivalence classes and single copy transformations*, New J. Phys. **15**, 033001 (2013).
[71] F. Ding, X. Hu, and H. Fan, *Amplifying asymmetry with correlating catalysts*, Phys. Rev. A **103**, 022403 (2021).
[72] M. Lostaglio and M. P. Müller, *Coherence and Asymmetry Cannot be Broadcast*, Phys. Rev. Lett. **123**, 020403 (2019).
[73] I. Marvian and R. W. Spekkens, *No-Broadcasting Theorem for Quantum Asymmetry and Coherence and a Trade-off Relation for Approximate Broadcasting*, Phys. Rev. Lett. **123**, 020404 (2019).
[74] H. Wilming, *Entropy and Reversible Catalysis*, Phys. Rev. Lett. **127**, 260402 (2021).
[75] I. Marvian Mashhad, *Symmetry, asymmetry and quantum information*, Ph.D. thesis (2012).
[76] L. Lami, B. Regula, and A. Streltsov, *Catalysis cannot overcome bound entanglement*, (2023), arXiv:2305.03489.
[77] M. Christandl and A. Winter, *"Squashed entanglement": An additive entanglement measure*, J. Math. Phys. **45**, 829 (2004).
[78] R. Alicki and M. Fannes, *Continuity of quantum conditional information*, J. Phys. A: Math. Gen. **37**, L55 (2004).
[79] H. Wilming, R. Gallego, and J. Eisert, *Axiomatic Characterization of the Quantum Relative Entropy and Free Energy*, Entropy **19**, 241 (2017).
[80] Z. Xi, Y. Li, and H. Fan, *Quantum coherence and correlations in quantum system*, Sci. Rep. **5**, 10922 (2015).
[81] A. Hickey and G. Gour, *Quantifying the imaginarity of quantum mechanics*, J. Phys. A: Math. Theor. **51**, 414009 (2018).

Entanglement purification with virtual local operation and classical communication

Kaoru Yamamoto[1] * Yuichiro Matsuzaki[2] Yasunari Suzuki[1] Yuuki Tokunaga[1] †
Suguru Endo[1 3] ‡

[1] *NTT Computer and Data Science Laboratories, NTT Corporation, Musashino 180-8585, Japan*
[2] *Department of Electrical, Electronic, and Communication Engineering, Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan*
[3] *JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan*

Abstract. A promising approach for scalable quantum computing involves distributed quantum computation, where quantum processing units (QPUs) are remotely connected through entanglement. Since entanglement is susceptible to noise, entanglement purification is often used to mitigate the impact of noise. However, this process produces entanglement with limited fidelity when each QPU is noisy. Here, we propose a new entanglement purification protocol, leveraging local operation and classical communication (LOCC) with a combination of classical post-processing, which we refer to virtual LOCC since it generates the purified entanglement in the expectation value. Owing to virtual LOCC, our demonstration finds that our protocol can break the fidelity limit of conventional purification protocols. Our results show the potential of virtual LOCC and pave the way for further scalability of quantum computer.

Keywords: Quantum error mitigation, Entanglement purification, Quantum computation

1 Introduction

Quantum computers hold the promise of outperforming classical computers in computational power, which necessitates a large number of physical qubits for fault-tolerant computation. Although current technological advancements have enabled an increase in the number of qubits on a single quantum processing unit (QPU), further scalability is crucial for achieving quantum supremacy. A promising approach to attaining this scalability is a modular architecture, where multiple QPUs are remotely connected via quantum links. Typically, this connection is realized through long-range entanglement [1, 2]. However, the current technology can generate noisy long-range entanglement insufficient for reliable quantum computations. Meanwhile, an alternative method involving classical links called circuit knitting has been actively explored to simulate quantum links [3, 4, 5]. Nonetheless, this approach requires an excessive number of additional circuit runs, known as the sampling cost, rendering it impractical for further scalability.

A feasible solution is entanglement purification, which generates high-fidelity entanglement by noisy entanglement with local operations and classical communication (LOCC) (Fig. 1(a)). The literature finds protocols that achieve perfect fidelity asymptotically as long as LOCC is noiseless [6, 7]. However, current QPUs have limited memory and are noisy, degrading the performance of entanglement purification. Although practical considerations of such hardware limitations and imperfections have investigated efficient protocols [8, 9, 10, 11, 12, 13], the fidelity is constrained by local noise in a QPU [8, 11]: For example, a 1% two-qubit gate error rate, which is achieved in current technology [14], enables a maximum fidelity of 99.5%, which is not enough for practical large-scale quantum computation on surface codes [15].

To overcome this challenge, we leverage the concept of *virtual* operations, a notion developed within the field of quantum error mitigation (QEM). QEM is initially proposed for noisy intermediate-scale quantum computers, which lack sufficient qubits for full quantum error correction [16, 17], to mitigate noise-induced biases in expectation values by post-processing outputs from a large number of noisy circuit runs. Since these operations used in QEM influence the expectation values rather than the quantum state itself, it is called *virtual* operations. Here, we introduce an entanglement distillation protocol that utilizes *virtual* operations confined LOCC, which we term *virtual* LOCC (vLOCC) (Fig. 1(b)). Our proposed protocol with vLOCC generates a purified Bell state from noisy Bell states in the expectation value. Remarkably, our demonstration shows that this protocol can surpass the fidelity limit of conventional protocols under noisy LOCC: a 1% two-qubit gate error rate enables a purified Bell state of 99.9% fidelity, only with single round purification from noisy Bell states with 90% fidelity. While our approach is used only for calculating expectation value at the cost of sampling shots, it is significantly more efficient than circuit knitting techniques. Our results show the potential of virtual LOCC and pave the way for further scalability by reducing the hardware requirement.

2 Construction of our protocol

To construct a protocol using vLOCC with high error tolerance, we can leverage recurrence protocols since these protocols have inherently higher error tolerance compared to other purification methods. Its high error tolerance may come from that they leverage quantum error detections by noisy parity measurements with noisy

*kaoru.yamamoto@ntt.com
†yuuki.tokunaga@ntt.com
‡suguru.endou@ntt.com

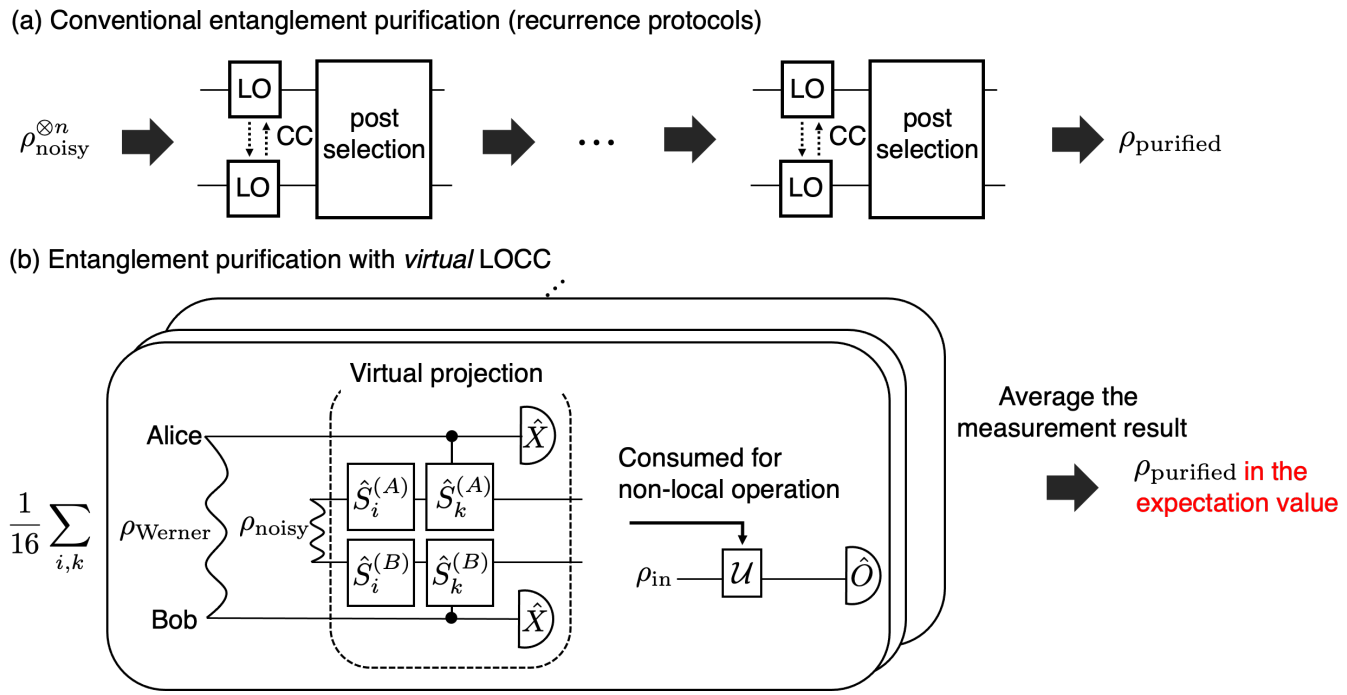


Figure 1: (a) Schematic illustration of conventional entanglement purification. This illustration mainly considers recurrence protocols, which purify noisy entanglement states with local operations and classical communication (LOCC). (b) Entanglement purification with virtual LOCC (vLOCC), where the LOCC is combined with post-processing of the measurement results. vLOCC purifies noisy entanglements in the expectation value of the measurement outcomes.

entanglement [7, 11]. In other words, they purify the Bell states by projecting noisy ones,

$$\frac{\hat{P}\rho_{\text{noisy}}\hat{P}}{\text{Tr}[\hat{P}\rho_{\text{noisy}}\hat{P}]} = \rho_{\text{Bell}} = |\Psi_{\text{Bell}}\rangle\langle\Psi_{\text{Bell}}|, \tag{1}$$

where $|\Psi_{\text{Bell}}\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and $\hat{P}$ is the projector onto $|\Psi_{\text{Bell}}\rangle$ described as $\hat{P} = \sum_{\hat{S}_i\in\mathcal{S}}\hat{S}_i/4$ with $\mathcal{S} = \{\hat{I}\hat{I}, \hat{X}\hat{X}, \hat{Z}\hat{Z}, -\hat{Y}\hat{Y}\}$ being the stabilizer operators of $|\Psi_{\text{Bell}}\rangle$, respectively.

Our protocol implement this projection *virtually* with the help of a Werner state [18] with error rate ϵ as an ancilla, $\rho_{\text{Werner}} = (1-3\epsilon/4)\rho_{\text{Bell}} + \epsilon\hat{I}/4$, as shown in the quantum circuit in Fig. 1(b), which can be prepared from any two-qubit state by implementing a suitable twirling operation, without changing its fidelity. [19], where $\hat{S}_{i(j,k)} \equiv \hat{S}^{(A)}_{i(j,k)}\hat{S}^{(B)}_{i(j,k)}$ with $\hat{S}_i\hat{S}_k = \hat{S}_j$ are uniformly sampled from the stabilizers $\mathbb{S} = \{\hat{I}\hat{I}, \hat{X}\hat{X}, \hat{Z}\hat{Z}, -\hat{Y}\hat{Y}\}$. In a practical quantum computation, Bell state is consumed for non-local operation $\mathcal{U}$, and we would like to obtain the expectation value of an operator $\hat{O}$ for the state $\mathcal{U}(\rho_{\text{in}})$. Fig. 1(b) provides the desired expectation value as follows. The $\hat{X}\hat{X}$ measurement on ρ_{Werner} and $\hat{O}$ measurement in Fig. 1(b) provides $1/2(\text{Tr}[\mathcal{U}(\hat{S}_i\rho_{\text{noisy}}\hat{S}_j \otimes \rho_{\text{in}})\hat{O}] + [i \leftrightarrow j])$, where $\hat{S}_i, \hat{S}_j \in \mathcal{S}$. Uniformly sampling i, j and averaging this expectation value provides $\text{Tr}[\mathcal{U}(\hat{P}\rho_{\text{noisy}}\hat{P}\otimes\rho_{\text{in}})\hat{O}]$ since $\hat{P} = \sum_{\hat{S}_i\in\mathcal{S}}\hat{S}_i/4$. Dividing the results of simultaneous measurement of $\hat{O}$ and $\hat{I}$ and $\mathcal{U}$ provides the desired expectation value,

$$\frac{\text{Tr}[\hat{O}\mathcal{U}(\rho_{\text{in}}\otimes(\hat{P}\rho_{\text{noisy}}\hat{P})]}{\text{Tr}[\hat{P}\rho_{\text{noisy}}\hat{P}]} = \text{Tr}[\hat{O}\mathcal{U}(\rho_{\text{in}}\otimes\rho_{\text{Bell}})]. \tag{2}$$

As with other QEM methods, our protocol incurs additional sampling shots $\mathcal{O}(\gamma^2 N)$ with $\gamma = \left(1-\frac{4}{3}\epsilon\right)^{-1}(1-F_{\text{noisy}})^{-1}$ being the sampling-cost factor [20, 21], where $F_{\text{noisy}} = \text{Tr}[\hat{P}\rho_{\text{noisy}}]$ is the fidelity of ρ_{noisy}.

3 Comparison with conventional protocols

Here, we compare the performance of our protocol with that of conventional entanglement purification protocols. The performance of entanglement purification is usually evaluated using the quantity called yield, which is defined as the ratio of the number of output purified Bell states per input noisy Bell state. Since we have extended entanglement purification protocols with vLOCC, we should define the yield for vLOCC. We require additional input noisy Bell states $\mathcal{O}[\gamma(F_{\text{noisy}})]^{2n_{\text{purified}}}$ for vLOCC to generate n_{purified} purified Bell states because we require $\mathcal{O}[\gamma(F_{\text{noisy}})]^{2n_{\text{purified}}}$ more sampling shots to achieve the same accuracy. This leads to the following definition of the yield for protocols with vLOCC:

$$Y_{\text{vLOCC}}(n_{\text{purified}}, F_{\text{noisy}}) = \frac{1}{K}\prod_{i=1}^{n_{\text{purified}}}\frac{1}{[\gamma_i(F_{\text{noisy}})]^2}, \tag{3}$$

where F_{purified} and F_{noisy} are the fidelity of the purified and noisy Bell states, respectively, K is the number of

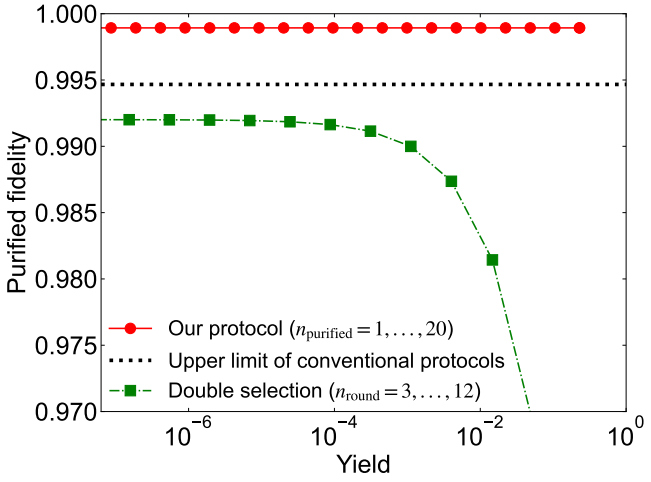


Figure 2: Purified fidelity with corresponding yield including local one-qubit, two-qubit and measurement noise. While the yield of our protocol decreases with increasing the number of distilled Bell states, n_{purified}, our protocol always generates much higher fidelity than that of the double selection protocol and brakes the conventional limit.

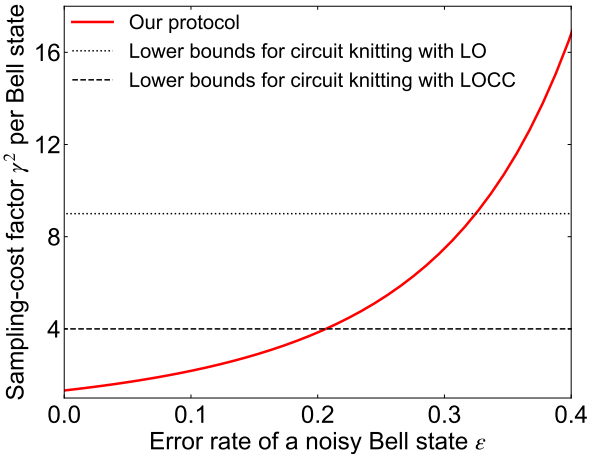


Figure 3: Sampling-cost factor γ^2 as a function of the error rate ϵ of the input noisy Bell state for our protocol, where the two horizontal lines represent the lower bounds of γ^2 for circuit knitting with LO and LOCC, respectively [3].

consumed Bell states (e.g. $K = 2$ for our protocol), and $\gamma_i(F_{\text{noisy}})$ is the sampling-cost factor for ith purified Bell state.

To compare our protocol with conventional entanglement purification protocols, we conducted numerical simulations using Qulacs [22] with introducing local noise as one-qubit and two-qubit depolarizing noise with error rates of 0.001 and 0.01, respectively, after each gate in the distillation circuit, as well as measurement noise with an error rate of 0.03 [8]. These error rates are consistent in the current superconducting quantum computer [14]. We set the initial state to be the Werner state $\rho_{\text{noisy}} = \rho_{\text{Werner}}$ with an initial fidelity of $F_{\text{noisy}} = 0.9$ for both our virtual and conventional distillation protocols. Note that the fidelity for our protocol is calculated $\text{Tr}[\rho_{\text{Bell}}\hat{P}\rho\hat{P}]/\text{Tr}[\hat{P}\rho]$ as the usual fidelity.

The purified fidelity and the corresponding yield are shown in Fig. 2. As shown, the purified fidelity of our protocol (the red circles) is much higher than that achieved by the conventional double-selection (the green dot-dashed curve with squares) protocols as well as the upper fidelity limit of the conventional protocols [8, 11] shown in black dotted line. Surprisingly, our protocol achieves a fidelity of 99.9%, while the maximum fidelity for many rounds of the double-selection protocol is only 99.2%. The reason for the high fidelity of our protocol is the error robustness of the ancilla. The effect of readout error is canceled by the division in our protocol. The most of the two-qubit errors are also canceled except for $\hat{Z}$ on the ancilla and $\hat{I}$ on the target state in the summation for calculating the trace, which contributes $p_2\epsilon$. Thus, the dominant contributions to the fidelity of our protocol are the one-qubit error p_1 and the term $p_2\epsilon$, leading to a much higher fidelity compared to conventional protocols. Although the yield for our protocol exponentially decreases with increasing n, the yield might still be acceptable for a few tens of n.

Figure 3 compares the sampling-cost factor between our protocol and the lower bounds of circuit knitting [3] and shows that our protocol can break the lower bounds of circuit knitting [3]. Given that a 10% error in Bell states can be generated using current technologies, our protocol can significantly reduce the sampling cost compared to circuit knitting. Since circuit knitting or cutting is regarded as a key technology for scaling quantum computers [1, 5], our protocol can serve as a more efficient alternative for simulation.

4 Conclusion and outlook

We have proposed a new entanglement purification protocol utilizing vLOCC and have demonstrated its higher error tolerance for noise in purification circuit than conventional protocols. In particular, our protocol can break the fidelity limit of the conventional purification protocols due to the classical post-processing in vLOCC. An interesting future direction of our work includes searching for more efficient protocols with vLOCC, extending it for multiple entangled states such as the n-qubit GHZ state [23] and the linear cluster state [24], and integrating our protocol with conventional entanglement purification protocols. Since vLOCC allows for more freedom, it would be intriguing to explore its limitations and possibilities mathematically.

Acknowledgments.— This work was supported by JST [Moonshot R&D] Grant Nos. JPMJMS2061 and JPMJMS226C; JST, PRESTO, Grant No. JPMJPR2114, Japan; MEXT Q-LEAP, Grant Nos. JPMXS0120319794 and JPMXS0118068682; JST CREST Grant No. JPMJCR23I4; JSPS KAKENHI, Grant No. 23H04390.

References

[1] Sergey Bravyi, Oliver Dial, Jay M. Gambetta, Darío Gil, and Zaira Nazario. The future of quantum computing with superconducting qubits. Journal of Applied Physics, 132(16):160902, 10 2022.

[2] James Ang, Gabriella Carini, Yanzhu Chen, Isaac Chuang, Michael Austin DeMarco, Sophia E Economou, Alec Eickbusch, Andrei Faraon, Kai-Mei Fu, Steven M Girvin, et al. Architectures for multinode superconducting quantum computers. arXiv:2212.06167, 2022.

[3] Christophe Piveteau and David Sutter. Circuit knitting with classical communication. IEEE Transactions on Information Theory, pages 1–1, 2023.

[4] Elisa Bäumer, Vinay Tripathi, Derek S Wang, Patrick Rall, Edward H Chen, Swarnadeep Majumder, Alireza Seif, and Zlatko K Minev. Efficient long-range entanglement using dynamic circuits. arXiv:2308.13065, 2023.

[5] Almudena Carrera Vazquez, Caroline Tornow, Diego Riste, Stefan Woerner, Maika Takita, and Daniel J Egger. Scaling quantum computing with dynamic circuits. arXiv:2402.17833, 2024.

[6] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. Phys. Rev. Lett., 76:722–725, Jan 1996.

[7] W Dür and H J Briegel. Entanglement purification and quantum error correction. Rep. Prog. Phys., 70(8):1381, jul 2007.

[8] Keisuke Fujii and Katsuji Yamamoto. Entanglement purification with double selection. Phys. Rev. A, 80:042308, Oct 2009.

[9] Naomi H. Nickerson, Joseph F. Fitzsimons, and Simon C. Benjamin. Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links. Phys. Rev. X, 4:041041, Dec 2014.

[10] Ramil Nigmatullin, Christopher J Ballance, Niel de Beaudrap, and Simon C Benjamin. Minimally complex ion traps as modules for quantum communication and computing. New J. Phys., 18(10):103028, oct 2016.

[11] Stefan Krastanov, Victor V. Albert, and Liang Jiang. Optimized Entanglement Purification. Quantum, 3:123, February 2019.

[12] Stefan Krastanov, Alexander Sanchez de la Cerda, and Prineha Narang. Heterogeneous multipartite entanglement purification for size-constrained quantum devices. Phys. Rev. Res., 3:033164, Aug 2021.

[13] Kenneth Goodenough, Sebastian De Bone, Vaishnavi Addala, Stefan Krastanov, Sarah Jansen, Dion Gijswijt, and David Elkouss. Near-term n to k distillation protocols using graph codes. IEEE Journal on Selected Areas in Communications, pages 1–1, 2024.

[14] J. C. Hoke, M. Ippoliti, E. Rosenberg, D. Abanin, R. Acharya, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, J. C. Bardin, A. Bengtsson, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, Z. Chen, B. Chiaro, D. Chik, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, A. G. Dau, D. M. Debroy, A. Del Toro Barba, S. Demura, A. Di Paolo, I. K. Drozdov, A. Dunsworth, D. Eppens, C. Erickson, E. Farhi, R. Fatemi, V. S. Ferreira, L. F. Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, R. Gosula, J. A. Gross, S. Habegger, M. C. Hamilton, M. Hansen, M. P. Harrigan, S. D. Harrington, P. Heu, M. R. Hoffmann, S. Hong, T. Huang, A. Huff, W. J. Huggins, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, K. Kechedzhi, T. Khattar, M. Khezri, M. Kieferová, S. Kim, A. Kitaev, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K. M. Lau, L. Laws, J. Lee, K. W. Lee, Y. D. Lensky, B. J. Lester, A. T. Lill, W. Liu, A. Locharla, O. Martin, J. R. McClean, M. McEwen, K. C. Miao, A. Mieszala, S. Montazeri, A. Morvan, R. Movassagh, W. Mruczkiewicz, M. Neeley, C. Neill, A. Nersisyan, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, T. E. O'Brien, S. Omonije, A. Opremcak, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, C. Rocque, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, N. Shutty, V. Shvarts, J. Skruzny, W. C. Smith, R. Somma, G. Sterling, D. Strain, M. Szalay, A. Torres, G. Vidal, B. Villalonga, C. V. Heidweiller, T. White, B. W. K. Woo, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, N. Zobrist, H. Neven, R. Babbush, D. Bacon, S. Boixo, J. Hilton, E. Lucero, A. Megrant, J. Kelly, Y. Chen, V. Smelyanskiy, X. Mi, V. Khemani, P. Roushan, Google Quantum AI, and Collaborators. Measurement-induced entanglement and teleportation on a noisy quantum processor. Nature, 622(7983):481–486, 2023.

[15] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. Phys. Rev. A, 86:032324, Sep 2012.

[16] Suguru Endo, Zhenyu Cai, Simon C. Benjamin, and Xiao Yuan. Hybrid quantum-classical algorithms and quantum error mitigation. J. Phys. Soc. Jpn., 90(3):032001, 2021.

[17] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. Quantum error mitigation. Rev. Mod. Phys., 95:045005, Dec 2023.

[18] Reinhard F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. Phys. Rev. A, 40:4277–4281, Oct 1989.

[19] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. Mixed-state entanglement and quantum error correction. Phys. Rev. A, 54:3824–3851, Nov 1996.

[20] Suguru Endo, Yasunari Suzuki, Kento Tsubouchi, Rui Asaoka, Kaoru Yamamoto, Yuichiro Matsuzaki, and Yuuki Tokunaga. Quantum error mitigation for rotation symmetric bosonic codes with symmetry expansion. arXiv:2211.06164, 2022.

[21] Kento Tsubouchi, Yasunari Suzuki, Yuuki Tokunaga, Nobuyuki Yoshioka, and Suguru Endo. Virtual quantum error detection. Phys. Rev. A, 108:042426, Oct 2023.

[22] Yasunari Suzuki, Yoshiaki Kawase, Yuya Masumura, Yuria Hiraga, Masahiro Nakadai, Jiabao Chen, Ken M. Nakanishi, Kosuke Mitarai, Ryosuke Imai, Shiro Tamiya, Takahiro Yamamoto, Tennin Yan, Toru Kawakubo, Yuya O. Nakagawa, Yohei Ibe, Youyuan Zhang, Hirotsugu Yamashita, Hikaru Yoshimura, Akihiro Hayashi, and Keisuke Fujii. Qulacs: a fast and versatile quantum circuit simulator for research purpose. Quantum, 5:559, October 2021.

[23] Géza Tóth and Otfried Gühne. Entanglement detection in the stabilizer formalism. Phys. Rev. A, 72:022340, Aug 2005.

[24] Chong Ying, Bin Cheng, Youwei Zhao, He-Liang Huang, Yu-Ning Zhang, Ming Gong, Yulin Wu, Shiyu Wang, Futian Liang, Jin Lin, et al. Experimental simulation of larger quantum circuits with fewer superconducting qubits. arXiv:2207.14142, 2022.

Simultaneous Measurement of Multiple Incompatible Observables and Tradeoff in Multiparameter Quantum Estimation

Hongzhen Chen[1] * Lingna Wang[2] Haidong Yuan[2] †

[1] *Institute of Quantum Precision Measurement, State Key Laboratory of Radio Frequency Heterogeneous Integration, College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China*

[2] *Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China*

Abstract. How well can multiple incompatible observables be implemented via a single measurement? This is a fundamental problem with wide implications in quantum information. While prior research substantially focused on two observables, our framework extends to any finite number, providing novel analytical error bounds of the implementations. Additionally, we introduce a stringent bound utilizing semi-definite programming that, in the context of two observables, generates an analytical bound tighter than any previously known bounds. These bounds have significant applications in assessing the trade-off among the precisions in multi-parameter quantum estimation. Experimental validation with a superconducting quantum processor confirms our theoretical results.

Keywords: Uncertainty Relations, Non-commutative Observables, Multi-parameter Quantum Metrology

1 Introduction

Quantum mechanics, distinct for its noncommutativity, poses challenges in measuring noncommuting observables simultaneously [1–3], necessitating approximation and inherently introducing trade-offs in measurement accuracy. This issue is underscored by the uncertainty principle and extends to quantum systems' preparation and measurement uncertainty relations [4–18]. Our paper focuses on state-dependent measurement uncertainty relations [4–15], particularly relevant to multi-parameter quantum estimation where simultaneous measurement of multiple observables is critical, as seen in applications like vector magnetometry and quantum imaging. We introduce methods that provide analytical and numerically stringent error-tradeoff relations for an arbitrary number of observables, offering insights into the calibration of the precision in multiparameter quantum metrology. These methods are empirically validated using a superconducting quantum processor.

2 Results

2.1 Analytical error-tradeoff relation

We commence by deriving an analytical measurement uncertainty relation for a general set of n observables. The objective is to use a single Positive Operator-Valued Measurement (POVM), denoted $\mathcal{M} = \{M_m\}$, to approximate the given n observables $X_1, X_2, \ldots, X_n$ when applied to a quantum state ρ and to determine relations that set limits on the minimum cumulative weighted approximation error. According to Neumark's dilation theorem [19], the POVM, $\mathcal{M} = \{M_m = K_m^\dagger K_m\}$, is equivalent to a projective measurement on $\rho \otimes \sigma$ in an extended Hilbert space $\mathcal{H}_\mathcal{S} \otimes \mathcal{H}_\mathcal{A}$, here $\sigma = |\xi_0\rangle\langle\xi_0|$ is an ancillary state such that $(I \otimes \langle\xi_0|)U^\dagger(I \otimes |\xi_m\rangle\langle\xi_m|)U(I \otimes |\xi_0\rangle) = M_m$, where $\{|\xi_m\rangle\}$ is an orthonormal basis for the ancillary system, U is a unitary operator on the extended space such that for any $|\psi\rangle$, $U|\psi\rangle \otimes |\xi_0\rangle = \sum_m K_m|\psi\rangle|\xi_m\rangle$. Denote $V_m = U^\dagger(I \otimes |\xi_m\rangle\langle\xi_m|)U$, we then have $\mathrm{Tr}[(\rho \otimes |\xi_0\rangle\langle\xi_0|)V_m] = \mathrm{Tr}(\rho M_m)$. From the measurement, we can construct a set of commuting observables, $\{F_j = \sum_m f_j(m)V_m | 1 \le j \le n\}$, to approximate $\{X_j \otimes I\}$ in the extended Hilbert space (see Fig.1). The mean squared error of the approximation on the state is given by [6, 7, 12]

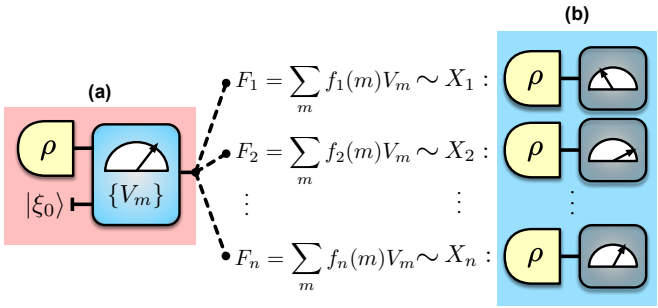


Figure 1: Simultaneous measurement of multiple observables $\{X_1, X_2, ..., X_n\}$ via a single measurement.

$$\epsilon_j^2 = \mathrm{Tr}\left[(F_j - X_j \otimes I)^2 (\rho \otimes \sigma)\right]. \quad (1)$$

In the case of two observables, Ozawa and Branciard obtained a series of error-tradeoff relations [6, 7, 12, 13], which is tight for pure states. While for mixed states, none of them is tight [15], and the geometrical method employed to derive these relations are not readily extendable to scenarios involving more than two observables. For general n observables, the error-tradeoff relation is little understood.

Here we present an approach that can lead to analytical tradeoff relations for an arbitrary number of observables.

*hzchen@szu.edu.cn

†hdyuan@mae.cuhk.edu.hk

Let

$$\mathcal{A}_u = \begin{pmatrix} \langle u|\sqrt{\rho\otimes\sigma}E_1 \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}E_n \\ \langle u|\sqrt{\rho\otimes\sigma}(X_1\otimes I) \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}(X_n\otimes I) \end{pmatrix} \begin{pmatrix} \langle u|\sqrt{\rho\otimes\sigma}E_1 \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}E_n \\ \langle u|\sqrt{\rho\otimes\sigma}(X_1\otimes I) \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}(X_n\otimes I) \end{pmatrix}^\dagger = \begin{pmatrix} Q_u & R_u \\ R_u^\dagger & S_u \end{pmatrix} \geq 0, \tag{2}$$

here $E_j = F_j - X_j \otimes I$ is the error operator, $|u\rangle$ is any vector, Q_u, R_u, S_u are $n \times n$ submatrices of $\mathcal{A}_u$. Given any set of states $\{|u_q\rangle\}$ such that $\sum_q |u_q\rangle\langle u_q| = I$, we can derive a corresponding set of matrices $\{\mathcal{A}_{u_q}\}$. We then construct a matrix $\tilde{\mathcal{A}}$ as the sum of these matrices with each $\tilde{\mathcal{A}}_{u_q}$ being either $\mathcal{A}_{u_q}$ or its transpose, $\mathcal{A}_{u_q}^T$. Since both $\mathcal{A}_{u_q}$ and $\mathcal{A}_{u_q}^T$ are positive semi-definite, it follows that:

$$\tilde{\mathcal{A}} = \sum_q \tilde{\mathcal{A}}_{u_q} = \begin{pmatrix} \tilde{Q} & \tilde{R} \\ \tilde{R}^\dagger & \tilde{S} \end{pmatrix} \geq 0, \tag{3}$$

where the components are defined as $\tilde{Q} = \sum_q \tilde{Q}_{u_q}$, $\tilde{R} = \sum_q \tilde{R}_{u_q}$, and $\tilde{S} = \sum_q \tilde{S}_{u_q}$, with every $(\tilde{Q}_{u_q}, \tilde{R}_{u_q}, \tilde{S}_{u_q})$ being either $(Q_{u_q}, R_{u_q}, S_{u_q})$ or their complex conjugate $(\bar{Q}_{u_q}, \bar{R}_{u_q}, \bar{S}_{u_q})$, here $\bar{M} = M_{\mathrm{Re}} - iM_{\mathrm{Im}}$ and for Hermitian matrix $\bar{M} = M^T$. We then have $(\tilde{Q}_{\mathrm{Re}})_{jk} = \frac{1}{2}\mathrm{Tr}\left[(\rho\otimes\sigma)\{E_j, E_k\}\right]$, $(\tilde{S}_{\mathrm{Re}})_{jk} = \frac{1}{2}\mathrm{Tr}(\rho\{X_j, X_k\})$. Specifically, the diagonal elements of $\tilde{Q}$ and $\tilde{S}$ are given as $(\tilde{Q})_{jj} = \epsilon_j^2$ and $(\tilde{S})_{jj} = \mathrm{Tr}(\rho X_j^2)$, respectively.

From Eq.(3), we derive an analytical error-tradeoff relation for approximating n observables [20]

$$\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}}) \geq \left(\sqrt{\|S_{\mathrm{Re}}^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}S_{\mathrm{Re}}^{-\frac{1}{2}}\|_F + 1} - 1\right)^2, \tag{4}$$

where $\|\cdot\|_F = \sqrt{\sum_{j,k}|(\cdot)_{jk}|^2}$ represents the Frobenius norm. In this inequality, the term Q_{Re} is the sole quantity dependent on the measurement strategy and its diagonal entries correspond to the mean-square errors of the approximation. Both S_{Re} and $\tilde{S}_{\mathrm{Im}}$ are independent of the specific measurement process; instead, they are entirely determined by the inherent properties of the observables when applied to the given quantum state.

The inequality in Eq.(4) establishes a fundamental limit on the minimum achievable errors for any POVM that approximates the given set of observables on a quantum state. It provides a bound that holds true for any choice of orthonormal basis $|u_q\rangle$, and the tightest bound can be obtained by optimizing over all possible $|u_q\rangle$. In the case of pure states, the selection of a specific $|u_q\rangle$ is not necessary. The derived analytical bound guarantees to be tighter than simply summing up Branciard's bounds for two observables pairwisely when the total number of observables exceeds four [20].

2.2 Error-tradeoff relation via semidefinite programming

We proceed to introduce a secondary approach that yields even tighter tradeoff relations. This method bypasses the need for selecting specific $\{|u_q\rangle\}$ and can be formulated as semi-definite programming (SDP), enabling efficient computation.

Again for any POVM, $\{M_m\} \in H_S$, it can be realized as projective measurement, $\{V_m\} \in H_S \otimes H_A$, with $(I \otimes \langle\xi_0|)V_m(I \otimes |\xi_0\rangle) = M_m$. We can then construct $\{F_j = \sum_m f_j(m)V_m\}$ to approximate $\{X_j \otimes I_A\}$. Let Q be an $n \times n$ Hermitian matrix, with its jkth element given as

$$\begin{aligned} Q_{jk} =& \mathrm{Tr}\left[(\rho\otimes\sigma)(F_j - X_j\otimes I)(F_k - X_k\otimes I)\right] \\ =& \mathrm{Tr}\left[\rho\sum_m f_j(m)M_m f_k(m)\right] - \mathrm{Tr}\left(\rho R_j X_k\right) \\ & - \mathrm{Tr}\left(\rho X_j R_k\right) + \mathrm{Tr}\left(\rho X_j X_k\right), \end{aligned} \tag{5}$$

here $R_j = \sum_m f_j(m)M_m$ is a Hermitian matrix in $\mathcal{H}_S$. We let $\mathbb{S}$ be a $n \times n$ block operator whose jk-th block is $\mathbb{S}_{jk} = \sum_m f_j(m)M_m f_k(m)$, and let $\mathbb{R} = \begin{pmatrix} R_1 & R_2 & \cdots & R_n \end{pmatrix}^\dagger$, $\mathbb{X} = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}^\dagger$. We have $\mathbb{S} \geq \mathbb{R}\mathbb{R}^\dagger$. $\mathcal{E}$ can then be rewritten as $\mathcal{E} = \mathrm{Tr}(WQ) = \mathrm{Tr}\left[(W\otimes\rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right]$, where $W \geq 0$ is a weighted matrix. The minimization of $\mathcal{E}$ is then readily to be formulated as a semi-definite programming with [20]

$$\begin{aligned} \mathcal{E}_0 = \min_{\mathbb{S},\{R_j\}_{j=1}^n} & \mathrm{Tr}\left[(W\otimes\rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right] \\ \text{subject to} \quad & \mathbb{S}_{jk} = \mathbb{S}_{kj} = \mathbb{S}_{jk}^\dagger, \ \forall j,k \\ & R_j = R_j^\dagger, \ \forall j \\ & \begin{pmatrix} I & \mathbb{R}^\dagger \\ \mathbb{R} & \mathbb{S} \end{pmatrix} \geq 0. \end{aligned} \tag{6}$$

The derived lower bound, $\mathcal{E} \geq \mathcal{E}_0$, offers a tighter constraint than the analytical bounds from the previous section for any selection of $\{|u_q\rangle\}$. Furthermore, an explicit construction detailing the optimal approximation strategy that attains this bound for pure states is provided, which demonstrates the tightness of the bound for any number of observables when applied to pure states [20].

2.3 Tighter analytical relation for two observables

By leveraging the SDP bound provided in Eq.(6) and employing a judicious selection of $|u_q\rangle$ analogues to the analytical bound in Eq.(4), we can derive analytical bounds on mixed states for two observables that are tighter than the Ozawa's relation, the tightest analytical bound previously known.

When $\rho = |\psi\rangle\langle\psi|$ is a pure state and $W = \mathrm{diag}\{w_1, w_2\}$, Eq.(6) can be analytically solved as [20]

$$w_1\epsilon_1^2 + w_2\epsilon_2^2 \geq \frac{1}{2}\left(\alpha - \sqrt{\alpha^2 - \beta^2}\right), \tag{7}$$

where

$$\begin{aligned} \alpha &= w_1(\Delta X_1)^2 + w_2(\Delta X_2)^2, \\ \beta &= i\sqrt{w_1 w_2}\langle\psi|[X_1, X_2]|\psi\rangle. \end{aligned} \tag{8}$$

For a mixed state, ρ, we can choose any $\{|u_q\rangle\}$ with $\sum_q |u_q\rangle\langle u_q| = I$ and write $\rho = \sum_q \sqrt{\rho}|u_q\rangle\langle u_q|\sqrt{\rho} = \sum_q \lambda_q |\phi_q\rangle\langle\phi_q|$, here $\lambda_q = \langle u_q|\rho|u_q\rangle$, $|\phi_q\rangle = \frac{\sqrt{\rho}|u_q\rangle}{\sqrt{\langle u_q|\rho|u_q\rangle}}$. For each $|\phi_q\rangle$ we can get a corresponding lower bound $\mathcal{E}_{|\phi_q\rangle}$ by substituting $|\phi_q\rangle\langle\phi_q|$ in Eq.(6), and solve it analytically to get $\mathcal{E}_{|\phi_q\rangle} = \frac{1}{2}\left(\alpha_q - \sqrt{\alpha_q^2 - \beta_q^2}\right)$, where α_q and β_q are obtained from Eq.(8) by substituting $|\psi\rangle$ with $|\phi_q\rangle$. Using the fact that $\mathcal{E}_0 \geq \sum_q \lambda_q \mathcal{E}_{|\phi_q\rangle}$, we obtain an analytical bound

$$w_1\epsilon_1^2 + w_2\epsilon_2^2 \geq \sum_q \frac{\lambda_q}{2}\left(\alpha_q - \sqrt{\alpha_q^2 - \beta_q^2}\right). \quad (9)$$

Specifically, by choosing $\{|u_q\rangle\}$ as the eigenstates of $\sqrt{\rho}[X_1, X_2]\sqrt{\rho}$, the analytical bound in Eq.(9) is tighter than the bound obtained from the Ozawa's relation [20]. The presented framework thus not only extends to scenarios involving an arbitrary number of observables but also provides improved analytical bounds in the case of two observables.

2.4 Experiment validation in a superconducting quantum processor

We conducted an experimental verification of the error-tradeoff relations on a superconducting quantum processor, utilizing the Quafu cloud quantum computing platform [21]. The selected processor, ScQ-P136, consists of 136 qubits with single-qubit gate fidelities surpassing 99% [21–23], and for our analysis, we focused exclusively on the first qubit. To experimentally quantify the error ϵ_j for each observable X_j when measured through a specific measurement set $\{M_m\}$, we adopt the "3-state method" [24, 25]. The essence of this approach is illustrated in Fig.2(a), which involves preparing and measuring three distinct quantum states:

$$\rho_1 = \rho,\ \rho_2 \simeq X_j\rho X_j,\ \rho_3 \simeq (I + X_j)\rho(I + X_j). \quad (10)$$

By analyzing the measurement statistics of $\{M_m\}$ on these three states—ρ_1, ρ_2, and ρ_3—we can obtain ϵ_j, the error of the approximation. The corresponding total mean-squared-errors are plotted in Fig.2(b)(c), where simulated results are also presented for comparison.

2.5 Tradeoff relations for multiparameter quantum estimation

By directly applying the tradeoff relations obtained above, we can readily obtain tradeoff relations for estimating multiple parameters by simply substituting the n observables with the n SLDs, $\{L_1, \cdots, L_n\}$. In this context, S_{Re} corresponds to the quantum Fisher information matrix, F_Q. For any POVM $\{M_m\}$, we construct $\{F_j = \sum_m f_j(m)V_m\}$ to approximate the SLDs. By minimizing $\epsilon_j^2 = \mathrm{Tr}\left[(F_j - L_j \otimes I)^2 (\rho_x \otimes \sigma)\right]$ under the given measurement, we have $Q_{\mathrm{Re}} = F_Q - F_C$, and $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}}) = n - \mathrm{Tr}(F_Q^{-1}F_C)$, $\mathcal{E} = \mathrm{Tr}(WQ) = \mathrm{Tr}(WQ_{\mathrm{Re}}) = \mathrm{Tr}[W(F_Q - F_C)]$ [20]. The error-tradeoff relations can thus be directly applied to assess the precision tradeoffs in multi-parameter quantum metrology.

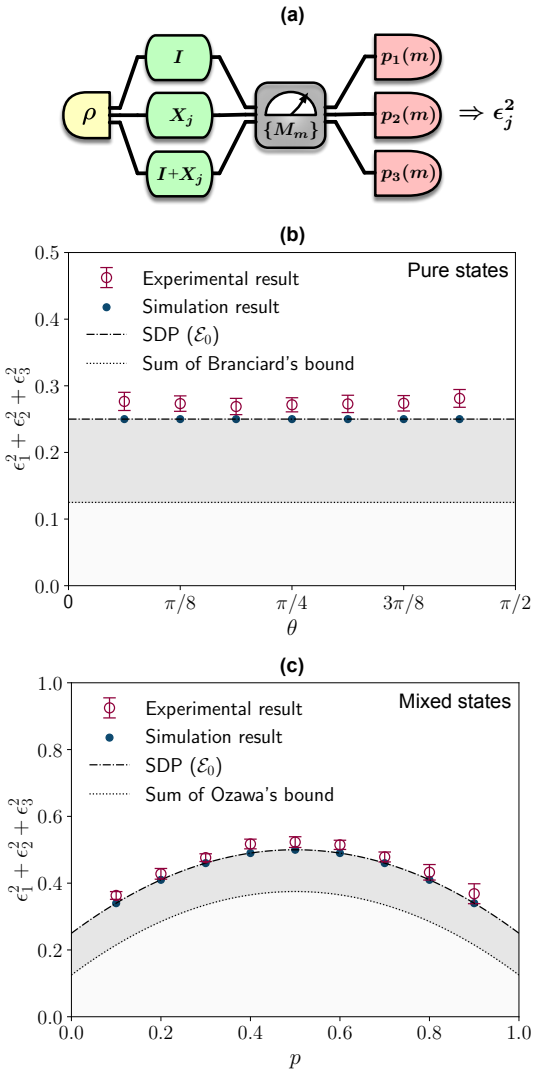


Figure 2: Experimental results for testing the error-tradeoff relations. (a) Scheme diagram to evaluate the mean squared error of each X_j using the "3-state method". (b) Error-tradeoff relations for the simultaneous measurement of three spin operators on a pure state $|\psi\rangle = R_z(\frac{\pi}{2})R_y(\theta)|0\rangle$. (c) Error-tradeoff relations for the simultaneous measurement of three spin operators on a mixed state $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$.

3 Discussion

We developed methodologies that establish tradeoff relations for approximating any number of observables using a single measurement, and we also refined the existing analytical bounds in scenarios involving two observables. Each of the error-tradeoff relations can be directly applied to assess the precision tradeoffs in multi-parameter quantum metrology. This paves the way for further exploration into state-dependent measurement uncertainty relations for multiple observables. Moreover, it reinforces the connection between measurement uncertainty and the incompatibility inherent to multi-parameter quantum estimation, thereby promoting deeper investigations across both domains.

References

[1] W. Heisenberg, "úber den anschaulichen inhalt der quantentheoretischen kinematik und mechanik," *Z. Phys.*, vol. 43, p. 172, 1927.

[2] H. P. Robertson, "The uncertainty principle," *Phys. Rev.*, vol. 34, pp. 163–164, 1929.

[3] E. Schrodinger, "About Heisenberg uncertainty relation," *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, vol. 19, pp. 296–303, 1930.

[4] E. Arthurs and J. L. Kelly Jr., "On the simultaneous measurement of a pair of conjugate observables," *Bell System Technical Journal*, vol. 44, no. 4, pp. 725–729, 1965.

[5] E. Arthurs and M. S. Goodman, "Quantum correlations: A generalized heisenberg uncertainty relation," *Phys. Rev. Lett.*, vol. 60, pp. 2447–2449, Jun 1988.

[6] M. Ozawa, "Universally valid reformulation of the heisenberg uncertainty principle on noise and disturbance in measurement," *Phys. Rev. A*, vol. 67, p. 042105, Apr 2003.

[7] M. Ozawa, "Uncertainty relations for joint measurements of noncommuting observables," *Physics Letters A*, vol. 320, no. 5, pp. 367–374, 2004.

[8] M. Ozawa, "Uncertainty relations for noise and disturbance in generalized quantum measurements," *Annals of Physics*, vol. 311, no. 2, pp. 350–416, 2004.

[9] M. Ozawa, "Physical content of heisenberg's uncertainty relation: limitation and reformulation," *Physics Letters A*, vol. 318, no. 1, pp. 21–29, 2003.

[10] M. Ozawa, "Heisenberg's uncertainty relation: Violation and reformulation," *Journal of Physics: Conference Series*, vol. 504, p. 012024, apr 2014.

[11] M. J. W. Hall, "Prior information: How to circumvent the standard joint-measurement uncertainty relation," *Phys. Rev. A*, vol. 69, p. 052113, May 2004.

[12] C. Branciard, "Error-tradeoff and error-disturbance relations for incompatible quantum measurements," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, p. 6742, 2013.

[13] C. Branciard, "Deriving tight error-trade-off relations for approximate joint measurements of incompatible quantum observables," *Phys. Rev. A*, vol. 89, p. 022124, Feb 2014.

[14] X.-M. Lu, S. Yu, K. Fujikawa, and C. H. Oh, "Improved error-tradeoff and error-disturbance relations in terms of measurement error components," *Phys. Rev. A*, vol. 90, p. 042113, Oct 2014.

[15] M. Ozawa, "Error-disturbance relations in mixed states," *Eprint Arxiv*, vol. arXiv:1404.3388, 2014.

[16] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde, "Noise and disturbance in quantum measurements: An information-theoretic approach," *Phys. Rev. Lett.*, vol. 112, p. 050401, Feb 2014.

[17] P. Busch, P. Lahti, and R. F. Werner, "Colloquium: Quantum root-mean-square error and measurement uncertainty relations," *Rev. Mod. Phys.*, vol. 86, pp. 1261–1281, Dec 2014.

[18] P. Busch, T. Heinonen, and P. Lahti, "Heisenberg's uncertainty principle," *Physics Reports*, vol. 452, no. 6, pp. 155–176, 2007.

[19] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge University Press, 2000.

[20] H. Chen and H. Yuan, "Tradeoff relations for simultaneous measurement of multiple incompatible observables and multi-parameter quantum estimation," *arXiv preprint arXiv:2310.11925*, 2023.

[21] "Baqis quafu group." `https://quafu.baqis.ac.cn`.

[22] BAQIS Quafu Group, "Quafu-RL: The Cloud Quantum Computers based Quantum Reinforcement Learning," *arXiv e-prints*, p. arXiv:2305.17966, May 2023.

[23] BAQIS Quafu Group, "Quafu-Qcover: Explore Combinatorial Optimization Problems on Cloud-based Quantum Computers," *arXiv e-prints*, p. arXiv:2305.17979, May 2023.

[24] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, "Experimental demonstration of a universally valid error–disturbance uncertainty relation in spin measurements," *Nature Physics*, vol. 8, pp. 185–189, Mar. 2012.

[25] M. Ringbauer, D. N. Biggerstaff, M. A. Broome, A. Fedrizzi, C. Branciard, and A. G. White, "Experimental joint quantum measurements with minimum uncertainty," *Phys. Rev. Lett.*, vol. 112, p. 020401, Jan 2014.

Ambient Stress Response of Spin Defects in Two-Dimensional Materials

Xiao-Dong Zeng[1 2 *]

[1] *CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei, Anhui 230026, China*

[2] *CAS Center For Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China*

Abstract. Spin defects in two-dimensional (2D) materials serve as promising systems for quantum sensing applications. The negatively charged boron vacancy (V_B^-) defect in hexagonal boron nitride (hBN) has emerged as the most extensively investigated spin defect. By designing a specific hBN suspension seal structure, external pressure induces strain in hBN films, which affects the properties of V_B^- defects near the strain structure. These V_B^- defects affected by the strain are selectively and site-specifically generated in hBN by using helium ion implantation. Therefore, we can establish the relationship between external pressure and the photoluminescence response properties of the spin defects. Our study presents the first demonstration of a quantum sensor designed for ambient pressure measurements, thereby opening a new avenue for the development of sensing strategies based on spin defects in 2D materials.

Keywords: hBN, 2D material, spin defect, quantum sensor

1 Introduction

Solid-state materials contain many defect structures, among which spin defects exhibit remarkable potential for quantum sensing applications [1, 2, 3, 4]. The most prominent systems are the nitrogen-vacancy (NV) center in diamond and various types of spin defects in silicon carbide (SiC) [5, 6, 7, 8]. These systems enable the initialization, manipulation, and optical readout of their electron spin states. This allows for the direct mapping of external stimuli, such as magnetic and electric fields, temperature, and pressure, onto the spin states [9, 10, 11, 12, 13, 14]. However, these spin defects are all embedded in bulk materials, whose three-dimensional (3D) nature makes it challenging to locate the spin defects near the surface of a sample. The sensitivity of quantum sensors is highly dependent on the proximity between the spin defect and the sensing target. The coherence of spin defects near the surface is greatly reduced, hindering their further applications [15, 16]

In this work, inspired by the fabrication techniques of nanoelectromechanical systems, we ultiized nanofabrication methods to initially fabricate hydrogen silsesquioxane (HSQ) circular annular cylinder on a gold coplanar waveguide. Subsequently, hBN was transferred onto these annular cylinder, forming hBN thin film sealed cavity structures. Leveraging the sealed cavity structure, the underlying coplanar waveguide configuration enhances the microwave efficiency radiated onto the color centers. Moreover, the pre-stress in the hBN thin films during fabrication leads to an increase in fluorescence intensity, thereby enhancing the sensitivity of this structure for application in quantum sensors. Subsequently, we showcased the potential of V_B^- spin defects in hBN as quantum sensors for ambient pressure, employing a custom-built vacuum pressure chamber.

*zengxiaodong@mail.ustc.edu.cn

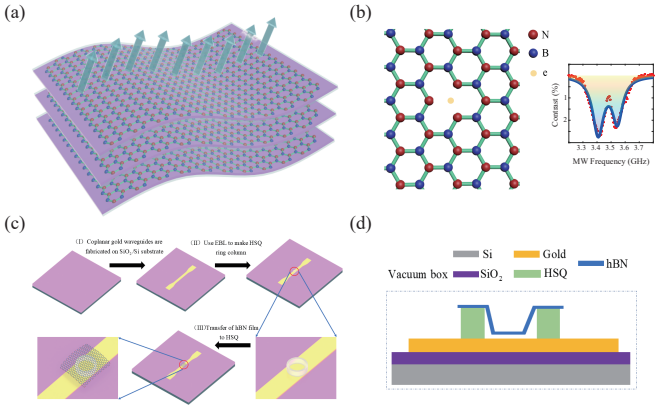


Figure 1: (a) Schematic of a flexible 2D material with spin defects. (b) Schematic diagram of the V_B^- spin defects structure. The typical ODMR signal is shown in the inset. (c) Schematic of sample fabrication step-by-step: hBN overlays HSQ which is positioned on the gold coplanar waveguide. (d) Schematic of the cross-section of a sealed hBN film cavity.

2 Schematic diagram of quantum pressure sensing

As shown in Fig. 1(a), we demonstrate a flexible hBN film embedded with spin defects. The purple planes represent single atomic layers of the hBN, and the arrows indicate the electron spins associated with atomic defects. Local stress variations induce deformation of the atomic layers, resulting in a noticeable shift of the resonant frequency of the spin. Fig. 1(b) illustrates the structure of the V_B^- spin defect, where a boron atom is replaced by an electron to form the defect. The inset on the right displays the V_B^- defect zero-field ground state optically detected magnetic resonance(ODMR), revealing zero-field splitting(ZFS) parameters of approximately $D_{gs} \approx 3.46$ GHz and $E_{gs} \approx 50$ MHz. Fig. 1(c) presents

the design process for the sealed cavity structure in hBN. First, a coplanar waveguide (CPW) is fabricated on a SiO_2/Si substrate. Then, the HSQ resist is shaped into a circular cylinder and patterned onto the CPW. Subsequently, a thin hBN film ($\sim$ 20 nm) is transferred onto the HSQ circular cylinder, Using a commonly employed dry transfer method, a sealed cavity structure is formed. This configuration can react to external pressure alterations, inducing strain in the hBN film. In Fig. 1(d), a cross-sectional schematic of the hBN film-sealed cavity is presented, demonstrating the structure ability to measure high ODMR contrast of the V_B^- spin defect, with potential applications in fabricating cavity structures for other 2D materials.

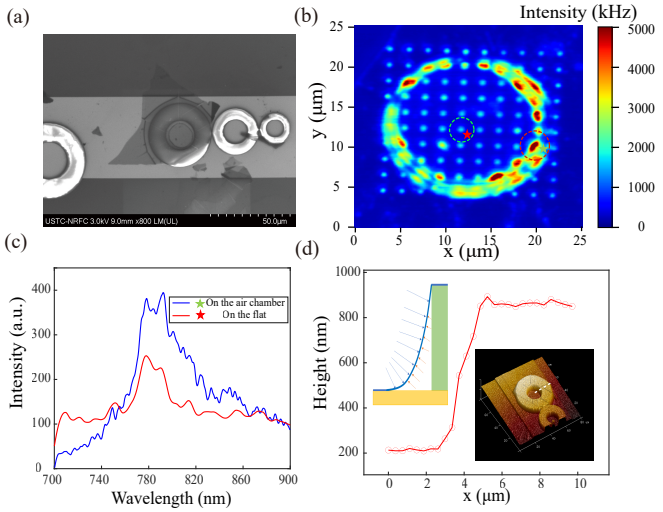


Figure 2: (a) SEM image of the sample with a scale bar of 50 µm in the lower right corner. (b) PL map scanned from the sample outside the vacuum chamber. (c) Spectra of the V_B^- spin defect measured, with the red line representing the spectrum on a flat region and the blue line representing the spectrum on a suspended region. (d) 3D AFM topographic map scanned from the sample, with an inset showing a schematic diagram of the internal and external pressures applied to the hBN film, where the red line indicates the height variation on the surface of the hBN film.

3 Spin defect array

Fig. 2(a) displays the scanning electron microscopy (SEM) image, showing HSQ circular annular cylinder above the CPW and a transferred thin film of hBN positioned above one of the annular cylinder to form the sealed cavity structure. We then used the helium ion microscope (HIM) to generate a V_B^- spin defect array, with the implantation dose of 10^{17} ions/cm^2 and the implantation energy of 30 keV, and the PL map depicted in Fig. 2(b). The distribution of photoluminescence (PL) spectra dinstinctly demonstrates the notable influence of stress at the edges of the circular annular cylinder, under continuous excitation from a 532 nm laser at power of 4 mW. A brighter ring of spin defects, corresponding to the edge of the HSQ circular annular cylinder[**?**], is observable in the array. The suspended region hBN film contributes to the enhanced photoluminescence. The PL spectra of V_B^- spin defects in both the flat and suspended regions were measured, as denoted by the green and red asterisks in Fig. 2(b), respectively. The measurement results reveal an enhancement in photoluminescence fluorescence, as shown in Fig. 2(c). We employed atomic force microscope (AFM) to analyze the topography of the hBN film on HSQ, as shown in Fig. 2(d). The 3D AFM image presents the surface morphology of hBN, with the height distribution position represented by a white dashed line. The schematic in the top-left corner illustrates the distribution of internal and external pressure on the hBN film under applied pressure.

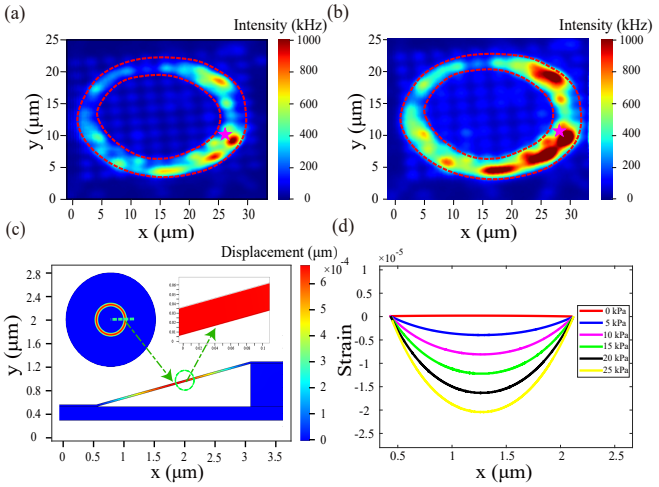


Figure 3: (a) PL map obtained from scanning the sample inside the vacuum chamber at atmospheric pressure, with red circles indicating the regions of the hBN film-sealed cavity. (b) PL map obtained from scanning the sample inside the vacuum chamber under a pressure of 27.2 kPa, with red circles indicating the regions of the hBN film-sealed cavity. (c) Simulation of the strain in the hBN film under external pressure, with a cross-sectional view showing the distribution of lateral stress. (d) Strain in the cross-section of the hBN film under various external pressures.

4 Fluorescent response to pressure

After fabricating the samples, we introduced them into a custom-made Sealed chamber (see Supporting Information). We selected the brightest point in the PL map for further experiments. Subsequently, we applied air pressure to the vacuum pressure chamber. Fig. 3(a) and (b) show the PL map obtained at atmospheric pressure and under an applied pressure of 27.2 kPa, respectively. In these figures, the strain caused by external pressure on the hBN film can be clearly observed, leading to an enhancement of the fluorescence intensity. Based on AFM measurements of the surface topography, we conducted numerical simulations. As shown in Fig. 3(c), when an external pressure of 30 kPa is applied, the suspended re-

gion of the hBN film undergoes deformation. Fig. 3(d) illustrates that as the external pressure increases, the strain in the hBN film also increases.

5 Conclusion

Utilizing the characteristics of hBN two-dimensional thin film, spin defects in 2D materials offer several advantages over traditional 3D solid-state spin systems, including layer-by-layer stacking, ease of integration, etc. Utilizing this property, we designed a sensing demonstration for detecting external pressure. Throughout the experiment, we observed an enhancement in fluorescence count due to stress. By leveraging this fluorescence enhancement, we improved the sensitivity of the targeted measurement points as quantum sensors. Subsequently, we varied the external air pressure and observed strain phenomena in the hBN film, along with corresponding changes in fluorescence count. Exploiting this response, we demonstrated the sensing of ambient pressure. Furthermore, we utilized finite element simulation software to investigate the actual strain of hBN films under external pressure.

References

[1] Childress, L.; Hanson, R. Diamond NV centers for quantum computing and quantum networks. *MRS bulletin*. **2013**, *38*(2), 134-138.

[2] Awschalom, D. D.; Bassett, L. C.; Dzurak, A. S.; Hu, E. L.; Petta, J. R. Quantum spintronics: engineering and manipulating atom-like spins in semiconductors. *Science*. **2013**, *339*(6124), 1174-1179.

[3] Atatüre, M.; Englund, D.; Vamivakas, N.; Lee, S. Y.; Wrachtrup, J. Material platforms for spin-based photonic quantum technologies. *Nat. Rev. Mater.* **2018**, *3*(5), 38-51.

[4] Togan, E.; Chu, Y.; Trifonov, A. S.; Jiang, L.; Maze, J.; Childress, L.; Dutt, M. V.; Sorensen, A. S.; Hemmer, P. R.; Zibrov, A. S.; Lukin, M. D. Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature*. **2010**, *466*(7307), 730-734.

[5] Barry, J. F.; Schloss, J. M.; Bauch, E.; Turner, M. J.; Hart, C. A.; Pham, L. M.; Walsworth, R. L. Sensitivity optimization for NV diamond magnetometry. *Rev. Mod. Phys.* **2020**, *92*(1), 015004.

[6] Jelezkoa, F.; Wrachtrup, J. Single defect centres in diamond: A review. *Physica Status Solidi (a)*. **2006**, *203*(13), 3207-3335.

[7] Riedel, D.; Fuchs, F.; Kraus, H.; Väth, S.; Sperlich, A.; Dyakonov, V.; Soltamova, A. A.; Baranov, P. G.; Ilyin, V. A.; Astakhov, G. V. Resonant addressing and manipulation of silicon vacancy qubits in silicon carbide. *Phys. Rev. Lett.* **2012**, *109*(22), 226402.

[8] Li, Q.; Wang, J. F.; Yan, F. F.; Zhou, J. Y.; Wang, H. F.; Liu, H.; Guo, L. P.; Zhou, X.; Gali, A.; Liu, Z. H.; Wang, Z. Q.; Sun, K.; Guo, G. P.; Tang, J. S.; Li, H.; You, L. X.; Xu, J. S.; Li, C. F.; Guo, G. C. Room temperature coherent manipulation of single-spin qubits in silicon carbide with a high readout contrast. *Natl. Sci. Rev.* **2021**, *9*(5), nwab122.

[9] Maze, J. R.; Stanwix, P. L.; Hodges, J. S.; Hong, S.; Taylor, J. M.; Cappellaro, P.; Jiang, L.; Dutt, M. V. G.; Togan, E.; Zibrov, A. S.; Yacoby, A.; Walsworth, R. L.; Lukin, M. D. Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature*. **2008**, *455*(7213), 644-647.

[10] Doherty, M. W.; Struzhkin, V. V.; Simpson, D. A.; McGuinness, L. P.; Meng, Y. F.; Stacey, Alastair.; Karle, T. J.; Hemley, R. J.; Manson, N. B.; Hollenberg, L. C. L.; Prawer, S. Electronic properties and metrology applications of the diamond NV-center under pressure. *Phys. Rev. Lett.* **2014**, *112*(4), 047601.

[11] Kucsko, G.; Maurer, P. C.; Yao, N. Y.; Kubo, M.; Noh, H. J.; Lo, P. K.; Park, H.; Lukin, M. D. Nanometre-scale thermometry in a living cell. *Nature*. **2013**, *500*(7460), 54-58.

[12] Kraus, H.; Soltamov, V. A.; Fuchs, F.; Simin, D.; Sperlich, A.; Baranov, PG and Astakhov, G. V.; Dyakonov, V. Magnetic feld and temperature sensing with atomic-scale spin defects in silicon carbide. *Sci. Rep.* **2014**, *4*(1), 1-8.

[13] Zhou, Y.; Wang, J. F.; Zhang, X.; Li, K.; Cai, J.; Gao, W. b. Self-Protected thermometry with infrared photons and defect spins in silicon carbide. *Phys. Rev. Appl.* **2017**, *8*(4), 044015.

[14] Simin, D.; Fuchs, F.; Kraus, H.; Sperlich, A.; Baranov, P. G.; Astakhov, G. V.; Dyakonov, V. High-precision angle-resolved magnetometry with uniaxial quantum centers in silicon carbide. *Phys. Rev. Appl.* **2015**, *4*(1), 014009.

[15] Zhang, W. L.; Zhang, J.; Wang, J. F.; Feng, F. P.; Lin, S. R.; Lou, L. R.; Zhu, W.; Wang, G. Z. Depth-dependent decoherence caused by surface and external spins for NV centers in diamond. *Phys. Rev. B.* **2017**, *96*(23), 235443.

[16] Tetienne, J. P. Quantum sensors go flat. *Nat. Phys.* **2021**, *17*(10), 1074–1075.

Quadratic speed-ups in quantum kernelized binary classification

Jungyun Lee[1] * Daniel K. Park[1 2] †

[1] *Department of Statistics and Data Science, Yonsei University, Seoul, Republic of Kore*
[2] *Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea*

Abstract. Quantum kernelized binary classifiers (QKCs) have recently emerged as a promising application at the intersection of quantum kernel methods and machine learning. However, current QKCs do not offer a quantum advantage in the number of data samples, despite requiring an initial quantum state that contains all data samples in superposition. This work demonstrates how leveraging the capability of superposing multiple data samples through Quantum Amplitude Estimation can achieve a quadratic speed-up. Furthermore, we propose simplified QKCs in which the number of qubits is reduced by one, and the circuit depth is reduced linearly with the number of data samples.

Keywords: quantum machine learning, quantum kernel methods, Quantum Amplitude Estimation

1 Introduction

Recent advancements in quantum hardware and simulation frameworks have led to the rise of quantum machine learning (QML), which merges machine learning (ML) with quantum information processing (QIP). This integration offers the potential to overcome the limitations of classical ML methods. A prominent approach within QML is the quantum kernel method (QKM) [1, 2, 3, 4], which leverages quantum computing to enhance the performance of kernel-based algorithms.

In ML, the kernel is a function that quantifies the similarity between two data points, enabling the learning of patterns within a dataset for effective classification or prediction. The computational advantages of the QKM stem from the state and measurement postulates of quantum mechanics, which allow for efficient computation of certain kernel functions on a quantum computer. In particular, the Hadamard or swap test can exponentially expedite the computing of fidelity between two quantum states compared to its classical counterpart [5, 6]. Therefore, they have been harnessed in several quantum kernelized binary classifiers (QKCs) for exponential speed-ups with respect to the number of features (dimensions) in the data when evaluating the classification score in supervised classification. These algorithms are known as the Hadamard classifier (HC) [7] and swap test classifier (SC) [8], respectively, and represent one of the simplest QML protocols with potential quantum speed-up.

The original QKCs prepare an initial quantum state containing entire data samples in a quantum superposition, which is a distinct feature of what quantum computers can do. This state preparation routine is implemented at the cost of increasing the circuit depth linearly and its width logarithmically with the number of data samples. However, QKCs do not utilize this unique property of quantum computing. Thus, there is no quantum advantage with respect to the number of data samples. In other words, previous QKCs failed to utilize the capability of placing training data in superposition and therefore merely increased the size of the quantum circuits without yielding any computational advantages.

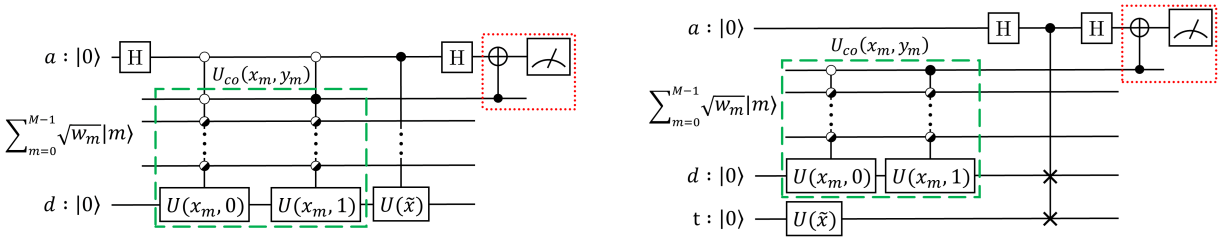
In this work, we first propose simplified QKCs (SQKCs) by modifying the encoding process and the measurement scheme. These modifications allow SQKCs to provide the same work as QKCs while using one less qubit and linear reduction in circuit depth concerning the number of training data. Then, as the main result, we present a protocol for integrating Quantum Amplitude Estimation (QAE) [9] into SQKCs (hence, QKCs), resulting in a quadratic speed-up with respect to the number of data samples used in superposition.

2 Results

2.1 Simplified quantum kernelized binary classifiers

As QKCs are supervised classification algorithms, they should be able to identify the class of the data after initial quantum state preparation. The original QKCs encode the class information into an extra qubit by linearly increasing the depth of the circuit

*ljy89017@yonsei.ac.kr
†dkd.park@yonsei.ac.kr



(a) Simplified Hadamard classifier (SHC)

(b) Simplified swap test classifier (SSC)

Figure 1: Quantum circuit diagrams for (a) Simplified Hadamard classifier (SHC) and (b) Simplified swap test classifier (SSC). The green dashed box, $U_{co}(x_m, y_m)$, indicates that the class-ordered encoding and the red dotted box is the measurement with the Clifford transformation.

Table 1: An overview of variables used for comparing the performance of SQKCs-QAE and SQKCs.

	Number of samples	Estimation error
SQKCs-QAE	$2^{t+1}N^q_{shot} := 2^{t+1}$	The 81% largest value among all errors, $\lvert a - \tilde{a}_i \rvert$, $i = 1, 2, ..., I$.
SQKCs	$N^c_{shot} := 2^{t+1}$	I is the number of repetitions for a given number of samples.

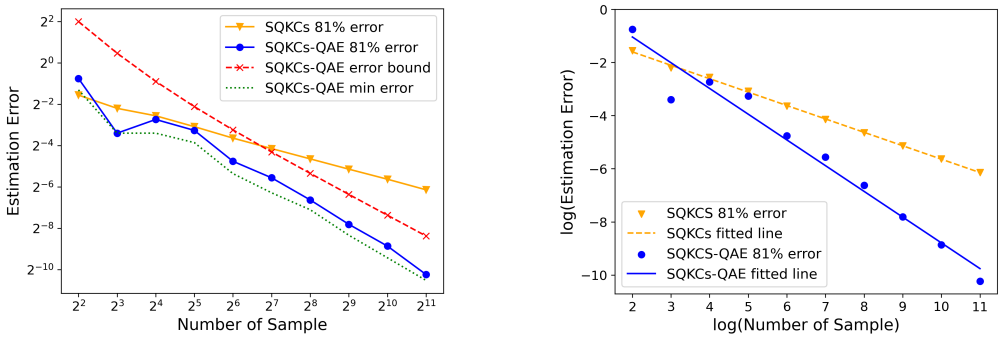
to training data. However, this can be simplified by the class-ordered encoding, shown in the green dashed box in Fig. 1, i.e. encode class 1 data after class 0 data are encoded. With this strategy, the class information is implicitly encoded into one of the qubits making the quantum superposition, $|m\rangle$. Thus, labeling the class of the data points can be achieved without increasing extra circuit width or depth by data size. Moreover, the Clifford transformation shown in the red dotted box in Fig. 1 can further reduce the measurement process from two-qubits to a single-qubit. This reduces the application of QAE by a factor of two, as the new measurement scheme allows classification with a single probability.

2.2 Main protocols

The main protocol of this work is to verify whether SQKCs with QAE (SQKCs-QAE) have the potential to get any quantum speed-up compared with normal SQKCs. Thus, it is crucial to establish a clear and rigorous evaluation criterion for comparing the performance of SQKCs-QAE and SQKCs. In this regard, we compare the rate at which the estimation errors decrease in SQKCs-QAE and SQKCs as the number of samples increases. Two variables, namely the number of samples and the estimation error, whose relationship is analyzed and compared in the subsequent section, are summarized in Table 1. Here "sample" is prepared by querying the circuit Fig. 1a or 1b. Thus, the "number of samples" corresponds to the instances of applying the circuit. The estimator of QAE satisfies the absolute error that is upper bounded by $\mathcal{O}(1/N_q)$ with a probability of at least $8/\pi^2 (\approx 81\%)$, where N_q is the number of samples (see Section 4 in [9] for more detail). Considering the success probability of QAE, "estimation error" is defined as the 81st percentile largest error value for each sample on both SQKCs-QAE and SQKCs. If, for instance, we generated 1000 error results on both SQKCs and SQKCs-QAE, respectively, the comparison is based on the 810th largest error. Note that, in this scenario, I in Table 1 is 1000, where a and $\tilde{a}_i$ denote the real value we want to estimate and its ith estimator, respectively. For a more concise comparison, we fitted each error result to a linear curve using the $\log_2$ function. The slope of the fitted line for SQKCs-QAE and SQKCs indicates how fast the estimation error decreases. Thus, we can verify the speed-up quantitatively by investigating the ratio between two slopes: (slope of the linear fit for SQKCs-QAE estimation error)/(slope of the linear fit for SQKCs estimation error).

2.3 Numerical simulation results

Numerical simulations were conducted on the IBM quantum simulator using the first and last classes, *setosa* and *virginica*, based on two features of the Iris dataset: *sepal width* and *petal length*. The error curves in Fig. 2a represent the mean value computed from a total of 12 results, with 6 for SHC and 6 for SSC. A total of 11 subsets from the Iris dataset are utilized, comprising one set used in both SHC and SSC, along with five independent random



(a) Average comparison result

(b) Linear fitting

Figure 2: (a) Comparison of average error scaling between SQKCs with QAE (SQKCs-QAE) and standard SQKCs. The results are derived from a total of 12 measurements, with 6 for SHC and 6 for SSC. (b) Linear fitting of (a) shows a slope ratio of approximately 1.9185, indicating a quadratic speed-up.

sets for each SHC and SSC. Fig. 2b displays the linear fitting results of Fig. 2a, and the ratio between two slopes is approximately 1.9185 (≈ 2). Comparative simulation results between SQKCs-QAE and SQKCs indicate that, for a given level of precision, the estimation speed of SQKCs can be quadratically enhanced by leveraging data superposition via QAE.

3 Future work

Since the classification score over the full dataset is computed entirely coherently on a quantum computer, the protocols we discussed are fully quantum algorithms. This work primarily focuses on achieving quantum speed-up in computing the classification score, expressed as a weighted sum of the kernel between new input and data samples. However, QAE can also expedite the estimation of a single kernel function, an element of the kernel matrix or Gram matrix (typically given by quantum state fidelity), by integrating QAE into the Hadamard or swap test. Thus, QAE is beneficial in quantum-classical hybrid ML, where the quantum kernel matrix is utilized in classical ML algorithms, such as the support vector machine (SVM). Therefore, estimating the kernel function faster through QAE could pave the way for interesting future work, since many near-term and fault-tolerant quantum models can be replaced or formulated by a general SVM with a quantum kernel [10].

While QKCs and SQKCs introduced in this paper focus on binary problems, they can also address multi-class classification problems using heuristic strategies such as one-vs-rest or one-vs-one. Moreover, we emphasize that QKCs can apply to any datasets, as long as they are supplied as quantum states that can be handled by a quantum computer, either by an inherently quantum-mechanical system or through quantum feature mapping. Therefore, extending our method to multi-class QKCs or investigating it on other datasets remains an interesting avenue for future research.

References

[1] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, vol. 567, no. 7747, pp. 209–212, 2019.

[2] M. Schuld and N. Killoran. Quantum machine learning in feature hilbert spaces. *Physical review letters*, vol. 122, no. 4, p. 040504, 2019.

[3] Y. Liu, S. Arunachalam, and K. Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, vol. 17, no. 9, pp. 1013–1017, 2021.

[4] T. Hur, I. F. Araujo, and D. K. Park. Neural quantum embedding: Pushing the limits of quantum supervised learning. *arXiv preprint arXiv:2311.11412*, 2023.

[5] H. Buhrman, R. Cleve, J. Watrous, and R. De Wolf. Quantum fingerprinting. *Physical Review Letters*, vol. 87, no. 16, p. 167902, 2001.

[6] D. Aharonov, V. Jones, and Z. Landau. A polynomial quantum algorithm for approximating the jones polynomial. in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pp. 427–436, 2006.

[7] M. Schuld, M. Fingerhuth, and F. Petruccione. Implementing a distance-based classifier with a

quantum interference circuit. *Europhysics Letters*, vol. 119, no. 6, p. 60002, 2017.

[8] C. Blank, D. K. Park, J.-K. K. Rhee, and F. Petruccione. Quantum classifier with tailored quantum kernel. *npj Quantum Information*, vol. 6, no. 1, p. 41, 2020.

[9] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, vol. 305, pp. 53–74, 2002.

[10] M. Schuld. Supervised quantum machine learning models are kernel methods. *arXiv preprint arXiv:2101.11020*, 2021.

Amplitude encoding of molecular orbitals in first-quantized systems

Taichi Kosugi[1 2 *] Shunsuke Daimon[3] Hirofumi Nishi[1 2] Yu-ichiro Matsushita[1 2 3]

[1] *Quemix Inc., Taiyo Life Nihombashi Building, 2-11-2, Nihombashi Chuo-ku, Tokyo 103-0027, Japan*
[2] *Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan*
[3] *Quantum Materials and Applications Research Center, National Institutes for Quantum Science and Technology (QST), 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan*

Abstract. Since the generic techniques for amplitude encoding often demand exponential cost in terms of qubit number, it is desirable for users to have moderately specialized techniques for practical tasks. We develop an encoding scheme that generates an arbitrary linear combination of localized functions. It is demonstrated that discrete Lorentzian functions as a basis set lead to more efficient probabilistic amplitude encoding than other localized functions. In addition, our scheme can be rendered deterministic analytically via quantum amplitude amplification. The new scheme will be a powerful tool especially for encoding molecular orbitals in first-quantized systems, as demonstrated by our resource estimation.

Keywords: Amplitude encoding, quantum chemistry

1 Introduction

A quantum algorithm that solves a problem more efficiently than classical algorithms often assumes that an initial many-qubit state has already been prepared in which the initial condition is appropriately encoded. Such generic techniques for preparing an arbitrary many-qubit state from initialized qubits as efficiently as possible are called the amplitude encoding. Since an n-qubit system has $\mathcal{O}(2^n)$ degrees of freedom, the encoding of a truly arbitrary state using predetermined circuit parameters inevitably suffers from exponential classical-computational cost (see, e.g., Refs. [1, 2]). When we tackle a specific kind of problems, however, such techniques are found to be too generic, that is, the degrees of freedom in a generic encoding technique is unnecessarily enormous compared to the amount of information specifying an initial state of practical use. We therefore develop moderately specialized encoding techniques in the present study: generation of an arbitrary linear combination (LC) of localized functions [3].

To this end, we design the main encoding technique by starting from the probabilistic operation [4, 5] for an LC of discrete Lorentzian functions (LFs), to which we apply the quantum amplitude amplification (QAA) technique [6, 7] to render the encoding deterministic. As is demonstrated later, the expansion of a target function in LFs is favorable for achieving efficient circuit implementation. Our encoding techniques possess applicability to diverse fields of quantum computation. Amongst them, quantum chemistry in real space [8, 9, 10, 11] is a promising one, where an initial state can be constructed from one-electron molecular orbitals (MOs), typically expressed as LCs of localized orbitals. We provide resource estimation for such calculations. Also, we perform the new scheme on the real quantum computers commercially provided by IBM.

*kosugi.taichi@gmail.com

2 Encoding scheme and resource estimation

Setup Let us consider a case where an n_q-qubit data register is available and we are provided with $n_{\rm loc}$ real functions $\{f_\ell\}_{\ell=0}^{n_{\rm loc}-1}$ localized at the origin as an expansion basis set in one-dimensional space. For the equidistant grid points $x_j \equiv j\Delta x$ ($j = 0, \ldots, 2^{n_q}-1$) of a spacing Δx on a range $[0, N\Delta x]$, we want to encode a normalized LC of the displaced basis functions on the data register as

$$|\psi_{\rm lc}\rangle = \sum_{j=0}^{N-1} \sum_{\ell=0}^{n_{\rm loc}-1} d_\ell f_\ell(x_j - k_{{\rm c}\ell}\Delta x)|j\rangle_{n_q}, \qquad (1)$$

where $N \equiv 2^{n_q}$. $|j\rangle_{n_q}$ is the computational basis for the data register. $k_{{\rm c}\ell}$ is the integer coordinate ($0 \le k_{{\rm c}\ell} \le N-1$) of the center of the displaced ℓth basis function. d_ℓ is the real coefficient for the LC. We assume that the basis functions are normalized over the range, that is, $\sum_{j=0}^{N-1} f_\ell(x_j)^2 = 1$ for each ℓ, and the displaced ones are not necessarily orthogonal to each other. We further assume that the n_q-qubit unitary for generating each basis function centered at the origin is known: $U_{\rm orig}^{(\ell)}|0\rangle_{n_q} = \sum_{j=0}^{N-1} f_\ell(j\Delta x)|j\rangle_{n_q}$ for each ℓ.

Generation of LC We define the phase shift gate $U_{\rm shift}(k)$ for an integer k that acts on a computational basis diagonally as

$$U_{\rm shift}(k)|j\rangle_{n_q} = \exp\left(-i\frac{2\pi k}{N}j\right)|j\rangle_{n_q}. \qquad (2)$$

This gate can be implemented as separate single-qubit gates. With this and the quantum Fourier transform (QFT), the operator

$$T(k) \equiv {\rm QFT}\cdot U_{\rm shift}(k)\cdot {\rm QFT}^\dagger \qquad (3)$$

is easily confirmed to perform modular addition for a computational basis as $T(k)|j\rangle_{n_q} = |(j+k) \bmod N\rangle_{n_q}$.

Each of the displaced basis functions can thus be generated by locating f_ℓ at the origin and translating it by $k_{c\ell}$ using the translation operator $T(k_{c\ell})$. The circuit for encoding the LC probabilistically can then be constructed by the method described in Ref. [5], where $n_{\rm A} \equiv \lceil \log_2 n_{\rm loc} \rceil$ ancillae are introduced and the circuit parameters are calculated from the coefficients $\{d_\ell\}_\ell$. The probabilistic circuit involves two QFT gates. Since the success probability can be calculated analytically, we can construct a unitary gate by introducing a single extra ancilla for QAA, so that the encoding is performed deterministically and thus there is no need for a measurement. For details, see the original paper [3].

Discrete Lorentzian functions The quantum computational cost of the encoding depends on the basis functions we adopt for expanding a target function. To find good basis functions, we first define, by recalling the work done by Klco and Savage [12], a discrete Slater function (SF) $S(n_q, a)$ having a decay rate $a > 0$ such that its value at an integer coordinate j is

$$S_j(n_q, a) \equiv \begin{cases} C_S(n_q, a) e^{-a\widetilde{j}} & 0 \le \widetilde{j} < N/2 \\ C_S(n_q, a) e^{-a(N-\widetilde{j})} & N/2 \le \widetilde{j} < N \end{cases} \quad (4)$$

with the normalization constant $C_S(n_q, a)$. We have introduced the tilde symbol as $\widetilde{j} = j \bmod N$. $S(n_q, a)$ is a period-N function having cusps at the origin and its duplicated points. A single SF can be thankfully encoded by a depth-$\mathcal{O}(\log n_q)$ unitary circuit $U^{(\rm S)}$ [3]. We define the discrete LF $L(n_q, a)$ corresponding to the SF defined above as

$$L_j(n_q, a) \equiv \frac{C_S(n_q, a)}{\sqrt{N}} \frac{(1 - e^{-2a})(1 - (-1)^j e^{-aN/2})}{1 - 2e^{-a}\cos(2\pi j/N) + e^{-2a}}, \quad (5)$$

which is also localized at the origin. One can confirm that the LF and SF are related via QFT, which means that the unitary gate $U^{(\rm L)} \equiv {\rm QFT} \cdot U^{(\rm S)}$ encodes the LF with depth $\mathcal{O}(n_q \log n_q)$. This expression tells us that, if we adopt the LFs as the expansion basis set, one of the two QFT gates in the probabilistic encoding of the LC cancels $U^{(\rm S)}$ for each basis function. The circuit $\mathcal{C}_{\rm lc}^{(\rm L)}$ in such a case is shown in Fig. 1. Even when we determinize this circuit by employing QAA, the QFT gate is performed only once, thanks to the Lorentzian basis set.

Resource estimation for quantum chemistry in real space One of the intriguing applications of our scheme is the encoding of molecular orbitals (MOs) for quantum chemistry in real space. Our scheme is straightforwardly extended to a three-dimensional case by employing product basis functions. Before we start a simulation of real-time dynamics or an energy minimization procedure such as the probabilistic imaginary-time evolution (PITE) [8] or the adiabatic time evolution (ATE) [11] or variational methods for a target molecule, an initial many-electron state has to be prepared. The efficient antisymmetrization scheme proposed by Berry et al. [13] assumes that all the occupied MOs have been prepared. Here we analyze the computational cost that is required for encoding MOs as LCs of LFs. As discussed in Ref. [8], the required number of data qubits for encoding an MO scales typically as $n_q = \mathcal{O}(\log(n_{\rm el}^{1/3}/\Delta x))$ for a molecule containing $n_{\rm el}$ electrons. If the desired MO is delocalized over the molecule, the number of basis functions for it thus scales the same way as the molecule size: $n_{\rm loc} = \mathcal{O}(n_{\rm el})$. The computational time with respect to $n_{\rm el}$ spent until the probabilistic encoding is done is thus estimated to be $\mathcal{O}(n_{\rm el}^2 \log n_{\rm el})$, while that for the determinized encoding is $\mathcal{O}(n_{\rm el}^{3/2} \log n_{\rm el})$. If the desired MO is localized, on the other hand, we have clearly $n_{\rm loc} = \mathcal{O}(1)$. The computational time in this case for the probabilistic encoding is estimated to be $\mathcal{O}((\log n_{\rm el}) \log\log n_{\rm el})$ as well as the determinized encoding.

3 Experiments

To confirm the validity of our encoding techniques, we generated LCs of LFs on the real quantum computers provided by IBM. We did not employ the determinization technique since we wanted to perform quantum computation using as few physical qubits as possible. We tried three combinations of decay rates and centers of two LFs for ibm_nairobi. The results are shown in Fig. 2.

4 Conclusions

In summary, we developed probabilistic and deterministic encoding techniques that generates an arbitrary LC of LFs. Since the ancillae increase only as $\mathcal{O}(n_{\rm loc})$ regardless of n_q and QFT is performed only once, this technique achieves efficient encoding. One of the intriguing applications of our generic scheme will be the encoding of MOs used in quantum chemistry in real space. We found that the computational time for encoding a localized MO is polynomial in terms of the logarithm of electron number. This results is encouraging since the encoding of MOs is a crucial part for the state preparation of an initial many-electron wave function. The encoding techniques developed in the present study thus make quantum chemistry in real space more promising on fault-tolerant quantum computers.

References

[1] G.-L. Long and Y. Sun. Efficient scheme for initializing a quantum register with an arbitrary superposed state. *Phys. Rev. A*, 64, 014303, 2001.

[2] V. Bergholm, J. J. Vartiainen, M. Möttönen, and M. M. Salomaa. Quantum circuits with uniformly controlled one-qubit gates. *Phys. Rev. A*, 71, 052330, 2005.

[3] T. Kosugi, S. Daimon, H. Nishi, and Y.-i. Matsushita. Qubit encoding for a mixture of localized functions. arXiv:2404.18529, 2024.

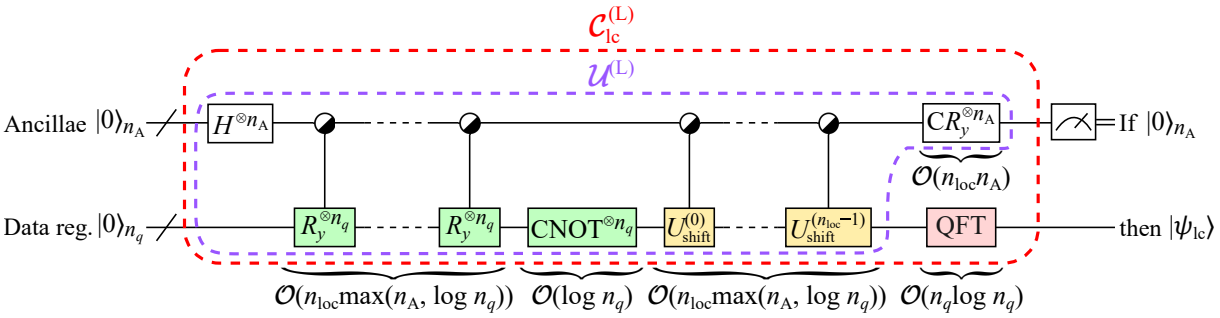


Figure 1: Circuit $\mathcal{C}_{\rm lc}^{(\rm L)}$ for probabilistic encoding of the LC of LFs.

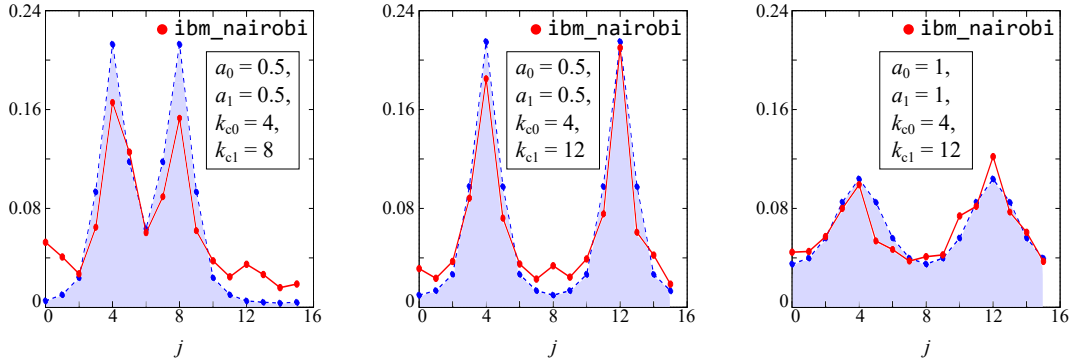


Figure 2: Observed probability distributions as red circles according to probabilistic encoding of $|\psi_{\rm lc}\rangle \propto |L; a_0, k_{\rm c0}\rangle_{n_q} + |L; a_1, k_{\rm c1}\rangle_{n_q}$ with various combinations of decay rates and centers of the Lorentzian function states for $n_q = 4$ qubits on ibm_nairobi. The ideal distributions are also shown as blue circles.

[4] T. Kosugi and Y.-i. Matsushita. Construction of green's functions on a quantum computer: Quasiparticle spectra of molecules. *Phys. Rev. A*, 101, 012330, 2020.

[5] T. Kosugi and Y.-i. Matsushita. Linear-response functions of molecules on a quantum computer: Charge and spin responses and optical absorption. *Phys. Rev. Research*, 2, 033043, 2020.

[6] G. Brassard and P. Hoyer. An exact quantum polynomial-time algorithm for simon's problem. *Proceedings of the Fifth Israeli Symposium on Theory of Computing and Systems*, 12, 1997.

[7] G. Brassard, P. Hoyer, M. Mosca, and A. Tapp. Quantum Amplitude Amplification and Estimation. arXiv:quant-ph/0005055.

[8] T. Kosugi, Y. Nishiya, H. Nishi, and Y.-i. Matsushita. Imaginary-time evolution using forward and backward real-time evolution with a single ancilla: First-quantized eigensolver algorithm for quantum chemistry. *Phys. Rev. Research*, 4, 033121, 2022.

[9] T. Kosugi, H. Nishi, and Y.-i. Matsushita. First-quantized eigensolver for ground and excited states of electrons under a uniform magnetic field. *Japanese Journal of Applied Physics*, 62, 062004, 2023.

[10] T. Kosugi, H. Nishi, and Y.-i. Matsushita. Exhaustive search for optimal molecular geometries using imaginary-time evolution on a quantum computer. *npj Quantum Information*, 9, 112, 2023.

[11] Y. Nishiya, H. Nishi, Y. Couzinié, T. Kosugi, and Y.-i. Matsushita. First-quantized adiabatic time evolution for the ground state of a many-electron system and the optimal nuclear configuration. *Phys. Rev. A*, 109, 022423, 2024.

[12] N. Klco and M. J. Savage. Minimally entangled state preparation of localized wave functions on quantum computers. *Phys. Rev. A*, 102, 012612, 2020.

[13] D. W. Berry, M. Kieferová, A. Scherer, Y. R. Sanders, G. H. Low, N. Wiebe, C. Gidney, and R. Babbush. Improved techniques for preparing eigenstates of fermionic hamiltonians. *npj Quantum Information* 4, 22, 2018.

Decoding Error Correction Codes with Boundaries

Mark Bryan Myers II[1] * Hui Khoon Ng[1 2] †

[1] *Center for Quantum Technologies, National University of Singapore, 117543, Singapore*
[2] *Yale-NUS College, Singapore 138527, Singapore*

Abstract. Quantum error correction (QEC) codes are crucial for mitigating noise in quantum systems, ensuring reliable quantum computations. Non-periodic boundaries in QEC codes introduce additional complexities in error correction. This paper examines how spatial boundaries affect QEC performance, focusing on decoding algorithms like Minimum-Weight Perfect-Matching (MWPM) and Union-Find (UF). Through numerical simulations, we analyze various boundary-handling strategies and their impact on the threshold of topological codes. Our findings show that these strategies can either improve or reduce error correction efficiency, providing valuable insights for optimizing QEC algorithms in the presence of physical boundaries.

Keywords: Fault-Tolerant Quantum Error Correction, Quantum Information

1 Introduction

Quantum computing stands at the forefront of computational science, offering unparalleled potential to solve problems that elude today's classical computers. However, the journey towards practical quantum computation is fraught with challenges, many of which stem from presence of noise in quantum devices. Quantum error correction (QEC) is needed to protect the fragile quantum information being processed; achieved by utilizing QEC codes to encode logical information across numerous physical qubits, allowing for the detection and correction of errors.

2 Quantum Error Correction

A core challenge in QEC is the complex interplay between theoretical models and experimental implementations. Theoretical frameworks like the toric code assume certain hardware capabilities, and overlook practical limitations. As experimental efforts progress, these constraints became clearer, emphasizing the need to align theory with practice. Significant milestones in this evolution include the shift from the toric code to the planar surface code, and from the planar to the rotated planar surface code [2, 1]. The toric code, which operates on a torus without physical boundaries, is conceptually elegant but impractical due to the requisite qubit topologies and long-range interactions. Recognizing some hardware architectures would be limited to nearest-neighbor interactions, the planar surface code emerged as a more practical solution. The rotated surface code further improves efficiency, offering the same error correction guarantees with fewer qubits. Notably, both planar codes have non-periodic boundaries [3]. These transitions highlight the necessity of accommodating the physical limitations of real-world quantum platforms.

3 Decoding with Boundaries

Decoding algorithms, such as Minimum Weight Perfect Matching using Sparse-Blossom (MWPM), Union-Find (UF), Maximum-Likelihood (ML), Neural Networks (NN), and Cellular Automaton (CA), are central to quantum error correction [4, 5]. Their primary goal is to identify and correct faults to preserve encoded quantum information. Decoders use syndrome measurements from error correction circuits to locate defects and apply corrective operations. However, certain fault-correction combinations can form non-trivial chains, leading to logical errors. In the planar surface codes, these chains must span the code's spatial boundaries to induce logical errors; whereas, for the toric code these chains must form closed-loops around the torus.

As we consider increasing code distance, the ratio of boundary qubits to interior qubits goes down; which may lead one to believe that boundary effects will diminish. This is natural, since fewer logical errors will be observed, due the need for longer fault-correction chains to induce each logical errors. However, when logical errors do occur, they always involve boundary interactions. When physical boundaries are present, there is unknowable information, with regards to the detection capabilities of the virtual ancillaries which could potentially identify defects – if physically present. Therefore, it is crucial that we carefully consider how physical boundaries can be utilized to improve the efficiency and accu-

* mbmyersii@u.nus.edu
† huikhoon.ng@nus.edu.sg

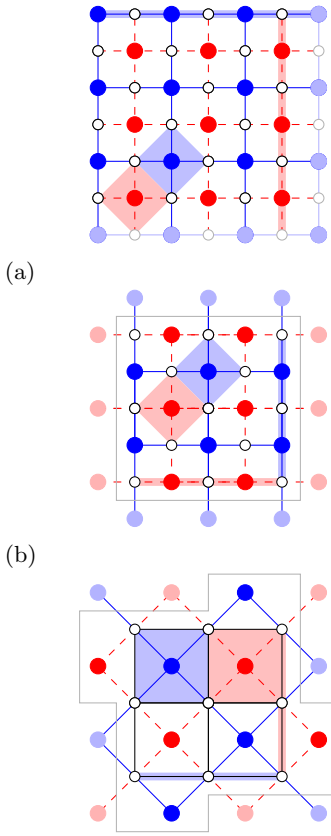


Figure 1: Visualization the distance $d = 3$ (a) toric code, (b) planar surface code, and (c) rotated surface code. X ancillaries (blue) are on each vertex, Z ancillaries (red) are on each plaquette, and data qubits (empty black circles) are on each edge. Example Z/X stabilizers are shown in the blue/red shaded regions. The red dashed lines illustrate the dual lattice. The blue and red highlighted row and column correspond to the logical X and Z operators. (a) The faded-out column and row indicate the periodic boundaries of the toric code lattice, where the bottom row and right column map to the top row and left column respectively. For clarity, this means that the X ancillaries (faded blue) in the three corners of the lattice, map to the same singular X ancillary qubit (blue) in the top left of the lattice. (b) and (c) The grey outline indicates where the physical boundaries of the lattices are located, and the faded blue and red circles represent virtual ancillary qubits which are missing from the stabilizers.

racy of decoding. See figure 2 for some examples of the strategies we investigated in this work.

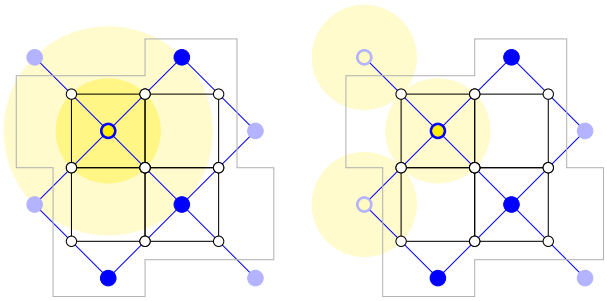


Figure 2: A yellow dot, inside an ancillary, indicates a detected defect. The yellow shaded regions indicate the cluster growths, with darker shading indicating multiple growth steps. (a) Treat Virtual Ancillaries as Stop Condition: You can see that during the second growth step, the cluster has reached two virtual ancillaries; however, it takes two growth rounds (b) Treat Virtual Ancillaries as Detected Defects with a Stop Condition: This method requires fewer growth steps; however, you do have to grow more clusters.

We have found that many still benchmark their decoding algorithms on toric code models – assuming that the algorithms straightforwardly generalize to codes with non-periodic boundaries. Some research has addressed decoding algorithms for codes with physical boundaries; however, these studies do not explore the rationale behind their specific boundary-handling strategies. This leaves a gap in understanding how spatial and temporal boundaries affect decoder performance. We aim to fill this gap by examining the impact of physical boundaries on decoders like MWPM and UF; where, through optimizing the boundary handling strategy, we were able to improve benchmarks, such as the fault-tolerance threshold and decoder runtime. Our findings highlight the importance of developing decoding algorithms that take full advantage of the spatial boundaries present in QEC codes.

References

[1] Fowler, A. G., Mariantoni, M., Martinis, J. M., & Cleland, A. N. (2012). Surface codes: Towards practical large-scale quantum computation, arXiv:1208.0928 [quant-ph].

[2] Kitaev, A. Yu., (1997). Fault-tolerant quantum computation by anyons, arXiv:quant-ph/9707021.

[3] Bravyi, S. B., & Kitaev, A. Y. (1998). Quantum codes on a lattice with boundary, arXiv:quant-ph/9811052.

[4] Higgott, O., & Gidney, C. (2023). Sparse Blossom: Correcting a million errors per core second with minimum-weight matching, arxiv:2303.15933 [quant-ph].

[5] Delfosse, N., & Nickerson, N. H. (2021). Almost-linear time decoding algorithm for topological codes, arXiv:1709.06218 [quant-ph].

The Magic in Qudit Shadow Estimation based on the Clifford Group

Chengsi Mao[1 2 3] Changhao Yi[1 2 3] Huangjun Zhu[1 2 3 *]

[1] *State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai 200433, China*
[2] *Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China*
[3] *Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China*

Abstract. Shadow estimation is a powerful tool in quantum characterization and verification. In this work, we extend the main theoretical results of Huang *et al.* to all prime local dimensions and general Clifford orbits, and bridge the gap between qudit and qubit systems in shadow estimation. Furthermore, adopting Clifford orbits based on magic states as measurement primitive, we provide rigorous bounds showing that a single magic gate can already boost the performance of shadow estimation. Specifically, the sample complexity corresponding to qudit stabilizer measurements is independent of system size, while its counterpart with qudit magic orbits is independent of both system size and local dimension.

Keywords: classical shadows, Clifford group, magic gates, qudit

1 Introduction

The shadow estimation is a sample-efficient protocol for learning the properties of a quantum system through randomized measurements [1]. Among existing protocols, Clifford-based measurements play a pivotal role. The Clifford group is one of the most important groups in quantum information processing, with extensive applications in quantum computation, quantum error correction, and randomized benchmarking. While most considerations focus on qubit systems, little is known about the efficiency of qudit shadow estimation previously.

In this work we perform the first systematic and in-depth study of qudit shadow estimation based on the Clifford group. In particular, we consider the case where local dimension d is an odd prime. Surprisingly, we find that although the qudit stabilizer states may deviate exponentially from a 3-design in terms of the third moment operator, the overhead of its associated sample complexity in shadow estimation, compared with qubit stabilizer measurements, is only $\mathcal{O}(d)$, which is independent of system size n.

Furthermore, we investigate shadow estimation with measurement primitive being the Clifford orbit based on qudit magic states, which we shall call 'qudit magic orbit', and prove rigorous upper bounds on its associated sample complexity. We emphasize that a single magic gate can already eliminate the $\mathcal{O}(d)$ overhead in qudit shadow estimation and bridge the gap between qudit and qubit systems. This provides new evidence for the power of a single magic gate in quantum information processing.

2 Classical shadows

Suppose ρ is an unknown n-qudit quantum state. To extract meaningful information, one repeatedly applies a random unitary U sampled from a pre-selected ensemble $\mathcal{E}$ to rotate the state ($\rho \mapsto U\rho U^\dagger$), followed by a computational-basis measurement with outcome $\mathbf{b} \in \mathbb{F}_d^n$, where d is the local dimension. A quantum channel related to this procedure is defined by $\mathcal{M}(\rho) := \mathbb{E}\left[U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\right]$, with its inverse $\mathcal{M}^{-1}$ called the reconstruction map. The ensemble of $\hat{\rho} := \mathcal{M}^{-1}\left(U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\right)$ stored in classical memory is called the classical shadows of ρ, which can then be used to estimate the expectation value of any observable.

Consider a linear operator $\mathfrak{O}$ in $\mathcal{H}_D := \mathcal{H}_d^{\otimes n}$, where $D = d^n$. Suppose we have N samples, using empirical means, an unbiased estimator $\hat{o}$ for $\mathrm{tr}(\mathfrak{O}\rho)$ can be constructed

$$\hat{o} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{tr}(\mathfrak{O}\hat{\rho}_j). \tag{1}$$

Its mean square error is upper bounded by

$$\langle \epsilon^2 \rangle \le \frac{\|\mathfrak{O}_0\|_{\mathrm{sh}}^2}{N}, \tag{2}$$

where $\mathfrak{O}_0 := \mathfrak{O} - \mathrm{tr}(\mathfrak{O})\mathbb{I}/D$ is the traceless part of $\mathfrak{O}$, and the (squared) shadow norm $\|\mathfrak{O}\|_{\mathrm{sh}}^2$ is defined as follows

$$\max_{\sigma} \mathbb{E}_{U\sim\mathcal{E}} \sum_{\mathbf{b}\in\mathbb{F}_d^n} \langle \mathbf{b}|U\sigma U^\dagger|\mathbf{b}\rangle \cdot \langle \mathbf{b}|U\mathcal{M}^{-1}(\mathfrak{O})U^\dagger|\mathbf{b}\rangle^2. \tag{3}$$

Alternatively, we can use the median of means estimation to decrease the probability of large deviation from the true value. The shadow norm plays a central role in the analysis of sample complexity, and is widely used as a figure of merit for evaluating the performance of shadow estimation protocols.

When the ensemble $\mathcal{E}$ of unitaries forms a 2-design, the associated reconstruction map takes on a simple form, $\mathcal{M}^{-1}(\mathfrak{O}) = (D+1)\mathfrak{O} - \mathrm{tr}(\mathfrak{O})\mathbb{I}$ for any linear operator $\mathfrak{O}$ acting on $\mathcal{H}_D$. If in addition the ensemble $\mathcal{E}$ forms a 3-design, then the shadow norm has upper bound $\|\mathfrak{O}_0\|_{\mathrm{sh}}^2 \le 3\|\mathfrak{O}_0\|_2^2$ [1]. Unfortunately, when the local dimension is an odd prime, the Clifford group is only a 2-design, but not 3-design. When the dimension is not a prime power, the Clifford group is not even a 2-design.

3 Shadow estimation based on stabilizer measurements

Denote the n-qudit Clifford group by $\mathrm{Cl}(n,d)$. We take the unitary ensemble to be $\mathcal{E} = \mathrm{Cl}(n,d)$, which is ex-

*zhuhuangjun@fudan.edu.cn

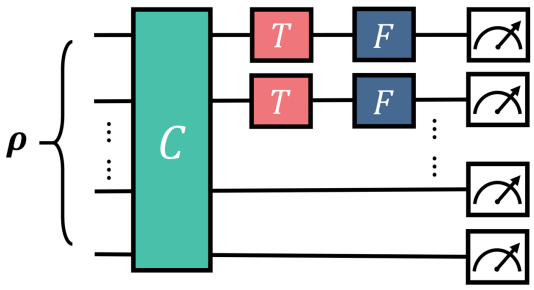


Figure 1: The quantum circuit for data acquisition in shadow estimation based on qudit magic orbits.

actly the qudit analogue of the 'random Clifford measurements' discussed in Ref. [1]. In this case, the corresponding measurement primitive is the Clifford orbit consisting of all stabilizer states. The associated shadow norm is bounded by the following theorem.

Theorem 1 *Suppose d is a prime, n is a positive integer, and $\mathfrak{O}$ is a linear operator on $\mathcal{H}_D$. Adopt stabilizer measurements, i.e., each random unitary is chosen uniformly from $\mathrm{Cl}(n,d)$. Then the shadow norm of its traceless part $\mathfrak{O}_0$ satisfies*

$$\|\mathfrak{O}_0\|_2^2 \le \|\mathfrak{O}_0\|_{\mathrm{sh}}^2 \le (2d-1)\|\mathfrak{O}_0\|_2^2. \tag{4}$$

The upper bound is asymptotically tight as $n \to \infty$, and $\mathfrak{O}$ is a projector onto a stabilizer state.

4 Shadow estimation based on qudit magic orbits

Magic gates supplement the set of Clifford gates so that universal quantum computation can be achieved. Here we are mainly interested in qudit magic gates that are diagonal in the computational basis, which were clarified by Howard and Vala [3]

$$T := \begin{cases} \sum_{u\in\mathbb{F}_d} \omega^{f(u)}|u\rangle\langle u| & d \ge 5, \\ \sum_{u\in\mathbb{F}_d} \omega_9^{f(u)}|u\rangle\langle u| & d = 3. \end{cases} \tag{5}$$

For $d \ge 5$, $\omega = e^{2\pi \mathbf{i}/d}$, and f is a cubic polynomial in $\mathbb{F}_d[x]$ with nonzero cubic coefficient. For $d = 3$, $\omega_9 = e^{2\pi\mathbf{i}/9}$, and $f(u) = c_3u^3 + 3c_2u^2$ ($c_3 \in \mathbb{Z}_9$, $c_2 \in \mathbb{F}_3$).

Assume that d is an odd prime. For shadow estimation based on qudit magic orbits, every element in unitary ensemble $\mathcal{E}$ consists of a random Clifford unitary $C \in \mathrm{Cl}(n,d)$ and k magic gates T ($1 \le k \le n$) acting on different qudits, each followed by a Fourier gate F

$$F := \frac{1}{\sqrt{d}} \sum_{x,y\in\mathbb{F}_d} e^{\mathbf{i}\cdot\frac{2\pi}{d}\cdot xy}|x\rangle\langle y|. \tag{6}$$

See Fig. 1 for an illustration. We prove the following upper bound on the shadow norm in this case.

Theorem 2 *Suppose d is an odd prime, n is a positive integer and $\mathfrak{O}$ is a linear operator on $\mathcal{H}_D$. Denote T as the d-dimensional magic gate defined in Eq. (5). Adopt a Clifford orbit based on magic states as the measurement primitive, i.e., each random unitary is chosen uniformly from $\mathrm{Cl}(n,d)$, followed by T gates and Fourier gates. Then the shadow norm of the traceless part $\mathfrak{O}_0$ of $\mathfrak{O}$ satisfies*

$$\|\mathfrak{O}_0\|_{\mathrm{sh}}^2 \le \gamma_{d,\mathcal{E}}\|\mathfrak{O}_0\|_2^2. \tag{7}$$

The coefficient $\gamma_{d,\mathcal{E}}$ is defined as

$$\gamma_{d,\mathcal{E}} = \begin{cases} 3 + \frac{2^{k+1}(d-2)}{d^k} & d = 2 \mod 3 \text{ or } d = 3, \\ 3 + \frac{9}{8}\cdot\frac{4^k}{d^{k-1}} & d = 1 \mod 3, \end{cases} \tag{8}$$

where k is the number of T gates.

Previously, Ref. [4] exploited the magic in the qubit Clifford group to perform multi-shot shadow estimation. Compared to the homeopathic circuit they use, which is composed of sequences of Clifford circuits, we only require one 'layer' of Clifford unitary, making the increase in circuit depth almost negligible.

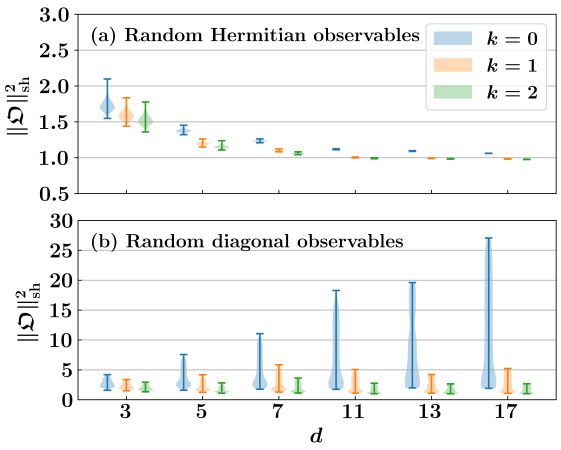


Figure 2: Distribution of the shadow norms of randomly sampled 2-qudit normalized Hermitian and diagonal observables for some odd prime dimensions. By 'normalized' we mean that $\mathfrak{O}$ is traceless and $\|\mathfrak{O}\|_2 = 1$. k is the number of T gates in the measurement primitive.

5 Numerical simulation

We demonstrate our theoretical findings with numerical simulations. In Fig. 2 we show the distribution of shadow norms of 1000 randomly sampled observables, associated with both stabilizer and magic orbit measurements. We see that one single T gate in the measurement primitive leads to a significant reduction in shadow norm, and thus the sample complexity. The effect is particularly strong for observables diagonal in stabilizer basis.

We then consider the task of fidelity estimation using shadow estimation based on magic and non-magic Clifford orbits. We test our protocols on both stabilizer state (n-qudit GHZ state $|\mathrm{GHZ}(n,d)\rangle$) and non-stabilizer state ($T|\mathrm{GHZ}(n,d)\rangle$), where

$$|\mathrm{GHZ}(n,d)\rangle = \frac{1}{\sqrt{d}}\sum_{x=0}^{d-1}|x\rangle^{\otimes n}. \tag{9}$$

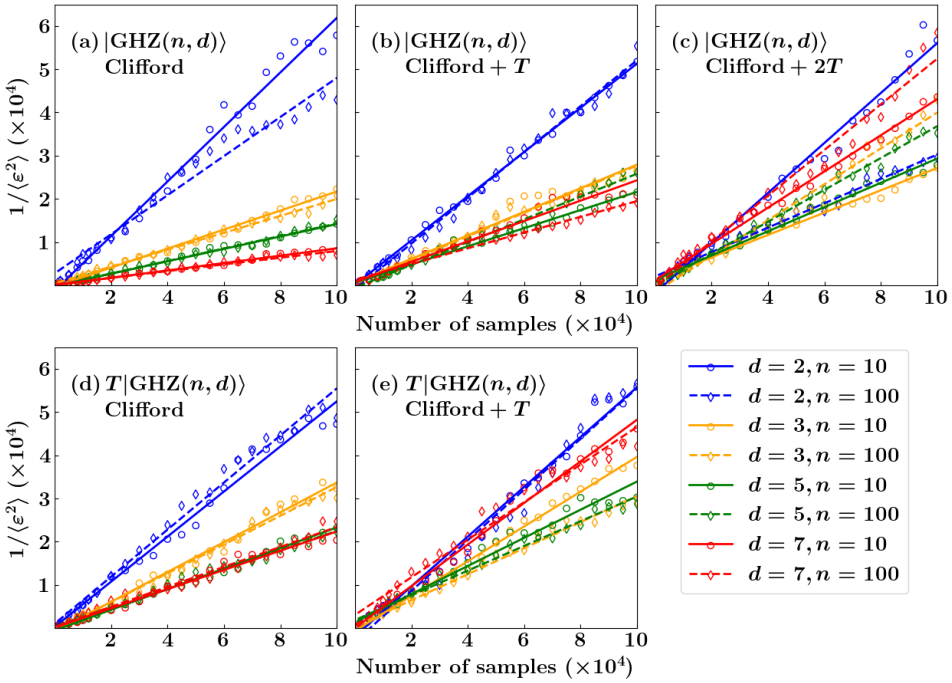


Figure 3: Fidelity estimation of qudit GHZ states and T-modified GHZ states with themselves based on Clifford orbits with different numbers of magic gates. We plot the inverse of average mean square error $1/\langle\epsilon^2\rangle$ over 100 runs versus the number of samples, and by solid (dashed) lines we denote the linear fittings of the data. The input states and measurement primitives are labeled in each figure.

Fig. 3 shows that the inverse of mean square error $1/\langle\epsilon^2\rangle$ increases almost linearly with the number of samples, which is consistent with Eq. (2). The larger the slope, the smaller the shadow norm. In all cases, system size n has almost no influence on sample complexity. When applying stabilizer measurements on stabilizer states (Fig. 3(a)), the sample complexity shows a clear dependence on local dimension d. The dependence becomes almost negligible once we switch to magic orbits. Moreover, the gap between the performance of shadow estimation in qubit systems and those of qudit cases narrows as the number of T gates increases. We also perform numerical simulations with cluster states, and use median of means instead of empirical means. The results are similar.

6 Summary

In this work, we generalize the shadow estimation based on the Clifford group to all prime local dimensions, as well as considering magic Clifford orbits as measurement primitives, which turns out to be a fruitful, yet little explored field of research. Leveraging a little magic-state resource, our protocol overcomes the gap between qudit and qubit systems in shadow estimation. Our work also provides valuable insights on the (qudit) Clifford groups and Clifford orbits, and highlights the power of a single magic gate in quantum information processing, which may have profound implications for various topics beyond shadow estimation.

References

[1] H. -Y. Huang, R. Kueng, and J. Preskill Predicting many properties of a quantum system from very few measurements. *Nat. Phys.*, 16(10):1050, 2020.

[2] D. Gross, S. Nezami, and M. Walter, Schur-Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations. *Commun. Math. Phys*, 385(3):1325–1393, 2021

[3] M. Howard and J. Vala. Qudit versions of the qubit $\pi/8$ gate. *Phys. Rev. A*, 86(2):022316, 2012.

[4] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth Efficient unitary designs with a system-size independent number of non-Clifford gates. *Commun. Math. Phys.*, 397(3): 995-1041, 2023.

[5] M. Heinrich. On stabiliser techniques and their application to simulation and certification of quantum devices. *Ph.D. thesis, University of Cologne*, 2021.

[6] H. Pashayan, O. Reardon-Smith, K. Korzekwa, and S. D. Bartlett Fast estimation of outcome probabilities for quantum circuits. *PRX Quantum*, 3(2): 020361, 2022.

The Magic in Qudit Shadow Estimation based on the Clifford Group

Chengsi Mao,[1, 2, 3] Changhao Yi,[1, 2, 3] and Huangjun Zhu[1, 2, 3]

[1] *State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai 200433, China*
[2] *Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China*
[3] *Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China*

The classical shadow estimation is a sample-efficient protocol for learning the properties of a quantum system through randomized measurements. In this work, we extend the main theoretical results of Huang *et al.* to all prime local dimensions as well as general Clifford orbits, and bridge the gap between qudit and qubit systems in shadow estimation. Furthermore, adopting Clifford orbits based on magic states as measurement primitive, we provide rigorous bounds showing that a single magic gate can already boost the performance of shadow estimation. Specifically, the sample complexity associated with measurements based on qudit stabilizer orbits is independent of system size, while its counterpart with qudit magic orbits is independent of both system size and local dimension.

CONTENTS

I. INTRODUCTION

Learning the properties of an unknown, but physically accessible quantum system is of both fundamental and practical importance in quantum science and technology. The classical shadow estimation was proposed as a sample-efficient protocol for fulfilling this task [1]. Not demanding a full classical description of the quantum state, this protocol circumvents the 'curse of dimensionality' for traditional state tomography [2, 3]. Several variants have been proposed by previous researchers. For example, robust shadow estimation [4, 5] and error-mitigation techniques [6, 7] are introduced to make the protocol noise-resilient. Novel schemes based on shallow circuits [8, 9], quench

dynamics [10] and generalized measurements [11, 12] also exhibit potential advantages in specific scenarios. Experimental implementations have also been reported [13, 14].

Among existing protocols, Clifford-based measurement plays a pivotal role. The Clifford group is one of the most important groups in quantum information processing, with extensive applications in quantum computation, quantum error correction, and randomized benchmarking. While most considerations focus on qubits, the underlying physical system as information carriers are not necessarily binary, but typically exhibit multilevel structures, which can be exploited as a resource in quantum simulation [15–17] and quantum computing algorithms [18, 19]. Developments in experimental control of qudit states with photonics, solid-state, trapped ion, and superconducting platforms have made universal, programmable qudit-based quantum processors possible [20–24]. Generalizing qubit classical shadow estimation to qudits is both theoretically interesting and potentially useful for verifying, calibrating and controlling qudit systems. However, little is known about the efficiency of qudit shadow estimation previously.

In this work we perform the first systematic and in-depth study of qudit shadow estimation based on the Clifford group. We consider the case where local dimension d is an odd prime. Surprisingly, we find that although the qudit stabilizer states may deviate exponentially from a 3-design in terms of the third moment operator, the overhead of its associated sample complexity in shadow estimation, compared with qubit stabilizer measurements, is only $\mathcal{O}(d)$, which is independent of system size n. In particular, with qudit stabilizer measurements, the shadow norm of any linear operator is upper bounded by $(2d-1)$ times its Hilbert-Schmidt norm.

Furthermore, we investigate shadow estimation with measurement primitive being the Clifford orbit based on qudit magic states, which we shall denote as the 'qudit magic orbit'. We prove rigorous upper bounds on the corresponding sample complexity. We emphasize that a single magic gate can already eliminate the $\mathcal{O}(d)$ overhead in qudit shadow estimation, making the sample complexity independent of both system size and local dimension. This can lead to a great reduction in resource cost, especially for high dimensional cases, and bridge the gap between qudit and qubit systems in terms of performance in shadow estimation. Our results also provide new evidence of the power of a single magic gate in quantum information processing.

This paper extracts the key results in our companion paper, which contains complete technical details and additional results, including the proofs of all statements presented here.

II. SHADOW ESTIMATION

Suppose the Hilbert space $\mathcal{H}_D$ is a tensor power of the form $\mathcal{H}_D = \mathcal{H}_d^{\otimes n}$, where the local dimension d is a prime, and the total dimension is $D = d^n$. The computational basis of $\mathcal{H}_d^{\otimes n}$ can be labeled by elements in $\mathbb{F}_d^n$, where $\mathbb{F}_d$ is the finite field composed of d elements. Our main task is to estimate the expectation values of certain linear operators on $\mathcal{H}_D$ with respect to an unknown n-qudit quantum state ρ. To this end, we can repeatedly apply a random unitary U sampled from a pre-selected ensemble $\mathcal{E}$ to rotate the state ($\rho \mapsto U\rho U^\dagger$), followed by a computational-basis measurement with outcome $\mathbf{b} \in \mathbb{F}_d^n$. This procedure defines a quantum channel as follows, $\mathcal{M}(\rho) := \mathbb{E}\left[U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\right]$, and the inverse $\mathcal{M}^{-1}$ is called the reconstruction map. By virtue of this map we can construct an estimator $\hat{\rho} := \mathcal{M}^{-1}\left(U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\right)$, called a (classical) shadow of ρ, in each run [1].

Suppose we want to estimate the expectation value of a linear operator $\mathfrak{O}$ on $\mathcal{H}_D$. If N samples are available, then an unbiased estimator $\hat{o}$ for $o := \mathrm{tr}(\mathfrak{O}\rho)$ can be constructed from the empirical mean as follows,

$$\hat{o} = \frac{1}{N}\sum_{j=1}^{N}\hat{o}_j = \frac{1}{N}\sum_{j=1}^{N}\mathrm{tr}(\mathfrak{O}\hat{\rho}_j), \tag{1}$$

where $\hat{\rho}_j$ is the estimator for ρ in the jth run. Since by construction $\mathrm{tr}(\hat{\rho}) = 1$, the variance of $\hat{o}$ only depends on the traceless part $\mathfrak{O}_0 = \mathfrak{O} - \mathrm{tr}(\mathfrak{O})\mathbb{I}/D$ of $\mathfrak{O}$ [1], where $\mathbb{I}$ is the identity operator on $\mathcal{H}_D$. Thus without loss of generality, sometimes we consider $\mathfrak{O}_0$ instead of $\mathfrak{O}$ to simplify the discussion. The mean square error of $\hat{o}$ is upper bounded by (see the supplement for a proof)

$$\langle\varepsilon^2\rangle \leq \frac{\|\mathfrak{O}_0\|_{\mathrm{sh}}^2}{N}, \tag{2}$$

where the (squared) shadow norm $\|\mathfrak{O}\|_{\mathrm{sh}}^2$ is defined as follows [1],

$$\max_{\sigma}\mathbb{E}_{U\sim\mathcal{E}}\sum_{\mathbf{b}\in\mathbb{F}_d^n}\langle\mathbf{b}|U\sigma U^\dagger|\mathbf{b}\rangle\cdot\left|\langle\mathbf{b}|U\mathcal{M}^{-1}(\mathfrak{O})U^\dagger|\mathbf{b}\rangle\right|^2. \tag{3}$$

Alternatively, we can use the median of means estimation to decrease the probability of large deviation from the true value. Generalization to two or more observables is also straightforward. In any case, the shadow norm plays a central role in the analysis of the sample complexity and is widely used as the key figure of merit for evaluating the performance of a shadow estimation protocol.

When the ensemble $\mathcal{E}$ of unitaries forms a 2-design, the associated reconstruction map takes on a simple form, $\mathcal{M}^{-1}(\mathfrak{O}) = (D+1)\mathfrak{O} - \mathrm{tr}(\mathfrak{O})\mathbb{I}$ for any linear

operator $\mathfrak{O}$ acting on $\mathcal{H}_D$. If in addition the ensemble $\mathcal{E}$ forms a 3-design, then the shadow norm satisfies $\|\mathfrak{O}_0\|_{\rm sh}^2 \leq 3\|\mathfrak{O}_0\|_2^2$ [1].

III. HEISENBERG-WEYL GROUP AND CLIFFORD GROUP

The phase operator Z and cyclic-shift operator X for a single qudit are defined as follows,

$$Z|j\rangle = \omega_d^j|j\rangle, \quad X|j\rangle = |j+1\rangle, \tag{4}$$

where $\omega_d := e^{\frac{2\pi \mathrm{i}}{d}}$, and the addition $j+1$ is modulo d. When d is an odd prime, the qudit Heisenberg-Weyl (HW) group $\mathcal{W}(d)$ is generated by Z and X; when $d = 2$, the HW group $\mathcal{W}(d)$ reduces to the Pauli group and is generated by Z, X, and $\mathrm{i}\mathbb{I}$. The n-qudit HW group $\mathcal{W}(n,d)$ is the tensor product of n copies of $\mathcal{W}(d)$, and its elements are called Weyl operators. A Weyl operator is trivial if it is proportional to the identity and nontrivial otherwise. A Weyl operator is m local if it is a tensor product of m nontrivial single-qudit Weyl operators and $(n-m)$ identity operators on $\mathcal{H}_d$.

The single-qudit Clifford group $\mathrm{Cl}(d)$ is the normalizer of the HW group $\mathcal{W}(d)$. The local Clifford group $\mathrm{Cl}(d)^{\otimes n}$ is the tensor product of n copies of $\mathrm{Cl}(d)$. By contrast, the (global) Clifford group $\mathrm{Cl}(n,d)$ is the normalizer of $\mathcal{W}(n,d)$. When $d = 2$, the Clifford group $\mathrm{Cl}(n,d)$ is a 3-design, [25, 26], and this is the only known infinite family of 3-designs based on discrete groups. Indeed, the performance guarantees of qubit shadow estimation is rooted from the 3-design property of $\mathrm{Cl}(n,2)$. When d is an odd prime, the Clifford group $\mathrm{Cl}(n,d)$ is only a 2-design, but not 3-design [27]. Consequently, whether the high efficiency of shadow estimation can be maintained in qudit case is unknown, which motivates the current study.

IV. SHADOW ESTIMATION BASED ON LOCAL CLIFFORD MEASUREMENTS

First, we choose the local Clifford group for shadow estimation, i.e., $\mathcal{E} = \mathrm{Cl}(d)^{\otimes n}$. This is the qudit generalization of the scheme based on 'random Pauli measurements' in qubit shadow estimation [1]. The associated reconstruction map reads

$$\mathcal{M}^{-1}\left(U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\right) = \bigotimes_{j=1}^{n}\left[(d+1)U_j^\dagger|b_j\rangle\langle b_j|U_j - \mathbb{I}\right]. \tag{5}$$

The basic properties of the shadow norm are summarized in Proposition 1 and Theorem 1; see Appendix D for proofs. Proposition 1 also follows from Theorem 4 in the companion paper. As in the qubit case, this strategy is efficient if $\mathfrak{O}$ only acts on a few qudits.

Proposition 1. *Suppose $\mathfrak{O}$ is an m-local Weyl operator. Then its shadow norm associated with local Clifford measurements reads*

$$\|\mathfrak{O}\|_{\rm sh}^2 = (d+1)^m. \tag{6}$$

Theorem 1. *Suppose $\mathfrak{O}$ is an m-local linear operator on $\mathcal{H}_d^{\otimes n}$, that is, $\mathfrak{O} = \tilde{\mathfrak{O}} \otimes \mathbb{I}^{\otimes(n-m)}$ with $\tilde{\mathfrak{O}} \in \mathcal{H}_d^{\otimes m}$. Then the shadow norm of $\mathfrak{O}$ with respect to local Clifford measurements satisfies*

$$\|\mathfrak{O}\|_{\rm sh}^2 \leq d^m\|\tilde{\mathfrak{O}}\|_2^2. \tag{7}$$

V. SHADOW ESTIMATION BASED ON STABILIZER MEASUREMENTS

Next, we turn to shadow estimation based on the global Clifford group, i.e., $\mathcal{E} = \mathrm{Cl}(n,d)$. Equivalently, the associated measurement primitive is the orbit of all n-qudit stabilizer states $\mathrm{Stab}(n,d)$. Since the Clifford group $\mathrm{Cl}(n,d)$ forms a unitary 2-design, any Clifford orbit forms a 2-design, and the corresponding reconstruction map is particularly simple,

$$\mathcal{M}^{-1}(U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U) = (D+1)U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U - \mathbb{I}. \tag{8}$$

However, $\mathrm{Stab}(n,d)$ does not form a 3-design when d is an odd prime [27], and it is substantially more difficult to determine the shadow norm of a generic observable. Actually, little is known about this issue although the counterpart for the qubit case is well known. Indeed, a deep understanding of the third moment of stabilizer states is indispensable to resolve this problem. In the companion paper we have clarified the properties of the third moment, which enable us to derive universal upper bounds for the shadow norm as shown in the following theorem. A sketch of the proof is presented in Appendix F.

Theorem 2. *Suppose $\mathfrak{O}$ is a linear operator on $\mathcal{H}_D$; then the shadow norm of its traceless part $\mathfrak{O}_0$ with respect to stabilizer measurements satisfies*

$$\|\mathfrak{O}_0\|_2^2 \leq \|\mathfrak{O}_0\|_{\rm sh}^2 \leq (2d-3)\|\mathfrak{O}_0\|_2^2 + 2\|\mathfrak{O}_0\|^2. \tag{9}$$

If in addition $\mathfrak{O}$ is diagonal in a stabilizer basis, then

$$\|\mathfrak{O}_0\|_{\rm sh}^2 \leq (d-1)\|\mathfrak{O}_0\|_2^2 + d\|\mathfrak{O}_0\|^2. \tag{10}$$

If in addition $n = 1$ and $\mathfrak{O}$ is diagonal in a stabilizer basis, then

$$\|\mathfrak{O}_0\|_{\rm sh}^2 = (d+1)\|\mathfrak{O}_0\|^2. \tag{11}$$

Incidentally, Eq. (11) implies Proposition 1. Thanks to Theorem 2, the ratio $\|\mathfrak{O}_0\|_{\rm sh}^2/\|\mathfrak{O}_0\|_2^2$ is upper bounded by $2d-1$, which is independent of

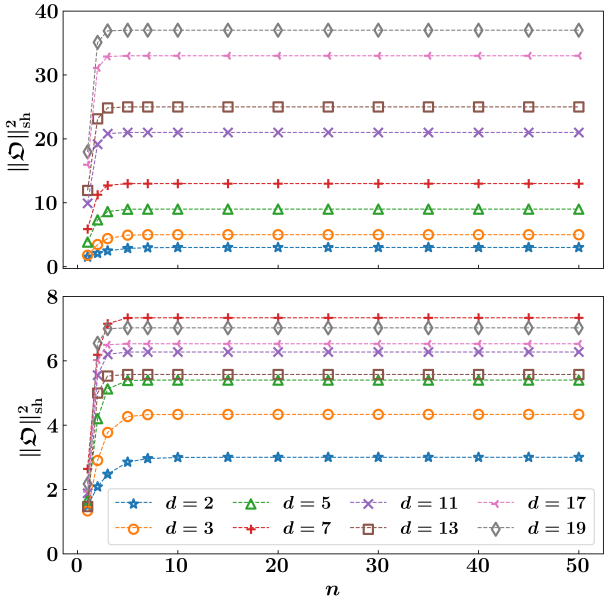


FIG. 1. Shadow norms of projectors onto stabilizer states with respect to Clifford measurements without T gates (upper) and with one canonical T gate (lower).

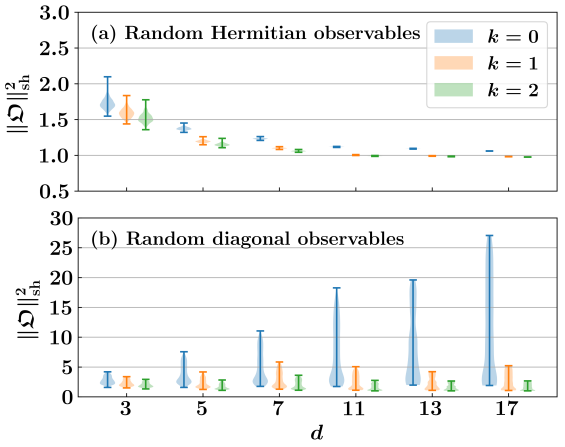


FIG. 2. Distribution of the shadow norms of randomly sampled 2-qudit normalized Hermitian and diagonal observables for some odd prime dimensions. By 'normalized' we mean that $\mathfrak{O}$ is traceless and $\|\mathfrak{O}\|_2 = 1$. k is the number of T gates in the measurement primitive. All T gates are canonical.

the number n of qudits, so the overhead of shadow estimation of a qudit system over a qubit system does not grow with n. This result may have profound implications for quantum information processing based on qudits, and is quite unexpected given that $\mathrm{Stab}(n,d)$ is not a 3-design. In fact, the operator norm of the third normalized moment operator associated with $\mathrm{Stab}(n,d)$ increases exponentially with n when $d = 3$ or $d = 1 \mod 3$ []. Surprisingly, measurement ensembles far from 3-designs (with respect to one of the most popular figures of merit) can achieve similar performance to 3-designs in shadow estimation. When $\mathfrak{O}$ is a stabilizer projector, we can even derive an analytic formula for $\mathfrak{O}$.

Proposition 2. *Suppose $\mathfrak{O}$ is a rank-K stabilizer projector on $\mathcal{H}_D$ with $1 \le K \le d^{n-1}$. Then its shadow norm associated with stabilizer measurements reads*

$$\frac{\|\mathfrak{O}\|_{\rm sh}^2}{\|\mathfrak{O}_0\|_2^2} = \frac{D+1}{D+d}\left(d - 1 - \frac{d}{D} + \frac{d}{K}\right). \tag{12}$$

The shadow norms of projectors onto stabilizer states are illustrated in Fig. 1. They can saturate the upper bounds in Eqs. (9) and (10) asymptotically as $n \to \infty$. Interestingly, they are also the most difficult to estimate by stabilizer measurements. For comparison, Fig. 2 shows the distributions of shadow norms of random Hermitian observables and random diagonal observables of two qudits that are normalized with respect to the Hilbert-Schmidt norm. In the former case, the shadow norms are usually much smaller than the upper bound in Eq. (9) and do not increase with the local dimension; in the later case, by contrast, the upper bound in Eq. (10) is nearly optimal.

VI. THE POWER OF MAGIC GATES IN SHADOW ESTIMATION

Clifford circuits supplemented by sufficiently many magic gates can realize universal quantum computation [], but little is known about the power of limited magic gates in quantum information processing. Here we propose a simple recipe for boosting the efficiency of qudit shadow estimation by virtue of a few magic gates and show that a single magic gate can already bridge the gap between a qudit system and a qubit system.

We are mainly interested in qudit magic gates that are diagonal in the computational basis and belong to the third Clifford hierarchy [28]. These magic gates as clarified by Howard and Vala [29] (see also our companion paper []) are referred to as T gates henceforth. Up to an irrelevant global phase factor, a qudit T gate takes the form

$$T := \sum_{u \in \mathbb{F}_d} \tilde{\omega}^{f(u)} |u\rangle\langle u|, \quad \tilde{\omega} := \begin{cases} \omega_d & d \ge 5, \\ \omega_9 & d = 3. \end{cases} \tag{13}$$

When $d \ge 5$, f is a cubic polynomial on $\mathbb{F}_d$ with nonzero cubic coefficient. When $d = 3$, $f(u) = c_3 u^3 + 3c_2 u^2$ (with $c_2 \in \mathbb{F}_3$, $c_3 \in \mathbb{Z}_9$, and $c_3 \neq 0 \mod 3$) is a function from $\mathbb{F}_3$ to $\mathbb{Z}_9$. The T gate associated with the function $f(u) = u^3$ is referred to as the canonical T gate henceforth.

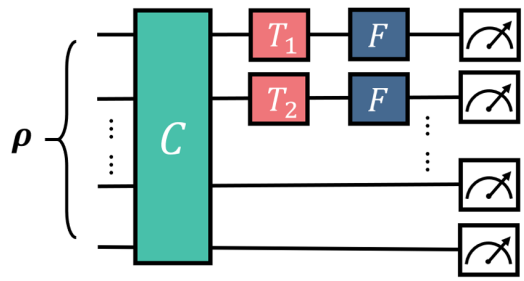


FIG. 3. Shadow estimation based on the Clifford circuit supplemented by a layer of T gates.

Using T gates we can construct an alternative unitary ensemble for shadow estimation as follows: First apply a random Clifford unitary C selected from $\mathrm{Cl}(n,d)$, then apply k T gates $(1 \le k \le n)$ on k different qudits, each followed by a Fourier gate F, where

$$F := \frac{1}{\sqrt{d}} \sum_{x,y\in\mathbb{F}_d} \omega_d^{xy} |x\rangle\langle y|. \tag{14}$$

Without loss of generality, we can assume that the T gates are applied on the first k qudits; denote by T_j the T gate applied to the jth qudit. The resulting unitary ensemble is denoted by $\mathcal{E}\left(\{T_j\}_{j=1}^k\right)$, and the effective measurement ensemble corresponds to a Clifford orbit generated from a magic state. The schematic diagram is illustrated in Fig. 3. The circuit we employ is substantially simpler than popular interleaved Clifford circuits: in each run we need to sample a random Clifford unitary only once, which is as economic as possible and is particularly appealing in the NISQ era [30]. Nevertheless, such simple circuits are surprisingly powerful in shadow estimation as shown in the following theorem.

Theorem 3. *Suppose $\mathfrak{O}$ is a linear operator on $\mathcal{H}_D$ and $T_1, T_2, \ldots, T_k$ are k T gates with $1 \le k \le n$. Then the shadow norm of its traceless part $\mathfrak{O}_0$ with respect to $\mathcal{E}\left(\{T_j\}_{j=1}^k\right)$ satisfies*

$$\|\mathfrak{O}_0\|_{\mathrm{sh}}^2 \le \gamma_{d,k} \|\mathfrak{O}_0\|_2^2, \tag{15}$$

where

$$\gamma_{d,k} := \begin{cases} 3 + \frac{2^{k+1}(d-2)}{d^k} & d \neq 1 \mod 3, \\ 3 + \frac{9}{8} \cdot \frac{4^k}{d^{k-1}} & d = 1 \mod 3. \end{cases} \tag{16}$$

Thanks to Theorem 3, the shadow norm associated with the unitary ensemble $\mathcal{E}\left(\{T_j\}_{j=1}^k\right)$ converges exponentially to the counterpart of a unitary 3-design as the number k of T gates increases. Moreover, a single T gate can already bridge the gap between a qudit system and a qubit system. Notably, if the observable $\mathfrak{O}$ has a bounded Hilbert–Schmidt norm, which is the case for many tasks, such as the fidelity estimation, then the sample complexity is independent of the local dimension and the number of qudits.

When $\mathfrak{O}$ is a stabilizer projector, we can even derive an analytical formula for the shadow norm; see our companion paper for details. The results for projectors onto stabilizer states are shown in Figs. 1 and 6; results for general stabilizer projectors are shown in Fig. 7. (Figures 6 and 7 are in Appendix H.) In addition, T gates can significantly reduce the shadow norms of random diagonal observables, as illustrated in Fig. 2, although their utility for random observables seems limited because Clifford circuits are already good enough. Theorem 3 is applicable irrespective of the specific choices of T gates. If we can make educated choices, then the shadow norm can be reduced further as shown in Appendix H.

VII. NUMERICAL SIMULATION

To corroborate our theoretical findings, we performed extensive numerical simulation on qudit shadow estimation based on Clifford circuits supplemented by T gates. To this end, we first generalize the tableau representation of stabilizer states [31] and gadgetization method for insertion of magic gates in Clifford circuits [32, 33] to the qudit setting (see Appendix I for more details). Together with the sampling algorithm of Clifford unitaries [34, 35], we are able to simulate a single shot in shadow estimation on a classical computer with an approximate computational cost of $\mathcal{O}\left((n+t)^3 + td^{t+1}\right)$, where t is the number of magic states in the gadgetized circuit.

As a showcase, we consider the task of fidelity estimation in which the input state ρ is identical to a pure target state $|\Psi\rangle\langle\Psi|$. In this case, the observable of interest is $\mathfrak{O} = |\Psi\rangle\langle\Psi|$ and the true fidelity is $\mathrm{tr}(\rho\mathfrak{O}) = 1$. Estimators for the fidelity can be constructed from empirical means as in Eq. (1). First, we test our protocols on the n-qudit GHZ state, that is $|\Psi\rangle = |\mathrm{GHZ}(n,d)\rangle$, where

$$|\mathrm{GHZ}(n,d)\rangle := \frac{1}{\sqrt{d}} \sum_{x=0}^{d-1} |x\rangle^{\otimes n}. \tag{17}$$

Figure 4 shows the simulation results on the inverse mean square error $1/\langle\varepsilon^2\rangle$, which increases linearly with the number of samples as expected by Eq. (2). The smaller the shadow norm, the larger the slope of the interpolation line. In all cases under consideration, the number n of qudits has little influence on the sample complexity. If we choose Clifford circuits without T gates for shadow estimation, then the slope is approximately inversely proportional to the local dimension d. If we add one canonical T gate, then the ratio of the maximum slope over the

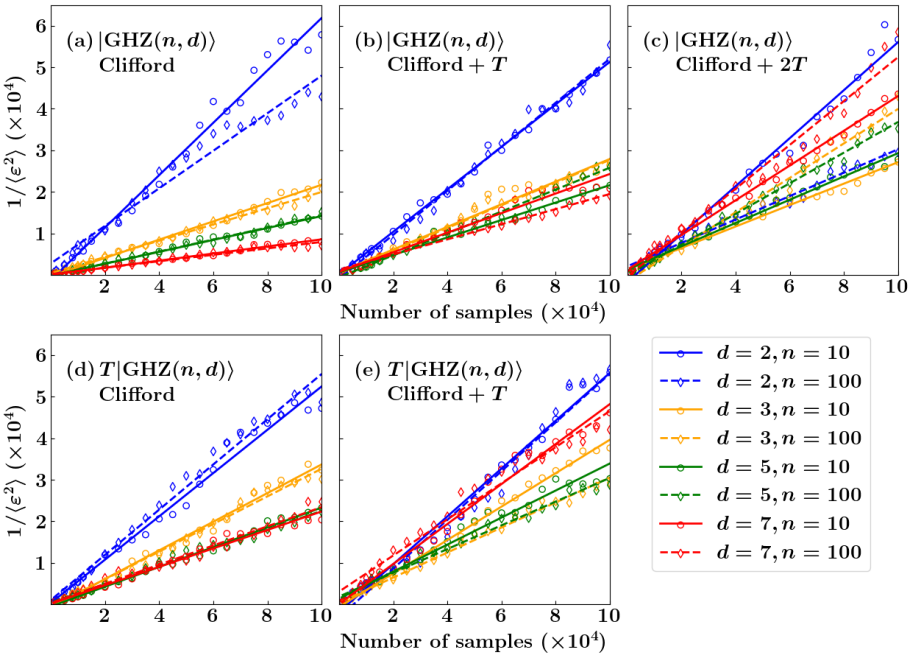


FIG. 4. Inverse mean square error in fidelity estimation of the n-qudit GHZ state $|\mathrm{GHZ}(n,d)\rangle$. The estimation protocols are based on the Clifford circuits supplemented by up to two canonical T gates. The mean square error $\langle\varepsilon^2\rangle$ for each data point is the average over 100 runs. The solid (dashed) lines are determined by interpolation. Results on the state $T|\mathrm{GHZ}(n,d)\rangle$ are also shown for comparison.

minimum slope is upper bounded by 3, so the dependence on the local dimension is insignificant. If more T gates are applied, then the ratio is even smaller. All these results are consistent with theoretical predictions.

For comparison, we also test our protocols on the state $T|\mathrm{GHZ}(n,d)\rangle$, where T is the canonical T gate acting on the first qudit, and the results are also shown in Fig. 4. Now the MSE gets smaller and we see a 'duality' between magic gates in state preparation and in measurements: the MSE is mainly determined by the total number of T gates. For example, Fig. 4(b) and Fig. 4(d) show similar behaviors; Fig. 4(c) and Fig. 4(e) show similar behaviors too. Additional simulation results on the median of means estimation and on cluster states can be found in Appendix H; the general conclusions are similar.

We then consider a more realistic case in which the input state is a depolarized GHZ state, i.e.,

$$\rho = p\mathbb{I}/D + (1-p)|\mathrm{GHZ}(n,d)\rangle\langle\mathrm{GHZ}(n,d)|, \quad (18)$$

where $p \in [0,1]$ is the noise probability, and we want to estimate its fidelity with $|\mathrm{GHZ}(n,d)\rangle$ using shadow estimation.

Figure 5 shows that our protocols give reasonably good estimators for fidelity using a relative small (10^4) number of samples. As more T gates are added to the Clifford circuit, the fluctuation (standard deviation) becomes smaller, meaning that the protocol has a better performance guarantee.This again confirms our theoretical prediction.

VIII. ESTIMATION OF QUADRATIC FUNCTIONS

Here we consider the estimation of quadratic functions based on Clifford circuits supplemented by a few T gates. In Appendix J, we derive an upper bound for the sample complexity of estimating quadratic functions of the form $\mathrm{tr}(\mathfrak{O}\rho\otimes\rho)$.

As a concrete example, consider the estimation of the purity of an m-qudit subsystem of interest. The associated observable is the swap operator $\mathbb{S}$: $\mathbb{S}|\psi\rangle\otimes|\phi\rangle = |\phi\rangle\otimes|\psi\rangle$ for all $|\psi\rangle,|\phi\rangle \in \mathcal{H}_D$. This task is tied to the probing of a Rényi entanglement entropy, which is of interest in many research fields. Suppose we employ the Clifford circuits supplemented by k T gates, then the sample complexity of this task is approximately $2d-1$ when $k=0$ and $8\sqrt{3}\gamma_{d,k}d^n/\varepsilon^2$ when $1 \le k \le n$, where $\gamma_{d,k}$ is defined in Eq. (16).

For local Clifford measurements, by contrast, the sample complexity is approximately $8(d^{2n}+d^n)/\varepsilon^2$. In addition, $\Omega\left(16d^{3n}/\varepsilon^2\right)$ samples are required to estimate the purity within precision ε using a traditional tomographic approach based on independent measurements [2]. So shadow estimation can reduce the resource cost by a factor of d^{2n}.

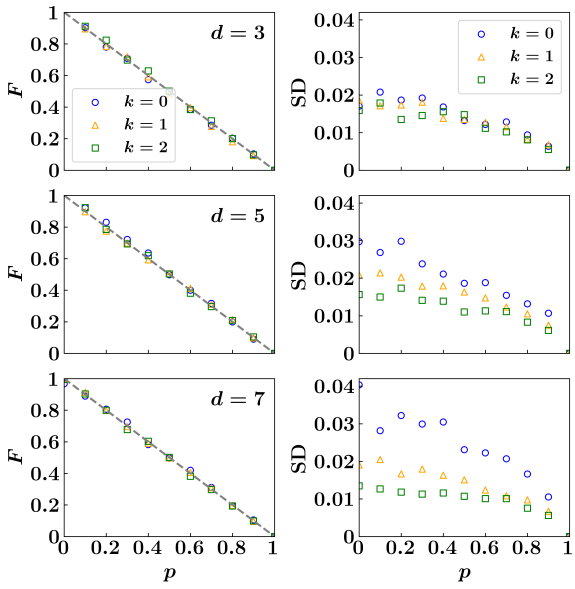


FIG. 5. Fidelity estimated using shadow estimation and the corresponding standard deviation. k denotes the number of canonical T gates in $\mathcal{E}\left(\{T_j\}_{j=1}^{k}\right)$. Here, the estimator is constructed from 10^4 samples, and the standard deviation (SD) is calculated from 100 independent runs. The grey lines in the left column show the true fidelity $F = 1 - p$.

IX. SUMMARY

In this work, we generalize the shadow estimation based on the Clifford group to all prime local dimensions, as well as considering qudit magic orbits as measurement primitives, which turns out to be a fruitful, yet little explored field of research.

Leveraging a little magic-state resource, our protocol resolves the dependence of sample complexity on local dimension and allows us to predict properties of large-scale, high-dimensional quantum states with comparable efficiency as in the qubit case. Since our protocol only require one layer of random Clifford circuit and one magic gate from the third Clifford hierarchy, it is promising to be implemented on NISQ devices.

Our work also provides valuable insights on the (qudit) Clifford group and the Clifford orbits, and highlights the power of a single magic gate in overcoming the gap between qudit and qubit systems in quantum information processing, which may have profound implications for various topics beyond shadow estimation.

[1] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020).
[2] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (2016) pp. 913–925.
[3] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Nat. Commun. **1**, 149 (2010).
[4] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, PRX Quantum **2**, 030348 (2021).
[5] D. E. Koh and S. Grewal, Quantum **6**, 776 (2022).
[6] A. Seif, Z.-P. Cian, S. Zhou, S. Chen, and L. Jiang, PRX Quantum **4**, 010303 (2023).
[7] H.-Y. Hu, R. LaRose, Y.-Z. You, E. Rieffel, and Z. Wang, arXiv: 2203.07263 (2022).
[8] C. Bertoni, J. Haferkamp, M. Hinsche, M. Ioannou, J. Eisert, and H. Pashayan, arXiv preprint arXiv:2209.12924 (2022).
[9] M. Ippoliti, Y. Li, T. Rakovszky, and V. Khemani, Phys. Rev. Lett. **130**, 230403 (2023).
[10] Z. Liu, Z. Hao, and H.-Y. Hu, arXiv preprint arXiv:2311.00695 (2023).
[11] H. C. Nguyen, J. L. Bönsel, J. Steinberg, and O. Gühne, Phys. Rev. Lett. **129**, 220502 (2022).
[12] L. Innocenti, S. Lorenzo, I. Palmisano, F. Albarelli, A. Ferraro, M. Paternostro, and G. M. Palma, PRX Quantum **4**, 040328 (2023).
[13] G. Struchalin, Y. A. Zagorovskii, E. Kovlakov, S. Straupe, and S. Kulik, PRX Quantum **2**, 010307 (2021).
[14] R. Stricker, M. Meth, L. Postler, C. Edmunds, C. Ferrie, R. Blatt, P. Schindler, T. Monz, R. Kueng, and M. Ringbauer, PRX Quantum **3**, 040310 (2022).
[15] R. Kaltenbaek, J. Lavoie, B. Zeng, S. D. Bartlett, and K. J. Resch, Nat. Phys. **6**, 850 (2010).
[16] M. Neeley, M. Ansmann, R. C. Bialczak, M. Hofheinz, E. Lucero, A. D. O'Connell, D. Sank, H. Wang, J. Wenner, A. N. Cleland, *et al.*, Science **325**, 722 (2009).
[17] M. S. Blok, V. V. Ramasesh, T. Schuster, K. O'Brien, J.-M. Kreikebaum, D. Dahlen, A. Morvan, B. Yoshida, N. Y. Yao, and I. Siddiqi, Phys. Rev. X **11**, 021010 (2021).
[18] Z. Jiang, Nat. Phys. **19**, 22 (2023).
[19] Y. Wang, Z. Hu, B. C. Sanders, and S. Kais, Front. Phys. **8**, 589504 (2020).
[20] M. Ringbauer, M. Meth, L. Postler, R. Stricker, R. Blatt, P. Schindler, and T. Monz, Nature Physics **18**, 1053 (2022).
[21] Y. Chi, J. Huang, Z. Zhang, J. Mao, Z. Zhou, X. Chen, C. Zhai, J. Bao, T. Dai, H. Yuan, M. Zhang, D. Dai, B. Tang, Y. Yang, Z. Li, Y. Ding, L. K. Oxenløwe, M. G. Thompson, J. L. O'Brien, Y. Li, Q. Gong, and J. Wang, Nat. Commun. **13**, 1166 (2022).
[22] M. Erhard, M. Krenn, and A. Zeilinger, Nat. Rev. Phys. **2**, 365 (2020).

[23] S. Choi, J. Choi, R. Landig, G. Kucsko, H. Zhou, J. Isoya, F. Jelezko, S. Onoda, H. Sumiya, V. Khemani, *et al.*, Nature **543**, 221 (2017).
[24] A. Cervera-Lierta, M. Krenn, A. Aspuru-Guzik, and A. Galda, Phys. Rev. Appl. **17**, 024062 (2022).
[25] H. Zhu, Phys. Rev. A **96**, 062336 (2017).
[26] Z. Webb, Quantum Info. Comput. **16**, 1379–1400 (2016).
[27] R. Kueng and D. Gross, Qubit stabilizer states are complex projective 3-designs (2015), poster at QIP 2013, arXiv:1510.02767.
[28] D. Gottesman and I. L. Chuang, Nature **402**, 390 (1999).
[29] M. Howard and J. Vala, Phys. Rev. A **86**, 022316 (2012).
[30] J. Preskill, Quantum **2**, 79 (2018).
[31] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).
[32] S. Bravyi and D. Gosset, Phys. Rev. Lett. **116**, 250501 (2016).
[33] H. Pashayan, O. Reardon-Smith, K. Korzekwa, and S. D. Bartlett, PRX Quantum **3**, 020361 (2022).
[34] R. Koenig and J. A. Smolin, J. Math. Phys. **55**, 122202 (2014).
[35] M. Heinrich, *On Stabiliser Techniques and Their Application to Simulation and Certification of Quantum Devices*, Ph.D. thesis, University of Cologne (2021).
[36] D. Gross, J. Math. Phys. **47**, 122107 (2006).
[37] G. B. Folland, *Harmonic analysis in phase space*, 122 (Princeton university press, 1989).
[38] A. Weil *et al.*, Acta math **111**, 14 (1964).
[39] D. Gottesman, in *NASA International Conference on Quantum Computing and Quantum Communications* (Springer, 1998) pp. 302–313.
[40] D. Gross, S. Nezami, and M. Walter, Commun. Math. Phys **385**, 1325 (2021).
[41] S. Anders and H. J. Briegel, Phys. Rev. A **73**, 022334 (2006).
[42] M. Nest, arXiv preprint arXiv:0811.0898 (2008).
[43] E. Van Den Berg, A simple method for sampling random Clifford operators (2021).
[44] S. Bravyi and D. Maslov, IEEE Trans. Inform. Theor. **67**, 4546 (2021).
[45] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, Quantum **3**, 181 (2019).
[46] E. M. Luks, F. Rákóczi, and C. R. Wright, Journal of Symbolic Computation **23**, 335 (1997).

Appendix A: Proof of Eq. (2)

Suppose in the jth run the random unitary is U and the measurement outcome is $\mathbf{b}$. Then the corresponding single-shot estimator $\hat{o}_j$ reads

$$\hat{o}_j := \mathrm{tr}\left(\mathfrak{O}\hat{\rho}_j\right) = \mathrm{tr}\left(\mathfrak{O}_0\hat{\rho}_j\right) = \mathrm{tr}\left[\mathfrak{O}_0 \cdot \mathcal{M}^{-1}(U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U)\right] = \mathrm{tr}\left[\mathcal{M}^{-1}(\mathfrak{O}_0)\cdot U^\dagger|\mathbf{b}\rangle\langle\mathbf{b}|U\right], \tag{A1}$$

where the last equality holds because the map $\mathcal{M}^{-1}$ is self-adjoint. The expectation value of $\hat{o}_j$ reads $\mathbb{E}[\hat{o}_j] = o = \mathrm{tr}(\mathfrak{O}\rho) = \mathrm{tr}(\mathfrak{O}_0\rho)$. The variance of $\hat{o}_j$ can be upper-bounded as follows,

$$\begin{aligned}\mathrm{Var}[\hat{o}_j] &= \mathbb{E}\left[|\hat{o}_j - o|^2\right] \le \mathbb{E}\left[|\hat{o}_j|^2\right] = \mathbb{E}\left[\left|\langle\mathbf{b}|U\mathcal{M}^{-1}(\mathfrak{O}_0)U^\dagger|\mathbf{b}\rangle\right|^2\right] \\ &= \mathbb{E}_{U\sim\mathcal{E}}\sum_{\mathbf{b}\in\mathbb{F}_d^n}\langle\mathbf{b}|U\rho U^\dagger|\mathbf{b}\rangle\cdot\left|\langle\mathbf{b}|U\mathcal{M}^{-1}(\mathfrak{O}_0)U^\dagger|\mathbf{b}\rangle\right|^2 \le \|\mathfrak{O}_0\|_{\mathrm{sh}}^2,\end{aligned} \tag{A2}$$

where the last inequality follows from the definition of the shadow norm in Eq. (3). Note that this derivation is applicable even if $\mathfrak{O}$ is not Hermitian.

By definition the empirical mean $\hat{o}$ is the average of N independent single-shot estimators. Therefore,

$$\langle\varepsilon^2\rangle := \mathrm{Var}[\hat{o}] = \frac{1}{N^2}\sum_{j=1}^N \mathrm{Var}[\hat{o}_j] \le \frac{\|\mathfrak{O}_0\|_{\mathrm{sh}}^2}{N}, \tag{A3}$$

which confirms Eq. (2).

Appendix B: Heisenberg-Weyl group and Clifford group

In this appendix, we provide a bit more details on the Heisenberg-Weyl group and Clifford group, assuming that the local dimension d is an odd prime. Recall that the HW group $\mathcal{W}(d)$ for a single qudit is generated by the two operators Z and X defined in Eq. (4), that is,

$$\mathcal{W}(d) := \langle Z, X\rangle = \left\{\omega_d^j Z^k X^l \,|\, j,k,l\in\mathbb{F}_d\right\}, \tag{B1}$$

where $\omega_d = e^{\frac{2\pi\mathrm{i}}{d}}$. Up to phase factors, the elements in $\mathcal{W}(d)$, known as Weyl operators, can be labeled by vectors in $\mathbb{F}_d^2$ [36]. Given $u = (p,q)\in\mathbb{F}_d^2$, define

$$W_u = W(p,q) := \chi(-2^{-1}pq)Z^pX^q, \tag{B2}$$

where $\chi(r) = \omega_d^r$. Then $\mathcal{W}(n,d) = \left\{\omega_d^j W_u \,|\, j\in\mathbb{F}_d, u\in\mathbb{F}_d^2\right\}$. Let u, v be two vectors in $\mathbb{F}_d^2$; then W_u commute with W_v iff u and v are linearly dependent. The n-qudit HW group $\mathcal{W}(n,d)$ is the tensor product of n copies of $\mathcal{W}(d)$. Up to phase factors, n-qudit Weyl operators can be labeled by vectors in $\mathbb{F}_d^{2n}$. Given $\mathbf{u} = (u_1^z, u_2^z, \ldots, u_n^z, u_1^x, u_2^x, \ldots, u_n^x)\in\mathbb{F}_d^{2n}$, define

$$W_{\mathbf{u}} := W_{u_1}\otimes W_{u_2}\otimes\cdots\otimes W_{u_n}, \tag{B3}$$

where $u_j = (u_j^z, u_j^x)$. Then $\mathcal{W}(n,d) = \left\{\omega_d^j W_{\mathbf{u}} \,|\, j\in\mathbb{F}_d, \mathbf{u}\in\mathbb{F}_d^{2n}\right\}$.

To better understand the structure of the HW group, we need to introduce a symplectic structure in $\mathbb{F}_d^{2n}$. Denote by $[\mathbf{u},\mathbf{v}]$ the symplectic product defined as follows,

$$[\mathbf{u},\mathbf{v}] := \mathbf{u}^\top J\mathbf{v}, \quad J = \begin{pmatrix} 0_n & \mathbb{I}_n \\ -\mathbb{I}_n & 0_n\end{pmatrix}. \tag{B4}$$

Then the composition and commutation relations of two Weyl operators $W_{\mathbf{u}}$ and $W_{\mathbf{v}}$ are determined by this symplectic product,

$$W_{\mathbf{u}}W_{\mathbf{v}} = \omega_d^{[\mathbf{u},\mathbf{v}]/2}W_{\mathbf{u}+\mathbf{v}},, \quad W_{\mathbf{u}}W_{\mathbf{v}} = \omega_d^{[\mathbf{u},\mathbf{v}]}W_{\mathbf{v}}W_{\mathbf{u}}, \quad \mathbf{u},\mathbf{v}\in\mathbb{F}_d^{2n}. \tag{B5}$$

In addition, this symplectic product defines the symplectic group $\mathrm{Sp}(2n,d)$, which is composed of all $2n\times 2n$ matrices on $\mathbb{F}_d$ that preserve the symplectic product, that is,

$$\mathrm{Sp}(2n,d) := \left\{S \in \mathbb{F}_d^{2n\times 2n} \mid S^\top J S = J\right\}. \tag{B6}$$

Elements in $\mathrm{Sp}(2n,d)$ are called symplectic matrices or symplectic transformations. If S is a symplectic matrix, then $[S\mathbf{u}, S\mathbf{v}] = [\mathbf{u},\mathbf{v}]$ for all $\mathbf{u},\mathbf{v}\in\mathbb{F}_d^{2n}$.

The Clifford group $\mathrm{Cl}(n,d)$ is the normalizer of the HW group $\mathcal{W}(n,d)$, that is,

$$\mathrm{Cl}(n,d) := \left\{U \in U(\mathcal{H}_d^{\otimes n}) \,|\, U\mathcal{W}(n,d)U^\dagger = \mathcal{W}(n,d)\right\}. \tag{B7}$$

By definition the commutation relations of Weyl operators are invariant under Clifford transformations. So every Clifford unitary induces a symplectic transformation on the space $\mathbb{F}_d^{2n}$. Conversely, for any symplectic transformation $M\in\mathrm{Sp}(2n,d)$, there exists a unitary operator $\mu(M)$ such that [36]

$$\mu(M) W_{\mathbf{u}} \mu(M)^\dagger = W_{M\mathbf{u}} \quad \forall \mathbf{u}\in\mathbb{F}_d^{2n}, \tag{B8}$$

where μ is called the Weil or metaplectic representation of the symplectic group [37, 38]. Note that this conclusion relies on the assumption that d is an odd prime and does not hold when $d=2$. Up to an overall phase factor, any Clifford unitary $C\in\mathrm{Cl}(n,d)$ is of the form

$$C = W_{\mathbf{a}}\mu(M), \quad \mathbf{a}\in\mathbb{F}_d^{2n}, \quad M\in\mathrm{Sp}(2n,d). \tag{B9}$$

In addition, we have

$$C W_{\mathbf{u}} C^\dagger = W_{\mathbf{a}}\mu(M) W_{\mathbf{u}} \mu(M)^\dagger W_{\mathbf{a}}^\dagger = W_{\mathbf{a}} W_{M\mathbf{u}} W_{\mathbf{a}}^\dagger = \chi\left([\mathbf{a}, M\mathbf{u}]\right) W_{M\mathbf{u}}. \tag{B10}$$

Incidentally, the Clifford group $\mathrm{Cl}(n,d)$ can be generated by the following unitary operators acting on individual qudits or individual pairs of qudits [35, 39]:

$$F := \frac{1}{\sqrt{d}} \sum_{x,y\in\mathbb{F}_d} \chi(x\cdot y)|x\rangle\langle y|, \qquad M(\nu) := \sum_{x\in\mathbb{F}_d} |\nu x\rangle\langle x|, \tag{B11}$$

$$S(\nu) := \sum_{x\in\mathbb{F}_d} \chi(2^{-1}\nu x^2)|x\rangle\langle x|, \qquad CX := \sum_{(x,y)\in\mathbb{F}_d^2} |x, x+y\rangle\langle x,y|, \tag{B12}$$

$$Z = \sum_{x\in\mathbb{F}_d} \chi(x)|x\rangle\langle x|, \qquad X = \sum_{x\in\mathbb{F}_d} |x+1\rangle\langle x|, \tag{B13}$$

where ν is a primitive element in $\mathbb{F}_d^\times$.

Appendix C: Stabilizer codes and stabilizer states

A stabilizer group $\mathcal{S}$ is an Abelian subgroup of $\mathcal{W}(n,d)$ that does not contain the operator $\omega_d\mathbb{I}$. Given a stabilizer group $\mathcal{S}$ generated by m ($1\le m\le n$) Weyl operators. The stabilizer projector $P_{\mathcal{S}}$ is defined as the projector onto the subspace $\mathcal{C}_{\mathcal{S}}$ stabilized by $\mathcal{S}$

$$\mathcal{C}_{\mathcal{S}} := \left\{|\psi\rangle \in \left(\mathbb{C}^d\right)^{\otimes n} \,|\, g|\psi\rangle = |\psi\rangle,\ \forall g\in\mathcal{S}\right\}, \tag{C1}$$

thus

$$P_{\mathcal{S}} = \frac{1}{|\mathcal{S}|}\sum_{g\in\mathcal{S}} g. \tag{C2}$$

It's not difficult to see that the rank K of $P_{\mathcal{S}}$ is d^{n-m}.

If the stabilizer group $\mathcal{S}$ has the maximum order of d^n, then the stabilizer code $\mathcal{C}_{\mathcal{S}}$ is one-dimensional and can be represented by a normalized state, known as a stabilizer state. A stabilizer state can be uniquely determined by its generating set, which contains n Weyl operators.

Appendix D: Local Clifford measurements

Before proving Proposition 1 and Theorem 1 we need to introduce some auxiliary notation and results. Let $\mathbf{u}, \mathbf{v}$ be two vectors in $\mathbb{F}_d^{2n}$. The weight $|\mathbf{u}|$ of $\mathbf{u}$ is defined as $|\mathbf{u}| := \big|\{j \,|\, (u_j^z, u_j^x) \neq (0,0)\}\big|$. As a generalization, we define

$$\begin{aligned} |\mathbf{u} \vee \mathbf{v}| &:= \big|\{j \,|\, (u_j^z, u_j^x) \neq (0,0) \text{ or } (v_j^z, v_j^x) \neq (0,0)\}\big|, \\ |\mathbf{u} \wedge \mathbf{v}| &:= \big|\{j \,|\, (u_j^z, u_j^x) \neq (0,0) \text{ and } (v_j^z, v_j^x) \neq (0,0)\}\big|. \end{aligned} \tag{D1}$$

Evidently, we have

$$|\mathbf{u}| + |\mathbf{v}| - |\mathbf{u} \wedge \mathbf{v}| = |\mathbf{u} \vee \mathbf{v}|. \tag{D2}$$

Let ϕ be a real phase, then the weight of a Weyl operator of the form $\mathrm{e}^{\mathrm{i}\phi} W_{\mathbf{u}}$ is defined as the weight of $\mathbf{u}$; the Weyl operator is m local if it has weight m, that is, $|\mathbf{u}| = m$. Two Weyl operators $W_{\mathbf{u}}, W_{\mathbf{v}}$ are locally commutative, denoted by $W_{\mathbf{u}} \bowtie W_{\mathbf{v}}$, if W_{u_j} and W_{v_j} are commutative for $j = 1, 2, \ldots, n$, where $u_j = (u_j^z, u_j^x)$ and $v_j = (v_j^z, v_j^x)$. In this case, we also use the notation $\mathbf{u} \bowtie \mathbf{v}$ to denote the induced relation on $\mathbb{F}_d^{2n}$. By definition $\mathbf{u} \bowtie \mathbf{v}$ iff u_j, v_j are linearly dependent for $j = 1, 2, \ldots, n$. In addition, we write $\mathbf{u} \triangleright \mathbf{v}$ if u_j is proportional to v_j for $j = 1, 2, \ldots, n$. If $\mathbf{u} \triangleright \mathbf{v}$, then $\mathbf{u} \bowtie \mathbf{v}$, but the converse is not guaranteed in general. If $\mathbf{u} \bowtie \mathbf{v}$, then we can deduce the following relations,

$$|\mathbf{u} \vee \mathbf{v}| = \min\{|\mathbf{s}| : \mathbf{s} \in \mathbb{F}_d^{2n}, \mathbf{u} \triangleright \mathbf{s}, \mathbf{v} \triangleright \mathbf{s}\}, \quad |\mathbf{u} \wedge \mathbf{v}| = \max\{|\mathbf{s}| : \mathbf{s} \in \mathbb{F}_d^{2n}, \mathbf{s} \triangleright \mathbf{u}, \mathbf{s} \triangleright \mathbf{v}\}. \tag{D3}$$

Define

$$\mathcal{V}_1^* := \{(0,1), \cdots, (d-1,1), (1,0)\}. \tag{D4}$$

$$\mathcal{V}_n^* := \big\{(s_{z,1}, s_{z,2}, ..., s_{z,n}; s_{x,1}, s_{x,2}, ..., s_{x,n}) \,|\, (s_j^z, s_j^x) \in \mathcal{V}_1^* \ \forall 1 \leq j \leq n\big\}. \tag{D5}$$

Lemma 1. *Suppose d is an odd prime, $\mathbf{b} \in \mathbb{F}_d^n$, and $\mathbf{u}, \mathbf{v} \in \mathbb{F}_d^{2n}$. Then we have*

$$\mathbb{E}_{U \sim \mathrm{Cl}(1,d)^{\otimes n}} U^\dagger |\mathbf{b}\rangle\langle\mathbf{b}| U \langle\mathbf{b}| U W_{\mathbf{u}} U^\dagger |\mathbf{b}\rangle\langle\mathbf{b}| U W_{\mathbf{v}} U^\dagger |\mathbf{b}\rangle^* = \begin{cases} d^{-n}(d+1)^{-|\mathbf{u}\vee\mathbf{v}|} W_{\mathbf{u}} W_{\mathbf{v}}^\dagger & \text{if } \mathbf{u} \bowtie \mathbf{v}, \\ 0 & \text{otherwise.} \end{cases} \tag{D6}$$

Proof. Thanks to the tensor structure of U, $W_{\mathbf{u}}$, $W_{\mathbf{v}}$, and $|\mathbf{b}\rangle$, it suffices to prove Eq. (D6) in the case $n = 1$. Then we can rewrite the LHS of Eq. (D6) as follows,

$$\mathbb{E}_{|\Psi\rangle \sim \mathrm{Stab}(1,d)} |\Psi\rangle\langle\Psi| \langle\Psi|W_{\mathbf{u}}|\Psi\rangle \langle\Psi|W_{\mathbf{v}}|\Psi\rangle^* = \frac{1}{d(d+1)} \sum_{|\Psi\rangle \in \mathrm{Stab}(1,d)} |\Psi\rangle\langle\Psi| \langle\Psi|W_{\mathbf{u}}|\Psi\rangle \langle\Psi|W_{\mathbf{v}}|\Psi\rangle^*. \tag{D7}$$

If $W_{\mathbf{u}}, W_{\mathbf{v}}$ do not commute, then they cannot belong to the stabilizer group of any stabilizer state simultaneously. Therefore, $\langle\Psi|W_{\mathbf{u}}|\Psi\rangle\langle\Psi|W_{\mathbf{v}}|\Psi\rangle^* = 0$, which implies Eq. (D6). If $W_{\mathbf{u}} = W_{\mathbf{v}} = \mathbb{I}$, then Eq. (D6) holds because $\mathbb{E}_{|\Psi\rangle \sim \mathrm{Stab}(1,d)} |\Psi\rangle\langle\Psi| = \mathbb{I}/d$.

Next, suppose $W_{\mathbf{u}}$ and $W_{\mathbf{v}}$ commute, and at least one of them is not proportional to the identity operator. Then we can find a Clifford unitary $C \in \mathrm{Cl}(d)$ such that

$$C^\dagger W_{\mathbf{u}} C = \omega_d^{r_1} Z^{s_1}, \quad C^\dagger W_{\mathbf{v}} C = \omega_d^{r_2} Z^{s_2}, \tag{D8}$$

where $r_1, s_1, r_2, s_2 \in \mathbb{F}_d$. Therefore,

$$\begin{aligned} &\sum_{|\Psi\rangle \in \mathrm{Stab}(1,d)} |\Psi\rangle\langle\Psi| \langle\Psi|W_{\mathbf{u}}|\Psi\rangle \langle\Psi|W_{\mathbf{v}}|\Psi\rangle^* = \sum_{|\Psi\rangle \in \mathrm{Stab}(1,d)} |\Psi\rangle\langle\Psi| \langle\Psi|C\omega_d^{r_1} Z^{s_1} C^\dagger|\Psi\rangle \langle\Psi|C\omega_d^{r_2} Z^{s_2} C^\dagger|\Psi\rangle^* \\ &= \sum_{|\Psi\rangle \in \mathrm{Stab}(1,d)} C|\Psi\rangle\langle\Psi|C^\dagger \langle\Psi|\omega_d^{r_1} Z^{s_1}|\Psi\rangle \langle\Psi|\omega_d^{r_2} Z^{s_2}|\Psi\rangle^* = \sum_{b \in \mathbb{F}_d} C|b\rangle\langle b|C^\dagger \langle b|\omega_d^{r_1} Z^{s_1}|b\rangle \langle b|\omega_d^{r_2} Z^{s_2}|b\rangle^* \\ &= C\omega_d^{r_1} Z^{s_1} \omega_d^{-r_2} Z^{-s_2} C^\dagger = W_{\mathbf{u}} W_{\mathbf{v}}^\dagger. \end{aligned} \tag{D9}$$

Together with Eq. (D7), this equation implies Eq. (D6) and completes the proof of Lemma 1. □

For local Clifford measurements, the reconstruction map in $\mathcal{M}^{-1}$ Eq. (5) can be expressed as

$$\mathcal{M}^{-1} = \left(\mathcal{D}^{-1}_{1/(d+1)}\right) d^{\otimes n}, \tag{D10}$$

where $\mathcal{D}_y^{-1}(\cdot)$ is the inverse of the single-qudit depolarizing channel $\mathcal{D}_y(\cdot) = y\cdot + (1-y)\frac{\mathrm{tr}(\cdot)}{d}\mathbb{I}$. Note that $\mathcal{D}^{-1}_{1/(d+1)}(\mathfrak{O}) = (d+1)\mathfrak{O} - \mathrm{tr}(\mathfrak{O})\mathbb{I}$ for any linear operator on $\mathcal{H}_d$. If $\mathfrak{O} = W_u$ is a Weyl operator with $u \in \mathbb{F}_d^2$, then

$$\mathcal{D}^{-1}_{1/(d+1)}(W_u) = \begin{cases} W_u & \text{if } W_u \propto \mathbb{I}, \\ (d+1)W_u & \text{otherwise.} \end{cases} \tag{D11}$$

1. Proof of Proposition 1

Proof of Proposition 1. If $d = 2$, then Proposition 1 holds according to Lemma 3 in Ref. [1].

Next, suppose d is an odd prime and $\mathfrak{O}$ is an m-local Weyl operator. Then $\mathfrak{O}$ has the form $\mathfrak{O} = \mathrm{e}^{\mathrm{i}\phi} W_{\mathbf{u}}$, where ϕ is a real phase and $\mathbf{u} \in \mathbb{F}_d^{2n}$ with $|\mathbf{u}| = m$. By virtue of Eqs. (3) and (D10) we can deduce that

$$\begin{aligned} \|\mathfrak{O}\|_{\mathrm{sh}}^2 = \|W_{\mathbf{u}}\|_{\mathrm{sh}}^2 &= \max_{\sigma} \sum_{\mathbf{b}\in\mathbb{F}_d^n} \mathbb{E}_{U\sim \mathrm{Cl}(d)^{\otimes n}} \langle \mathbf{b}|U\sigma U^\dagger|\mathbf{b}\rangle \cdot |\langle \mathbf{b}|U(\mathcal{D}^{-1}_{1/(d+1)})^{\otimes n}(W_{\mathbf{u}})U^\dagger|\mathbf{b}\rangle|^2 \\ &= (d+1)^{2m} \max_{\sigma} \sum_{\mathbf{b}\in\mathbb{F}_d^n} \mathbb{E}_{U\sim \mathrm{Cl}(d)^{\otimes n}} \langle \mathbf{b}|U\sigma U^\dagger|\mathbf{b}\rangle \cdot |\langle \mathbf{b}|U(W_{\mathbf{u}})U^\dagger|\mathbf{b}\rangle|^2 \\ &= d^n(d+1)^{2m} \times \frac{1}{d^n(d+1)^m} \max_{\sigma} \mathrm{tr}\left(\sigma W_{\mathbf{u}} W_{\mathbf{u}}^\dagger\right) = (d+1)^m, \end{aligned} \tag{D12}$$

which confirms Proposition 1. Here the second equality follows from Eq. (D11) and the fact that $W_{\mathbf{u}}$ is m-local, and the last equality follows from Lemma 1. □

2. Proof of Theorem 1

Proof of Theorem 1. If $d = 2$, then Eq. (7) follows from Proposition 3 in Ref. [1], so it remains to consider the case in which d is an odd prime.

We expand $\tilde{\mathfrak{O}}$ in the Weyl basis as $\tilde{\mathfrak{O}} = \sum_{\mathbf{u}} \alpha_{\mathbf{u}} W_{\mathbf{u}}$, where $\mathbf{u} \in \mathbb{F}_d^{2m}$. According to Eq. (D11), we have

$$(\mathcal{D}^{-1}_{1/(d+1)})^{\otimes m}(\tilde{\mathfrak{O}}) = \sum_{\mathbf{u}\in\mathbb{F}_d^{2m}} (d+1)^{|\mathbf{u}|} \alpha_{\mathbf{u}} W_{\mathbf{u}}. \tag{D13}$$

In conjunction with Eqs. (3) and (D10) we can deduce that

$$\begin{aligned} \|\mathfrak{O}\|_{\mathrm{sh}}^2 = \|\tilde{\mathfrak{O}}\|_{\mathrm{sh}}^2 &= \max_{\sigma} \mathbb{E}_{U\sim \mathrm{Cl}(1,d)^{\otimes m}} \sum_{\mathbf{b}\in\mathbb{F}_d^m} \langle \mathbf{b}|U\sigma U^\dagger|\mathbf{b}\rangle \left|\langle \mathbf{b}|U(\mathcal{D}^{-1}_{1/(d+1)})^{\otimes m}(\tilde{\mathfrak{O}})U^\dagger|\mathbf{b}\rangle\right|^2 \\ &= \max_{\sigma} \sum_{\mathbf{b}\in\mathbb{F}_d^m} \sum_{\mathbf{u},\mathbf{v}\in\mathbb{F}_d^{2m}} (d+1)^{|\mathbf{u}|+|\mathbf{v}|} \alpha_{\mathbf{u}}\alpha_{\mathbf{v}}^* \,\mathrm{tr}\left[\sigma \cdot \mathbb{E}_{U\sim \mathrm{Cl}(1,d)^{\otimes m}} U^\dagger|\mathbf{b}\rangle\langle \mathbf{b}|U\langle \mathbf{b}|UW_{\mathbf{u}}U^\dagger|\mathbf{b}\rangle\langle \mathbf{b}|UW_{\mathbf{v}}U^\dagger|\mathbf{b}\rangle^*\right] \\ &= \max_{\sigma} \sum_{\mathbf{u},\mathbf{v}\in\mathbb{F}_d^{2m},\mathbf{u}\bowtie\mathbf{v}} \frac{(d+1)^{|\mathbf{u}|+|\mathbf{v}|}}{(d+1)^{|\mathbf{u}\vee\mathbf{v}|}} \alpha_{\mathbf{u}}\alpha_{\mathbf{v}}^* \,\mathrm{tr}\left(\sigma W_{\mathbf{u}} W_{\mathbf{v}}^\dagger\right) = \left\| \sum_{\mathbf{u},\mathbf{v}\in\mathbb{F}_d^{2m},\mathbf{u}\bowtie\mathbf{v}} \frac{(d+1)^{|\mathbf{u}|+|\mathbf{v}|}}{(d+1)^{|\mathbf{u}\vee\mathbf{v}|}} \alpha_{\mathbf{u}}\alpha_{\mathbf{v}}^* W_{\mathbf{u}} W_{\mathbf{v}}^\dagger \right\| \\ &= \left\| \sum_{\mathbf{s}\in\mathcal{V}_n^*} \sum_{\mathbf{u},\mathbf{v}\rhd\mathbf{s}} \frac{(d+1)^{|\mathbf{u}|+|\mathbf{v}|}}{(d+1)^m} \alpha_{\mathbf{u}}\alpha_{\mathbf{v}}^* W_{\mathbf{u}} W_{\mathbf{v}}^\dagger \right\| \le \frac{1}{(d+1)^m} \sum_{\mathbf{s}\in\mathcal{V}_m^*} \left\| \sum_{\mathbf{u},\mathbf{v}\rhd\mathbf{s}} (d+1)^{|\mathbf{u}|+|\mathbf{v}|} \alpha_{\mathbf{u}}\alpha_{\mathbf{v}}^* W_{\mathbf{u}} W_{\mathbf{v}}^\dagger \right\| \\ &= \frac{1}{(d+1)^m} \sum_{\mathbf{s}\in\mathcal{V}_m^*} \left\| \sum_{\mathbf{u}\rhd\mathbf{s}} (d+1)^{|\mathbf{u}|} \alpha_{\mathbf{u}} W_{\mathbf{u}} \right\|^2 \le \frac{1}{(d+1)^m} \sum_{\mathbf{s}\in\mathcal{V}_m^*} \left| \sum_{\mathbf{u}\rhd\mathbf{s}} (d+1)^{|\mathbf{u}|} \alpha_{\mathbf{u}} \right|^2. \end{aligned} \tag{D14}$$

Here the fourth equality follows from Lemma 1 and the sixth equaity follows from the fact that

$$|\{\mathbf{s} \in \mathcal{V}_m^* \,|\, \mathbf{u}, \mathbf{v} \triangleright \mathbf{s}\}| = (d+1)^{m-|\mathbf{u}\vee\mathbf{v}|}. \tag{D15}$$

If $\tilde{\mathfrak{O}} = W_{\mathbf{q}}$ is an m-local Weyl operator, say, $\tilde{\mathfrak{O}} = W_{\mathbf{q}}$ with $\mathbf{q} \in \mathbb{F}_d^{2m}$ and $|\mathbf{q}| = m$, then $\alpha_{\mathbf{u}} = 1$ if $\mathbf{u} = \mathbf{q}$ and $\alpha_{\mathbf{u}} = 0$ otherwise, so both inequalities in Eq. (D14) are saturated and we recover Proposition 1.

Next, apply the Cauchy-Schwartz inequality to Eq. (D14) we can deduce that

$$\begin{aligned}\|\mathfrak{O}\|_{\mathrm{sh}}^2 &\leq \frac{1}{(d+1)^m} \sum_{\mathbf{s}\in\mathcal{V}_m^*} \left|\sum_{\mathbf{u}\triangleright\mathbf{s}} (d+1)^{|\mathbf{u}|}\alpha_{\mathbf{u}}\right|^2 \leq \frac{1}{(d+1)^m} \sum_{\mathbf{s}\in\mathcal{V}_m^*} \left(\sum_{\mathbf{u}\triangleright\mathbf{s}} (d+1)^{|\mathbf{u}|}\right)\left(\sum_{\mathbf{u}\triangleright\mathbf{s}} (d+1)^{|\mathbf{u}|}|\alpha_{\mathbf{u}}|^2\right) \\ &= d^{2m} \sum_{\mathbf{s}\in\mathcal{V}_m^*} \sum_{\mathbf{u}\triangleright\mathbf{s}} (d+1)^{|\mathbf{u}|-m}|\alpha_{\mathbf{u}}|^2 = d^{2m} \sum_{\mathbf{u}\in\mathbb{F}_d^{2m}} |\alpha_{\mathbf{u}}|^2 = d^m \|\tilde{\mathfrak{O}}\|_2^2,\end{aligned} \tag{D16}$$

which confirms Theorem 1. In deriving the above equalities we have taken into account the following facts,

$$\sum_{\mathbf{u}\triangleright\mathbf{s}} (d+1)^{|\mathbf{u}|} = \sum_{j=0}^{m} \binom{m}{j} (d-1)^j (d+1)^j = d^{2m}, \quad |\{\mathbf{s} \in \mathcal{V}_m^* \,|\, \mathbf{u} \triangleright \mathbf{s}\}| = (d+1)^{m-|\mathbf{u}|}. \tag{D17}$$

□

Appendix E: The third moment operator

Recently, Gross, Nezami and Walter (GNW) developed a variant of Schur-Weyl duality for the Clifford group [40], which states that when $n \geq t-1$ the commutant of $\mathrm{Cl}(n,d)^{\otimes t}$ is spanned by the set of operators $R(\mathcal{T})$ for $\mathcal{T} \in \Sigma_{t,t}(d)$, where the operator $R(\mathcal{T})$ is defined as

$$r(\mathcal{T}) := \sum_{(\mathbf{x};\mathbf{y})\in\mathcal{T}} |\mathbf{x}\rangle\langle\mathbf{y}|, \quad R(\mathcal{T}) := r(\mathcal{T})^{\otimes n} \tag{E1}$$

for each $\mathcal{T} \leq \mathbb{F}_d^{2t}$, and $\Sigma_{t,t}(d)$ is called the set of stochastic Lagrangian subspaces. For the purpose of our discussion, let $t = 3$, d be an odd prime, and $\Sigma(d) := \Sigma_{3,3}(d)$. The cardinality of $\Sigma(d)$ is $2d+2$.

Let $S_3 := \{\mathbb{1}, \zeta, \zeta^2, \tau_{12}, \tau_{23}, \tau_{13}\}$ be the third order symmetric group, where $\mathbb{1}$ is the identity, ζ is the cyclic permutation and τ_{ij} is the transposition that interchanges i and j. For all $O \in S_3$,

$$\mathcal{T}_O := \left\{(O\mathbf{x};\mathbf{x}) \,|\, \mathbf{x} \in \mathbb{F}_d^3\right\} \tag{E2}$$

is an element in $\Sigma(d)$. Define

$$\mathscr{T}_{\mathrm{sym}} := \{\mathcal{T}_O \,|\, O \in S_3\}, \quad \mathscr{T}_{\mathrm{ns}} := \Sigma(d) \backslash \mathscr{T}_{\mathrm{sym}}. \tag{E3}$$

Specifically, when $d = 2$, $\Sigma(d) = \mathscr{T}_{\mathrm{sym}}$.

Define the third moment operator of the Clifford orbit generated from any fiducial state $|\Psi\rangle \in \mathcal{H}_D$ as

$$Q\left(\mathrm{orb}(\Psi)\right) := \mathbb{E}_{U\sim\mathrm{Cl}(n,d)} \left[U^\dagger|\Psi\rangle\langle\Psi|U\right]^{\otimes 3}. \tag{E4}$$

Evidently, $Q\left(\mathrm{orb}(\Psi)\right)$ belongs to the commutant of $\mathrm{Cl}(n,d)^{\otimes 3}$, and can be expanded as a linear combination of $\{R(\mathcal{T})\}_{\mathcal{T}\in\Sigma(d)}$. Specifically, when $|\Psi\rangle$ is a stabilizer state, $Q\left(\mathrm{orb}(\Psi)\right)$ takes the following simple form [40]

$$Q(n,d,3) := \mathbb{E}_{\mathrm{Stab}(n,d)} \left[|S\rangle\langle S|^{\otimes 3}\right] = \frac{1}{D(D+1)(D+d)} \sum_{\mathcal{T}\in\Sigma(d)} R(\mathcal{T}), \tag{E5}$$

where $\mathrm{Stab}(n,d)$ denotes the ensemble of n-qudit stabilizer states. We also define the 'normalized third moment operator' as

$$\bar{Q}\left(\mathrm{orb}(\Psi)\right) := \pi_{[3]} Q\left(\mathrm{orb}(\Psi)\right), \tag{E6}$$

where

$$\pi_{[3]} = \mathrm{tr}\big(P_{[3]}\big) = \binom{D+2}{3}, \tag{E7}$$

with $P_{[3]}$ being the projector onto the tripartite symmetric subspace in $\mathcal{H}_D^{\otimes 3}$. The third moment operator is closely related with the shadow norm. Suppose the measurement primitive corresponds to a Clifford orbit based on $|\Psi\rangle \in \mathcal{H}_D$. For any linear operator $\mathfrak{O}$, define

$$\bar{\mathcal{Q}}_\Psi(\mathfrak{O}) := \mathrm{tr}_{BC}\left[\bar{Q}\left(\mathrm{orb}(\Psi)\right)\left(\mathbb{I}\otimes\mathfrak{O}\otimes\mathfrak{O}^\dagger\right)\right], \tag{E8}$$

then the shadow norm of $\mathfrak{O}$ reads

$$\|\mathfrak{O}\|_{\mathrm{sh}}^2 = \frac{6(D+1)}{D+2}\|\bar{\mathcal{Q}}_\Psi(\mathfrak{O})\|, \tag{E9}$$

This is the case for both scenarios discussed in Sec. V and Sec. VI. When there are more than one Clifford orbits in the measurement primitive, Eq. (E9) can be generalized by taking the average over all orbits.

Appendix F: Proof of Theorem 2

Proof of Theorem 2. We focus on the proof of Eq. (9), and Eqs. (10) and (11) follows by a similar reasoning. For a detailed derivation, see our companion paper. In the case of stabilizer measurements, define

$$\bar{Q}(n,d,3) := \pi_{[3]}Q(n,d,3), \quad \bar{\mathcal{Q}}_{n,d}(\mathfrak{O}) := \mathrm{tr}_{BC}\left[\bar{Q}(n,d,3)(\mathbb{I}\otimes\mathfrak{O}\otimes\mathfrak{O}^\dagger)\right], \tag{F1}$$

then

$$\|\mathfrak{O}\|_{\mathrm{sh}}^2 = \frac{6(D+1)}{D+2}\|\bar{\mathcal{Q}}_{n,d}(\mathfrak{O})\|. \tag{F2}$$

First consider the case $d=2$, the set of stabilizer states forms a 3-design [25], which means $\bar{Q}(n,d,3) = P_{[3]}$. Then for any linear operator $\mathfrak{O}\in\mathcal{H}_D$,

$$6\,\mathrm{tr}_{BC}[P_{[3]}(\mathbb{I}\otimes\mathfrak{O}\otimes\mathfrak{O}^\dagger)] = \mathfrak{O}\mathfrak{O}^\dagger + \mathfrak{O}^\dagger\mathfrak{O} + \mathrm{tr}\big(\mathfrak{O}^\dagger\mathfrak{O}\big)\mathbb{I} + |\,\mathrm{tr}\,\mathfrak{O}|^2\mathbb{I} + \mathrm{tr}(\mathfrak{O})\mathfrak{O}^\dagger + \mathrm{tr}\big(\mathfrak{O}^\dagger\big)\mathfrak{O}. \tag{F3}$$

Consider the shadow norm of $\mathfrak{O}_0$. Since $\mathfrak{O}_0$ is traceless, we have

$$\|\mathfrak{O}_0\|_2^2 \le 6\|\bar{\mathcal{Q}}_{n,d}(\mathfrak{O}_0)\| = 6\|\,\mathrm{tr}_{BC}[P_{[3]}(\mathbb{I}\otimes\mathfrak{O}_0\otimes\mathfrak{O}_0^\dagger)]\| = \|\mathfrak{O}_0\|_2^2 + \|\mathfrak{O}_0\mathfrak{O}_0^\dagger + \mathfrak{O}_0^\dagger\mathfrak{O}_0\| = \|\mathfrak{O}_0\|_2^2 + 2\|\mathfrak{O}_0\|^2, \tag{F4}$$

which implies Eq. (9).

From now on, assume that d is an odd prime, then

$$\begin{aligned} 6\|\bar{\mathcal{Q}}_{n,d}(\mathfrak{O}_0)\| &= 6\max_\sigma \mathrm{tr}\left[\bar{Q}(n,d,3)(\sigma\otimes\mathfrak{O}_0\otimes\mathfrak{O}_0^\dagger)\right] \ge \frac{6}{D}\,\mathrm{tr}\left[\bar{Q}(n,d,3)(\mathbb{I}\otimes\mathfrak{O}_0\otimes\mathfrak{O}_0^\dagger)\right] \\ &= \frac{D+2}{D}\,\mathrm{tr}\Big[(\mathbb{I}+\mathbb{S})(\mathfrak{O}_0\otimes\mathfrak{O}_0^\dagger)\Big] = \frac{D+2}{D}\,\mathrm{tr}\Big(\mathfrak{O}_0\mathfrak{O}_0^\dagger\Big) = \frac{D+2}{D}\|\mathfrak{O}_0\|_2^2, \end{aligned} \tag{F5}$$

where $\mathbb{S}$ is the swap operator. The second equality holds because $6\,\mathrm{tr}_A\,\bar{Q}(n,d,3) = (D+2)(\mathbb{I}+\mathbb{S})$ given that the set of stabilizer states forms a 2-design.

Define the 'shadow map' associated with $\mathcal{T}\in\Sigma(d)$ for any linear operator $\mathfrak{O}\in\mathcal{H}_D$

$$\mathcal{R}_\mathcal{T}(\mathfrak{O}) := \mathrm{tr}_{BC}\left[R(\mathcal{T})(\mathbb{I}\otimes\mathfrak{O}\otimes\mathfrak{O}^\dagger)\right]. \tag{F6}$$

It follows that

$$\bar{\mathcal{Q}}_{n,d}(\mathfrak{O}) = \frac{D+2}{6(D+d)}\sum_{\mathcal{T}\in\Sigma(d)}\mathcal{R}_\mathcal{T}(\mathfrak{O}). \tag{F7}$$

If $\mathcal{T} \in \mathscr{T}_{\rm sym}$, then $\mathcal{T} = \mathcal{T}_O$ with $O \in S_3$. By straightforward calculation we have

$$\|\mathcal{R}_{\mathcal{T}}(\mathfrak{O})\| = \begin{cases} |\operatorname{tr}(\mathfrak{O})|^2 & \text{if } O = \mathbb{1}; \\ |\operatorname{tr}(\mathfrak{O})|\|\mathfrak{O}\| & \text{if } O = \tau_{12}, \tau_{13}; \\ \|\mathfrak{O}\|_2^2 & \text{if } O = \tau_{23}; \\ \|\mathfrak{O}\|^2 & \text{if } O = \zeta, \zeta^2. \end{cases} \tag{F8}$$

then

$$\sum_{\mathcal{T} \in \mathscr{T}_{\rm sym}} \|\mathcal{R}_{\mathcal{T}}(\mathfrak{O}_0)\| = \|\mathfrak{O}_0\|_2^2 + 2\|\mathfrak{O}_0\|^2. \tag{F9}$$

For $\mathcal{T} \in \mathscr{T}_{\rm ns}$, we have the following lemma. The detailed proof is presented in our companion paper.

Lemma 2 (Lemma 17 in Ref.)**.** *Suppose d is an odd prime and $\mathfrak{O}$ is a linear operator in $\mathcal{H}_D$. Then*

$$\|\mathcal{R}_{\mathcal{T}}(\mathfrak{O})\| \le \|\mathfrak{O}\|_2^2 \quad \forall \mathcal{T} \in \mathscr{T}_{\rm ns}. \tag{F10}$$

According to Lemma 2, we have

$$\sum_{\mathcal{T} \in \mathscr{T}_{\rm ns}} \|\mathcal{R}_{\mathcal{T}}(\mathfrak{O}_0)\| = (2d-4)\|\mathfrak{O}_0\|_2^2. \tag{F11}$$

In conjunction with Eq. (F7), we can deduce that

$$\begin{aligned} \frac{6(D+d)}{D+2}\|\bar{\mathcal{Q}}_{n,d}(\mathfrak{O}_0)\| &\le \sum_{\mathcal{T} \in \Sigma(d)} \|\mathcal{R}_{\mathcal{T}}(\mathfrak{O}_0)\| = \sum_{\mathcal{T} \in \mathscr{T}_{\rm sym}} \|\mathcal{R}_{\mathcal{T}}(\mathfrak{O}_0)\| + \sum_{\mathcal{T} \in \mathscr{T}_{\rm ns}} \|\mathcal{R}_{\mathcal{T}}(\mathfrak{O}_0)\| \\ &\le (2d-3)\|\mathfrak{O}_0\|_2^2 + 2\|\mathfrak{O}_0\|^2. \end{aligned} \tag{F12}$$

Eqs. (F5) and (F12) give the lower and upper bound in Eq. (9), respectively, according to Eq. (F7).

□

Appendix G: Magic gates and magic states

We have introduced the qudit magic gate T in the main text. The single-qudit magic state associated with T reads

$$|T\rangle := T|+\rangle = \frac{1}{\sqrt{d}} \sum_{u \in \mathbb{F}_d} \tilde{\omega}_d^{f(u)} |u\rangle, \tag{G1}$$

where $|+\rangle := \sum_{u \in \mathbb{F}_d} |u\rangle/\sqrt{d}$, T is defined in Eq. (13). By definition, $|T\rangle$ can be generated from $|0\rangle$ by the Fourier gate followed by the magic gate T.

When $d = 1 \bmod 3$, we need to distinguish three types of T gates depending on the cubic coefficient c ($c \neq 0$) of the underlying cubic polynomial f. It turns out the cubic character of the cubic coefficient can identify the essence. Let ν be a primitive element in the field $\mathbb{F}_d$. Here we choose the cubic character η_3 to be

$$\eta_3(\nu^k) = \omega_3^k, \text{ where } \omega_3 = e^{2\pi \mathbf{i}/3}. \tag{G2}$$

This cubic character depends on the choice of the primitive element ν, the choice of which for some small primes is shown in Table I. The cubic character of f, denoted by $\eta_3(f)$, is defined as the cubic character of c, that is, $\eta_3(f) = \eta_3(c)$. The cubic character of T and that of $|T\rangle$ are defined as the cubic character of f.

From Eq. (G2) we see that $\eta_3(c) = 0$ if and only if c is a cubic residue, that is, a cube of another element in $\mathbb{F}_d$. Thus, when $d = 2 \bmod 3$, every element in $\mathbb{F}_d$ is a cubic residue, so only one type of T gate exists. When $d = 3$, it can be verified by direct calculation that there is only one type of T gate.

TABLE I. Specific choices of primitive elements for $\mathbb{F}_d$, $d < 50$

d	3	5	7	11	13	17	19	23	29	31	37	41	43	47
ν	2	2	3	2	2	3	2	5	2	3	2	6	3	5

Appendix H: Additional numerical results

1. Shadow norms of stabilizer projectors

While Fig. 1 shows that one single T gate can already reduce the shadow norm of a stabilizer state significantly, and that the system size n has little influence once $n \geq 10$, Fig. 6 shows that $\|\mathfrak{O}\|_{\text{sh}}^2 - 3$ actually decrease exponentially with respect to the number of T gates. Recall that $\|\mathfrak{O}\|_{\text{sh}}^2 \approx 3$ with the measurement primitive is a 3-design, this result indicates that as more T gates are included, the underlying magic orbit approaches exponentially to a 3-design.

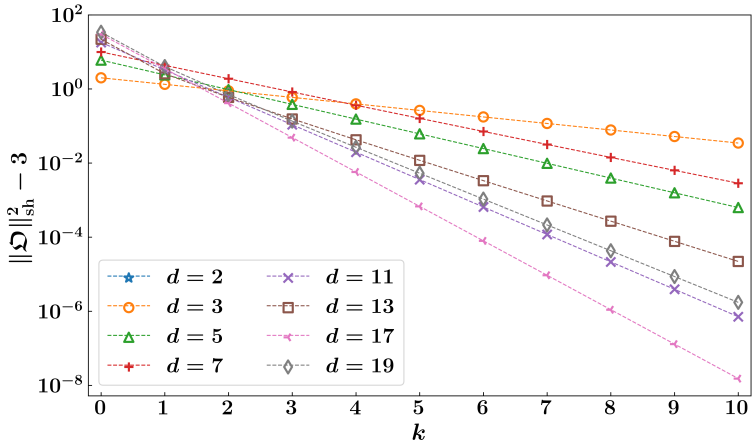


FIG. 6. Shadow norms of stabilizer states with respect to Clifford measurements supplemented by k canonical T gates. Here $n = 50$.

Besides stabilizer states ($K = 1$), we also calculate the shadow norm for stabilizer projectors of general rank K according to the analytical expressions given in the companion paper. In Fig. 7, we show the shadow norm of normalized stabilizer projector with respect to its rank, where $K = d^\eta$, and the system size is set to be $n = 50$. We consider normalized stabilizer projectors mainly to remove the influence of $\|P_{\mathcal{S}}\|_2$.

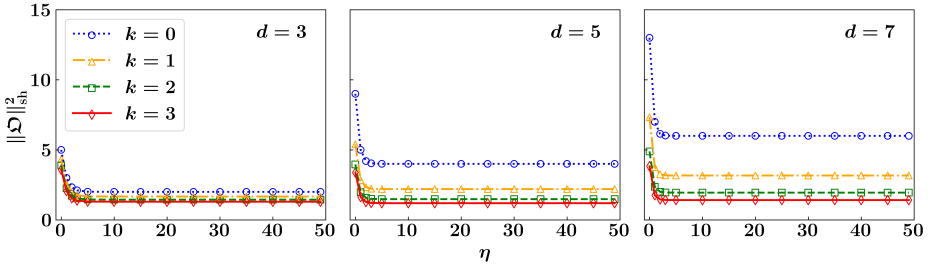


FIG. 7. Shadow norm of normalized stabilizer projectors. Suppose P is a rank-K stabilizer projector ($K = d^\eta$). Then $\mathfrak{O} = P_0/\|P_0\|_2$, where $P_0 = P - K\mathbb{I}/D$ and $\|P_0\|_2 = \sqrt{K - K^2/D}$. We set the number of qudits to be $n = 6$. k is the number of canonical T gates.

2. Median of means estimation

To compare different estimators, we also show the mean square error versus sample number using median of means estimation in Fig. 8, in comparison with Fig. 4 in the main text, which uses empirical means. For each sample number N, we divide the samples equally into c parts, each containing $h = N/c$ samples. We then calculate the means of every h samples, and take the median of c means as the estimator,

$$\hat{o}(h,c) = \text{median}\left(\hat{o}^{(1)}(h,1),\cdots,\hat{o}^{(c)}(h,1)\right), \text{ where } \hat{o}^{(j)}(h,1) = \frac{1}{h}\sum_{i=h(j-1)+1}^{hj} \text{tr}(\mathfrak{O}\hat{\rho}_i). \tag{H1}$$

Here we take $c = 10$, which, according to Theorem 1 in Ref. [1], implies that an error larger then ε occurs with probability less than $\delta \approx 0.013$. We see in Fig. 8 that the inverse mean square error $1/\langle\varepsilon^2\rangle$ for median of means is approximately 30% lower than that of empirical means.

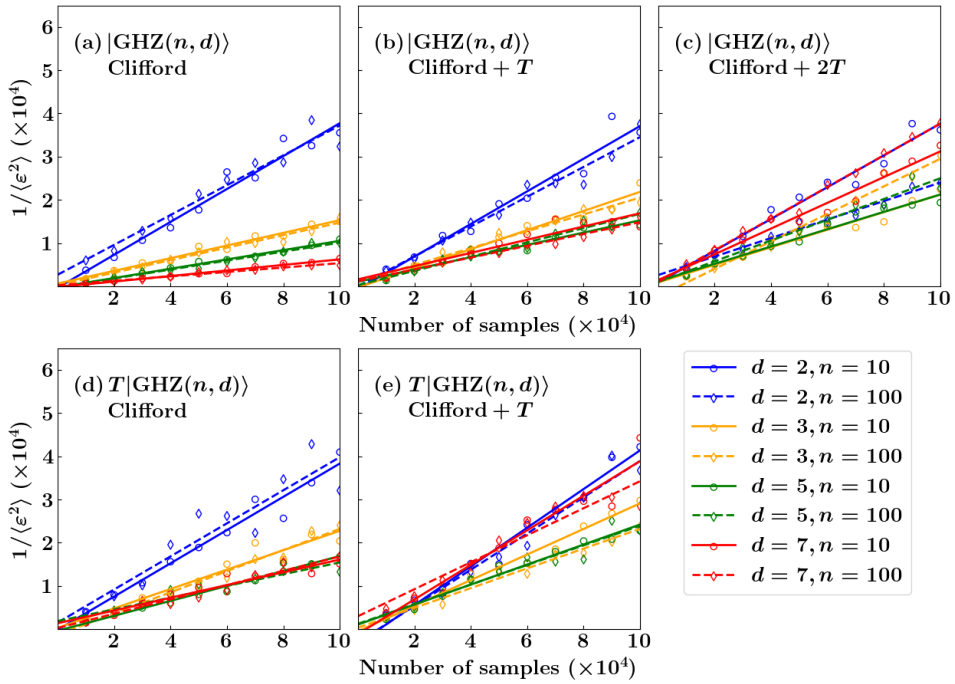


FIG. 8. Fidelity estimation of qudit GHZ states and T-modified GHZ states with themselves based on Clifford orbits with different numbers of magic gates, using median of means estimation with $c = 10$. We plot the inverse of average mean square error $1/\langle\varepsilon^2\rangle$ over 100 runs versus the number of samples, and by solid (dashed) lines we denote the linear fitting of the data.

3. Performance of shadow estimation on higher local dimensions and other states

In addition to Fig. 4, we show the performance of shadow estimation for different local dimensions in Fig. 9. Here we fix the systems size to be $n = 10$ and consider shadow estimation based on stabilizer measurements, and magic orbits with 1 and 2 T gates, respectively. We test on all prime dimensions under 20, and the type of T gates are all optimal choices under the specific settings. (See Table II). In all three settings, $d = 2$ has the smallest mean square error, but the gap between qubit and higher dimension cases is narrowing as the number of T gates increases. The dependence of $1/\langle\varepsilon^2\rangle$ on local dimension is also diminishing as more T gates are added to the circuit.

Similar behavior as in Fig. 4 also holds if we replace qudit GHZ state with other stabilizer states. As another widely interested case, we consider 2d cluster state on square lattice (with periodic boundary conditions). Suppose the square lattice is $a \times a$, and the local dimension is d. We denote the X (Z) operator of the

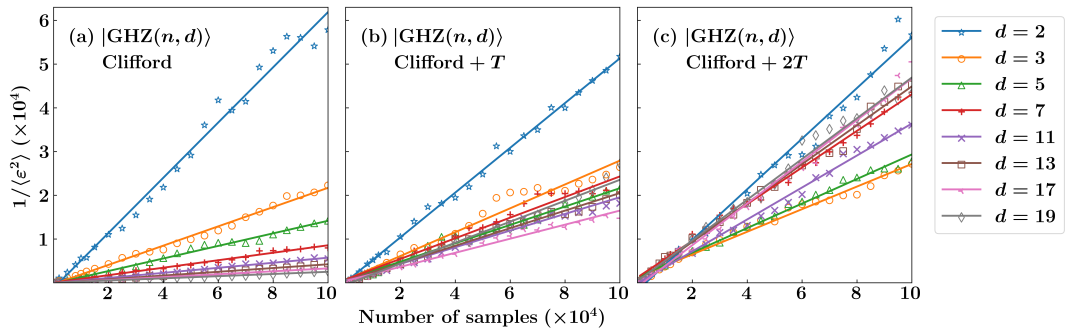


FIG. 9. Fidelity estimation of qudit GHZ states with itself based on Clifford orbits with different numbers of magic gates. Here we fix the number of qudits to be $n = 10$ and plot $1/\langle\varepsilon^2\rangle$ versus sample number for different dimensions. The data points are average over 100 runs, and the straight lines denote the linear fittings. The input state is $|\mathrm{GHZ}(10, d)\rangle$, and the measurement primitives are labeled in each figure.

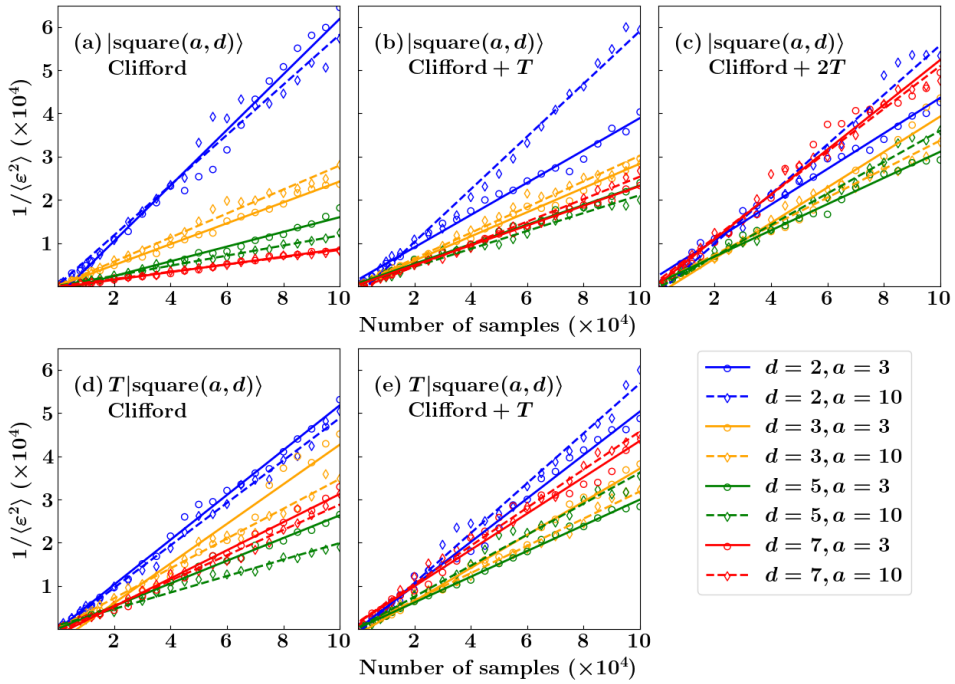


FIG. 10. Fidelity estimation of qudit cluster states (with periodic boundary conditions) and T-modified cluster states with themselves based on Clifford orbits with different numbers of magic gates. We plot the inverse of average mean square error $1/\langle\varepsilon^2\rangle$ over 100 runs versus the number of samples, and by solid (dashed) lines we denote the linear fittings of the data. The input states and measurement primitives are labeled in each figure.

qudit on the i-th row, j-th column by $X_{i,j}$ ($Z_{i,j}$). The stabilizer generators of the corresponding cluster state $|\mathrm{square}(a, d)\rangle$ are

$$g = X_{i,j} Z_{i-1,j} Z_{i+1,j} Z_{i,j-1} Z_{i,j+1}, \quad 1 \le i, j \le a, \tag{H2}$$

where the arithmetics on indices is done modulo a. The results for fidelity estimation are shown in Fig. 10.

4. Influence of T gates types on shadow norms

The type of T gates in $\mathcal{E}(\{T_j\}_{j=1}^k)$ may also affect the performance of shadow estimation. In Table II, we list the optimal choices of types of T gates for some dimensions satisfying $d = 1 \mod 3$. The ith number

in the bracket denotes the cubic coefficient of the T gate acting on the ith qudit. To numerically show the difference between these measurement settings, in Fig. 11 we calculate the distributions of shadow norms for randomly sampled 2-qudit diagonal and Hermitian observables for $d = 7$, with all possible combinations of cubic coefficients. For $d = 7$, one choice of the primitive element is $\nu = 3$. From Fig. 11 we see that for random diagonal observables in stabilizer basis, the settings $[0, \nu^2]$, $[\nu^2, \nu^2]$, $[\nu, \nu^2]$ stand out for their small shadow norm, which is consistent with the result in Table II. For random Hermitian observables, typical shadow norms are much smaller, making the difference negligible.

TABLE II. Optimal choices of types of T gates for different dimensions

	T	$2T$ (identical)	$2T$
$d = 7, \nu = 3$	$[\nu^2]$	$[\nu^2, \nu^2]$	$[\nu, \nu^2]$
$d = 13, \nu = 2$	$[1]$	$[1, 1]$	$[1, \nu]$
$d = 19, \nu = 2$	$[\nu^2]$	$[\nu^2, \nu^2]$	$[1, \nu^2]$
$d = 31, \nu = 3$	$[1]$	$[1, 1]$	$[1, \nu^2]$

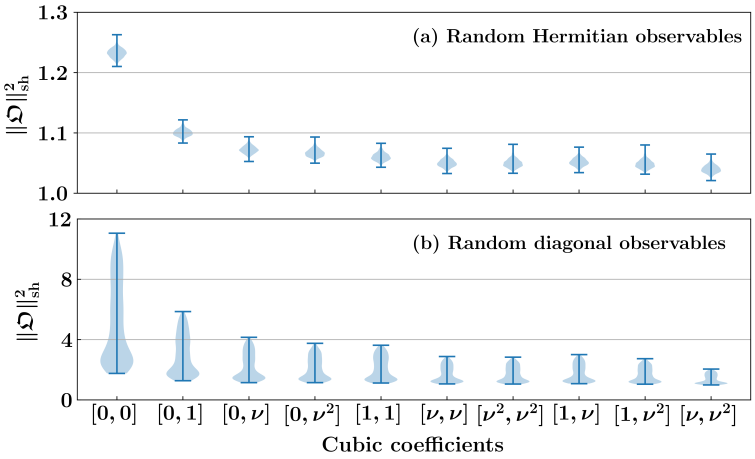


FIG. 11. Distribution of shadow norm of randomly sampled 2-qudit observables for $d = 7$ under different measurement settings. The X-axis denotes the cubic coefficients of the T gates on two qudits, $\nu = 3$.

Appendix I: Details for numerical simulations

Classical simulation of a quantum circuit has long been an important task for verification and validation of quantum devices. Notably, the Gottesman-Knill theorem states that a stabilizer circuit can be efficiently simulated on a classical computer. Later on, Aaronson and Gottesman gave a practical implementation. Using the tableau representation, their algorithm runs in time $\mathcal{O}(n^2)$ for both deterministic and random measurements, and $\mathcal{O}(n^3)$ for computing the inner product between two stabilizer states, where n is the qubit number [31]. Other simulation methods based on graph state representation [41] and canonical form of Clifford unitary [42] were also proposed. However, as far as we are concerned, no explicit algorithm for simulating qudit stabilizer circuits has been implemented so far.

There has been an extended collection of works on sampling algorithm for (qubit) Clifford group [34, 43, 44]. Ref. [35] generalized the subgroup algorithm in Ref. [34] to all prime dimensions.

The data acquisition phase of shadow estimation is essentially a random Clifford circuit. When no magic gates are involved, we can simulate this procedure on a classical computer by generalizing the tableau method in Ref. [31] to qudit case, and combining it with the Clifford sampling in Ref. [35].

Towards the simulation of universal quantum circuits, there has been several works focusing on the simulation of (qubit) stabilizer circuits interspersed with magic gates, among which two main approaches for dealing with T gates were proposed: first is by low-rank stabilizer decomposition [45], and second is through gadgetization [32, 33]. For both methods, the computational cost is polynomial in qubit number and exponential with respect to the number of magic gates. Here we take the latter one and generalize it to qudits.

The simulation method for qudit circuits we developed in this section may be of interest in itself, and useful beyond the current task.

1. Generating matrix of qudit stabilizer states

Before presenting the simulation algorithms, we shall first briefly review the generating matrix representation of qudit stabilizer states. Since for qubit case ($d = 2$) various techniques have been developed, from now on we assume $d \geq 3$, d is an odd prime.

Suppose $|\Psi\rangle \in \mathrm{Stab}(n, d)$ is an n-qudit stabilzer state, with stabilizer group $\mathcal{S} = \langle S_1, S_2, ..., S_n\rangle$, then $|\Psi\rangle$ can be uniquely determined by its generating matrix $\mathcal{G}_\Psi$

$$\mathcal{G}_\Psi = \left[\, \mathcal{G}_\Psi^z \,|\, \mathcal{G}_\Psi^x \,|\, r \,\right] = \left[\begin{array}{ccc|ccc|c} u_{1,1}^z & \cdots & u_{1,n}^z & u_{1,1}^x & \cdots & u_{1,n}^x & r_1^u \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ u_{n,1}^z & \cdots & u_{n,n}^z & u_{n,1}^x & \cdots & u_{n,n}^x & r_n^u \end{array}\right], \tag{I1}$$

where every row represents a stabilizer generator

$$S_i = \chi(r_i^u) W_{\mathbf{u}_i} = \chi(r_i^u) \bigotimes_{k=1}^{n} W(u_{i,k}^z, u_{i,k}^x). \tag{I2}$$

We shall call $\mathbf{u}_i := (u_{i,1}^z, \ldots, u_{i,n}^z; u_{i,1}^x, \ldots, u_{i,n}^x)$ a 'stabilizer vector', and $\omega_d^{r_i^u}$ its corresponding phase factor. According to the properties of stabilizer generators, if we swap two rows of $\mathcal{G}_\Psi$, add one row to another, or multiply one row by any integers in $\mathbb{F}_d \backslash \{0\}$, the underlying stabilizer state $|\Psi\rangle$ is unchanged. We shall call these 'row actions'.

According to Eq. (B10), it's not hard to see that applying a Clifford unitary $C = W_{\mathbf{a}}\mu(M) \in \mathrm{Cl}(n, d)$ on $|\Psi\rangle$ is equivalent to the mapping

$$\mathbf{u}_i \mapsto M\mathbf{u}_i, \tag{I3}$$

followed by a change in phase factors

$$r_i^u \mapsto r_i^u + [\mathbf{a}, M\mathbf{u}_i], \tag{I4}$$

for $i = 1, \ldots, n$ in the generating matrix $\mathcal{G}_\Psi$, where M, $\mathbf{a}$ are the symplectic matrix and vector corresponding to C in Eq. (B9), respectively.

Our shadow estimation protocol involves sampling of a random Clifford unitary. In generating matrix representation, this naturally reduces to the sampling of $M \in \mathrm{Sp}(2n, d)$ plus sampling of $\mathbf{a} \in \mathbb{F}_d^n$. We directly use the algorithm in Ref. [35] for sampling of $M \in \mathrm{Sp}(2n, d)$, which is a generalization of the subgroup algorithm proposed in Ref. [34] for $\mathrm{Sp}(2n, 2)$.

2. Simulating stabilizer circuit

For the simulation of stabilizer circuit (with no T gates), using tableau representation can be more efficient, especially for computing inner products. The overall simulation algorithm for stabilizer circuits in data acquisition phase is presented in Algorithm 1.

The sampling of random symplectic matrix in $\mathrm{Sp}(2n, d)$ is described in detail in Section 5.3.1 of Ref. [35], which we shall not repeat here. Associated notations and subroutines we developed will be introduced in the following subsections.

We generalize the tableau method for simulating stabilizer circuits in Ref. [31] to qudit cases mainly in the following aspects. Firstly, we develop an algorithm for constructing the tableau representation for arbitrary qudit stabilizer states. Secondly, inspired by the representation of stabilizer state with its associated Lagrangian subspace introduced in [35, 36], we develop a new method based on Gaussian elimination for computing the inner product, as well as simulating measurement, which runs in time $\mathcal{O}(n^3)$.

Algorithm 1: Data acquisition for stabilizer circuit

Input : Representation of the input state $|\Psi\rangle$, in the form of $\mathbf{s}_1, \ldots, \mathbf{s}_n$, and $\phi_1, \ldots, \phi_n$
Output: n-dit measurement outcome $\mathbf{x}$
1 Construct tableau $\mathcal{J}_\Psi$; `// See Appendix I 2 b`
2 Sample a random $M \in \mathrm{Sp}(2n, d)$ and $\mathbf{a} \in \mathbb{F}_d^n$;
3 Update $\mathcal{J}_\Psi$ with M and $\mathbf{a}$; `// See Appendix I 2 a`
4 Computational basis measurement on $\mathcal{J}_\Psi$ with outcome $\mathbf{x}$. `// See Appendix I 2 d`

a. Tableau representation

The tableau $\mathcal{J}_\Psi$ of a stabilizer state $|\Psi\rangle$ is a $2n \times (2n+1)$ matrix over $\mathbb{F}_d$. The upper half of $\mathcal{J}_\Psi$ is exactly $\mathcal{G}_\Psi$. The lower half of $\mathcal{J}_\Psi$, which we shall denote as $[v_{j,1}^z \cdots v_{j,n}^z \,|\, v_{j,1}^x \cdots v_{j,n}^x \,|\, r_j^v]$ $(j = 1, ..., n)$, should satisfy the following relations where the bracket denotes the symplectic inner product [31]

$$[\mathbf{u}_i, \mathbf{u}_j] = 0, \qquad [\mathbf{v}_i, \mathbf{v}_j] = 0, \qquad [\mathbf{u}_i, \mathbf{v}_j] = \delta_{ij}, \tag{I5}$$

for all $i, j = 1, \ldots, n$. Analogous to stabilizer generators $\{S_i\}_{i=1}^n$, the Weyl operators correspond to $\mathbf{v}_j$ and r_j^v are called 'destabilzer' generators $\{D_j\}_{j=1}^n$, where

$$D_j := \chi(r_j^v) W_{\mathbf{v}_j} = \chi(r_j^v) \bigotimes_{k=1}^n W(v_{j,k}^z, v_{j,k}^x). \tag{I6}$$

Similarly, we shall call $\mathbf{v}_j := (v_{j,1}^z, \ldots, v_{j,n}^z; v_{j,1}^x, \ldots, v_{j,n}^x)$ a 'destabilizer vector'. Apparently, row actions on the upper (or lower) half of tableau $\mathcal{J}_\Psi$ will not change the underlying state $|\Psi\rangle$. Destabilizer vectors $\mathbf{v}_j$ and r_j^v in phase factors satisfy the same update rules Eqs. (I3) and (I4) as $\mathbf{u}_i$ and r_i^u.

b. Constructing tableau for a stabilizer state

Suppose that a generating set of the stabilizer group of stabilizer state $|\Psi\rangle$ is $\langle \chi(\phi_1) W_{\mathbf{s}_1}, \cdots, \chi(\phi_n) W_{\mathbf{s}_n} \rangle$, where $\mathbf{s}_1, \cdots, \mathbf{s}_n \in \mathbb{F}_d^n$, $\phi_1, \cdots \phi_n \in \mathbb{F}_d$. Given $\mathbf{s}_1, \cdots, \mathbf{s}_n, \phi_1, \cdots \phi_n$, we can construct the tableau representation $\mathcal{J}_\Psi$ of $|\Psi\rangle$ with the following two steps: (1) finding stabilizer and destabilizer vectors; (2) tracking phase factors.

(1) Finding stabilizer and destabilizer vectors: Leave out the last column in $\mathcal{J}_\Psi$ for now and denote the resulting $2n \times 2n$ matrix by $\tilde{\mathcal{J}}_\Psi$. We shall call $\tilde{\mathcal{J}}_\Psi$ the *reduced* tableau of $|\Psi\rangle$. In this step, we want to find a set of n stabilizer vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and n destabilizer vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that Eq. (I5) are satisfied, and that $\{\mathbf{u}_i\}_{i=1}^n$ spans the same Lagrangian subspace as $\{\mathbf{s}_i\}_{i=1}^n$. This is done by Algorithm 2 introduced below.

Consider any eigenstate in the computational basis. Its corresponding reduced tableau can be written as $\tilde{\mathcal{J}}_0 = \mathbb{I}_{2n}$, which we shall call the 'standard' reduced tableau. We denote the rows of $\tilde{\mathcal{J}}_0$ the 'standard basis' of reduced tableau, denoted as $\{\mathbf{e}_{2n}(i)\}_{i=1}^{2n}$, where $\mathbf{e}_{2n}(i)$ is a vector of length $2n$ with all zero entries except a 1 on the i-th entry.

In this step, we take the 'local' convention and write the tableau row $(z_1, \ldots, z_n, x_1, \ldots, x_n)$ as $(z_1, x_1, \ldots, z_n, x_n)$. The basic idea of Algorithm 2 is that, starting from $\tilde{\mathcal{J}}_0$, if we can find a symplectic transformation that takes $\mathbf{e}_{2n}(2i-1)$ to $\mathbf{u}_i$, then it also transforms $\mathbf{e}_{2n}(2i)$ to $\mathbf{v}_i$ such that the symplectic inner products Eq. (I5) are preserved.

Denote $\mathrm{Sp}(2m, d)$ by G_m. For symplectic groups, the following nested subgroup chain holds

$$G_1 \subset G_2 \subset \cdots \subset G_{n-1} \subset G_n =: G. \tag{I7}$$

Let $M_{m-1} \in G_{m-1}$. The embedding $G_{m-1} \mapsto G_m$ is given by $M_{m-1} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \oplus M_{m-1}$, and the map

$$G_n/G_{n-1} \times G_{n-1}/G_{n-2} \times \cdots \times G_2/G_1 \times G_1 \to G \tag{I8}$$

$$([M_n], [M_{n-1}], \ldots, [M_2], M_1) \to M_n M_{n-1} \cdots M_1 \tag{I9}$$

is an isomorphism. In particular, each symplectic transformation $M \in G$ has a unique representation as $M_n M_{n-1} \cdots M_1$ with $[M_m] \in G_m/G_{m-1}$ for $j = 2, \ldots, n$ and $M_1 \in G_1$. Thus, to obtain the desire M, we can construct $M_n, M_{n-1}, ..., M_1$ sequentially.

The major tool for finding a desired symplectic transformation is symplectic transvections [35]. Given a vector $\mathbf{h} \in \mathbb{F}_d^{2m}$ (in our case $m = 1, 2, ..., n$) and a scalar $\lambda \in \mathbb{F}_d \backslash \{0\}$, a symplectic transvection is a symplectic map $T_{\lambda,\mathbf{h}}$ such that

$$T_{\lambda,\mathbf{h}}(\mathbf{x}) = \mathbf{x} + \lambda[\mathbf{x}, \mathbf{h}]\mathbf{h}. \tag{I10}$$

The inverse transvection is given by

$$T^{-1}_{\lambda,\mathbf{h}}(\mathbf{x}) = T_{-\lambda,\mathbf{h}}(\mathbf{x}). \tag{I11}$$

A basic fact from symplectic geometry is that transvections generate the symplectic group. More concretely, let $\mathbf{x}, \mathbf{y} \in \mathbb{F}_d^{2m} \backslash \{\mathbf{0}_{2m}\}$ be two vectors. Then there exist vectors $\mathbf{h}_1, \mathbf{h}_2$ and scalars λ_1, λ_2 such that

$$\mathbf{y} = T_{\lambda_1,\mathbf{h}_1} T_{\lambda_2,\mathbf{h}_2}(\mathbf{x}). \tag{I12}$$

This is Lemma 5.1 of Ref. [35], in which a constructive proof is also given and it should be clear how the subroutine 'find transvections' in Algorithm 2 Line 3 works. The overall computational complexity of Algorithm 2 is $\mathcal{O}(n^3)$.

For convenience of notation, we will write $T_{\lambda_1,\mathbf{h}_1} T_{\lambda_2,\mathbf{h}_2}(\mathbf{x})$ simply as $T(\mathbf{x})$ and $\left(T_{\lambda_1,\mathbf{h}_1} T_{\lambda_2,\mathbf{h}_2}\right)^{-1}(\mathbf{x}) = T^{-1}_{\lambda_2,\mathbf{h}_2} T^{-1}_{\lambda_1,\mathbf{h}_1}(\mathbf{x})$ as $T^{-1}(\mathbf{x})$. Here we take the 'local convention', that is, the $(2j-1)$-th, $2j$-th entry of $\mathbf{u}$ denotes the power of Z_j, X_j respectively.

Algorithm 2: Constructing tableau for stabilizers

Input : Stabilizer vectors $\mathbf{s}_1, \ldots, \mathbf{s}_n$
Output: Tableau rows $\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}_1, \ldots, \mathbf{v}_n$

```
1  Set {u_1^(1), ..., u_n^(1)} = {s_1, ..., s_n};
2  for i = 1 : n do
3  |  Find transvections, construct T_i such that T_i(e_{2(n+1-i)}(1)) = u_i^(i);
4  |  for j = i + 1 : n do
5  |  |  u_j^(i+1) = T_i^{-1}(u_j^(i)), and discard the first two entries of u_j^(i+1);
6  |  end
7  end
8  for i = n : 1 do
9  |  M = I_2 ⊕ M (M = I_2 for i = n);
10 |  Apply T_i to every row of M;
11 end
12 u_1, ..., u_n, v_1, ..., v_n are the corresponding rows of M.
```

Since the commutation relations Eq. (I5) are preserved under symplectic transformations, in each iteration, $\mathbf{u}_j^{(i+1)} = T_i^{-1}(\mathbf{u}_j^{(i)})$ is guaranteed to commute with both $\mathbf{e}_{2(n+1-i)}(1)$ and $\mathbf{e}_{2(n+1-i)}(2)$, thus the first two entries of $\mathbf{u}_j^{(i+1)}$ must be zero. So we can discard them and continue to look for transvections that correspond to symplectic transformations in $\mathrm{Sp}(2(n-i), d)$ in the next iteration.

(2) Tracking phase factors: Now we determine the phase factors in the last column of $\mathcal{J}_\Psi$. Since every $\mathbf{s}_i$ $(i = 1, \ldots, n)$ is a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_n$, i.e., $\mathbf{s}_i = \sum_{j=1}^n c_i^j \mathbf{u}_j$, with coefficients

$$c_i^j = [\mathbf{s}_i, \mathbf{v}_j], \tag{I13}$$

we can solve for $r_1^u, \ldots, r_n^u$ from n linear equations

$$\sum_{j=1}^n c_i^j r_j^u = \phi_i, \quad (i = 1, \ldots, n). \tag{I14}$$

This is essentially a Gaussian elimination, and takes time $\mathcal{O}(n^3)$.

The phase factors for destabilizers $\{r_i^v\}_{i=1}^n$ are actually redundant. For simplicity, we set as convention $r_i^v = 0$ for $i = 1, \ldots, n$.

Algorithm 3: Simulation of computational basis measurement

Input : The Lagrangian subspace $\mathcal{L}$ and characteristic vector $\mathbf{m}$ of pre-measurement state
Output: Measurement outcome $\mathbf{x}$
1 Find a basis $\{\mathbf{w}_k\}_{k=1}^{\zeta}$ of $\mathcal{L}_0 \cap \mathcal{L}$ with the Zassenhaus algorithm;
2 Construct the equations according to Eq. (I20), and solve for $\tilde{\mathbf{x}}$;
3 Return the latter half of $\tilde{\mathbf{x}}$ as $\mathbf{x}$.

c. Lagrangian subspace and characteristic vector

An alternative representation of stabilizer state $|\Psi\rangle$ is to use its Lagrangian subspace, which would be very useful when computing the inner product of two stabilizer states in next subsection.

Given the stabilizer vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$ of state $|\Psi\rangle$, they span a subspace $\mathcal{L}$ of $\mathbb{F}_d^n$. Note that the symplectic inner product vanish on $\mathcal{L}$. Such subspaces are called isotropic. One can show that the maximal dimension of an isotropic subspace is n. Such a maximal isotropic subspace is called a Lagrangian subspace. Apparently, $\mathcal{L}$ is a Lagrangian subspace. Let $\mathcal{U}$ be a subspace of $\mathcal{V}$. We use $\mathcal{U}^\perp$ to denote the symplectic complement of $\mathcal{U}$, i.e., for any $\mathbf{v} \in \mathcal{U}^\perp$, $[\mathbf{u}, \mathbf{v}] = 0,\ \forall \mathbf{v} \in \mathcal{U}$. If in addition $\mathcal{U}$ is a Lagrangian subspace, then $\mathcal{U}^\perp = \mathcal{U}$.

We can express $|\Psi\rangle$ as [36]

$$|\Psi\rangle = |\mathcal{L}, \mathbf{m}\rangle := |\mathcal{L}|^{-1} \sum_{\mathbf{u}\in\mathcal{L}} \chi([\mathbf{m}, \mathbf{u}]) W_{\mathbf{u}}, \tag{I15}$$

where $\mathbf{m} \in \mathbb{F}_d^n$ is called the characteristic vector.

Note that the choice of $\mathbf{m}$ is not unique. In fact, if $\mathbf{m}' - \mathbf{m} \in \mathcal{L}$, then $|\mathcal{L}, \mathbf{m}'\rangle$ represents the same stabilizer state as $|\mathcal{L}, \mathbf{m}\rangle$, that is to say, $\mathbf{m}$ has a d^n-fold degeneracy.

The conversion from tableau $\mathcal{J}_\Psi$ to $|\mathcal{L}, \mathbf{m}\rangle$ is straightforward. Observe that $\mathbf{m}$ should satisfy

$$[\mathbf{m}, \mathbf{u}_i] = r_i^u, \quad \forall\ i = 1, \ldots, n. \tag{I16}$$

Thus we can choose $\mathbf{m} = -\sum_i r_i^u \cdot \mathbf{v}_i$. In fact, instead of tableau $\mathcal{J}_\Psi$, according to Eq. (I15), we can represent $|\Psi\rangle$ by $\mathbf{u}_1, \cdots, \mathbf{u}_n$ and $\mathbf{m}$, and simulate Clifford transformation $C = W_{\mathbf{a}}\mu(M)$ by the following map

$$\begin{aligned} &\mathbf{u}_i \mapsto M\mathbf{u}_i, \\ &\mathbf{m} \mapsto M\mathbf{m} + \mathbf{a}. \end{aligned} \tag{I17}$$

However, to determine the characteristic vector $\mathbf{m}$, we first need to construct a symplectic basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{v}_1, \ldots, \mathbf{v}_n\}$, so the overall computational cost of this approach is almost the same as that of tableau representation.

d. Computational basis measurement

The simulation of computational basis measurement can be seen as a variant of computing inner product of two stabilizer states.

Given two stabilizer states $|\mathcal{L}_1, \mathbf{m}_1\rangle$ and $|\mathcal{L}_2, \mathbf{m}_2\rangle$, their overlap reads [35]

$$|\langle \mathcal{L}_1, \mathbf{m}_1 | \mathcal{L}_2, \mathbf{m}_2 \rangle|^2 = \begin{cases} d^{\dim|\mathcal{L}_1\cap\mathcal{L}_2| - n} & \text{if } \mathbf{m}_1 - \mathbf{m}_2 \in (\mathcal{L}_1 \cap \mathcal{L}_2)^\perp \\ 0 & \text{otherwise.} \end{cases} \tag{I18}$$

Denote $\mathcal{L}_0$ as the Lagrangian subspace spanned by $\mathbf{e}_{2n}(1), ..., \mathbf{e}_{2n}(n)$, that is, the upper half of $\tilde{\mathcal{J}}_0$. Suppose $\mathbf{x} = (x_1, ..., x_n)$ is the measurement outcome. The corresponding characteristic vector can be chosen as $\tilde{\mathbf{x}} = (\mathbf{0}, \mathbf{x}) = (0, ..., 0, x_1, ..., x_n)$. Then the probability of obtaining outcome $\mathbf{x}$ when performing computational basis measurement on stabilizer state $|\mathcal{L}, \mathbf{m}\rangle$ is

$$p(\mathbf{x}) = |\langle \mathcal{L}, \mathbf{m} | \mathcal{L}_0, \tilde{\mathbf{x}} \rangle|^2. \tag{I19}$$

We then describe the scheme for sampling an n-dit measurement outcome $\mathbf{x}$. Firstly, find a basis of $\mathcal{L}_0 \cap \mathcal{L}$. This can be done with the Zassenhaus algorithm [46]. Suppose the basis we find is $\{\mathbf{w}_k\}_{k=1}^{\zeta}$, where

$\zeta = \dim|\mathcal{L}_0 \cap \mathcal{L}|, 0 \le \zeta \le n$. Then we can sample a characteristic vector $\tilde{\mathbf{x}}$ by solving the equations

$$\begin{aligned} [\tilde{\mathbf{x}}, \mathbf{w}_i] &= [\mathbf{m}, \mathbf{w}_i] \quad \text{for } i = 1, \ldots, \zeta, \\ [\tilde{\mathbf{x}}, \mathbf{e}_{2n}(i)] &= \alpha_i \quad \text{for } i = \zeta + 1, \ldots, n, \end{aligned} \tag{I20}$$

where α_i is a random integer in $\mathbb{F}_d$. Finally, we take the latter half of $\tilde{\mathbf{x}}$, which is exactly $\mathbf{x}$. The overall procedure is summarized in Algorithm 3.

3. Simulating Clifford+T gate circuit

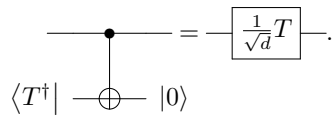
In the main text, we focus on fidelity estimation using shadow estimation as a prototypical task. In this section, we shall describe how to classically simulate the procedure when the unitary ensemble corresponds to a Clifford circuit with magic gates.

In the data acquisition stage, an input state goes through a layer of random Clifford gate $C_m \in \mathrm{Cl}(n, d)$, a layer of T gates and a layer of Fourier gates, followed by computational basis measurement. If the input state can be prepared by applying a few layers of Clifford and T gates to $|0\rangle^{\otimes n}$, then this procedure can be efficiently simulated. That is to say, we can sample the measurement outcome $\mathbf{x} := (x_1, x_2, \ldots, x_n)$ of this quantum circuit according to probability distribution $\{p(\mathbf{x})\}_{\mathbf{x}}$ on a classical computer, with computational cost polynomial in n, and exponential with respect to $(t_1 + t_2)$, where t_1 is the number of T gates used in preparation of input state, and t_2 is the number of T gates in measurement primitive.

We use the same simulation scheme as in Ref. [33], in which each T gate is replaced equivalently by a 'reversed' gadget. For simplicity, we denote the magic gates liberally as T, regardless of its specific type. Define a single qudit ancillary state

$$|T^\dagger\rangle := \begin{cases} \frac{1}{\sqrt{d}} \sum_{u \in \mathbb{F}_d} \omega_d^{-f(u)} |u\rangle & d \ge 5, \\ \frac{1}{\sqrt{d}} \sum_{u \in \mathbb{F}_d} \omega_9^{-f(u)} |u\rangle & d = 3. \end{cases} \tag{I21}$$

Diagrammatically, we can express the generalized reversed gadget for qudit circuits as



(I22)

Circuits in this section are read from right to left, so that we can derive the expression of some intermediate quantum states directly from the circuit diagram.

Suppose the input state is $C_{p,l}...T^{\otimes s_2} C_{p,1} T^{\otimes s_1} C_{p,0} |0\rangle^{\otimes n}$, where $\sum_{i=1}^k s_i = t_1$, $C_{p,0}, C_{p,1}, ..., C_{p,l} \in \mathrm{Cl}(n, d)$, then the circuit U we are to simulate can be expressed diagrammatically as in Fig. 12 (a). (p for state preparation and m for measurement). Since C_m is selected randomly from $\mathrm{Cl}(n, d)$, the last $C_{p,l}$ can be absorbed into C_m, and the t_2 T gates in the measurement primitive are applied to the first t_2 different qudits without loss of generality. Its equivalent circuit using reversed gadget is shown in Fig. 12 (b).

To obtain outcome $\mathbf{x}$ and its corresponding probability $p(\mathbf{x})$, we sample the outcome on each qudit subsequently [32]. For each $j = 1, \ldots, n-1$, define the conditional probabilities

$$p(y \,|\, x_1, \ldots, x_{j-1}) := \frac{p(x_1, \ldots, x_{j-1}, y)}{p(x_1, \ldots, x_{j-1})}, \tag{I23}$$

where $y \in \mathbb{F}_d$. Thus, to simulate each experiment shot, we only need to calculate the outcome probabilities $p(\mathbf{x})$ for nd sequences $\mathbf{x}$ of length varying from 1 to n, instead of calculating all d^n probabilities $p(\mathbf{x})$ for $\mathbf{x} \in \mathbb{F}_d^n$. Evidently, the corresponding (single shot) fidelity estimator can be expressed as

$$\hat{f}(\mathbf{x}) = (d^n + 1) p(\mathbf{x}) - 1. \tag{I24}$$

In what follows, we will illustrate how to evaluate an outcome probability $p(\mathbf{x})$ for $\mathbf{x} = (x_1, \ldots, x_w)$ of length w ($1 \le w \le n$). We refer to the first w qudits as the measured register 'a' and the remaining $(n - w)$ qudits as the marginalized register 'b', then the probability reads

$$p(\mathbf{x}) = \left\| \langle \mathbf{x}|_a U |0\rangle_{ab}^{\otimes n} \right\|_2^2, \tag{I25}$$

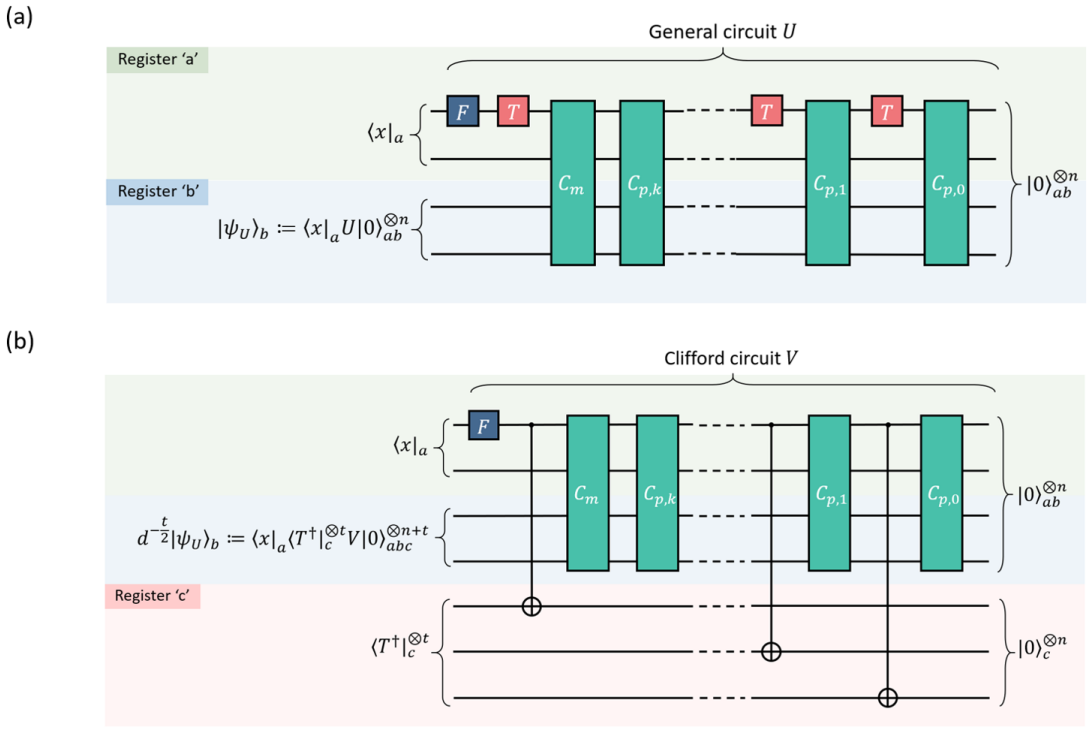


FIG. 12. (a) Quantum circuit for Clifford+T measurement on an input state prepared from Clifford unitaries plus a few T gates. (b) Post-selected circuit obtained by reverse gadgetization of each T gate.

After gadgetization, circuit U acting on $|0\rangle_{ab}^{\otimes n}$ is re-expressed as a Clifford circuit V acting on $|0\rangle_{abc}^{\otimes n+t}$, with t ancillary qudits in register 'c' post-selected on the state $|T^\dagger\rangle$. Thus the above probability can be re-expressed as

$$p(\mathbf{x}) = d^t \left\| \langle \mathbf{x}|_a \langle T^\dagger|_c^{\otimes t} V|0\rangle_{abc}^{\otimes n+t} \right\|_2^2, \tag{I26}$$

which can be further considered as the trace of the product of two projectors $p(\mathbf{x}) = \mathrm{tr}\left(\Pi_{\mathcal{G}} \Pi_{T,\mathbf{x}}\right)$, where

$$\Pi_{\mathcal{G}} = V|0\rangle\langle 0|_{abc}^{\otimes n+t} V^\dagger, \quad \Pi_{T,\mathbf{x}} = |x\rangle\langle x|_a \otimes \left(\mathbb{I}^{\otimes n-w}\right)_b \otimes |T^\dagger\rangle\langle T^\dagger|_c^{\otimes t}. \tag{I27}$$

We express stabilizer state $V|0\rangle_{abc}^{\otimes n+t}$ with its generating matrix $\mathcal{G}$.

In order to contribute non-trivially to $p(\mathbf{x})$, S_i must satisfy the following constraints: **(i)** S_i is identity on qudits in register 'b', and contains only Z components in qudits in register 'a', and **(ii)** the phase factor r_i of S_i subject to certain constraints imposed by $|x\rangle\langle x|_a$.

Before presenting the algorithm for calculating $p(\mathbf{x})$, we first introduce two useful functions.

(1) Slicing of $\mathcal{G}$: We denote by $\mathcal{G}^z[\alpha:\beta, \gamma:\delta]$ the sub-matrix of $\mathcal{G}^z$ the contains rows from α to β, and columns form γ to δ in $\mathcal{G}^z$, and similarly for $\mathcal{G}^x[\alpha:\beta, \gamma:\delta]$ and $\mathcal{G}^x$. For $\mathcal{G}$, we define

$$\mathcal{G}[\alpha:\beta, \gamma:\delta] := \left[\begin{array}{ccc|ccc|c} u^z_{\alpha,\gamma} & \cdots & u^z_{\alpha,\delta} & u^x_{\alpha,\gamma} & \cdots & u^x_{\alpha,\delta} & r^u_\alpha \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ u^z_{\beta,\gamma} & \cdots & u^z_{\beta,\delta} & u^x_{\beta,\gamma} & \cdots & u^x_{\beta,\delta} & r^u_\beta \end{array}\right]. \tag{I28}$$

When arguments in position α (γ) are omitted, we mean starting from the first row (column). When arguments in position β (δ) are omitted, we mean stopping on the last row (column). Suppose $\alpha < \beta$, $\gamma < \delta$, $\mathcal{G}[\beta:\alpha, \gamma:\delta]$ ($\mathcal{G}[\alpha:\beta, \delta:\gamma]$) means rearranging the rows (columns) of ($\mathcal{G}[\alpha:\beta, \gamma:\delta]$) in inverse order.

(2) Echelon forms of $\mathcal{G}$: We denote by *Echelon-z*($\mathcal{G}$, m) (*Echelon-x*($\mathcal{G}$, m)) the generating matrix $\mathcal{G}'$ obtained by row actions (swap, addition and multiplication by integers in $\mathbb{F}_d \backslash \{0\}$) on $\mathcal{G}$ such that $\mathcal{G}'^z[:, 1:m]$ ($\mathcal{G}'^x[:, 1:m]$) is in echelon form.

For convenience of expression, we also define an 'inverse echelon form': *inv-Echelon-z*($\mathcal{G}$, m) (*inv-Echelon-x*($\mathcal{G}$, m)) is the generating matrix $\mathcal{G}'$ derived from $\mathcal{G}$ by row actions such that $\mathcal{G}'^z[:, m:1]$ ($\mathcal{G}'^x[:, m:1]$) is in echelon form.

Moreover, we define a 'cut' function for (inverse) echelon form, which returns a row index. When $\mathcal{G}'^z$ is in echelon form (or inverse echelon form), *cut-z*$(\mathcal{G}'; k_1, k_2)$ is defined as row index of $\mathcal{G}'$ such that for all $j \geq c =$*cut-z*$(\mathcal{G}'; k_1, k_2)$, $\mathcal{G}'^z[j, k_1 : k_2]$ has all zero entries; for all $1 \leq j < c$, $\mathcal{G}'^z[j, k_1 : k_2]$ has at least one non-zero entry. Similarly we can define *cut-x*$(\mathcal{G}'; k_1, k_2)$.

Algorithm 4 takes as input the generating matrix $\mathcal{G}$ of $V|0\rangle_{abc}^{\otimes n+t}$, and outputs the 'constrained stabilizers' $\mathcal{G}_c$ and an integer ξ. Essentially, Algorithm 4 takes the partial trace in Eq. (I26), as the first step for calculating $p(\mathbf{x})$. When the overlap is zero, the algorithm returns None. $\mathcal{G}_c$ is in a form similar to the generating matrix, with every row representing a Weyl operator g_i on register 'c' (see Eq. (I2)). Suppose the rank of $\mathcal{G}_c$ is k_c. $\mathcal{G}_c$ and ξ satisfy the following equation

$$p(\mathbf{x}) = d^t \operatorname{tr}\left(\Pi_{\mathcal{G}} \Pi_{T,\mathbf{x}}\right) = d^{t+\xi} \operatorname{tr}\left(\Pi_{\mathcal{G}_c} \cdot |T^\dagger\rangle\langle T^\dagger|_c^{\otimes t}\right), \tag{I29}$$

where $\Pi_{\mathcal{G}_c} = \prod_{i=1}^{k_c} \left(\sum_{j=0}^{d-1} g_i^j\right)$.

Algorithm 4: Constrain stabilizers

Input : Generating matrix $\mathcal{G}$, w-dit outcome $\mathbf{x}$
Output: Constrained stabilizers $\mathcal{G}_c$, integer ξ

1 $\mathcal{G}' =$*Echelon-x*$(\mathcal{G}, n+t)$;
2 $c_1 =$*cut-x*$(\mathcal{G}'; 1, n)$, $\mathcal{G}_1 = \mathcal{G}'[c_1:, :]$, $k_1 = \text{len}(\mathcal{G}_1)$;
3 $\mathcal{G}_1' =$*inv-Echelon-z*$(\mathcal{G}_1, n)$;
4 $c_2 =$*cut-z*$(\mathcal{G}_1'; w+1, n)$, $\mathcal{G}_2 = \mathcal{G}_1'[c_2:, :]$, $k_2 = \text{len}(\mathcal{G}_2)$;
5 **for** $i = 1 : k_2$ **do**
6 $\quad r_i' = r_i - \mathcal{G}_2^z[i, 1:w] \cdot \mathbf{x}$;
7 $\quad$ **if** $\mathcal{G}_2^z[i, n+1 : n+t]$, $\mathcal{G}_2^x[i, n+1 : n+t]$ *all zero* **then**
8 $\quad\quad$ **if** $r_1' \neq 0$ **then**
9 $\quad\quad\quad$ **return** None;
10 $\quad\quad$ **else**
11 $\quad\quad\quad$ Continue;
12 $\quad\quad$ **end**
13 $\quad$ **else**
14 $\quad\quad$ Add $\left(\mathcal{G}_2^z[i, n+1:n+t] \,|\, \mathcal{G}_2^x[i, n+1:n+t] \,|\, r_i'\right)$ to $\mathcal{G}_c$;
15 $\quad$ **end**
16 **end**
17 $k_c = \text{len}(\mathcal{G}_c)$;
18 **return** $\mathcal{G}_c$, $\xi = k_2 - k_c$.

Note that $\mathcal{G}_2$ in Algorithm 4 line 4 already satisfies condition **(i)**, and we check for condition **(ii)** in lines 5 to 16. The $\cdot$ in line 6 denotes vector inner product, and all arithmetic is done modulo d. The time cost for Algorithm 4 is approximately $\mathcal{O}\left((n+t)^3 + k_1^3\right)$.

With $\mathcal{G}_c$ obtained, it follows immediately that $p(\mathbf{x}) = d^{t+\xi} \operatorname{tr}\left(\Pi_{\mathcal{G}_c} \cdot |T^\dagger\rangle\langle T^\dagger|_c^{\otimes t}\right)$ can be computed directly in time $\mathcal{O}\left(td^{k_c+1}\right)$. Usually, k_c equals to t, so the computational cost increase exponentially with t. The procedure for simulating a Clifford circuit with T gates is summarized in Algorithm 5.

Notice that some intermediate results (such as $\mathcal{G}_1$, $\mathcal{G}_2$) can be used repeatedly, we made some optimization to our algorithm. Finally, the overall time cost for simulation of a single experiment shot, that is, the time cost for sampling an outcome $\mathbf{x}$ of length n, is approximately $\mathcal{O}\left((n+t)^3 + k_1^3 + td^{k_1+1} + ndk_1\right)$. Since typical k_1 is little larger than t, we conclude that a single experiment shot can be simulated on classical computer in time $\mathcal{O}\left((n+t)^3 + td^{t+1}\right)$, where t is the number of T gates in the gadgetized circuit.

Appendix J: Predicting quadratic functions

Here we focus on the task of predicting a quadratic function of the form $f(\rho) := \operatorname{tr}(\mathfrak{O}\rho \otimes \rho)$ using shadow estimation based on the Clifford group. The estimator constructed from U-statistic reads

$$\hat{o} = \frac{1}{N(N-1)} \sum_{i \neq j} \operatorname{tr}(\mathfrak{O}\hat{\rho}_i \otimes \hat{\rho}_j), \tag{J1}$$

Algorithm 5: Data acquisition for Clifford circuit

Input : Circuit representation U
Output: n-dit measurement outcome $\mathbf{x}$

Construct the generating matrix $\mathcal{G}$ for $|0\rangle^{\otimes n}$, apply $C_{p,0}, \mathrm{CNOT}, C_{p,1}, ...$ sequentially;
Sample a random $M \in \mathrm{Sp}(2n, d)$ and $\mathbf{a} \in \mathbb{F}_d^{2n}$;
Update $\mathcal{G}$ with M and $\mathbf{a}$;
Apply CNOT and F gates to the corresponding qudits in $\mathcal{G}$;
for $w = 1 : n$ **do**
  for $y = 0 : d-1$ **do**
    $\mathbf{y} = \mathrm{concatenate}(\mathbf{x}, y)$;
    Constrain stabilizers($\mathcal{G}$,$\mathbf{x}$);
    if *result*= None **then**
      $p(\mathbf{y}) = 0$;
    else
      $p(\mathbf{y}) = d^{t+\xi} \operatorname{tr}\left(\Pi_{\mathcal{G}_c} \cdot |T^\dagger\rangle\langle T^\dagger|_c^{\otimes t}\right)$;
    end
  end
  Sample $\mathbf{y}$ according to $\{p(\mathbf{y})/p(\mathbf{x})\}_{y=0}^{d-1}$, $\mathbf{x} \leftarrow \mathbf{y}$;
end
return $\mathbf{x}$

which is the minimum-variance unbiased estimator. A general upper bound for the sample complexity was established in Ref. [1], as reproduced here.

Lemma 3. *Fix a measurement primitive $\mathcal{E}$. For a quadratic target function $\mathfrak{O}$, a measurement budget of size*

$$8/\varepsilon^2 \times \max\left\{\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\rho\otimes\hat{\rho}_1)], \mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\rho)], \sqrt{\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\hat{\rho}_2)]}\right\} \tag{J2}$$

allows for predicting the expectation value of $\mathfrak{O}$ via U-statistic estimators within error ε.

This is a simple corollary of Lemma 5 in Ref. [1]. Next, we consider variance bounds for the three measurement primitives discussed in the main text, respectively.

1. Variance bounds for measurement based on Clifford orbits

Let ρ be an n-qudit quantum state, and $D = d^n$. First we consider shadow estimation based on magic and non-magic Clifford orbits. Based on Eq. (16), we generalize the definition of $\gamma_{d,k}$

$$\gamma_{d,k} = \begin{cases} 2d-1 & k = 0, \\ 3 + \frac{2^{k+1}(d-2)}{d^k} & 1 \le k \le n, d \neq 1 \mod 3, \\ 3 + \frac{9}{8}\cdot\frac{4^k}{d^{k-1}} & 1 \le k \le n, d = 1 \mod 3, \end{cases} \tag{J3}$$

where k is the number of T gates in the measurement primitive. From Theorem 2 and Theorem 3, we know that $\|\mathfrak{O}\|_{\mathrm{sh}}^2 \le \gamma_{d,k}\|\mathfrak{O}\|_2^2$ for any traceless linear observable $\mathfrak{O} \in \mathcal{H}_D$.

Proposition 3. *Suppose $\mathfrak{O}$ describes a quadratic function $\mathrm{tr}(\mathfrak{O}\rho\otimes\rho)$ and $\|\mathfrak{O}\|_2 \ge 1$. Adopting shadow estimation based on magic or non-magic Clifford orbits, the associated variance can be upper bounded as follows:*

$$\max\left\{\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\rho\otimes\hat{\rho}_1)], \mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\rho)], \sqrt{\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\hat{\rho}_2)]}\right\} \lesssim \sqrt{3}\gamma_{d,k}\left(\|\mathfrak{O}\|_2^2 + \vartheta\right), \tag{J4}$$

where $\gamma_{d,k}$ is defined in Eq. (J3)*, and*

$$\vartheta = \max\left(0, \mathrm{tr}(\mathfrak{O})^2 - \sum_{a=1,2} \|\mathrm{tr}_a(\mathfrak{O})\|_2^2\right), \tag{J5}$$

and by $\lesssim$ we have omitted terms of order $\mathcal{O}(D^{-1})$.

Proof of Proposition 3. Note that

$$\mathbb{E}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1 \otimes \rho) = \mathbb{E}[\mathrm{tr}(\mathfrak{O}\rho \otimes \hat{\rho}_1)] = \mathbb{E}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1 \otimes \hat{\rho}_2)] = \mathrm{tr}(\mathfrak{O}\rho \otimes \rho), \tag{J6}$$

so we can drop the square of expectation in variance and only consider the expectation of the square.

For the first variance, define $\mathfrak{O}_\rho = \mathrm{tr}_1(\rho \otimes \mathbb{I} \cdot \mathfrak{O})$, then

$$\mathbb{E}[\mathrm{tr}(\mathfrak{O}_\rho\hat{\rho}_1)^2] \le \|\mathfrak{O}_\rho\|_{\mathrm{sh}}^2 \le \gamma_{d,k} \left\|\mathfrak{O}_\rho - \frac{\mathrm{tr}(\mathfrak{O}_\rho)}{D}\mathbb{I}\right\|_2^2 \le \gamma_{d,k}\|\mathfrak{O}_\rho\|_2^2. \tag{J7}$$

Since $\mathrm{tr}(\rho) = 1$, $\mathrm{tr}\big(\rho^2\big) < 1$, we have

$$\|\mathfrak{O}_\rho\|_2^2 = \mathrm{tr}\left(\mathrm{tr}_1(\rho \otimes \mathbb{I} \cdot \mathfrak{O})^2\right) \le \mathrm{tr}\big(\mathfrak{O}^2\big) = \|\mathfrak{O}\|_2^2. \tag{J8}$$

Thus we conclude that

$$\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\rho \otimes \hat{\rho}_1)] \le \gamma_{d,k}\|\mathfrak{O}\|_2^2. \tag{J9}$$

A similar result holds for the second variance $\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1 \otimes \rho)]$. Lemma 4 below provides a bound for the square of the final contribution. We have assumed that $\|\mathfrak{O}\|_2 \ge 1$, so $\|\mathfrak{O}\|_2 \le \|\mathfrak{O}\|_2^2$, and the claim follows. □

Lemma 4. *Suppose d is a prime, n is a positive integer, ρ is an operator on $\mathcal{H}_D$, and $\mathfrak{O}$ is an operator on $\mathcal{H}_D^{\otimes 2}$. Consider measurements based on magic or non-magic Clifford orbits, we have*

$$\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1 \otimes \hat{\rho}_2)] \lesssim 3\gamma_{d,k}^2 \left(\|\mathfrak{O}\|_2^2 - \sum_{a=1,2} \|\,\mathrm{tr}_a(\mathfrak{O})\|_2^2 + \mathrm{tr}(\mathfrak{O})^2\right) \tag{J10}$$

Proof of Lemma 4. To bound the variance of $\mathrm{tr}(\mathfrak{O}\hat{\rho}_1 \otimes \hat{\rho}_2)$, we follow the same approach as in Ref. [1] and define the following traceless variants of $\mathfrak{O}$:

$$\begin{aligned} \mathfrak{O}_0^{(1)} &= \mathrm{tr}_2(\mathfrak{O}) - \frac{\mathrm{tr}(\mathfrak{O})}{D}\mathbb{I}, \quad \mathfrak{O}_0^{(2)} = \mathrm{tr}_1(\mathfrak{O}) - \frac{\mathrm{tr}(\mathfrak{O})}{D}\mathbb{I}, \\ \mathfrak{O}_0^{(1,2)} &= \mathfrak{O} - \mathrm{tr}_2(\mathfrak{O}) \otimes \frac{\mathbb{I}}{D} - \frac{\mathbb{I}}{D} \otimes \mathrm{tr}_1(\mathfrak{O}) + \mathrm{tr}(\mathfrak{O})\frac{\mathbb{I}}{D} \otimes \frac{\mathbb{I}}{D}, \end{aligned} \tag{J11}$$

where $\mathrm{tr}_a(\cdot)$ with $a = 1, 2$ denotes the partial trace over the first and second system respectively. All three operators are traceless. Recall that the reconstruction map reads

$$\hat{\rho}_a = (D+1)U_a^\dagger|\mathbf{b}_a\rangle\langle\mathbf{b}_a|U_a - \mathbb{I} \quad \text{for } a = 1, 2. \tag{J12}$$

Using these expressions, we can rewrite $\mathrm{tr}(\mathfrak{O}\hat{\rho}_1 \otimes \hat{\rho}_2)$ as follows:

$$\begin{aligned} \mathrm{tr}(\mathfrak{O}\hat{\rho}_1 \otimes \hat{\rho}_2) =& \mathrm{tr}\left(\mathfrak{O} \cdot [(D+1)U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1 - \mathbb{I}] \otimes [(D+1)U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2 - \mathbb{I}]\right) \\ =& (D+1)^2\,\mathrm{tr}\left(\mathfrak{O}_0^{(1,2)}U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1 \otimes U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2\right) + \frac{\mathrm{tr}(\mathfrak{O})}{D^2} \\ &+ \frac{D+1}{D}\,\mathrm{tr}\left(\mathfrak{O}_0^{(1)}U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1\right) + \frac{D+1}{D}\,\mathrm{tr}\left(\mathfrak{O}_0^{(2)}U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2\right). \end{aligned} \tag{J13}$$

The second term is constant and does not contribute to the variance. For the remaining terms, same as before, we consider the expected square, i.e.,

$$\mathbb{E}\left[\left((D+1)^2\,\mathrm{tr}\left(\mathfrak{O}_0^{(1,2)}U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1 \otimes U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2\right) + \frac{D+1}{D}\sum_{a=1,2}\mathrm{tr}\left(\mathfrak{O}_0^{(a)}U_a^\dagger|\mathbf{b}_a\rangle\langle\mathbf{b}_a|U_a\right)\right)^2\right], \tag{J14}$$

and analyze on a case-by-case basis.

Linear terms: A direct calculation shows that

$$\mathbb{E}\left[\left(\frac{D+1}{D}\,\mathrm{tr}\left(\mathfrak{O}_0^{(a)}U_a^\dagger|\mathbf{b}_a\rangle\langle\mathbf{b}_a|U_a\right)\right)^2\right] = \frac{1}{D^2}\mathbb{E}\left[\mathrm{tr}\left(\mathfrak{O}_0^{(a)}\hat{\rho}_a\right)^2\right] \le \frac{1}{D^2}\left\|\mathfrak{O}_0^{(a)}\right\|_{\mathrm{sh}}^2 \le \frac{\gamma_{d,\mathcal{E}}}{D^2}\left\|\mathfrak{O}_0^{(a)}\right\|_2^2. \tag{J15}$$

Leading-order term: In this case, the derivation is more involved. We calculate that

$$\begin{aligned}
&\mathbb{E}\left[\left((D+1)^2\operatorname{tr}\left(\mathfrak{O}_0^{(1,2)}U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1\otimes U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2\right)\right)^2\right]\\
=&(D+1)^4\mathbb{E}_{U_1,U_2\sim\mathrm{Cl}(n,d)}\sum_{\mathbf{b}_1,\mathbf{b}_2\in\mathbb{F}_d^n}\langle\mathbf{b}_1|U_1\rho U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_2|U_2\rho U_2^\dagger|\mathbf{b}_2\rangle\operatorname{tr}\left[\left(\mathfrak{O}_0^{(1,2)}\right)^{\otimes 2}\cdot(U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1\otimes U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2)^{\otimes 2}\right]\\
=&(D+1)^4\mathbb{E}_{U_1,U_2\sim\mathrm{Cl}(n,d)}\sum_{\mathbf{b}_1,\mathbf{b}_2\in\mathbb{F}_d^n}\operatorname{tr}\left[\rho\otimes\rho\otimes\mathfrak{O}_0^{(1,2)}\otimes\mathfrak{O}_0^{(1,2)}\cdot(U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1\otimes U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2)^{\otimes 3}\right]\\
=&(D+1)^4D^2\operatorname{tr}\left[\rho\otimes\rho\otimes\mathfrak{O}_0^{(1,2)}\otimes\mathfrak{O}_0^{(1,2)}\cdot Q\left(\mathrm{orb}(\Psi)\right)^{(\mathrm{odd})}\otimes Q\left(\mathrm{orb}(\Psi)\right)^{(\mathrm{even})}\right],
\end{aligned}\tag{J16}$$

where $|\Psi\rangle$ is the fiducial state of the Clifford orbit. The superscript 'even' and 'odd' indicate on which subset of tensor factors the projectors act. Using the results in our companion paper, we can rewrite (J16) as

$$\begin{aligned}
&\frac{(D+1)^2}{(D+2)^2}\sum_{\mathcal{T},\mathcal{T}'\in\Sigma(d)}\hat{\kappa}(\Psi,\mathcal{T})\hat{\kappa}(\Psi,\mathcal{T}')\operatorname{tr}\left(\rho\otimes\rho\otimes\mathfrak{O}_0^{(1,2)}\otimes\mathfrak{O}_0^{(1,2)}\cdot R(\mathcal{T})^{(\mathrm{odd})}\otimes R(\mathcal{T}')^{(\mathrm{even})}\right)\\
=&\frac{(D+1)^2}{(D+2)^2}\sum_{\mathcal{T},\mathcal{T}'\in\Sigma(d)}\hat{\kappa}(\psi,\mathcal{T})\hat{\kappa}(\psi,\mathcal{T}')\operatorname{tr}\left(\rho\otimes\rho\cdot\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{O}_0^{(1,2)})\right),
\end{aligned}$$

where we have defined the 'shadow map'

$$\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A}):=\operatorname{tr}_{BCB'C'}\left[\mathbb{I}\otimes\mathbb{I}\otimes\mathfrak{A}\otimes\mathfrak{A}^\dagger\cdot R(\mathcal{T})^{(\mathrm{odd})}\otimes R(\mathcal{T}')^{(\mathrm{even})}\right],\tag{J17}$$

for any linear operator $\mathfrak{A}$ on $\mathcal{H}_D^{\otimes 2}$. We use A,B,C (A',B',C') to denote the first, second and third party $R(\mathcal{T})^{(\mathrm{odd})}$ $(R(\mathcal{T})^{(\mathrm{even})})$ acts on. If $|\Psi\rangle\in\mathrm{Stab}(n,d)$, $\hat{\kappa}(\Psi,\mathcal{T})=\frac{D+2}{D+d}$ for all $\mathcal{T}\in\Sigma(d)$.

Since in the case under consideration $\mathfrak{A}=\mathfrak{O}_0^{(1,2)}$ is traceless, evidently, if either $\mathcal{T}$ or $\mathcal{T}'$ corresponds to an element in $\{\mathbb{1},\tau_{12},\tau_{13}\}$, then $\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A})=0$. Define $\tilde{\Sigma}(d)=\Sigma(d)\backslash\{\mathbb{1},\tau_{12},\tau_{13}\}$. According to Lemma 5, we have

$$\|\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{O}_0^{(1,2)})\|\leq\|\mathfrak{O}_0^{(1,2)})\|_2^2\quad\mathcal{T},\mathcal{T}'\in\Sigma(d).\tag{J18}$$

Then the leading-order term is upper bounded by

$$\frac{(D+1)^2}{(D+2)^2}\sum_{\mathcal{T},\mathcal{T}'\in\tilde{\Sigma}(d)}\hat{\kappa}(\Psi,\mathcal{T})\hat{\kappa}(\Psi,\mathcal{T}')\|\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{O}_0^{(1,2)})\|\leq\left(\sum_{\mathcal{T}\in\tilde{\Sigma}(d)}|\hat{\kappa}(\Psi,\mathcal{T})|\right)^2\|\mathfrak{O}_0^{(1,2)}\|_2^2\leq\gamma_{d,k}^2\|\mathfrak{O}_0^{(1,2)}\|_2^2.\tag{J19}$$

In the case $|\Psi\rangle\in\mathrm{Stab}(n,d)$, the last inequality is saturated. When the measurement primitive is a qudit magic orbit, we have

$$\sum_{\mathcal{T}\in\tilde{\Sigma}(d)}|\hat{\kappa}(\Psi,\mathcal{T})|\leq\sum_{\mathcal{T}\in\Sigma(d)}|\hat{\kappa}(\Psi,\mathcal{T})|\leq\gamma_{d,k}.\tag{J20}$$

See our companion paper for a detailed proof.

Bounds on cross terms: Recall that

$$\mathbb{E}[U_a^\dagger|b_a\rangle\langle b_a|U_a]=\frac{\rho+\mathbb{I}}{D+1},\tag{J21}$$

we can effectively get rid of the linear contribution:

$$\left(\frac{D+1}{D}\right)^2\mathbb{E}\left[\prod_{a=1,2}\operatorname{tr}\left(\mathfrak{O}_0^{(a)}U_a^\dagger|\mathbf{b}_a\rangle\langle\mathbf{b}_a|U_a\right)\right]=\frac{1}{D^2}\operatorname{tr}\left(\mathfrak{O}_0^{(1)}\rho\right)\operatorname{tr}\left(\mathfrak{O}_0^{(2)}\rho\right)\leq\frac{1}{2D^2}\left(\|\mathfrak{O}_0^{(1)}\|^2+\|\mathfrak{O}_0^{(2)}\|^2\right).\tag{J22}$$

We use Eqs. (J15) (J16) and (J17) to upper bound the remaining terms:

$$\begin{aligned}
&\frac{(D+1)^3}{D}\mathbb{E}\left[\mathrm{tr}\left(\mathfrak{O}_0^{(1,2)}U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1\otimes U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2\right)\mathrm{tr}\left(\mathfrak{O}_0^{(a)}U_a^\dagger|\mathbf{b}_a\rangle\langle\mathbf{b}_a|U_a\right)\right]\\
\leq&\frac{1}{2}\left\{\mathbb{E}\left[\left((D+1)^2\,\mathrm{tr}\left(\mathfrak{O}_0^{(1,2)}U_1^\dagger|\mathbf{b}_1\rangle\langle\mathbf{b}_1|U_1\otimes U_2^\dagger|\mathbf{b}_2\rangle\langle\mathbf{b}_2|U_2\right)\right)^2\right]+\mathbb{E}\left[\left(\frac{D+1}{D}\,\mathrm{tr}\left(\mathfrak{O}_0^{(a)}U_a^\dagger|\mathbf{b}_a\rangle\langle\mathbf{b}_a|U_a\right)\right)^2\right]\right\}\\
\leq&\frac{\gamma_{d,k}}{2}\left(\gamma_{d,k}\|\mathfrak{O}_0^{(1,2)}\|_2^2+\frac{1}{D^2}\|\mathfrak{O}_0^{(a)}\|_2^2\right).
\end{aligned}\tag{J23}$$

Full variance bound: Combine all individual bounds together, we have

$$\mathrm{Var}\left[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\hat{\rho}_2)\right]\leq 3\gamma_{d,k}^2\|\mathfrak{O}_0^{(1,2)}\|_2^2+\frac{2\gamma_{d,k}}{D}\sum_{a=1,2}\|\mathfrak{O}_0^{(a)}\|_2^2+\frac{1}{D^2}\sum_{a=1,2}\|\mathfrak{O}_0^{(a)}\|^2\tag{J24}$$

Note that

$$\begin{aligned}
\|\mathfrak{O}_0^{(1,2)})\|_2^2&=\left(1-\frac{2}{D}\right)\left(\|\mathfrak{O}\|_2^2-\sum_{a=1,2}\|\,\mathrm{tr}_a(\mathfrak{O})\|_2^2\right)+\mathrm{tr}(\mathfrak{O})^2\\
\|\mathfrak{O}_0^{(a)}\|_2^2&\leq\|\,\mathrm{tr}_{\bar{a}}(\mathfrak{O})\|_2^2-\frac{\mathrm{tr}(\mathfrak{O})^2}{D},\\
\|\mathfrak{O}_0^{(a)}\|&\leq\|\,\mathrm{tr}_{\bar{a}}(\mathfrak{O})\|-\frac{|\,\mathrm{tr}(\mathfrak{O})|}{D},
\end{aligned}\tag{J25}$$

where $\bar{a}$ denotes the party other than a.

Substituting these expressions into Eq. (J24), finally we have

$$\begin{aligned}
&\mathrm{Var}\left[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\hat{\rho}_2)\right]\\
\leq&3\gamma_{d,k}^2\|\mathfrak{O}\|_2^2+\left(-3\gamma_{d,k}^2+\frac{2}{D}(\gamma_{d,k}+1)\right)\sum_{a=1,2}\|\,\mathrm{tr}_a(\mathfrak{O})\|_2^2+\frac{2}{D^2}\sum_{a=1,2}\|\,\mathrm{tr}_a(\mathfrak{O})\|^2+3\gamma_{d,k}^2\,\mathrm{tr}(\mathfrak{O})^2\\
\lesssim&3\gamma_{d,k}^2\left(\|\mathfrak{O}\|_2^2-\sum_{a=1,2}\|\,\mathrm{tr}_a(\mathfrak{O})\|_2^2+\mathrm{tr}(\mathfrak{O})^2\right),
\end{aligned}\tag{J26}$$

where by $\lesssim$ we have omitted terms of order $\mathcal{O}(D^{-1})$. □

Lemma 5. *Suppose d is an odd prime and $\mathfrak{A}$ is a linear operator on $\mathcal{H}_D^{\otimes 2}$. Then*

$$\|\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A})\|\leq\|\mathfrak{A}\|_2^2\quad\forall\mathcal{T},\mathcal{T}'\in\mathscr{T}_{\mathrm{ns}}.\tag{J27}$$

If in addition $\mathfrak{A}$ is traceless, then

$$\|\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A})\|\leq\|\mathfrak{A}\|_2^2\quad\forall\mathcal{T}\in\Sigma(d).\tag{J28}$$

Proof of Lemma 5. The proof of Lemma 5 is similar to that of Lemma 17 in the companion paper.

First we consider the case where $\mathcal{T},\mathcal{T}'\in\mathscr{T}_{\mathrm{ns}}$. To simplify notation, we also focus on the case $n=1$. The basic idea admits straightforward generalization.

$$\begin{aligned}
\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A}):&=\mathrm{tr}_{BCB'C'}\left[\mathbb{I}\otimes\mathbb{I}\otimes\mathfrak{A}\otimes\mathfrak{A}^\dagger\cdot R(\mathcal{T})^{(\mathrm{odd})}\otimes R(\mathcal{T}')^{(\mathrm{even})}\right]\\
&=\sum_{\substack{(\mathbf{x};\mathbf{y})\in\mathcal{T}\\(\mathbf{x}';\mathbf{y}')\in\mathcal{T}'}}\mathrm{tr}_{BCB'C'}\left[|\mathbf{x},\mathbf{x}'\rangle\langle\mathbf{y},\mathbf{y}'|\left(\mathbb{I}\otimes\mathbb{I}\otimes\mathfrak{A}\otimes\mathfrak{A}^\dagger\right)\right]\\
&=\sum_{\substack{(\mathbf{x};\mathbf{y})\in\mathcal{T}\\(\mathbf{x}';\mathbf{y}')\in\mathcal{T}'}}\mathfrak{A}_{y_2,x_2;y_2',x_2'}(\mathfrak{A}^\dagger)_{y_3,x_3;y_3',x_3'}|x_1,x_1'\rangle\langle y_1,y_1'|,
\end{aligned}\tag{J29}$$

where x_1, x_2, x_3 (x_1', x_2', x_3') are the three entries of $\mathbf{x}$ ($\mathbf{x}'$), and y_1, y_2, y_3 (y_1', y_2', y_3') are the three entries of $\mathbf{y}$ ($\mathbf{y}'$). We introduce the following two linear maps from $\mathcal{T}$ to $\mathbb{F}_d^3$,

$$f_i : (\mathbf{x}; \mathbf{y}) \mapsto (x_1; x_i; y_i), \quad i = 2, 3. \tag{J30}$$

Both f_2 and f_3 are injective and surjective. See Lemma 17 in the companion paper for a proof. Consequently,

$$\{(y_2; x_2) \,|\, (\mathbf{x}; \mathbf{y}) \in \mathcal{T}, x_1 = a\} = \{(y_3; x_3) \,|\, (\mathbf{x}; \mathbf{y}) \in \mathcal{T}, x_1 = a\} = \mathbb{F}_d^2 \quad \forall a \in \mathbb{F}_d. \tag{J31}$$

Therefore,

$$\sum_{b,b' \in \mathbb{F}_d} |\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A})_{a,b;a',b'}| \leq \sum_{b,c,b',c' \in \mathbb{F}_d} |\mathfrak{A}_{b,c;b',c'}||\mathfrak{A}_{\alpha(b,c);\alpha'(b',c')}| \leq \sum_{b,c,b',c' \in \mathbb{F}_d} |\mathfrak{A}_{b,c;b',c'}|^2 = \|\mathfrak{A}\|_2^2 \quad \forall a \in \mathbb{F}_d, \tag{J32}$$

where α, α' are permutations on $\mathbb{F}_d^2$, which may depend on a and a', respectively. By a similar reasoning we can deduce that

$$\sum_{b,b' \in \mathbb{F}_d} |\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A})_{b,a;b',a'}| \leq \|\mathfrak{A}\|_2^2 \quad \forall a \in \mathbb{F}_d. \tag{J33}$$

The above two equations together imply Eq. (J27).

Since $\Sigma(d) = \mathscr{T}_{\text{sym}} \cup \mathscr{T}_{\text{ns}}$, now we consider the $\mathcal{T}$ in $\mathscr{T}_{\text{sym}} := \{\mathcal{T}_O \,|\, O \in S_3\}$. If $\mathfrak{A}$ is traceless, direct calculation shows that if either $\mathcal{T}$ or $\mathcal{T}'$ corresponds to an element in $\{\mathbb{1}, \tau_{12}, \tau_{13}\}$, then $\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A}) = 0$. Note that in addition to $\mathscr{T}_{\text{ns}}$, Eq. (J31) also holds for $\mathcal{T}_{\tau_{23}}$. Thus for all $\mathcal{T}, \mathcal{T}' \in \mathscr{T}_{\text{ns}} \cup \{\mathcal{T}_{\tau_{23}}\}$, we have

$$\|\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A})\| \leq \|\mathfrak{A}\|_2^2. \tag{J34}$$

If $\mathcal{T}, \mathcal{T}' \in \{\mathcal{T}_\zeta, \mathcal{T}_{\zeta^2}\}$, according to Lemma 6 in Ref. [1], Eq. (J34) also holds.

If only one of $\mathcal{T}, \mathcal{T}'$ is in $\{\mathcal{T}_\zeta, \mathcal{T}_{\zeta^2}\}$, say $\mathcal{T} \in \mathscr{T}_{\text{ns}} \cup \{\mathcal{T}_{\tau_{23}}\}, \mathcal{T}' = \mathcal{T}_\zeta$, assuming $(x_1', x_2', x_3') = (a, b, c)$, we can rewrite Eq. (J29) as

$$\begin{aligned} \mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A}) &= \sum_{\substack{(\mathbf{x},\mathbf{y}) \in \mathcal{T} \\ a,b,c \in \mathbb{F}_d}} \mathfrak{A}_{y_2,x_2;c,b} (\mathfrak{A}^\dagger)_{y_3,x_3;a,c} |x_1, a\rangle\langle y_1, b|, \\ &= \sum_{\substack{x_1,x_2,y_2 \in \mathbb{F}_d \\ a,b,c \in \mathbb{F}_d}} \mathfrak{A}_{y_2,x_2;c,b} (\mathfrak{A}^\dagger)_{\alpha(y_2,x_2);a,c} |x_1, a\rangle\langle \beta(x_1), b|, \end{aligned} \tag{J35}$$

where α, β are permutations on $\mathbb{F}_d^2$ and $\mathbb{F}_d$, respectively.

Define the vectorization of submatrices of $\mathfrak{A}$ as

$$v(\mathfrak{A})^{(i,j)} = (\mathfrak{A}_{0,0;i,j}, \cdots, \mathfrak{A}_{0,d-1;i,j}, \mathfrak{A}_{1,0;i,j}, \cdots, \mathfrak{A}_{1,d-1;i,j}, \cdots, \mathfrak{A}_{d-1,0;i,j}, \cdots, \mathfrak{A}_{d-1,d-1;i,j}), \tag{J36}$$

for $i, j \in \mathbb{F}_d$, then $v(\mathfrak{A})^{(i,j)} \in \mathbb{C}^{d^2}$. Define an inner product on $\mathbb{C}^{d^2}$

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\top P_\alpha \mathbf{v}, \quad \mathbf{u}, \mathbf{v} \in \mathbb{C}^{d^2}, \tag{J37}$$

where P_α is a permutation matrix associated with α. Then we have

$$\sum_{x_2,y_2,c \in \mathbb{F}_d} \mathfrak{A}_{y_2,x_2;c,b} (\mathfrak{A}^\dagger)_{\alpha(y_2,x_2);a,c} = \sum_{c \in \mathbb{F}_d} \langle v(\mathfrak{A})^{(a,c)}, v(\mathfrak{A})^{(c,b)} \rangle \leq \sum_{c \in \mathbb{F}_d} \|v(\mathfrak{A})^{(a,c)}\|_2 \cdot \|v(\mathfrak{A})^{(c,b)}\|_2. \tag{J38}$$

We introduce a matrix $\mathfrak{M} \in \mathbb{R}_+^{d \times d}$ such that

$$\mathfrak{M}_{ij} = \|v(\mathfrak{A})^{(i,j)}\|_2. \tag{J39}$$

Then

$$\begin{aligned} \mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A}) &\leq \sum_{a,b,c,x_1 \in \mathbb{F}_d} \mathfrak{M}_{ac} \mathfrak{M}_{cb} |x_1, a\rangle\langle \beta(x_1), b| = \left(\sum_{a,b \in \mathbb{F}} (\mathfrak{M}^2)_{ab} |a\rangle\langle b| \right) \otimes \left(\sum_{x_1 \in \mathbb{F}_d} |x_1\rangle\langle \beta(x_1)| \right) \\ &= \mathfrak{M}^2 \otimes P_\beta, \end{aligned} \tag{J40}$$

where P_β is a permutation matrix associated with β. Finally, we can calculate that

$$\|\mathcal{R}_{\mathcal{T},\mathcal{T}'}(\mathfrak{A})\| \leq \|\mathfrak{M}^2\| \cdot \|P_\beta\| = \|\mathfrak{M}\|^2 \leq \|\mathfrak{M}\|_2^2. \tag{J41}$$

It can be easily verified that

$$\|\mathfrak{M}\|_2^2 = \sum_{i,j\in\mathbb{F}_d} \|v(\mathfrak{A})^{(i,j)}\|_2^2 = \sum_{i,j,k,l\in\mathbb{F}_d} |\mathfrak{A}_{k,l;i,j}|^2 = \|\mathfrak{A}\|_2^2. \tag{J42}$$

Thus we conclude that in this case Eq. (J34) is still valid.

The above discussion, together with Eq. (J27), implies Eq. (J28).

□

2. Variance bounds for local Clifford measurements

Proposition 4. *Suppose that $\mathfrak{O}$ describes a quadratic function $\mathrm{tr}(\mathfrak{O}\rho\otimes\rho)$ that acts on at most m-qubits in both the first and the second system. Let $\tilde{\mathfrak{O}}$ denote the non-trivial part of $\mathfrak{O}$, $\mathfrak{O} = \tilde{\mathfrak{O}}\otimes\mathbb{I}^{\otimes 2(n-m)}$, and $\|\tilde{\mathfrak{O}}\|_2 \geq 1$. Adopting local Clifford measurements, the associated variance can be upper bounded as follows:*

$$\max\left(\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\rho\otimes\hat{\rho}_1)], \mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\rho)], \sqrt{\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\hat{\rho}_2)]}\right) \leq d^m\|\tilde{\mathfrak{O}}\|_2^2. \tag{J43}$$

Proof. Same as the previous case, we omit the square of expectation and only consider the expectation of square. For the first and second variance, we have

$$\mathbb{E}[(\mathrm{tr}(\mathfrak{O}\rho\otimes\hat{\rho}_1))^2] \leq \|\mathfrak{O}_\rho\|_{\mathrm{sh}}^2 \leq d^m\|\tilde{\mathfrak{O}}_\rho\|_2^2, \tag{J44}$$

where $\mathfrak{O}_\rho = \mathrm{tr}_1(\rho\otimes\mathbb{I}\cdot\mathfrak{O})$, and $\tilde{\mathfrak{O}}_\rho$ denotes its non-trivial part. Note that

$$\|\tilde{\mathfrak{O}}_\rho\|_2^2 = \mathrm{tr}\left(\tilde{\mathfrak{O}}_\rho^2\right) = \mathrm{tr}\left(\mathrm{tr}_1(\rho_m\otimes\mathbb{I}^{\otimes m}\cdot\tilde{\mathfrak{O}})^2\right) \leq \mathrm{tr}\left(\tilde{\mathfrak{O}}^2\right) = \|\tilde{\mathfrak{O}}_\rho\|_2^2, \tag{J45}$$

where ρ_m is the reduced density matrix of the qudits $\mathfrak{O}$ acts non-trivially on. Then

$$\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\rho\otimes\hat{\rho}_1)] \leq d^m\|\tilde{\mathfrak{O}}\|_2^2. \tag{J46}$$

As for the third variance, recall that the reconstruction map associated with Weyl measurement has single-qudit tensor product structure, thus the tensor product of two snapshots $\hat{\rho}_1\otimes\hat{\rho}_2$ of state ρ may be viewed as a single snapshot of the tensor product state $\varrho := \rho\otimes\rho$:

$$\begin{aligned}\hat{\rho}_1\otimes\hat{\rho}_2 =& \bigotimes_{i=1}^{n}\left(\mathcal{M}^{-1}(U_1^{(i)}|\mathbf{b}_1^{(i)}\rangle\langle\mathbf{b}_1^{(i)}|U_1^{(i)\dagger})\right)\bigotimes_{i=1}^{n}\left(\mathcal{M}^{-1}(U_2^{(i)}|\mathbf{b}_2^{(i)}\rangle\langle\mathbf{b}_2^{(i)}|U_2^{(i)\dagger})\right)\\ =& \bigotimes_{i=1}^{2n}\mathcal{M}^{-1}(U^{(i)}|\mathbf{b}^{(i)}\rangle\langle\mathbf{b}^{(i)}|U^{(i)\dagger}) = \hat{\varrho}.\end{aligned} \tag{J47}$$

By assumption, $\mathfrak{O}$ is an observable acting on at most $2m$ qudits. From Theorem 2 we obtain

$$\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\hat{\rho}_1\otimes\hat{\rho}_2)] \leq \|\mathfrak{O}\|_{\mathrm{sh}}^2 \leq d^{2m}\|\tilde{\mathfrak{O}}\|_2^2, \tag{J48}$$

and the variance bound in Proposition 4 immediately follows. □

3. Sample complexities for purity and 2nd Rényi entropy

We focus on estimating the purity of a subsystem, as one of the most widely concerned quadratic functions. The associated observable is the swap operator $\mathbb{S}$ (defined as $\mathbb{S}|\psi\rangle\otimes|\phi\rangle = |\phi\rangle\otimes|\psi\rangle$ for all states $|\psi\rangle$ and $|\phi\rangle$):

$$\mathrm{tr}\left(\rho^2\right) = \mathrm{tr}\left(\mathbb{S}\cdot\rho\otimes\rho\right). \tag{J49}$$

Suppose $\max\left(\mathrm{Var}[\mathrm{tr}(\mathbb{S}\rho\otimes\hat{\rho}_1)],\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\rho)],\sqrt{\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\hat{\rho}_2)]}\right)$ is upper bounded by σ_{pu}. The sample complexity for estimating purity is $8\sigma_{pu}/\varepsilon^2$. The second Rényi entropy is closely related with purity

$$S_2(\rho) = -\log\mathrm{tr}\big(\rho^2\big). \tag{J50}$$

Using the law of propagation of uncertainties, we can calculate the variance associated with S_2:

$$\sigma_{S_2} = \frac{1}{\mathrm{tr}(\rho^2)}\cdot\sigma_{pu}, \tag{J51}$$

and the sample complexity for estimating S_2 immediately follows:

$$\frac{8\sigma_{pu}}{\mathrm{tr}(\rho^2)\varepsilon^2}. \tag{J52}$$

Measurement based on Clifford orbits: We shall calculate σ_{pu} explicitly based on Proposition 3. Several simplifications can be made using the properties of the swap operator.

For the first and second variance, note that

$$\mathbb{S}_\rho := \mathrm{tr}_1\left(\rho\otimes\mathbb{I}\cdot\mathbb{S}\right) = \mathrm{tr}(\mathbb{I})\cdot\rho = D\rho, \tag{J53}$$

and we obtain

$$\mathrm{Var}[\mathrm{tr}(\mathbb{S}\rho\otimes\hat{\rho}_1)] \le \|\mathbb{S}_\rho\|_{\mathrm{sh}}^2 \le \gamma_{d,k}\|D\rho-\mathbb{I}\|_2^2 \le \gamma_{d,k}D\left(D\,\mathrm{tr}\big(\rho^2\big)-1\right). \tag{J54}$$

For the third variance, we have

$$\mathrm{tr}(\mathbb{S}) = D, \quad \mathrm{tr}_1(\mathbb{S}) = \mathrm{tr}_2(\mathbb{S}) = \mathbb{I}, \tag{J55}$$

and consequently,

$$\mathbb{S}_0^{(1)} = \mathbb{S}_0^{(2)} = 0, \quad \mathbb{S}_0^{(1,2)} = \mathbb{S} - \frac{1}{D}\cdot\mathbb{I}\otimes\mathbb{I}. \tag{J56}$$

Substituting into Eq. (J24), we obtain

$$\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\hat{\rho}_2)] \le 3\gamma_{d,k}^2\left\|\mathbb{S}-\frac{1}{D}\cdot\mathbb{I}\otimes\mathbb{I}\right\|_2^2 = 3\gamma_{d,k}^2(D-1)^2. \tag{J57}$$

Finally, we conclude that

$$\max\left\{\mathrm{Var}[\mathrm{tr}(\mathbb{S}\rho\otimes\hat{\rho}_1)],\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\rho)],\sqrt{\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\hat{\rho}_2)]}\right\} \le \sqrt{3}\gamma_{d,k}D =: \sigma_{pu}, \tag{J58}$$

and the sample complexities for estimating purity and 2nd Rényi entropy follow immediately.

Measurement based on local Clifford measurements: In this case we have

$$\mathrm{Var}[\mathrm{tr}(\mathfrak{O}\rho\otimes\hat{\rho}_1)] \le \mathbb{E}[(\mathrm{tr}(\mathbb{S}\rho\otimes\hat{\rho}_1))^2] = \mathbb{E}[(\mathrm{tr}(\rho\hat{\rho}_1))^2] \le D^2, \tag{J59}$$

and

$$\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\hat{\rho}_2)] \le \mathbb{E}[(\mathrm{tr}(\hat{\rho}_1\hat{\rho}_2))^2] \le (D^2+D-1)^2. \tag{J60}$$

Thus in this case we have

$$\max\left\{\mathrm{Var}[\mathrm{tr}(\mathbb{S}\rho\otimes\hat{\rho}_1)],\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\rho)],\sqrt{\mathrm{Var}[\mathrm{tr}(\mathbb{S}\hat{\rho}_1\otimes\hat{\rho}_2)]}\right\} \le D^2+D-1 =: \sigma_{pu}, \tag{J61}$$

Optimal quantum metrology of two-photon absorption parameter and related physics with photon number statistics

Changhyoup Lee[1] *

[1] *Korea Research Institute of Standards and Science, Daejeon 34113, Korea*

Abstract. Two-photon absorption (TPA) is a crucial nonlinear optical process with significant applications, yet its precise measurement is challenging. We explore using quantum light to enhance TPA parameter estimation via Quantum Fisher Information, showing that optimized discrete-variable quantum states exhibit a quantum advantage over classical benchmarks. Comparisons with a squeezed vacuum state and the use of photon counting as a nearly optimal detection scheme are discussed. Additionally, we find that TPA can be enhanced without entanglement by adjusting the photon number statistics of the injected light, offering new insights into the related physics of TPA.

Keywords: Quantum metrology, Two-photon absorption, Photon number statistics

1 Optimal quantum metrology [1]

Two-photon absorption (TPA) is a crucial nonlinear optical process with significant applications. Despite its importance, precisely measuring and characterizing TPA parameters is challenging due to the weak nature of the process and the discrete nature of light.

In this work, we study the use of quantum light to enhance TPA parameter estimation precision. Quantum Fisher information (QFI) is employed to quantify the information about the parameter, leading to a fundamental precision bound through the quantum Cramer-Rao inequality. We optimize discrete variable (DV) quantum states to maximize QFI for given losses, revealing a quantum advantage compared to classical benchmarks. Our results show that the Fock state is optimal for large TPA losses, while a superposition of vacuum and a particular Fock state is optimal for small losses. This differs from single-photon absorption, where the Fock state is optimal across all parameters.

Although optimal DV quantum states are theoretically interesting, they are challenging to realize. Therefore, we investigate the performance of a practical quantum state, the single-mode squeezed vacuum state. In comparison with the coherent state, the squeezed state outperforms for small TPA losses but underperforms in the intermediate regime and becomes comparable in the large loss limit. These behaviors can be understood by the difference between even and odd number Fock states, which are also analyzed. Interestingly, the QFI for even number states diverges in both large and small loss limits, while that for odd number states diverges only in the small loss limit.

We also examine the photon number counting scheme as a practical measurement setup, demonstrating that it offers nearly optimal performance compared to the QFI bound for the studied states in a wide range of TPA losses. This work provides valuable insights into quantum-enhanced TPA parameter estimation, and paves the way for its potential application in TPA imaging techniques.

*changhyoup.lee@gmail.com

2 Enhanced two-photon absorption [2]

TPA has been extensively studied, revealing that entangled photon pairs from spontaneous parametric down-conversion (SPDC) enhance TPA rates linearly with photon flux, unlike the quadratic scaling with coherent light. This enhancement, due to correlated photon arrival times, holds potential for TPA-based imaging and spectroscopy. Despite numerous studies, the role of entanglement in TPA enhancement remains unclear, and current experiments have yet to achieve it.

We are thus motivated to explore quantum-enhanced TPA without entanglement by engineering the input state's photon number distribution. Using a single-mode approximation, it demonstrates that optimizing the probe state can achieve similar or greater TPA enhancement without entanglement. The key factor influencing TPA is the zero-delay second-order coherence function of the input state. The squeezed vacuum state, with its higher auto-correlation function, outperforms coherent light, which is less effective than even thermal light in TPA.

By confirming that entanglement is not essential for quantum-enhanced TPA, the findings provide significant insights into the quantum features driving this process and pave the way for more effective TPA-based applications.

References

[1] A. Karsa, R. Nair, A. Chia, K. -G. Lee, and C. Lee. Optimal quantum metrology for two-photon absorption. *Accepted in Quantum Science and Technology*

[2] J. Park, A. Karsa, S. Panahiyan, F. Schlawin, K. -G. Lee, and C. Lee. Qnautum-enhanced two-photon absorption without entanglement. *in preparation*

Nonstabilizerness enhances the thrifty shadow estimation

Datong Chen[1 2 3 *] Huangjun Zhu[1 2 3 †]

[1] *State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai 200433, China.*

[2] *Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China.*

[3] *Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China.*

Abstract. Classical shadow has emerged as an effective method for efficiently estimating quantum systems through randomized measurements. Many works have been presented to study the classical shadow. One of the improved protocols is the thrifty shadow, which reduces resource consumption by employing multiple independent measurements per unitary. Here, we find that the performance of the thrifty shadow is directly related to the nonstabilizerness in the protocol. Specifically, we analyze the effects of nonstabilizerness in both the states and the unitary ensembles. Our findings revise the previous understanding of thrifty shadows with Clifford measurements, highlighting their broad applicability in certifying quantum states. Moreover, we propose a new configuration of the non-Clifford gates interleaved Clifford circuits. It reduces the requirements of an abundance of Clifford layers and shows prospects in other applications.

Keywords: Shadow Estimation, Nonstabilizerness, Interleaved circuits

1 Introduction

As a significant advancement in quantum state learning, the classical shadow [1] has attracted wide attention. In contrast to the inefficiencies of traditional quantum tomography for large quantum systems, classical shadow offers a general framework for efficiently estimating quantum systems through randomized measurements. The fundamental step in this protocol involves randomly applying a unitary transformation to the unknown state and performing a measurement in each round, serving as the foundation for efficient information extraction from the unknown state. However, making multiple changes to the unitary requires resetting the experimental configuration, leading to significant resource consumption in practice. To address this challenge, recent research has proposed a protocol with reduced resource consumption, known as thrifty shadow estimation [2, 3, 4]. The primary modification of this protocol is straightforward: we apply several independent measurements for each unitary. Given that repeating the same unitary transformation is less resource-consuming, this protocol is anticipated to be more feasible compared to the original approach.

In contrast to the original classical shadow, the thrifty shadow with Clifford measurements proves ineffective in certain scenarios. As demonstrated in Ref. [2], for a thrifty shadow protocol to universally outperform the classical shadow, its unitary ensemble must form a unitary 4-design. However, the lack of an exact construction scheme for unitary 4-designs makes implementing an efficient thrifty shadow protocol highly challenging.

2 Main Results

In this work, we explore how nonstabilizerness enhances the performance of thrifty shadows from two aspects: nonstabilizerness of the state and that of the unitary ensemble.

We start with some important preliminaries about this work. Compared to the original classical shadow, each unitary in the thrifty shadow protocol is reused R times. As a consequence, the variance using the thrifty shadow to estimate an observable O with respect to the state ρ is now given by:

$$V_R(O,\rho) = \frac{1}{R}V(O,\rho) + \frac{R-1}{R}V_*(O,\rho), \quad (1)$$

where $V(O,\rho)$ is the variance using the original classical shadow. It's important to note that the sample complexity is proportional to $RV_R(O,\rho)$, which is greater than $V(O,\rho)$. Therefore, $V_*(O,\rho)$ characterizes the additional sample complexity introduced by thrifty shadow estimation. On the other hand, the resource cost of reusing a circuit is considerably lower than that of introducing a new circuit. If $V_*(O,\rho)$ is negligible, achieving the same accuracy with fewer resources becomes feasible.

*22110190002@m.fudan.edu.cn

†zhuhuangjun@fudan.edu.cn

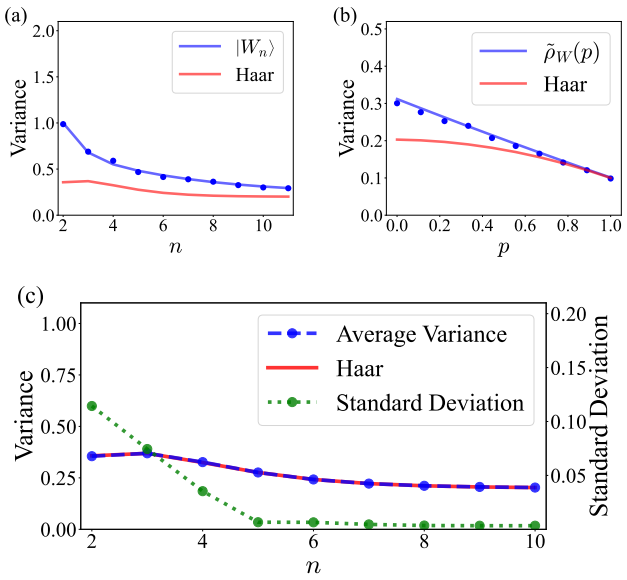


Figure 1: The statistical variance of the thrifty shadow to certify various families of states with respect to Clifford group. The circled dots are the numerical results. In (a) and (b), the lines represent the theoretical results. (a) Estimation of the fidelity of W states, parameterized by the qubits number n. (b) Estimation of the fidelity of the depolarized W states: $\tilde{\rho} = (1-p)\rho + pI/d$ and the ideal state is chosen as the 10-qubit W state. (c) Estimation of the fidelity of a random state. The blue dashed line represents the average variance of estimating 1000 random states from the Haar measure. The green dotted line indicates the standard deviation of the variances.

2.1 Nonstablizerness from the state

The previous work argued that the thrifty shadow with Clifford measurements is inefficient based on the results of certifying a stabilizer state. Surprisingly, we find that the Clifford group is suitable for certifying most states instead. We demonstrate that the performance of the thrifty shadow with respect to the Clifford group hinges on the stabilizer 2-Rényi entropy of the target state [5], which characterizes its nonstabilizerness:

Theorem 1 *Suppose $\rho = |\phi\rangle\langle\phi|$ is any n-qubit state and the observable $O = |\phi\rangle\langle\phi| - I/d$, then the variance of the thrifty shadow using the Clifford group reads*

$$V_R(O,\rho) = \frac{1}{R}\frac{2(d-1)}{d+2} + \frac{R-1}{R}\frac{2(d+1)2^{-M_2(|\phi\rangle)} - 4}{d+2}, \tag{2}$$

where $M_2(|\phi\rangle)$ is the stabilizer 2-Rényi entropy of state $|\phi\rangle$ defined by [5]

$$M_2(|\phi\rangle) = -\log_2 \frac{1}{d}\sum_{P\in\mathcal{P}_n} \langle\phi|P|\phi\rangle^4. \tag{3}$$

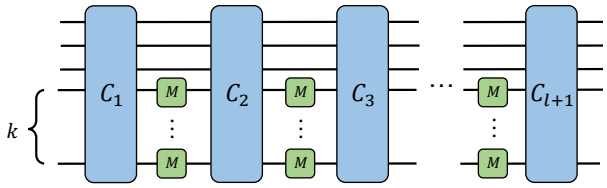


Figure 2: The circuits model of l-layer $\hat{M}_k$-interleaved Clifford circuits. C_1, C_2, ...,C_{l+1} are randomly selected in the Clifford group. In this configuration, the M gates can be replaced by any single-phase gate.

In the worst case that the target state is a stabilizer state, the variance $V_R(O,\rho)$ is equivalent to $V(O,\rho)$, indicating that reusing the same circuits does not obtain extra information. However, the thrifty shadow with respect to the Clifford group begins to show benefits when $M_2(|\phi\rangle)$ is large. Given that most states are expected to have a high stabilizer 2-Rényi entropy [5], the Clifford group is generally suitable for certifying states. Besides the theoretical conclusion, this result can also be verified by numerical experiment in Fig. 1(c). We find that using the thrifty shadow with the Clifford measurements to certify a random state from the Haar measure is efficient with high probability. We also examine various state families, including W states, Greenberger–Horne–Zeilinger (GHZ) states, and a family of states constructed by the non-Clifford gates. Here we take the W state as an example, seeing in Fig. 1(a). For W states, the stabilizer 2-Rényi entropy is of order $\log_2 n^2$, indicating that the thrifty shadow with Clifford measurements also benefits the fidelity estimation. Since the prepared states in realistic experiments may not be ideal, we also consider states affected by depolarization noise, defined as $\tilde{\rho} = (1-p)\rho + pI/d$. Interestingly, we observe a decrease in variance as the error increases in Fig. 1(b). It shows the robustness of the thrifty shadow against noises.

2.2 Nonstabilizerness from the unitary ensemble

The power of the nonstabilizerness of the unitary ensemble is shown by the non-Clifford interleaved Clifford circuit [2, 6, 7, 8]. Ref. [2] exhibited that this circuits will enhance the thrifty shadow. Inspired by Ref. [6], the authors used the interleaved circuits

with only one non-Clifford gate between two Clifford layers. Despite the convenience of the mathematical analysis, this approach poses practical challenges due to the multiple use of Clifford gates. Our work proposes a novel circuit model that employs multiple non-Clifford gates within each interleaved layer, shown in Fig. 2. Here we denote the $\pi/8$ gate by M and simplify the notation $I^{\otimes n-1} \otimes M$ as $\hat{M}_1$, indicating the action of the M gate on the nth qubit. Generally, we can denote $I^{\otimes n-k} \otimes M^{\otimes k}$ by $\hat{M}_k$. We discover that the effectiveness of this model in thrifty shadow depends primarily on the total number of non-Clifford gates, rather than their specific locations. We present two simple case studies to illustrate our primary findings.

Theorem 2 *Suppose $\mathcal{U}$ is ensemble of one layer $\hat{M}_k$-interleaved Clifford circuits, which means $U \in \mathcal{U}$ satisfies $U = U_1 \hat{M}_k U_2$ where U_1, U_2 are randomly chosen in the Clifford group and $\hat{M}_k = I^{\otimes n-k} \otimes M^{\otimes k}$. $V_*(O, \rho)$ is defined in Eq. (1) where O is a traceless observable. Then*

$$V_*(O, \rho) < \left[2 \left(\frac{3}{4} \right)^k + \frac{4}{d} \right] \mathrm{tr}(O^2). \tag{4}$$

Proposition 3 *Suppose $\mathcal{U}$ is ensemble of l-layer $\hat{M}_1$-interleaved Clifford circuits, which means $U \in \mathcal{U}$ satisfies $U = U_1 \hat{M}_1 U_2 \dots \hat{M}_1 U_{l+1}$ where $U_1, \dots, U_{l+1}$ are randomly chosen in the Clifford group and $\hat{M}_1 = I^{\otimes n-1} \otimes M$. $V_*(O, \rho)$ is defined in Eq. (1) where O is a traceless observable. Then*

$$V_*(O, \rho) < \left[2 \left(\frac{3}{4} \right)^l + \frac{8}{d} \right] \mathrm{tr}(O^2). \tag{5}$$

These circuits, despite their differing structures, perform comparably in leading-order terms. However, the depth required for one layer $\hat{M}_k$-interleaved Clifford circuits is considerably less than the alternative. This disparity provides a method to substantially reduce the number of necessary Clifford gates, which simplifies the experimental procedures and lessens the computational demands of classical simulations.

We further argue that the effectiveness of the thrifty shadow with respect to the ensemble of l-layer $\hat{M}_k$-interleaved Clifford circuits mainly depends on the number of the non-Clifford gates:

$$V_*(O, \rho) = \left[2 \left(\frac{3}{4} \right)^{kl} + \mathcal{O}(d^{-1}) \right] \mathrm{tr}(O^2). \tag{6}$$

Given that this count naturally measures the nonstabilizerness of the unitary ensemble, such as the T-count [5, 9, 10], it substantiates the idea that the nonstabilizerness of the unitary ensemble can significantly improve thrifty shadow performance.

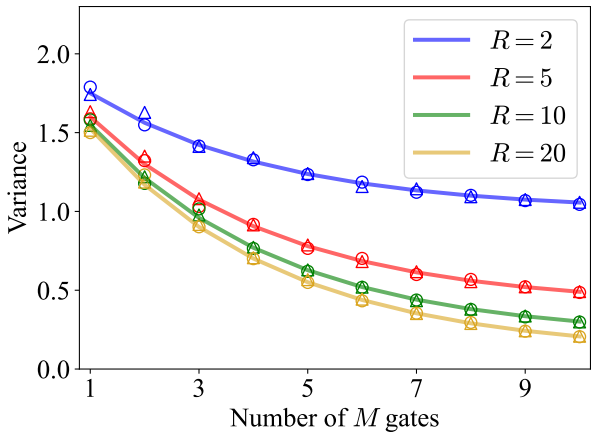


Figure 3: The statistical variance of the thrifty shadow to estimate the fidelity of a 50-qubit GHZ state. The unitary ensembles are chosen as the one layer $\hat{M}_k$-interleaved Clifford circuits and the l-layer $\hat{M}_1$-interleaved Clifford circuits, depicted with circled and triangular dots, respectively.

The analytical results can be found in our paper, and we use the numerical results to show our findings. When the dimension of the quantum system is large enough, the difference in variance between these two unitary ensembles becomes negligible. Our results also indicate that increasing the number of times a unitary is reused leads to a reduction in variance. These findings are corroborated by the numerical simulations presented in Fig. 3.

2.3 Technical contribution

Additionally, our main technical contribution is the investigation of the mathematical structure underlying thrifty shadow. To our knowledge, it is the first time such a structure has appeared. Except for its intrinsic properties, we investigate the connection between the fourth moment.

By virtue of our thorough understanding of the structure, the general upper bounds of the variances of the thrifty shadows with respect to various unitary ensembles are tighter than the results in the preceding work. Also, the fidelity estimation task can be analyzed completely.

At last, the mathematical structure underlying the thrifty shadow is brand new. It deserves further study to find other applications for the mathematical structure.

References

[1] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Phys.*, 16(10):1050–1057, June 2020.

[2] Jonas Helsen and Michael Walter. Thrifty shadow estimation: Reusing quantum circuits and bounding tails. *Phys. Rev. Lett.*, 131:240602, Dec 2023.

[3] You Zhou and Qing Liu. Performance analysis of multi-shot shadow estimation. *Quantum*, 7:1044, June 2023.

[4] Alireza Seif, Ze-Pei Cian, Sisi Zhou, Senrui Chen, and Liang Jiang. Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors. *PRX Quantum*, 4:010303, Jan 2023.

[5] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. Stabilizer Rényi Entropy. *Phys. Rev. Lett.*, 128:050402, Feb 2022.

[6] Jonas Haferkamp, Felipe Montealegre-Mora, Markus Heinrich, Jens Eisert, David Gross, and Ingo Roth. Efficient unitary designs with a system-size independent number of non-Clifford gates. *Commun. Math. Phys.*, 397(3):995–1041, 2023.

[7] Jonas Haferkamp. Random quantum circuits are approximate unitary t-designs in depth $O\left(nt^{5+o(1)}\right)$. *Quantum*, 6:795, September 2022.

[8] Lorenzo Leone, Salvatore F. E. Oliviero, You Zhou, and Alioscia Hamma. Quantum Chaos is Quantum. *Quantum*, 5:453, May 2021.

[9] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Phys. Rev. Lett.*, 116:250501, Jun 2016.

[10] Sergey Bravyi, Graeme Smith, and John A. Smolin. Trading classical and quantum computational resources. *Phys. Rev. X*, 6:021043, Jun 2016.

Complete version: Nonstabilizerness enhances the thrifty shadow estimation

Datong Chen[1 2 3 *] Huangjun Zhu[1 2 3 †]

[1] *State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai 200433, China.*
[2] *Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China.*
[3] *Center for Field Theory and Particle Physics, Fudan University, Shanghai 200433, China.*

Contents

1 Introduction

As a significant advancement in quantum state learning, the classical shadow [1] has attracted wide attention. In contrast to the inefficiencies of traditional quantum tomography for large quantum systems [2, 3, 4], classical shadow offers a general framework for efficiently estimating quantum systems through randomized measurements [5]. This field has seen numerous theoretical studies [6, 7, 8, 9, 10] and experimental breakthroughs [11, 12, 13, 14], highlighting its increasing relevance and application. The fundamental step in this protocol involves randomly applying a unitary transformation to the unknown state and performing a measurement in each round, serving as the foundation for efficient information extraction from the unknown state. However, making multiple changes to the unitary requires resetting the experimental configuration, leading to significant resource consumption in practice. To address this challenge, recent research has proposed a protocol with reduced resource consumption, known as thrifty shadow estimation [10, 15, 16]. The primary modification of this protocol is straightforward: we apply several independent measurements for each unitary. Given that repeating the same unitary transformation is less resource-consuming, this protocol is anticipated to be more feasible compared to the original approach.

In contrast to the original classical shadow, the thrifty shadow with Clifford measurements proves ineffective in certain scenarios. As demonstrated in Ref. [10], for a thrifty shadow to universally outperform the classical shadow, its unitary ensemble must form a unitary 4-design. However, the lack of an exact construction scheme for unitary 4-designs makes implementing an efficient thrifty shadow protocol highly challenging.

In this work, we explore how nonstabilizerness enhances the performance of thrifty shadows from two aspects: nonstabilizerness of the state and that of the unitary ensemble. The former is confirmed through quantum fidelity estimation tasks. We demonstrate that the performance of the thrifty shadow with respect to the Clifford group hinges on the stabilizer 2-Rényi entropy of the target state [17], which characterizes its nonstabilizerness. Surprisingly, we find that the Clifford group is suitable for certifying most states.

The power of the nonstabilizerness of the unitary ensemble is shown by the non-Clifford interleaved Clifford circuit [10, 18, 19, 20]. Ref. [10] exhibited that this circuits will enhance the thrifty shadow. Inspired by Ref. [18], the authors used the interleaved circuits with only one non-Clifford gate between two Clifford layers. Despite the convenience of the mathematical analysis, this approach poses practical challenges due to the multiple use of Clifford gates. Our work proposes a novel circuit model that employs multiple non-Clifford gates within each interleaved layer. We discover that the effectiveness of this model in thrifty shadow depends primarily on the total number of non-Clifford gates, rather than their specific locations. Thus, it can serve as a unitary

*22110190002@m.fudan.edu.cn
†zhuhuangjun@fudan.edu.cn

ensemble combining the experimental feasibility and the simulation efficiency for the thrifty shadow.

2 Setup

The standard protocol for the classical shadow of a state ρ is constructed by a unitary ensemble $\mathcal{U}$ and a measurement basis, typically the standard computational basis $|\boldsymbol{b}\rangle$ where $\boldsymbol{b} \in \{0,1\}^n$. Initially, a random unitary transformation from $\mathcal{U}$ is applied to ρ. Then an outcome $\boldsymbol{b}$ is obtained from the computational-basis measurement, yielding a record $U^\dagger|\boldsymbol{b}\rangle\langle\boldsymbol{b}|U$. The average over the records can be viewed as a post-processing channel:

$$\mathcal{M}(\rho) = \sum_{\boldsymbol{b}} \mathbb{E}_{U\sim\mathcal{U}} \langle\boldsymbol{b}|U\rho U^\dagger|\boldsymbol{b}\rangle U^\dagger|\boldsymbol{b}\rangle\langle\boldsymbol{b}|U. \tag{1}$$

The shadow channel $\mathcal{M}$ depends on the unitary ensemble. Moreover, if the measurement ensemble $\{U^\dagger|\boldsymbol{b}\rangle\}_{U\sim\mathcal{U},\boldsymbol{b}}$ is informationally complete, the shadow channel is invertible, and its inverse is termed the reconstruction map. Thus, for a state ρ, the snapshot corresponding to outcome $\boldsymbol{b}$ and unitary U is written as $\hat{\rho} = \mathcal{M}^{-1}(U^\dagger|\boldsymbol{b}\rangle\langle\boldsymbol{b}|U)$. One can efficiently extract information from the state through this snapshot. Previous studies have demonstrated that the number of required measurements to predict a certain property of the target state within a given error remains independent of the system's dimension, which overcomes the most serious problem of traditional quantum tomography. In particular, for the global Clifford group, the variance in estimating a traceless observable O is bounded by $3\mathrm{tr}(O^2)$

Compared to the original classical shadow, each unitary in the thrifty shadow protocol is reused R times. As a consequence, the snapshot of the state in this protocol is $\hat{\rho}_R = \sum_{i=1}^R \mathcal{M}^{-1}(U^\dagger|\boldsymbol{b}_i\rangle\langle\boldsymbol{b}_i|U)/R$. The variance of estimating an observable O with respect to the estimator $\mathrm{tr}(O\hat{\rho}_R)$ is now given by:

$$V_R(O,\rho) = \frac{1}{R}V(O,\rho) + \frac{R-1}{R}V_*(O,\rho), \tag{2}$$

where $V(O,\rho)$ is the variance using the original classical shadow. It's important to note that the sample complexity is proportional to $RV_R(O,\rho)$, which is greater than $V(O,\rho)$. Therefore, $V_*(O,\rho)$ characterizes the additional sample complexity introduced by thrifty shadow estimation. On the other hand, the resource cost of reusing a circuit is considerably lower than that of introducing a new circuit. If $V_*(O,\rho)$ is negligible, achieving the same accuracy with fewer resources becomes feasible.

Due to Ref. [15], the variance of the thrifty shadow estimation is only concerned with the traceless part of the observable. Hence, we assume that the observable O is traceless. Under this assumption, $V_*(O,\rho)$ defined in Eq. (2) with respect to some unitary ensemble $\mathcal{U}$ can be specifically formulated as:

$$\begin{aligned} V_*(O,\rho) &= (d+1)^2\mathrm{tr}\left[\mathbb{E}_{U\sim\mathcal{U}}\sum_{i,j} U^{\dagger\otimes 4}|\boldsymbol{b}_i\rangle\langle\boldsymbol{b}_i|^{\otimes 2}\otimes|\boldsymbol{b}_j\rangle\langle\boldsymbol{b}_j|^{\otimes 2}U^{\otimes 4}O\otimes\rho\otimes O\otimes\rho\right] - \mathrm{tr}(O\rho)^2. \\ &= (d+1)^2\mathrm{tr}\left[\Omega_{\mathcal{U}}(\{|\boldsymbol{b}_i\rangle\}_i)O\otimes\rho\otimes O\otimes\rho\right] - \mathrm{tr}(O\rho)^2. \end{aligned} \tag{3}$$

Since we focus on the qubit case, we denote 2^n by d for simplicity, where n is the number of the qubits. The performance of the thrifty shadow with respect to a unitary ensemble $\mathcal{U}$ completely depends on this operator $\Omega_{\mathcal{U}}(\{|\boldsymbol{b}_i\rangle\}_i)$. However, its algebraic structure remains unclear. To gain a deeper understanding of the thrifty shadow estimation, we investigate this operator analytically for various unitary ensembles. The discussion of this operator included an extension to a general measurement basis $\{|\psi_i\rangle\}_i$ can be found in Appendix B. Thereafter, we will use the properties to analyze the performance of the thrifty shadow and show the relationship between the nonstabilizerness.

3 The nonstabilizerness from the states

Ref. [10] claims that if U is drawn uniformly from Haar measure or any unitary 4-design, $V_*(O,\rho)$ is of order $\mathcal{O}(d^{-1}\mathrm{tr}(O^2))$, which is much less than $V(O,\rho)$. Here we give a more accurate proposition.

Proposition 1 *Suppose $\mathcal{U}$ is a Haar random ensemble or a unitary 4-design and $V_*(O,\rho)$ is defined in Eq. (3) where O is a traceless observable. Then*

$$V_*(O,\rho) \le \frac{4d^2+28d+26}{d(d+2)(d+3)}\mathrm{tr}(O^2) = \mathcal{O}(d^{-1}\mathrm{tr}(O^2)). \tag{4}$$

This proposition shows that in consideration of a large dimensional quantum system, the thrifty shadow with respect to the unitary 4-design can decrease the number of the required circuits but only introduce little impact on the accuracy for any observable.

Then, we focus on the case when $\mathcal{U}$ is the Clifford group. We get a general upper bound of the variance with respect to the Clifford group, and this bound is tight in the leading order.

Theorem 2 *Suppose $\mathcal{U}$ is the Clifford group and $V_*(O,\rho)$ is defined in Eq. (3) where O is a traceless observable. Then*

$$V_*(O,\rho) \le \frac{2(d+1)}{(d+2)}\mathrm{tr}(O^2) < 2\mathrm{tr}(O^2). \tag{5}$$

This implies that there is a significant gap between the performances of these two ensembles in some cases. If $V_*(O,\rho)$ is of order $\mathrm{tr}(O^2)$, the thrifty shadow using the

Clifford group performs even worse than the original classical shadow in the sense that it introduces extra unnecessary experimental costs. Refs. [10, 15] have shown that in the task of stabilizer state fidelity estimation, the variance of the thrifty shadow with respect to the Clifford group approximately achieves the upper bound in Theorem 2. Here we extend our investigation to the fidelity estimation of general states. We find that the Clifford group performs well in certifying states with high non-stabilizerness.

Theorem 3 *Suppose $\rho = |\phi\rangle\langle\phi|$ is any n-qubit state and the observable $O = |\phi\rangle\langle\phi| - I/d$, then the variance of the thrifty shadow using the Haar random ensemble or a unitary 4-design reads*

$$V_R(O,\rho) = \frac{1}{R}\frac{2(d-1)}{d+2} + \frac{R-1}{R}\frac{4(d-1)}{(d+2)(d+3)}. \quad (6)$$

Theorem 4 *Suppose $\rho = |\phi\rangle\langle\phi|$ is any n-qubit state and the observable $O = |\phi\rangle\langle\phi| - I/d$, then the variance of the thrifty shadow using the Clifford group reads*

$$V_R(O,\rho) = \frac{1}{R}\frac{2(d-1)}{d+2} + \frac{R-1}{R}\frac{2(d+1)2^{-M_2(|\phi\rangle)} - 4}{d+2}, \quad (7)$$

where $M_2(|\phi\rangle)$ is the stabilizer 2-Rényi entropy of state $|\phi\rangle$ defined by [17]

$$M_2(|\phi\rangle) = -\log_2 \frac{1}{d}\sum_{P\in\mathcal{P}_n} \langle\phi|P|\phi\rangle^4. \quad (8)$$

This theorem provides a practical criterion for selecting the thrifty shadow with Clifford measurements. Compared to Theorem 3, the performance of the thrifty shadow with respect to the Clifford group is dependent on the stabilizer 2-Rényi entropy of the target state. In the worst case that the target state is a stabilizer state, the variance $V_R(O,\rho)$ is equivalent to $V(O,\rho)$, indicating that reusing the same circuits does not obtain extra information. However, the thrifty shadow with respect to the Clifford group begins to show benefits when $M_2(|\phi\rangle)$ is large. Particularly, if $M_2(|\phi\rangle) > \log_2(d+3)/4$, using the Clifford group becomes even more efficient than the unitary 4-design. Given that most states are expected to have a high stabilizer 2-Rényi entropy [17, 21], the Clifford group is generally suitable for certifying states, also shown by the random states results in Fig. 1(e).

We verify our conclusions with several numerical experiments [22, 23, 24, 25] in Fig. 1, including W states [26], Greenberger–Horne–Zeilinger (GHZ) states, a family of states constructed by the non-Clifford gates, and random states. For W states, the stabilizer 2-Rényi entropy is of order $\log_2 n^2$, seeing in Appendix E, indicating that the thrifty shadow with Clifford measurements also benefits the fidelity estimation. For GHZ states, we consider GHZ states with a phase factor:

$$|GHZ_\theta\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle^{\otimes n} + e^{i\theta}|1\rangle^{\otimes n}\right). \quad (9)$$

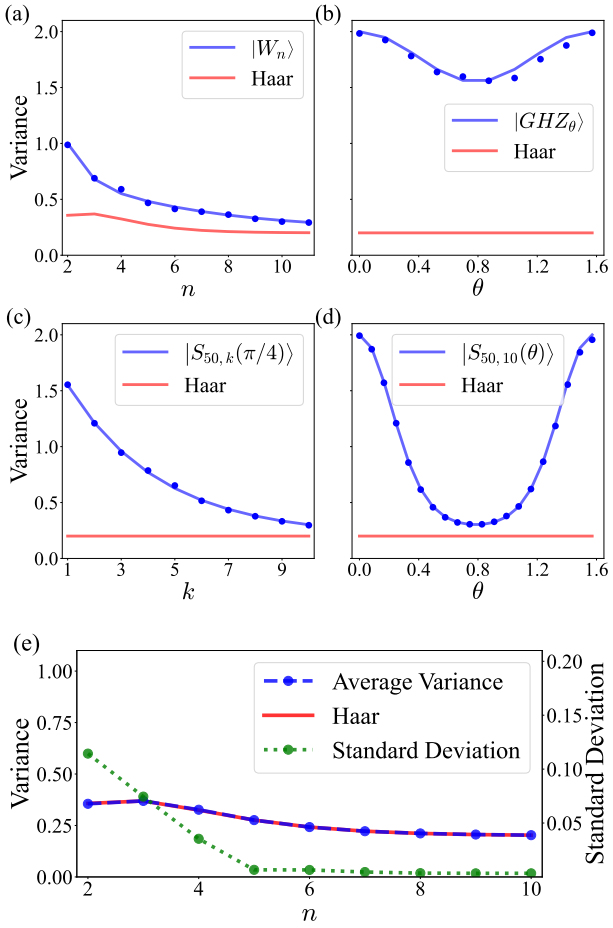


Figure 1: The statistical variance of the thrifty shadow to certify various families of states with respect to Clifford group. In (a)-(d), the lines represent the analytical results in Theorems 3 and 4 for different state families. The circled dots are the numerical results with 100,000 random samplings from the Clifford group and 10 repeated times. (a) Estimation of the fidelity of W states, parameterized by the qubits number n. (b) Estimation of the fidelity of 50-qubit GHZ states, parameterized by the phase factor θ shown in Eq. (9). (c) Estimation of the fidelity of $|S_{50,k}(\pi/4)\rangle$ defined in Eq. (10), parameterized by the number of magic states k. (d) Estimation of the fidelity of $|S_{50,10}(\theta)\rangle$, parameterized by the phase factor θ. In (e), the blue dashed line represents the average variance of estimating 1000 random states from the Haar measure. The green dotted line indicates the standard deviation of the variances.

Unfortunately, the advantages of the thrifty shadow are inconspicuous for any θ. At last, we construct a state family with high nonstabilizerness [27]:

$$|S_{n,k}(\theta)\rangle = |0\rangle^{\otimes n-k} \otimes \left[\frac{1}{\sqrt{2}}\left(|0\rangle + e^{i\theta}|1\rangle\right)\right]^{\otimes k}. \quad (10)$$

This can also be viewed as applying the gates $I^{\otimes n-k} \otimes (HP)^{\otimes k}$ on $|0\rangle^{\otimes n}$, where H is the Hadamard gate and P is a general single-phase gates. As k increases, the variance will decrease exponentially. The exact stabilizer 2-Rényi entropies of these states can be found in Appendix E. Since the prepared states in realistic experiments may not be ideal, we also examine states affected by depolarization noise, defined as $\tilde{\rho} = (1-p)\rho + pI/d$. Interestingly, we observe a decrease in variance as the error increases, as shown in Fig. 2. It shows the robustness of the thrifty shadow against noises. Detailed results are available in Appendix D.

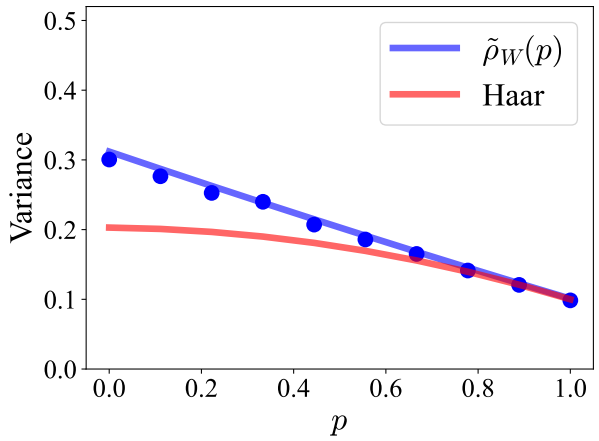


Figure 2: The statistical variance of the thrifty shadow to certify a W state with depolarized noise. Here $\tilde{\rho} = (1-p)\rho + pI/d$ and the ideal state is chosen as the 10-qubit W state. The lines are the analytical results and the circled dots represent the numerical results with 20,000 random samplings from the Clifford group and 10 repeated times.

4 The nonstabilizerness from unitary ensemble

When the Clifford group fails to benefit the thrifty shadow, such as certifying the states close to stabilizer states, we can use an approximate 4-design to improve the performance of the thrifty shadow. Considering both the practical costs and the complexity of classical simulation, the non-Clifford gates interleaved Clifford circuits are recommended. In contrast to the preceding model [18, 19, 20, 10], our work presents a modified circuit model. This new circuit architecture diverges from prior models by incorporating multiple non-Clifford gates within each layer, which is shown in Fig. 3.

We assume the non-Clifford gates as the $\pi/8$ gates since this gate has been identified as the most effective single-qubit phase gate, seeing in Appendix B. Denote

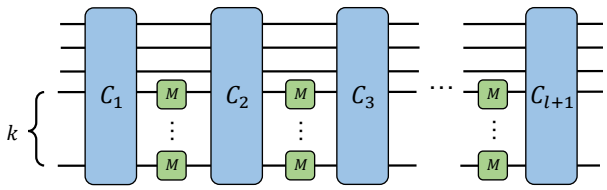


Figure 3: The circuits model of l-layer $\hat{M}_k$-interleaved Clifford circuits. C_1, C_2, ...,C_{l+1} are randomly selected in the Clifford group. In this configuration, the M gates can be replaced by any single-phase gate.

the $\pi/8$ gate by M and simplify the notation $I^{\otimes n-1} \otimes M$ as $\hat{M}_1$, indicating the action of the M gate on the nth qubit. Generally, we can denote $I^{\otimes n-k} \otimes M^{\otimes k}$ by $\hat{M}_k$. Next, we present two simple case studies to illustrate our primary findings.

Theorem 5 *Suppose $\mathcal{U}$ is ensemble of one layer $\hat{M}_k$-interleaved Clifford circuits, which means $U \in \mathcal{U}$ satisfies $U = U_1 \hat{M}_k U_2$ where U_1, U_2 are randomly chosen in the Clifford group and $\hat{M}_k = I^{\otimes n-k} \otimes M^{\otimes k}$. $V_*(O,\rho)$ is defined in Eq. (3) where O is a traceless observable. Then*

$$V_*(O,\rho) < \left[2\left(\frac{3}{4}\right)^k + \frac{4}{d}\right] \mathrm{tr}(O^2). \quad (11)$$

Proposition 6 *Suppose $\mathcal{U}$ is ensemble of l-layer $\hat{M}_1$-interleaved Clifford circuits, which means $U \in \mathcal{U}$ satisfies $U = U_1 \hat{M}_1 U_2 \ldots \hat{M}_1 U_{l+1}$ where $U_1, \ldots, U_{l+1}$ are randomly chosen in the Clifford group and $\hat{M}_1 = I^{\otimes n-1} \otimes M$. $V_*(O,\rho)$ is defined in Eq. (3) where O is a traceless observable. Then*

$$V_*(O,\rho) < \left[2\left(\frac{3}{4}\right)^l + \frac{8}{d}\right] \mathrm{tr}(O^2). \quad (12)$$

These circuits, despite their differing structures, perform comparably in leading-order terms. However, the depth required for one layer $\hat{M}_k$-interleaved Clifford circuits is considerably less than the alternative. This disparity provides a method to substantially reduce the number of necessary Clifford gates. Since any Clifford gates can be synthesized by $\mathcal{O}(n^2/\log n)$ gates from the gate set of Hadamard gate, Phase gate, and the CNOT gate [22], using one layer $\hat{M}_k$-interleaved Clifford circuits allows for a reduction of up to $\mathcal{O}((k-1)(n^2/\log n))$ elementary gates. We further argue that the effectiveness of the thrifty shadow with respect to the ensemble of l-layer $\hat{M}_k$-interleaved Clifford circuits mainly depends on the number of the non-Clifford gates, proved in Appendix B:

$$V_*(O,\rho) = \left[2\left(\frac{3}{4}\right)^{kl} + \mathcal{O}(d^{-1})\right] \mathrm{tr}(O^2). \quad (13)$$

Given that this count naturally measures the nonstabilizerness of the unitary ensemble, such as the T-count [17, 28, 29], it substantiates the idea that the nonstabilizerness of the unitary ensemble can significantly improve thrifty shadow performance.

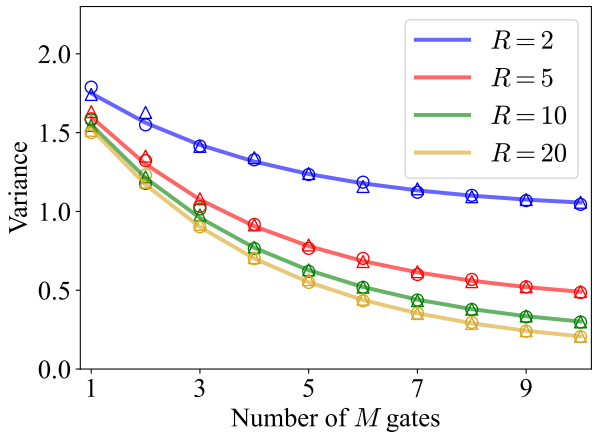


Figure 4: The statistical variance of the thrifty shadow to estimate the fidelity of a 50-qubit GHZ state. The unitary ensembles are chosen as the one layer $\hat{M}_k$-interleaved Clifford circuits and the l-layer $\hat{M}_1$-interleaved Clifford circuits, depicted with circled and triangular dots, respectively. For each ensemble, we perform $100,000$ times random sampling and R repeated times. Lines represent analytical results for different repeated times R.

Additionally, we note that the upper bounds introduced in Theorem 5 and Proposition 6 are not tight. Therefore, one cannot assert that the performance of l-layer $\hat{M}_1$ interleaved Clifford circuits is worse than the one-layer $\hat{M}_k$ interleaved Clifford circuits when the total number of M gates is the same. In the subsequent example, we find that the variance with respect to the l-layer $\hat{M}_1$ interleaved Clifford circuits is even smaller.

Theorem 7 *Suppose $\rho = |S\rangle\langle S|$ is a stabilizer state and the observable $O = |S\rangle\langle S| - I/d$, then the variance of the thrifty shadow using the ensemble of one layer $\hat{M}_k$-interleaved Clifford circuits reads*

$$V_R(O,\rho) = \frac{1}{R}\frac{2(d-1)}{d+2} + \frac{R-1}{R}\left[2\left(\frac{3}{4}\right)^k + \mathcal{O}(d^{-1})\right]. \tag{14}$$

Theorem 8 *Suppose $\rho = |S\rangle\langle S|$ is a stabilizer state and the observable $O = |S\rangle\langle S| - I/d$, then the variance of the thrifty shadow using the ensemble of l-layer $\hat{M}_1$-interleaved Clifford circuits reads*

$$V_R(O,\rho) = \frac{1}{R}\frac{2(d-1)}{d+2} + \frac{R-1}{R}\left[2\left(\frac{3}{4}\right)^l + \mathcal{O}(d^{-1})\right]. \tag{15}$$

It can be verified that the variance of the l-layer $\hat{M}_1$ interleaved Clifford circuits is smaller than the other configuration by the exact results, seeing in Appendix D. When the dimension of the quantum system is large enough, the difference in variance between these two unitary ensembles becomes negligible. Our results also indicate that increasing the number of times a unitary is reused leads to a reduction in variance. These findings are corroborated by the numerical simulations presented in Fig. 4.

5 Conclusions

We have presented a complete description of the thrifty shadow. Generalizing to the preceding work, our findings demonstrate that thrifty shadow estimation with the Clifford measurements can offer significant advantages over the original classical shadow in some cases. This is particularly evident in fidelity estimation tasks where the performance of thrifty shadow is enhanced by the nonstabilizerness of the target states. For almost all states, selecting the Clifford group for thrifty estimation is advantageous. Conversely, for states close to stabilizer states, interleaved Clifford circuits effectively address the limitations of the Clifford group. Moreover, our study reveals that interleaved Clifford circuits, enhanced by incorporating multiple non-Clifford gates within each layer, can reduce the need for numerous Clifford layers. This adaptation simplifies the experimental procedures and lessens the computational demands of classical simulations. We believe that it may play a role in wide applications.

Our work also analytically discusses the mathematical structure underlying thrifty shadow. To our knowledge, it is the first time such a structure has appeared. Except for its intrinsic properties, we investigate the connection between the fourth moment. It deserves further study to find other applications for the mathematical structure.

References

[1] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Phys.*, 16(10):1050–1057, June 2020.

[2] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017.

[3] Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.*, 14(9):095022, sep 2012.

[4] Martin Kliesch and Ingo Roth. Theory of quantum system certification. *PRX Quantum*, 2:010201, Jan 2021.

[5] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, et al. The randomized measurement toolbox. *Nat. Rev. Phys.*, 5:9–24, 2023.

[6] Hong-Ye Hu, Soonwon Choi, and Yi-Zhuang You. Classical shadow tomography with locally scrambled quantum dynamics. *Phys. Rev. Res.*, 5:023027, Apr 2023.

[7] H. Chau Nguyen, Jan Lennart Bönsel, Jonathan Steinberg, and Otfried Gühne. Optimizing shadow tomography with generalized measurements. *Phys. Rev. Lett.*, 129:220502, Nov 2022.

[8] Jonas Helsen, Marios Ioannou, Jasper Kitzinger, et al. Shadow estimation of gate-set properties from random sequences. *Nat. Commun.*, 14:5039, 2023.

[9] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. *PRX Quantum*, 2:030348, Sep 2021.

[10] Jonas Helsen and Michael Walter. Thrifty shadow estimation: Reusing quantum circuits and bounding tails. *Phys. Rev. Lett.*, 131:240602, Dec 2023.

[11] G.I. Struchalin, Ya. A. Zagorovskii, E.V. Kovlakov, S.S. Straupe, and S.P. Kulik. Experimental estimation of quantum state properties from classical shadows. *PRX Quantum*, 2:010307, Jan 2021.

[12] Ting Zhang, Jinzhao Sun, Xiao-Xu Fang, Xiao-Ming Zhang, Xiao Yuan, and He Lu. Experimental quantum state measurement with classical shadows. *Phys. Rev. Lett.*, 127:200501, Nov 2021.

[13] Roman Stricker, Michael Meth, Lukas Postler, Claire Edmunds, Chris Ferrie, Rainer Blatt, Philipp Schindler, Thomas Monz, Richard Kueng, and Martin Ringbauer. Experimental single-setting quantum state tomography. *PRX Quantum*, 3:040310, Oct 2022.

[14] William J. Huggins, Bryan A. O'Gorman, Nicholas C. Rubin, et al. Unbiasing fermionic quantum monte carlo with a quantum computer. *Nature*, 603:416–420, 2022.

[15] You Zhou and Qing Liu. Performance analysis of multi-shot shadow estimation. *Quantum*, 7:1044, June 2023.

[16] Alireza Seif, Ze-Pei Cian, Sisi Zhou, Senrui Chen, and Liang Jiang. Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors. *PRX Quantum*, 4:010303, Jan 2023.

[17] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. Stabilizer Rényi Entropy. *Phys. Rev. Lett.*, 128:050402, Feb 2022.

[18] Jonas Haferkamp, Felipe Montealegre-Mora, Markus Heinrich, Jens Eisert, David Gross, and Ingo Roth. Efficient unitary designs with a system-size independent number of non-Clifford gates. *Commun. Math. Phys.*, 397(3):995–1041, 2023.

[19] Jonas Haferkamp. Random quantum circuits are approximate unitary t-designs in depth $O\left(nt^{5+o(1)}\right)$. *Quantum*, 6:795, September 2022.

[20] Lorenzo Leone, Salvatore F. E. Oliviero, You Zhou, and Alioscia Hamma. Quantum Chaos is Quantum. *Quantum*, 5:453, May 2021.

[21] Andi Gu, Lorenzo Leone, Soumik Ghosh, Jens Eisert, Susanne Yelin, and Yihui Quek. A little magic means a lot. *arXiv preprint arXiv:2308.16228*, 2023.

[22] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. A*, 70:052328, Nov 2004.

[23] Robert Koenig and John A. Smolin. How to efficiently select an arbitrary Clifford group element. *J. Math. Phys.*, 55(12):122202, 12 2014.

[24] Hakop Pashayan, Oliver Reardon-Smith, Kamil Korzekwa, and Stephen D. Bartlett. Fast estimation of outcome probabilities for quantum circuits. *PRX Quantum*, 3:020361, Jun 2022.

[25] Sergey Bravyi and Dmitri Maslov. Hadamard-free circuits expose the structure of the clifford group. *IEEE Trans. Inf. Theory*, 67(7):4546–4563, 2021.

[26] H. Häffner, W. Hänsel, C. Roos, et al. Scalable multiparticle entanglement of trapped ions. *Nature*, 438:643–646, 2005.

[27] Sergey Bravyi and Alexei Kitaev. Universal quantum computation with ideal clifford gates and noisy ancillas. *Phys. Rev. A*, 71:022316, Feb 2005.

[28] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by clifford gates. *Phys. Rev. Lett.*, 116:250501, Jun 2016.

[29] Sergey Bravyi, Graeme Smith, and John A. Smolin. Trading classical and quantum computational resources. *Phys. Rev. X*, 6:021043, Jun 2016.

[30] Goodman Roe and Nolan R. Wallach. *Symmetry, Representations, and Invariants*, volume 255. Springer New York, 2009.

[31] Pavel Etingof, Oleg Golberg, Sebastian Hensel, et al. Introduction to representation theory. *arXiv preprint arXiv:0901.0827*, 2009.

[32] David Gross, Sepehr Nezami, and Michael Walter. Schur–Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations. *Commun. Math. Phys.*, 385(3):1325–1393, 2021.

[33] A. R. Calderbank and Peter W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54:1098–1105, Aug 1996.

[34] Andrew Steane. Multiple-particle interference and quantum error correction. *Proc. R. Soc. Lond. A.*, 452(1954):2551–2577, 1996.

[35] Huangjun Zhu. Multiqubit Clifford groups are unitary 3-designs. *Phys. Rev. A*, 96:062336, Dec 2017.

[36] Huangjun Zhu, Richard Kueng, Markus Grassl, and David Gross. The Clifford group fails gracefully to be a unitary 4-design. *arXiv preprint arXiv:1609.08172*, 2016.

A Commutant of the Clifford tensor powers

We first introduce the Schur-Weyl duality of the Clifford tensor powers as the main mathematical basis of this paper. The most established Schur-Weyl duality applicable to the unitary group states that the commutant of the unitary tensor powers is generated by the symmetric group [30, 31]. Since the Clifford group is a subgroup of the unitary group, the commutant of Clifford tensor powers $\mathrm{Cl}^{\otimes t}$ should be larger. In the following context, we focus on the qubit case and assume the Hilbert space to be $\mathcal{H}_2^{\otimes n}$. The central concept for analyzing the commutant of $\mathrm{Cl}^{\otimes t}$ is the notion of stochastic Lagrangian subspace [32]. A subspace $T \leq \mathbb{Z}_2^{2t}$ is a *stochastic Lagrangian subspace* if it satisfies the following three conditions:

1. $\boldsymbol{x} \cdot \boldsymbol{x} = \boldsymbol{y} \cdot \boldsymbol{y} \mod 4$ for all $(\boldsymbol{x}, \boldsymbol{y}) \in T$.
2. T has dimension t.
3. $\mathbf{1}_{2t} = (1, 1, \ldots, 1) \in T$.

The first condition implies that T is totally isotropic with respect to the quadratic form $\mathfrak{q}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{x} - \boldsymbol{y} \cdot \boldsymbol{y} \mod 4$. Together with the second condition, such subspace is referred to as Lagrangian. The set comprising all such subspaces is denoted by $\Sigma_{t,t}$. A special subset of the $\Sigma_{t,t}$ can be determined by the stochastic orthogonal group O_t. A $t \times t$ matrix with entries in $\mathbb{Z}_2$ is called *stochastic orthogonal* if $O\boldsymbol{x} \cdot O\boldsymbol{x} = \boldsymbol{x} \cdot \boldsymbol{x} \mod 4$ for all $\boldsymbol{x} \in \mathbb{Z}_2^t$. This group can be directly embedded into $\Sigma_{t,t}$ by defining $T_O := \{(O\boldsymbol{x}, \boldsymbol{x}) |\ \boldsymbol{x} \in \mathbb{Z}_2^t\}$.

To further understand the structure of the stochastic Lagrangian subspace, we introduce two additional concepts. A *defect subspace* is a subspace $N \leq \mathbb{Z}_2^t$ satisfying $\boldsymbol{x} \cdot \boldsymbol{x} = 0 \mod 4$ for all $\boldsymbol{x} \in N$. Such subspaces are totally q-isotropic according to the quadratic form $q(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{x} \mod 4$. The orthogonal complement $N^\perp$ is taken with respect to the inner product modulo 2. Given two defect subspaces M, N, a linear map $J : N^\perp/N \to M^\perp/M$ is called a *defect isomorphism* if it satisfies:

1. $q\left(J[\boldsymbol{x}]\right) = q\left([\boldsymbol{x}]\right)$ for all $[\boldsymbol{x}] \in N^\perp/N$.
2. $J[\mathbf{1}_t] = [\mathbf{1}_t]$.

According to Ref. [32], any stochastic Lagrangian subspace is induced by two defect subspaces M, N and a defect isomorphism J.

For each stochastic Lagrangian subspace T, we define the corresponding operators $r(T)$ and $R(T)$ as follows:

$$r(T) := \sum_{(\boldsymbol{x},\boldsymbol{y}) \in T} |\boldsymbol{x}\rangle\langle\boldsymbol{y}|, \quad R(T) := r(T)^{\otimes n}. \tag{16}$$

We also denote $r(T_O)$ by $r(O)$. It has been proved that when $n \geq t - 1$, the set of $R(T)$ is linearly independent and spans the commutant of $\mathrm{Cl}(n)^{\otimes t}$. Moreover, $r(T)$ and $R(T)$ are closely related to the Calderbank-Shor-Steane (CSS) code [33, 34]. For a defect subspace N, we can define a stabilizer group:

$$\mathrm{CSS}(N) := \{Z_{\boldsymbol{p}} X_{\boldsymbol{q}} |\ \boldsymbol{p}, \boldsymbol{q} \in N\}, \tag{17}$$

which corresponds to a CSS code. The projector onto the corresponding code space is given by

$$P_N = \frac{1}{|N|^2} \sum_{\boldsymbol{p},\boldsymbol{q} \in N} Z_{\boldsymbol{p}} X_{\boldsymbol{q}}, \tag{18}$$

where $|N|$ is the cardinality of N. Based on the CSS code, given a stochastic Lagrangian subspace T, if the left and right defect subspaces of T coincide, denoted as N, and the defect isomorphism is trivial, $r(T)$ can be expressed as

$$r(T) = |N| P_N. \tag{19}$$

Furthermore, any $r(T)$ can be determined by a stochastic orthogonal matrix O and a defect subspace N [18].

Then, we consider the case $t = 4$ in detail. For the sake of simplicity, we denote 2^n by d in the following. By virtue of Ref. [32], for $t = 4$, we have

$$\Sigma_{4,4} = S_4 \cup S_4 \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{array}\right) S_4 = S_4 \cup S_4 T_4 S_4. \tag{20}$$

Moreover, the set $S_4 T_4 S_4$, which comprises six subspaces, exhibits a simpler form, as demonstrated the following lemma.

Lemma 9 *Suppose S_t is the permutation group and S_3 is the subset of S_4 generated by $\{(12),(123)\}$. Let T_4 be defined as in Eq. (20). Then*

$$S_4T_4S_4 = S_3T_4. \tag{21}$$

Proof. It is straightforward to verify that $N = \{|0000\rangle, |1111\rangle\}$ is the only nontrivial defect subspace in $\mathbb{Z}_2^4$. Thus, the stochastic Lagrangian subspace in the set $S_4T_4S_4$ is determined by the defect isomorphism J. Specifically, for $T \in S_4T_4S_4$,

$$T = \{(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}+\boldsymbol{w}) |\ [\boldsymbol{y}] \in N^\perp/N, J[\boldsymbol{y}] = [\boldsymbol{x}], \boldsymbol{z}, \boldsymbol{w} \in N\}. \tag{22}$$

By definition, the set of defect isomorphisms is the set of automorphisms of $N^\perp/N$. Given that $\dim N^\perp/N = 2$ and $J[\mathbf{1}_4] = [\mathbf{1}_4]$, the set of automorphisms is isomorphic to S_3. Actually, it is exactly the permutation group S_3. For instance, for $(12) \in S_3$, the action of (12) induces an automorphism of $N^\perp/N$, J, such that

$$J[(0000)] = [(0000)],\ J[(1010)] = [(0110)],\ J[(0110)] = [(1010)],\ J[(0011)] = [(0011)]. \tag{23}$$

The remaining automorphisms are induced by other elements in S_3, thus concluding the proof. □

As mentioned above, for $n \geq 3$, the commutant of $\mathrm{Cl}(n)^{\otimes 4}$ is spanned by the linearly independent set $\{R(T)\}_{T\in\Sigma_{4,4}}$. Actually, the set of $\{R(T)\}_{T\in\Sigma_{4,4}}$ spans the commutant of $\mathrm{Cl}(n)^{\otimes 4}$ for all n, indicating that the restriction on n can be relaxed.

Lemma 10 *Suppose $\Sigma_{4,4}$ is the set of stochastic Lagrangian subspace in $\mathbb{Z}_2^8$. Then $\{R(T) = r(T)^{\otimes n}\}_{T\in\Sigma_{4,4}}$ spans the commutant of $\mathrm{Cl}(n)^{\otimes 4}$.*

Proof. Let $\{T_i\}_{i=1}^{|\Sigma_{4,4}|}$ be an enumeration of $\Sigma_{4,4}$. Denote by Γ the Gram matrix of the set $\{R(T)\}_{T\in\Sigma_{4,4}}$, where the entries is defined as

$$\Gamma_{ij} := \mathrm{tr}\left[R(T_i)^\dagger R(T_j)\right]. \tag{24}$$

The rank of Γ is precisely the dimension of the span of $\{R(T)\}_{T\in\Sigma_{4,4}}$, i.e.,

$$\mathrm{rank}\ \Gamma = \dim \mathrm{span}\left(\{R(T)\}_{T\in\Sigma_{4,4}}\right). \tag{25}$$

Calculating the eigenvalue of Γ yields

$$d(d-1)(d-2)(d-4),\ d(d+1)(d+2)(d+4),\ d(d-1)(d+1)(d-2),\ d(d-1)(d+1)(d+2), \tag{26}$$

with multiplicities 1, 1, 14, and 14 respectively. Consequently,

$$\mathrm{rank}\ \Gamma = \begin{cases} 15, & n = 1, \\ 29, & n = 2, \\ 30, & n \geq 3. \end{cases} \tag{27}$$

This matches the dimension of the commutant of $\mathrm{Cl}(n)^{\otimes 4}$ [35], thereby finishing the proof. □

Considering the following requirements, we are interested in the projectors onto the representation spaces of the Clifford group. Based on the Schur-Weyl duality, the total space $(\mathbb{C}^d)^{\otimes t}$ can be decomposed into multiplicity-free irreducible representations of $U(d) \times S_t$:

$$(\mathbb{C}^d)^{\otimes t} = \bigoplus_\lambda W_\lambda \otimes S_\lambda, \tag{28}$$

where λ represents the non-increasing partitions of t into nor more than d parts, W_λ is the Weyl module carrying the irreducible representation of $U(d)$ associated with λ, and S_λ is the Specht module on which S_t acts irreducibly. Denote the projector onto $W_\lambda \otimes S_\lambda$ by P_λ. Obviously, it is the weighted summation of $R(T)$ over S_4, which has the form

$$P_\lambda = \frac{d_\lambda}{24} \sum_{T\in S_4} \chi_\lambda(T) R(T), \tag{29}$$

where d_λ is the multiplicity of the Weyl module W_λ, χ_λ is the character of the irrep of S_4 corresponding to the partition λ.

As we mentioned before, the operator $R(T_4)$ is proportional to a projector onto a CSS code space [36, 32]:

$$R(T_4) = dP_N^{(n)} = \frac{1}{d}\left(I^{\otimes 4} + X^{\otimes 4} + Y^{\otimes 4} + Z^{\otimes 4}\right)^{\otimes n} = \frac{1}{d} \sum_{P\in\mathcal{P}_n} P^{\otimes 4}, \tag{30}$$

where $N = \{|0000\rangle, |1111\rangle\}$ is the nontrivial defect subspace in $\mathbb{Z}_2^4$ and the subscript (n) represents the number of qubits. Hereafter, N specifically refers to this space. Since $R(T_4)$ or $P_N^{(n)}$ commutes with all permutations in S_4, $P_\lambda P_N^{(n)}$ is also a projector, denoted by $P_{N,\lambda}^{(n)}$, which is composed by $R(T)$ with $T \in S_4 T_3$:

$$P_{N,\lambda}^{(n)} = \frac{d_\lambda}{24d} \sum_{T \in S_4} \chi_\lambda(T) R(T) R(T_4). \tag{31}$$

The exact dimensions of each projector can be found in Ref. [36].

B Mathematical structure

According to Eq. (3), we would like to analyze the property of the operator $\Omega_{\mathcal{U}}$. For a general orthogonal basis $\{|\psi_i\rangle\}_i$, we can define $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ as

$$\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i) := \mathbb{E}_{U\sim\mathcal{U}} \sum_{i,j} U^{\dagger\otimes 4} |\psi_i\rangle\langle\psi_i|^{\otimes 2} \otimes |\psi_j\rangle\langle\psi_j|^{\otimes 2} U^{\otimes 4}. \tag{32}$$

Next, we consider the specific examples of this operator.

B.1 Haar uniform ensemble and Clifford group

We start by discussing the concrete form of Ω for some simple cases. If the unitary ensemble is a Haar random ensemble or a unitary 4-design, Ω_{Haar} is independent of the choice of the basis. According to the Schur-Weyl duality and Schur's lemma, we obtain the following proposition:

Proposition 11 *Suppose that $\mathcal{U}$ is a Haar random ensemble or a unitary 4-design, $\{|\psi_i\rangle\}_i$ is any set of orthogonal bases, and Ω_{Haar} is defined in Eq. (32). Then*

$$\Omega_{\text{Haar}} = \kappa\left(\text{Haar}, P_{[4]}\right) P_{[4]} + \kappa\left(\text{Haar}, SP_{[2,2]}\right) SP_{[2,2]} + \kappa\left(\text{Haar}, SP_{[3,1]}\right) SP_{[3,1]}, \tag{33}$$

where $SP_\lambda = P_\lambda + (12)P_\lambda + (34)P_\lambda + (12)(34)P_\lambda$ and $(12), (34), (12)(34)$ are the shorthand for the corresponding $R(T)$. The coefficients κ are

$$\begin{cases} \kappa\left(\text{Haar}, P_{[4]}\right) = \dfrac{4(d+5)}{(d+1)(d+2)(d+3)}, \\ \kappa\left(\text{Haar}, SP_{[2,2]}\right) = \dfrac{1}{d(d+1)}, \\ \kappa\left(\text{Haar}, SP_{[3,1]}\right) = \dfrac{1}{(d+1)(d+2)}. \end{cases} \tag{34}$$

Proof. By definition, the operator $\Omega_{\text{Haar}}(|\psi_i\rangle i)$ is expressed in two components:

$$\begin{aligned} \Omega_{\text{Haar}} &= \mathbb{E}_{U\sim\mathcal{U}} \sum_i (U^\dagger |\psi_i\rangle\langle\psi_i| U)^{\otimes 4} + \mathbb{E}_{U\sim\mathcal{U}} \sum_{i\neq j} U^{\dagger\otimes 4} |\psi_i\rangle\langle\psi_i|^{\otimes 2} \otimes |\psi_j\rangle\langle\psi_j|^{\otimes 2} U^{\otimes 4} \\ &= d\Phi_{\text{Haar}} + \sum_{i\neq j} \Phi'_{\text{Haar},(i,j)} = d\Phi_{\text{Haar}} + d(d-1)\Phi'_{\text{Haar}}. \end{aligned} \tag{35}$$

where Φ_{Haar} is the fourth moment over the Haar uniform ensemble, calculated as:

$$\Phi_{\text{Haar}} = \frac{24 P_{[4]}}{d(d+1)(d+2)(d+3)}. \tag{36}$$

The term $\Phi'_{\text{Haar},(i,j)}$ corresponds to different states $|\psi_i\rangle$ and $|\psi_j\rangle$. After the scrambling over the Haar uniform ensemble, all such terms for $i \neq j$ become equivalent, thus allowing us to simplify notation by omitting the indices (i, j).

Unlike the fourth moment, Φ'_{Haar} incorporates more than just a combination of the projectors P_λ. Generally, it can be written as

$$\Phi'_{\text{Haar}} = \sum_{\sigma\in S_4} \sum_\lambda \alpha_{ij} R(\sigma) P_\lambda = \sum_{\sigma\in S_4} \sum_\lambda \alpha_{ij} \sigma P_\lambda, \tag{37}$$

where we sum over all the possible permutations σ and partitions λ.For simplicity, we replace $R(\sigma)$ with σ. It is important to note that σP_λ remains within the support of P_λ. Given that $\text{tr}(\Phi'_{\text{Haar}} P_{[1^4]}) = 0$ and $\text{tr}(\Phi'_{\text{Haar}} P_{[2,1^2]}) = 0$, and considering the positivity of Φ'_{Haar}, we claim that the terms $\sigma P_{[1^4]}, \sigma P_{[2,1^2]}$ are excluded from the expansion of Φ'_{Haar}. Moreover, due to the symmetry of the $S = \{(e), (12), (34), (12)(34)\}$, the permutation group is categorized into three distinct parts:

1. $(e), (12), (34), (12)(34)$;
2. $(13), (23), (14), (24), (123), (132), (124), (142), (134), (143), (234), (243), (1234), (1243), (1342), (1432)$;
3. $(13)(24), (14)(23), (1324), (1423)$;

The coefficients of σP_λ where σ belongs to the same group should be equivalent. Denote

$$SP_\lambda := P_\lambda + (12)P_\lambda + (34)P_\lambda + (12)(34)P_\lambda. \tag{38}$$

It is straightforward to verify that:

$$(SP_\lambda)^2 = 4(SP_\lambda). \tag{39}$$

With the fact that SP_λ is hermitian, it implies that $(SP_\lambda)/4$ is a projector. Then, we can verify that $(13)(24)P_\lambda + (14)(23)P_\lambda + (1324)P_\lambda + (1423)P_\lambda = (13)(24)SP_\lambda$ is in the support of SP_λ:

$$(13)(24)SP_\lambda \ (SP_\lambda) = 4(13)(24)SP_\lambda. \tag{40}$$

Moreover, we can find that

$$((13)(24)SP_\lambda)^2 = 4(SP_\lambda). \tag{41}$$

By virtue of the fact that $(13)(24)SP_\lambda$ is also hermitian, we derive that

$$|(13)(24)SP_\lambda| = SP_\lambda. \tag{42}$$

Particularly, in the case $\lambda = [2,2], [3,1]$, we have

$$\begin{aligned} \mathrm{tr}(SP_{[2,2]}) &= \frac{d^2(d^2-1)}{3}, \qquad & (13)(24)SP_{[2,2]} &= SP_{[2,2]}, \\ \mathrm{tr}(SP_{[3,1]}) &= \frac{d(d+2)(d^2-1)}{2}, \qquad & (13)(24)SP_{[3,1]} &= -(SP_{[3,1]}). \end{aligned} \tag{43}$$

Considering that $P_{[4]}P_\lambda = 0$ for any $\lambda \neq [4]$, the sum of σP_λ over the σ in the second group is also proportional to SP_λ for $\lambda = [2,2], [3,1]$. Thus, we express Φ'_{Haar} as:

$$\Phi'_{\mathrm{Haar}} = \alpha_1 P_{[4]} + \alpha_2 SP_{[2,2]} + \alpha_3 SP_{[3,1]}. \tag{44}$$

Using the fact that $\mathrm{tr}(\Phi'_{\mathrm{Haar}}P_{[4]}) = 1/6$, $\mathrm{tr}(\Phi'_{2,2}P_{[2,2]}) = 1/3$, $\mathrm{tr}(\Phi'_{\mathrm{Haar}}P_{[3,1]}) = 1/2$, we can obtain the desired equation. □ We can also compute the Schatten l-norm of Ω_{Haar} analytically. Particularly, $\|\Omega_{\mathrm{Haar}}\|_1 = d^2$:

$$\begin{aligned} \|\Omega_{\mathrm{Haar}}\|_l^l = &\frac{d(d+1)(d+2)(d+3)}{24}\left(\frac{4(d+5)}{(d+1)(d+2)(d+3)}\right)^l \\ &+ \frac{d^2(d^2-1)}{12}\left(\frac{4}{d(d+1)}\right)^l + \frac{d(d+2)(d^2-1)}{8}\left(\frac{4}{(d+1)(d+2)}\right)^l. \end{aligned} \tag{45}$$

Given the ubiquitous application of the Clifford group in quantum information tasks, particularly in shadow estimation, we now delve into the specific case of the Clifford group.

Proposition 12 *Suppose that $\mathcal{U}$ is the Clifford group, $\{|\psi_i\rangle\}_i$ is any orthogonal basis, and $\Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i)$ is defined in Eq. (32). Then*

$$\begin{aligned} \Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i) = &\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{[4]}\right) P_{[4]} + \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right) P^{(n)}_{N,[4]} + \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) SP_{[2,2]} \\ &+ \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right) SP^{(n)}_{N,[2,2]} + \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right) SP_{[3,1]}, \end{aligned} \tag{46}$$

where $SP_\lambda = P_\lambda + (12)P_\lambda + (34)P_\lambda + (12)(34)P_\lambda$ and $(12), (34), (12)(34)$ are the shorthand for the corresponding $R(T)$. P_N is defined in Eq. (30). The coefficients κ are

$$\begin{cases} \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{[4]}\right) = \dfrac{4d^3(d+5) - 8G_1(\{|\psi_i\rangle\})}{d^2(d-1)(d+1)(d+2)(d+4)}, \\ \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right) = \dfrac{-4d^2(d+5) + 2(d+3)G_1(\{|\psi_i\rangle\})}{d(d-1)(d+1)(d+2)(d+4)}, \\ \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) = \dfrac{d^3(d-1) - 2G_2(\{|\psi_i\rangle\})}{d^2(d^2-1)(d^2-4)}, \\ \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right) = \dfrac{G_2(\{|\psi_i\rangle\}) - 2d(d-1)}{2(d^2-1)(d^2-4)}, \\ \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right) = \dfrac{1}{(d+1)(d+2)}, \end{cases} \tag{47}$$

where $G_1(\{|\psi_i\rangle\})$ and $G_2(\{|\psi_i\rangle\})$ are defined as

$$\begin{aligned} G_1(\{|\psi_i\rangle\}) &= \sum_{i,j}\sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^2\langle\psi_j|P|\psi_j\rangle^2 + 2|\langle\psi_i|P|\psi_j\rangle|^2\langle\psi_i|P|\psi_i\rangle\langle\psi_j|P|\psi_j\rangle, \\ G_2(\{|\psi_i\rangle\}) &= \sum_{i,j}\sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^2\langle\psi_j|P|\psi_j\rangle^2 - |\langle\psi_i|P|\psi_j\rangle|^2\langle\psi_i|P|\psi_i\rangle\langle\psi_j|P|\psi_j\rangle. \end{aligned} \tag{48}$$

In particular, assuming that $\{|\psi_i\rangle\}_i$ is the standard computational basis, denoted by $\{|\boldsymbol{b}_i\rangle\}_i$, the coefficients κ are

$$\begin{cases} \kappa\left(\mathrm{Cl}, \{|\boldsymbol{b}_i\rangle\}_i, P_{[4]}\right) = \dfrac{4}{(d+1)(d+2)}, \\ \kappa\left(\mathrm{Cl}, \{|\boldsymbol{b}_i\rangle\}_i, P^{(n)}_{N,[4]}\right) = \dfrac{2d}{(d+1)(d+2)}, \\ \kappa\left(\mathrm{Cl}, \{|\boldsymbol{b}_i\rangle\}_i, SP_{[2,2]}\right) = \dfrac{1}{(d+1)(d+2)}, \\ \kappa\left(\mathrm{Cl}, \{|\boldsymbol{b}_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right) = \dfrac{d}{2(d+1)(d+2)}, \\ \kappa\left(\mathrm{Cl}, \{|\boldsymbol{b}_i\rangle\}_i, SP_{[3,1]}\right) = \dfrac{1}{(d+1)(d+2)}. \end{cases} \tag{49}$$

Proof. According to Refs. [32, 36] and Proposition 11, $\Omega_{\mathrm{Cl}}(|\psi_i\rangle_i)$ can be expressed using a combination of projectors including $P_{[4]}$,$P^{(n)}_{N,[4]}$,$SP_{[2,2]}$,$SP^{(n)}_{N,[2,2]}$, and $SP_{[3,1]}$. The dimensions of these operators $P^{(n)}_{N,[4]}$ and $SP^{(n)}_{N,[2,2]}$ are given by:

$$\mathrm{tr}\left(P^{(n)}_{N,[4]}\right) = \frac{(d+1)(d+2)}{6}, \quad \mathrm{tr}\left(SP^{(n)}_{N,[2,2]}\right) = \frac{4(d^2-1)}{3}. \tag{50}$$

It is straightforward to verify that

$$\begin{aligned} \mathrm{tr}\left[P^{(n)}_{N,[4]}\Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i)\right] &= \frac{1}{6d^2}\sum_{i,j}\sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^2\langle\psi_j|P|\psi_j\rangle^2 \\ &\quad + 4|\langle\psi_i|P|\psi_j\rangle|^2\langle\psi_i|P|\psi_i\rangle\langle\psi_j|P|\psi_j\rangle + |\langle\psi_j|P|\psi_i\rangle|^4, \\ \mathrm{tr}\left[SP^{(n)}_{N,[2,2]}\Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i)\right] &= \frac{1}{6d^2}\sum_{i,j}\sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^2\langle\psi_j|P|\psi_j\rangle^2 \\ &\quad - 2|\langle\psi_i|P|\psi_j\rangle|^2\langle\psi_i|P|\psi_i\rangle\langle\psi_j|P|\psi_j\rangle + |\langle\psi_j|P|\psi_i\rangle|^4. \end{aligned} \tag{51}$$

Then we prove the following lemma, which shows the summation of the first term equals to the third one:

Lemma 13 *Suppose that A, B are two operators acting on the $\mathcal{H}_2^{\otimes n}$, then the following relation holds*

$$\sum_{P\in\mathcal{P}_n} \mathrm{tr}(APBP)^2 = \sum_{P\in\mathcal{P}_n} \mathrm{tr}(AP)^2\mathrm{tr}(BP)^2, \tag{52}$$

where $\mathcal{P}_n$ is the set of Pauli operators in $\mathcal{H}_2^{\otimes n}$.

Proof. We give two separate proofs of this lemma, where the first one is based on the property of the commutant of the Clifford group, while the second one is more general.

As detailed in Lemma 9, certain products $R(\sigma)R(T_4)$ for $\sigma \in S_4$ are actually equivalent. Specifically, we find that $(13)(24)R(T_4) = R(T_4)$ where $(13)(24)$ is the shorthand for the corresponding operator. Therefore, we can deduce that

$$\mathrm{tr}(A\otimes A\otimes B\otimes BR(T_4)) = \mathrm{tr}(A\otimes A\otimes B\otimes B(13)(24)R(T_4)). \tag{53}$$

Using the explicit form of $R(T_4)$ as given in Eq. (30), we can obtain the desired result.

The second proof does not require knowledge of the Clifford commutant. Noting that $\{1/d^{1/2}P\}_{P\in\mathcal{P}_n}$ forms an orthogonal basis of the space of the operators acting on $\mathcal{H}_2^{\otimes n}$, we express A, B as:

$$A = \frac{1}{d}\sum_{P\in\mathcal{P}_n} \mathrm{tr}(AP)P, \quad B = \frac{1}{d}\sum_{P\in\mathcal{P}_n} \mathrm{tr}(BP)P. \tag{54}$$

Substituting these expansions into $\mathrm{tr}(APBP)^2$ allows for straightforward verification of the lemma. □

Following from Lemma 13, the overlap between Ω_{Cl} and $P^{(n)}_{N,[4]}, SP^{(n)}_{N,[2,2]}$, simplifies to

$$\mathrm{tr}\left[P^{(n)}_{N,[4]}\Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i)\right] = \frac{1}{3d^2}G_1, \quad \mathrm{tr}\left[SP^{(n)}_{N,[2,2]}\Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i)\right] = \frac{1}{3d^2}G_2. \tag{55}$$

A detailed but straightforward calculation reveals the coefficients for each term, which confirms the first part of the proposition. As for the second part, G_1 and G_2 can be calculated specifically

$$G_1 = d^3 + 2d^2, \quad G_2 = d^3 - d^2. \tag{56}$$

Thus, the specific form of $\Omega_{\mathrm{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i)$ can be computed in this case. □

In order to quantify the difference between Ω with respect to the unitary 4-design and the Clifford group, we can calculate the 1-norm

$$\begin{aligned}\left\|\Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i) - \Omega_{\mathrm{Haar}}\right\|_1 = &\frac{(d-1)(d+1)(d+2)(d+4)}{3d(d+3)}\left|\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right)\right| \\ &+ \frac{8(d^2-4)(d^2-1)}{3d^2}\left|\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right)\right|.\end{aligned} \tag{57}$$

Assuming the set of bases as the standard computational basis, the 1-norm is given by

$$\left\|\Omega_{\mathrm{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i) - \Omega_{\mathrm{Haar}}\right\|_1 = \frac{2(d-1)\left(3d^2+6d-10\right)}{3d(d+3)} = \mathcal{O}(d). \tag{58}$$

Rescaling it by dividing $\left\|\Omega_{\mathrm{Haar}}\right\|_1$, one find that $\Omega_{\mathrm{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i)$ is close to Ω_{Haar} when the dimension is large enough.

B.2 Interleaved Clifford circuits

We have redefined the structure of non-Clifford gates interleaved with Clifford circuits in Sec. 4, as shown in Fig. 3. Here, these non-Clifford gates are specifically single-phase gates.

We start with the analysis of the M gate and then extend to the general case. For the M-interleaved Clifford circuits, we obtain the following theorem

Theorem 14 *Suppose $\mathcal{U}$ is ensemble of l-layer $\hat{M}_k$-interleaved Clifford circuits, which means $U \in \mathcal{U}$ satisfies $U = U_1\hat{M}_kU_2\ldots\hat{M}_kU_{l+1}$ where $U_1,\ldots,U_{l+1}$ is randomly chosen in the Clifford group and $\hat{M}_k = I^{\otimes n-k}\otimes M^{\otimes k}$. $\{|\psi_i\rangle\}_i$ is any orthogonal basis and $\Omega_{l\hat{M}_k}(\{|\psi_i\rangle\}_i)$ is defined in Eq. (32). Then*

$$\begin{aligned}\Omega_{l\hat{M}_k}(\{|\psi_i\rangle\}_i) = &\kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, P_{[4]}\right)P_{[4]} + \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right)P^{(n)}_{N,[4]} \\ &+ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right)SP_{[2,2]} + \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right)SP^{(n)}_{N,[2,2]} \\ &+ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right)SP_{[3,1]},\end{aligned} \tag{59}$$

where $SP_\lambda = P_\lambda + (12)P_\lambda + (34)P_\lambda + (12)(34)P_\lambda$ and $(12), (34), (12)(34)$ are the shorthand for the corresponding $R(T)$. The coefficients κ are

$$\begin{cases}\kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, P_{[4]}\right) = \left[\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{[4]}\right) - \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right)\frac{\alpha_0^{(k)}}{\alpha_1^{(k)}}\left(1-\left(1-\alpha_1^{(k)}\right)^l\right)\right], \\ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right)\left(1-\alpha_1^{(k)}\right)^l, \\ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) = \left[\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) - \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right)\frac{\beta_0^{(k)}}{\beta_1^{(k)}}\left(1-\left(1-\beta_1^{(k)}\right)^l\right)\right], \\ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right)\left(1-\beta_1^{(k)}\right)^l, \\ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right),\end{cases} \tag{60}$$

$\alpha_0^{(k)}$, $\alpha_1^{(k)}$, $\beta_0^{(k)}$, *and* $\beta_1^{(k)}$ *write*

$$\begin{cases} \alpha_0^{(k)} = -\dfrac{4\left\{d^2\left[1-\left(\frac{3}{4}\right)^k\right]+3d\left[1-\left(\frac{1}{2}\right)^k\right]+2\right\}}{(d-1)(d+1)(d+2)(d+4)}, \\ \alpha_1^{(k)} = \dfrac{d(d+3)\left\{d^2\left[1-\left(\frac{3}{4}\right)^k\right]+3d\left[1-\left(\frac{1}{2}\right)^k\right]+2\right\}}{(d-1)(d+1)(d+2)(d+4)}, \\ \beta_0^{(k)} = -\dfrac{4\left\{d^2\left[1-\left(\frac{3}{4}\right)^k\right]-1\right\}}{(d^2-1)(d^2-4)}, \\ \beta_1^{(k)} = \dfrac{d^2\left\{d^2\left[1-\left(\frac{3}{4}\right)^k\right]-1\right\}}{(d^2-1)(d^2-4)}. \end{cases} \tag{61}$$

Proof. We begin our analysis with one-layer $\hat{M}_k$ Clifford circuits. Considering the action of $\hat{M}_1^{\otimes 4}$ on the commutant of $\mathrm{Cl}(n)^{\otimes 4}$, if $T \in S_4$, it is straightforward to verify that $\hat{M}_1^{\otimes 4} R(T) \hat{M}_1^{\dagger \otimes 4} = R(T)$. Otherwise, after the conjugate action of $\hat{M}_1^{\otimes 4}$, we can derive that

$$\begin{aligned} \hat{M}_1^{\otimes 4} R(T) \hat{M}_1^{\dagger \otimes 4} &= r(T)^{\otimes n-1} \otimes \left[r(T) - 2\left(|0000\rangle\langle 1111| + |1111\rangle\langle 0000|\right)\right] \\ &= r(T)^{\otimes n-1} \otimes \left[r(T) - 2\left(|\mathbf{0}\rangle\langle \mathbf{1}| + |\mathbf{1}\rangle\langle \mathbf{0}|\right)\right]. \end{aligned} \tag{62}$$

For the sake of simplicity, we denote $|0000\rangle$ and $|1111\rangle$ by $|\mathbf{0}\rangle$ and $|\mathbf{1}\rangle$. Therefore, after the action of $\hat{M}_k$, the projector P_λ and $P_{N,\lambda}^{(n)}$ will transform to

$$\begin{cases} \hat{M}_k^{\otimes 4} P_\lambda \hat{M}_k^{\dagger \otimes 4} = P_\lambda, \\ \hat{M}_k^{\otimes 4} P_{N,\lambda}^{(n)} \hat{M}_k^{\dagger \otimes 4} = \displaystyle\sum_{j=0}^{k} (-1)^j \binom{k}{j} P_{N,\lambda}^{(n-j)} \otimes \left(|\mathbf{0}\rangle\langle \mathbf{1}| + |\mathbf{1}\rangle\langle \mathbf{0}|\right)^{\otimes j}. \end{cases} \tag{63}$$

Remark that $P_{N,\lambda}^{(n-j)}$ may project onto different subspaces with $n-j$ qubits. We denote them by the same notation for the sake of simplicity. By virtue of Eq. (63), we obtain the following equation after the conjugation action by

$$\begin{aligned} \Omega_{\hat{M}_k}(\{|\psi_i\rangle\}_i) = {} & \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{[4]}\right) P_{[4]} + \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{N,[4]}^{(n)}\right)\left(P_{N,[4]}^{(n)} - \Delta_{[4]}\right) \\ & + \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) SP_{[2,2]} + \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{N,[2,2]}^{(n)}\right)\left(SP_{N,[2,2]}^{(n)} - \Delta_{[2,2]}\right)) \\ & + \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right) SP_{[3,1]}, \end{aligned} \tag{64}$$

where Δ_λ is defined as

$$\Delta_\lambda := -\mathbb{E}\, U^{\dagger \otimes 4} \left\{ \sum_{j=1}^{k} (-1)^j \binom{k}{j} P_{N,[\lambda]}^{(n-j)} \otimes \left(|\mathbf{0}\rangle\langle \mathbf{1}| + |\mathbf{1}\rangle\langle \mathbf{0}|\right)^{\otimes j} \right\} U^{\otimes 4}. \tag{65}$$

Obviously, it is also combined by the projectors P_λ and $SP_{N,\lambda}^{(n)}$:

$$\Delta_{[4]} = \alpha_0^{(k)} P_{[4]} + \alpha_1^{(k)} P_{N,[4]}^{(n)}, \quad \Delta_{[2,2]} = \beta_0^{(k)} P_{[2,2]} + \beta_1^{(k)} SP_{N,[2,2]}^{(n)}. \tag{66}$$

The coefficients can be determined by solving a linear system of equations. Take $\lambda = [4]$ as an example:

$$\begin{cases} \mathrm{tr} P_{[4]} \alpha_0^{(k)} + \mathrm{tr} P_{N,[4]}^{(n)} \alpha_1^{(k)} = \mathrm{tr}\left[P_{[4]} \Delta_{[4]}\right] = 0, \\ \mathrm{tr}\left[P_N P_{[4]}\right] \alpha_0^{(k)} + \mathrm{tr}\left[P_N P_{N,[4]}^{(n)}\right] \alpha_1^{(k)} = \mathrm{tr}\left[P_N \Delta_{[4]}\right] = \dfrac{d^2\left[1-\left(\frac{3}{4}\right)^k\right]+3d\left[1-\left(\frac{1}{2}\right)^k\right]+2}{6}. \end{cases} \tag{67}$$

Then we can derive that

$$\begin{aligned} \alpha_0^{(k)} &= -\frac{4\left\{d^2\left[1-\left(\frac{3}{4}\right)^k\right]+3d\left[1-\left(\frac{1}{2}\right)^k\right]+2\right\}}{(d-1)(d+1)(d+2)(d+4)} \\ \alpha_1^{(k)} &= \frac{d(d+3)\left\{d^2\left[1-\left(\frac{3}{4}\right)^k\right]+3d\left[1-\left(\frac{1}{2}\right)^k\right]+2\right\}}{(d-1)(d+1)(d+2)(d+4)}. \end{aligned} \tag{68}$$

Similarly, $\beta_0^{(k)}$ and $\beta_1^{(k)}$ can be computed exactly. Therefore, we obtain

$$\begin{cases} \kappa\left(\hat{M}_k, \{|\psi_i\rangle\}_i, P_{[4]}\right) = \left[\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{[4]}\right) - \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{N,[4]}^{(n)}\right)\alpha_0^{(k)}\right], \\ \kappa\left(\hat{M}_k, \{|\psi_i\rangle\}_i, P_{N,[4]}^{(n)}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{N,[4]}^{(n)}\right)\left(1-\alpha_1^{(k)}\right), \\ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) = \left[\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) - \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{N,[2,2]}^{(n)}\right)\beta_0^{(k)}\right], \\ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{N,[2,2]}^{(n)}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{N,[2,2]}^{(n)}\right)\left(1-\beta_1^{(k)}\right), \\ \kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right), \end{cases} \tag{69}$$

Repeating the above step recursively, we get the result. □

Also, the 1-norm of the difference from $\Omega_{l\hat{M}_k}$ to Ω_{Haar} is expressed as

$$\begin{aligned} \left\|\Omega_{l\hat{M}_k}(\{|\psi_i\rangle\}_i) - \Omega_{\mathrm{Haar}}\right\|_1 &= \frac{(d-1)(d+1)(d+2)(d+4)}{3d(d+3)}\left|\kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, P_{N,[4]}^{(n)}\right)\right| \\ &\quad + \frac{8(d^2-4)(d^2-1)}{3d^2}\left|\kappa\left(l\hat{M}_k, \{|\psi_i\rangle\}_i, SP_{N,[2,2]}^{(n)}\right)\right| \\ &= \left\|\Omega_{\mathrm{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i) - \Omega_{\mathrm{Haar}}\right\|_1\left[\left(\frac{3}{4}\right)^{kl} + \mathcal{O}(d^{-1})\right]. \end{aligned} \tag{70}$$

The previous work mainly focuses on the effect of l-layer $\hat{M}_1$-interleaved Clifford circuits [18], which is a corollary of Theorem 14. Due to Eq. (70), when n is large, $\Omega_{l\hat{M}_1}(\{|\psi_i\rangle\}_i)$ will converge to Ω_{Haar} with factor $(3/4)^l$. Strikingly, by virtue of this theorem, we find that the closeness degree of the l-layer $\hat{M}_k$-interleaved Clifford circuits to the unitary 4-design only depends on the number of M gates in the circuits approximately. Particularly, we compare two cases, one layer $\hat{M}_k$-interleaved Clifford circuits and l-layer $\hat{M}_1$-interleaved Clifford circuits.

Proposition 15 *$\alpha_1, \alpha_1^{(k)}, \beta_1, \beta_1^{(k)}$ are defined in Eq. (61) for the $\hat{M}_1$ and $\hat{M}_k$-interleaved Clifford circuits. Then*

$$-k\left(\frac{3}{4}\right)^k d^{-1} < (1-\alpha_1)^k - (1-\alpha_1^{(k)}) \le 0, \quad 0 \le (1-\beta_1)^k - (1-\beta_1^{(k)}) < k\left(\frac{3}{4}\right)^k d^{-1}. \tag{71}$$

where the equality is saturated when $k = 1$.

Proof. First, we prove the inequality about α. Considering the upper bound, when $k = 1$, the equality holds automatically. For $k = 2$, we have

$$(1-\alpha_1)^k - (1-\alpha_1^{(k)}) = \frac{d(-3d^3 - 2d^2 + 29d + 24)}{16(d+2)(d^2-1)} < 0 \tag{72}$$

when $d \ge 4$. For $k = 3$,

$$(1-\alpha_1)^k - (1-\alpha_1^{(k)}) = -\frac{d\left(24d^6 + 59d^5 - 357d^4 - 1071d^3 + 57d^2 + 1480d + 768\right)}{64(d+2)(d+4)\left(d^2-1\right)^3} < 0 \tag{73}$$

when $d \ge 8$. For $k \ge 4$,

$$(1-\alpha_1)^k = \left(\frac{3}{4}\right)^k\left(1 - \frac{3d+1}{3(d-1)(d+1)}\right)^k \le \left(\frac{3}{4}\right)^k\left(1 - \frac{3d+1}{3(d-1)(d+1)}\right)^4. \tag{74}$$

Therefore, we obtain

$$(1-\alpha_1)^k - (1-\alpha_1^{(k)}) \le \frac{f(d) + g(d)(3/4)^k + h(d)(1/2)^k}{81(d-1)^4(d+1)^4(d+2)(d+4)}, \tag{75}$$

where

$$\begin{aligned} f(d) &= 324d^8 + 972d^7 - 324d^6 - 2916d^5 - 972d^4 + 2916d^3 + 1620d^2 - 972d - 648, \\ g(d) &= -81d^9 - 999d^8 - 567d^7 + 5670d^6 + 3645d^5 - 10407d^4 - 8781d^3 + 5632d^2 + 7680d + 2048, \\ h(d) &= -243d^9 - 729d^8 + 729d^7 + 2187d^6 - 729d^5 - 2187d^4 + 243d^3 + 729d^2. \end{aligned} \tag{76}$$

It is straightforward to check that $g(d), h(d) < 0$ when $d \geq 16$. Due to the relation $d^{-1} \leq (1/2)^k < (3/4)^k$ and $1/d < 1$, we deduce

$$f(d) + g(d)\left(\frac{3}{4}\right)^k + h(d)\left(\frac{1}{2}\right)^k < -756d^7 - 162d^6 + 4941d^5 + 1944d^4 - 9678d^3 - 6918d^2 + 5389d + 9080 < 0, \quad (77)$$

when $d \geq 16$. Combining all the situations, we complete the proof of the second inequality.

Next, we consider the lower bound. According to the inequality $(1-x)^k \geq 1 - kx$ when $x \in (0,1)$ and $k \geq 1$, we have

$$\begin{aligned}(1-\alpha_1)^k - (1-\alpha_1^{(k)}) + \frac{k}{d}\left(\frac{3}{4}\right)^k &\geq \frac{\sum_{i=0}^4 \omega_i d^i}{d(d-1)(d+1)(d+2)(d+4)} \\ &\geq \frac{\omega_0 + (\omega_1 + 2\omega_2)d + \omega_3 d^3 + \omega_4 d^4}{d(d-1)(d+1)(d+2)(d+4)},\end{aligned} \quad (78)$$

where

$$\begin{aligned}\omega_0 &= -8k\left(\frac{3}{4}\right)^k, \\ \omega_1 &= 8 - \left(8 + \frac{26k}{3}\right)\left(\frac{3}{4}\right)^k, \\ \omega_2 &= 12 - (6+3k)\left(\frac{3}{4}\right)^k, \\ \omega_3 &= 4 + \left(7 - \frac{k}{3}\right)\left(\frac{3}{4}\right)^k - 9\left(\frac{1}{2}\right)^k > 4 - \left(2 + \frac{k}{3}\right)\left(\frac{3}{4}\right)^k > 0, \\ \omega_4 &= 3\left(\frac{3}{4}\right)^k - 3\left(\frac{1}{2}\right)^k > 0.\end{aligned} \quad (79)$$

Also,

$$\omega_1 + 2\omega_2 = 28 - \left(20 + \frac{44k}{3}\right)\left(\frac{3}{4}\right)^k > 0. \quad (80)$$

Therefore,

$$\begin{aligned}(1-\alpha_1)^k - (1-\alpha_1^{(k)}) + \frac{k}{d}\left(\frac{3}{4}\right)^k &\geq \frac{\omega_0 + (\omega_1 + 2\omega_2)2 + \omega_3 2^3 + \omega_4 2^4}{d(d-1)(d+1)(d+2)(d+4)} \\ &= \frac{96 + (64 - 40k)(3/4)^k - 120(1/2)^k}{d(d-1)(d+1)(d+2)(d+4)} \\ &> \frac{96 - (56 + 40k)(3/4)^k}{d(d-1)(d+1)(d+2)(d+4)} > 0.\end{aligned} \quad (81)$$

The inequality about β can also be proved in this way, which is not shown here. □

According to this proposition, the difference between the one layer $\hat{M}_k$-interleaved Clifford circuits and l-layer $\hat{M}_1$-interleaved Clifford circuits is sufficiently small when the number of the magic gates is the same, and the dimension of the quantum system is large. As a consequence, one can use one layer $\hat{M}_k$-interleaved Clifford circuits to achieve the same performance of l-layer $\hat{M}_1$-interleaved Clifford circuits considering a large quantum system.

More generally, we can replace M by any single phase gate, denoted by P where P is defined as

$$P = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}. \quad (82)$$

Then, we obtain the corresponding result

Proposition 16 *Suppose $\mathcal{U}$ is ensemble of l-layer $\hat{P}_k$-interleaved Clifford circuits, which means $U \in \mathcal{U}$ satisfies $U = U_1 \hat{P}_k U_2 \dots \hat{P}_k U_{l+1}$ where $U_1, \dots, U_{l+1}$ is randomly chosen in the Clifford group and $\hat{P}_k = I^{\otimes n-k} \otimes P^{\otimes k}$. $\Omega_{l\hat{P}_k}(\{|\psi\rangle\}_i)$ is defined in Eq. (32). Then*

$$\begin{aligned}\Omega_{l\hat{P}_k}(\{|\psi_i\rangle\}_i) = &\kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, P_{[4]}\right) P_{[4]} + \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right) P^{(n)}_{N,[4]} \\ &+ \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) SP_{[2,2]} + \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right) SP^{(n)}_{N,[2,2]} \\ &+ \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right) SP_{[3,1]},\end{aligned} \quad (83)$$

where $SP_\lambda = P_\lambda + (12)P_\lambda + (34)P_\lambda + (12)(34)P_\lambda$ *and* $(12), (34), (12)(34)$ *are the shorthand for the corresponding* $R(T)$*. The coefficients* κ *are*

$$\begin{cases} \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, P_{[4]}\right) = \left[\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P_{[4]}\right) - \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right)\frac{\chi_0^{(k)}}{\chi_1^{(k)}}\left(1-\left(1-\chi_1^{(k)}\right)^l\right)\right], \\ \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right)\left(1-\chi_1^{(k)}\right)^l, \\ \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) = \left[\kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) - \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right)\frac{\eta_0^{(k)}}{\eta_1^{(k)}}\left(1-\left(1-\eta_1^{(k)}\right)^l\right)\right], \\ \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right)\left(1-\eta_1^{(k)}\right)^l, \\ \kappa\left(l\hat{P}_k, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right) = \kappa\left(\mathrm{Cl}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right), \end{cases} \tag{84}$$

$\chi_0^{(k)}$, $\chi_1^{(k)}$, $\eta_0^{(k)}$, *and* $\eta_1^{(k)}$ *write*

$$\begin{cases} \chi_0^{(k)} = -\dfrac{4\left\{d^2\left[1-\left(\frac{7+\cos 4\phi}{8}\right)^k\right] + 3d\left[1-\left(\frac{3+\cos 4\phi}{4}\right)^k\right] + 2\left[1-\left(\frac{1+\cos 4\phi}{2}\right)^k\right]\right\}}{(d-1)(d+1)(d+2)(d+4)}, \\ \chi_1^{(k)} = \dfrac{d(d+3)\left\{d^2\left[1-\left(\frac{7+\cos 4\phi}{8}\right)^k\right] + 3d\left[1-\left(\frac{3+\cos 4\phi}{4}\right)^k\right] + 2\left[1-\left(\frac{1+\cos 4\phi}{2}\right)^k\right]\right\}}{(d-1)(d+1)(d+2)(d+4)}, \\ \eta_0^{(k)} = -\dfrac{4\left\{d^2\left[1-\left(\frac{7+\cos 4\phi}{8}\right)^k\right] - 4\left[1-\left(\frac{1+\cos 4\phi}{2}\right)^k\right]\right\}}{(d^2-1)(d^2-4)}, \\ \eta_1^{(k)} = \dfrac{d^2\left\{d^2\left[1-\left(\frac{7+\cos 4\phi}{8}\right)^k\right] - 4\left[1-\left(\frac{1+\cos 4\phi}{2}\right)^k\right]\right\}}{(d^2-1)(d^2-4)}. \end{cases} \tag{85}$$

Proof. When M is replaced by P, the projectors undergo the following transformation:

$$\begin{cases} \hat{P}_k^{\otimes 4} P_\lambda \hat{P}_k^{\dagger\otimes 4} = P_\lambda, \\ \hat{P}_k^{\otimes 4} P^{(n)}_{N,\lambda} \hat{P}_k^{\dagger\otimes 4} = \sum_{j=0}^{k}\left(-\frac{1}{2}\right)^j \binom{k}{j} P^{(n-j)}_{N,\lambda} \otimes \left[(1-e^{-4i\phi})|\mathbf{0}\rangle\langle\mathbf{1}| + (1-e^{4i\phi})|\mathbf{1}\rangle\langle\mathbf{0}|\right]^{\otimes j}. \end{cases} \tag{86}$$

We apply the same methodology of the Theorem 14 to define the term Δ'_λ writes

$$\Delta'_\lambda := -\mathbb{E}\, U^{\dagger\otimes 4}\left\{\sum_{j=1}^{k}\left(-\frac{1}{2}\right)^j \binom{k}{j} P^{(n-j)}_{N,\lambda} \otimes \left[(1-e^{-4i\phi})|\mathbf{0}\rangle\langle\mathbf{1}| + (1-e^{4i\phi})|\mathbf{1}\rangle\langle\mathbf{0}|\right]^{\otimes j}\right\} U^{\otimes 4}. \tag{87}$$

Subsequently, by deducing the trace relationships and solving the corresponding linear equations, we obtain expressions for $\chi_0^{(k)}$, $\chi_1^{(k)}$, $\eta_0^{(k)}$, and $\eta_1^{(k)}$. From these, the final result is straightforwardly derived. □

By virtue of this proposition, the effect of different phase gates can be discussed in detail. The phase gate with $\phi = \pi/4$ or $3\pi/4$ is the most efficient considering all phase gates. It convinces us that the $\hat{M}$-interleaved Clifford circuit is the best choice in this sense. When $\phi = 0$, $\pi/2$, π, the $\hat{P}$-interleaved Clifford circuit is futile. In fact, these phase gates are included in the Clifford group exactly. Furthermore, the coefficients also show a nonstabilizing power of a non-Clifford gate. Compared to the definition in Ref. [17] which corresponds to the stabilizer 2-Rényi entropy, the nonstabilizing power here is characterized by the operator Ω.

B.3 The relation between the fourth moment

In this subsection, we are devoted to showing the discrepancy between $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ and the fourth moment for the state ensemble $\{U^\dagger|\psi_i\rangle\}_{U\sim\mathcal{U},i}$. Also, we find although these two notions are different, they have a close relationship with each other.

Given that the unitary ensembles of interest are invariant under conjugation, we will rewrite the state ensemble as $\{U|\psi_i\rangle\}_{U\sim\mathcal{U},i}$ in subsequent discussions. For the state ensemble $\{U|\psi_i\rangle\}_{U\sim\mathcal{U},i}$, a well-known concept is the projective

t-design, which characterizes the closeness of the first t-th moments from the state ensemble to the Haar measure. It has been widely used in the quantum information. Besides, the operator $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ can also be viewed as the intrinsic property of the state ensemble $\{U|\psi_i\rangle\}_{U\sim\mathcal{U},i}$. Note that only part of $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ is related to the fourth moment, we suppose that the projective 4-design is not enough for $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ to achieve Ω_{Haar}. It means even if $\{U|\psi_i\rangle\}_{U\sim\mathcal{U},i}$ forms a projective 4-design, $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ does not equal to Ω_{Haar}. In particular, considering the state ensemble generated by the Clifford group, i.e., $\{U|\psi_i\rangle\}_{U\sim\text{Cl},i}$, we can analyze this intuition specifically.

As a result of the Proposition 12, $\Omega_{\text{Cl}}(\{|\psi_i\rangle\}_i)$ coincides with Ω_{Haar} iff $G_1 = 2d^2(d+5)/(d+3)$ and $G_2 = 2d(d-1)$. To study the fourth moment of the state ensemble, one can refer to the magic of a single state. Adopting the idea in Reference [17], we define the stabilizer 2-Rényi entropy for an orthogonal basis:

$$M_2(\{|\psi_i\rangle\}_i) = -\log_2 \frac{1}{d}\sum_i \text{tr}\left[R(T_4)|\psi_i\rangle\langle\psi_i|^{\otimes 4}\right] = -\log_2 \sum_i \sum_{P\in\mathcal{P}_n} \frac{1}{d^2}\langle\psi_i|P|\psi_i\rangle^4. \tag{88}$$

It is straightforward to verify that the properties of the stabilizer 2-Rényi entropy also apply here: (i) $M_2(\{|\psi_i\rangle\}_i) = 0$ iff any $|\psi_i\rangle$ in the ensemble is a stabilizer state, otherwise $M_2(\{|\psi_i\rangle\}_i) > 0$; (ii) $M_2(\{U|\psi_i\rangle\}_i) = M_2(\{|\psi_i\rangle\}_i)$ for any $U \in \text{Cl}$; (iii) $M_2(\{|\psi_i\rangle\}_i \otimes \{|\phi_j\rangle\}_j) = M_2(\{|\psi_i\rangle\}_i) + M_2(\{|\phi_j\rangle\}_j)$. Based on the conclusions in Ref. [36], the state ensemble $\{U|\psi_i\rangle\}_{U\sim\text{Cl},i}$ forms a projective 4-design iff $M_2(\{|\psi_i\rangle\}_i) = \log_2(d+3)/4$. We claim that this condition is independent of the constraints of $\Omega_{\text{Cl}}(\{|\psi_i\rangle\}_i)$ in some sense. For example, we exhibit a simple two qubits example. Assume that the orthogonal basis is separable, which means it can be written as

$$\{|\psi_i\rangle\}_i = \{\{\cos\theta_1|0\rangle + \sin\theta_1|1\rangle, -\sin\theta_1|0\rangle + \cos\theta_1|1\rangle\} \otimes \{\cos\theta_2|0\rangle + \sin\theta_2|1\rangle, -\sin\theta_2|0\rangle + \cos\theta_2|1\rangle\}\}. \tag{89}$$

In this case, we can calculate the stabilizer 2-Rényi entropy, G_1 and G_2 exactly:

$$\begin{aligned}
M_2(\{|\psi_i\rangle\}_i) &= -\log_2\left[\frac{1}{4}(1+\sin^4 2\theta_1+\cos^4 2\theta_1)(1+\sin^4 2\theta_2+\cos^4 2\theta_2)\right],\\
G_1(\{|\psi_i\rangle\}_i) &= 16(1+\sin^4 2\theta_1+\cos^4 2\theta_1)(1+\sin^4 2\theta_2+\cos^4 2\theta_2)\\
&\quad + 8\left[1+(\sin^2 2\theta_1-\cos^2 2\theta_1)^2\right]\left[1+(\sin^2 2\theta_2-\cos^2 2\theta_2)^2\right],\\
G_2(\{|\psi_i\rangle\}_i) &= 16(1+\sin^4 2\theta_1+\cos^4 2\theta_1)(1+\sin^4 2\theta_2+\cos^4 2\theta_2)\\
&\quad - 4\left[1+(\sin^2 2\theta_1-\cos^2 2\theta_1)^2\right]\left[1+(\sin^2 2\theta_2-\cos^2 2\theta_2)^2\right].
\end{aligned} \tag{90}$$

The constraint of projective 4-design confines that $M_2(\{|\psi_i\rangle\}_i) = \log_2 7/4$, while the restriction on Ω_{Haar} indicates that $M_2(\{|\psi_i\rangle\}_i) = \log_2 28/13$. In this example, if the state ensemble $\{U|\psi_i\rangle\}_{U\sim\text{Cl},i}$ constructs a projective 4-design, then $\Omega_{\text{Cl}}(\{|\psi_i\rangle\}_i)$ will never achieve the Haar random case, and vice versa. The contradiction implies that $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ is a brand-new structure compared to the fourth moment for such the state ensemble.

Although $M_2(\{|\psi_i\rangle\}_i)$ is not suitable for analyzing $\Omega_{\text{Cl}}(\{|\psi_i\rangle\}_i)$, it still gives an bound for G_1 and G_2, which is shown in the following context. Moreover, we claim that when the Clifford orbits of $\{|\psi_i\rangle\}_i$ form a projective 4-design, $\Omega_{\text{Cl}}(\{|\psi_i\rangle\}_i)$ will be closer to the Haar random case than the one with respect to the standard computational basis:

Theorem 17 *Suppose that $\mathcal{U}$ is the Clifford group, $\{|\psi_i\rangle\}_i$ is an orthogonal basis satisfying that $M_2(\{|\psi_i\rangle\}_i) = \log_2(d+3)/4$. Then*

$$\|\Omega_{\text{Cl}}(\{|\psi_i\rangle\}_i) - \Omega_{\text{Haar}}\|_1 \,/\, \|\Omega_{\text{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i) - \Omega_{\text{Haar}}\|_1 \le \mathcal{O}(d^{-1}). \tag{91}$$

Proof. We initiate our proof by establishing bounds for G_1 and G_2 the stabilizer 2-Rényi entropy $M_2(\{|\psi_i\rangle\}_i)$. First, we have

$$\begin{aligned}
&\left|\sum_{i,j}\sum_{P\in\mathcal{P}_n} |\langle\psi_i|P|\psi_j\rangle|^2\langle\psi_i|P|\psi_i\rangle\langle\psi_j|P|\psi_j\rangle\right| \le \sum_{i,j}\sum_{P\in\mathcal{P}_n} |\langle\psi_i|P|\psi_j\rangle|^2\,|\langle\psi_i|P|\psi_i\rangle\langle\psi_j|P|\psi_j\rangle|\\
&\le \sum_{i,j}\left[\sum_{P\in\mathcal{P}_n}|\langle\psi_i|P|\psi_j\rangle|^4\sum_{P\in\mathcal{P}_n}\langle\psi_i|P|\psi_i\rangle^2\langle\psi_j|P|\psi_j\rangle^2\right]^{1/2} = \sum_{i,j}\sum_{P\in\mathcal{P}_n}\langle\psi_i|P|\psi_i\rangle^2\langle\psi_j|P|\psi_j\rangle^2.
\end{aligned} \tag{92}$$

where we use the Triangle inequality, Cauchy-Schwarz inequality, and Lemma 13. Furthermore, we can obtain that

$$\begin{aligned}
&\sum_{i,j}\sum_{P\in\mathcal{P}_n}\langle\psi_i|P|\psi_i\rangle^2\langle\psi_j|P|\psi_j\rangle^2 \le \sum_{i,j}\left[\sum_{P\in\mathcal{P}_n}\langle\psi_i|P|\psi_i\rangle^4\sum_{P\in\mathcal{P}_n}\langle\psi_j|P|\psi_j\rangle^4\right]^{1/2}\\
&= \left[\sum_i\left(\sum_{P\in\mathcal{P}_n}\langle\psi_i|P|\psi_i\rangle^4\right)^{1/2}\right]^2 \le d\sum_i\sum_{P\in\mathcal{P}_n}\langle\psi_i|P|\psi_i\rangle^4 = d^3 2^{-M_2(\{|\psi_i\rangle\}_i)}.
\end{aligned} \tag{93}$$

From this, we obtain

$$-d^3 2^{-M_2(\{|\psi_i\rangle\}_i)} \leq G_1 \leq 3d^3 2^{-M_2(\{|\psi_i\rangle\}_i)}, \quad 0 \leq G_2 \leq 2d^3 2^{-M_2(\{|\psi_i\rangle\}_i)}. \tag{94}$$

If the state ensemble $\{U|\psi_i\rangle\}_{U\sim\mathcal{U},i}$ form a projective 4-design, then the stabilizer 2-Rényi entropy is $\log_2(d+3)/4$. By substituting this into Eq. (57), we deduce that:

$$\|\Omega_{\mathrm{Cl}}(\{|\psi_i\rangle\}_i) - \Omega_{\mathrm{Haar}}\|_1 \leq \max\{\frac{4(d-1)(7d+6)}{3d(d+3)}, \frac{4(9d^2+d+6)}{3d(d+3)}\} = \frac{4(9d^2+d+6)}{3d(d+3)}. \tag{95}$$

This completes the proof by dividing the case for the standard computational basis. □

Remark that the established bounds of G_1 and G_2 in Eq. (94) are not tight when considering the standard computational basis. Generally, these bounds are more suitable for $M_2(\{|\psi_i\rangle\}_i)$ of order $\log_2 \mathcal{O}(d)$ and less stringent for values of the order order $\log_2 \mathcal{O}(1)$. Here we propose another set of bounds that are tighter for smaller values of $M_2(\{|\psi_i\rangle\}_i)$.

We first find that for any $i \neq j$

$$|\langle\psi_i|P|\psi_i\rangle| \leq \sqrt{1 - |\langle\psi_i|P|\psi_j\rangle|^2}, \tag{96}$$

where we use the completeness of the set $\{|\psi_i\rangle\}_i$. Therefore,

$$\begin{aligned}
&\left|\sum_{i,j}\sum_{P\in\mathcal{P}_n} |\langle\psi_i|P|\psi_j\rangle|^2 \langle\psi_i|P|\psi_i\rangle\langle\psi_j|P|\psi_j\rangle\right| \\
\leq &\sum_i \sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^4 + \sum_{i\neq j}\sum_{P\in\mathcal{P}_n} |\langle\psi_i|P|\psi_j\rangle|^2 \left(1 - |\langle\psi_i|P|\psi_j\rangle|^2\right) \\
= &d^2(d-1) + 2d^2 2^{-M_2(\{|\psi_i\rangle\}_i)} - \sum_{i,j}\sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^2 \langle\psi_j|P|\psi_j\rangle^2.
\end{aligned} \tag{97}$$

where the last equality uses $\sum_{P\in\mathcal{P}_n} |\langle\psi_i|P|\psi_j\rangle|^2 = d$ and Lemma 13. Furthermore,

$$\sum_{i,j}\sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^2 \langle\psi_j|P|\psi_j\rangle^2 \geq \sum_i \sum_{P\in\mathcal{P}_n} \langle\psi_i|P|\psi_i\rangle^4 = d^2 2^{-M_2(\{|\psi_i\rangle\}_i)}. \tag{98}$$

Combining with Eq. (93), we obtain

$$\begin{aligned}
-d^2 2^{-M_2(\{|\psi_i\rangle\}_i)} - 2d^2(d-1) \leq &G_1 \leq 2d^2(d-1) + 3d^2 2^{-M_2(\{|\psi_i\rangle\}_i)} \\
-d^2(d-1) \leq &G_2 \leq d^2(d-1) + 2d^2 2^{-M_2(\{|\psi_i\rangle\}_i)}.
\end{aligned} \tag{99}$$

For an orthogonal basis composed of stabilizer states, this formulation offers a tighter upper bound. In conclusion, we deduce that

$$\begin{aligned}
G_1 \in &\Big[\max\{-d^3 2^{-M_2(\{|\psi_i\rangle\}_i)}, -d^2 2^{-M_2(\{|\psi_i\rangle\}_i)} - 2d^2(d-1)\}\,, \\
&\min\{3d^3 2^{-M_2(\{|\psi_i\rangle\}_i)}, 2d^2(d-1) + 3d^2 2^{-M_2(\{|\psi_i\rangle\}_i)}\}\Big], \\
G_2 \in &\Big[0, \min\{2d^3 2^{-M_2(\{|\psi_i\rangle\}_i)}, d^2(d-1) + 2d^2 2^{-M_2(\{|\psi_i\rangle\}_i)}\}\Big].
\end{aligned} \tag{100}$$

C Proofs of Propositions 1, 6 and Theorems 2, 5

Here, we first prove the propositions and theorems about the general bounds for the variance of the thrifty shadow. For the convenience of the following discussion, we start by rewriting Propositions 11, 12, and Theorem 14 in the form of the summation of $R(T)$ with $T \in \Sigma_{4,4}$. By the definition, $\Omega_{\mathcal{U}}$ is invariant under the action of $S \times S$ and the conjugation action, where $S = \{(e), (12), (34), (12)(34)\}$. So all stochastic Lagrangian subspaces $T \in \Sigma_{4,4}$ can be classified as five distinct sets, denoted by $\mathbb{G}_i$ respectively:

1. $(e), (12), (34), (12)(34)$;
2. $(13), (23), (14), (24), (123), (132), (124), (142), (134), (143), (234), (243), (1234), (1243), (1342), (1432)$;
3. $(13)(24), (14)(23), (1324), (1423)$;
4. $T_4, (12)T_4$;

5. $(13)T_4, (23)T_4, (123)T_4, (132)T_4$.

Thus, we define

$$\mathcal{R}_i = \sum_{T\in\mathbb{G}_i} R(T). \tag{101}$$

According to the symmetry,

$$\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i) = \sum_{i=1}^{5} r_i\left(\mathcal{U}, \{|\psi_i\rangle\}_i\right)\mathcal{R}_i. \tag{102}$$

In view of the unitary 4-design, the Clifford group, and the interleaved Clifford circuits, the coefficients r_i have explicit relationships with κ:

$$\begin{cases} r_1\left(\mathcal{U}, \{|\psi_i\rangle\}_i\right) = \dfrac{1}{24}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, P_{[4]}\right) + \dfrac{1}{3}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) + \dfrac{1}{2}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right), \\ r_2\left(\mathcal{U}, \{|\psi_i\rangle\}_i\right) = \dfrac{1}{24}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, P_{[4]}\right) - \dfrac{1}{6}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right), \\ r_3\left(\mathcal{U}, \{|\psi_i\rangle\}_i\right) = \dfrac{1}{24}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, P_{[4]}\right) + \dfrac{1}{3}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, SP_{[2,2]}\right) - \dfrac{1}{2}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, SP_{[3,1]}\right), \\ r_4\left(\mathcal{U}, \{|\psi_i\rangle\}_i\right) = \dfrac{1}{6d}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right) + \dfrac{4}{3d}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right), \\ r_5\left(\mathcal{U}, \{|\psi_i\rangle\}_i\right) = \dfrac{1}{6d}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, P^{(n)}_{N,[4]}\right) - \dfrac{2}{3d}\kappa\left(\mathcal{U}, \{|\psi_i\rangle\}_i, SP^{(n)}_{N,[2,2]}\right). \end{cases} \tag{103}$$

Then, we present alternative formulations of $\Omega_{\mathcal{U}}(\{|\psi_i\rangle\}_i)$ employing various unitary ensembles. Specifically for applications in the thrifty shadow, we primarily assume the orthogonal basis as the standard computational basis.

Lemma 18 *Suppose that $\mathcal{U}$ is a Haar random ensemble or a unitary 4-design, $\{|\psi_i\rangle\}_i$ is any orthogonal basis and Ω_{Haar} is defined in Eq. (32). Then*

$$\Omega_{\text{Haar}} = \frac{(d^2+4d+2)}{d(d+1)(d+2)(d+3)}\mathcal{R}_1 - \frac{1}{d(d+1)(d+2)(d+3)}\mathcal{R}_2 + \frac{1}{d(d+1)(d+3)}\mathcal{R}_3. \tag{104}$$

Lemma 19 *Suppose that $\mathcal{U}$ is the Clifford group, $\{|\boldsymbol{b}_i\rangle\}_i$ is the standard computational basis and $\Omega_{\text{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i)$ is defined in Eq. (32). Then*

$$\Omega_{\text{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i) = \frac{1}{(d+1)(d+2)}\mathcal{R}_1 + \frac{1}{(d+1)(d+2)}\mathcal{R}_4. \tag{105}$$

Lemma 20 *Suppose $\mathcal{U}$ is ensemble of one layer $\hat{M}_k$-interleaved Clifford circuits, which means $U\in\mathcal{U}$ satisfies $U = U_1\hat{M}_kU_2$ where U_1, U_2 is randomly chosen in the Clifford group and $\hat{M}_k = I^{\otimes n-k}\otimes M^{\otimes k}$. $\{|\boldsymbol{b}_i\rangle\}_i$ is the standard computational basis and $\Omega_{\hat{M}_k}(\{|\boldsymbol{b}_i\rangle\}_i)$ is defined in Eq. (32). Then*

$$\begin{aligned} \Omega_{\hat{M}_k}(\{|\boldsymbol{b}_i\rangle\}_i) = {} & \frac{(d^4+4d^3-6d^2-16d+16) - d^3(d+2)(3/4)^k - (d-2)d^2(1/2)^k}{(d-2)(d-1)(d+1)(d+2)^2(d+4)}\mathcal{R}_1 \\ & - \frac{d^2\left[d+1-2d(3/4)^k + (d-2)(1/2)^k\right]}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)}\mathcal{R}_2 \\ & + \frac{d\left[(d+1)(d^2+2d-4) - (d+2)d^2(3/4)^k - (d-2)d(1/2)^k\right]}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)}\mathcal{R}_3 \\ & + \frac{(d^2+3d-2)d^3(3/4)^k + \left(d^2+d-6\right)d^2(1/2)^k - 4(d+1)(d^2+2d-4)}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)}\mathcal{R}_4 \\ & + \frac{d\left[-(d+2)d^2(3/4)^k + \left(d^2+d-6\right)d(1/2)^k + 4(d+1)\right]}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)}\mathcal{R}_5. \end{aligned} \tag{106}$$

Lemma 21 *Suppose $\mathcal{U}$ is ensemble of l-layer $\hat{M}_1$-interleaved Clifford circuits, which means $U\in\mathcal{U}$ satisfies $U = U_1\hat{M}_1U_2\ldots\hat{M}_1U_{l+1}$ where $U_1,\ldots,U_{l+1}$ is randomly chosen in the Clifford group and $\hat{M}_1 = I^{\otimes n-1}\otimes M.\{|\boldsymbol{b}_i\rangle\}_i$ is the*

standard computational basis and $\Omega_{l\hat{M}_1}(\{|\boldsymbol{b}_i\rangle\}_i)$ *is defined in Eq. (32). Then*

$$\begin{aligned}
\Omega_{l\hat{M}_1}(\{|\boldsymbol{b}_i\rangle\}_i) = &\left[\frac{d^2+4d+2}{d(d+1)(d+2)(d+3)} - \frac{\left(1-\frac{d(d+3)}{4(d^2-1)}\right)^l}{3(d+1)(d+2)(d+3)} - \frac{2\left(1-\frac{d^2}{4(d^2-1)}\right)^l}{3d(d+1)(d+2)}\right]\mathcal{R}_1 \\
&+\left[-\frac{1}{d(d+1)(d+2)(d+3)} - \frac{\left(1-\frac{d(d+3)}{4(d^2-1)}\right)^l}{3(d+1)(d+2)(d+3)} + \frac{\left(1-\frac{d^2}{4(d^2-1)}\right)^l}{3d(d+1)(d+2)}\right]\mathcal{R}_2 \\
&+\left[\frac{1}{d(d+1)(d+3)} - \frac{\left(1-\frac{d(d+3)}{4(d^2-1)}\right)^l}{3(d+1)(d+2)(d+3)} - \frac{2\left(1-\frac{d^2}{4(d^2-1)}\right)^l}{3d(d+1)(d+2)}\right]\mathcal{R}_3 \\
&+\left[\frac{\left(1-\frac{d(d+3)}{4(d^2-1)}\right)^l}{3(d+1)(d+2)} + \frac{2\left(1-\frac{d^2}{4(d^2-1)}\right)^l}{3(d+1)(d+2)}\right]\mathcal{R}_4 \\
&+\left[\frac{\left(1-\frac{d(d+3)}{4(d^2-1)}\right)^l}{3(d+1)(d+2)} - \frac{\left(1-\frac{d^2}{4(d^2-1)}\right)^l}{3(d+1)(d+2)}\right]\mathcal{R}_5.
\end{aligned} \tag{107}$$

Using these lemmas, we can prove the propositions and theorems in Sections. 3 and 4. *Proof.* [Proof of Proposition 1] By virtue of Ref. [10], for any state ρ, traceless operator O, and $T \in \Sigma_{4,4}$, we have

$$|\mathrm{tr}(R(T)O\otimes\rho\otimes O\otimes\rho)| \le \mathrm{tr}(O^2). \tag{108}$$

In addition,

$$\mathrm{tr}(\mathcal{R}_1 O\otimes\rho\otimes O\otimes\rho) = \mathrm{tr}(O\rho)^2 \le \mathrm{tr}(O^2), \tag{109}$$

where the last inequality is followed by the Cauchy-Schwarz inequality and $\|\rho\|_2 \le 1$. Therefore, due to Lemma 18, we deduce that

$$\begin{aligned}
V_*(O,\rho) &= (d+1)^2\mathrm{tr}\left[\Omega_{\mathrm{Haar}}O\otimes\rho\otimes O\otimes\rho\right] - \mathrm{tr}(O\rho)^2 \\
&= \left(\frac{(d+1)(d^2+4d+2)}{d(d+2)(d+3)} - 1\right)\mathrm{tr}\left(\mathcal{R}_1 O\otimes\rho\otimes O\otimes\rho\right) \\
&- \frac{d+1}{d(d+2)(d+3)}\mathrm{tr}\left(\mathcal{R}_2 O\otimes\rho\otimes O\otimes\rho\right) + \frac{d+1}{d(d+3)}\mathrm{tr}\left(\mathcal{R}_3 O\otimes\rho\otimes O\otimes\rho\right) \\
&\le \left(\frac{(d+1)(d^2+4d+2)}{d(d+2)(d+3)} - 1 + \frac{16(d+1)}{d(d+2)(d+3)} + \frac{4(d+1)}{d(d+3)}\right)\mathrm{tr}(O^2) \\
&= \frac{4d^2+28d+26}{d(d+2)(d+3)}\mathrm{tr}(O^2) = \mathcal{O}(d^{-1}\mathrm{tr}(O^2)).
\end{aligned} \tag{110}$$

□

Proof. [Proof of Theorem 2] Using Lemma 19 and the same conclusion in the proof of Proposition 1,

$$V_*(O,\rho) = -\frac{1}{d+2}\mathrm{tr}\left(\mathcal{R}_1 O\otimes\rho\otimes O\otimes\rho\right) + \frac{d+1}{d+2}\mathrm{tr}\left(\mathcal{R}_4 O\otimes\rho\otimes O\otimes\rho\right) \le \frac{2(d+1)}{d+2}\mathrm{tr}(O^2). \tag{111}$$

□

Proof. [Proof of Theorem 5] By virtue of the analysis in the proof of Proposition 1, the expression for evolves as:

$$\begin{aligned}
&V_*(O,\rho) = (d+1)^2\mathrm{tr}\left[\Omega_{\hat{M}_k}(\{|\boldsymbol{b}_i\rangle\}_i)O\otimes\rho\otimes O\otimes\rho\right] - \mathrm{tr}(O\rho)^2 \\
&= \left[(d+1)^2 r_1\left(\hat{M}_k,\{|\boldsymbol{b}_i\rangle\}_i\right) - 1\right]\mathrm{tr}(O\rho)^2 + \sum_{i=2}^{5}(d+1)^2 r_i\left(\hat{M}_k,\{|\boldsymbol{b}_i\rangle\}_i\right)\mathrm{tr}\left[\mathcal{R}_i O\otimes\rho\otimes O\otimes\rho\right] \\
&< \left[16\left|r_2\left(\hat{M}_k,\{|\boldsymbol{b}_i\rangle\}_i\right)\right| + 4\left|r_3\left(\hat{M}_k,\{|\boldsymbol{b}_i\rangle\}_i\right)\right| + 2\left|r_4\left(\hat{M}_k,\{|\boldsymbol{b}_i\rangle\}_i\right)\right| + 4\left|r_5\left(\hat{M}_k,\{|\boldsymbol{b}_i\rangle\}_i\right)\right|\right](d+1)^2\mathrm{tr}(O^2).
\end{aligned} \tag{112}$$

In the last inequality, we use the fact that

$$\begin{aligned}(d+1)^2 r_1\left(\hat{M}_k, \{|\boldsymbol{b}_i\rangle\}_i\right) - 1 &= -\frac{(d+2)d^3(3/4)^k + (d-2)d^2(1/2)^k - 2(3d+4)d + 16}{(d-2)(d-1)(d+2)^2(d+4)} \\ &\leq -\frac{(d+2)d^3(3/4)^n + (d-2)d^2(1/2)^n - 2(3d+4)d + 16}{(d-2)(d-1)(d+2)^2(d+4)} < 0.\end{aligned} \tag{113}$$

Similarly, we analyze all the coefficients r_i

$$r_2\left(\hat{M}_k, \{|\boldsymbol{b}_i\rangle\}_i\right) = -\frac{d^2\left[d+1-2d(3/4)^k + (d-2)(1/2)^k\right]}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)} \leq 0. \tag{114}$$

$$r_3\left(\hat{M}_k, \{|\boldsymbol{b}_i\rangle\}_i\right) = \frac{d\left[(d+1)(d^2+2d-4) - (d+2)d^2(3/4)^k - (d-2)d(1/2)^k\right]}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)} \geq 0. \tag{115}$$

$$r_4\left(\hat{M}_k, \{|\boldsymbol{b}_i\rangle\}_i\right) = \frac{(d^2+3d-2)d^3(3/4)^k + \left(d^2+d-6\right)d^2(1/2)^k - 4(d+1)(d^2+2d-4)}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)} \geq 0. \tag{116}$$

$$r_5\left(\hat{M}_k, \{|\boldsymbol{b}_i\rangle\}_i\right) = \frac{d\left[-(d+2)d^2(3/4)^k + \left(d^2+d-6\right)d(1/2)^k + 4(d+1)\right]}{(d-2)(d-1)(d+1)^2(d+2)^2(d+4)} \leq 0. \tag{117}$$

Remark that $\Omega_{\hat{M}_k}(\{|\boldsymbol{b}_i\rangle\}_i)$ is closer to Ω_{Haar} than $\Omega_{\text{Cl}}(\{|\boldsymbol{b}_i\rangle\}_i)$, which gives an intuition about how large these values are. Therefore,

$$V_*(O,\rho) < \frac{2\left(d^2+3d-18\right)d^3(3/4)^k - 2\left(d^2-5d+6\right)d^2(1/2)^k + 4\left(d^4+5d^3-8d^2-4d+8\right)}{(d-2)(d-1)(d+2)^2(d+4)}\text{tr}(O^2). \tag{118}$$

Moreover, we can verify that

$$\begin{aligned}&\frac{2\left(d^2+3d-18\right)d^3(3/4)^k - 2\left(d^2-5d+6\right)d^2(1/2)^k + 4\left(d^4+5d^3-8d^2-4d+8\right)}{(d-2)(d-1)(d+2)^2(d+4)} - 2\left(\frac{3}{4}\right)^k - \frac{4}{d} \\ &= \frac{2\left(-d^5+5d^4-6d^3\right)2^{1-k} + 4\left(\frac{3}{4}\right)^k\left(-d^5-8d^4+14d^3+4d^2-16d\right) + 8\left(-3d^3+12d^2+8d-16\right)}{(d-2)(d-1)d(d+2)^2(d+4)} \\ &\leq \frac{2\left(-d^5+5d^4-6d^3\right)2^{1-n} + 4\left(\frac{3}{4}\right)^n\left(-d^5-8d^4+14d^3+4d^2-16d\right) + 8\left(-3d^3+12d^2+8d-16\right)}{(d-2)(d-1)d(d+2)^2(d+4)} \\ &\leq 0.\end{aligned} \tag{119}$$

where the first inequality follows from $-d^5+5d^4-6d^3 \leq 0$, $-d^5-8d^4+14d^3+4d^2-16d < 0$, and $k \leq n$. Since the numerator in the last step is monotone decreasing and the value is zero when $n=1$, we deduce that the second inequality, thus concluding the proof □

Proof. [Proof of Proposition 6] Following the similar approach of the proof of Proposition 5, we find that

$$V_*(O,\rho) < \frac{12d^2+84d+78+\left(-2d^3+3d^2+5d\right)a^k + \left(8d^3+38d^2+48d+18\right)b^k}{3d(d+2)(d+3)}\text{tr}(O^2). \tag{120}$$

Since the derivation is direct, we will not show it in detail. Here $a = 1-d(d+3)/4(d^2-1)$ and $b = 1-d^2/4(d^2-1)$ for the sake of simplicity. Due to the inequalities $a, b < 3/4$, we find that

$$\begin{aligned}&\frac{12d^2+84d+78+\left(-2d^3+3d^2+5d\right)a^k + \left(8d^3+38d^2+48d+18\right)b^k}{3d(d+2)(d+3)} - 2\left(\frac{3}{4}\right)^k - \frac{8}{d} \\ &= \frac{\left(11d^2+17d+18\right)(3/4)^k}{3d(d+2)(d+3)} + \frac{-12d^2-36d-66}{3d(d+2)(d+3)} \leq -\frac{5d^2+31d+70}{4d(d+2)(d+3)} < 0.\end{aligned} \tag{121}$$

Thereby, we finish the proof of the proposition. □

D Proofs of Theorems 3, 4, 7, and 8

In this section, we prove the theorems and propositions corresponding to the fidelity estimation. *Proof.* [Proof of Theorem 3] By the definition of variance in the original shadow estimation [1], one obtains that

$$V(O,\rho)=\frac{d+1}{d+2}\left[\mathrm{tr}(O^2)+2\mathrm{tr}(\rho O^2)\right]-\mathrm{tr}(O\rho)^2=\frac{2(d-1)}{d+2}. \tag{122}$$

We now proceed to compute $V_*(O,\rho)$. Based on the definition of $V_*(O,\rho)$ and Lemma 18, we express $V_*(O,\rho)$ as

$$\begin{aligned}V_*(O,\rho)=&\frac{(d+1)(d^2+4d+2)}{d(d+2)(d+3)}\mathrm{tr}\left(\mathcal{R}_1O\otimes\rho\otimes O\otimes\rho\right)-\frac{d+1}{d(d+2)(d+3)}\mathrm{tr}\left(\mathcal{R}_2O\otimes\rho\otimes O\otimes\rho\right)\\&+\frac{d+1}{d(d+3)}\mathrm{tr}\left(\mathcal{R}_3O\otimes\rho\otimes O\otimes\rho\right)-\mathrm{tr}(O\rho)^2.\end{aligned} \tag{123}$$

For each term, it is easy to calculate that

$$\begin{aligned}&\mathrm{tr}\left(\mathcal{R}_1O\otimes\rho\otimes O\otimes\rho\right)=\mathrm{tr}(O\rho)^2=\frac{(d-1)^2}{d^2},\\&\mathrm{tr}\left(\mathcal{R}_2O\otimes\rho\otimes O\otimes\rho\right)=\mathrm{tr}(O^2)+4\mathrm{tr}(O^2\rho)+2\mathrm{tr}(O^2\rho^2)+2\mathrm{tr}(O\rho O\rho)=\frac{(d-1)(9d-8)}{d^2},\\&\mathrm{tr}\left(\mathcal{R}_3O\otimes\rho\otimes O\otimes\rho\right)=\mathrm{tr}(O^2)\mathrm{tr}(\rho^2)+\mathrm{tr}(O\rho)^2+2\mathrm{tr}(O^2\rho^2)=\frac{(d-1)(4d-3)}{d^2}.\end{aligned} \tag{124}$$

Upon substituting these calculations into Eq. (123) and applying the definition of $V_R(O,\rho)$, we obtain the desired result. □

Proof. [Proof of Theorem 4] Similar to the proof of Proposition 3, we need to calculate the $V_*(O,\rho)$. The expression for $V_*(O,\rho)$ is given by

$$\begin{aligned}V_*(O,\rho)&=\frac{d+1}{d+2}\mathrm{tr}\left(\mathcal{R}_1O\otimes\rho\otimes O\otimes\rho\right)+\frac{d+1}{d+2}\mathrm{tr}\left(\mathcal{R}_4O\otimes\rho\otimes O\otimes\rho\right)-\mathrm{tr}(O\rho)^2\\&=\frac{d+1}{d+2}\mathrm{tr}\left(\mathcal{R}_4O\otimes\rho\otimes O\otimes\rho\right)-\frac{1}{d+2}\mathrm{tr}(O\rho)^2.\end{aligned} \tag{125}$$

To further refine this calculation, we substitute the specific choice of O and ρ, yielding

$$O\otimes\rho\otimes O\otimes\rho=|\phi\rangle\langle\phi|^{\otimes4}-\frac{1}{d}I\otimes|\phi\rangle\langle\phi|^{\otimes3}-\frac{1}{d}|\phi\rangle\langle\phi|^{\otimes2}\otimes I\otimes|\phi\rangle\langle\phi|+\frac{1}{d^2}I\otimes|\phi\rangle\langle\phi|\otimes I\otimes|\phi\rangle\langle\phi|. \tag{126}$$

For the first part,

$$\mathrm{tr}(\mathcal{R}_4|\phi\rangle\langle\phi|^{\otimes4})=2^{1-M_2(|\phi\rangle)}, \tag{127}$$

where we use the fact $\mathrm{tr}\left(R(T)|\phi\rangle\langle\phi|^{\otimes4}\right)=2^{-M_2(|\phi\rangle)}$ for any $T\in S_3T_4$.

For the second part,

$$\begin{aligned}\mathrm{tr}(\mathcal{R}_4I\otimes|\phi\rangle\langle\phi|^{\otimes3})&=\mathrm{tr}(R(T_4)I\otimes|\phi\rangle\langle\phi|^{\otimes3})+\mathrm{tr}((12)R\left(T_4\right)I\otimes|\phi\rangle\langle\phi|^{\otimes3})\\&=\frac{1}{d}\sum_{P\in P_n}\left[\mathrm{tr}(P)\mathrm{tr}(P|\phi\rangle\langle\phi|)^3+\mathrm{tr}(|\phi\rangle\langle\phi|)\mathrm{tr}(P|\phi\rangle\langle\phi|)^2\right]\\&=1+\frac{1}{d}\sum_{P\in P_n}\mathrm{tr}(P|\phi\rangle\langle\phi|)^2=1+\mathrm{tr}(|\phi\rangle\langle\phi|^2)=2.\end{aligned} \tag{128}$$

In the third equality, we use the fact that $\mathrm{tr}(P)=d\delta_{P,I}$. The last equality is followed by the fact that $\{1/\sqrt{d}P\}_{P\in P_n}$ is a set of orthogonal bases.

For the remaining terms,

$$\mathrm{tr}(\mathcal{R}_4|\phi\rangle\langle\phi|^{\otimes2}\otimes I\otimes|\phi\rangle\langle\phi|)=\frac{1}{d}\sum_{P\in P_n}\left[\mathrm{tr}(P)\mathrm{tr}(P|\phi\rangle\langle\phi|)^3+\mathrm{tr}\left((P|\phi\rangle\langle\phi|)^2\right)\mathrm{tr}(P)\mathrm{tr}(P|\phi\rangle\langle\phi|)\right]=2, \tag{129}$$

$$\mathrm{tr}(\mathcal{R}_4I\otimes|\phi\rangle\langle\phi|\otimes I\otimes|\phi\rangle\langle\phi|)=\frac{1}{d}\sum_{P\in P_n}\left[\mathrm{tr}(P)^2\mathrm{tr}(P|\phi\rangle\langle\phi|)^2+\mathrm{tr}(|\phi\rangle\langle\phi|)\mathrm{tr}(P)\mathrm{tr}(P|\phi\rangle\langle\phi|)\right]=d+1, \tag{130}$$

where we use the fact that $\mathrm{tr}(P) = d\delta_{P,I}$. Thus, substituting these values into the equation for $V_*(O,\rho)$ results in

$$V_*(O,\rho) = \frac{d+1}{d+2}\frac{2^{1-M_2(|\phi\rangle)}d^2 - 3d + 1}{d^2} - \frac{(d-1)^2}{d^2(d+2)} = \frac{2(d+1)2^{-M_2(|\phi\rangle)} - 4}{d+2}. \tag{131}$$

This completes the proof, as substituted into Eq. (2). □

Then, we consider the fidelity estimation of a state with noise.

Proposition 22 *Suppose $\rho = (1-p)|\phi\rangle\langle\phi| + pI/d$ is any n-qubit state with noises and the observable $O = |\phi\rangle\langle\phi| - I/d$, then the variance of the thrifty shadow using the Haar random ensemble or a unitary 4-design reads*

$$\begin{aligned} V(O,\rho) = &\frac{1}{R}\left\{\left(\frac{d-1}{d}\right)^2\left[-p^2 + \frac{4d}{(d-1)(d+2)}p + \frac{d^2-3d-2}{(d-1)(d+2)}\right] + \frac{d^2-1}{d^2}\right\} \\ &+ \frac{R-1}{R}\left\{\frac{4\left(d^3-d^2+d+1\right)(1-p)^2}{d^2(d+2)(d+3)} - \frac{8d+8}{d^4(d+2)(d+3)}\right\}, \end{aligned} \tag{132}$$

Proposition 23 *Suppose $\rho = (1-p)|\phi\rangle\langle\phi| + pI/d$ is any n-qubit state with noises and the observable $O = |\phi\rangle\langle\phi| - I/d$, then the variance of the thrifty shadow using the Clifford group reads*

$$\begin{aligned} V(O,\rho) = &\frac{1}{R}\left\{\left(\frac{d-1}{d}\right)^2\left[-p^2 + \frac{4d}{(d-1)(d+2)}p + \frac{d^2-3d-2}{(d-1)(d+2)}\right] + \frac{d^2-1}{d^2}\right\} \\ &+ \frac{R-1}{R}\left\{\frac{d+1}{(d+2)d}\left[2d(1-p)^2 2^{-M_2|\phi\rangle} - p^2 + 6p - 3\right] - \frac{1}{d+2}\left(\frac{d-1}{d}\right)^2(1-p)^2\right\}, \end{aligned} \tag{133}$$

where $M_2(|\phi\rangle)$ is the stabilizer 2-Rényi entropy of state $|\phi\rangle$.

Proof. [Proof of Propositions 22 and 23] Here we prove the above two propositions together. We first calculate the variance of the classical shadow

$$V(O,\rho) = \left(\frac{d-1}{d}\right)^2\left[-p^2 + \frac{4d}{(d-1)(d+2)}p + \frac{d^2-3d-2}{(d-1)(d+2)}\right] + \frac{d^2-1}{d^2}. \tag{134}$$

Then, by virtue of Eq. (125), we calculate the following four terms:

$$\begin{aligned} \mathrm{tr}\left(\mathcal{R}_1 O\otimes\rho\otimes O\otimes\rho\right) &= \left[\frac{d-1}{d}(1-p)\right]^2, \\ \mathrm{tr}\left(\mathcal{R}_2 O\otimes\rho\otimes O\otimes\rho\right) &= \frac{4((d-3)(d-2)d-3)p^2}{d^3} + \frac{4(d-2)(d-1)(d+2)p}{d^3} + \frac{d^3+3d^2+4}{d^3}, \\ \mathrm{tr}\left(\mathcal{R}_2 O\otimes\rho\otimes O\otimes\rho\right) &= \frac{2(d-1)(2(d-2)d+3)p^2}{d^3} + \frac{4(d-2)(d-1)p}{d^3} + \frac{(d-1)(d+2)}{d^3}, \\ \mathrm{tr}\left(\mathcal{R}_4 O\otimes\rho\otimes O\otimes\rho\right) &= \frac{2(d+1)}{d+2}(1-p)^2 2^{-M_2(|\phi\rangle)} + \frac{d+1}{d(d+2)}(-p^2+6p-3). \end{aligned} \tag{135}$$

By combining these results, we can get the desired conclusions. □

Proof. [Proof of Theorem 7] In the proof of Propositions 3 and 7, we have already calculated that

$$\begin{aligned} \mathrm{tr}\left(\mathcal{R}_1 O\otimes\rho\otimes O\otimes\rho\right) &= \frac{(d-1)^2}{d^2}, \\ \mathrm{tr}\left(\mathcal{R}_2 O\otimes\rho\otimes O\otimes\rho\right) &= \frac{(d-1)(9d-8)}{d^2}, \\ \mathrm{tr}\left(\mathcal{R}_3 O\otimes\rho\otimes O\otimes\rho\right) &= \frac{(d-1)(4d-3)}{d^2}, \\ \mathrm{tr}\left(\mathcal{R}_4 O\otimes\rho\otimes O\otimes\rho\right) &= \frac{(d-1)(2d-1)}{d^2}. \end{aligned} \tag{136}$$

For $\mathcal{R}_5$, we calculate each $R(T)$ for $T \in \mathbb{G}_5$ separately.

$$\begin{cases} \mathrm{tr}\left((13)R(T_4)O\otimes\rho\otimes O\otimes\rho\right) = \dfrac{d-1}{d}, \\ \mathrm{tr}\left((23)R(T_4)O\otimes\rho\otimes O\otimes\rho\right) = \left(\dfrac{d-1}{d}\right)^2, \\ \mathrm{tr}\left((123)R(T_4)O\otimes\rho\otimes O\otimes\rho\right) = \left(\dfrac{d-1}{d}\right)^2, \\ \mathrm{tr}\left((132)R(T_4)O\otimes\rho\otimes O\otimes\rho\right) = \left(\dfrac{d-1}{d}\right)^2, \end{cases} \tag{137}$$

which gives the result that

$$\mathrm{tr}\left(\mathcal{R}_5 O\otimes\rho\otimes O\otimes\rho\right) = \frac{(d-1)(4d-3)}{d^2}. \tag{138}$$

Then, due to Lemma 20 and the definition of $V_R(O,\rho)$, one can derive the conclusion:

$$V(O,\rho) = \frac{1}{R}\frac{2(d-1)}{d+2} + \frac{R-1}{R}\frac{2\left(d^3(3/4)^k + 3d^2(1/2)^k + 2d^2 - 8\right)}{(d+2)^2(d+4)}. \tag{139}$$

□

The proof of Theorem 8 is a clear extension of Lemma 21, whose detail is not displayed. Here we give the exact form of the variance in Theorem 8:

$$\begin{aligned} V(O,\rho) = \frac{1}{R}\frac{2(d-1)}{d+2} + \frac{R-1}{R}\Bigg[&\frac{4(d-1)}{(d+2)(d+3)} + \frac{2(d-1)(d+1)}{(d+2)(d+3)}\left(1-\frac{d(d+3)}{4(d^2-1)}\right)^l \\ &-\frac{2(d-1)(d+1)(9d-8)}{3d^3(d+2)}\left(1-\frac{d^2}{4(d^2-1)}\right)^l\Bigg] \end{aligned} \tag{140}$$

E Stabilizer 2-Rényi entropies of various state families

E.1 W states

W states are a special case of Dicke states, expressed as

$$|W_n\rangle = \frac{1}{\sqrt{n}}(|10\ldots0\rangle + |01\ldots0\rangle + \cdots + |00\ldots1\rangle). \tag{141}$$

Thus, $|W_n\rangle^{\otimes 4}$ can be written as

$$|W_n\rangle^{\otimes 4} = \frac{1}{n^2}\sum_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\boldsymbol{d}\in\mathbb{S}\subset\{0,1\}^n} |a_1b_1c_1d_1\rangle\otimes|a_2b_2c_2d_2\rangle\otimes\cdots\otimes|a_nb_nc_nd_n\rangle. \tag{142}$$

By the definition of W states, for each vector $\boldsymbol{v}\in\mathbb{S}$, only one element equals 1.

According to Eq. (30) and the discussion in Appendix A, we can compute the stabilizer 2-Rényi entropy as

$$\begin{aligned} M_2(|W_n\rangle) &= -\log_2 \mathrm{tr}\left(R(T_4)|W_n\rangle\langle W_n|^{\otimes 4}\right) \\ &= -\log_2\frac{1}{n^4}\sum_{\boldsymbol{a},\boldsymbol{a}',\ldots,\boldsymbol{d}'\in\mathbb{S}} \langle a_1b_1c_1d_1|r(T_4)|a_1'b_1'c_1'd_1'\rangle\ldots\langle a_nb_nc_nd_n|r(T_4)|a_n'b_n'c_n'd_n'\rangle \\ &= -\log_2\frac{1}{n^4}\left[n^2 + \frac{n(n-1)}{2}\times 2\times 2\times 3\right] = \log_2\frac{n^3}{7n-6}. \end{aligned} \tag{143}$$

E.2 GHZ states

For GHZ states with a phase factor, it is equivalent to computing the stabilizer 2-Rényi entropy of a single qubit state:

$$|\psi_\theta\rangle = \frac{1}{2}\left(|0\rangle + e^{i\theta}|1\rangle\right). \tag{144}$$

The stabilizer 2-Rényi entropy can be calculated directly

$$M_2(|GHZ_\theta\rangle) = M_2(|\psi_\theta\rangle) = -\log_2\frac{\cos 4\theta + 7}{8}. \tag{145}$$

E.3 $|S_{n,k}(\theta)\rangle$

Given that the additivity of the stabilizer α-Rényi entropy and Eq. (144), we can derive the stabilizer 2-Rényi entropy of $|S_{n,k}(\theta)\rangle$ easily

$$M_2(|S_{n,k}(\theta)\rangle) = kM_2(|\psi_\theta\rangle) = -k\log_2\frac{\cos 4\theta + 7}{8}. \tag{146}$$

An efficient Julia framework for hierarchical equations of motion in open quantum systems

Yi-Te Huang[1 2] Po-Chen Kuo[1 2] Neill Lambert[3 *] Mauro Cirio[4 †] Simon Cross[3]
Shen-Liang Yang[1 2] Franco Nori[3 5 6] Yueh-Nan Chen[1 2 7 ‡]

[1] *Department of Physics, National Cheng Kung University, Tainan 701401, Taiwan*
[2] *Center for Quantum Frontiers of Research and Technology (QFort), Tainan 701401, Taiwan*
[3] *Theoretical Quantum Physics Laboratory, Cluster for Pioneering Research, RIKEN, Wakoshi, Saitama 351-0198, Japan*
[4] *Graduate School of China Academy of Engineering Physics, Haidian District, Beijing, 100193, China*
[5] *Center for Quantum Computing (RQC), RIKEN, Wakoshi, Saitama 351-0198, Japan*
[6] *Physics Department, The University of Michigan, Ann Arbor, Michigan 48109-1040, USA.*
[7] *Physics Division, National Center for Theoretical Sciences, Taipei 106319, Taiwan*

Abstract. The hierarchical equations of motion (HEOM) approach can describe the reduced dynamics of a system simultaneously coupled to multiple bosonic and fermionic environments. The complexity of exactly describing the system-environment interaction with the HEOM method usually results in time-consuming calculations and a large memory cost. Here, we introduce an open-source software package called HierarchicalEOM.jl: a Julia framework integrating the HEOM approach. HierarchicalEOM.jl features a collection of methods to compute bosonic and fermionic spectra, stationary states, and the full dynamics in the extended space of all auxiliary density operators (ADOs). The required handling of the ADOs multi-indexes is achieved through a user-friendly interface.

Keywords: Julia, open quantum system, hierarchical equations of motion, non-Markovianity

1 Introduction

The time evolution of a closed quantum system can offer important insights about its nature and properties. However, the dynamics are inevitable affected by interactions with external environments [1, 2], which can involve the exchange of energy or particles, and the suppression of quantum coherence. Due to the effective continuum of degrees of freedom present in these external baths [3, 4], modeling the dynamics of an open quantum system can be challenging. This is especially the case when perturbative approaches [5] are no longer valid due to non-Markovian effects emerging in the presence of strong interaction with the bath [6, 7]. In this regime, standard Markovian master equations are no longer applicable, and non-perturbative techniques are required [8–12].

In particular, here we consider the hierarchical equations of motion (HEOM) approach, which offers a non-perturbative [13] characterization of all the environmental effects on the system. This is achieved by using a hierarchy of auxiliary density operators (ADOs) to model system-bath correlations and entanglement [14]. This makes the HEOM suitable for studying complex systems strongly coupled to either bosonic [15–19] or fermionic [20–22] environments. Additionally, quantum systems interacting simultaneously with both bosonic and fermionic environments can be found in the study of electron transport through both natural and artificial molecules [23, 24]. Naturally, such an increase in the complexity of the environment leads to an increase in computational resources which, in the case of the HEOM method, corresponds to an increase in the size of the HEOM Liouvillian superoperator (HEOMLS) matrix. To deal with this issue, it is beneficial to explore the numerical efficiency of programming languages designed to optimize different computational resources.

`HierarchicalEOM.jl` is developed following the design philosophy of `Julia` programming language [25, 26]: one can have machine performance without sacrificing human convenience [25]. While integrating many of the features presented in other open-source HEOM packages [27–31], `HierarchicalEOM.jl` also includes other functionalities, such as the estimation of importance values for all ADOs, the calculation of spectra for both bosonic and fermionic systems, the construction of HEOMLS matrices for even- or odd-parity auxiliary density operators, and a user-friendly interface (which interrogates the ADOs multi-indexes) for gaining access to bath properties. By wrapping some functions from other `Julia` packages [32–34], we could further optimize the computations of the dynamics and stationary states for all ADOs.

2 Total Hamiltonian

Throughout this work, we consider an open quantum system (s) interacting with *fermionic* (f) and *bosonic* (b) *environments* described by the following total (T) *Hamiltonian* ($\hbar$ is set to unity throughout this work):

$$H_{\mathrm{T}} = H_{\mathrm{s}}(t) + H_{\mathrm{f}} + H_{\mathrm{b}} + H_{\mathrm{sf}} + H_{\mathrm{sb}}, \tag{1}$$

where $H_{\mathrm{s}}(t)$ is the (possibly time-dependent) system Hamiltonian containing boson and fermion particles. Here, we allow the fermionic environment to be composed

*nwlambert@gmail.com
†cirio.mauro@gmail.com
‡yuehnan@mail.ncku.edu.tw

by multiple baths of non-interacting fermionic degrees of freedom described by the Hamiltonian

$$H_{\mathrm{f}} = \sum_{\alpha} \sum_{k} \epsilon_{\alpha,k} c^{\dagger}_{\alpha,k} c_{\alpha,k}, \tag{2}$$

where $c_{\alpha,k}$ ($c^{\dagger}_{\alpha,k}$) annihilates (creates) a fermion (f) in the state k (with energy $\epsilon_{\alpha,k}$) of the αth fermionic bath. Analogously,

$$H_{\mathrm{b}} = \sum_{\beta} \sum_{k} \omega_{\beta,k} b^{\dagger}_{\beta,k} b_{\beta,k}, \tag{3}$$

describes a generic bosonic environment which can accommodate multiple non-interacting bosonic baths in which $b_{\beta,k}$ ($b^{\dagger}_{\beta,k}$) is the bosonic annihilation (creation) operator associated to the kth mode (with frequency $\omega_{\beta,k}$) in the βth bosonic bath. The interaction Hamiltonian between a fermionic system and the fermionic environments can be written as

$$H_{\mathrm{sf}} = \sum_{\alpha,k} \left(g_{\alpha,k} c^{\dagger}_{\alpha,k} d_{\mathrm{s}} + g^{*}_{\alpha,k} d^{\dagger}_{\mathrm{s}} c_{\alpha,k} \right), \tag{4}$$

in terms of the *coupling strengths* $g_{\alpha,k}$. Analogously, the interaction between a bosonic or fermionic system and the exterior bosonic environments can be modeled by

$$H_{\mathrm{sb}} = V_{\mathrm{s}} \sum_{\beta,k} g_{\beta,k} (b_{\beta,k} + b^{\dagger}_{\beta,k}), \tag{5}$$

in terms of the coupling strengths $g_{\beta,k}$. Here, d_{s} and V_{s} refer to the coupling operators acting on the system's degrees of freedom. In particular, d_{s} is a odd-parity operator destroying a fermion in the system, while V_{s} is in general a Hermitian operator which can act on both fermionic and bosonic systems. When V_{s} is acting on the fermionic system, it must have even-parity to be compatible with charge conservation. Furthermore, one can easily generalize to the case where the system contains multiple bosonic and fermionic quantum numbers (such as frequency, energy, or spin) interacting with either individual or shared environment(s).

We assume the following three conditions in HEOM approach: (1) The system and the environments (baths) are initialized in a separable state. (2) Each of the fermionic (bosonic) baths is initially in thermal equilibrium characterized by a Fermi-Dirac (Bose-Einstein) distribution. (3) The bath operator within the system-bath interaction Hamiltonian should be linear in the bath annihilation and creation operators, as shown in Eq. (4) and Eq. (5). In this case, the effects of fermionic and bosonic environments [initially in thermal equilibrium (eq) and linearly coupled to the system] are completely encoded in the *two-time correlation functions* $C(t_1, t_2)$ [35]. In the fermionic case, they depend on the spectral density $J_{\alpha}(\omega) = 2\pi \sum_k |g_{\alpha,k}|^2 \delta(\omega - \omega_k)$ and the Fermi–Dirac distribution $n^{\mathrm{eq}}_{\alpha}(\omega) = \{\exp[(\omega - \mu_{\alpha})/k_{\mathrm{B}} T_{\alpha}] + 1\}^{-1}$ as

$$C^{\nu}_{\alpha}(t_1, t_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega J_{\alpha}(\omega) \left[\frac{1-\nu}{2} + \nu n^{\mathrm{eq}}_{\alpha}(\omega) \right] e^{\nu i \omega (t_1 - t_2)}. \tag{6}$$

Analogously, in the bosonic case, they depend on the spectral density $J_{\beta}(\omega) = 2\pi \sum_k g^2_{\beta,k} \delta(\omega - \omega_k)$ and the Bose–Einstein distribution $n^{\mathrm{eq}}_{\beta}(\omega) = \{\exp[\omega/k_{\mathrm{B}} T_{\beta}] - 1\}^{-1}$ as

$$C_{\beta}(t_1, t_2) = \frac{1}{2\pi} \int_{0}^{\infty} d\omega J_{\beta}(\omega) \Big[n^{\mathrm{eq}}_{\beta}(\omega) e^{i\omega(t_1 - t_2)} + (n^{\mathrm{eq}}_{\beta}(\omega) + 1) e^{-i\omega(t_1 - t_2)} \Big]. \tag{7}$$

Here, k_{B} is the Boltzmann constant and T_{α} (T_{β}) represents the absolute temperature of the α-fermionic (β-bosonic) bath. A non-zero chemical potential ($\mu_{\alpha} \neq 0$) in the α-fermionic bath can account for non-equilibrium physics.

In Ref. [35], by expressing the bath correlation functions in Eqs. (6-7) as a sum of exponential terms (exponents), one can define an iterative procedure which leads to the celebrated hierarchical equations of motion. It can also be expressed as the HEOM *Liouvillian superoperator* (HEOMLS) matrix which characterizes the dynamics in the full *auxiliary density operators* (ADOs) space.

3 Package architecture

The package `HierarchicalEOM.jl` is designed to integrate the efficiency of `Julia` with the functionalities provided by other existing HEOM packages [27–31]. This leads to an intuitive interface to construct arbitrary Hamiltonians and initial states. We now introduce the ecosystem of the package as summarized in Fig. 1.

Following Fig. 1(a) and Fig. 1(b), users should specify the system Hamiltonian $H_{\mathrm{s}}(t)$, system coupling operators V_{s} (d_{s}) describing the interaction with bosonic (fermionic) baths, and the bath correlation functions. For the bath correlation functions in Eq. (6) and Eq. (7), users can specify them by an exponential series (list of exponents). `HierarchicalEOM.jl` provides built-in functions to construct these series from physical parameters (e.g., coupling strength, temperature, etc.). Alternatively, `HierarchicalEOM.jl` offers the possibility to manually define the correlation functions (simply supplying the list of exponents). Additional spectral densities could be incorporated into the built-in functions in future releases. We explicitly note that the package allows for any combination of fermionic and bosonic baths.

After the definition of the bath, users can further construct the HEOMLS matrix $\hat{\mathcal{M}}$ together with the system Hamiltonian $H_{\mathrm{s}}(t)$, see Fig. 1(c).

As shown in Fig. 1(d), with this information it is possible to proceed with solving for the dynamics of all the ADOs on either CPUs or GPUs. `HierarchicalEOM.jl` provides two distinct methods to compute the dynamics. The first one relies on `DifferentialEquations.jl` [32] which provides a set of low-level solvers for ordinary differential equations. The second method directly builds the propagator when the system Hamiltonian is time-independent: $\hat{\mathcal{G}}(t) = \exp(\hat{\mathcal{M}} t)$ using `FastExpm.jl` [34] which is optimized for the exponentiation of either large-dense or sparse matrices. In addition, `HierarchicalEOM.jl` can also directly

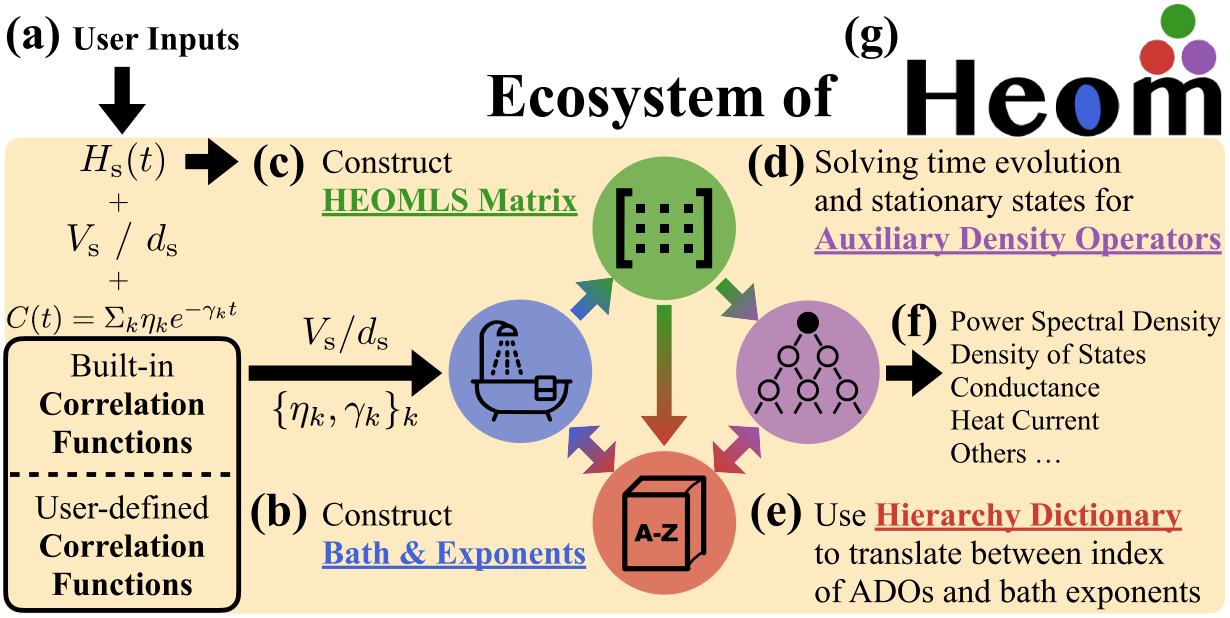


Figure 1: **(a)** Users should specify the system Hamiltonian $H_s(t)$, coupling operators (V_s or d_s), and the bath correlation function $C(t)$. For the exponent $\{\eta_k, \gamma_k\}$, users can either specify the physical parameters characterizing the spectral density of the bath by built-in functions, or directly providing a list of exponents. **(b)** Construction of the bath-object which includes the system coupling operator and a list of exponents characterizing the bath correlation function. **(c)** Construction of the HEOM Liouvillian superoperator (HEOMLS) matrix $\hat{\mathcal{M}}$ which defines the hierarchical equations of motion from the system Hamiltonian and the bath-objects. **(d)** Computation of the dynamics and stationary states for all auxiliary density operators using $\hat{\mathcal{M}}$. **(e)** The hierarchy dictionary translates the index of each ADO into the corresponding multi-index ensembles together with the exponents of the bath, and vice-versa. **(f)** The hierarchy dictionary allows a high-level interpretation of the ADOs to compute some physical properties. **(g)** Logo of `HierarchicalEOM.jl` package.

solve for the stationary states of all the ADOs using `LinearSolve.jl` [33], which offers a unified interface to solve linear equations in `Julia`.

To allow further analysis of specific physical properties, `HierarchicalEOM.jl` provides a hierarchy dictionary, as depicted in Fig. 1(e) and Fig. 1(f). This dictionary translates the index of each ADO in terms of the corresponding exponential terms of the bath, and vice-versa. This feature is designed to allow a high-level description of the ADOs, which can be useful in the analysis of electronic currents [20, 22], heat currents [18, 36, 37], and higher-order moments of heat currents [38]. Moreover, `HierarchicalEOM.jl` can calculate the spectrum for both bosonic and fermionic systems using `LinearSolve.jl`.

4 Conclusion

In conclusion, the `HierarchicalEOM.jl` software package provides a user-friendly and efficient tool for simulating complex open quantum systems, including non-Markovian effects due to non-perturbative interaction with one, or multiple, environments. It takes advantage of other available packages [27–31] for the following features:

- It supports different choices of spectral densities and spectral decomposition methods to accurately compute bath correlation functions.
- It supports time-dependent system Hamiltonians.
- It constructs the HEOMLS matrices for different types (bosonic, fermionic, or hybrid) of baths.
- The HEOMLS matrices are constructed using multi-threading.
- It provides different methods based on `DifferentialEquations.jl` [32], `LinearSolve.jl` [33], and `FastExpm.jl` [34] to compute the dynamics and stationary state of all ADOs with either CPUs or GPUs.

As a result, we believe that `HierarchicalEOM.jl` will be a valuable tool for researchers working in different fields such as quantum biology, quantum optics, quantum thermodynamics, quantum information, quantum transport, and condensed matter physics.

5 Code Availability

The `HierarchicalEOM.jl` package is available through a public GitHub repo (https://github.com/NCKU-QFort/HierarchicalEOM.jl). It is also registered in the `Julia` package registry and can be installed by the `Julia` package manager. Moreover, detailed information (documentation and other examples) is available through a public website (https://ncku-qfort.github.io/HierarchicalEOM.jl).

References

[1] R. Zwanzig, "Ensemble method in the theory of irreversibility", J. Chem. Phys. **33**, 1338 (1960).

[2] R. Feynman and F. Vernon, "The theory of a general quantum system interacting with a linear dissipative system", Ann. Phys. **24**, 118 (1963).

[3] A. Caldeira and A. Leggett, "Path integral approach to quantum Brownian motion", Physica A **121**, 587 (1983).

[4] P. Hedegård and A. O. Caldeira, "Quantum dynamics of a particle in a Fermionic environment", Phys. Scripta **35**, 609 (1987).

[5] H.-B. Chen, N. Lambert, Y.-C. Cheng, Y.-N. Chen, and F. Nori, "Using non-Markovian measures to evaluate quantum master equations for photosynthesis", Sci. Rep. **5**, 12753 (2015).

[6] Y. Tanimura and R. Kubo, "Time evolution of a quantum system in contact with a nearly Gaussian-Markoffian noise bath", J. Phys. Soc. Jpn. **58**, 101 (1989).

[7] Y. Tanimura, "Nonperturbative expansion method for a quantum system coupled to a harmonic-oscillator bath", Phys. Rev. A **41**, 6676 (1990).

[8] R. Bulla, T. A. Costi, and T. Pruschke, "Numerical renormalization group method for quantum impurity systems", Rev. Mod. Phys. **80**, 395 (2008).

[9] W.-M. Zhang, P.-Y. Lo, H.-N. Xiong, M. W.-Y. Tu, and F. Nori, "General non-Markovian dynamics of open quantum systems", Phys. Rev. Lett. **109**, 170402 (2012).

[10] P. Strasberg, G. Schaller, N. Lambert, and T. Brandes, "Nonequilibrium thermodynamics in the strong coupling and non-Markovian regime based on a reaction coordinate mapping", New J. Phys. **18**, 073007 (2016).

[11] M. Brenes, J. J. Mendoza-Arenas, A. Purkayastha, M. T. Mitchison, S. R. Clark, and J. Goold, "Tensor-network method to simulate strongly interacting quantum thermal machines", Phys. Rev. X **10**, 031040 (2020).

[12] J. K. Sowa, N. Lambert, T. Seideman, and E. M. Gauger, "Beyond Marcus theory and the Landauer–Büttiker approach in molecular junctions. II. A self-consistent Born approach", J. Chem. Phys. **152**, 064103 (2020).

[13] Z. Li, N. Tong, X. Zheng, D. Hou, J. Wei, J. Hu, and Y. Yan, "Hierarchical Liouville-space approach for accurate and universal characterization of quantum impurity systems", Phys. Rev. Lett. **109**, 266403 (2012).

[14] Y. Tanimura, "Numerically "exact" approach to open quantum dynamics: The hierarchical equations of motion (HEOM)", J. Chem. Phys. **153**, 020901 (2020).

[15] N. Lambert, S. Ahmed, M. Cirio, and F. Nori, "Modelling the ultra-strongly coupled spin-boson model with unphysical modes", Nat. Commun. **10**, 3721 (2019).

[16] T. P. Fay and D. T. Limmer, "Coupled charge and energy transfer dynamics in light harvesting complexes from a hybrid hierarchical equations of motion approach", J. Chem. Phys. **157**, 174104 (2022).

[17] J. Ma, Z. Sun, X. Wang, and F. Nori, "Entanglement dynamics of two qubits in a common bath", Phys. Rev. A **85**, 062323 (2012).

[18] A. Kato and Y. Tanimura, "Quantum heat current under non-perturbative and non-Markovian conditions: applications to heat machines", J. Chem. Phys. **145**, 224105 (2016).

[19] X.-Y. Chen, N.-N. Zhang, W.-T. He, X.-Y. Kong, M.-J. Tao, F.-G. Deng, Q. Ai, and G.-L. Long, "Global correlation and local information flows in controllable non-Markovian open quantum dynamics", npj Quantum Inf. **8**, 22 (2022).

[20] J. Jin, X. Zheng, and Y. Yan, "Exact dynamics of dissipative electronic systems and quantum transport: hierarchical equations of motion approach", J. Chem. Phys. **128**, 234703 (2008).

[21] A. Ishizaki and G. R. Fleming, "Unified treatment of quantum coherent and incoherent hopping dynamics in electronic energy transfer: reduced hierarchy equation approach", J. Chem. Phys. **130**, 234111 (2009).

[22] R. Härtle, G. Cohen, D. R. Reichman, and A. J. Millis, "Decoherence and lead-induced interdot coupling in nonequilibrium electron transport through interacting quantum dots: a hierarchical quantum master equation approach", Phys. Rev. B **88**, 235426 (2013).

[23] C. Schinabeck, R. Härtle, and M. Thoss, "Hierarchical quantum master equation approach to electronic-vibrational coupling in nonequilibrium transport through nanosystems: reservoir formulation and application to vibrational instabilities", Phys. Rev. B **97**, 235429 (2018).

[24] J. Bätge, Y. Ke, C. Kaspar, and M. Thoss, "Nonequilibrium open quantum systems with multiple bosonic and fermionic environments: a hierarchical equations of motion approach", Phys. Rev. B **103**, 235413 (2021).

[25] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman, "Julia: a fast dynamic language for technical computing", arXiv preprint arXiv:1209.5145 (2012).

[26] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: a fresh approach to numerical computing", SIAM Review **59**, 65 (2017).

[27] J. Strümpfer and K. Schulten, "Open quantum dynamics calculations with the hierarchy equations of motion on parallel computers", J. Chem. Theory Comput. **8**, 2808 (2012).

[28] L. Ye, X. Wang, D. Hou, R.-X. Xu, X. Zheng, and Y. Yan, "HEOM-quick: a program for accurate, efficient, and universal characterization of strongly correlated quantum impurity systems", WIREs Comput. Mol. Sci. **6**, 608 (2016).

[29]T. Kramer, M. Noack, A. Reinefeld, M. Rodríguez, and Y. Zelinskyy, "Efficient calculation of open quantum system dynamics and time-resolved spectroscopy with distributed memory HEOM (DM-HEOM)", J. Comput. Chem. **39**, 1779 (2018).

[30]T. Ikeda and G. D. Scholes, "Generalization of the hierarchical equations of motion theory for efficient calculations with arbitrary correlation functions", J. Chem. Phys. **152**, 204101 (2020).

[31]N. Lambert, T. Raheja, S. Cross, P. Menczel, S. Ahmed, A. Pitchford, D. Burgarth, and F. Nori, "QuTiP-BoFiN: a bosonic and fermionic numerical hierarchical-equations-of-motion library with applications in light-harvesting, quantum control, and single-molecule electronics", Phys. Rev. Res. **5**, 013181 (2023).

[32]C. Rackauckas and Q. Nie, "DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia", J. Open Res. Software **5**, 15 (2017).

[33]W. Kimmerer, V. Puri, and C. Rackauckas, *Linearsolve.jl*.

[34]H. Hogben, M. Krzystyniak, G. Charnock, P. Hore, and I. Kuprov, "Spinach – A software library for simulation of spin dynamics in large spin systems", J. Magn. Reson. **208**, 179 (2011).

[35]Y.-T. Huang, P.-C. Kuo, N. Lambert, M. Cirio, S. Cross, S.-L. Yang, F. Nori, and Y.-N. Chen, "An efficient Julia framework for hierarchical equations of motion in open quantum systems", Communications Physics **6**, 313 (2023).

[36]K. A. Velizhanin, H. Wang, and M. Thoss, "Heat transport through model molecular junctions: a multilayer multiconfiguration time-dependent hartree approach", Chem. Phys. Lett. **460**, 325 (2008).

[37]A. Kato and Y. Tanimura, "Quantum heat transport of a two-qubit system: interplay between system-bath coherence and qubit-qubit coherence", J. Chem. Phys. **143**, 064107 (2015).

[38]L. Song and Q. Shi, "Hierarchical equations of motion method applied to nonequilibrium heat transport in model molecular junctions: transient heat current and high-order moments of the current operator", Phys. Rev. B **95**, 064308 (2017).

Steering-enhanced quantum metrology using superpositions of noisy phase shifts

Kuan-Yi Lee[1] Jhen-Dong Lin[1] Adam Miranowicz[2 3] Franco Nori[2 4 5]
Huan-Yu Ku[6] Yueh-Nan Chen[1 2 8 *]

[1] *Department of Physics and Center for Quantum Frontiers of Research & Technology (QFort), National Cheng Kung University, Tainan 701, Taiwan*
[2] *Theoretical Quantum Physics Laboratory, Cluster for Pioneering Research, RIKEN, Wakoshi, Saitama 351-0198, Japan*
[3] *Institute of Spintronics and Quantum Information, Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland*
[5] *Center for Quantum Computing, RIKEN, Wakoshi, Saitama 351-0198, Japan*
[6] *Department of Physics, The University of Michigan, Ann Arbor, 48109-1040 Michigan, USA*
[7] *Department of Physics, National Taiwan Normal University, Taipei 11677, Taiwan*
[8] *Physics Division, National Center for Theoretical Sciences, Taipei 106319, Taiwan*

Abstract. Quantum steering, an essential concept in quantum information theory, has been shown to enhance quantum metrology [Nat. Commun. 12, 2410 (2021)]. In this work, we extend steering-enhanced metrology from single noiseless phase shifts to superpositions of noisy phase shifts. We examine a control system guiding a target through superpositions of dephased or depolarized phase shift channels. Our findings indicate that such superpositions can mitigate noise and enhance metrology. Proof-of-principle experiments conducted on the IBM Quantum Experience for dephased phase shifts demonstrate notable improvements, underscoring the practical benefits of this approach.

Keywords: Quantum Fisher information, Quantum correlations, Quantum metrology

1 Motivation

Recently, Ref. [1] showed that Reid's criterion [2] can be extended to the domain of quantum metrology, where Bob aims to estimate an unknown phase shift θ generated by a Hamiltonian H. An important result is that there exists a complementary relation between the variance of H and the precision of the θ estimation quantified by the quantum Fisher information (QFI). This complementary relation can be regarded as not only a metrological steering inequality (MSI), but also a generalized local uncertainty relation.

The metrological steering task has so far only been investigated under a noiseless scenario, where the phase shift is generated by a perfect unitary evolution. However, in a real experimental setup, the effects of noise are ubiquitous, such that the phase shifts could deviate from a perfect unitary and, thus, neutralize quantum advantages in metrology [3]. A typical source of noise comes from the inevitable interaction between a given system and its uncontrollable environments. A question arises on how to mitigate the effects of these undesired interactions [4, 5].

Recently, a novel approach, termed *superposition of quantum channels*, has been used to enhance quantum capacity in communication tasks [6–8]. In this framework, multiple quantum channels can be used. Furthermore, an additional quantum control was introduced to determine which channel for the target system to pass through. Hence, when the control system is prepared in a superposition state, the target system can go through these channels in a quantum-superposed manner. One can take advantage of the quantum interference between these channels to alleviate the effects of noise [9, 10].

In this work, we consider the cases where the phase shifts are distorted by either pure dephasing noise or depolarizing noise. In this sense, we denote the corresponding noise-distorted phase shifts as dephased and depolarized phase shifts, respectively. Intuitively, the enhancement of the estimation precision decreases when the noise strength increases. Furthermore, we investigate the influences of a superposition of both dephased and depolarized phase shifts by comparing different (coherent and incoherent) states of the control system. We show that the control system in a coherent state can mitigate the noise and enhance the violation of the MSI. Finally, we experimentally implemented a metrological steering task with a superposition of dephased phase shifts on the IBM Quantum (IBM Q) Experience. Our experimental results clearly show that the enhancement of the MSI violation is due to the initial coherence of the control system. We also provide noise simulations that take into account the inherent errors of the IBM Q device.

2 A Metrological steering task

We start by formulating the noiseless metrological task, where the phase shift θ is generated by a unitary channel $\exp(-iH\theta)$, with a "generating" Hamiltonian H. We consider a bipartite state ρ_{AB} shared by Alice and Bob. In each round of the experiment, Alice performs a measurement labeled by A. The probability to obtain the result a is denoted as $p(a|A)$; and the conditional

*yuehnan@mail.ncku.edu.tw

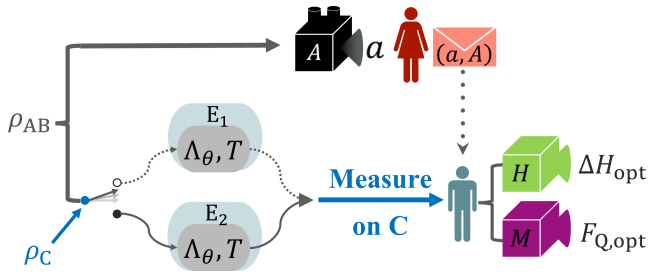


Figure 1: Illustration of steering-enhanced quantum metrology with a superposition of quantum channels.

reduced state of Bob's subsystem is $\rho_{B,a|A}$. After generating a local phase shift θ, Bob's conditional reduced state becomes $\rho_{B,a|A}(\theta) = \exp(-iH\theta)\rho_{B,a|A}\exp(iH\theta)$.

As reported in Ref. [1], when an assemblage is unsteerable, the MSI can be derived as $F_{Q,opt} \leq 4\Delta H_{opt}$. Here, we define the violation V of the MSI, i.e.,

$$V := \max\left(F_{Q,opt} - 4\Delta H_{opt},\ 0\right). \tag{1}$$

Therefore, $V > 0$ implies that the assemblage is steerable.

3 A Superposition of noisy phase shifts

We now consider a scenario for superposing two identical noisy phase shifts, as shown in Fig. 1. A control system C to determine which environment (i.e., E_1 or E_2,) is introduced, affecting the system B. The total system is initially prepared in $\rho_{tot} = |j\rangle\langle j|_C \otimes \rho \otimes \mathcal{E}_{E_1} \otimes \mathcal{E}_{E_2}$ for j being either 0 or 1. In this case, the total evolution can be described by

$$U_{tot} = |0\rangle\langle 0|_C \otimes\ U_{BE_1}(\theta) + |1\rangle\langle 1|_C \otimes\ U_{BE_2}(\theta). \tag{2}$$

In other words, when C is prepared in the state $|j\rangle$, B interacts with the corresponding environment E_j. Thus, if C is prepared in an incoherent mixed state, i.e., $(|0\rangle\langle 0|_C + |1\rangle\langle 1|_C)/2$, the system B has equal probabilities to interact with either E_0 or E_1. For simplicity, we consider that $U_{BE_1}(\theta)$ and $U_{BE_2}(\theta)$ are isomorphic to each other (so $\mathcal{E}_{E_1} = \mathcal{E}_{E_2}$); that is, two phase shifts are implemented in the same way.

On the other hand, when the control C is prepared in $|+\rangle_C = (|0\rangle_C + |1\rangle_C)/\sqrt{2}$, we obtain

$$\rho_{CB}(\theta) = \frac{\mathbb{1}_C}{2}\otimes\Lambda_\theta(\rho) + \frac{(|0\rangle\langle 1|_C + |1\rangle\langle 0|_C)}{2}\otimes T\rho\, T^\dagger, \tag{3}$$

where $T = \mathrm{Tr}_E\left[U_{BE}\ (\mathbb{1}\otimes\mathcal{E})\right]$ characterizes the quantum interference effect between these two channels [7]. The interference effect occurs simultaneously with the non-zero off-diagonal terms in C. In this case, the target passes through a "superposition of noisy phase shift channels".

4 Implementation on the IBM Cairo

To further decrease the circuit depth, we consider a scenario known as temporal steering [11, 12]. Therein, the initial maximally entangled state shared by Alice and Bob can be replaced by a prepare-and-measure scenario [13, 14]. As shown in Fig. 2, we provide a circuit model to experimentally implement the metrological steering task with the superposition of dephased phase shifts, which involves four qubits: the control C, the system B, E_1, and E_2, respectively.

Now, we perform a set of projective measurements, $\{|+\rangle\langle+|_C, |-\rangle\langle-|_C\}$, with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$, on the quantum control C with probabilities $P_\pm = \mathrm{Tr}\left[(|\pm\rangle\langle\pm|_C \otimes \mathbb{1}_B)\ \rho_{CB}(\theta)\right]$ are the probabilities of the outcomes $\pm$ for the projective measurements.

The main result of this paper that: the superposition of phase shifts can enhance the violation of the MSI. To highlight this point, we compare the two cases: (1) control C is prepared in an incoherent mixed state $\mathbb{1}_C/2$ (*without* a superposition of phase shifts); (2) control C is in a superposition $|+\rangle\langle+|_C$ state (*with* a superposition of phase shifts). We show that case (1), in general, cannot improve the violation of MSI; nevertheless, for case (2), it is possible to observe an enhancement of the MSI violation under the dephased phase shifts characterized by the following system-environment unitary evolution:

$$\begin{aligned} U_w^{deph}|\psi\rangle\otimes|0\rangle_E = &\sqrt{1-\frac{w}{2}}\ |\psi_\theta\rangle\otimes|0\rangle_E \\ &+\sqrt{\frac{w}{2}}\ \sigma_z|\psi_\theta\rangle\otimes|1\rangle_E, \end{aligned} \tag{4}$$

where $|\psi_\theta\rangle = \exp(-iZ\theta)|\psi\rangle$, and w is the visibility for the dephased phase shift.

To further discuss the coherent-control-enhanced violation of the MSI, we consider the average optimal FI (the lower bound of QFI) and variance by taking into account their probabilities [15], namely:

$$F_{opt}^{avg} := \sum_\pm P_\pm F_{opt,\pm}; \quad \Delta H_{opt}^{avg} := \sum_\pm P_\pm \Delta H_{opt,\pm}. \tag{5}$$

This circuit can be divided into three parts: (i) state preparation, (ii) the superposition of dephased phase shifts, and (iii) measurement on the qubits C and B. In part (i), the qubits C, B, and $E_{1,(2)}$ are prepared in the states $|+\rangle\langle+|$, $\rho_{a|A}$, and $|0\rangle\langle 0|_{1,(2)}$, respectively. In part (ii), the circuit model of the superposition of dephased phase shifts is shown in Fig. 2(a). The qubit topology of the four qubits that we chose in IBM-Cairo is shown in Fig. 2(b). Through the control qubit C, the system B can interact with alternative environments. We divide the total unitary in Fig. 2(a) into a gate sequence, which is shown in Fig. 2(c). In this sequence, we use control-rotation with angle ϕ on the system B and its corresponding environment such that $\phi = 2\sin^{-1}(\sqrt{w/2})$, with $\phi \in [0, \pi/2]$. In part (iii), we measure σ_x on qubit C and measure σ_z or σ_y on qubit B.

As shown in Fig. 3, the experimental result with the dephased phase shifts: the control C is prepared in (a) $\rho_C = \mathbb{1}_C/2$ (*without* a superposition of phase shifts) and (b) $\rho_C = |+\rangle\langle+|_C$ (*with* a superposition of phase shifts). Specifically, the red-cross (blue-circle) data points are the experimental results of the average optimal Fisher infor-

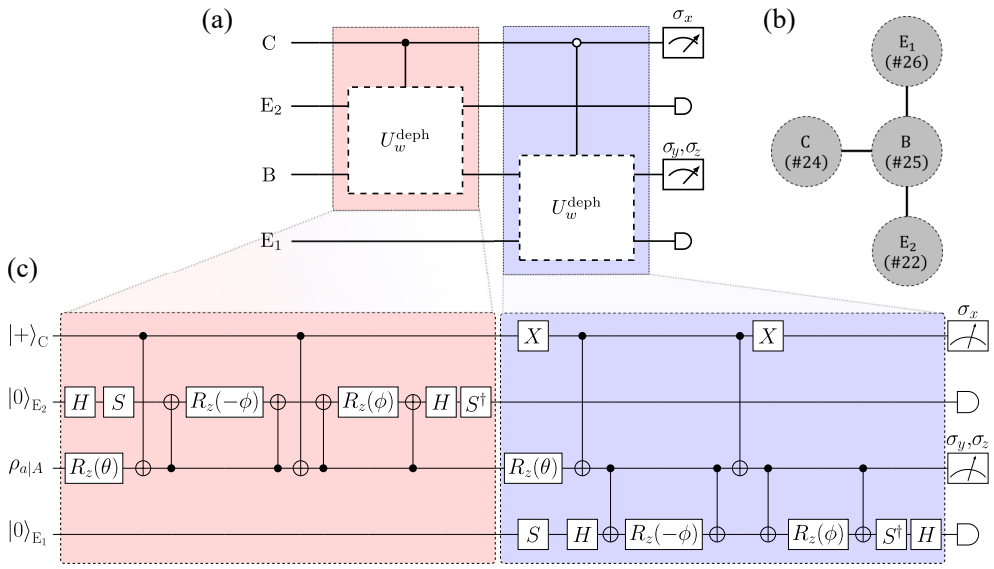


Figure 2: Circuit model for steering-enhanced quantum metrology with a superposition of dephased phase shifts.

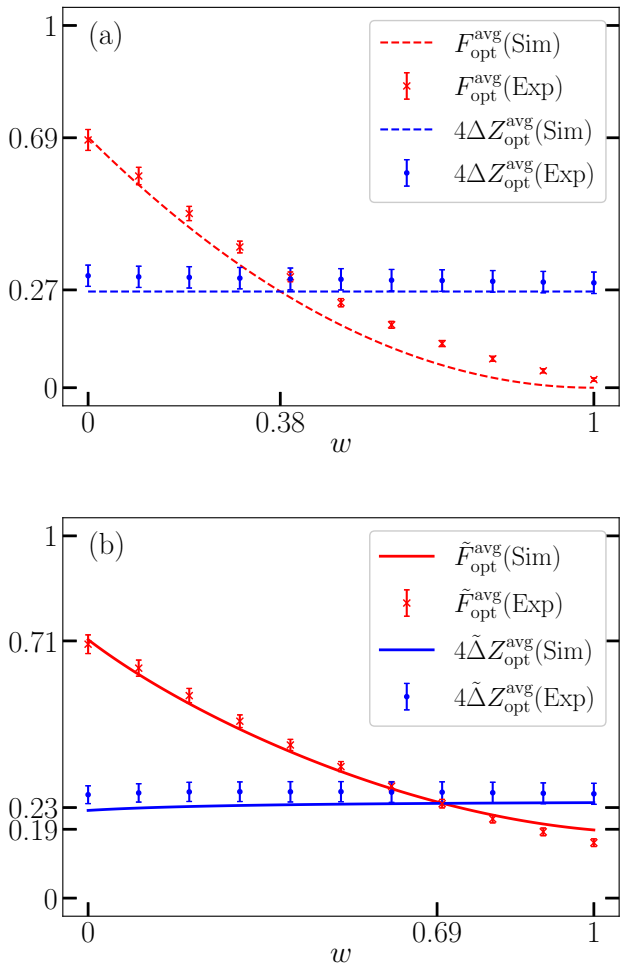


Figure 3: Experimental results and noise simulations (including qubit relaxation and qubit dephasing, gate error, and readout error [16]) of the metrological tasks.

mation (the optimal variance) with respect to the visibility w. Note that $\theta = 0$. the error bars are obtained from 40 individual rounds of experiments; each experimental data point consists of 10,000 individual runs performed on about ten different dates. Therefore, the error bars represent the variance of the IBM-Cairo device to conduct these experiments in long timescales. The solid curves represent the noise simulations for the dephased phase shifts with $\rho_{\rm C} = |+\rangle\langle+|_{\rm C}$; the dashed curves represent the noise simulations with $\rho_{\rm C} = \mathbb{1}_{\rm C}/2$. Although we consider only several common noisy resources [16], the tendencies and magnitudes of noise simulations approach the actual experiment in both cases (a) and (b). We clearly observe that the control in a superposition state, i.e., $|+\rangle\langle+|_{\rm C}$, can enhance the optimal Fisher information and decrease the optimal variance; thus, it extends the violations of the metrological steering inequality from $w \approx 0.38$ to $w \approx 0.69$.

5 Summary

In this paper, we generalize the metrological steering task described in Ref. [1] to a scenario with superpositions of noisy phase shifts. We show that the control in $|+\rangle\langle+|_{\rm C}$ (i.e., via a superposition of dephased and depolarized phase shifts) can alleviate the noisy effect and enhance the average violations of the MSI in comparison with the case where the control is in an incoherent mixed state (i.e., without superposition of dephased and depolarized phase shifts).

Moreover, we proposed a circuit model for superposing two dephased phase shifts and experimentally implemented the circuit on the IBM Quantum Experience. We clearly observe the violations of the MSI, and the experimental results agree with our noise simulations.

Finally, it is known that the order of channels can also be coherently controlled [17, 18]. Therefore, it would be promising to apply this framework to the noisy metrological steering task.

References

[1] B. Yadin, M. Fadel, and M. Gessner, "Metrological complementarity reveals the Einstein-Podolsky-Rosen paradox", Nat. Commun. **12**, 2410 (2021).

[2] M. D. Reid, P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen, and G. Leuchs, "Colloquium: the Einstein-Podolsky-Rosen paradox: from concepts to applications", Rev. Mod. Phys. **81**, 1727–1751 (2009).

[3] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum metrology", Phys. Rev. Lett. **96**, 010401 (2006).

[4] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, "Learning-based quantum error mitigation", PRX Quantum **2**, 040330 (2021).

[5] B. Regula and R. Takagi, "Fundamental limitations on distillation of quantum channel resources", Nat. Commun. **12**, 4411 (2021).

[6] G. Chiribella and H. Kristjánsson, "Quantum Shannon theory with superpositions of trajectories", Proc. R. Soc. A: Math. Phys. Eng. Sci. **475**, 20180903 (2019).

[7] A. A. Abbott, J. Wechs, D. Horsman, M. Mhalla, and C. Branciard, "Communication through coherent control of quantum channels", Quantum **4**, 333 (2020).

[8] K. Goswami, Y. Cao, G. A. Paz-Silva, J. Romero, and A. G. White, "Increasing communication capacity via superposition of order", Phys. Rev. Res. **2**, 033292 (2020).

[9] D. K. L. Oi, "Interference of quantum channels", Phys. Rev. Lett. **91**, 067902 (2003).

[10] G. Rubino et al., "Experimental quantum communication enhancement by superposing trajectories", Phys. Rev. Res. **3**, 013093 (2021).

[11] Y.-N. Chen, C.-M. Li, N. Lambert, S.-L. Chen, Y. Ota, G.-Y. Chen, and F. Nori, "Temporal steering inequality", Phys. Rev. A **89**, 032112 (2014).

[12] K. Bartkiewicz, A. Černoch, K. Lemr, A. Miranowicz, and F. Nori, "Temporal steering and security of quantum key distribution with mutually unbiased bases against individual attacks", Phys. Rev. A **93**, 062345 (2016).

[13] C.-M. Li, Y.-N. Chen, N. Lambert, C.-Y. Chiu, and F. Nori, "Certifying single-system steering for quantum-information processing", Phys. Rev. A **92**, 062310 (2015).

[14] A. Tavakoli, J. Pauwels, E. Woodhead, and S. Pironio, "Correlations in entanglement-assisted prepare-and-measure scenarios", PRX Quantum **2**, 040357 (2021).

[15] D. R. M. Arvidsson-Shukur, N. Y. Halpern, H. V. Lepage, A. A. Lasek, C. H. W. Barnes, and S. Lloyd, "Quantum advantage in postselected metrology", Nat. Commun. **11**, 3775 (2020).

[16] K.-Y. Lee, J.-D. Lin, A. Miranowicz, F. Nori, H.-Y. Ku, and Y.-N. Chen, "Steering-enhanced quantum metrology using superpositions of noisy phase shifts", Phys. Rev. Res. **5**, 013103 (2023).

[17] D. Ebler, S. Salek, and G. Chiribella, "Enhanced communication with the assistance of indefinite causal order", Phys. Rev. Lett. **120**, 120502 (2018).

[18] J. Barrett, R. Lorenz, and O. Oreshkov, "Cyclic quantum causal models", Nat. Commun. **12**, 885 (2021).

Robust Error Mitigation for Physical and Algorithmic Errors by Trotter Subspace Expansion in a Hamiltonian Simulation

Shigeo Hakkaku[1] * Yuuki Tokunaga[1] Suguru Endo[1 2] †

[1] *NTT Computer and Data Science Laboratories, NTT Corporation, 3-9-11 Midori-cho, Musashino, Tokyo 180-8585, Japan*
[2] *JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan*

Abstract. Quantum dynamics simulation via Hamilton simulation algorithms is one of the most crucial applications in the quantum computing field. While this task has been relatively considered the target in the fault-tolerance era, the experiment for demonstrating utility by an IBM team simulates the dynamics of an Ising-type quantum system with the Trotter-based Hamiltonian simulation algorithm with the help of quantum error mitigation. In this study, we propose the Trotter subspace expansion method to mitigate not only physical errors but also algorithmic errors of Trotterized quantum circuits in both the near-term and early fault-tolerant eras. This method inherits the advantages of the two existing quantum error mitigation methods: two-dimensional error extrapolation, which considers both physical and Trotter error, and virtual distillation. Using the 1D transverse-field Ising model, we numerically demonstrate that we can suppress both physical and algorithmic errors.

Keywords: quantum computation, quantum simulation, quantum algorithm, quantum error mitigation, Trottrization, early-fault tolerant quantum computation

1 Overview

One of the most promising applications of a fault-tolerant quantum computer is simulating the dynamics of many-body problems, such as condensed matter physics and quantum chemistry [1]. Hamiltonian simulation algorithms have been mainly discussed in the regime of fault tolerance, and numerous improvements have been made in this area. Although most such algorithms are beyond the capacity of the current quantum computers, the Trotter-based Hamiltonian simulation algorithm has already been performed in the current quantum device by an IBM team, with the utility of quantum devices being investigated [2].

In the above experiment, the error-extrapolation quantum error mitigation (QEM) [3–5] has played a crucial role in improving the accuracy of the simulation result by suppressing the physical errors due to coupling to the environment. Besides the near-term application, it has been pointed out that QEM can significantly improve the computation accuracy of quantum algorithms in the early fault-tolerant quantum computing (FTQC) era and can contribute to reducing resources such as code distances and T gate counts.

While the primary target of QEM has been the suppression of physical errors, the computation accuracy of quantum algorithms is also restricted by algorithmic errors because of the insufficiency of the circuit depth. Ref. [6] elaborated a QEM method tailored to noisy Trotterized quantum circuits to suppress both the algorithmic and physical errors. This method uses error-extrapolation QEM methods to mitigate physical error rates and then applies the algorithmic error extrapolation method for Trotter-based quantum simulation by leveraging the outcome for the fewer Trotter step number. Although this method does not impose any additional hardware requirements and may enhance computation accuracy even in current quantum devices, the heuristic nature of error-extrapolation QEM methods does not guarantee that the result corresponds to the physical quantum state. It may increase the bias of the estimation.

In this work, we propose the Trotter subspace expansion, an even more robust QEM for Trotter-based quantum algorithms. This QEM method also ensures that the error-mitigated computation outcome leads to the physical one. We construct the error-mitigated state $\rho_{\mathrm{QEM}} = \frac{\rho_{\mathrm{TS}}^2}{\mathrm{Tr}[\rho_{\mathrm{TS}}^2]}$ for the effective state $\rho_{\mathrm{TS}} = \sum_i c_i \rho(p_i, M_i)$ ($c_i \in \mathbb{R}$) based on the purification-based QEM methods. Here, $\rho(p_i, M_i)$ are noisy states with p_i and M_i being the different physical error rates and Trotter step numbers. We find the optimal coefficients c_i with two-dimensional hypersurface extrapolation, which gives a more accurate result than sequentially applying extrapolation for physical and algorithmic errors. Then, purification further filters out the residual noise and results in the outcome for the physical density operator, i.e., positive-semidefinite operators.

To benchmark the Trotter subspace expansion, we numerically compare it with the existing error mitigation method proposed in Ref. [6] when the shot noise is free. We find that the Trotter subspace expansion can mitigate more bias than the existing error mitigation methods from the numerical simulation with a one-dimensional transverse-filed Ising model with 10 qubits without the shot noise effect.

2 Preliminaries

In this section, we review the Trotterization and then how to suppress the algorithmic error introduced by the Trotterization according to Ref. [6]. The Trotterization approximates an exponential of a sum of operators, such as a time evolution operator, by a product of elementary exponentials. In the following, we explain the Trotterization using a time evolution operator as an example. Let $H = \sum_j H_j$, where each H_j acts on a constant number of qubits. The time evolution operator e^{-iHt} can be approximated by a product of unitaries e^{-it/MH_j}, i.e., $\exp(-iHt) = \left[\prod_j \exp(-i\frac{t}{M}H_j)\right]^M + \mathcal{O}\left(\frac{t^2}{M}\right)$, where M is the Trotter step number and $\mathcal{O}\left(\frac{t^2}{M}\right)$ is an algorithmic error. Therefore, M determines the level of al-

* shigeo.hakkaku@ntt.com
† suguru.endo@ntt.com

gorithmic error. If quantum circuits are error-free, the algorithmic error can be arbitrarily suppressed as the Trotter step number increases. Denoting $\epsilon_M = 1/M$ and U_{ϵ_M} as $U_{\epsilon_M}(t) := \left[\prod_j \exp(-i\epsilon_M t H_j)\right]^{1/\epsilon_M}$, then we can obtain $\lim_{\epsilon_M \to +0} U_{\epsilon_M}(t) = \exp(-iHt)$.

However, we cannot increase M infinitely because current quantum devices suffer from quantum noise. In such a device, a number of Trotter step number introduce more physical errors, and thus, there exists the optimal number of Trotter step number M_{opt} [7]. Thus, we cannot reduce the algorithmic error any further in a naive way.

To alleviate this issue, Ref. [6] elaborated the method that suppresses the algorithmic error by regarding an inverse of the Trotter step number as an ''error rate'' in the standard polynomial error extrapolation method [3–5]. Suppose that one applies the unitary operator $U_{\epsilon_M}(t)$ to the state $|\psi\rangle$, then the expectation value of an observable A is given by

$$\langle A(t)\rangle(\epsilon_M) = \left\langle \psi \middle| U^\dagger_{\epsilon_M}(t) A U_{\epsilon_M}(t) \middle| \psi \right\rangle.$$

Expanding $\langle A(t)\rangle(\epsilon_M)$ as a function of ϵ_M gives

$$\langle A(t)\rangle(\epsilon_M) = \langle A(t)\rangle(0) + \sum_{i=1}^{n'} A(t)_i \epsilon_M^i + \mathcal{O}\left(\epsilon_M^{n'+1}\right), \quad (1)$$

where $\langle A(t)\rangle(0) = \langle\psi|\exp(iHt)A\exp(-iHt)|\psi\rangle$ i.e., the algorithmic-error-free expectation value of A.

From Eq. (1), we can confirm that the error extrapolation method can be applied with algorithmic errors by reducing the Trotter step number. Normally, we increase the Trotter step up to M_{opt}. Combining the set of data points with different Trotter step number $\left\{\langle A(t)\rangle\left(\epsilon_{M_i}\right)\right\}_{i=0}^{n'}$, where $M_{n'} \leq \cdots M_1 \leq M_0 = M_{\text{opt}}$, with Eq. (1), we obtain

$$\begin{aligned}\langle A\rangle_{\text{est}}(0) &= \sum_{i=0}^{n'} \langle A\rangle\left(\epsilon_{M_i}\right) \prod_{k\neq i} \frac{\epsilon_{M_k}}{\epsilon_{M_k} - \epsilon_{M_i}} \\ &= \langle A\rangle(0) + \mathcal{O}\left(\epsilon_{M_{\text{opt}}}^{n'+1}\right).\end{aligned}$$

Thus, we see that the algorithmic error could be suppressed to $\mathcal{O}\left(\epsilon_{M_{\text{opt}}}^{n'+1}\right) = \mathcal{O}\left(M_{\text{opt}}^{-(n'+1)}\right)$.

The authors of Ref. [6] have numerically confirmed that the above method suppresses the algorithmic error if the physical errors are well mitigated in advance by a physical error extrapolation method [3–5]. However, as in the error extrapolation for physical errors, algorithmic error extrapolation is a heuristic method, and the physicality of the result is not generally guaranteed.

3 Trotter Subspace Expansion

Here, we propose Trotter subspace expansion that robustly mitigates both algorithmic and physical errors and certifies the physicality of the result. We prepare the $n'+1$ output states of noisy Trotterized circuit $\sigma_i = \rho(p_i, M_i)(i = 0, \ldots, n')$, where p_i and M_i are the error rates of physical noise p and the Trotter step number of the noisy Trotterized circuit, respectively.

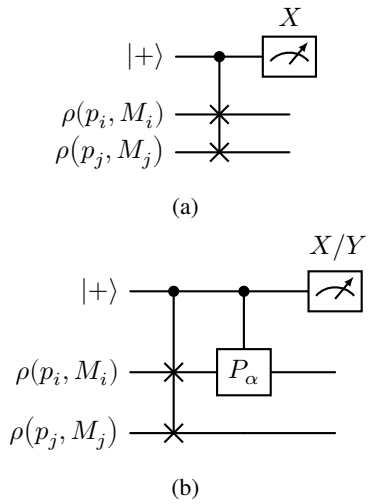


Figure 1: Quantum circuits for evaluating the expectation value of an observable A for the ansatz state of the Trotter subspace ρ_{QEM} shown in Eq. (2). (a): Swap test circuit to evaluate $\text{Tr}\left(\rho(p_i, M_i)\rho(p_j, M_j)\right)$ by measuring the Pauli X of the most upper line. (b): Quanutm circuit to evaluate $\text{Tr}\left(\rho(p_i, M_i)\rho(p_j, M_j)P_\alpha\right)$ by measuring the Pauli X and Pauli Y, where $A = \sum_\alpha c_\alpha P_\alpha$ and P_α is a Pauli operator.

Then, we virtually construct the following state:

$$\begin{aligned}\rho_{\text{TS}} &= \sum_i c_i \rho(p_i, M_i) \\ \rho_{\text{QEM}} &= \frac{\rho_{\text{TS}}^2}{\text{Tr}(\rho_{\text{TS}})}\end{aligned} \quad (2)$$

where c_i is the coefficient of the multidimensional extrapolation [8, 9] parametrized by the physical error p_i and the inverse of the Trotter number $1/M_i$ and is calculated by classical computers. The ansatz state ρ_{QEM} satisfies $\text{Tr}\left(\rho_{\text{QEM}}\right) = 1$ and $\rho \geq 0$, which ensures that ρ_{QEM} is physical. Then, the corresponding error-mitigated expectation value of A is given by

$$\langle A\rangle_{\text{est}} = \frac{\sum_{i,j=0}^{n'} e_i e_j \text{Tr}\left(\rho(p_i, M_i)\rho(p_j, M_j)A\right)}{\sum_{i,j=0}^{n'} e_i e_j \text{Tr}\left(\rho(p_i, M_i)\rho(p_j, M_j)\right)}. \quad (3)$$

To obtain each term constituting $\langle A\rangle_{\text{est}}$, i.e., $\text{Tr}\left(\rho(p_i, M_i)\rho(p_j, M_j)A\right)$ and $\text{Tr}\left(\rho(p_i, M_i)\rho(p_j, M_j)\right)$ in Eq. (3), we use a modified quantum circuit of purification-based QEM methods [10, 11]. The quantum circuits for Trotter subspace expansion for copy-based purification are shown in Fig. 1 [12].

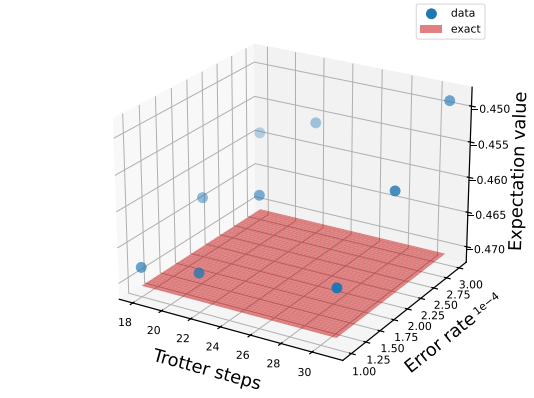
To benchmark the Trotter subspace expansion, we present numerical simulations using the one-dimensional transverse Ising model. We approximate the time evolution under this Hamiltonian by the 1st-order Trotterization. We assume there is single- or two-qubit depolarizing noise after a single- or two-qubit gate in a Trotterized circuit, respectively. Using the data points shown in Fig. 2a, we plot the estimation bias between each error mitigation method and an exact expectation value of the Pauli operator X_0 without the shot noise effect in Fig. 2b. From Fig. 2b, we confirm that the bias of the Trotter subspace

expansion is about 10 times smaller than those of the existing methods.

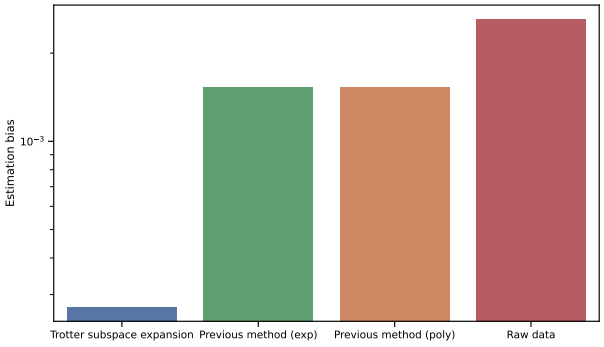
4 Conclusion

We have presented an error mitigation method for noisy Trotterized quantum circuits that suppresses both physical and algorithmic errors. We call the method the Trotter subspace expansion. The Trotter subspace expansion uses the ansatz constructed with the multi-dimensional extrapolation, where physical and algorithmic errors are parametrized, as well as the VD. We have numerically confirmed that the Trotter subspace expansion well mitigates both physical and algorithmic errors without shot noise, and it has decreased more bias than the proposed method [6].

Our work leaves several open questions. Although we have considered the first-order deterministic Trotterization, it can be extended to the higher-order Trotterization or randomized Trotterization. In particular, the extension to the qDRIFT [13] would be interesting for investigating the dynamics of electronic structure Hamiltonians. Recently, post-Trotter methods, such as truncated Taylor series [14–16] and quantization [17], have been paid much attention. This is because such methods exponentially improve as a function of the desired accuracy compared to the Trotter ones [1]. We also expect that one could extend our work or invention for the above post-Trotter methods. Although we only consider the dynamics of a Hamiltonian of condensed-matter physics as a practical case in our paper, it would be interesting to apply our proposed method to the statistical phase estimation [18], which is one of the most promising algorithms for the early-fault tolerant era.



(a)



(b)

Figure 2: Benchmark of Trotter subspace expansion and the previous method proposed in Ref. [6]. (a): Expectation values of X_0 obtained by noisy Trotterized circuits with different physical error rates and Trotter steps. These data points are used for the quantum error mitigation methods. The blue makers indicate the expectation values, and the red plane indicates the exact expectation value. (b): Comparison of the biases in the error-mitigated expectation values. The red bar represents the estimation bias of the data point whose absolute error is the smallest. The yellow and green bars represent the biases of the previous methods proposed in Ref. [6] using the polynomial and the exponential physical error extrapolation, respectively. The blue bar represents the bias of the Trotter subspace expansion.

References

[1] N. Yoshioka, T. Okubo, Y. Suzuki, Y. Koizumi, and W. Mizukami, ''Hunting for quantum-classical crossover in condensed matter problems'', npj Quantum Inf **10**, 1–10 (2024).

[2] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, ''Evidence for the utility of quantum computing before fault tolerance'', Nature **618**, 500–505 (2023).

[3] K. Temme, S. Bravyi, and J. M. Gambetta, ''Error Mitigation for Short-Depth Quantum Circuits'', Phys. Rev. Lett. **119**, 180509 (2017).

[4] Y. Li and S. C. Benjamin, ''Efficient Variational Quantum Simulator Incorporating Active Error Minimization'', Phys. Rev. X **7**, 021050 (2017).

[5] S. Endo, S. C. Benjamin, and Y. Li, ''Practical Quantum Error Mitigation for Near-Future Applications'', Phys. Rev. X **8**, 031027 (2018).

[6] S. Endo, Q. Zhao, Y. Li, S. Benjamin, and X. Yuan, ''Mitigating algorithmic errors in a Hamiltonian simulation'', Phys. Rev. A **99**, 012334 (2019).

[7] G. C. Knee and W. J. Munro, ''Optimal Trotterization in universal quantum simulators under faulty control'', Physical Review A **91**, 10.1103/PhysRevA.91.052327 (2015).

[8] M. Otten and S. K. Gray, ''Recovering noise-free quantum observables'', Phys. Rev. A **99**, 012338 (2019).

[9] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, ''Hybrid Quantum-Classical Algorithms and Quantum Error Mitigation'', J. Phys. Soc. Jpn. **90**, 032001 (2021).

[10] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, ''Virtual Distillation for Quantum Error Mitigation'', Phys. Rev. X **11**, 041036 (2021).

[11] B. Koczor, ''Exponential Error Suppression for Near-Term Quantum Devices'', Phys. Rev. X **11**, 031057 (2021).

[12] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, ''Generalized Quantum Subspace Expansion'', Phys. Rev. Lett. **129**, 020502 (2022).

[13] E. Campbell, ''Random Compiler for Fast Hamiltonian Simulation'', Phys. Rev. Lett. **123**, 070503 (2019).

[14] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, ''Exponential improvement in precision for simulating sparse Hamiltonians'', in Proceedings of the forty-sixth annual ACM symposium on Theory of computing, STOC '14 (May 31, 2014), pp. 283–292.

[15] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, ''Simulating Hamiltonian Dynamics with a Truncated Taylor Series'', Phys. Rev. Lett. **114**, 090502 (2015).

[16] R. Babbush, D. W. Berry, I. D. Kivlichan, A. Y. Wei, P. J. Love, and A. Aspuru-Guzik, ''Exponentially more precise quantum simulation of fermions in second quantization'', New J. Phys. **18**, 033032 (2016).

[17] G. H. Low and I. L. Chuang, ''Hamiltonian Simulation by Qubitization'', Quantum **3**, 163 (2019).

[18] L. Lin and Y. Tong, ''Heisenberg-Limited Ground-State Energy Estimation for Early Fault-Tolerant Quantum Computers'', PRX Quantum **3**, 010318 (2022).

Designing Elegant Bell Inequalities

Kwangil Bae[1] * Junghee Ryu[2 3] Ilkwon Sohn[1] Wonhyuk Lee[1]

[1] *Quantum Network Research Center, Division of Science and Technology Digital Convergence, Korea Institute of Science and Technology Information, Daejeon 34141, Republic of Korea*

[2] *Center for Quantum Information R&D, Division of National Supercomputing, Korea Institute of Science and Technology Information, Daejeon 34141, Republic of Korea*

[3] *Division of Quantum Information, Korea National University of Science and Technology (UST)*

Abstract. Elegant Bell inequality is well known for its much exploited property, being maximally violated by maximal entanglement, mutually unbiased bases, and symmetric informationally complete potitive operator-valued measure elements. It is the only one with such property known so far. We present a method to construct Bell inequalities with violation feature analogous to original elegant Bell inequality in high dimension from a simple analytic quantum bound. A Bell inequality with such feature is derived in three dimension for the first time. It shows larger violation compared to known Bell inequalities of similar classes while requiring arguably small number of measurements.

Keywords: Bell inequality, Mutually unbiased bases, Symmetric informationally complete POVM

1 Introduction

Performing local measurements on the shared system, distant subsystems can reveal a quantum correlation that cannot be reproduced by any local hidden variable theoretical explanation. Such nonlocality of the correlation can be tested via the violation of Bell inequalities [1]. What Bell inequality can tell us about the system goes beyond the scope of the mere existence of nonlocal correlation. It has been investigated the method to exploit Bell's theorem for further certifying the properties of given system in so-called device-independent (DI) approach [2]. This approach leads to various applications, including DI quantum key distribution [3, 4, 5] and DI certification of randomness [6, 7].

Reflecting on the aforementioned possibilities provided by Bell inequalities, it is of prime importance to derive them such that their maximal violations are obtained from desirable features, for example maximal entanglement, mutually unbiased bases (MUBs), or symmetric informationally complete POVM elements (SICs). As well known, Clauser-Horne-Shimony-Holts (CHSH) Bell inequality is the case in which the maximal violation can be obtained from maximal entanglement and Pauli measurements having MUBs as their eigenstates. There exists another bipartite Bell inequality so-called Gisin's elegant Bell inequality (EBI) [8],

$$\begin{aligned} &E_{11} + E_{12} - E_{13} - E_{14} \\ &+ E_{21} - E_{22} + E_{23} - E_{24} \\ &+ E_{31} - E_{32} - E_{33} + E_{34} \leq 6, \end{aligned}$$

also optimally violated by maximally entangled state. The expectation value of the product outcomes of Alice's x-th measurement and Bob's y-th measurement is denoted by E_{xy}. Quantum bound for EBI is proven as $4\sqrt{3} \simeq 6.9282 > 6$ [9]. What makes EBI more interesting is that its violation is obtained from the case where the local measurements of each subsystem is constructed by MUBs and SICs respectively. As first remarked by Gisin, the geometrical symmetries in the optimal measurement settings of EBI are clearly revealed when they are mapped on Bloch's sphere [8]. Three eigenstates defining three projective measurements of Alice form three mutually orthogonal vectors on Bloch sphere and four eigenstates of Bob's side make a tetraheron [8]. Original EBI has been used in several fields of quantum information including DI randomness certification [9, 16], DI certification of non-projective measurement [11], and for other tasks [12, 13]. Self-testing property of EBI has also been studied in [14].

MUBs and SICs are important resources in quantum cryptography [9, 15, 16]. There has been efforts to derive Bell inequalities for MUBs or SICs exploiting the advantage of high dimension [17, 18]. For example, for DI certification of optimal amount of randomness, high dimension provides natural advantage [9]. EBI is distinguished from Bell inequalities constructed either for MUBs or SICs as it is optimally violated by both of them.

We address the problem of deriving Bell inequalities with violation feature analogous to original EBI in high dimension. The problem is closely related to the question of what benefit for Bell nonlocality would come from simultaneous consideration of the symmetries of MUBs and SICs. It is worth noting that although EBI is originally derived in the different context [8], we focus on the generalization of the property, being maximally violated by maximal entanglement, MUBs and SICs. Such generalization of EBI has not much been investigated so far to our knowledge. We approach the raised problem, by defining optimal correlation as follows.

Definition. For prime $d \geq 3$, *generalized elegant correlation*(GEC) is defined as the bipartite correlation obtained from the quantum realization: (i) maximally entangled state $\left|\phi_d^+\right\rangle := (1/\sqrt{d}) \sum_{\alpha=0}^{d-1} |\alpha\alpha\rangle$, (ii) Alice's $d^2 - 1$ observables comprising $\mathcal{W}_d \setminus \mathbb{1}$ and (iii) Bob's d^2 observables having d^2 SICs in their eigenspaces.

*kibae@kisti.re.kr

In condition (ii), $\mathcal{W}_d := \{X^m Z^n | m, n \in [0, d)\}$ is the set of d^2 Weyl-Heisenberg (WH) operators defined with $X := \sum_{\alpha=0}^{d-1} |\alpha+1\rangle\langle\alpha|$, $Z := \sum_{\alpha=0}^{d-1} \omega^\alpha |\alpha\rangle\langle\alpha|$ and $\omega := e^{2\pi i/d}$. The integer domain $\{0, 1, \ldots, d-1\}$ is denoted by $[0, d)$ and similar abbreviation is to be used throughout this work.

2 Preliminaries

We consider the measurement scenario of two subsystems Alice and Bob respectively performing one of possible d^2-1 and d^2 number of d outcome measurements A_x and B_y on the shared system. Measurements are labelled with $x \in \{1, 2, \ldots, d^2-1\}$ and $y \in \{0, 1, \ldots, d^2-1\}$. Outcomes of A_x and B_y are respectively distinguished with indices $\alpha, \beta \in \{0, 1, \ldots, d-1\}$. Probability of obtaining the outcomes α and β respectively from settings x and y is denoted by $P(\alpha\beta|xy)$. A statistics of the Bell test can be described with correlation, $\mathbf{p} := \{P(\alpha\beta|xy)|\forall\alpha, \beta, x, y\}$. Quantum correlation allows

$$P(\alpha\beta|xy) = \mathrm{tr}\left[\rho_{AB}(G_\alpha^x \otimes H_\beta^y)\right]$$

where ρ_{AB} is the bipartite density matrix of Alice and Bob. Measurement operators, $\{G_\alpha^x\}$ and $\{H_\beta^y\}$, define the measurements of Alice and Bob respectively. Bell inequality is used to show that there is quantum correlation which does not allow the local hidden variable (LHV) theoretical description of the correlation [1]. Geometrically, LHV correlation region forms a polytope, $\mathcal{L}$, and quantum region, $\mathcal{Q}$, is known to include it.

The generic form of so-called Bell expression without marginal probabilities can be defined as,

$$S = \sum_{\alpha,\beta}\sum_{x,y} g_{x,y}^{\alpha,\beta} P(\alpha\beta|xy) \tag{1}$$

$$= \sum_{n}\sum_{x,y} f_{x,y}^n \langle A_x^n B_y^n\rangle \tag{2}$$

where $n \in (1, d)$ and $g_{x,y}^{\alpha,\beta} \in \mathbb{R}$ is the real weight for each probability. For simplicity, the summations over all possible values are to be denoted without limits. The correlator $\langle A_x^n B_y^n\rangle = \sum_{\alpha,\beta=0}^{d-1} \omega^{n(\alpha+\beta)} P(\alpha\beta|xy)$, where the outcomes of the measurements A_x and B_y are respectively set as $\omega^\alpha, \omega^\beta \in \{1, \omega, \ldots, \omega^{d-1}\}$. And the coefficient $g_{x,y}^{\alpha,\beta} := \sum_n f_{x,y}^n \omega^{n(\alpha+\beta)}$. The condition $f_{x,y}^{d-n} := (f_{x,y}^n)^*$ guarantees that the expression is real-valued. A certain expression is specified by a $(d^2-1)\times d^2$ coefficient matrices, $\{F_n\}$, whose entries are $\{f_{x,y}^n\}$. We define x-th row and y-th column of F_n respectively as column vectors $\mathbf{r}_x^n := (f_{x,0}^n, f_{x,1}^n, \ldots, f_{x,d^2-1}^n)^\intercal$ and $\mathbf{c}_y^n := (f_{1,y}^n, f_{2,y}^n, \ldots, f_{d^2-1,y}^n)^\intercal$ such that $F_n = [\mathbf{r}_1^n, \mathbf{r}_2^n, \ldots, \mathbf{r}_{d^2-1}^n]^\intercal = [\mathbf{c}_0^n, \mathbf{c}_1^n, \ldots, \mathbf{c}_{d^2-1}^n]$. In terms of F_n, real-valuedness condition is re-written as $F_n = F_{d-n}^*$.

Maximal value of S under LHV theory is, $L := \max_{\mathbf{p}\in\mathcal{L}} S(\mathbf{p})$. Violation of the Bell inequality by quantum mechanics, $L < Q$, is shown with the maximal quantum value, $Q := \max_{\mathbf{p}\in\mathcal{Q}} S(\mathbf{p})$. Quantum mechanical expectation can be evaluated with the operator,

$$\mathcal{B} = \sum_n \sum_{x,y} f_{x,y}^n A_x^n \otimes B_y^n \tag{3}$$

Here, we consider the quantum measurements, $A_x = \sum_\alpha \omega^\alpha G_\alpha^x$ and $B_y = \sum_\beta \omega^\beta H_\beta^y$ where G_α^x, H_β^y are rank-one projectors.

3 Result

As one of the main results, we proved the below theorem. Here, we explain the meaning of the theorem in brief manner. It states that the correlation function maximized by GEC always can be found from Eq.(4) with Weyl-Heisenberg(WH) group covariant SICs. One has freedom in the choice of SICs, consequently determining the coefficients $\{f_{x,0}^n\}$ in our method. One can construct different correlation functions maximized by GEC by varying $\{f_{x,0}^n\}$ of Eq.(4) with different choices of SICs.

Theorem. Consider prime d, in which WH group covariant SICs are defined. Then, there always exists the expression S defined with,

$$f_{dp+q,dr+s}^n = \omega^{-n(ps+qr)} f_{dp+q,0}^n \tag{4}$$

for $p, q, r, s \in [0, d)$, $dp+q \neq 0$ whose quantum maximum

$$\tilde{Q} = \frac{1}{2}\left[\sum_n \|F_n\|^2 + d^2(d-1)\right] \tag{5}$$

is obtained with *generalized elegant correlation.*

In the Eq.(5), $\|\cdot\|$ is the Frobeniuns norm of a matrix. The upper bound (5) can be obtained with the maximally entangled state $|\phi_d^+\rangle$ and the optimal measurements,

$$A_{dp+q} = \tau W_{dp+q} \quad \forall p, q \tag{6}$$

$$B_y = W_y B_0 W_y^\dagger \quad \forall y \tag{7}$$

where $\tau := \omega^{pq\delta(d,2)/2}$ is defined with kronecker delta, δ and the parameterization $x = dp + q$ is used in Eq.(6). The index $x \in (0, d^2)$ is parameterized with $p, q \in [0, d)$ satisfying $dp + q \neq 0$ and $y \in [0, d^2)$. The constant τ is especially introduced for $d = 2$, the only even prime dimension, to satisfy $A_3^2 = \mathbb{1}$. Alice's observables in (6) are given as $d^2 - 1$ WH operators. Bob's d^2 observables are generated from B_0 in Eq.(7).

From the $d = 2$ case of our framework, the Gisin's EBI can be generated. In $d = 3$, we derive a Bell inequality maximally violated by generalized elegant correlation,

$$S_3 := \sum_{n=1}^{2}\sum_{x=1}^{8}\sum_{y=0}^{8} f_{x,y}^n \langle A_x B_y\rangle \leq 15 \tag{8}$$

where the coefficients for $n = 1$ is defined from Eq.(4) along with the first column of F_1, $(f_{1,0}^1, f_{2,0}^1, \ldots, f_{8,0}^1)^\intercal = (1/2)(1, 1, \lambda, \nu, \mu, \lambda, \mu, \nu)^\intercal$ with $\lambda := -i/\sqrt{3}$, $\mu := \omega\lambda$ and $\nu := \omega^2\lambda$. The coefficient matrix is expressed as,

$$F_1 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & \omega^2 & \omega^2 & \omega^2 & \omega & \omega & \omega \\ 1 & 1 & 1 & \omega & \omega & \omega & \omega^2 & \omega^2 & \omega^2 \\ \lambda & \nu & \mu & \lambda & \nu & \mu & \lambda & \nu & \mu \\ \nu & \mu & \lambda & \mu & \lambda & \nu & \lambda & \nu & \mu \\ \mu & \lambda & \nu & \nu & \mu & \lambda & \lambda & \nu & \mu \\ \lambda & \mu & \nu & \lambda & \mu & \nu & \lambda & \mu & \nu \\ \mu & \nu & \lambda & \lambda & \mu & \nu & \nu & \lambda & \mu \\ \nu & \lambda & \mu & \lambda & \mu & \nu & \mu & \nu & \lambda \end{bmatrix}.$$

All the coefficients are determined with the above F_1 and the real-valuedness condition, $F_1 = F_2^*$. The violation of the Bell inequality (8) is obtained from the maximum quantum value, $\tilde{Q} = 18$, evaluated from (5) with $\|F_n\| = 9$ for $n = 1, 2$. The optimal violation is obtained from maximally entangled state $\left|\phi_d^+\right\rangle$, $A_x = W_x$ and $B_y = W_y B_0 W_y^\dagger$ with

$$B_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1/2 & i\sqrt{3}/2 \\ 0 & i\sqrt{3}/2 & -1/2 \end{pmatrix}. \tag{9}$$

Detailed derivation is to be given in the presentation.

We remark that the Bell inequality defined in (8) shows stronger violation than known Bell inequalities [8, 17, 18] of similar classes. The critical visibility ν^c, the smallest ν of the isotropic state $\rho = \nu \left|\phi_d^+\right\rangle \left\langle\phi_d^+\right| + (1-\nu)/d^2$ to show the violation, is 83.33% for S_3. It is smaller than that of similar types of Bell inequalities including Bell inequality for $d = 3$ MUBs, 96.77% [17], 96.63% [18], for SICs, 96.41% [18], and original EBI, 86.6% . In addition, the number of settings required for Alice is 8 for S_3 and it is smaller than the case of knwon Bell inequality for $d = 3$ SICs which requires $\binom{d^2}{2} = 36$ settings for Alice when the number of Bob's setting is same as 9 [18].

4 Conclusion

We generalize the original feature of EBI, being maximally violated by maximal entanglement, MUBs and SICs, in arbitrary prime local dimension. Based on it, we presented a method to construct a correlation function for which the quantum maximum is always obtained in a simple analytic form from complete MUBs and SICs. A Bell inequality for two-qutrit system derived with our method manifests larger violation compared to known Bell inequalities of similar classes, including original EBI. Our work lays a foundation for the future studies of device independent cryptographic protocols exploiting MUBs or SICs in high dimension.

References

[1] J. S. Bell. On the Einstein Podolsky Rosen paradox Phys. Phys. Fiz., pages 195–200, 1964.

[2] D. Mayers and A. Yao Self testing quantum apparatus arxiv: 0307205v3, 2004.

[3] U. Vazirani and T. Vidick Fully Device-Independent Quantum Key Distribution Phys. Rev. Lett. 113, 140501, 2014

[4] C. A. Miller and Y. Shi Robust Protocols for Securely Expanding Randomness and Distributing Keys Using Untrusted Quantum Devices J. ACM 63(4):0004-5411, 2016

[5] R. Schwonnek et al. Device-independent quantum key distribution with random key basis Nat. Comm. 12(1), 2880, 2021.

[6] S. Pironio et al. Random numbers certified by Bell's theorem Nature, 464, 1021-1024, 2010.

[7] Y. Liu et al. Device-independent quantum random-number generation Nature 562, 548–551, 2018.

[8] N. Gisin Bell Inequalities: Many Questions, a Few Answers *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle* Springer, 2009.

[9] A. Acín and S. Pironio Optimal randomness certification from one entangled bit Phys. Rev. A, 93, 040102, 2016.

[10] O. Andersson et al. Device-independent certification of two bits of randomness from one entangled bit and Gisin's elegant Bell inequality Phys. Rev. A, 97, 012314, 2018.

[11] M. Smania et al. Experimental certification of an informationally complete quantum measurement in a device-independent protocol Optica, 7(2), 123–128, 2020.

[12] M. Smania et al. Robust self-testing of steerable quantum assemblages and its applications on device-independent quantum certification Quantum, 5, 552, 2021.

[13] A. Tavakoli et al. Correlations in star networks: from Bell inequalities to network inequalities New Journal of Physics, 19, 073003, 2017.

[14] O. Andersson et al. Self-testing properties of Gisin's elegant Bell inequality Phys. Rev. A, 96, 032119, 2017.

[15] C. H. Bennett and G. Brassard In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* IEEE, 1984

[16] O. Andersson, P. Badziág, I. Dumitru and A. Cabello Device-independent certification of two bits of randomness from one entangled bit and Gisin's elegant Bell inequality Phys. Rev. A, 97, 012314, 2018.

[17] J. Kaniewski et al. Maximal nonlocality from maximal entanglement and mutually unbiased bases, and self-testing of two-qutrit quantum systems Quantum, 3, 198, 2019.

[18] A. Tavakoli et al. Mutually unbiased bases and symmetric informationally complete measurements in Bell experiments Science Advances, 7(7), eabc3847, 2021.

Interplay among entanglement, measurement incompatibility, and nonlocality

Yuwei Zhu[1 2 *] Xingjian Zhang[1 †] Xiongfeng Ma[1 ‡]

[1] *Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, P. R. China*

[2] *Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, P. R. China*

Abstract. Nonlocality within quantum systems guarantees the presence of entanglement, prompting the question of how much entanglement is necessary for specific nonlocal behaviors. This study answers this by examining generalized Clauser-Horne-Shimony-Holt-type (CHSH) Bell inequalities to quantify entanglement. We derive analytical bounds on the entanglement of formation and negativity and provide numerical estimates of one-way distillable entanglement. Without neglecting the necessity of measurement incompatibility, our study reveals a counterintuitive interplay among entanglement, incompatibility and nonlocality in qubit-qubit systems. The study also applies its findings to experimental states, demonstrating the potential to enhance entanglement estimation through optimized Bell inequality selection. This submission is based on Ref. [1].

Keywords: entanglement, nonlocality, measurement incompatibility

1 Introduction

In the early stages of quantum mechanics, Einstein, Podolsky, and Rosen identified "spooky action" between observables [2], later formalized by Bell [3]. If the CHSH value exceeds 2, it indicates a violation of local realism, ensuring Bell nonlocality and the presence of entanglement, a non-classical feature essential for quantum information tasks. Entanglement is characterized by a joint state that cannot be generated through local operations and classical communication (LOCC) [4, 5]. It is quantified using appropriate measures like entanglement of formation and one-way distillable entanglement [6–9]. However, state tomography, which fully reconstructs a quantum state, is prone to untrust results due to detection loss, noise, and attacks from adversaries [10, 11]. Quantum nonlocality allows us to detect entanglement while not needing to characterize the quantum devices *a priori* [12, 13]. This observation leads to the question of what the minimum amount of entanglement is necessary for a given nonlocal behavior, which serves as a device-independent (DI) entanglement estimation tool [14–20]. Despite the significant role of measurement incompatibility in this narrative, it has been largely overlooked in previous studies. This oversight makes exploring the interplay among entanglement, nonlocality, and measurement incompatibility a captivating and worthwhile endeavor. The illustration is in Fig. 1.

In this work, we first explore the minimum entanglement necessary for a given nonlocal behavior manifested by violating generalized CHSH-type Bell inequalities. Then, we investigate the interplay among entanglement, nonlocality, and measurement incompatibility, revealing a complex relationship rather than a simple trade-off. In the third part, we analyze the statistics that arise from pure entangled states and Werner states and examine the performance of our results.

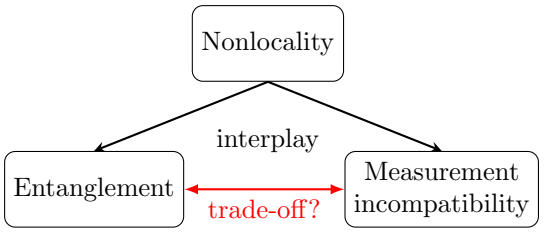


Figure 1: The interplay among nonlocality, entanglement, and measurement incompatibility.

2 Preliminaries

To establish the interplay among nonlocality, entanglement, and measurement incompatibility, in our study, we refer to the generalized CHSH-type Bell inequality for quantifiers of nonlocality. The corresponding Bell expression under quantum measurements is given by:

$$S = \mathrm{Tr}\left(\rho_{AB}\hat{S}_\alpha\right), \tag{1}$$

where $\hat{S}_\alpha = \alpha\hat{A}_0 \otimes \hat{B}_0 + \alpha\hat{A}_0 \otimes \hat{B}_1 + \hat{A}_1 \otimes \hat{B}_0 - \hat{A}_1 \otimes \hat{B}_1$. A Bell value within $(2\alpha, 2\sqrt{\alpha^2+1}]$ indicates nonlocality, further indicating entanglement and measurement incompatibility [21, 22].

We focus on entanglement estimation in bipartite systems, represented by density operators ρ_{AB}. We employ various entanglement measures. For qubit pairs, the concurrence $C(\rho_{AB})$ is considered [23, 24]. For general bipartite states, we also consider:

- Entanglement of formation (EOF), $E_\mathrm{F}(\rho_{AB})$, defined via the convex-roof construction from pure states. For two-qubit states, it simplifies to a closed form in terms of concurrence [23].
- One-way distillable entanglement (ODE), given

*zhuyw18@mails.tsinghua.edu.cn
†thuxjzhang@gmail.com
‡xma@tsinghua.edu.cn

by the negative conditional entropy $E_D^{\rightarrow}(\rho_{AB}) = -H(A|B)_\rho = H(\rho_B) - H(\rho_{AB})$ [25].

- Negativity of entanglement $\mathcal{N}(\rho_{AB})$, determined by the violation of the positive partial transpose (PPT) criterion [26].

3 DI entanglement quantification

Our goal is to estimate the underlying entanglement of a state from the observed α-CHSH Bell value. For a given entanglement measure E, the problem is formulated as:

$$\begin{aligned} E_{\text{est}} = & \min_{\rho_{AB},\hat{A}_0,\hat{A}_1,\hat{B}_0,\hat{B}_1} E(\rho_{AB}), \\ \text{s.t.} \quad & S = \text{Tr}\left(\rho_{AB}\hat{S}_\alpha\right), \\ & \rho_{AB} \geq 0, \\ & \text{Tr}(\rho_{AB}) = 1. \end{aligned} \tag{2}$$

The optimization problem is challenging to solve directly. We degenerate this problem into four steps.

1. Set up the entanglement estimation with the observed Bell value S as the constraint, as in Eq. (2).
2. Apply duality to find optimal measurements that maximize the Bell value for a given state ρ_{AB}.
3. Use Jordan's Lemma to simplify the measurement process as a convex combination of qubit pairs.
4. Narrow down to Bell-diagonal states for the optimization, maintaining the α-CHSH Bell value and non-increasing entanglement through LOCC.

We systematically obtain a tight estimation result for qubit pairs through these steps. Based on Jordan's lemma, when generalizing the results from a qubit-pair scenario to unknown dimensions, if the qubit-pair entanglement estimation E_{est} is convex in the Bell value S, the process is straightforward. On the other hand, if it is concave, a "convex closure" of the function is necessary for general bipartite states. A detailed discussion can be found in Ref. [1].

By applying different entanglement measures, we obtain the following results.

Theorem 1 *Suppose the underlying quantum state is a pair of qubits. For a given tilted CHSH expression in Eq.* (1) *parametrized by* α*, if the Bell expression value is* S*, then the amount of concurrence in the underlying state can be lower-bounded,*

$$C(\rho_{AB}) \geq \sqrt{\frac{S^2}{4} - \alpha^2}. \tag{3}$$

Theorem 1 is under the assumption that the underlying state is a pair of qubits. Applying the convex closure technique and the closed-form expression of EOF in terms of the concurrence, the EOF of an unknown dimension system can be lower-bounded.

Theorem 2 *For a given tilted CHSH expression in Eq.* (1)*, if the Bell expression value is* S*, then the EOF in the underlying state can be lower-bounded,*

$$E_{\text{F}}(\rho_{AB}) \geq \frac{S - 2\alpha}{2\sqrt{1+\alpha^2} - 2\alpha}. \tag{4}$$

With similar techniques, the negativity of entanglement is lower-bounded without system-dimension assumption.

Corollary 3 *For a given tilted CHSH expression in Eq.* (1)*, if the Bell expression value is* S*, then the amount of negativity in the underlying state can be lower-bounded,*

$$\mathcal{N}(\rho_{AB}) \geq \frac{S - 2\alpha}{4(\sqrt{1+\alpha^2} - \alpha)}. \tag{5}$$

Notably, when $\alpha = 1$, Eq. (5) analytically confirms the observed linear relation between negativity and Bell value. Our result proves the conjecture of Eq. (5) in Ref. [16], obtained by using the third level of a Navascués-Pironio-Acín-type numerical algorithm [27].

Additionally, we provide DI numerical estimation results for ODE under α-CHSH Bell nonlocality. We present numerical results for some values of α in Fig. 2.

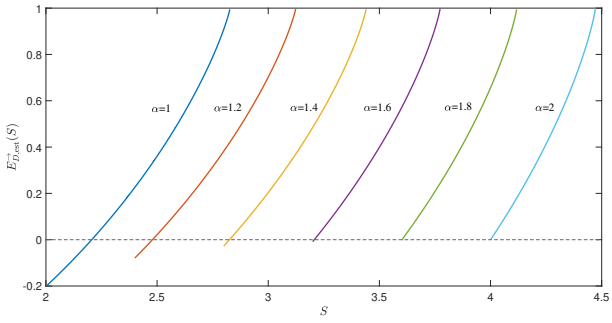


Figure 2: ODE estimation diagram using CHSH-type Bell expressions with increasing discrete α. Estimation $E_{D,\text{est}}^{\rightarrow}(S)$ is convex over valid S range. $E_{D,\text{est}}^{\rightarrow}(S)$ at $S = 2\alpha$ increases and converges to 0 as α grows.

4 Interplay among nonlocality, entanglement, and measurement incompatibility

We further explore the interplay among nonlocality, entanglement, and measurement incompatibility for a pair of qubits. We aim to determine the minimum entanglement required for nonlocal behavior given a level of measurement incompatibility. The optimization problem is defined based on Eq. (2) with a few additional assumptions: (1) the underlying system is a pair of qubits, and (2) the measurement operators are qubit observables. These assumptions aim to provide straightforward parameterization in the study. Specifically, we assume measurements as $\hat{A}_0 = \sigma_z, \hat{A}_1 = \sigma_x, \hat{B}_0 = \cos\theta\sigma_z + \sin\theta\sigma_x, \hat{B}_1 = \cos\theta\sigma_z - \sin\theta\sigma_x$. This parametrization determines the measurement incompatibility of Bob as

$|\sin(2\theta)|$: when $\theta = 0$, $\hat{B}_0$ and $\hat{B}_1$ commute; when $\theta = \pi/4$,$\hat{B}_0$ and $\hat{B}_1$ exhibit maximal incompatibility. We analyze the relation between entanglement and measurement incompatibility varying θ within the range $[0, \pi/4]$.

Taking concurrence and ODE as the target entanglement measures, we simulate the trajectories between θ and E_{est} for discrete α-CHSH violations when $\alpha = 1.2$, as shown in Fig. 3. The figure shows a non-trivial relation between the necessary amount of entanglement with respect to measurement incompatibility under different Bell values S. For both entanglement measures, initially, entanglement decreases with increasing incompatibility. After reaching a threshold θ_E^* ($\theta_C^* = \arctan\left(\sqrt{S^2/4 - \alpha^2}/\alpha\right)$ for concurrence), entanglement increases counterintuitively. The range of θ for which nonlocality is possible shrinks as S increases, with the underlying state self-testing maximally entangled state and non-maximally incompatible measurements (with $\theta = \arctan 1/\alpha$) at the highest Bell value, $2\sqrt{1+\alpha^2}$. These trajectories indicate a complex interplay among the three quantities: the most nonlocality, in general, is not indicated by the largest amount of entanglement and measurement incompatibility.

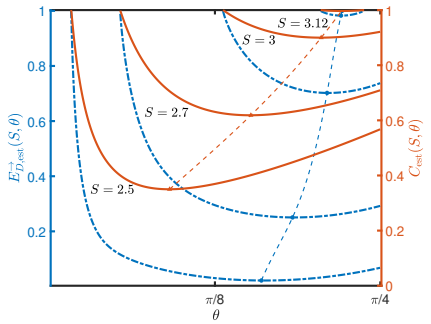


Figure 3: Interplay among Bell nonlocality, measurement incompatibility, and entanglement using the α-CHSH Bell expression with $\alpha = 1.2$. Blue curves: ODE; red curves: concurrence. Before θ reaches a threshold θ_E^*, there is a trade-off: less entanglement is needed for nonlocality as measurements become more incompatible, and vice versa; when $\theta \in (\theta_C^*, \pi/4]$, entanglement increases with θ. At the peak Bell value of $2\sqrt{1.2^2+1}$, the system self-tests a single point, showing that maximal violation occurs with maximal entanglement and non-maximal incompatibility.

5 Application: optimizing entanglement estimation in realistic settings

Finally, we examine the impact of various Bell expressions on entanglement estimation, particularly under conditions of experimental imperfections such as loss and noise. The α-CHSH expression is used for its enhanced capacity to estimate entanglement over the standard CHSH expression.

We delve into the analysis of Werner state,

$$\rho_{\text{W}}(p) = (1-p)\left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + p\frac{I}{4}, \tag{6}$$

where we write $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. The Werner state is entangled when $p < 2/3$. We parametrize the measurements as follows for Alice and Bob: $\hat{A}_0 = \sigma_z, \hat{A}_1 = \cos\theta_1\sigma_z + \sin\theta_1\sigma_x, \hat{B}_0 = \cos\theta_2\sigma_z + \sin\theta_2\sigma_x, \hat{B}_1 = \cos\theta_3\sigma_z + \sin\theta_3\sigma_x$. Fig. 4 indicates that based on the entanglement estimation result in Sec. 3, the optimal value for estimating EOF and ODE of a Werner state are 1.2 and 1.4, respectively. This insight is instrumental for strategically selecting CHSH-type inequalities in entanglement estimation, as shown in the analytical results for EOF.

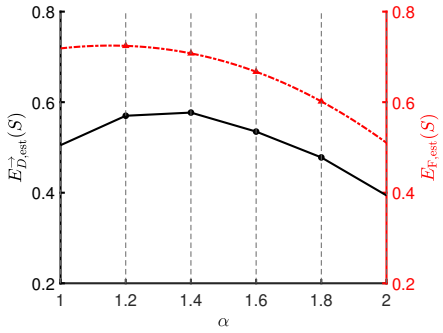


Figure 4: Entanglement estimation results for nonlocal correlations arising from Werner states. The experimental setting is given by $p = 0.05, \theta_1 = \pi/2, \theta_2 = \pi/6$, and $\theta_3 = -\pi/6$. We depict the entanglement estimation results using different α-CHSH Bell expressions. We plot the estimated values of ODE and EOF with the black solid line and the red dashed line, respectively.

Theorem 4 *In a Bell test experiment, suppose the underlying state of the system takes the form of Eq.* (6)*, and the observables are parameterized as above. For EOF estimation solely from the violation values of α-CHSH Bell inequalities, if $\theta_1, \theta_2, \theta_3$ and p satisfy*

$$(1-p)[\sin\theta_1(\sin\theta_2 - \sin\theta_3) + \cos\theta_2(\sqrt{2}+1+\cos\theta_1) + \cos\theta_3(\sqrt{2}+1-\cos\theta_1)] > 2(1+\sqrt{2}), \tag{7}$$

then there exists $\alpha > 1$, where a better estimation of $E_{\text{F,est}}(S)$ can be obtained by using the α-CHSH inequality parameterized by this value than by using the original CHSH inequality (corresponding to $\alpha = 1$).

Theorem 4 analytically confirms that nonlocality depicted by the original CHSH Bell value does not always provide the EOF estimation that approaches the real value. Similarly, examples and conclusions for the non-maximally entangled state are derived in Ref. [1]. These results guarantee that more optimal choices of Bell inequality exist for entanglement quantification in various scenarios.

References

[1] Y. Zhu, X. Zhang, and X. Ma, *Entanglement quantification via nonlocality* (2023), 2303.08407.

[2] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935), URL https://link.aps.org/doi/10.1103/PhysRev.47.777.

[3] J. S. Bell, Phys. Phys. Fiz. **1**, 195 (1964), URL https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195.

[4] O. Gühne and G. Tóth, Phys. Rep. **474**, 1 (2009), URL https://www.sciencedirect.com/science/article/pii/S0370157309000623.

[5] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009), URL https://link.aps.org/doi/10.1103/RevModPhys.81.865.

[6] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997), URL https://link.aps.org/doi/10.1103/PhysRevLett.78.2275.

[7] H.-K. Lo and H. F. Chau, Science **283**, 2050 (1999), URL https://www.science.org/doi/10.1126/science.283.5410.2050.

[8] P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000), URL https://link.aps.org/doi/10.1103/PhysRevLett.85.441.

[9] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996), URL https://link.aps.org/doi/10.1103/PhysRevA.54.3824.

[10] P. Xu, X. Yuan, L.-K. Chen, H. Lu, X.-C. Yao, X. Ma, Y.-A. Chen, and J.-W. Pan, Phys. Rev. Lett. **112**, 140506 (2014), URL https://link.aps.org/doi/10.1103/PhysRevLett.112.140506.

[11] X. Yuan, Q. Mei, S. Zhou, and X. Ma, Phys. Rev. A **93**, 042317 (2016), URL https://link.aps.org/doi/10.1103/PhysRevA.93.042317.

[12] D. Mayers and A. Yao, in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Washington, DC, USA, 1998), FOCS '98, pp. 503–509, ISBN 0-8186-9172-7, URL http://dl.acm.org/citation.cfm?id=795664.796390.

[13] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007), URL https://link.aps.org/doi/10.1103/PhysRevLett.98.230501.

[14] F. Verstraete and M. M. Wolf, Phys. Rev. Lett. **89**, 170401 (2002), URL https://link.aps.org/doi/10.1103/PhysRevLett.89.170401.

[15] Y.-C. Liang, T. Vértesi, and N. Brunner, Phys. Rev. A **83**, 022108 (2011), URL https://link.aps.org/doi/10.1103/PhysRevA.83.022108.

[16] T. Moroder, J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne, Phys. Rev. Lett. **111**, 030501 (2013), URL https://link.aps.org/doi/10.1103/PhysRevLett.111.030501.

[17] G. Tóth, T. Moroder, and O. Gühne, Phys. Rev. Lett. **114**, 160501 (2015), URL https://link.aps.org/doi/10.1103/PhysRevLett.114.160501.

[18] R. Arnon-Friedman and H. Yuen, arXiv:1712.09368 (2017), URL https://arxiv.org/abs/1712.09368.

[19] S.-L. Chen, C. Budroni, Y.-C. Liang, and Y.-N. Chen, Phys. Rev. A **98**, 042127 (2018), URL https://link.aps.org/doi/10.1103/PhysRevA.98.042127.

[20] R. Arnon-Friedman and J.-D. Bancal, New J. Phys. **21**, 033010 (2019), URL https://iopscience.iop.org/article/10.1088/1367-2630/aafef6.

[21] A. Acín, S. Massar, and S. Pironio, Phys. Rev. Lett. **108**, 100402 (2012), URL https://link.aps.org/doi/10.1103/PhysRevLett.108.100402.

[22] E. Woodhead, A. Acín, and S. Pironio, Quantum **5**, 443 (2021), ISSN 2521-327X, URL https://doi.org/10.22331/q-2021-04-26-443.

[23] S. A. Hill and W. K. Wootters, Phys. Rev. Lett. **78**, 5022 (1997), URL https://link.aps.org/doi/10.1103/PhysRevLett.78.5022.

[24] P. Rungta, V. Bužek, C. M. Caves, M. Hillery, and G. J. Milburn, Phys. Rev. A **64**, 042315 (2001), URL https://link.aps.org/doi/10.1103/PhysRevA.64.042315.

[25] M. M. Wilde, M. Tomamichel, and M. Berta, IEEE Trans. Inf. Theory **63**, 1792 (2017).

[26] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002), URL https://link.aps.org/doi/10.1103/PhysRevA.65.032314.

[27] M. Navascués, S. Pironio, and A. Acín, Phys. Rev. Lett. **98**, 010401 (2007), URL https://link.aps.org/doi/10.1103/PhysRevLett.98.010401.

Explicit decoders using quantum singular value transformation

Takeru Utsumi[1] * Yoshifumi Nakata[2] †

[1] *Graduate School of Arts and Sciences, University of Tokyo, Japan.*
[2] *Yukawa Institute for Theoretical Physics, Kyoto University, Japan.*

Abstract. Explicitly constructing a quantum decoder is key to reliably transmitting quantum information. In this submission, we provide two explicit decoders capable of recovering quantum information whenever it is in principle recoverable, i.e., when the decoupling condition is satisfied. These decoders particularly achieve quantum capacities with a suitable encoder. They are constructed using the amplitude amplification algorithm based on the quantum singular value transformation (QSVT), revealing the power of the quantum algorithmic approach to quantum decoders. The proposed decoders also have practical advantages since they reduce the computational cost for implementation compared to a previous explicit decoder.

Keywords: quantum decoder, decoupling, Yoshida-Kitaev decoder, Petz recovery map, amplitude amplification, quantum singular value transformation (QSVT)

1 Introduction and summary of results

Reliably transmitting quantum information via a noisy quantum channel is crucial in quantum information theory, and is commonly investigated by the decoupling approach [1–3]. The decoupling offers the necessary and sufficient condition for reliable transmission and has an outstanding feature that one can address the problem *without* explicitly considering a decoder. This feature is typically considered to be a strong advantage of the decoupling approach as it significantly simplifies the analysis, but it can also be a drawback as one cannot achieve the task in practice without an explicit decoder. In recent years, the importance of explicit decoders also increases in fundamental physics to better understand quantum chaos and quantum black holes [4–14]. Hence, it is important in quantum information and fundamental physics to explicitly construct a decoder that can recover quantum information.

As explicitly constructing a decoder is, in general, a highly non-trivial task, only a handful of results are known so far. A commonly used decoder is the Petz recovery map [15,16]. It is known that the Petz recovery map is capable of recovering quantum information with a nearly optimal recovery error [17], and can be implemented by a quantum circuit [18], implying that it can be used as an explicit decoder. However, the circuit complexity for implementing the Petz recovery map is fairly high. Decoders with smaller computational cost have been desired.

A possible approach for constructing a decoder with smaller computational cost is to extend a construction in [19], known as a Yoshida-Kitaev (YK) decoder. The YK decoder is for decoding quantum information in the Hayden-Preskill (HP) protocol [7], a specific model based on the qubit-erasure noise with a Haar random encoding, and the decoder has relatively small circuit complexity. It is constructed in two steps: first, a decoding protocol with quantum measurement and post-selection is considered. Then, the measurement is replaced with a non-trivial application of the amplitude amplification (AA) algorithm. This two-step construction was shown to work well for decoding the HP protocol, but the reason why it works strongly relies on the simple properties of the specific encoding and noise in the HP protocol. It has been unclear if the two-step construction can be extended to general situations.

In this submission based on [19], we explore the two-step construction with a certain modification and provide two explicit quantum decoders. One is a *generalized YK decoder*, and the other is a simplified Petz recovery map that we call a *Petz-like* decoder. The crucial modification for the construction of these decoders is to use the fixed-point AA (FPAA) based on the quantum singular value transformation (QSVT) [20–22] instead of the standard AA algorithm in the original YK approach. Due to the flexibility of the QSVT-based FPAA, all the issues that may arise in extending the original approach can be circumvented, and the two-step construction becomes applicable to more general situations.

*takeru-utsumi@g.ecc.u-tokyo.ac.jp
†yoshifumi.nakata@yukawa.kyoto-u.ac.jp

We show that both the generalized YK and Petz-like decoders have high decoding performance in the sense that they can recover quantum information when the decoupling condition is satisfied. As decoupling is necessary and sufficient for the recovery of quantum information, this implies that the decoders work well whenever quantum information is in principle recoverable. In particular, the decoders can achieve quantum capacity, both entanglement-assisted [23–25] and non-assisted [26–28] ones, when it is combined with a suitable encoder meeting the decoupling condition.

We also investigate the circuit complexity of the proposed decoders. While the complexity depends on various factors, a simple criterion can be obtained for the generalized YK decoder to have smaller complexity than the Petz-like decoder. The criterion is in terms of the number of logical qubits, the amount of entanglement pre-shared between the sender and the receiver, the size of the output of the noisy channel, and the number of Kraus operators of the noisy channel. Furthermore, we show that both decoders have smaller complexity than the Petz recovery map [18] in most parameter range.

Our results extend and demonstrate the power of the two-step construction of quantum decoders that lifts up a decoding protocol with post-selection to a decoding quantum circuit by using the QSVT-based FPAA. This approach is of theoretical interest and paves the way for exploring decoders with better performance by a quantum algorithmic approach. The approach may also be of practical use since we have shown that the constructed decoders have high decoding performance and reduce the computational cost. We expect that our results, opening a new research direction in the intersection of quantum information and quantum algorithms, contribute to the further development of quantum information science.

2 Setting

We use superscripts to indicate the system on which operators and maps are defined. A reduced density operator on, e.g., A of φ^{AB} is denoted by φ^A. We also use the notation that d_A is the dimension of a Hilbert space $\mathcal{H}^A$ and that a Hilbert space, such as $\mathcal{H}^{A'}$ or $\mathcal{H}^{\hat{A}}$, is isomorphic to $\mathcal{H}^A$. A maximally entangled state (MES) in the computational basis $\{|i\rangle\}_i$ is denoted by $|\Phi\rangle^{A\hat{A}} = \frac{1}{\sqrt{d_A}}\sum_{i=1}^{d_A}|i\rangle^A|i\rangle^{\hat{A}}$ with the corresponding density operator $\Phi^{A\hat{A}}$. The completely mixed state (CMS) by π, such as $\pi^A = \mathbb{I}^A/d_A$, where $\mathbb{I}^A$ is the identity operator.

We consider the following standard situation. A sender aims to transmit $(\log d_A)$-qubit quantum information to a receiver using a noisy channel $\mathcal{N}^{C\to D}$. The sender and the receiver may share entanglement $|\Phi\rangle^{BB'}$ in advance, where B (B') is with the sender (receiver). When they share no entanglement, we set $d_B = 1$. The sender encodes the system A with B using an encoder $\mathcal{E}^{AB\to C}$ and sends the encoded system C to the receiver by $\mathcal{N}^{C\to D}$. The receiver applies a decoder $\mathcal{D}^{DB'\to R'}$ onto the system DB' for recovering the quantum information to be transmitted as much as possible. Our goal is to explicitly construct a decoder $\mathcal{D}^{DB'\to R'}$ for a given encoder $\mathcal{E}$ and noisy channel $\mathcal{N}$,

The recovery error $\Delta(\mathcal{D}|\mathcal{E},\mathcal{N})$ by a decoder $\mathcal{D}$ is defined by introducing a reference system R isomorphic to A with $d_R = d_A$ and by preparing the systems A and R to be in a MES $|\Phi\rangle^{AR}$. More precisely,

$$\Delta(\mathcal{D}|\mathcal{E},\mathcal{N}) := \frac{1}{2}\|\Phi^{RR'} - \mathcal{D}^{DB'\to R'}(\omega^{RDB'})\|_1, \quad (1)$$

where $\omega^{DRB'} := \mathcal{N}^{C\to D}\circ\mathcal{E}^{AB\to C}(\Phi^{AR}\otimes\Phi^{BB'})$.

3 Main results

We describe only the two-step construction of the generalized YK decoder. See [29] for the results of the Petz-like decoder and the in-depth analysis of these decoders.

We start with a decoding protocol with post-selection consisting of three steps. Note that the receiver has the output D of the noisy channel and the system B' of the pre-shred entanglement.

1. The MES $\Phi^{A'R'}$ is prepared in ancillary systems $A'R'$.

2. An isometry $(V_{\mathcal{N}\circ\mathcal{E}}^{A'B'\to E'D'})^*$ is applied onto $A'B'$. Here, $V_{\mathcal{N}\circ\mathcal{E}}$ is any Stinespring isometry of $\mathcal{N}\circ\mathcal{E}$ and $*$ is the complex conjugate in the computational basis.

3. The system DD' is measured by $\mathcal{M} := \{|\Phi\rangle\langle\Phi|^{DD'}, \mathbb{I}^{DD'} - |\Phi\rangle\langle\Phi|^{DD'}\}$. If the former outcome is obtained, the protocol succeeds.

This decoding protocol generalizes the original YK decoding protocol in a straightforward manner. It is based on the idea to emulate the inverse dynamics of the encoding and the noisy channel in the receiver's local system and to measure the actual output D of the noisy channel and the emulated one D' in the maximally entangled basis. When the desired measurement outcome is obtained, all the

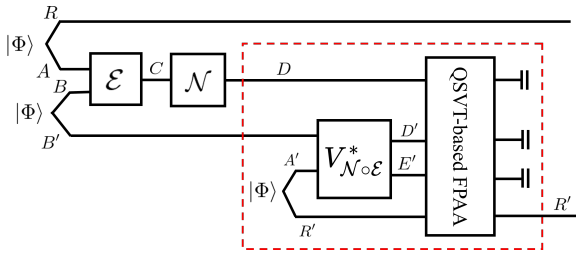


Figure 1: A diagram of the generalized YK decoder in our setting. The decoder is indicated by the dashed box. In the decoding protocol with post-selection, the QSVT-based FPAA is replaced by the measurement $\mathcal{M}$ on DD' and the partial trace over E'.

"information" in D is transferred to D' and then, the effect of the noise is canceled by the emulated inverse that was applied in advance. As a result, under the condition that the post-selection is successful, the MES between R and R' is obtained, i.e., the recovery of quantum information is succeeded.

In fact, we can show that the fidelity between the MES $\Phi^{RR'}$ and the state after the post-selection $\zeta_{\rm succ}^{RR'}$ is given by

$$\mathrm{F}\big(\zeta_{\rm succ}^{RR'}, \Phi^{RR'}\big) = \frac{1}{d_A} 2^{H_2(RE)_\omega - H_2(E)_\omega}. \tag{2}$$

We observe from Eq. (2) that, if $\omega^{RE} \approx \pi^R \otimes \omega^E$, which is nothing but the decoupling condition, $\mathrm{F}\big(\zeta_{\rm succ}^{RR'}, \Phi^{RR'}\big) \approx 1$. Hence, when the decoupling condition is satisfied, this protocol succeeds in recovering quantum information. However, the success probability of this protocol is $p_{\rm succ} = 2^{-H_2(RE)_\omega} d_B/d_D$. As this is exponentially small even if the decoupling is satisfied, this protocol fails to recover quantum information in most cases.

The key of our construction is to use the QSVT-based FPAA algorithm [20–22] instead of the measurement $\mathcal{M}$. By doing so, one can amplify the success probability and achieve a decoding quantum channel without post-selection. More concretely, we replace the step 3 in the above protocol with

3'. The QSVT-based FPAA algorithm is applied on $DD'E'R'$, which is characterized by the number t of repetitions of certain unitaries with phases $\phi \in (-\pi, \pi]^t$.

See also Fig. 1. By taking the partial trace over $DD'E'$, the generalized YK decoder $\mathcal{D}_{t,\phi}^{DB'\to R'}$ is constructed. Note that the value of ϕ are independent of $\mathcal{N}\circ\mathcal{E}$ and there exist classical algorithms to compute that in running time $\mathcal{O}\big(\mathrm{poly}(t)\big)$ [30–34]

We emphasize that the QSVT-based FPAA is crucial. If we use the AA algorithm as in the original YK approach, decoding fails due to an issue related to the 'overcook' problem of the AA algorithm. Similarly, the original FPAA algorithm [35–37] does not work, which is elaborated on in our paper [29].

Our main result about the generalized YK decoder is the following. It reveals the trade-off between the circuit complexity $\mathcal{C}(\mathcal{D}_{t,\phi})$ and the recovery error $\Delta(\mathcal{D}_{t,\phi}|\mathcal{N}\circ\mathcal{E})$.

Theorem 1 *Let ω^{RE} be the state on the reference system R and the environment E of the encoder $\mathcal{E}$ and the noisy channel $\mathcal{N}$, and $\lambda_{\min}(\omega^{RE})$ be the non-zero minimum eigenvalue of ω^{RE}. Suppose that there exists a state τ^E such that $\|\omega^{RE} - \pi^R \otimes \tau^E\|_1 \le \epsilon$. For any $\delta \in (0,1]$ and any odd integer t satisfying*

$$t = \Omega\left(\sqrt{\frac{d_D}{d_B \lambda_{\min}(\omega^{RE})}} \log(1/\delta)\right), \tag{3}$$

there exist $\phi \in (-\pi, \pi]^t$ such that the recovery error $\Delta(\mathcal{D}_{t,\phi}|\mathcal{N}\circ\mathcal{E})$ is given by

$$\Delta(\mathcal{D}_{t,\phi}|\mathcal{N}\circ\mathcal{E}) \le \sqrt{\epsilon} + \sqrt{2\delta}, \tag{4}$$

and its circuit complexity $\mathcal{C}\big(\mathcal{D}_{t,\phi}\big)$ satisfies

$$\mathcal{C}\big(\mathcal{D}_{t,\phi}\big) = \mathcal{O}\Big(t\big(\mathcal{C}(V_{\mathcal{N}\circ\mathcal{E}}) + \log(d_D^2 d_E/d_B)\big)\Big), \tag{5}$$

where $\mathcal{C}(V_{\mathcal{N}\circ\mathcal{E}})$ is a circuit complexity of the Stinespring isometry of $\mathcal{N}\circ\mathcal{E}$.

In Theorem 1, the recovery error $\Delta(\mathcal{D}_{t,\phi}|\mathcal{N}\circ\mathcal{E})$ depends on δ as well as the degree of decoupling ϵ. As the circuit complexity is proportional to $\log(1/\delta)$ (see Eqs.(3) and (5)), we can take δ exponentially small without significantly increasing the complexity. Thus, the generalized YK decoder can recover quantum information with small error if the decoupling condition is satisfied.

We observe that the dominant term in the complexity is t which arises from the QSVT-based FPAA algorithm. As Eq. (3) is known to be an optimal order for implementing the AA algorithm [21, 36, 38], the quantum circuit for the generalized YK decoder cannot be significantly improved. Note that the number t in Eq. (3) depends on $\lambda_{\min}(\omega^{RE})$. When the decoupling condition is satisfied, we have $\lambda_{\min}(\omega^{RE}) \approx \lambda_{\min}(\tau^E)/d_A$. Hence, the complexity depends on the minimum non-zero eigenvalue of the state in the environment. For instance, if τ^E is pure and the CMS, we have the complexity with $t = \Theta\big(\sqrt{d_A d_D/d_B}\log(1/\delta)\big)$ and $t = \Theta\big(\sqrt{d_A d_D d_E/d_B}\log(1/\delta)\big)$, respectively.

References

[1] P. Hayden, M. Horodecki, A. Winter, and J. YardA decoupling approach to the quantum capacity, Open Syst. Inf. Dyn. **15**, 7 (2008).

[2] F. Dupuis, *The decoupling approach to quantum information theory*, Ph.D. thesis, University of Montreal (2010).

[3] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, The decoupling theprem, arXiv:1012.6044 (2010).

[4] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. **43**, 4452 (2002).

[5] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. **303**, 2 (2003).

[6] A. Kitaev, Anyons in an exactly solved model and beyond, Ann. Phys. **321**, 2 (2006).

[7] P. Hayden and J. Preskill, Black holes as mirrors: quantum information in random subsystems, J. High Energy Phys. **2007**, 120 (2007).

[8] D. Harlow and P. Hayden, Quantum computation vs. firewalls, J. High Energy Phys. **2013**, 1 (2013).

[9] A. Almheiri, X. Dong, and D. Harlow, Bulk locality and quantum error correction in AdS/CFT, J. High Energy Phys. **2015**, 1 (2015).

[10] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence, J. High Energy Phys. **2015**, 1 (2015).

[11] P. Hosur, X.-L. Qi, D. A. Roberts, and B. Yoshida, Chaos in quantum channels, J. High Energy Phys. **2016**, 1 (2016).

[12] D. A. Roberts and B. Yoshida, Chaos and complexity by design, J. High Energy Phys. **2017**, 1 (2017).

[13] Y. Nakata, E. Wakakuwa, and M. Koashi, Black holes as clouded mirrors: the Hayden-Preskill protocol with symmetry, Quantum **7**, 928 (2023).

[14] Y. Nakata and M. Tezuka, Hayden-Preskill recovery in Hamiltonian systems, Phys. Rev. Res. **6**, L022021 (2024).

[15] D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, Commun. Math Phys. **105**, 123 (1986).

[16] D. Petz, Sufficiency of channels over von Neumann algebras, Q. J. Math. **39**, 97 (1988).

[17] H. Barnum and E. Knill, Reversing quantum dynamics with near-optimal quantum and classical fidelity, J. Math. Phys. **43**, 2097 (2002).

[18] A. Gilyén, S. Lloyd, I. Marvian, Y. Quek, and M. M. Wilde, Quantum algorithm for Petz recovery channels and pretty good measurements, Phys. Rev. Lett. **128**, 220502 (2022).

[19] B. Yoshida and A. Kitaev, Efficient decoding for the Hayden-Preskill protocol, arXiv:1710.03363 (2017).

[20] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, Quantum Singular Value Transformation and beyond: Exponential Improvements for Quantum Matrix Arithmetics, in *Proc. of the 51st ACM SIGACT STOC*, ACM (2019), pp. 193–204.

[21] A. Gilyén, *Quantum Singular Value Transformation & Its Algorithmic Applications*, Ph.D. thesis, University of Amsterdam (2019).

[22] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, Grand Unification of Quantum Algorithms, PRX Quantum **2**, 040203 (2021).

[23] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem, IEEE Trans. Inf. Theory **48**, 2637 (2002).

[24] M. M. Wilde, From classical to quantum Shannon theory, arXiv:1106.1445 (2011).

[25] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels, IEEE Trans. Inf. Theory **60**, 2926 (2014).

[26] S. Lloyd, Capacity of the noisy quantum channel, Phys. Rev. A **55**, 1613 (1997).

[27] P. W. Shor, *The quantum channel capacity and coherent information*, Lecture notes, MSRI Workshop on Quantum Computation (2002).

[28] I. Devetak, The private classical capacity and quantum capacity of a quantum channel, IEEE Trans. Inf. Theory **51**, 44 (2005).

[29] T. Utsumi and Y. Nakata, Explicit decoders using quantum singular value transformation, arXiv:2405.06051 (2024).

[30] J. Haah, Product Decomposition of Periodic Functions in Quantum Signal Processing, Quantum **3**, 190 (2019).

[31] R. Chao, D. Ding, A. Gilyén, C. Huang, and M. Szegedy, Finding Angles for Quantum Signal Processing with Machine Precision., arXiv: Quantum Physics (2020).

[32] Y. Dong, X. Meng, K. B. Whaley, and L. Lin, Efficient phase-factor evaluation in quantum signal processing, Phys. Rev. A **103**, 042419 (2021).

[33] L. Lin, Lecture Notes on Quantum Algorithms for Scientific Computation, arXiv:2201.08309 (2022).

[34] K. Mizuta and K. Fujii, Recursive Quantum Eigenvalue/Singular-Value Transformation: Analytic Construction of Matrix Sign Function by Newton Iteration, arXiv:2304.13330 (2023).

[35] L. K. Grover, Fixed-point quantum search, Phys. Rev. Lett. **95**, 150501 (2005).

[36] T. J. Yoder, G. H. Low, and I. L. Chuang, Fixed-point quantum search with an optimal number of queries, Phys. Rev. Lett. **113**, 210501 (2014).

[37] B. Yan, S. Wei, H. Jiang, H. Wang, Q. Duan, Z. Ma, and G.-L. Long, Fixed-point oblivious quantum amplitude-amplification algorithm, Sci. Rep. **12**, 14339 (2022).

[38] S. Aaronson and P. Christiano, Quantum money from hidden subspaces, in *Proc. of the 44th ACM STOC*, ACM (2012), pp. 41–60.

YITP-24-60

Explicit decoders using quantum singular value transformation

Takeru Utsumi[1,*] and Yoshifumi Nakata[2,†]

[1] *Graduate School of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan.*
[2] *Yukawa Institute for Theoretical Physics, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan.*

(Dated: May 13, 2024)

Recovering quantum information from a noisy quantum system is one of the central challenges in quantum information science and fundamental physics. The key to this goal is explicitly constructing a decoder. In this paper, we provide two explicit decoding quantum circuits that are both capable of recovering quantum information when a decoupling condition is satisfied, i.e., when quantum information is in principle recoverable. The decoders are constructed by using the fixed-point amplitude amplification algorithm based on the quantum singular value transformation, which significantly extends an approach by Yoshida and Kitaev in a specific noise model to general situations. We also show that the proposed decoding circuits reduce the computational cost compared to a previously known explicit decoder. Our constructions not only show an intriguing intersection between decoders and quantum algorithms but also reveal the power of an algorithmic approach to recovering quantum information.

I. INTRODUCTION

Recovering quantum information from a noisy system is crucial for transmitting quantum information over noisy quantum channels. A standard technique is to use quantum error correction, in which quantum information is encoded before the system experiences noise and is decoded afterward. Recovery of quantum information is also of significant importance in fundamental physics to understanding complicated quantum many-body phenomena. By analyzing the recovery of quantum information, various novel insights into the black hole information paradox [1–3], the AdS/CFT correspondence [4, 5], topological orders [6–8], and quantum chaos [9–11], have been obtained.

The recovery of quantum information is commonly investigated by the *decoupling* approach [12–14]. Decoupling refers to the situation, where the environmental system of the noisy channel is decoupled from the reference system that keeps track of the quantum information, and is necessary and sufficient for the quantum information to be recoverable. While decoupling provides a useful theoretical approach to the problem of information recovery without referring to the recovery process, from a practical viewpoint, it is important to explicitly construct a recovery protocol, or a decoder. An explicit decoder also advances the understanding of the recovery process of quantum information.

Only a handful of results about explicit constructions of a decoder are known so far [15–17]. A standard explicit decoder is the *Petz recovery* map [18, 19]. While the map was originally introduced in a different context, it is known that the map is applicable to recovering quantum information, resulting in a close-to-optimal recovery error [15]. However, a quantum circuit for implementing the Petz recovery map requires high computational complexity [20]. Hence, simplifications of the Petz recovery map, focusing on its use as a decoder, have been studied [21, 22].

Another explicit decoder is the *Yoshida-Kitaev (YK)* decoder [16], which is capable of decoding the so-called Hayden-Preskill (HP) protocol [1]. The HP protocol is a toy model of the qubit-erasure noise with a specific unitary encoding, and has a good interpretation in the black hole information paradox. The YK decoder can decode the HP protocol, and its quantum circuit is explicitly given. The decoder is also of interest from an algorithmic perspective: first a recovery protocol with post-selection by measurement is considered, and then a decoder is constructed by replacing the measurement with a non-trivial use of the *amplitude amplification (AA)* algorithm [23–25], which is for amplifying the success probability. The YK decoder, however, strongly relies on the specific setting of the HP protocol. It is highly non-trivial if such a two-step construction of a decoder using a AA-type algorithm can be extended to more general situations.

In this paper, we explore the use of the AA-type algorithms for recovering quantum information and provide two explicit decoding quantum circuits. One is a generalization of the YK decoder, in which we crucially modify the decoder by replacing the AA algorithm with the fixed-point amplitude amplification (FPAA) algorithm based on the *quantum singular value transformation (QSVT)* [26–28]. Due to the flexibility of the QSVT-based FPAA algorithm, the issues that arise with the AA algorithm can be circumvented, and the generalized YK decoder is applicable to general encoding and noisy channels. The other is a simplification of the Petz

* takeru-utsumi@g.ecc.u-tokyo.ac.jp
† yoshifumi.nakata@yukawa.kyoto-u.ac.jp

recovery map, which we call a *Petz-like* decoder. Similarly to the generalized YK decoder, the simplification is achieved by using the QSVT-based FPAA algorithm, and the Petz-like decoder is also applicable to any situation.

We show that both decoders have high recovery performance in the sense that they succeed in recovering quantum information if the decoupling condition is satisfied. As the decoupling is a necessary and sufficient condition for the information recovery, this immediately implies that quantum information can be recovered by the proposed decoders whenever it is in principle recoverable. Important applications of the decoders are to the independent and identically distributed (i.i.d.) asymptotic setting. In the i.i.d. setting, both decoders with suitably chosen encoders achieve the *quantum capacity* [29–31]. This is also true when the sender and the receiver share entanglement in advance. This situation is called an entanglement-assisted setting. The proposed decoders with suitable encoders achieve the entanglement-assisted quantum capacity [32–34] as well.

Taking advantage of our explicit constructions, we also investigate the circuit complexity of the generalized YK and the Petz-like decoders. While the complexity depends on various factors, the dominant factor is in general the complexity for implementing the QSVT-based FPAA algorithm. We provide a simple criterion for the generalized YK decoder to have smaller complexity than the Petz-like decoder. The criterion is in terms of the number of qubits of the encoded quantum information, the amount of pre-shared entanglement, the number of output qubits of the noisy channel, and also the number of Kraus operators of the channel. It turns out that the generalized YK decoder typically has less complexity when more entanglement is shared in advance. We additionally compare the complexity with the algorithmic implementation of the original Petz recovery map and show that the proposed decoders have smaller complexity in a large parameter region.

This paper is organized as follows. We start with preliminaries in II. Our main results are summarized in III. The proofs of our results are provided in IV. We conclude with a summary and outlooks in V, and provide a derivation of a technical statement in Appendix A.

II. PRELIMINARIES

We here introduce our notation and our setting. We then briefly overview an implicit decoder commonly used in the decoupling approach. We also provide quick overviews of the Petz recovery map and the YK decoder.

A. Notation

Throughout this paper, we denote by $\mathcal{S}(\mathcal{H})$ a set of all quantum states on a Hilbert space $\mathcal{H}$. While we usually denote a pure state by $|\varphi\rangle$, the corresponding density operator is sometimes described as φ, namely, $\varphi = |\varphi\rangle\langle\varphi|$. We use a superscript to represent a system on which operators and maps are defined. For instance, an operator on a system AB and a superoperator from A to B are denoted by φ^{AB} and $\mathcal{T}^{A\to B}$, respectively. The superscript is omitted when it is clear from the context. A reduced density operator on A of φ^{AB} is described as φ^A, i.e., $\varphi^A = \mathrm{Tr}_B\,\varphi^{AB}$, where Tr_B is the partial trace over B.

For an operator M, we denote the complex conjugate and the transpose in a given basis by M^* and M^T, respectively, and denote the Hermitian conjugate by $M^\dagger$. The identity operation is denoted by $\mathbb{I}$ and id for operators and superoperators, respectively. We often omit the identity operators and superoperators for simplicity.

A Hilbert space, such as $\mathcal{H}^{A'}$ or $\mathcal{H}^{\hat{A}}$, is isomorphic to $\mathcal{H}^A$: it has the same dimension, and we fix the same basis as $\mathcal{H}^A$. This applies not only to the system A, but also to any systems, such as $\mathcal{H}^{B'}$ and $\mathcal{H}^{\hat{C}}$. We write the dimension of a Hilbert space $\mathcal{H}$ as d, and for instance, denote by d_A the dimension of $\mathcal{H}^A$.

We omit the symbol of the tensor product between vectors and denote it as $|\varphi\rangle \otimes |\psi\rangle = |\varphi\rangle|\psi\rangle$, for simplicity, when it is clear from the context. We denote by $|\Phi\rangle$ a maximally entangled state (MES) defined in the orthonormal computational basis. For instance, the MES between A and $\hat{A}$ is

$$|\Phi\rangle^{A\hat{A}} = \frac{1}{\sqrt{d_A}} \sum_{i=1}^{d_A} |i\rangle^A |i\rangle^{\hat{A}}, \tag{1}$$

where $\{|i\rangle\}_i$ is the computational basis in A and $\hat{A}$, respectively. Note that a MES in an arbitrary basis can be transformed into the MES in the computational basis by applying an appropriate unitary to one of the local systems. We also denote the completely mixed state (CMS) by π, such as $\pi^A = \mathbb{I}^A/d_A$.

The circuit complexity of $\mathcal{T}$ is denoted by $\mathcal{C}(\mathcal{T})$. It is the minimum total number of one- and two-qubit unitary gates required to perform $\mathcal{T}$ with ancillae polynomial in qubits.

For a matrix M, the trace norm is defined by $\|M\|_1 := \mathrm{Tr}\left[\sqrt{M^\dagger M}\,\right]$. The trace norm has the contraction property such that for $\varphi^{AB} \in \mathcal{S}(\mathcal{H}^{AB})$ and $\psi^{AB} \in \mathcal{S}(\mathcal{H}^{AB})$,

$$\|\varphi^A - \psi^A\|_1 \le \|\varphi^{AB} - \psi^{AB}\|_1. \tag{2}$$

The fidelity between $\varphi \in \mathcal{S}(\mathcal{H})$ and $\psi \in \mathcal{S}(\mathcal{H})$ is defined as $\mathrm{F}(\varphi,\psi) := \left\|\sqrt{\varphi}\sqrt{\psi}\right\|_1^2$. The fidelity is rephrased using the purified states of φ and ψ as

$$\mathrm{F}(\varphi^A,\psi^A) = \max_V \left|\langle\varphi|^{AC} V^{B\to C} |\psi\rangle^{AB}\right|^2, \tag{3}$$

where the maximization is taken over all isometries $V^{B\to C}$. Here, we supposed $d_C \ge d_B$ without loss of generality. This is called the *Uhlmann's theorem* [35]. The trace norm and the fidelity are related by the Fuchs-van

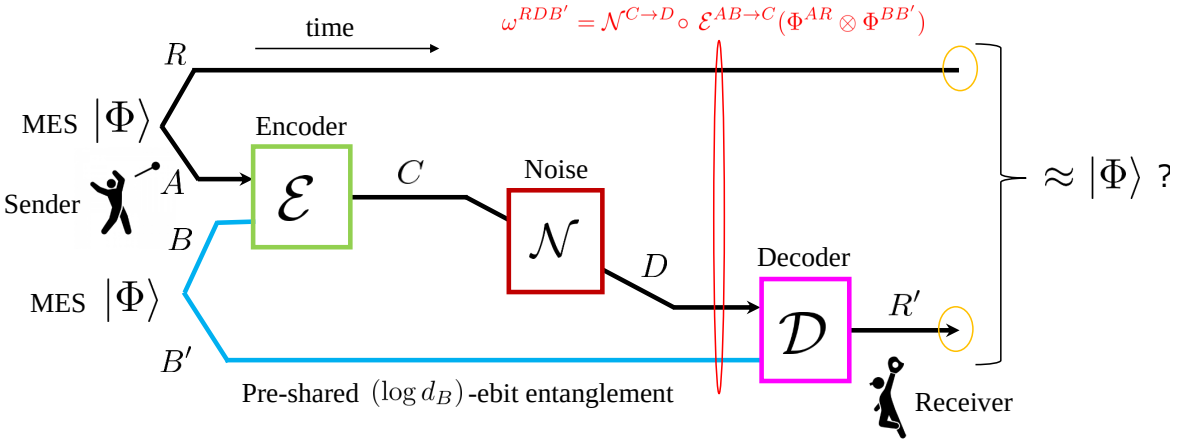


FIG. 1. A diagram of our setting. Time flows from left to right. The boxes represent quantum channels. The purpose of the sender and the receiver is to transmit quantum information via noisy channel $\mathcal{N}^{C\to D}$, which is equivalent to preserving the maximally entangled state between A and R. They may share $(\log d_B)$-ebit entanglement in advance, which is used during the encoding and decoding operations.

de Graaf inequalities [36, 37]

$$1 - \sqrt{\mathrm{F}(\varphi,\psi)} \le \frac{1}{2}\|\varphi - \psi\|_1 \le \sqrt{1 - \mathrm{F}(\varphi,\psi)}. \tag{4}$$

We use the quantum collision entropy. For $\varphi^A \in \mathcal{S}(\mathcal{H}^A)$ it is given by

$$H_2(A)_\varphi = -\log \mathrm{Tr}[(\varphi^A)^2]. \tag{5}$$

This satisfies $0 \le H_2(A)_\varphi \le d_A$.

B. Our setting

We consider the following setting. Suppose that a sender aims to transmit $(\log d_A)$-qubit quantum information using a given noisy channel $\mathcal{N}^{C\to D}$ and possibly a pre-shared entanglement $|\Phi\rangle^{BB'}$, where B and B' are with the sender and receiver, respectively. When they share no entanglement, we set $d_B = 1$. The sender encodes the system A with B using an encoding channel $\mathcal{E}^{AB\to C}$. The qubits in C are then transmitted to the receiver through the noisy channel $\mathcal{N}^{C\to D}$. The receiver obtains the output system D of the noisy channel and applies a recovery channel, i.e., a decoder $\mathcal{D}^{DB'\to R'}$ onto the system DB'. For simplicity, we denote by $\mathcal{F}^{AB\to D}$ the composite channel $\mathcal{N}^{C\to D} \circ \mathcal{E}^{AB\to C}$. The main concern in this paper is to explicitly construct a decoder $\mathcal{D}^{DB'\to R'}$ for a given channel $\mathcal{F}^{AB\to D}$. We assume that the descriptions of the encoding map $\mathcal{E}$ and the noisy channel $\mathcal{N}$ are known, so that the decoder can depend on their details.

Following the convention, we introduce a reference system R isomorphic to A with $d_R = d_A$, and prepare the systems A and R to be in a MES $|\Phi\rangle^{AR}$. We denote by $\omega^{RDB'}$ the state just before the decoder is applied:

$$\omega^{RDB'} := \mathcal{F}^{AB\to D}(\Phi^{AR} \otimes \Phi^{BB'}). \tag{6}$$

See also Fig. 1. The recovery error of quantum information by a decoder $\mathcal{D}^{DB'\to R'}$ in this protocol is defined as [38]

$$\Delta(\mathcal{D}|\,\mathcal{F}) := \frac{1}{2}\| \Phi^{RR'} - \mathcal{D}^{DB'\to R'}(\omega^{RDB'}) \|_1. \tag{7}$$

C. Decoupling and the Uhlmann decoder

A standard approach to evaluating the recovery error is to estimate how much quantum information is leaked to an "environment" of the noisy channel. This is specifically quantified by the degree of decoupling.

We denote by $V_{\mathcal{F}}^{AB\to ED}$ a Stinespring isometry of the channel $\mathcal{F}^{AB\to D} = \mathcal{N}^{C\to D} \circ \mathcal{E}^{AB\to C}$ by an environment E. That is, the channel $\mathcal{F}^{AB\to D}$ is represented as

$$\mathcal{F}^{AB\to D}(\,\cdot\,) = \mathrm{Tr}_E\left[\, V_{\mathcal{F}}^{AB\to ED}(\,\cdot\,)(V_{\mathcal{F}}^{AB\to ED})^\dagger \,\right]. \tag{8}$$

For convenience, we also introduce a purified state of $\omega^{RDB'}$ in Eq. (6) as

$$|\omega\rangle^{REDB'} := V_{\mathcal{F}}^{AB\to ED}|\Phi\rangle^{AR}|\Phi\rangle^{BB'}. \tag{9}$$

The following is called the decoupling approach.

Proposition 1 (Decoupling approach [12–14])**.** *Suppose $|\omega\rangle^{REDB'}$ is a pure state. If there exists a state τ^E such that $\|\omega^{RE} - \pi^R \otimes \tau^E\|_1 \le \epsilon$, then there exists a CPTP map $\mathcal{D}_{\mathrm{Uhlmann}}^{DB'\to R'}$ that satisfies*

$$\frac{1}{2}\| \Phi^{RR'} - \mathcal{D}_{\mathrm{Uhlmann}}^{DB'\to R'}(\omega^{RDB'}) \|_1 \le \sqrt{\epsilon}. \tag{10}$$

The proof of this proposition follows from Eqs. (2), (3), and (4). See, e.g., [12–14]. We refer to the decoder

$\mathcal{D}_{\mathrm{Uhlmann}}$ as the *Uhlmann decoder*. The condition that there exists τ^E such that

$$\|\omega^{RE} - \pi^R \otimes \tau^E\|_1 \leq \epsilon, \tag{11}$$

is known as a *decoupling condition*. While the decoupling approach implicitly indicates the existence of a decoder when the decoupling condition is satisfied, it does not provide an explicit procedure to construct a decoder. For this reason, all the details about decoders, such as the computational cost for the construction, are open.

The decoupling approach is particularly strong in the study of the maximum possible rate for transmitting quantum information. Let N be the number of uses of a noisy channel $\mathcal{N}^{C\to D}$ to transmit quantum information. The transmission rate for a fixed N is defined by $\mathrm{R}_N \coloneqq \frac{1}{N}\log d_A$. An asymptotically-achievable rate is then defined by $\mathrm{R} \coloneqq \lim_{N\to\infty}\mathrm{R}_N$ under the assumption that there exists a sequence of pairs of an encoder and a decoder such that the recovery error tends to zero as $N \to \infty$. The supremum of asymptotically-achievable rates for the channel is called the quantum capacity $Q(\mathcal{N})$. It is known by the technique of the *random encoding* that if $\mathrm{R} < Q(\mathcal{N})$, there exists an isometric encoder that asymptotically achieves decoupling, i.e., $\epsilon \to 0$ [39]. Hence, the recovery error of the Uhlmann decoder also asymptotically tends to zero. That is, the Uhlmann decoder with suitably chosen encoders achieves the quantum capacity.

D. Petz recovery map

One of the explicit decoders we use is the Petz recovery map [18, 19], which has been a useful tool in quantum information theory and has been intensely studied [33, 40]. The Petz recovery map is developed from a quantum analog of Bayes theorem based on the idea that there can be a reverse channel that recovers an effect of noise. The general form of the Petz recovery map is determined by a map $\mathcal{T}$ and a reference state σ, and given by

$$\begin{aligned}&\mathcal{P}^{B\to A}_{\sigma,\mathcal{T}}(\,\cdot\,)\\ &= (\sigma^A)^{\frac{1}{2}}(\mathcal{T}^{A\to B})^\dagger\big([\mathcal{T}(\sigma^A)]^{-\frac{1}{2}}(\,\cdot\,)[\mathcal{T}(\sigma^A)]^{-\frac{1}{2}}\big)(\sigma^A)^{\frac{1}{2}},\end{aligned} \tag{12}$$

where $(\mathcal{T}^{A\to B})^\dagger$ is the adjoint map of $\mathcal{T}^{A\to B}$ with respect to the Hilbert-Schmidt inner product. The Petz recovery map is composed of three CP maps:

$$(\,\cdot\,) \to [\mathcal{T}(\sigma^A)]^{-\frac{1}{2}}(\,\cdot\,)[\mathcal{T}(\sigma^A)]^{-\frac{1}{2}}, \tag{13}$$
$$(\,\cdot\,) \to (\mathcal{T}^{A\to B})^\dagger(\,\cdot\,), \tag{14}$$
$$(\,\cdot\,) \to (\sigma^A)^{\frac{1}{2}}(\,\cdot\,)(\sigma^A)^{\frac{1}{2}}. \tag{15}$$

It achieves the perfect recovery for the reference state σ^A, i.e., $\mathcal{P}^{B\to A}_{\sigma,\mathcal{T}}(\mathcal{T}^{A\to B}(\sigma^A)) = \sigma^A$.

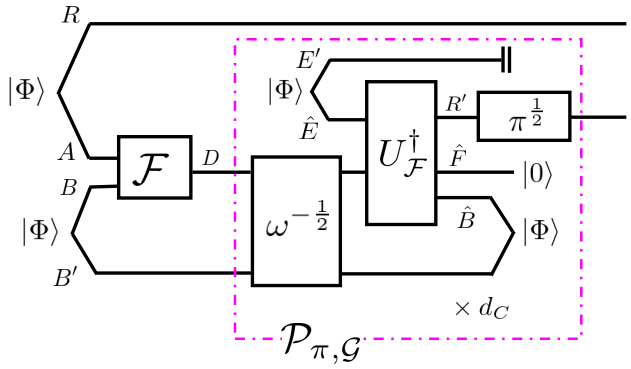


FIG. 2. A diagram of the Petz recovery map applied to our setting. The dash-dotted box corresponds to the Petz recovery map $\mathcal{P}_{\pi,\mathcal{G}}$ given in Eq. (19). The boxes of $(\omega^{B'D})^{-1/2}$ and $(\pi^{R'})^{1/2}$ represent that $(\,\cdot\,) \to (\omega^{B'D})^{-1/2}(\,\cdot\,)(\omega^{B'D})^{-1/2}$ and $(\,\cdot\,) \to (\pi^{R'})^{1/2}(\,\cdot\,)(\pi^{R'})^{1/2}$, respectively. The double vertical lines represent that the qubits of that system are traced out.

For the recovery error of the Petz recovery map, the following is known, stating that, if there exists a decoder that recovers information with a small error, the Petz recovery map also recovers it with a small error.

Proposition 2 (Barnum-Knill's theorem [15])**.** *For any state ρ^A and any channel $\mathcal{T}^{A\to B}$, it holds that*

$$\begin{aligned}&\mathrm{F}\big(\rho^{AR}, \mathcal{P}^{B\to A}_{\rho,\mathcal{T}}\circ\mathcal{T}^{A\to B}(\rho^{AR})\big)\\ &\qquad\geq \big[\max_{\mathcal{R}}\mathrm{F}\big(\rho^{AR}, \mathcal{R}^{B\to A}\circ\mathcal{T}^{A\to B}(\rho^{AR})\big)\big]^2,\end{aligned} \tag{16}$$

where $\rho^{AR} = |\rho\rangle\langle\rho|^{AR}$ is a purified state of ρ^A. The maximum is taken over all quantum channels $\mathcal{R}^{B\to A}$.

To apply the Petz recovery map to our setting, let F be the system such that $ABF = ED$, and a unitary $U^L_{\mathcal{F}}$ be defined by

$$V^{AB\to ED}_{\mathcal{F}} = U^L_{\mathcal{F}}|0\rangle^F, \tag{17}$$

where $L = ABF = ED$. Using this unitary, Eq. (8) is rephrased as

$$\mathcal{F}^{AB\to D}(\,\cdot\,) = \mathrm{Tr}_E\left[U^L_{\mathcal{F}}(\,\cdot\,\otimes|0\rangle\langle 0|^F)(U^L_{\mathcal{F}})^\dagger\right]. \tag{18}$$

We use $\mathcal{G}^{A\to DB'}(\,\cdot\,) \coloneqq \mathcal{F}^{AB\to D}(\,\cdot\,\otimes\Phi^{BB'})$ and fix the reference state to be the CMS π^A. The explicit form of the Petz recovery map in our setting is then given by

$$\begin{aligned}&\mathcal{P}^{DB'\to R'}_{\pi,\mathcal{G}}(\omega^{RDB'})\\ &= d_E(\pi^{R'})^{1/2}\langle\Phi|^{\hat{B}B'}\langle 0|^{\hat{F}}(U^{\hat{L}}_{\mathcal{F}})^\dagger\big[(\omega^{DB'})^{-1/2}\omega^{RDB'}\\ &\qquad(\omega^{DB'})^{-1/2}\otimes\Phi^{\hat{E}E'}\big]U^L_{\mathcal{F}}|\Phi\rangle^{\hat{B}B'}|0\rangle^{\hat{F}}(\pi^{R'})^{1/2},\end{aligned} \tag{19}$$

where $\hat{L}$ is equal to $R'\hat{B}\hat{F} = \hat{E}D$. See also the diagram in Fig. 2.

By combining Proposition 2 with Proposition 1 and the Fuchs-van de Graaf inequalities, we derive the following statement, which relates the recovery error of the Petz recovery map $\mathcal{P}_{\pi,\mathcal{G}}^{DB'\to R'}$ against $\mathcal{F}^{AB\to D}$ to the decoupling condition: if there exists a state τ^E such that $\|\omega^{RE} - \pi^R \otimes \tau^E\|_1 \leq \epsilon$, then the recovery error of the Petz recovery map in the above setting is given by

$$\Delta(\,\mathcal{P}_{\pi,\mathcal{G}}\,|\,\mathcal{F}\,) \leq 2\,\epsilon^{1/4}. \tag{20}$$

As discussed in II C, the decoupling is asymptotically achieved by an appropriately chosen encoder. Since the upper bound on the recovery error of $\mathcal{P}_{\pi,\mathcal{G}}$ tends to zero with such an encoder, hence, the Petz recovery map also archives the quantum capacity.

An algorithmic implementation of the Petz recovery map using the QSVT is provided in [20]. Using the algorithm, one can obtain an explicit decoder. However, its circuit complexity is generally inefficient.

E. Yoshida-Kitaev decoder in the Hayden-Preskill protocol

The YK decoder [16] was proposed for recovering quantum information in the toy model of the black hole information paradox, i.e., the HP protocol [1]. The HP protocol formulates the information paradox based on the qubit-erasure noise with a restriction that the encoding operation is given by a unitary dynamics of a black hole, typically assumed to be sufficiently random. More specifically, the encoder $\mathcal{E}$ and the noisy channel $\mathcal{N}$ in Fig. 1 are given by a random unitary and the partial trace over a subsystem E of C, respectively, where $AB = C$. It is further assumed that the receiver, i.e., the person who applies a decoder, knows what unitary was applied and which qubits were traced over.

The YK decoder provides an explicit algorithm for decoding the HP protocol, and is based on the idea of "emulating" the inverse dynamics of the encoding unitary and the erasure noise in the receiver's local system. The receiver then measures the output of the erasure noise and the corresponding "emulated" output in the maximally-entangled basis. If a desired outcome were obtained, the emulated output becomes as if it were in the quantum state same as the state of the actual input of noise. In this case, the effect of the erasure noise is canceled by the emulated inverse in the local system, and the receiver succeeds in recovering quantum information. While this protocol does not succeed with certainty as it requires post-selection, the probability of obtaining the desired outcome can be amplified by a non-trivial use of the AA algorithm, completing the construction of the YK decoder.

Although the YK decoder provides an insight that the two-step approach, i.e., the approach of considering the protocol with post-selection and combining it with the AA algorithm, may be useful for constructing a decoder, the reason for the decoder to work strongly relies on the specific properties of the HP protocol. In particular, it is crucial that, when the decoupling condition is met, the unitary encoding and the erasure noise make the eigenvalues of the quantum state on the reference R and the environment E of the noise completely uniform, namely, $\omega^{RE} \approx \pi^R \otimes \pi^E$. Without this uniform property, the AA algorithm in the YK decoder does not work. Since the uniform condition is not satisfied for general encoding operations and noises, extending the YK decoder to a general situation is highly non-trivial. We comment on this point in more detail in the proof of our main result in IV A 2.

III. MAIN RESULTS

In this section, we summarize our results. We provide explicit quantum circuit constructions of two decoders and evaluate their performance. One is the generalized Yoshida-Kitaev decoder presented in III A, and the other is the Petz-like decoder given in III B. We investigate the complexity of the decoders in III C and III D.

Both decoders are constructed by the two-step approach similar to the YK decoder: we first consider a protocol with post-selection and then transform the protocol into the one without post-selection. Unlike the YK decoder, however, we use the QSVT-based FPAA algorithm [26–28] instead of the standard AA algorithm. The QSVT-based FPAA algorithm is a slight extension of the standard FPAA algorithm [41–43] and is crucial for circumventing the issues arisen when the standard AA algorithm is used.

A. Generalized Yoshida-Kitaev decoder

We below propose a generalization of the YK decoder. In III A 1, we investigate a decoding protocol with post-selection that works for general encoding maps and noisy channels. We then show in III A 2 that the protocol can be transformed into a decoder using the QSVT-based FPAA algorithm.

1. *Decoding protocol with post-selection*

The decoding protocol with post-selection consists of the following three steps. See Fig. 3 as well.

1. The receiver prepares ancilla qubits in the system $A'R'$, and then generates a MES $\Phi^{A'R'}$, which is a copy of the MES Φ^{AR}.

2. The receiver applies an isometry $(V_{\mathcal{F}}^{A'B'\to E'D'})^*$ onto $A'B'$, where $V_{\mathcal{F}}^{AB\to ED}$ is a Stinespring isometry of $\mathcal{F}^{AB\to D}$. The complex conjugate is taken in the computational basis.

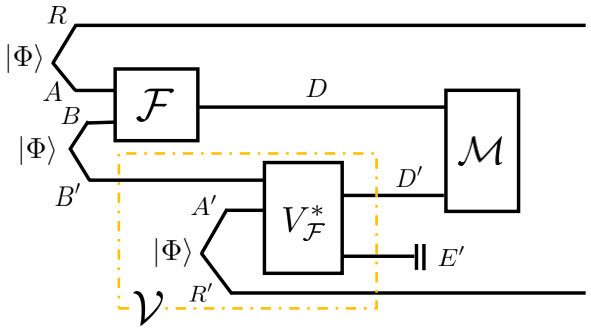


FIG. 3. A diagram of the protocol with post-selection for the generalized YK decoder. The double vertical lines represent that the qubits of that system are traced out. The dash-dotted box corresponds to the isometry map $\mathcal{V}^{B'\to D'E'R'}$ defined in Eq. (21).

3. The receiver performs a binary measurement $\mathcal{M} \coloneqq \{|\Phi\rangle\langle\Phi|^{DD'}, \mathbb{I}^{DD'} - |\Phi\rangle\langle\Phi|^{DD'}\}$ on DD'. When the former result of the measurement $\mathcal{M}$ is obtained, this protocol succeeds.

In this protocol, all the systems with a prime, i.e., A', B' R', D', and E', in addition to the output system D of the channel $\mathcal{F}$ are in the hands of the receiver. Hence, the above protocol can be executed by the receiver.

The Stinespring dilation $V_{\mathcal{F}}^{AB\to CD}$ in the step 2 is not uniquely determined from a given channel $\mathcal{F}^{AB\to D}$: the dilation has a freedom of applying additional isometries on the environment E. However, the protocol works for any choice of $V_{\mathcal{F}}^{AB\to CD}$. Hence, the receiver can choose arbitrary Stinespring dilation of the channel $\mathcal{F}^{AB\to D}$.

For future use, we denote the operation up to the step 2 of the above protocol by an isometry map $\mathcal{V}^{B'\to D'E'R'}$. That is,

$$\begin{aligned}&\mathcal{V}^{B'\to D'E'R'}(\,\cdot\,)\\ &\quad\coloneqq (V_{\mathcal{F}}^{A'B'\to E'D'})^*(\,\cdot\,\otimes\Phi^{A'R'})(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}.\end{aligned} \tag{21}$$

We denote by p_{succ} and ζ_{succ} the success probability and the output state with the success in the step 3, respectively. The reduced state on RR' of ζ_{succ} is given by

$$\zeta_{\mathrm{succ}}^{RR'} = \mathrm{Tr}_{DD'E'}\Big[\frac{1}{p_{\mathrm{succ}}}|\Phi\rangle\langle\Phi|^{DD'}\mathcal{V}^{B'\to D'E'R'}(\omega^{RDB'})\Big]. \tag{22}$$

In IV A 1, we compute p_{succ} and the fidelity between $\zeta_{\mathrm{succ}}^{RR'}$ and $\Phi^{RR'}$, and then obtain

$$p_{\mathrm{succ}} = \frac{d_B}{d_D}2^{-H_2(RE)_\omega}, \tag{23}$$

$$\mathrm{F}\big(\zeta_{\mathrm{succ}}^{RR'}, \Phi^{RR'}\big) = \frac{1}{d_A}2^{H_2(RE)_\omega - H_2(E)_\omega}. \tag{24}$$

This implies that if ω^{RE} decouples as $\omega^{RE}\approx\pi^R\otimes\omega^E$, the fidelity after post-selection becomes $\mathrm{F}\big(\zeta_{\mathrm{succ}}^{RR'}, \Phi^{RR'}\big)\approx 1$. Namely, the recovery of the MES is succeeded if the measurement is successful under the decoupling is satisfied.

2. Construction of the generalized YK decoder

We now explain how the above decoding protocol with post-selection can be transformed into a decoder without post-selection by the QSVT-based FPAA algorithm. The QSVT is a quantum algorithm described by a unitary $G_{t,\phi}$ with parameters $\phi = (\phi_1, \phi_2, \dots, \phi_t) \in (-\pi, \pi]^t$ and $t \in \mathbb{N}$. It is to apply a polynomial transformation to the singular values of a linear operator embedded in a submatrix of the unitary [26–28]. The polynomial is determined by the phase sequence ϕ, and the integer t which corresponds to the degree of the polynomial. The key point of the QSVT-based FPAA algorithm is that, by choosing an appropriate t and ϕ for approximating the sign function, we amplify the success probability of the measurement $\mathcal{M}$ in the step 3 to nearly unity.

To elucidate the structure of the unitary $G_{t,\phi}$ in our setting, we introduce two projectors:

$$\begin{aligned}\Pi_1^{D'E'R'} &\coloneqq (V_{\mathcal{F}}^{A'B'\to E'D'})^*\\ &\qquad(\mathbb{I}^{B'}\otimes|\Phi\rangle\langle\Phi|^{A'R'})(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}},\end{aligned} \tag{25}$$

$$\Pi_2^{DD'} \coloneqq |\Phi\rangle\langle\Phi|^{DD'}, \tag{26}$$

and unitaries:

$$W_m(\theta) \coloneqq e^{i\theta(2\Pi_m - \mathbb{I})}, \tag{27}$$

where $m = 1, 2$ and $\theta \in (-\pi, \pi]$. Let $W_{t,\phi}^{DD'E'R'}$ be a unitary given by

$$\begin{aligned}&W_{t,\phi}^{DD'E'R'}\\ &\quad\coloneqq W_2(\phi_t)^{DD'}\prod_{j=1}^{(t-1)/2}W_1(\phi_{2j})^{D'E'R'}W_2(\phi_{2j-1})^{DD'}.\end{aligned} \tag{28}$$

The unitary $G_{t,\phi}^{DD'E'R'H}$ is then defined by

$$\begin{aligned}G_{t,\phi}^{DD'E'R'H} &\coloneqq W_{t,\phi}^{DD'E'R'}\otimes|+\rangle\langle+|^H\\ &\qquad + W_{t,-\phi}^{DD'E'R'}\otimes|-\rangle\langle-|^H,\end{aligned} \tag{29}$$

where H is a single-qubit auxiliary system.

When we construct the generalized YK decoder, the measurement step 3 in the previous protocol is replaced with the application of $G_{t,\phi}$ as follows.

3'. The receiver prepares an auxiliary single-qubit state $|0\rangle^H$ in a system H, and then applies a unitary $G_{t,\phi}^{DD'E'R'H}$, with appropriate t and ϕ to approximate the sign function.

All together, the decoder is given by $\mathcal{D}_{t,\phi}^{DB'\to R'}$ as

$$\begin{aligned}&\mathcal{D}_{t,\phi}^{DB'\to R'}(\,\cdot\,)\\ &\quad\coloneqq \mathrm{Tr}_{DD'E'H}\big[G_{t,\phi}^{DD'E'R'H}\big(\mathcal{V}^{B'\to D'E'R'}(\,\cdot\,)\\ &\qquad\qquad\otimes|0\rangle\langle0|^H\big)(G_{t,\phi}^{DD'E'R'H})^\dagger\big],\end{aligned} \tag{30}$$

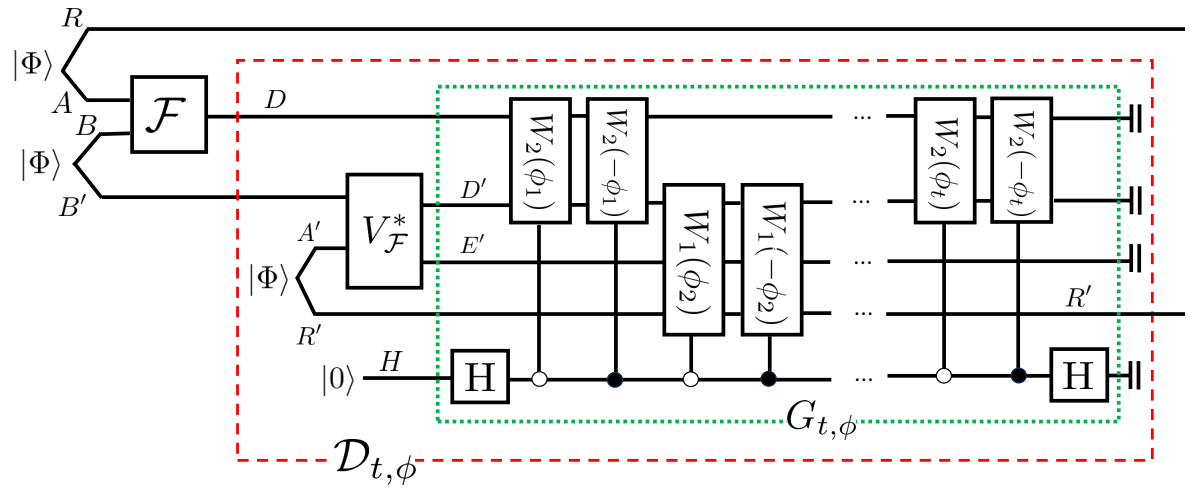


FIG. 4. A diagram of the generalized YK decoder. Open circles imply that the gates are controlled by $|0\rangle$, while closed circles indicate the ones controlled by $|1\rangle$. The gate H is the single-qubit Hadamard gate. The red dashed and green dotted boxes correspond to the generalized YK decoder $\mathcal{D}_{t,\phi}$ defined in Eq. (30), and the unitary $G_{t,\phi}$ by the QSVT-based FPAA algorithm given in Eq. (29), respectively.

where $\mathcal{V}^{B'\to D'E'R'}$ is defined in Eq. (21). See Fig. 4 as well.

The following theorem holds.

Theorem 3 (Performance of the generalized YK decoder)**.** *For a given channel $\mathcal{F}^{AB\to D}$, let $\bar{\mathcal{F}}^{AB\to E}$ be a complementary channel of $\mathcal{F}^{AB\to D}$, ω^{RE} be given by*

$$\omega^{RE} = \bar{\mathcal{F}}^{AB\to E}(\Phi^{AR}\otimes\pi^{B}), \tag{31}$$

and $\lambda_{\min}(\omega^{RE})$ be the non-zero minimum eigenvalue of ω^{RE}. Suppose that there exists a state τ^{E} such that $\|\omega^{RE}-\pi^{R}\otimes\tau^{E}\|_1\le\epsilon$. For any $\delta\in(0,1]$, there exist $t\in\mathbb{N}$ and $\phi=(\phi_1,\phi_2,\ldots,\phi_t)\in(-\pi,\pi]^t$ such that the recovery error $\Delta(\,\mathcal{D}_{t,\phi}\,|\,\mathcal{F}\,)$ of the generalized YK decoder $\mathcal{D}^{DB'\to R'}_{t,\phi}$ is given by

$$\Delta(\,\mathcal{D}_{t,\phi}\,|\,\mathcal{F}\,)\le\sqrt{\epsilon}+\sqrt{2\delta}, \tag{32}$$

where t is an odd integer satisfying

$$t=\Theta\left(\sqrt{\frac{d_D}{d_B\lambda_{\min}(\omega^{RE})}}\,\log(1/\delta)\right). \tag{33}$$

The circuit complexity of the decoder $\mathcal{D}^{DB'\to R'}_{t,\phi}$ is

$$\mathcal{C}\big(\mathcal{D}_{t,\phi}\big)=\mathcal{O}\Big(t\,\big(\mathcal{C}(U_{\mathcal{F}})+\log(d_D^2d_Ed_B)\big)\Big), \tag{34}$$

and the number of ancilla qubits is $\mathcal{O}\big(\log(d_D^2d_Ed_B)\big)$. Here, $\mathcal{C}(U_{\mathcal{F}})$ is a circuit complexity of a unitary $U^L_{\mathcal{F}}$ such that $U^L_{\mathcal{F}}|0\rangle^F$ is a Stinespring isometry of $\mathcal{F}^{AB\to D}$, and $L=ABF=ED$.

Theorem 3 shows in Eq. (32) that the recovery error is dependent on ϵ and δ. While ϵ is an upper bound on the degree of decoupling and depends only on the channel $\mathcal{F}$, δ can be chosen arbitrarily small. One may hence think that the limit $\delta\to 0$ should be taken. This is true if the recovery error is the only concern. However, there is a trade-off relation between the recovery error and the circuit complexity. The parameter δ is to characterize the trade-off. In fact, Eqs. (33) and (34) show that the circuit complexity of the generalized YK decoder depends on δ, such as $\log(1/\delta)$. Hence, the complexity increases if one wishes to achieve small errors. This trade-off is naturally expected due to the nature of the AA-type algorithm. Note that the dependence of the complexity on $1/\delta$ is only logarithmic, and so, exponentially small δ is feasible.

One needs to know the value of each ϕ_j for $j=1,2,\ldots,t$ to implement the generalized YK decoder, which requires additional computational cost, apart from the circuit complexity. However, the computational cost for this is not high since the values are independent of $\mathcal{F}$ and there exist classical algorithms to compute such ϕ_j in running time $\mathcal{O}\big(\mathrm{poly}(t)\big)$ [44–48].

As $\mathcal{F}^{AB\to D}=\mathcal{N}^{C\to D}\circ\mathcal{E}^{AB\to C}$, Theorem 3 states that when the encoding map $\mathcal{E}$ is appropriately chosen against a given noise $\mathcal{N}$, or equivalently when the encoder $\mathcal{E}$ is chosen to satisfy the decoupling condition with small error, then the generalized YK decoder achieves a small error in recovering quantum information. As explained in II C, if the rate is below the quantum capacity, there exists such a good encoder that achieves the decoupling condition with a vanishing ϵ in the i.i.d. asymptotic limit. Hence, by setting δ in Theorem 3 to the values vanishing in the i.i.d. asymptotic limit, the generalized YK decoder can be used as a decoder that achieves the quantum capacity, which can be entanglement-non-assisted or -assisted. In this sense, the generalized YK decoder is a

capacity-achieving decoder.

We make a couple of comments on the complexity $\mathcal{C}(\mathcal{D}_{t,\phi})$. First, as the number t depends on $\lambda_{\min}(\omega^{RE})$, the receiver needs to know that value. When the decoupling condition is satisfied with small ϵ, the minimum eigenvalue $\lambda_{\min}(\omega^{RE}) \approx \lambda_{\min}(\tau^E)/d_A$, where $\lambda_{\min}(\tau^E)$ is minimum eigenvalue of the state τ^E in the environment. Since the number t is proportional to $[\lambda_{\min}(\tau^E)]^{-1/2}$, the larger $\lambda_{\min}(\tau^E)$ is, the smaller the complexity becomes. In the case that τ^E is a pure state, for instance, $\lambda_{\min}(\tau^E) = 1$. We then have a minimal complexity with $t = \Theta\big(\sqrt{d_A d_D/d_B}\,\log(1/\delta)\big)$. On the other hand, when τ^E is the CMS, $\lambda_{\min}(\tau^E) = 1/d_E$ and then $t = \Theta\big(\sqrt{d_A d_D d_E/d_B}\,\log(1/\delta)\big) = \Theta\big(d_A\sqrt{d_F}\,\log(1/\delta)\big)$.

From these observations and Eqs. (33) and (34), the number t is dominant in the complexity unless $\mathcal{C}(U_{\mathcal{F}})$ is exponentially large. The number t arises from the QSVT-based FPAA algorithm and is known to be an optimal order [27, 42, 49]. Hence, the quantum circuit implementation for the generalized YK decoder given in Fig. 4 cannot be significantly improved. Note that, while t is independent of the choice of the dilation of $\mathcal{F}^{AB\to D}$, the whole complexity is dependent on the choice due to the factor $\mathcal{C}(U_{\mathcal{F}}) + \log(d_D^2 d_E/d_B)$ in Eq. (34). Hence, using the unitary $U_{\mathcal{F}}^L$ which minimizes $\mathcal{C}(U_{\mathcal{F}}) + \log(d_D^2 d_E/d_B)$ results in the smallest complexity.

Another important factor to be noted in the complexity is $\sqrt{d_D/d_B}$, where d_D is the dimension of the output of the noisy channel $\mathcal{N}^{C\to D}$ and d_B is that of the pre-shared entanglement. In the simplest case, where the encoding map is given by a unitary on AB that is set to the same size as the input system C of the noisy channel $\mathcal{N}^{C\to D}$, we have $\sqrt{d_D/d_B} = \sqrt{d_A d_D/d_C}$. In this case, the complexity depends on d_A and the ratio d_D/d_C between the dimensions of the input C and the output D of the noisy channel. If the encoding is non-unitary, this is not the case, and one may expect that the complexity could be decreased by increasing d_B. This might be done by, e.g., factitiously adding more entanglement at the outset, and by discarding it in the encoding process. This trick, however, does not change the total complexity due to the other factor $[\lambda_{\min}(\omega^{RE})]^{-1/2}$. As $|\omega\rangle^{REDB'}$ is pure, $\lambda_{\min}(\omega^{RE}) = \lambda_{\min}(\omega^{DB'})$, where $\lambda_{\min}(\omega^{DB'})$ is non-zero minimum eigenvalue of $\omega^{DB'}$. This implies that, even if we factitiously add extra entanglement of dimension d_{extra} for increasing d_B, the value of $\lambda_{\min}(\omega^{DB'})$ changes by factor $1/d_{\text{extra}}$, which cancels the increase of d_B in the complexity.

B. Petz-like decoder

Using a similar technique, we can construct another decoder, which is thought of as a simplification of the Petz recovery map. We call this decoder the Petz-like decoder [50]. We first introduce a decoding protocol with post-selection in III B 1. Combining it with the QSVT-based FPAA algorithm, we explicitly construct the Petz-like decoder in III B 2.

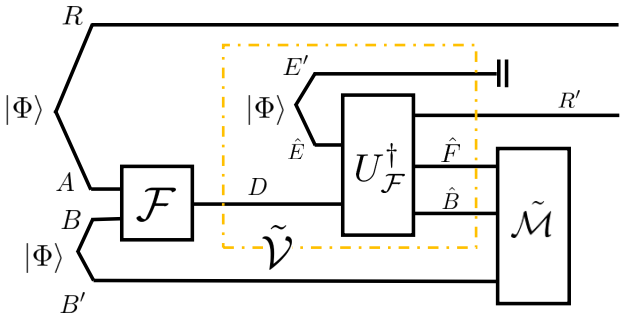


FIG. 5. A diagram of the protocol with post-selection for the Petz-like decoder. The dash-dotted box represents the isometry map $\tilde{\mathcal{V}}$ in Eq. (35).

1. *Decoding protocol with post-selection*

The decoding protocol with post-selection is as follows. See Fig. 5 as well. Similarly to the generalized YK decoder, we denote a Stinespring isometry of $\mathcal{F}^{AB\to D}$ by $U_{\mathcal{F}}^L|0\rangle^F$ as given in Eqs. (17) and (18). Note that the protocol works for any choice of $U_{\mathcal{F}}$.

1. The receiver prepares ancilla qubits in the system $\hat{E}E'$, and then generates a MES $\Phi^{\hat{E}E'}$.
2. The receiver applies the unitary $(U_{\mathcal{F}}^{\hat{L}})^\dagger$, where $\hat{L} = R'\hat{F}\hat{B} = \hat{E}D$.
3. the receiver performs a binary measurement $\tilde{\mathcal{M}} := \{|0\rangle\langle 0|^{\hat{F}} \otimes |\Phi\rangle\langle\Phi|^{\hat{B}B'}, \mathbb{I}^{\hat{F}\hat{B}B'} - |0\rangle\langle 0|^{\hat{F}} \otimes |\Phi\rangle\langle\Phi|^{\hat{B}B'}\}$ on $\hat{F}\hat{B}B'$. When the former result of the measurement $\tilde{\mathcal{M}}$ is obtained, this protocol succeeds.

In this protocol, all the systems with a prime or a hat, and the channel output D, are in the hands of the receiver. Below, we denote by $\tilde{\mathcal{V}}^{D\to E'R'\hat{F}\hat{B}}$ an isometry map of the operation up to the step 2. That is

$$\tilde{\mathcal{V}}^{D\to E'R'\hat{F}\hat{B}}(\,\cdot\,) := (U_{\mathcal{F}}^{\hat{L}})^\dagger(\,\cdot\otimes\Phi^{\hat{E}E'}\,)\,U_{\mathcal{F}}^{\hat{L}}. \tag{35}$$

Conditioned by the success of the measurement $\tilde{\mathcal{M}}$, the reduced state on the system RR' is given by

$$\tilde{\zeta}_{\text{succ}}^{RR'} = \mathrm{Tr}_{E'\hat{E}\hat{B}B'}\Big[\frac{1}{\tilde{p}_{\text{succ}}}(|0\rangle\langle 0|^{\hat{F}} \otimes |\Phi\rangle\langle\Phi|^{\hat{B}B'}) \tilde{\mathcal{V}}^{D\to E'\hat{F}R'\hat{B}}(\omega^{RDB'})\Big], \tag{36}$$

where $\tilde{p}_{\text{succ}}$ is the success probability of $\tilde{\mathcal{M}}$, and $\omega^{RDB'} = \mathcal{F}^{AB\to D}(\Phi^{AR}\otimes\Phi^{BB'})$. It is straightforward to show that

$$\tilde{p}_{\text{succ}} = \frac{d_A}{d_E}2^{-H_2(RE)_\omega}, \tag{37}$$

$$\mathrm{F}\big(\tilde{\zeta}_{\text{succ}}^{RR'}, \Phi^{RR'}\big) = \frac{1}{d_A}2^{H_2(RE)_\omega - H_2(E)_\omega}. \tag{38}$$

See Sec. III B 1 for the details.

As mentioned before, $U_{\mathcal{F}}^{L}$ is not uniquely determined from $\mathcal{F}^{AB\to D}$. Although this decoding protocol works for any choice of $U_{\mathcal{F}}$, the decoding performance depends on the choice, which is unlike the generalized YK decoder. In fact, the success probability $\tilde{p}_{\mathrm{succ}}$ is inverse-proportional to d_E, which implies that it succeeds with higher probability if a smaller environment of the channel $\mathcal{F}^{AB\to D}$ is chosen. On the other hand, the fidelity is the same as the generalized YK decoder. It is independent of the choice of $U_{\mathcal{F}}$, and we have $\mathrm{F}\big(\tilde{\zeta}_{\mathrm{succ}}^{RR'}, \Phi^{RR'}\big) \approx 1$ when the decoupling is satisfied as $\omega^{RE} \approx \pi^R \otimes \omega^E$.

2. *Construction of the Petz-like decoder*

We now use the QSVT-based FPAA algorithm to amplify the success probability of the measurement $\tilde{\mathcal{M}}$. To describe the corresponding unitary $\tilde{G}_{t,\phi}$, let us define two projectors as

$$\tilde{\Pi}_1^{E'R'\hat{F}\hat{B}} := (U_{\mathcal{F}}^{\hat{L}})^\dagger \big(|\Phi\rangle\langle\Phi|^{\hat{E}E'} \otimes \mathbb{I}^D \big) U_{\mathcal{F}}^{\hat{L}}, \tag{39}$$

$$\tilde{\Pi}_2^{\hat{F}\hat{B}B'} := |0\rangle\langle 0|^{\hat{F}} \otimes |\Phi\rangle\langle\Phi|^{\hat{B}B'}. \tag{40}$$

By replacing Π_m, in the definition of $W_m(\theta)$ $(m = 1, 2)$ in Eq. (27) and the following the constructions by (28) and (29), with $\tilde{\Pi}_1$ and $\tilde{\Pi}_2$, we define the unitary $\tilde{G}_{t,\phi}^{E'R'\hat{F}\hat{B}B'H}$.

The Petz-like decoder $\tilde{\mathcal{D}}_{t,\phi}^{DB'\to R'}$ is given by changing the step 3 in the protocol with post-selection to the following. See Fig. 6 as well.

3'. The receiver prepares an auxiliary state $|0\rangle^H$ in the system H and applies the unitary $\tilde{G}_{t,\phi}^{E'R'\hat{F}\hat{B}B'H}$.

With this modification, the Petz-like decoder is explicitly given as

$$\begin{aligned}&\tilde{\mathcal{D}}_{t,\phi}^{DB'\to R'}(\,\cdot\,)\\ &:= \mathrm{Tr}_{E'\hat{F}\hat{B}B'H}\big[\tilde{G}_{t,\phi}^{E'R'\hat{F}\hat{B}B'H}\big(\tilde{\mathcal{V}}^{D\to E'R'\hat{F}\hat{B}}(\,\cdot\,)\\ &\qquad\otimes |0\rangle\langle 0|^H\big)(\tilde{G}_{t,\phi}^{E'R'\hat{F}\hat{B}B'H})^\dagger\big].\end{aligned} \tag{41}$$

The number $t \in \mathbb{N}$ and the phases $\phi \in (-\pi, \pi]^t$ are chosen such that the QSVT realizes an approximation of the sign function.

The following theorem provides the performance of the Petz-like decoder.

Theorem 4 (Performance of the Petz-like decoder)**.** *For a given channel $\mathcal{F}^{AB\to D}$, let $\bar{\mathcal{F}}^{AB\to E}$ be a complementary channel of $\mathcal{F}^{AB\to D}$, ω^{RE} be*

$$\omega^{RE} = \bar{\mathcal{F}}^{AB\to E}(\Phi^{AR} \otimes \pi^B), \tag{42}$$

and $\lambda_{\min}(\omega^{RE})$ be the non-zero minimum eigenvalue of ω^{RE}. Suppose that there exists a state τ^E such that $\|\omega^{RE} - \pi^R \otimes \tau^E\|_1 \le \epsilon$. For any $\delta \in (0, 1]$, there exist $t \in \mathbb{N}$ and $\phi = (\phi_1, \phi_2, \dots, \phi_t) \in (-\pi, \pi]^t$ such that the recovery error $\Delta(\tilde{\mathcal{D}}_{t,\phi}\,|\,\mathcal{F})$ of the Petz-like decoder $\tilde{\mathcal{D}}_{t,\phi}^{DB'\to R'}$ is given by

$$\Delta(\tilde{\mathcal{D}}_{t,\phi}\,|\,\mathcal{F}) \le \sqrt{\epsilon} + \sqrt{2\delta}, \tag{43}$$

where t is an odd integer t satisfying

$$t = \Theta\left(\sqrt{\frac{d_E}{d_A \lambda_{\min}(\omega^{RE})}}\,\log(1/\delta)\right). \tag{44}$$

The circuit complexity of the decoder $\tilde{\mathcal{D}}_{t,\phi}^{DB'\to R'}$ is

$$\mathcal{C}(\tilde{\mathcal{D}}_{t,\phi}) = \mathcal{O}\Big(t\,\big(\mathcal{C}(U_{\mathcal{F}}) + \log(d_D d_E^2/d_A)\big)\Big), \tag{45}$$

and the number of ancilla qubits is $\mathcal{O}\big(\log(d_D d_E^2/d_A)\big)$. Here, $\mathcal{C}(U_{\mathcal{F}})$ is a circuit complexity of a unitary $U_{\mathcal{F}}^L$ such that $U_{\mathcal{F}}^L|0\rangle^F$ is the Stinespring isometry of $\mathcal{F}^{AB\to D}$, and $L = ABF = ED$.

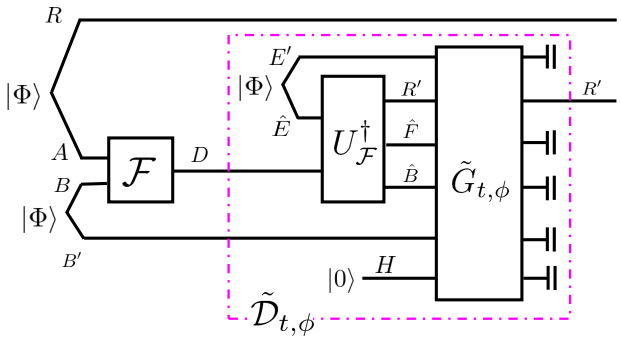


FIG. 6. A diagram of the Petz-like decoder $\tilde{\mathcal{D}}_{t,\phi}$, which is given in Eq. (41), corresponds to the dash-dotted box. Note that $\tilde{G}_{t,\phi}$ consists of repeated applications of unitaries, which is similar to Fig. 4.

Theorem 4 has many similarities to Theorem 3 for the generalized YK decoder, such as that the recovery error depends on the degree ϵ of the decoupling as well as the parameter δ that characterizes the trade-off relation between the recovery error and the circuit complexity of the decoder. Also, from the upper bound on the recovery error in Eq. (43), we observe that the Petz-like decoder achieves quantum capacity in the asymptotic i.i.d. limit if the encoder and δ are suitably chosen.

On the other hand, the complexity of the Petz-like decoder differs from that of the generalized YK decoder. The number t, as well as the remaining part in $\mathcal{C}(\tilde{\mathcal{D}}_{t,\phi})$, explicitly depends on d_E. This implies that the complexity depends on the choice of the dilation of $\mathcal{F}^{AB\to D}$, which reflects the aforementioned fact that the success probability of the protocol with post-selection is dependent on d_E. Hence, it is desirable to use a dilated unitary $U_{\mathcal{F}}^L$ with a small environment E.

In the next section, we compare the complexities of decoders and clarify the cases in which one decoder has

TABLE I. A table of notation that we use in III C. Instead of the dimensions, we use the numbers of qubits in the systems.

k	The number of logical qubits in A: $k = \log d_A$.
$n_{\rm in}$	The number of input qubits of the channel $\mathcal{N}$: $n_{\rm in} = \log d_C$.
$n_{\rm out}$	The number of output qubits of the channel $\mathcal{N}$: $n_{\rm out} = \log d_D$
e	The number of ebits shared by the sender and the receiver in advance: $e = \log d_B$
κ	The number of qubits in the environment E, which is equal to the logarithm of #Kraus ops.: $\kappa = \log d_E = \log(\#\text{Kraus ops.})$.

smaller complexity than the other. As explained, our decoder has a better circuit complexity than the algorithmic implementation of the original Petz recovery map [20], if δ is appropriately chosen. This is because we are interested in applying the Petz recovery map to decoding quantum information. When this is the case, it is not necessary to exactly implement the Petz recovery map.

C. Comparision of the circuit complexities

We compare the circuit complexities of the generalized YK decoder, the Petz-like decoder, and the algorithmic implementation of the original Petz recovery map [20]. In the comparison, we use the number of qubits in each system instead of the dimensions. Specifically, we denote the number of qubits in A, B, C, D, and E by k, e, $n_{\rm in}$, $n_{\rm out}$, and κ, respectively. See Table I as well. Note that κ is the logarithm of the number of the Kraus operators of the channel $\mathcal{F}^{AB\to D}$, i.e., $\kappa = \log d_E = \log(\#\text{Kraus ops.})$. This number depends on how the channel is dilated. As we are interested in minimizing the complexity, we take the minimum possible number of Kraus operators in the comparison below.

We here compare the complexity of the generalized YK decoder and that of the Petz-like decoder. As explained in Sec. III A 2, the number t is the significant factor in the complexity. We denote the numbers t for the generalized YK decoder and for the Petz-like decoder by $t_{\rm gYK}$ and $t_{\rm Pl}$, respectively. That is,

$$t_{\rm gYK} = \Theta\Big(\big[\,2^{e-n_{\rm out}}\lambda_{\min}(\omega^{RE})\big]^{-1/2}\log(1/\delta)\Big), \tag{46}$$

$$t_{\rm Pl} = \Theta\Big(\big[\,2^{k-\kappa}\lambda_{\min}(\omega^{RE})\big]^{-1/2}\log(1/\delta)\Big). \tag{47}$$

See Eq. (33) and Eq. (44). Comparing $t_{\rm gYK}$ and $t_{\rm Pl}$, we find that

$$t_{\rm gYK} \le t_{\rm Pl} \tag{48}$$

$$\iff\quad k - e \le \kappa - n_{\rm out}. \tag{49}$$

The left-hand side of Eq. (49) is given by the number k of qubits that the sender intends to transmit and the number e of pre-shared ebits. On the other hand, the right-hand side depends on the quantities κ and $n_{\rm out}$ that are the properties of the channel $\mathcal{F}^{AB\to D}$. To better understand the condition (49), we below consider a couple of explicit instances, in which we assume an isometric encoder for convenience. In these cases, κ corresponds to the number of Kraus operators of the noisy channel $\mathcal{N}^{C\to D}$.

For a given noisy channel $\mathcal{N}^{C\to D}$, the right-hand side of Eq. (49) is fixed as a property of the noise. Hence, the number of logical qubits, k, and that of pre-shared entanglement, e, determines which decoder has smaller complexity. In general, the generalized YK decoder has an advantage when e is large, and as e becomes smaller, the advantage shifts to the Petz-like decoder. To observe this more concretely, we note that $0 \le e \le n_{\rm in} - k$. When the sender and the receiver pre-share the maximal number of entanglement, i.e., $e = n_{\rm in} - k$, Eq. (49) is rephrased as $k \le \frac{1}{2}(n_{\rm in} - n_{\rm out} - \kappa)$. In particular, if the input and the output systems of the channel $\mathcal{N}^{C\to D}$ are identical, i.e., $n_{\rm in} = n_{\rm out}$, it reduces to $k \le \frac{1}{2}\kappa$. In this case, unless the number of logical qubits exceeds half of the number of the Kraus operators of the noisy channel, the generalized YK decoder has smaller complexity than the Petz-like decoder. In contrast, when no entanglement is shared in advance and $e = 0$, Eq. (49) reduces to $k \le \kappa - n_{\rm out}$. Although whether this holds or not depends on details, there exist cases where the inequality is violated, such as the amplitude damping noise on each qubit independently. For such noises or the choice of large k, the Petz-like decoder has smaller complexity than the generalized YK decoder.

We may also use the fact that k should necessarily satisfy $k \le n_{\rm in}$ for the recovery to be possible. This leads to a trivial inequality $k + n_{\rm out} - \kappa \le n_{\rm in} + n_{\rm out} - \kappa$. Furthermore, κ always satisfies $\kappa \le n_{\rm in} + n_{\rm out}$, since κ is the logarithm of the number of Kraus operators. If a given noisy channel $\mathcal{N}$ has the property that $\kappa = n_{\rm in} + n_{\rm out}$, it follows that

$$k + n_{\rm out} - \kappa \le 0 \le e, \tag{50}$$

for any e. Hence, for the noise with the maximum possible number of Kraus operators, the generalized YK decoder has smaller complexity than the Petz-like decoder no matter how much entanglement is pre-shared.

We next compare the complexity of the Petz-like decoder with an algorithmic implementation of the original

Petz recovery map provided in [20]. The following Corollary can be derived by applying this algorithmic implementation to our setting.

Corollary 5 (Algorithmic implementation of the Petz recovery map [20]). *Let $\mathcal{P}_{\pi,\mathcal{G}}^{DB'\to R'}$ be the decoder based on the Petz recovery map defined in Eq. (19). There exists a quantum algorithm realizing the map $\tilde{\mathcal{P}}_{\pi,\mathcal{G}}^{DB'\to R'}$, which satisfies*

$$\|\tilde{\mathcal{P}}_{\pi,\mathcal{G}}^{DB'\to R'} - \mathcal{P}_{\pi,\mathcal{G}}^{DB'\to R'}\|_\diamond \le \varepsilon, \tag{51}$$

with a circuit complexity

$$\mathcal{C}(\,\tilde{\mathcal{P}}_{\pi,\mathcal{G}}\,) = \mathcal{O}\Big(\, t_{\,\mathrm{Petz}}\big(\,\mathcal{C}(U_{\mathcal{F}}) + \log d_B d_E + \frac{\mathcal{C}(U_\omega)}{\lambda_{\min}(\omega^{RE})} \times \log\frac{d_E}{\varepsilon} + d_A \log d_A \log\frac{d_E}{\varepsilon\lambda_{\min}(\omega^{RE})}\big)\Big), \tag{52}$$

where $t_{\,\mathrm{Petz}}$ is an integer satisfying

$$t_{\,\mathrm{Petz}} = \Theta\left(\sqrt{\frac{d_E}{\lambda_{\min}(\omega^{RE})}}\right), \tag{53}$$

and $\mathcal{C}(U_\omega)$ is a circuit complexity of a unitary $U_\omega^{DB'P}$ such that, for any system P,

$$\omega^{DB'} = \mathrm{Tr}_P[U_\omega^{DB'P}|0\rangle\langle 0|^{DB'P}(U_\omega^{DB'P})^\dagger]. \tag{54}$$

From Eqs. (20) and (51), the recovery error of $\tilde{\mathcal{P}}_{\pi,\mathcal{G}}$ is bounded as

$$\Delta(\,\tilde{\mathcal{P}}_{\pi,\mathcal{G}}\,|\,\mathcal{F}\,) \le 2\,\epsilon^{1/4} + \varepsilon, \tag{55}$$

when there exists τ^E such that $\|\omega^{RC} - \pi^R\otimes\tau^E\|_1 \le \epsilon$.

We clarify the condition that the Petz-like decoder has smaller complexity than the algorithmic implementation of the Petz recovery map. First, when $\mathcal{C}(U_{\mathcal{F}})$ is larger than other terms, Eqs. (45) and (52) approximately reduce to

$$\mathcal{C}(\tilde{\mathcal{D}}_{t,\phi}) \approx \mathcal{O}(\,t_{\,\mathrm{Pl}}\,\mathcal{C}(U_{\mathcal{F}})\,), \tag{56}$$

$$\mathcal{C}(\tilde{\mathcal{P}}_{\pi,\mathcal{G}}) \approx \mathcal{O}(\,t_{\,\mathrm{Petz}}\,\mathcal{C}(U_{\mathcal{F}})\,), \tag{57}$$

respectively. When this is the case, we only need to compare $t_{\,\mathrm{Pl}}$ with $t_{\,\mathrm{Petz}}$, which satisfies $t_{\,\mathrm{Pl}} = \Theta\big(\frac{\log(1/\delta)}{\sqrt{d_A}}\,t_{\,\mathrm{Petz}}\big)$. Hence, as far as

$$\delta = \Omega(2^{-\sqrt{d_A}}), \tag{58}$$

the Petz-like decoder has smaller complexity than the algorithmic implementation of the original Petz recovery map. Note that δ is also related to the recovery error of the Petz-like decoder, as in Eq. (43). However, the choice of δ such as Eq. (58) is sufficiently small and can be negligible in the recovery error.

The advantage of the Petz-like decoder remains even when $\mathcal{C}(U_{\mathcal{F}})$ is not dominant. To see this, suppose that ε in Eq. (52) is $\varepsilon = \mathcal{O}(\sqrt{\delta}\,)$ with sufficiently small δ. The complexity of the algorithmic implementation of the Petz recovery map reduces to

$$\mathcal{C}(\tilde{\mathcal{P}}_{\pi,\mathcal{G}}) \approx \mathcal{O}\Big(t_{\,\mathrm{Petz}}\log(1/\delta)\mathrm{poly}(n_{\mathrm{in}},n_{\mathrm{out}},k) \times\Big(\frac{\mathrm{poly}(n_{\mathrm{in}},n_{\mathrm{out}},k)}{\lambda_{\min}(\omega^{RE})} + k2^k\Big)\Big) \tag{59}$$

$$= \mathcal{O}\Big(t_{\,\mathrm{Pl}}\,\mathrm{poly}(n_{\mathrm{in}},n_{\mathrm{out}},k) \times 2^{k/2}\Big(\frac{\mathrm{poly}(n_{\mathrm{in}},n_{\mathrm{out}},k)}{\lambda_{\min}(\omega^{RE})} + k2^k\Big)\Big). \tag{60}$$

Here, we used in the second equation that $t_{\,\mathrm{Pl}} = \Theta\big(\frac{\log(1/\delta)}{\sqrt{d_A}}\,t_{\,\mathrm{Petz}}\big)$ and assumed that $\mathcal{C}(U_{\mathcal{F}}^L)$ is polynomial in qubits, which further implies that $\mathcal{C}(U_\omega^{DB'P})$ is polynomial. On the other hand, the complexity of the Petz-like decoder in this case is

$$\mathcal{C}(\tilde{\mathcal{D}}_{t,\phi}) = \mathcal{O}(\,t_{\,\mathrm{Pl}}\,\mathrm{poly}(n_{\mathrm{in}},n_{\mathrm{out}},k)\,). \tag{61}$$

Since this corresponds to the first line of Eq. (60), the Petz-like decoder has smaller circuit complexity than the algorithmic implementation of the Petz recovery map.

D. Application to concrete noisy models

We consider several noises for demonstration. We investigate the noises that independently act on each qubit, such as the independent Pauli noise, the independent amplitude damping noise, and the qubit-erasure noise. If the input system C of the noisy channel $\mathcal{N}^{C\to D}$ is equal to the output system D of it, we denote by S the system as $S = C = D$, and by n the number of these qubits as $n = n_{\mathrm{in}} = n_{\mathrm{out}}$.

- Independent Pauli noise

The first example is the independent Pauli noise. A Stinespring isometry of the single-qubit Pauli noise is given by

$$V_{\mathcal{N}}^{S\to ES} = \sum_{i=0}^{3}\sqrt{p_i}\,|e_i\rangle^E\otimes\sigma_i^S, \tag{62}$$

where $\sum_{i=0}^3 p_i = 1$ and $(\sigma_i^S) = (\mathbb{I}^S, X^S, Y^S, Z^S)$. Since the number of qubits of the system S is n, and the logarithm of the number of the Kraus operators $\kappa = 2n$, we can rephrase Eqs. (48) and (49) as

$$n - k - e \ge 0 \tag{63}$$

$$\iff\quad t_{\,\mathrm{gYK}} \le t_{\,\mathrm{Pl}}. \tag{64}$$

Since $k+e \le n$ is always satisfied, the generalized YK decoder has smaller complexity than the Petz-like decoder for the independent Pauli noise.

TABLE II. The circuit complexity of our decoders to particular noise models. We denote $\min_i\{p_i\}$ by $p_{\rm min}$. The constant γ is assumed to be $1/2$ or less. We have assumed a unitary encoding, so $k+e=n_{\rm in}$. The part $\mathrm{poly}(\cdots)$ comes from the term of unitary dilation of the noise, and from the term logarithmic in dimensions in Eqs. (34) and (45).

	Generalized YK decoder $\mathcal{C}(\mathcal{D}_{t,\phi})$	Petz-like decoder $\mathcal{C}(\tilde{\mathcal{D}}_{t,\phi})$
Pauli noise	$\left[\left(2^k/p_{\rm min}^{n/2}\right)\log(1/\delta)\right]\mathrm{poly}(n,k)$	$\left[\left(2/p_{\rm min}^{1/2}\right)^n\log(1/\delta)\right]\mathrm{poly}(n,k)$
Amplitude damping noise	$\left[2^k(2/\gamma)^{n/2}\log(1/\delta)\right]\mathrm{poly}(n,k)$	$\left[(4/\gamma)^{n/2}\log(1/\delta)\right]\mathrm{poly}(n,k)$
Erasure noise	$\left[2^k\log(1/\delta)\right]\mathrm{poly}(n_{\rm in},n_{\rm out},k)$	$\left[2^{n_{\rm in}-n_{\rm out}}\log(1/\delta)\right]\mathrm{poly}(n_{\rm in},n_{\rm out},k)$

- Independent amplitude damping noise

The second example is the amplitude damping noise for $\{|0\rangle^S,|1\rangle^S\}$, which independently acts on each qubit. The single-qubit noise is represented by an isometry

$$V_{\mathcal{N}}^{S\to ES}=\sqrt{\gamma}\,|e_0\rangle^E\otimes|0\rangle\langle 1|^S + |e_1\rangle^E\otimes(|0\rangle\langle 0|^S+\sqrt{1-\gamma}\,|1\rangle\langle 1|^S), \tag{65}$$

where $\gamma\in[0,1]$. As $n=\kappa$, Eqs. (48) and (49) become

$$e-k\ge 0 \tag{66}$$
$$\iff\quad t_{\rm gYK}\le t_{\rm Pl}. \tag{67}$$

Hence, when the number of pre-shared entanglement e is more than the number of the logical qubits k, the generalized YK decoder has smaller complexity than the Petz-like decoder.

- Qubit-erasure noise

The third example is the qubit-erasure noise, which erases κ qubits out of $n_{\rm in}$ input qubits. The erased qubits are randomly chosen, but it is assumed that the receiver knows which qubits were erased. In this case, it holds that $n_{\rm in}=n_{\rm out}+\kappa$. Thus, Eqs. (48) and (49) become

$$n_{\rm in}-2n_{\rm out}-k+e\ge 0 \tag{68}$$
$$\iff\quad t_{\rm gYK}\le t_{\rm Pl}. \tag{69}$$

Especially, when there is no pre-shared entanglement, $e=0$ and the encoding rate $k/n_{\rm in}$ is given by $k/n_{\rm in}=n_{\rm out}/n_{\rm in}-1/2$, which is the value near the quantum capacity, Eq. (68) does not hold, and the Petz-like decoder has smaller complexity than the generalized YK decoder. On the other hand, when the maximal amount of entanglement is pre-shared, i.e., $e=n_{\rm in}-k$, Eq. (68) is rephrased as $k\le n_{\rm in}-n_{\rm out}=\kappa$. Hence, if more than k qubits are erased by the noise, the generalized YK decoder has smaller complexity than the Petz-like decoder.

In Table II, we explicitly provide the circuit complexities of our decoders against these noise models. For simplicity, the values in Table II are restricted to those for a unitary encoder by a polynomial-sized quantum circuit. Moreover, we assume the decoupling $\omega^{RE}\approx\pi^R\otimes\omega^E$, which leads to

$$\lambda_{\rm min}(\omega^{RE})\approx\lambda_{\rm min}(\omega^E)/d_A, \tag{70}$$

where $\lambda_{\rm min}(\omega^E)$ is the non-zero minimum eigenvalue of $\omega^E=\bar{\mathcal{N}}^{C\to E}(\pi^C)$.

From these results, we find that, when $p_{\rm min}=\min_{i=0,1,2,3}\{p_i\}$ or γ is larger, the complexities becomes smaller. Hence, from the viewpoint of the computational cost, both the generalized YK decoder and the Petz-like decoder are more advantageous in moderately noisy situations.

IV. PROOFS

In this section, we provide proofs of the main results. In IV A and IV B, we show the statements about the generalized YK decoder and the Petz-like decoder, respectively.

A. Proofs: the generalized YK decoder

We first consider the decoding protocol with post-selection, and provide the success probability and the fidelity after the post-selection. We then prove Theorem 3.

1. *Success probability and fidelity in the decoding protocol with post-selection*

The input state of the decoding protocol is

$$\omega^{RDB'}=\mathcal{F}^{AB\to D}(\Phi^{AR}\otimes\Phi^{BB'}). \tag{71}$$

When necessary, we consider the state including the environment E, namely, a purified state

$$|\omega\rangle^{REDB'}=V_{\mathcal{F}}^{AB\to ED}|\Phi\rangle^{AR}|\Phi\rangle^{BB'}. \tag{72}$$

We use the following lemma. The proof of this lemma is straightforward. See Fig. 7 for the diagram of the statement.

Lemma 6 (Transpose of a matrix sandwiched by two MESs)**.** *For any linear operator $L^{AB\to ED}$, i.e., $d_Ed_D\times d_Ad_B$ matrix, it holds that*

$$\langle\Phi|^{EE'}\big(\mathbb{I}^{B'E'}\otimes L^{AB\to ED}\big)|\Phi\rangle^{BB'} = \sqrt{\frac{d_Ad_D}{d_Bd_E}}\,\langle\Phi|^{AA'}\big((L^{A'B'\to E'D'})^{\mathsf{T}}\otimes\mathbb{I}^{AD}\big)|\Phi\rangle^{DD'}. \tag{73}$$

Note that this is a liner operator from AE' to $B'D$. The transpose is taken with respect to the basis that defines each MES.

Using Lemma 6 for $L=V_{\mathcal{F}}^*$, the state $\zeta_{\rm succ}^{RR'}$ on the system RR' after the post-selection is rewritten as

$$\begin{aligned}
\zeta_{\rm succ}^{RR'} &= \frac{1}{p_{\rm succ}}\mathrm{Tr}_{E'}\big[\langle\Phi|^{DD'}(V_{\mathcal{F}}^{A'B'\to E'D'})^*\\
&\qquad(\omega^{RDB'}\otimes\Phi^{A'R'})(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}|\Phi\rangle^{DD'}\big] &(74)\\
&= \frac{1}{p_{\rm succ}}\mathrm{Tr}_{E'}\big[\langle\Phi|^{DD'}(V_{\mathcal{F}}^{A'B'\to E'D'})^*|\Phi\rangle^{A'R'}\\
&\qquad\omega^{RDB'}\langle\Phi|^{A'R'}(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}|\Phi\rangle^{DD'}\big] &(75)\\
&= \frac{1}{p_{\rm succ}}\frac{d_Bd_E}{d_Ad_D}\mathrm{Tr}_{E'}\big[\langle\Phi|^{\hat{B}B'}(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D})^\dagger|\Phi\rangle^{\hat{E}E'}\\
&\qquad\omega^{RDB'}\langle\Phi|^{\hat{E}E'}V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}|\Phi\rangle^{\hat{B}B'}\big] &(76)\\
&= \frac{1}{p_{\rm succ}}\frac{d_B}{d_Ad_D}\langle\Phi|^{\hat{B}B'}(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D})^\dagger\\
&\qquad(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}|\Phi\rangle^{\hat{B}B'}. &(77)
\end{aligned}$$

Here, we used Lemma 6 in the third equation. The success probability of the measurement $\mathcal{M}$ is then given as

$$\begin{aligned}
p_{\rm succ} &= \frac{d_B}{d_Ad_D}\mathrm{Tr}[\langle\Phi|^{\hat{B}B'}\big(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}\big)^\dagger\\
&\qquad(\,\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}}\,)V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}|\Phi\rangle^{\hat{B}B'}] &(78)\\
&= \frac{d_B}{d_D}\mathrm{Tr}\big[\big(\mathbb{I}^R\otimes V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}(\pi^{R'}\otimes\Phi^{\hat{B}B'})\\
&\qquad(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D})^\dagger\big)(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})\big] &(79)\\
&= \frac{d_B}{d_D}\mathrm{Tr}[(\,\mathbb{I}^R\otimes\omega^{DB'\hat{E}})(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})] &(80)\\
&= \frac{d_B}{d_D}\mathrm{Tr}[(\omega^{DB'})^2] &(81)\\
&= \frac{d_B}{d_D}2^{-H_2(RE)_\omega}. &(82)
\end{aligned}$$

Since the state $|\omega\rangle^{REDB'}$ is pure, we here used that $\mathrm{Tr}[(\omega^{DB'})^2]=\mathrm{Tr}[(\omega^{RE})^2]=2^{-H_2(RE)_\omega}$.

The fidelity after the post-selection is calculated from

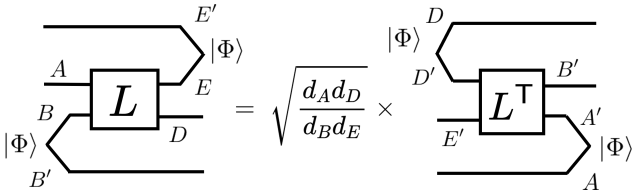


FIG. 7. A diagram of the transpose of a matrix L sandwiched by two MESs

$\zeta_{\rm succ}^{RR'}$ as follows:

$$\begin{aligned}
&\mathrm{F}(\,\zeta_{\rm succ}^{RR'},\Phi^{RR'})\\
&= \frac{1}{p_{\rm succ}}\frac{d_B}{d_Ad_D}\mathrm{Tr}[\Phi^{RR'}\langle\Phi|^{\hat{B}B'}\big(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}\big)^\dagger\\
&\qquad(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}|\Phi\rangle^{\hat{B}B'}] &(83)\\
&= \frac{1}{p_{\rm succ}}\frac{d_B}{d_Ad_D}\mathrm{Tr}[V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}(\Phi^{RR'}\otimes\Phi^{\hat{B}B'})\\
&\qquad(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D})^\dagger(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})] &(84)\\
&= \frac{1}{p_{\rm succ}}\frac{d_B}{d_Ad_D}\mathrm{Tr}[\,\omega^{R\hat{E}DB'}(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})] &(85)\\
&= \frac{1}{p_{\rm succ}}\frac{d_B}{d_Ad_D}\mathrm{Tr}[(\omega^{RDB'})^2] &(86)\\
&= \frac{1}{p_{\rm succ}}\frac{d_B}{d_Ad_D}2^{-H_2(RDB')_\omega}. &(87)
\end{aligned}$$

Substituting Eq. (82), we obtain that

$$\begin{aligned}
\mathrm{F}(\,\zeta_{\rm succ}^{RR'},\phi^{RR'}) &= \frac{1}{d_A}2^{H_2(RE)-H_2(RDB')} &(88)\\
&= \frac{1}{d_A}2^{H_2(RE)-H_2(E)}, &(89)
\end{aligned}$$

where we used $H_2(RDB')_\omega=H_2(E)_\omega$ since $|\omega\rangle^{REDB'}$ is pure. Thus, we obtain Eqs. (23) and (24).

2. *Proof of Theorem 3*

To show Theorem 3, we use the QSVT-based FPAA algorithm instead of the measurement $\mathcal{M}$. We here mention that our situation differs from the common situation for the AA algorithm since the receiver has access only to a part of the whole system: the reference R and environment E are not with the receiver. This issue will be circumvented by Jordan's lemma, which we explain below.

We denote the input state of the QSVT-based FPAA algorithm by

$$\omega_0^{RDD'E'R'} := \mathcal{V}^{B'\to D'E'R'}(\omega^{RDB'}), \tag{90}$$

where $\mathcal{V}^{B'\to D'E'R'}$ is the isometry map such that

$$\begin{aligned}
&\mathcal{V}^{B'\to D'E'R'}(\,\cdot\,)\\
&\quad= (V_{\mathcal{F}}^{A'B'\to E'D'})^*(\,\cdot\,\otimes\Phi^{A'R'})(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}.
\end{aligned} \tag{91}$$

Note that $\omega_0^{RE} = \omega^{RE}$. Let $|\omega_0\rangle^{REDD'E'R'}$ be the purified state of $\omega_0^{RDD'E'R'}$ given by

$$|\omega_0\rangle^{REDD'E'R'} = (V_{\mathcal{F}}^{A'B'\to E'D'})^* |\omega\rangle^{REDB'} |\Phi\rangle^{A'R'}. \tag{92}$$

We first check relations between this state ω_0, the state after the post-selection $\zeta_{\rm succ}$, and the two projectors Π_1 and Π_2. Here, the state $\zeta_{\rm succ}$ on $REE'R'$ after post-selection is given by

$$|\zeta_{\rm succ}\rangle^{REE'R'} = \frac{1}{\sqrt{p_{\rm succ}}} \langle\Phi|^{DD'} |\omega_0\rangle^{REDD'E'R'}. \tag{93}$$

To this end, we use the following lemma.

Lemma 7 (Jordan's lemma [51–53]). *For any two projectors Π and Π' on a Hilbert space $\mathcal{H}$, there exists an orthogonal decomposition of $\mathcal{H}$ into one- and two-dimensional subspaces $\mathcal{H}_\mu$. Each subspace $\mathcal{H}_\mu$ is invariant under Π and Π'. Moreover, in each subspace, Π and Π' act as rank-one projectors, such as $\Pi|_{\mathcal{H}_\mu} = |\psi_\mu\rangle\langle\psi_\mu|$ and $\Pi'|_{\mathcal{H}_\mu} = |\xi_\mu\rangle\langle\xi_\mu|$, respectively. Each subspace is hence given by $\mathcal{H}_\mu = \mathrm{span}\{|\psi_\mu\rangle, |\xi_\mu\rangle\}$.*

This lemma states that, as $|\psi_\mu\rangle \perp |\xi_\nu\rangle$ for $\mu \neq \nu$, namely, they are in different subspaces, the products of Π and Π' are given by

$$\Pi\Pi'\Pi = \sum_{\mu=1}^{r} q_\mu |\psi_\mu\rangle\langle\psi_\mu|, \tag{94}$$

$$\Pi'\Pi\Pi' = \sum_{\mu=1}^{r} q_\mu |\xi_\mu\rangle\langle\xi_\mu|, \tag{95}$$

where $q_\mu = |\langle\xi_\mu|\psi_\mu\rangle|^2$, and we arranged them such as $q_1 \geq q_2 \geq \ldots \geq q_r > 0$. The whole Hilbert space can be decomposed as

$$\mathcal{H} = \bigoplus_{\mu=1}^{r} \mathcal{H}_\mu \oplus \mathcal{H}_\perp. \tag{96}$$

Here, the Hilbert spaces $\mathcal{H}_\mu = \mathrm{span}\{|\psi_\mu\rangle, |\xi_\mu\rangle\}$ are either common one-dimensional subspaces spanned by $|\psi_\mu\rangle = |\xi_\mu\rangle$ or two-dimensional subspaces. The Hilbert space $\mathcal{H}_\perp$ is the remaining orthogonal complement to the others.

We apply the Jordan's lemma to our projectors

$$\mathbb{I}^D \otimes \Pi_1^{D'E'R'} = \mathbb{I}^D \otimes (V_{\mathcal{F}}^{A'B'\to E'D'})^* (\mathbb{I}^{B'} \otimes |\Phi\rangle\langle\Phi|^{A'R'}) (V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}, \tag{97}$$

$$\Pi_2^{DD'} \otimes \mathbb{I}^{E'R'} = |\Phi\rangle\langle\Phi|^{DD'} \otimes \mathbb{I}^{E'R'}. \tag{98}$$

Then, the Hilbert space $\mathcal{H}^{DD'E'R'}$ is decomposed into a direct sum of one- and two-dimensional subspaces, each of which is invariant under $\Pi_1^{D'E'R'}$ and $\Pi_2^{DD'}$. The products of these projectors can be computed as

$$(\Pi_1\Pi_2\Pi_1)^{DD'E'R'} = \frac{d_B}{d_D} \omega_0^{DD'E'R'}, \tag{99}$$

$$(\Pi_2\Pi_1\Pi_2)^{DD'E'R'} = |\Phi\rangle\langle\Phi|^{DD'} \otimes \left(\frac{d_B p_{\rm succ}}{d_D} \zeta_{\rm succ}^{E'R'}\right)^{1/2}, \tag{100}$$

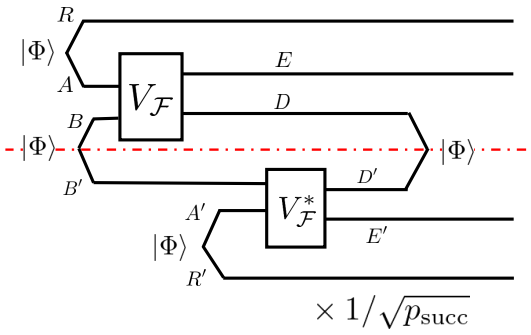


FIG. 8. A diagram of the state $|\zeta_{\rm succ}\rangle^{REE'R'}$. This is symmetrical with respect to the red dash-dotted line, up to the complex conjugate. Due to this symmetry, the Schmidt basis of $|\zeta_{\rm succ}\rangle^{REE'R'}$ is given by $\{|\eta_\mu\rangle^{RE} |\eta_\mu^*\rangle^{E'R'}\}_\mu$.

which are derived in Appendix A.

Let q_μ and $|\psi_\mu\rangle^{DD'E'R'}$ for $\mu = 1, 2, \ldots, r$ be nonzero eigenvalues and the corresponding eigenstates of $(\Pi_1\Pi_2\Pi_1)^{DD'E'R'}$, respectively. From Eq. (99), the Schmidt decomposition of $|\omega_0\rangle^{REDD'E'R'}$, divided into RE and $DD'E'R'$, is given by

$$|\omega_0\rangle^{REDD'E'R'} = \sum_{\mu=1}^{r} \sqrt{\frac{d_D}{d_B}} \sqrt{q_\mu} |\eta_\mu\rangle^{RE} |\psi_\mu\rangle^{DD'E'R'}, \tag{101}$$

where $\{|\eta_\mu\rangle^{RE}\}_\mu$ is an orthonormal basis. From Eq. (93), the state $|\zeta_{\rm succ}\rangle^{REE'R'}$ is then given by

$$|\zeta_{\rm succ}\rangle^{REE'R'} = \sum_{\mu=1}^{r} \sqrt{\frac{d_D q_\mu}{d_B p_{\rm succ}}} |\eta_\mu\rangle^{RE} \langle\Phi|^{DD'} |\psi_\mu\rangle^{DD'E'R'}. \tag{102}$$

It is important to notice the symmetry of $|\zeta_{\rm succ}\rangle^{REE'R'}$ between RE and $R'E'$. From Fig. 8, we observe that taking the complex conjugate of this state is equal to swapping RE for $R'E'$. Hence, the Schmidt basis of $|\zeta_{\rm succ}\rangle^{REE'R'}$ in RE and that in $E'R'$ are the same up to the complex conjugate. Together with Eq. (102), we see that $\langle\Phi|^{DD'} |\psi_\mu\rangle^{DD'E'R'}$ is proportional to $|\eta_\mu^*\rangle^{E'R'}$ with a real coefficient. Moreover, substituting Eq. (102) to Eq. (100) and noting that the eigenvalues of $\Pi_2\Pi_1\Pi_2$ are q_μ by the Jordan's lemma, the coefficient turns out to be $\sqrt{q_\mu}$. Thus,

$$\langle\Phi|^{DD'} |\psi_\mu\rangle^{DD'E'R'} = \sqrt{q_\mu} |\eta_\mu^*\rangle^{E'R'}. \tag{103}$$

From Eqs. (102) and (103), the Schmidt decomposition of $|\zeta_{\rm succ}\rangle^{REE'R'}$ is given by

$$|\zeta_{\rm succ}\rangle^{REE'R'} = \sum_{\mu=1}^{r} \sqrt{\frac{d_D}{d_B}} \frac{q_\mu}{\sqrt{p_{\rm succ}}} |\eta_\mu\rangle^{RE} |\eta_\mu^*\rangle^{E'R'}. \tag{104}$$

Using the states $\{|\psi_\mu\rangle^{DD'E'R'}\}_\mu$ and $\{|\eta_\mu^*\rangle^{E'R'}\}_\mu$, the products of projectors Π_1 and Π_2 are rephrased as

$$(\Pi_1\Pi_2\Pi_1)^{DD'E'R'} = \sum_{\mu=1}^{r} q_\mu |\psi_\mu\rangle\langle\psi_\mu|^{DD'E'R'}, \tag{105}$$

$$(\Pi_2\Pi_1\Pi_2)^{DD'E'R'} = \sum_{\mu=1}^{r} q_\mu |\xi_\mu^*\rangle\langle\xi_\mu^*|^{DD'E'R'}, \tag{106}$$

where $|\xi_\mu^*\rangle^{DD'E'R'} := |\Phi\rangle^{DD'}|\eta_\mu^*\rangle^{E'R'}$, and the Hilbert space $\mathcal{H}^{DD'E'R'}$ is decomposed into

$$\mathcal{H}^{DD'E'R'} = \bigoplus_{\mu=1}^{r} \mathcal{H}_\mu^{DD'E'R'} \oplus \mathcal{H}_\perp^{DD'E'R'}, \tag{107}$$

where $\mathcal{H}_\mu^{DD'E'R'} = \mathrm{span}\{|\psi_\mu\rangle^{DD'E'R'}, |\xi_\mu^*\rangle^{DD'E'R'}\}$ and $\mathcal{H}_\perp^{DD'E'R'}$ is remaining orthogonal complement to $\oplus_{\mu=1}^r \mathcal{H}_\mu^{DD'E'R'}$.

In the following, we focus only on the subspaces $\oplus_{\mu=1}^r \mathcal{H}_\mu^{DD'E'R'}$ and ignore $\mathcal{H}_\perp^{DD'E'R'}$. This does not cause any issue since Eqs. (99) and (100) guarantee that all eigenstates of $\omega_0^{DD'E'R'}$ and $\Phi^{DD'} \otimes \zeta_{\mathrm{succ}}^{E'R'}$ are in $\oplus_{\mu=1}^r \mathcal{H}_\mu^{DD'E'R'}$. As we will explain later, our goal is to transform the eigenvectors $|\psi_\mu\rangle^{DD'E'R'}$ to the corresponding eigenvectors $|\xi_\mu^*\rangle^{DD'E'R'}$ within each subspace $\mathcal{H}_\mu^{DD'E'R'}$ by the QSVT-based FPAA algorithm. Thus, it is sufficient that we focus only on the subspaces that contain all eigenstates.

For the sake of analysis, we define an auxiliary state $|\omega_{\mathrm{targ}}\rangle^{RR'EE'}$ as

$$|\omega_{\mathrm{targ}}\rangle^{RR'EE'} := \sum_\mu \sqrt{\frac{d_D}{d_B}} \sqrt{q_\mu} |\eta_\mu\rangle^{RE} |\eta_\mu^*\rangle^{E'R'}. \tag{108}$$

This state is useful due to the following lemma.

Lemma 8. *If there exists a state τ^E such that $\|\omega^{RE} - \pi^R \otimes \tau^E\|_1 \le \epsilon$, it holds that*

$$\frac{1}{2}\|\omega_{\mathrm{targ}}^{RR'} - \Phi^{RR'}\|_1 \le \sqrt{\epsilon}. \tag{109}$$

This lemma is shown by a *vectorization* operation. A vectorization in a given basis $\{|i\rangle\}_i$ is a linear map **Vec** such that

$$\mathbf{Vec}(|\psi\rangle\langle\varphi|) = |\psi\rangle|\varphi^*\rangle, \tag{110}$$

where the complex conjugate is taken in the basis $\{|i\rangle\}_i$, i.e., $|\varphi^*\rangle = \sum_i c_i^*|i\rangle$ when $|\varphi\rangle = \sum_i c_i|i\rangle$. A vectorization has the property that

$$\|L - M\|_2 = \|\mathbf{Vec}(L) - \mathbf{Vec}(M)\|, \tag{111}$$

for any matrix L and M, where $\|\cdot\|_2$ is the Hilbert-Schmidt norm for matrices and $\|\cdot\|$ is the Euclidian norm for vectors.

Proof. (Proof of Lemma 8) Let τ_α and $\{|e_\alpha\rangle^E\}_\alpha$ be eigenvalues and eigenstates of τ^E, respectively, and let $|\tau\rangle^{EE'} := \sum_\alpha \sqrt{\tau_\alpha}|e_\alpha\rangle^E|e_\alpha^*\rangle^{E'}$, where the complex conjugate is taken in the computational basis $|\alpha\rangle^E$.

We regard the two pure states $|\omega_{\mathrm{targ}}\rangle^{RR'EE'}$ and $|\Phi\rangle^{RR'}|\tau\rangle^{EE'}$ as the states after the vectorization of operators on RE, which is taken in the computational basis $\{|i\rangle^R|\alpha\rangle^E\}_{i,\alpha}$. That is,

$$\begin{aligned}
&\| |\omega_{\mathrm{targ}}\rangle^{RR'EE'} - |\Phi\rangle^{RR'}|\tau\rangle^{EE'}\| \\
&= \Big\| \sum_\mu \sqrt{\frac{d_D}{d_B}} \sqrt{q_\mu} |\eta_\mu\rangle^{RE} |\eta_\mu^*\rangle^{R'E'} \\
&\qquad - \sum_{i,\alpha} \sqrt{\frac{\tau_\alpha}{d_A}} |i\rangle^R |e_\alpha\rangle^E |i\rangle^{R'} |e_\alpha^*\rangle^{E'} \Big\|
\end{aligned} \tag{112}$$

$$\begin{aligned}
&= \Big\| \mathbf{Vec}\Big(\sum_\mu \sqrt{\frac{d_D}{d_B}} \sqrt{q_\mu} |\eta_\mu\rangle\langle\eta_\mu|^{RE} \Big) \\
&\qquad - \mathbf{Vec}\Big(\sum_{i,\alpha} \sqrt{\frac{\tau_\alpha}{d_A}} |i\rangle\langle i|^R \otimes |e_\alpha\rangle\langle e_\alpha|^E \Big) \Big\|.
\end{aligned} \tag{113}$$

Using the property of the vectorization in Eq. (111). Eq. (113) is equal to

$$\begin{aligned}
&\Big\| \sum_\mu \sqrt{\frac{d_D}{d_B}} \sqrt{q_\mu} |\eta_\mu\rangle\langle\eta_\mu|^{RE} \\
&\qquad - \sum_{i,\alpha} \sqrt{\frac{\tau_\alpha}{d_A}} |i\rangle\langle i|^R \otimes |e_\alpha\rangle\langle e_\alpha|^E \Big\|_2
\end{aligned} \tag{114}$$

$$= \big\| \big(\omega_{\mathrm{targ}}^{RE}\big)^{1/2} - \big(\pi^R \otimes \tau^E\big)^{1/2} \big\|_2 \tag{115}$$

$$\le \|\omega_{\mathrm{targ}}^{RE} - \pi^R \otimes \tau^E\|_1^{1/2} \tag{116}$$

$$= \|\omega^{RE} - \pi^R \otimes \tau^E\|_1^{1/2} \tag{117}$$

$$\le \sqrt{\epsilon}. \tag{118}$$

In the first inequality we used the Powers-Størmer inequality [54, 55]: $\|L^{1/2} - M^{1/2}\|_2^2 \le \|L - M\|_1$ for Hermite operators L and M. The last equation follows as $\omega_{\mathrm{targ}}^{RE} = \omega^{RE}$, and the last inequality is by assumption.

From $\||v\rangle\langle v| - |w\rangle\langle w|\|_1 \le 2\||v\rangle - |w\rangle\|$ for any pure states $|v\rangle$ and $|w\rangle$, it follows that

$$\frac{1}{2}\|\omega_{\mathrm{targ}}^{RR'EE'} - \Phi^{RR'} \otimes \tau^{EE'}\|_1 \le \sqrt{\epsilon}. \tag{119}$$

Using the contraction property of the trace norm against the partial trace, we complete the proof. □

We now turn to investigate the QSVT-based FPAA algorithm. From Lemma 8, it suffices to show that the output state $\mathcal{D}_{t,\phi}^{DB'\to R'}(\omega^{RDB'})$ is closed to $\omega_{\mathrm{targ}}^{RR'}$. This is achieved by the operation such that $|\psi_\mu\rangle^{DD'E'R'} \to |\xi_\mu^*\rangle^{DD'E'R'} = |\Phi\rangle^{DD'}|\eta_\mu^*\rangle^{E'R'}$ for all μ. In fact, we

observe from Eqs. (101) and (108) that this operation achieves

$$|\omega_0\rangle^{REDD'E'R'} \to |\omega_{\rm targ}\rangle^{REE'R'}|\Phi\rangle^{DD'}, \tag{120}$$

whose reduced state on RR' is $\omega_{\rm targ}^{RR'}$. The goal below is to show that this operation is ahieved by the QSVT-based FPAA algoriothm with high accuracy.

Before we start, we comment on the crucial role of the QSVT-based FPAA algorithm rather than the standard AA algorithm. As we will soon show, when the QSVT-based FPAA algorithm or the AA algorithm is applied, $|\psi_\mu\rangle^{DD'E'R'}$ rotates toward $|\xi_\mu^*\rangle^{DD'E'R'}$ in each two-dimensional subspace $\mathcal{H}_\mu^{DD'E'R'}$. Hence, the decoding succeeds by stopping the rotation when all the states $|\psi_\mu\rangle^{DD'E'R'}$ *simultaneously* get close to the corresponding $|\xi_\mu^*\rangle^{DD'E'R'}$. If q_μ differs from each other, this simultaneous condition is hard to satisfied by the standard AA algorithm since it can over-rotate the state. This is the reason why we need to use the QSVT-based FPAA algorithm.

We make another small comment on the difference between the QSVT-based FPAA algorithm and the standard FPAA algorithm in [42]. When we use the standard FPAA algorithm instead of the QSVT-based FPAA algorithm, $|\psi_\mu\rangle^{DD'E'R'}$ still rotates in each subspace. However, the algorithm may end up with undesirable phases θ_μ, such as $|\psi_\mu\rangle^{DD'E'R'} \to e^{i\theta_\mu}|\xi_\mu^*\rangle^{DD'E'R'}$. In our case, these phases act as relative phases (see Eq. (101)), and results in the failure of the recovery. This issue is also circumvented by the QSVT-based FPAA algorithm [26, 27].

The following is an important lemma about the QSVT in our setting.

Lemma 9 (Quantum singular value transformation to real odd polynomials [26, 27, 47])**.** *Suppose that $Q_t(x)$ is any degree-t odd real polynomial satisfying $|Q_t(x)| \le 1$ for all $x \in [-1,1]$. Then, there exists $\phi \in (-\pi,\pi]^t$ such that*

$$\begin{aligned}(\Pi_2^{DD'} \otimes \langle 0|^H) G_{t,\phi}^{DD'E'R'H} (\Pi_1^{D'E'R'} \otimes |0\rangle^H) \\ = Q_t(\Pi_2^{DD'}\Pi_1^{D'E'R'}).\end{aligned} \tag{121}$$

The unitary $G_{t,\phi}^{DD'E'R'H}$ is given by Eq. (29), and $\Pi_1^{D'E'R'}$ and $\Pi_2^{DD'}$ are given by Eqs. (25) and (26), respectively. The system H is a single-qubit system.

By the Jordan's lemma, $\mathcal{H}_\mu^{DD'C'R'}$ is invariant under the action of $\Pi_1^{D'E'R'}$ and $\Pi_2^{DD'}$. Hence, it suffices to consider the action of $G_{t,\phi}^{DD'E'R'H}$ in each subspace $\mathcal{H}_\mu^{DD'E'R'H} := \mathrm{span}\{|\psi_\mu\rangle^{DD'E'R'}|0\rangle^H, |\xi_\mu^*\rangle^{DD'E'R'}|0\rangle^H\}$. We use a notation such as $|\check{\varphi}\rangle^{DD'E'R'H} = |\varphi\rangle^{DD'E'R'}|0\rangle^H$ for a state $|\varphi\rangle^{DD'E'R'}$. From Eq. (103), the state $|\check{\psi}_\mu\rangle^{DD'E'R'H}$ is expanded as

$$\begin{aligned}|\check{\psi}_\mu\rangle^{DD'E'R'H} = \sqrt{q_\mu}|\check{\xi}_\mu^*\rangle^{DD'E'R'H} \\ + \sqrt{1-q_\mu}|\check{\perp}_\mu\rangle^{DD'E'R'H},\end{aligned} \tag{122}$$

where $|\check{\perp}_\mu\rangle^{DD'E'R'H}$ is a state in $\mathcal{H}_\mu^{DD'E'R'H}$ orthogonal to $|\check{\xi}_\mu^*\rangle^{DD'E'R'H}$. From Lemma 9, the QSVT achieves the matrix transformation in $\mathcal{H}_\mu^{DD'E'R'H}$ such as

$$\mathbb{I}^{DD'E'R'H}|_{\mathcal{H}_\mu} = \begin{matrix}\\ |\check{\xi}_\mu^*\rangle \\ |\check{\perp}_\mu\rangle\end{matrix}\begin{matrix}\begin{matrix}\langle\check{\psi}_\mu| & \langle\check{\psi}_\mu^\perp|\end{matrix}\\ \begin{pmatrix}\sqrt{q_\mu} & \sqrt{1-q_\mu}\\ \sqrt{1-q_\mu} & -\sqrt{q_\mu}\end{pmatrix}\end{matrix} \tag{123}$$

$$\overset{\rm QSVT}{\longrightarrow} G_{t,\phi}^{DD'E'R'H}|_{\mathcal{H}_\mu} = \begin{matrix}\\ |\check{\xi}_\mu^*\rangle \\ |\check{\perp}_\mu\rangle\end{matrix}\begin{matrix}\begin{matrix}\langle\check{\psi}_\mu| & \langle\check{\psi}_\mu^\perp|\end{matrix}\\ \begin{pmatrix}Q_t(\sqrt{q_\mu}) & \cdot\\ \cdot & \cdot\end{pmatrix}\end{matrix}. \tag{124}$$

Here, $|\check{\psi}_\mu^\perp\rangle$ is the state in $\mathcal{H}_\mu^{DD'E'R'H}$ orthogonal to $|\check{\psi}_\mu\rangle^{DD'E'R'H}$.

It is clear from this representation that, if one chooses the polynomial $Q_t(\,\cdot\,)$ such that $Q_t(\sqrt{q_\mu}) \approx 1$ for all μ, the desired operation that transforms $|\psi_\mu\rangle$ into $|\xi_\mu^*\rangle$ is realized. A possible choice of such a polynomial is a polynomial approximating the sign function:

$$\mathrm{sign}(x) = \begin{cases} 1 & (x>1)\\ 0 & (x=0)\\ -1 & (x<0).\end{cases} \tag{125}$$

The following lemma shows that there exists such a polynomial approximating the sign function.

Lemma 10 (Polynomial approximation of the sign function [26–28, 56, 57])**.** *For any $\beta,\delta \in (0,1]$, there exists an odd integer $t = \Theta\big(\frac{1}{\beta}\log(1/\delta)\big)$ and a real polynomial $Q_t^{\rm sign}(x)$ of degree t such that*

- $x \in [-1,1] : |Q_t^{\rm sign}(x)| \le 1$,
- $x \in [-1,-\beta)\cup(\beta,1] : |Q_t^{\rm sign}(x) - \mathrm{sign}(x)| \le \delta$.

Given a polynomial, the corresponding ϕ can be computed in $\mathcal{O}(\mathrm{poly}(t))$ time by a classical computer [44–48], where t is the degree of the polynomial. We take the phase sequence $\phi = (\phi_1,\dots,\phi_t)$ so that the polynomial $Q_t(\,\cdot\,)$ in Eqs. (121) and (124) becomes $Q_t^{\rm sign}(\,\cdot\,)$. From Lemma 10, for $Q_t^{\rm sign}(\sqrt{q_\mu})$ to be larger than $1-\delta$ for all $\mu = 1,\dots,r$, it is necessary that $\sqrt{q_{\min}} \ge \beta$, where $q_{\min} := \min_{\mu\in[1,r]} q_\mu$. From $\omega_0^{RE} = \omega^{RE}$ and Eq. (101), the non-zero minimum eigenvalue of ω^{RE} is $\lambda_{\min}(\omega^{RE}) = \frac{d_D}{d_B}q_{\min}$. Hence, we take the odd integer t such that

$$t = \Theta\left(\frac{1}{\sqrt{q_{\min}}}\log(1/\delta)\right) \tag{126}$$

$$= \Theta\left(\sqrt{\frac{d_D}{d_B\lambda_{\min}(\omega^{RE})}}\log(1/\delta)\right). \tag{127}$$

We finally combine all together. We denote the output state of the QSVT-based FPAA algorithm by

$$|\breve{\omega}_t\rangle^{REDD'E'R'H} := G_{t,\phi}^{DD'E'R'H}|\omega_0\rangle^{REDD'E'R'}|0\rangle^H. \tag{128}$$

By taking t and ϕ as mentioned above to approximate the sign function, we obtain the overlap between this output state and the state $|\omega_{\rm targ}\rangle^{REE'R'}|\Phi\rangle^{DD'}|0\rangle^H$ as

$$\begin{aligned}
&\langle\omega_{\rm targ}|^{REE'R'}\langle\Phi|^{DD'}\langle 0|^H|\breve{\omega}_t\rangle^{REDD'E'R'H} && (129)\\
&= \frac{d_D}{d_B}\sum_{\mu=1}^r q_\mu \langle\check{\xi}_\mu^*|^{DD'E'R'H} G_{t,\phi}|\check{\psi}_\mu\rangle^{DD'E'R'H} && (130)\\
&= \frac{d_D}{d_B}\sum_{\mu=1}^r q_\mu Q_t^{\rm sign}(\sqrt{q_\mu}) && (131)\\
&\geq (1-\delta)\frac{d_D}{d_B}\sum_{\mu=1}^r q_\mu && (132)\\
&= 1-\delta, && (133)
\end{aligned}$$

where we use $\frac{d_D}{d_B}\sum_{\mu=1}^r q_\mu = 1$. Using the Fuchs-van de Graaf inequities and the contraction property of the trace norm, it follows that

$$\frac{1}{2}\|\breve{\omega}_t^{RR'} - \omega_{\rm targ}^{RR'}\|_1 \leq \sqrt{1-(1-\delta)^2} \leq \sqrt{2\delta}. \tag{134}$$

Note that the state $\breve{\omega}_t^{RR'}$ is the output state of the generalized YK decoder: $\breve{\omega}_t^{RR'} = \mathcal{D}_{t,\phi}^{DB'\to R'}(\omega^{RDB'})$. By Lemma 8, Eq. (134), and the triangle inequality, we have

$$\frac{1}{2}\|\breve{\omega}_t^{RR'} - \Phi^{RR'}\|_1 \leq \sqrt{\epsilon} + \sqrt{2\delta}, \tag{135}$$

completing the evaluation of the recovery error by the generalized YK decoder.

We next investigate the circuit complexity of the generalized YK decoder. Since the non-trivial part is to implement the unitary $G_{t,\phi}$ by the QSVT-based FPAA algorithm, we focus on $\mathcal{C}(G_{t,\phi})$.

We start with a circuit implementation of $W_m(\theta)$ for $m = 1, 2$:

$$\begin{aligned}
W_m(\theta) &= e^{i\theta(2\Pi_m - \mathbb{I})} && (136)\\
&= e^{-i\theta}\,\mathbb{I} - (e^{-i\theta} - e^{i\theta})\Pi_m. && (137)
\end{aligned}$$

To implement the unitary $W_m(\theta)$, we use the *projector-controlled NOT gate* [26, 27] that is in general defined for a projector Π on the system P as

$$\mathrm{C}_\Pi\mathrm{NOT}^{P\text{-}G} := \Pi^P \otimes X^G + (\mathbb{I}^P - \Pi^P)\otimes \mathbb{I}^G. \tag{138}$$

The order of the superscripts in the left-hand side indicates the controlling and controlled systems. The gate

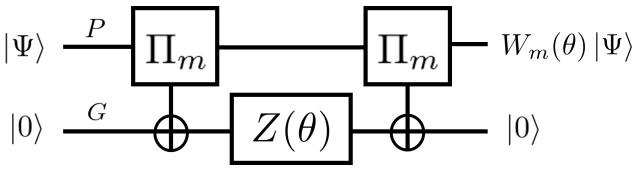


FIG. 9. A quantum circuit for implementing a unitary $W_m(\theta)^P$. The box in which a projector is written implies that this projector controls the gate. The circle drawn inside the intersecting lines represents the NOT gate, i.e., the Pauli-X gate.

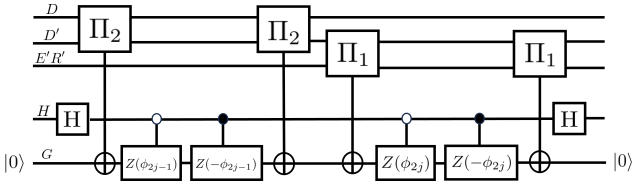


FIG. 10. A quantum circuit for implementing $W_1(\phi_{2j})^{D'E'R'}W_2(\phi_{2j-1})^{DD'}\otimes|+\rangle\langle+|^H + W_1(-\phi_{2j})^{D'E'R'}W_2(-\phi_{2j-1})^{DD'}\otimes|-\rangle\langle-|^H$. Open circles implies that the gates are controlled by $|0\rangle$, while closed circles indicate the ones controlled by $|1\rangle$.

X is the single-qubit Pauli-X gate. We also use a single-qubit rotation-Z gate:

$$\begin{aligned}
Z(\theta) &:= e^{-i\theta Z} && (139)\\
&= e^{-i\theta}|0\rangle\langle 0| + e^{i\theta}|1\rangle\langle 1|. && (140)
\end{aligned}$$

It is straightforward to check that, for any state $|\Psi\rangle^P$,

$$\begin{aligned}
&(\mathrm{C}_\Pi\mathrm{NOT}^{P\text{-}G}Z(\theta)^G\mathrm{C}_\Pi\mathrm{NOT}^{P\text{-}G})(|\Psi\rangle^P\otimes|0\rangle^G)\\
&\quad = \left[e^{-i\theta}\,\mathbb{I}^P - (e^{-i\theta} - e^{i\theta})\Pi^P)|\Psi\rangle^P\right]\otimes|0\rangle^G.
\end{aligned} \tag{141}$$

Hence, we can implement $W_m(\theta)^P$ by preparing a single-qubit system G and by operating a quantum circuit in Fig. 9.

To construct a circuit for $G_{t,\phi}$, we prepare another single-qubit system H for the controlled implementation of $W_m(\theta)^P$. For instance, a quantum circuit implementing

$$\begin{aligned}
&W_1(\phi_{2j})^{D'E'R'}W_2(\phi_{2j-1})^{DD'}\otimes|+\rangle\langle+|^H\\
&\quad + W_1(-\phi_{2j})^{D'E'R'}W_2(-\phi_{2j-1})^{DD'}\otimes|-\rangle\langle-|^H,
\end{aligned} \tag{142}$$

is given in Fig. 10. By applying the circuit $(t-1)/2$ times with various phases and finally applying $W_2(\phi_t)^{DD'}\otimes \mathrm{H}^H$, the unitary $G_{t,\phi}$ is realized. Here, the gate H^H is the single-qubit Hadamard gate on the system H.

In this construction, the unitary $G_{t,\phi}$ is decomposed into two unitaries $\mathrm{C}_{\Pi_1}\mathrm{NOT}^{D'E'R'\text{-}G}$ and $\mathrm{C}_{\Pi_2}\mathrm{NOT}^{DD'\text{-}G}$. A quantum circuit for $\mathrm{C}_{\Pi_1}\mathrm{NOT}^{D'E'R'\text{-}G}$ is given in Fig. 11. The unitary $\mathrm{C}_{|0\rangle\langle 0|}\mathrm{NOT}^{P\text{-}G}$ can be implemented using $\mathcal{O}(\log d_P)$ single- and two-qubit gates and $\mathcal{O}(\log d_P)$ ancilla qubits [58], and the unitary $U_\Phi^{A'R'}$,

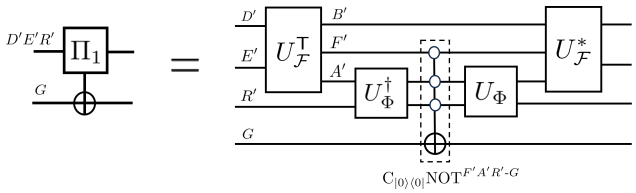


FIG. 11. A quantum circuit for implementing the protector-controlled NOT gate $\mathrm{C}_{\Pi_1}\mathrm{NOT}^{D'E'R'\text{-}G}$. The dashed box represents the gete $\mathrm{C}_{|0\rangle\langle 0|}\mathrm{NOT}^{F'A'R'\text{-}G}$.

which is given by

$$U_{\Phi}^{A'R'}|0\rangle^{A'}|0\rangle^{R'} = |\Phi\rangle^{A'R'}, \tag{143}$$

can be implemented using $\mathcal{O}(\log d_A)$ gates. Hence, in total, $\mathrm{C}_{\Pi_1}\mathrm{NOT}^{D'E'R'\text{-}G}$ can be implemented by

$$\mathcal{O}\big(\mathcal{C}(U_{\mathcal{F}}) + \log\left(d_A d_F\right)\big) \tag{144}$$

gates and $\mathcal{O}(\log d_A d_F)$ ancilla qubits. Similarly, $\mathrm{C}_{\Pi_2}\mathrm{NOT}^{DD'\text{-}G}$ can be implemented using $\mathcal{O}(\log d_D)$ gates and $\mathcal{O}(\log d_D)$ ancilla qubits.

In the unitary $G_{t,\phi}$, these projector-controlled NOT gates are used $\mathcal{O}(t)$ times. Thus, the total complexity of the generalized YK decoder is given by

$$\begin{aligned}\mathcal{C}(\mathcal{D}_{t,\phi}) &= \mathcal{O}\Big(t\left(\mathcal{C}(U_{\mathcal{F}}) + \log(d_A d_F d_D)\right)\Big) \\ &\quad + \mathcal{C}(U_{\mathcal{F}}) + \mathcal{O}(\log d_A) \qquad (145)\\ &= \mathcal{O}\Big(t\left(\mathcal{C}(U_{\mathcal{F}}) + \log(d_D^2 d_E/d_B)\right)\Big), \qquad (146)\end{aligned}$$

with $\mathcal{O}\big(\log\left(d_D^2 d_E/d_B\right)\big)$ ancilla qubits. Here, we used $d_A d_B d_F = d_E d_D$. In Eq. (145), the first line in the right-hand side comes from $G_{t,\phi}$ and the second line comes from $\mathcal{V}^{A'B'\to E'D'}$, which is applied before $G_{t,\phi}$.

B. Proofs: the Petz-like decoder

Similarly to the YK decoder, we first consider the decoding protocol with post-selection and then provide a sketch of a proof of Theorem 4.

From Eqs. (17), (35), and (36), the success probability $\tilde{p}_{\mathrm{succ}}$ is computed as

$$\begin{aligned}\tilde{p}_{\mathrm{succ}} &= \frac{d_A}{d_E}\operatorname{Tr}\left[V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}(\Phi^{\hat{B}B'}\otimes\pi^{R'})\right.\\ &\qquad\left.(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D})^{\dagger}(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})\right] \qquad (147)\\ &= \frac{d_A}{d_E}\operatorname{Tr}\left[(\omega^{DB'})^2\right] \qquad (148)\\ &= \frac{d_A}{d_E}2^{-H_2(DB')_\omega} \qquad (149)\\ &= \frac{d_A}{d_E}2^{-H_2(RE)_\omega}, \qquad (150)\end{aligned}$$

where we used $H_2(DB')_\omega = H_2(RE)_\omega$ as $|\omega\rangle^{REDB'}$ is pure. The fidelity between $\tilde{\zeta}_{\mathrm{succ}}^{RR'}$ and $\Phi^{RR'}$ is computed as

$$\begin{aligned}&\mathrm{F}(\tilde{\zeta}_{\mathrm{succ}}^{RR'}, \Phi^{RR'})\\ &= \frac{1}{d_E\tilde{p}_{\mathrm{succ}}}\operatorname{Tr}\left[V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D}(\Phi^{RR'}\otimes\Phi^{\hat{B}B'})\right.\\ &\qquad\left.(V_{\mathcal{F}}^{R'\hat{B}\to\hat{E}D})^{\dagger}(\omega^{RDB'}\otimes\mathbb{I}^{\hat{E}})\right] \qquad (151)\\ &= \frac{1}{d_A}2^{H_2(RE)_\omega}\operatorname{Tr}\left[\omega^{R\hat{E}DB'}(\omega^{RB'D}\otimes\mathbb{I}^{\hat{E}})\right] \qquad (152)\\ &= \frac{1}{d_A}2^{H_2(RE)_\omega - H_2(RDB')_\omega} \qquad (153)\\ &= \frac{1}{d_A}2^{H_2(RE)_\omega - H_2(E)_\omega}, \qquad (154)\end{aligned}$$

by using $H_2(RDB')_\omega = H_2(E)_\omega$ for $|\omega\rangle^{REDB'}$. Hence, we obtained Eqs. (37) and (38).

Let us now turn to the proof of Theorem 4. Since it can be shown similar to Theorem 3, we provide only an outline of the proof.

We denote the input state of the QSVT-based FPAA algorithm by

$$\tilde{\omega}_0^{RE'R'\hat{F}\hat{B}B'} := \tilde{\mathcal{V}}^{D\to E'R'\hat{F}\hat{B}}(\omega^{RDB'}), \tag{155}$$

where $\tilde{\mathcal{V}}^{D\to E'R'\hat{F}\hat{B}}$ is the isometry map such that

$$\tilde{\mathcal{V}}^{D\to E'R'\hat{F}\hat{B}} = (U_{\mathcal{F}}^{\hat{L}})^{\dagger}(\,\cdot\otimes\Phi^{\hat{E}E'})U_{\mathcal{F}}^{\hat{L}}, \tag{156}$$

and $\hat{L} = R'\hat{F}\hat{B} = \hat{E}D$. Note that $\tilde{\omega}_0^{RE} = \omega^{RE}$. Let $|\tilde{\omega}_0\rangle^{REE'R'\hat{F}\hat{B}B'}$ be the purified state which is given by

$$|\tilde{\omega}_0\rangle^{REE'R'\hat{F}\hat{B}B'} = (U_{\mathcal{F}}^{\hat{L}})^{\dagger}|\omega\rangle^{REDB'}|\Phi\rangle^{\hat{E}E'}. \tag{157}$$

The state on $REE'R'$ after the post-selection is then given by

$$|\tilde{\zeta}_{\mathrm{succ}}\rangle^{REE'R'} = \frac{1}{\sqrt{\tilde{p}_{\mathrm{succ}}}}\langle 0|^{\hat{F}}\langle\Phi|^{\hat{B}B'}|\tilde{\omega}_0\rangle^{REE'R'\hat{F}\hat{B}B'}. \tag{158}$$

It is important to observe that

$$\tilde{\zeta}_{\mathrm{succ}}^{REE'R'} = \zeta_{\mathrm{succ}}^{REE'R'}, \tag{159}$$

where the right-hand side is the state after the post-selection in the generalized YK protocol. Although it may be hard to observe this relation from its construction in Fig. 5, it can be readily shown using Lemma 6 as in Fig. 12. From this relation, it turns out that the state $\tilde{\zeta}_{\mathrm{succ}}^{REE'R'}$ is also symmetrical between RE and $E'R'$ up to the complex conjugate, and thus, the Schmidt basis in RE and that in $E'R'$ are complex conjugate of each other.

We next compute the products of projectors $\tilde{\Pi}_1^{E'R'\hat{F}\hat{B}}$ and $\tilde{\Pi}_2^{\hat{F}\hat{B}B'}$, which are defined in Eqs.(39) and (40). Sim-

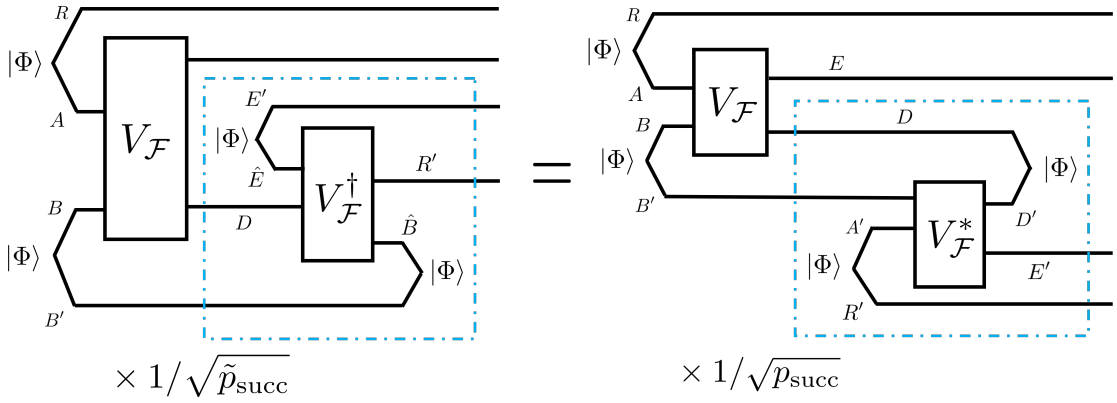


FIG. 12. The equivalence of the states $\tilde{\zeta}_{\mathrm{succ}}^{REE'R'}$ and $\zeta_{\mathrm{succ}}^{REE'R'}$, which are obtained after the post-selection in the Petz-like protocol and in the generalized YK protocol, respectively. We can derive this equivalence by applying Lemma 6 onto the portion enclosed by the blue dash-dotted lines.

ilarly to Eqs. (99) and (100), we obtain

$$(\tilde{\Pi}_1\tilde{\Pi}_2\tilde{\Pi}_1)^{E'R'\hat{F}\hat{B}B'} = \frac{d_A}{d_E}\tilde{\omega}_0^{E'R'\hat{F}\hat{B}B'}, \tag{160}$$

$$(\tilde{\Pi}_2\tilde{\Pi}_1\tilde{\Pi}_2)^{E'R'\hat{F}\hat{B}B'} = \Big(\frac{d_A\tilde{p}_{\mathrm{succ}}}{d_E}\tilde{\zeta}_{\mathrm{succ}}^{E'R'}\Big)^{1/2} \otimes |0\rangle\langle 0|^{\hat{F}} \otimes |\Phi\rangle\langle\Phi|^{BB'}. \tag{161}$$

Let $\tilde{q}_\mu$ and $|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}$ for $\mu = 1, 2, \ldots, r$ be non-zero eigenvalues and corresponding eigenstates of $(\tilde{\Pi}_1\tilde{\Pi}_2\tilde{\Pi}_1)^{E'R'\hat{F}\hat{B}B'}$, respectively. From Eq. (160), the Schmidt decomposition of $|\tilde{\omega}_0\rangle^{REE'R'\hat{F}\hat{B}B'}$, divided into RE and $E'R'\hat{F}\hat{B}B'$, is given by

$$|\tilde{\omega}_0\rangle^{REE'R'\hat{F}\hat{B}B'} = \sum_{\mu=1}^{r}\sqrt{\frac{d_E}{d_A}}\sqrt{\tilde{q}_\mu}\,|\eta_\mu\rangle^{RE}|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}, \tag{162}$$

where $\{|\eta_\mu\rangle^{RE}\}_\mu$ is an orthonormal basis. As $\tilde{\omega}_0^{RE}$ is equal to ω_0^{RE}, we have that $\tilde{q}_\mu = \frac{d_Ad_D}{d_Bd_E}q_\mu$.

Since the state $|\tilde{\zeta}_{\mathrm{succ}}\rangle$ is defined by using $|\tilde{\omega}_0\rangle$ as Eq. (158), it follows that

$$|\tilde{\zeta}_{\mathrm{succ}}\rangle^{REE'R'} = \sum_{\mu=1}^{r}\sqrt{\frac{d_E\tilde{q}_\mu}{d_A\tilde{p}_{\mathrm{succ}}}}\,|\eta_\mu\rangle^{RE}\langle 0|^{\hat{F}}\langle\Phi|^{\hat{B}B'}|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}. \tag{163}$$

From Eq. (104) for $|\zeta_{\mathrm{succ}}\rangle$ in the generalized YK protocol with post-selection and Eq. (159), we have

$$\langle 0|^{\hat{F}}\langle\Phi|^{\hat{B}B'}|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'} = \sqrt{\frac{d_Ad_D\tilde{p}_{\mathrm{succ}}}{d_Bd_Ep_{\mathrm{succ}}}}\frac{q_\mu}{\sqrt{\tilde{q}_\mu}}|\eta_\mu^*\rangle^{E'R'} = \sqrt{\tilde{q}_\mu}\,|\eta_\mu^*\rangle^{E'R'}. \tag{164}$$

Here, we substituted the success probabilities p_{succ} and $\tilde{p}_{\mathrm{succ}}$ in the generalized YK and Petz-like protocols with post-selection, which are given by Eqs. (23) and (37). We have also used $\tilde{q}_\mu = \frac{d_Ad_D}{d_Bd_E}q_\mu$.

Applying the Jordan's lemma (Lemma 7) to the projectors $\tilde{\Pi}_1^{E'R'\hat{F}\hat{B}}$ and $\tilde{\Pi}_2^{\hat{F}\hat{B}B'}$, the Hilbert space $\mathcal{H}^{E'R'\hat{F}\hat{B}B'}$ is decomposed into a direct sum of one- and two-dimensional subspaces $\mathcal{H}_\mu^{E'R'\hat{F}\hat{B}B'}$ and the remaining orthogonal complement $\mathcal{H}_\perp^{E'R'\hat{F}\hat{B}B'}$ such that

$$\mathcal{H}^{E'R'\hat{F}\hat{B}B'} = \oplus_{\mu=1}^{r}\mathcal{H}_\mu^{E'R'\hat{F}\hat{B}B'} \oplus \mathcal{H}_\perp^{E'R'\hat{F}\hat{B}B'}, \tag{165}$$

where $\mathcal{H}_\mu^{E'R'\hat{F}\hat{B}B'}$ is given by

$$\mathcal{H}_\mu^{E'R'\hat{F}\hat{B}B'} = \mathrm{span}\{|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}, |\eta_\mu^*\rangle^{E'R'}|0\rangle^{\hat{F}}|\Phi\rangle^{\hat{B}B'}\}. \tag{166}$$

From Eqs. (160) and (161), all eigenstates of $\tilde{\omega}_0^{E'R'\hat{F}\hat{B}B'}$ and $\tilde{\zeta}_{\mathrm{succ}}^{E'R'}$ are in $\oplus_{\mu=1}^{r}\mathcal{H}_\mu^{E'R'\hat{F}\hat{B}B'}$, on which we focus in the following.

In each subspace $\mathcal{H}_\mu^{E'R'\hat{F}\hat{B}B'}$, the state $|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}$ is decomposed as

$$|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'} = \sqrt{\tilde{q}_\mu}\,|\eta_\mu^*\rangle^{E'R'}|0\rangle^{\hat{F}}|\Phi\rangle^{\hat{B}B'} + \sqrt{1-\tilde{q}_\mu}\,|\tilde{\perp}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}, \tag{167}$$

where $|\tilde{\perp}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}$ is a state in $\mathcal{H}_\mu^{E'R'\hat{F}\hat{B}B'}$ orthogonal to $|\eta_\mu^*\rangle^{E'R'}|0\rangle^{\hat{F}}|\Phi\rangle^{\hat{B}B'}$. By the QSVT-based FPAA algorithm with appropriately chosen $\phi \in (-\pi, \pi]^t$, $|\tilde{\psi}_\mu\rangle^{E'R'\hat{F}\hat{B}B'}$ is transformed to $|\eta_\mu^*\rangle^{E'R'}|0\rangle^{\hat{F}}|\Phi\rangle^{\hat{B}B'}$ in each subspace. Hence, it approximately achieves the transformation that

$$|\tilde{\omega}_0\rangle^{REE'R'\hat{F}\hat{B}B'} \mapsto |\omega_{\mathrm{targ}}\rangle^{REE'R'}|0\rangle^{\hat{F}}|\Phi\rangle^{\hat{B}B'}, \tag{168}$$

where $|\omega_{\mathrm{targ}}\rangle^{REE'R'}$ is defined as Eq. (108). Thus, by a similar technique to the generalized YK decoder, we obtain that the Petz-like decoder $\tilde{\mathcal{D}}_{t,\phi}$ achieves

$$\frac{1}{2}\|\tilde{\mathcal{D}}_{t,\phi}^{DB'\to R'}(\omega^{RDB'}) - \omega_{\mathrm{targ}}^{RR'}\| \le \sqrt{2\delta}, \tag{169}$$

where t is an odd number satisfying

$$t = \Theta\left(\frac{1}{\sqrt{\tilde{q}_{\min}}}\log\left(1/\delta\right)\right) \tag{170}$$

$$= \Theta\left(\sqrt{\frac{d_A}{d_E\lambda_{\min}(\omega^{RE})}}\log\left(1/\delta\right)\right). \tag{171}$$

From Lemma 8, $\omega_{\rm targ}^{RR'} \approx \Phi^{RR'}$ when the decoupling condition is satisfied. Hence, using the triangle inequality, the recovery error by the Petz-like decoder is evaluated as

$$\frac{1}{2}\|\tilde{\mathcal{D}}_{t,\phi}^{DB'\to R'}(\omega^{RDB'}) - \Phi^{RR'}\|_1 \le \sqrt{\epsilon} + \sqrt{2\delta}. \tag{172}$$

Finally, since $\tilde{\Pi}_1^{E'R'\hat{F}\hat{B}}$ and $\tilde{\Pi}_2^{\hat{F}\hat{B}B'}$ are explicitly given by Eqs. (39) and (40), respectively, the complexity of the Petz-like decoder can be evaluated similarly to the generalized YK decoder. The circuit complexity of $\mathrm{C}_{\tilde{\Pi}_1}\mathrm{NOT}$ is

$$\mathcal{O}\Big(\mathcal{C}(U_{\mathcal{F}}) + \log d_E\Big), \tag{173}$$

and that of $\mathrm{C}_{\tilde{\Pi}_2}\mathrm{NOT}$ is

$$\mathcal{O}\big(\log d_B d_F\big). \tag{174}$$

Since they are applied $\mathcal{O}(t)$ times in the Petz-like decoder, the total complexity is given by

$$\mathcal{O}\Big(t\left(\mathcal{C}(U_{\mathcal{F}}) + \log\left(d_E d_B d_F\right)\right)\Big), \tag{175}$$

with $\mathcal{O}\big(\log\left(d_E d_B d_F\right)\big)$ ancilla qubits. Using $d_A d_B d_F = d_E d_D$, Theorem 4 is obtained.

V. SUMMARY AND OUTLOOKS

In this paper, we have provided two explicit decoders that are applicable to any encoding and noisy channels: one is the generalized YK decoder, and the other is the Petz-like decoder. Both are constructed by two steps: first we consider a decoding protocol with measurement and post-selection, and then we construct a decoder by replacing the measurement with the QSVT-based FPAA algorithm, which is for amplifying the success probability of the post-selection. These decoders have been shown to have high recovering performance in the sense that they can recover quantum information when the recovery is guaranteed to be in principle possible, which is formulated in terms of the decoupling condition. An important implication is that the decoders with a suitable choice of encoding are capacity-achieving.

We have then investigated the circuit complexity of the generalized YK decoder and the Petz-like decoder. While the complexity depends on various factors, we have shown that the generalized YK decoder has smaller complexity in general if the sender and the receiver share more entanglement in advance. This conclusion was obtained by comparing the dominant term, i.e., the one that comes from the implementation of the QSVT-based FPAA algorithm.

Our approach extends the powerful use of the QSVT to the problem of recovering quantum information, which bridges quantum algorithms to quantum information theory, and is of conceptual interest. As mentioned, this approach was proposed in the original work by Yoshida and Kitaev [16] with a limited use in a specific model, where the standard AA algorithm was used. Our work shows that, if one uses the QSVT-based FPAA algorithm instead of the standard AA algorithm, the approach can be extended to general situations. The constructed decoder is still inefficient in general, but it would be an interesting open problem to see if an efficient decoder can be constructed by this approach.

It may also be interesting to address the question about whether a similar approach may work for recovering *classical* [59, 60] or *hybrid* [61–64] information. In the former, the encoded information is classical, and the decoder is simply given by quantum measurement. In the latter, the information is a mixture of classical and quantum, which can be decoded by a simultaneous use of quantum measurement and quantum decoder. Both use quantum measurement, and a couple of quantum measurements are known to work well, such as the pretty-good measurement [60, 65]. Our approach adapted to these settings may provide a better decoder.

From a technical viewpoint, another direction is a relaxation of the assumptions about the knowledge of the noisy channel [66] and the non-zero minimum eigenvalue of the noisy state. While general decoders, as well as the proposed decoders in this paper, are constructed based on such knowledge, it would not be realistic to obtain complete knowledge of the noise. If we can relax these assumptions, the decoders become more practical ones. An intriguing future challenge lies in understanding to what extent we can relax those restrictions.

These decoders may also have potential use in fundamental physics for exploring exotic quantum many-body phenomena that are related to the recovery of quantum information. For instance, the proposed decoders could be potentially applied to reconstructing the internal structure of a black hole from the noisy Hawking radiation [67], and to recovering the bulk structure from a part of boundaries, such as the entanglement wedge reconstruction [68]. This is also an intriguing direction of study with the decoders.

ACKNOWLEDGMENTS

T. U. and Y. N. were supported by JST CREST Grant Number JPMJCR23I3. T. U. was supported by JST SPRING Grant Number JPMJSP2108. Y. N. was supported by MEXT-JSPS Grant-in-Aid for Transformative Research Areas (A) "Extreme Universe" Grant Numbers

JP21H05182 and JP21H05183, and by JSPS KAKENHI Grant Number JP22K03464. The authors thank Takaya Matsuura, Shiro Tamiya, and Ryuji Takagi for valuable discussions.

[1] P. Hayden and J. Preskill, Black holes as mirrors: quantum information in random subsystems, J. High Energy Phys. **2007**, 120 (2007).

[2] D. Harlow and P. Hayden, Quantum computation vs. firewalls, J. High Energy Phys. **2013**, 1 (2013).

[3] Y. Nakata, E. Wakakuwa, and M. Koashi, Black holes as clouded mirrors: the Hayden-Preskill protocol with symmetry, Quantum **7**, 928 (2023).

[4] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence, J. High Energy Phys. **2015**, 1 (2015).

[5] A. Almheiri, X. Dong, and D. Harlow, Bulk locality and quantum error correction in AdS/CFT, J. High Energy Phys. **2015**, 1 (2015).

[6] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. **43**, 4452 (2002).

[7] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. **303**, 2 (2003).

[8] A. Kitaev, Anyons in an exactly solved model and beyond, Ann. Phys. **321**, 2 (2006).

[9] P. Hosur, X.-L. Qi, D. A. Roberts, and B. Yoshida, Chaos in quantum channels, J. High Energy Phys. **2016**, 1 (2016).

[10] D. A. Roberts and B. Yoshida, Chaos and complexity by design, J. High Energy Phys. **2017**, 1 (2017).

[11] Y. Nakata and M. Tezuka, Hayden-Preskill recovery in Hamiltonian systems, Phys. Rev. Res. **6**, L022021 (2024).

[12] P. Hayden, M. Horodecki, A. Winter, and J. Yard, A decoupling approach to the quantum capacity, Open Syst. Inf. Dyn. **15**, 7 (2008).

[13] F. Dupuis, *The decoupling approach to quantum information theory*, Ph.D. thesis, University of Montreal (2010).

[14] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, The decoupling theprem, arXiv:1012.6044 (2010).

[15] H. Barnum and E. Knill, Reversing quantum dynamics with near-optimal quantum and classical fidelity, J. Math. Phys. **43**, 2097 (2002).

[16] B. Yoshida and A. Kitaev, Efficient decoding for the Hayden-Preskill protocol, arXiv:1710.03363 (2017).

[17] Y. Nakata, T. Matsuura, and M. Koashi, Constructing quantum decoders based on complementarity principle, arXiv:2210.06661 (2022).

[18] D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, Commun. Math Phys. **105**, 123 (1986).

[19] D. Petz, Sufficiency of channels over von Neumann algebras, Q. J. Math. **39**, 97 (1988).

[20] A. Gilyén, S. Lloyd, I. Marvian, Y. Quek, and M. M. Wilde, Quantum algorithm for Petz recovery channels and pretty good measurements, Phys. Rev. Lett. **128**, 220502 (2022).

[21] B. Yoshida, Decoding the entanglement structure of monitored quantum circuits, arXiv:2109.08691 (2021).

[22] Y. Nakayama, A. Miyata, and T. Ugajin, The Petz (lite) recovery map for scrambling channel, arXiv:2310.18991 (2023).

[23] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, ACM (1996), pp. 212–219.

[24] G. Brassard and P. Høyer, An exact quantum polynomial-time algorithm for Simon's problem, in *Proceedings of the Fifth Israeli Symposium on Theory of Computing and Systems*, ISTCS '97, IEEE C. S. (1997), pp. 12–23.

[25] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, Quantum amplitude amplification and estimation, Contemp. Math. **305**, 53 (2002).

[26] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, Quantum Singular Value Transformation and beyond: Exponential Improvements for Quantum Matrix Arithmetics, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC '19, ACM (2019), pp. 193–204.

[27] A. Gilyén, *Quantum Singular Value Transformation & Its Algorithmic Applications*, Ph.D. thesis, University of Amsterdam (2019).

[28] J. M. Martyn, Z. M. Rossi, A. K. Tan, and I. L. Chuang, Grand Unification of Quantum Algorithms, PRX Quantum **2**, 040203 (2021).

[29] S. Lloyd, Capacity of the noisy quantum channel, Phys. Rev. A **55**, 1613 (1997).

[30] P. W. Shor, *The quantum channel capacity and coherent information*, Lecture notes, MSRI Workshop on Quantum Computation (2002).

[31] I. Devetak, The private classical capacity and quantum capacity of a quantum channel, IEEE Trans. Inf. Theory **51**, 44 (2005).

[32] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem, IEEE Trans. Inf. Theory **48**, 2637 (2002).

[33] M. M. Wilde, From classical to quantum Shannon theory, arXiv:1106.1445 (2011).

[34] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels, IEEE Trans. Inf. Theory **60**, 2926 (2014).

[35] A. Uhlmann, The "transition probability" in the state space of a$*$-algebra, Rep. Math. Phys. **9**, 273 (1976).

[36] C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[37] J. Watrous, *The Theory of Quantum Information*, Cambridge University Press (2018).

[38] It is known that the recovery errors defined by other metrics, such as the diamond norm, are closely related [31, 69].

[39] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, One-shot decoupling, Commun. Math Phys. **328**, 251 (2014).

[40] S. Beigi, N. Datta, and F. Leditzky, Decoding quantum information via the Petz recovery map, J. Math. Phys.

57 (2016).
[41] L. K. Grover, Fixed-point quantum search, Phys. Rev. Lett. **95**, 150501 (2005).
[42] T. J. Yoder, G. H. Low, and I. L. Chuang, Fixed-point quantum search with an optimal number of queries, Phys. Rev. Lett. **113**, 210501 (2014).
[43] B. Yan, S. Wei, H. Jiang, H. Wang, Q. Duan, Z. Ma, and G.-L. Long, Fixed-point oblivious quantum amplitude-amplification algorithm, Sci. Rep. **12**, 14339 (2022).
[44] J. Haah, Product Decomposition of Periodic Functions in Quantum Signal Processing, Quantum **3**, 190 (2019).
[45] R. Chao, D. Ding, A. Gilyén, C. Huang, and M. Szegedy, Finding Angles for Quantum Signal Processing with Machine Precision., arXiv: Quantum Physics (2020).
[46] Y. Dong, X. Meng, K. B. Whaley, and L. Lin, Efficient phase-factor evaluation in quantum signal processing, Phys. Rev. A **103**, 042419 (2021).
[47] L. Lin, Lecture Notes on Quantum Algorithms for Scientific Computation, arXiv:2201.08309 (2022).
[48] K. Mizuta and K. Fujii, Recursive Quantum Eigenvalue/Singular-Value Transformation: Analytic Construction of Matrix Sign Function by Newton Iteration, arXiv:2304.13330 (2023).
[49] S. Aaronson and P. Christiano, Quantum money from hidden subspaces, in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, STOC '12, ACM (2012), pp. 41–60.
[50] In this scenario, while we assume that the state on BB' is the MES $|\Phi\rangle^{BB'}$, the Petz-like decoder works even for an arbitrary state $|\rho\rangle^{BB'}$ (one example: the thermofield double state). In such cases, replacing every $|\Phi\rangle^{BB'}$ which appears in this section with $|\rho\rangle^{BB'}$ should suffice.
[51] C. Jordan, Essai sur la géométrie à n dimensions, Bull. Soc. Math. Fr. **3**, 103 (1875).
[52] O. Regev, *Witness-preserving Amplification of QMA*, Lecture Notes, Tel Aviv University (2006).
[53] A. Prakash, *Quantum Algorithms for Linear Algebra and Machine Learning*, Ph.D. thesis, University of California (2014).
[54] R. T. Powers and E. Størmer, Free states of the canonical anticommutation relations, Commun. Math Phys. **16**, 1 (1970).
[55] F. Kittaneh and H. Kosaki, Inequalities for the Schatten p-norm V, Publ. Res. Inst. Math. Sci. **23**, 433–443 (1987).
[56] G. H. Low and I. L. Chuang, Hamiltonian Simulation by Uniform Spectral Amplification, arXiv: Quantum Physics (2017).
[57] L. Lin and Y. Tong, Near-optimal ground state preparation, Quantum **4**, 372 (2020).
[58] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, Cambridge University Press (2010).
[59] B. Schumacher and M. D. Westmoreland, Sending classical information via noisy quantum channels, Phys. Rev. A **56**, 131 (1997).
[60] A. S. Holevo, The capacity of the quantum channel with general signal states, IEEE Trans. Inf. Theory **44**, 269 (1998).
[61] I. Devetak and P. W. Shor, The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information, Commun. Math Phys. **256**, 287 (2005).
[62] M.-H. Hsieh and M. M. Wilde, Entanglement-assisted communication of classical and quantum information, IEEE Trans. Inf. Theory **56**, 4682 (2010).
[63] Y. Nakata, E. Wakakuwa, and H. Yamasaki, One-shot quantum error correction of classical and quantum information, Phys. Rev. A **104**, 012408 (2021).
[64] E. Wakakuwa and Y. Nakata, One-Shot Triple-Resource Trade-Off in Quantum Channel Coding, IEEE Trans. Inf. Theory **69**, 2400 (2023).
[65] P. Hausladen and W. K. Wootters, A 'Pretty Good' Measurement for Distinguishing Quantum States, J. Mod. Opt. **41**, 2385 (1994).
[66] I. Bjelaković, H. Boche, and J. Nötzel, Entanglement transmission and generation under channel uncertainty: Universal quantum channel coding, Commun. Math Phys. **292**, 55 (2009).
[67] N. Bao and Y. Kikuchi, Hayden-Preskill decoding from noisy Hawking radiation, J. High Energy Phys. **02**, 017 (2021).
[68] C.-F. Chen, G. Penington, and G. Salton, Entanglement wedge reconstruction using the Petz map, J. High Energy Phys. **2020**, 1 (2020).
[69] D. Kretschmann and R. F. Werner, Tema con variazioni: quantum channel capacity, New J. Phys. **6**, 26 (2004).

Appendix A: Derivation of Eqs. (99) and (100)

In this section, we derive Eqs. (99) and (100). The calculations are as follows.

$$
\begin{aligned}
&(\Pi_1\Pi_2\Pi_1)^{DD'E'R'} \\
&= (V_{\mathcal{F}}^{A'B'\to E'D'})^*|\Phi\rangle\langle\Phi|^{A'R'}(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}|\Phi\rangle^{DD'} \\
&\quad (V_{\mathcal{F}}^{A'B'\to E'D'})^*|\Phi\rangle\langle\Phi|^{A'R'}(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}} && \text{(A1)} \\
&= \frac{d_B d_E}{d_A d_D}(V_{\mathcal{F}}^{A'B'\to E'D'})^*|\Phi\rangle^{A'R'}\langle\Phi|^{EE'}V_{\mathcal{F}}^{AB\to ED} \\
&\quad |\Phi\rangle\langle\Phi|^{BB'}(V_{\mathcal{F}}^{AB\to ED})^\dagger|\Phi\rangle^{EE'}\langle\Phi|^{A'R'}(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}} && \text{(A2)} \\
&= \frac{d_B}{d_D}(V_{\mathcal{F}}^{A'B'\to E'D'})^*(\omega^{DB'}\otimes\Phi^{A'R'})(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}} && \text{(A3)} \\
&= \frac{d_B}{d_D}\omega_0^{DD'E'R'}, && \text{(A4)}
\end{aligned}
$$

where we used Lemma 6 in the second equation. Note that $\omega^{DB'}$ is given by $\omega^{DB'} = \mathrm{Tr}_E[\,V_{\mathcal{F}}^{AB\to ED}(\pi^A\otimes\Phi^{BB'})(V_{\mathcal{F}}^{AB\to ED})^{\mathsf{T}}\,]$.

The other one is calculated as

$$
\begin{aligned}
&\left[(\Pi_2\Pi_1\Pi_2)^{DD'E'R'}\right]^2 \\
&= |\Phi\rangle\langle\Phi|^{DD'}(V_{\mathcal{F}}^{A'B'\to E'D'})^*|\Phi\rangle\langle\Phi|^{A'R'} \\
&\quad |\Phi\rangle\langle\Phi|^{DD'}(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}(V_{\mathcal{F}}^{A'B'\to E'D'})^* \\
&\quad |\Phi\rangle\langle\Phi|^{A'R'}(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}|\Phi\rangle\langle\Phi|^{DD'} && \text{(A5)} \\
&= \Phi^{DD'}\otimes\frac{d_B}{d_D}\langle\Phi|^{DD'}(V_{\mathcal{F}}^{A'B'\to E'D'})^* \\
&\quad (\omega^{DB'}\otimes\Phi^{A'R'})(V_{\mathcal{F}}^{A'B'\to E'D'})^{\mathsf{T}}|\Phi\rangle^{DD'} && \text{(A6)} \\
&= \Phi^{DD'}\otimes\frac{d_B p_{\mathrm{succ}}}{d_D}\zeta_{\mathrm{succ}}^{E'R'}. && \text{(A7)}
\end{aligned}
$$

Taking the square root of both sides concludes the derivation. Here, we also used Lemma 6 in the second equation.

Security analyses for practical mistrustful quantum cryptography based on quantum state discrimination games

Adrian Kent[1 2 *] Damián Pitalúa-García[1 †]

[1] *Centre for Quantum Information and Foundations, DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, U.K.*

[2] *Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON N2L 2Y5, Canada*

Abstract. In many ideal quantum cryptographic protocols, including relativistic quantum bit commitment [1] and quantum money tokens [2], Bob sends Alice random states from the BB84 [3] or another given set. In practice, the states are prepared with misalignment, not uniformly distributed, are mixed, and include some multi-photon states. To cheat, Alice must produce statistically plausible results for measurements in both BB84 bases, allowing for a given error level. We present a general security analysis based on maximum confidence quantum measurements [4] that strongly bounds Alice's probability of winning games of this type with arbitrary quantum strategies, and discuss applications to specific protocols.

Keywords: practical quantum cryptography, maximum confidence quantum measurement, quantum tokens, quantum bit commitment

1 Summary of results

Our main technical results are twofold.

First, we consider a broad class of quantum tasks in which Alice receives quantum states from a given set in N independent rounds and is required to obtain particular classical information about the prepared states for all rounds, with the possibility of failing in no more than n rounds, for a given $0 \leq n \leq N$. Effectively, Alice is playing a multi-round game which she wins if she succeeds in a sufficiently high proportion of the rounds.

We show that if Alice's success probability in the kth round is upper bounded by P^k_{bound}, conditioned on any quantum inputs ρ_j and classical outputs x_j for rounds $j \neq k$ and on any extra measurement outcome o_{extra} obtained by Alice, for all $k \in [N]$, then Alice's success probability $P_{\text{win}}(n, N|o_{\text{extra}})$ in the task conditioned on the extra outcome o_{extra} is upper bounded by the probability $P^{\text{coins}}_{\text{bound}}(n, N)$ of having no more than n errors in N independent coin tosses with success probabilities $P^1_{\text{bound}}, P^2_{\text{bound}}, \ldots, P^N_{\text{bound}}$. Thus, we have

$$\begin{aligned} P_{\text{win}}(n, N|o_{\text{extra}}) &\leq P^{\text{coins}}_{\text{bound}}(n, N) \\ &\leq \sum_{l=0}^{n} \binom{N}{l} (1 - P_{\text{bound}})^l (P_{\text{bound}})^{N-l}, \end{aligned} \tag{1}$$

where $P^k_{\text{bound}} \leq P_{\text{bound}} < 1$ for all $k \in [N]$. This further implies that we can upper bound the right hand side by a Chernoff bound decreasing exponentially with N if $n < N(1 - P_{\text{bound}})$.

This result is quite useful for a great variety of quantum cryptography protocols in which Alice's cheating probability reduces to wining the described task. In this case the security proof can be reduced to finding the upper bound P^k_{bound} for the round k conditioned on any quantum inputs ρ_j and classical outputs x_j for rounds $j \neq k$ and on any extra measurement outcome o_{extra} obtained by Alice, for all $k \in [N]$. Crucially, we note that the result applies to arbitrary quantum strategies by Alice, including arbitrary joint quantum measurements on the quantum states received in all N rounds.

Examples where this result is useful include relativistic quantum bit commitment protocols (e.g., [1]), quantum money schemes (e.g., [5]), quantum S-money token schemes [2]. It can also be used for security proofs in other mistrustful quantum cryptography protocols, for example, quantum spacetime-constrained oblivious transfer protocols [6, 7].

Second, we deduce the bound P^k_{bound} for an important and cryptographically relevant subset of the quantum tasks described above, in which Alice's task in each round can be shown to be equivalent to a quantum state discrimination task. In this case, we show that Alice's probability to win the task in round k, conditioned on any quantum input states ρ_i and classical outputs x_i for rounds $i \neq k$ and on any extra measurement outcomes o_{extra}, is upper bounded by her maximum confidence quantum measurement $\max_{j \in S_k} P_{\text{MC}}(\rho^k_j)$ [4], where

$$P_{\text{MC}}(\rho^k_j) = \max_{Q \geq 0} \frac{p^k_j Tr[Q\rho^k_j]}{Tr[Q\rho^k]}, \tag{2}$$

where in the relevant state discrimination task Alice receives the quantum state ρ^k_j with probability p^k_j, for all $j \in S_k$, and where $\rho^k = \sum_{j \in S_k} p^k_j \rho^k_j$.

Because P_{MC} can be shown to increase relatively little for small variations from the ideal protocol, this result allows us to derive significantly tighter and more general security bounds for S-money quantum tokens of Ref. [2], in which we allow the prepared states to deviate from the target BB84 state up to an angle θ on the Bloch sphere. Previous security analyses [2] assumed that the four states belonged to two qubit orthonormal bases, which cannot be precisely guaranteed in a realistic experimental setup. Similarly, we note that this result applies to the experimental implementation of relativis-

*apak@damtp.cam.ac.uk

†D.Pitalua-Garcia@damtp.cam.ac.uk

tic quantum bit commitment reported in Ref. [1], as the security analysis provided in Ref. [1] also assumed that the prepared states belonged to two qubit orthonormal bases deviating from the BB84 basis by an angle θ.

We further refine the security analysis for the S-money quantum tokens of [2] by allowing a small probability P_θ that the qubit prepared states deviate from the intended BB84 states by an angle greater than θ in the Bloch sphere. This allows security to be proven based on experimental data that sample the distribution of deviations from BB84 states.

2 Application to a refined security analysis of quantum S-money token schemes

2.1 The quantum S-money scheme

We describe the quantum S-money scheme of Ref. [2] for two presentation regions R_0 and R_1. Alice and Bob agree in advance on a spacetime reference frame F, and define R_0 and R_1 in F. Let P be the intersection of the causal pasts of R_0 and R_1. Alice and Bob have laboratories A_i and B_i able to communicate within R_i, for $i = 0, 1$. Alice (Bob) and her (his) laboratories communicate via secure and authenticated channels, which can be implemented via predistributed secret keys, for instance. An ideal quantum S-money scheme comprises the following steps.

1. Bob sends Alice N random states from the BB84 set $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$ [3]. Alice chooses a random bit z and measures all the states in the qubit orthonormal basis $\mathcal{D}_z$, where $\mathcal{D}_0 = \{|0\rangle, |1\rangle\}$ and $\mathcal{D}_1 = \{|+\rangle, |-\rangle\}$. Let $\mathbf{t}$ denote Bob's encoded bits and $\mathbf{u}$ the preparation bases. Let $\mathbf{x}$ denote Alice's outcomes.

2. Alice sends $\mathbf{x}$ to A_i; while Bob sends $\mathbf{t}$ and $\mathbf{u}$ to B_i, for $i = 0, 1$.

   The previous steps comprise the quantum phase, which can be performed arbitrarily in advance of the following stage. The following steps comprise the classical phase.

3. Within P, Alice obtains the bit b, labelling the presentation region R_b; she sends b to A_i, for $i = 0, 1$, and $c = b \oplus z$ to Bob.

4. Within P, after receiving c, Bob sends c to B_i, for $i = 0, 1$.

5. After receiving b, A_b sends $\mathbf{x}$ to B_b within R_b.

6. For $i = 0, 1$, within R_i, assuming that B_i receives a token $\mathbf{x}^i$, B_i calculates $d_i = c \oplus i$, and computes the number $N_{\mathrm{errors},i}$ of entries $k \in \Delta_i$ that do not satisfy $x^i_k = t_k$, where $\Delta_i = \{k \in [N] | u_k = d_i\}$. B_i locally validates the token if

$$\frac{N_{\mathrm{errors},i}}{N_i} \leq \gamma_{err} \,, \tag{3}$$

   or rejects it otherwise, where $N_i = |\Delta_i|$.

The practical scheme deviates from the ideal scheme above by allowing the experimental imperfections described in Table 5, and making the assumptions of Table 6, of Ref. [2]. However, we note that we relax the assumption A of table 6, as now we allow the prepared states to deviate from the intended BB84 states up to an angle θ on the Bloch sphere without constraining them to define two qubit orthonormal bases. We also allow a small probability $P_\theta > 0$ that such uncertainty angle is larger than θ.

Furthermore, the scheme can be extended to allow for a big amount of losses, by requiring that Alice reports to Bob a set $\Lambda \subseteq [N]$ of received pulses, and constraining the scheme to these pulses. In this case Bob does not abort if and only if $|\Lambda| \geq \gamma_{\mathrm{det}} N$, for a predetermined $\gamma_{\mathrm{det}} \in (0, 1]$ (the case $\gamma_{\mathrm{det}} = 1$ corresponds to Alice not reporting any losses). We can consider a situation in which the experimental setup has sufficiently small losses so that Alice does not need to report any losses to Bob. In this case, we do not need to make the assumptions C, D and F of Ref. [2], and we can guarantee perfect protection against multi-photon attacks [9], which as discussed in Ref. [9] applies to various implementations of mistrustful quantum cryptography.

The ideal and practical schemes extend straightforwardly to an arbitrary number of presentation spacetime regions. Our security analysis applies to this general case. The unforgeability proof holds even if Alice is required to report losses and applies for arbitrarily powerful dishonest Alice who may detect all quantum states received from Bob and choose to report an arbitrary subset of states as lost.

2.2 Security analysis

We consider the general case in which Alice reports losses to Bob. We define the following parameters.

P_{det} is the probability that a quantum state $|\psi_k\rangle$ transmitted by Bob is reported by Alice as being successfully measured, with label $k \in \Lambda$, for $k \in [N] = \{1, 2, \ldots, N\}$; E is the probability that Alice obtains a wrong measurement outcome when she measures a quantum state $|\psi_k\rangle$ in the basis of preparation by Bob; if the error rates E_{tu} are different for different prepared states, labelled by $t = 0, 1$, and for different measurement bases, labelled by $u = 0, 1$, we simply take $E = \max_{t,u}\{E_{tu}\}$; P_{noqub} is the probability that a prepared quantum state has dimension greater than two (by comprising two or more qubits, for instance), which arises due to an imperfect single-photon source; P_θ is is the probability that a prepared quantum state has uncertainty angle greater than θ in the Bloch sphere, due to errors in state preparation; $P_{\mathrm{noqub},\theta}$ is the probability that a prepared quantum state has dimension greater than two or its uncertainty angle in the Bloch sphere is greater than θ; We have

$$P_{\mathrm{noqub},\theta} = 1 - (1 - P_{\mathrm{noqub}})(1 - P_\theta); \tag{4}$$

N is the number of quantum states that Bob sends Alice; Ω_{qub} is the set of labels $k \in [N]$ for quantum states $|\psi_k\rangle$ of dimension two that Bob sends Alice; β_{PB} (β_{PS}) is the

maximum deviation away from probability $\frac{1}{2}$ when Bob chooses a preparation basis (state); $\beta_{\rm E}$ is the maximum deviation away from probability $\frac{1}{2}$ when Alice chooses the bit z denoting her measurement basis; $\gamma_{\rm det}$ is the minimum rate of states reported by Alice as successfully measured for Bob not abort; $\gamma_{\rm err}$ is the maximum error rate allowed by Bob when validating Alice's token; $\nu_{\rm cor}$ is a security parameter chosen by Alice to compute a guaranteed degree of correctness; $\nu_{\rm unf}$ is a security parameter chosen by Alice to compute a guaranteed degree of unforgeability.

Our refined unforgeability proof is based on computing the maximum confidence quantum measurement in the following quantum state discrimination task. Let ρ^k_{tu} denote the density matrix for the qubit state that Bob sends Alice with label $k \in \Omega_{\rm qub}$, when Bob aims to prepare the state encoding the bit t in the basis $\mathcal{D}_u$, for $t,u \in \{0,1\}$. That is, $\rho^k_{t0} \approx |t\rangle\langle t|$ for $t=0,1$, $\rho^k_{01} \approx |+\rangle\langle +|$ and $\rho^k_{11} \approx |-\rangle\langle -|$. For $k \in \Omega_{\rm qub}$, let $P^k_{\rm PS}(t)P^k_{\rm PB}(u)$ be the probability that Bob prepares the state ρ^k_{tu}, where $\{P^k_{\rm PS}(t)\}^1_{t=0}$ and $\{P^k_{\rm PB}(u)\}^1_{u=0}$ are bit probability distributions. For $k \in \Omega_{\rm qub}$, we define $\rho^k_1 = \rho^k_{00}$, $\rho^k_2 = \rho^k_{01}$, $\rho^k_3 = \rho^k_{10}$, $\rho^k_4 = \rho^k_{11}$, $q^k_1 = P^k_{\rm PS}(0)P^k_{\rm PB}(0)$, $q^k_2 = P^k_{\rm PS}(0)P^k_{\rm PB}(1)$, $q^k_3 = P^k_{\rm PS}(1)P^k_{\rm PB}(0)$, $q^k_4 = P^k_{\rm PS}(1)P^k_{\rm PB}(1)$, and

$$r^k_i = \frac{q^k_i + q^k_{i+1}}{2}, \chi^k_i = \frac{q^k_i\rho^k_i + q^k_{i+1}\rho^k_{i+1}}{q^k_i + q^k_{i+1}}, \rho^k = \sum_{i=1}^{4} r^k_i\chi^k_i, \quad (5)$$

for all $i \in [4]$, where we use the notation $4+1=1$. Let $P_{\rm MC}(\chi^k_j)$ be the maximum confidence quantum measurement that the received state was χ^k_j when Alice's outcome is $j \in [4]$ [4], where the maximum is taken over all positive operators Q acting on a two dimensional Hilbert space. That is, we have

$$P_{\rm MC}(\chi^k_j) = \max_{Q\geq 0} \frac{r^k_j Tr[Q\chi^k_j]}{Tr[Q\rho^k]}. \quad (6)$$

We define $P_{\rm bound}$ as a probability satisfying

$$\max_{j\in[4],k\in\Omega_{\rm qub}} 2P_{\rm MC}(\chi^k_j) \leq P_{\rm bound} < 1. \quad (7)$$

A bound $P_{\rm bound}$ can be computed analytically or numerically (e.g., [8]).

We say the scheme is $\epsilon_{\rm unf}-$unforgeable if the probabilty that Alice can make Bob validate a token at more than one presentation region is not greater than $\epsilon_{\rm unf}$.

Theorem 1 *Suppose that the following constraints hold:*

$$N\gamma_{det} \leq n \leq N,$$
$$0 < P_{noqub,\theta} < \nu_{unf} < \gamma_{det}\Big(1 - \frac{\gamma_{err}}{1-P_{bound}}\Big), \quad (8)$$

for predetermined $\gamma_{det} \in (0,1]$ and $\gamma_{err} \in [0,1)$ and for some $\nu_{unf} \in (0,1)$, where $n = |\Lambda|$, and where P_{bound} satisfies (7). The quantum token scheme is $\epsilon_{unf}-$unforgeable with

$$\epsilon_{unf} = \sum_{l=0}^{\lfloor N(1-\nu_{unf})\rfloor} \binom{N}{l} (1-P_{noqub,\theta})^l (P_{noqub,\theta})^{N-l} + \sum_{l=0}^{\lfloor n\gamma_{err}\rfloor} \binom{n-\lfloor N\nu_{unf}\rfloor}{l} (1-P_{bound})^l (P_{bound})^{n-\lfloor N\nu_{unf}\rfloor - l}, \quad (9)$$

which decreases exponentially with N from the conditions (8). In the case that losses are not reported we take $\gamma_{det} = 1$ and $n = N$.

We note that theorem 1 is improved with respect to theorem 1 of Ref. [2] in two main ways: 1) it allows Bob's prepared states to deviate arbitrarily from the intended BB84 states up to an angle θ in the Bloch sphere without restricting the prepared states to form qubit orthonormal bases; and 2) it replaces $P_{\rm noqub}$ by $P_{{\rm noqub},\theta}$. That is, in the security analysis of Ref. [2], θ was considered an upper bound on the uncertainty angle in the Bloch sphere for state preparation. But, here we relax this assumption by allowing the uncertainty angle to be greater than θ with a probability P_θ. The probability $P_{{\rm noqub},\theta}$ considers this via equation (4). Crucially, we note that by using the maximum confidence quantum measurement, our scheme is proved secure against arbitrary attacks by Alice, comprising an arbitrary quantum measurement on the whole quantum system of N pulses received from Bob and an ancilla of arbitrary finite Hilbert space dimension held by Alice, which also include loss dependent attacks in which Alice selects what set of pulses to report as lost to her convenience.

3 Discussion

Our refined security proof can be helpful for implementations of other tasks in mistrustful quantum cryptography, for instance, relativistic quantum bit commitment. For example, one of the first experimental demonstrations of realtivistic quantum bit commitment [1] based its security analysis on the assumption that states belong to two qubit orthonormal basis, deviating from the BB84 bases by an angle θ in the Bloch sphere. Our refined security analysis allows this assumption to be discarded and for security to be proved based directly on experimental estimates for deviations from BB84 states.

More broadly, we believe our security analysis can be helpful to analyse the security of practical implementations of mistrustful quantum cryptography. Together with the analysis of multiphoton attacks in Ref. [9], these results provide a more rigorous security analysis of implementations of mistrustful quantum cryptography with realistic experimental setups. This is crucial for developing the secure mistrustful quantum cryptographic applications envisaged for free space and fibre optic quantum networks and the eventual quantum internet [10, 11].

References

[1] T. Lunghi, J. Kaniewski, F. Bussières, R. Houlmann, M. Tomamichel, A. Kent, N. Gisin, S. Wehner and H. Zbinden. Experimental bit commitment based on quantum communication and special relativity. *Phys. Rev. Lett.*, 111:180504, 2013.

[2] A. Kent, D. Lowndes, D. Pitalúa-García and J. Rarity. Practical quantum tokens without quantum memories and experimental tests. *npj Quantum Inf.*, 8:28, 2022.

[3] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proc. of IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India*, pages 175–179, 1984.

[4] S. Croke, E. Andersson, S. M. Barnett, C. R. Gilson and J. Jeffers. Maximum confidence quantum measurements. *Phys. Rev. Lett.*, 96(7):070401, 2006.

[5] M. Bozzio, A. Orieux, L. T. Vidarte, I. Zaquine, I. Kerenidis and E. Diamanti. Experimental investigation of practical unforgeable quantum money. *npj Quantum Inf.*, 4:5, 2018.

[6] D. Pitalúa-García and I. Kerenidis. Practical and unconditionally secure spacetime-constrained oblivious transfer. *Phys. Rev. A*, 98(3):032327, 2018.

[7] D. Pitalúa-García One-out-of-m spacetime-constrained oblivious transfer. *Phys. Rev. A*, 100(1):012302, 2019.

[8] H. Lee, K. Flatt, C. Roch i Carceller, J. B. Brask and J. Bae Maximum-confidence measurement for qubit states. *Phys. Rev. A*, 106(3):032422, 2022.

[9] M. Bozzio, A. Cavaillès, E. Diamanti, A. Kent and D. Pitalúa-García. Multiphoton and side-channel attacks in mistrustful quantum cryptography. *PRX Quantum*, 2(3):030338, 2021.

[10] H. J. Kimble. The quantum internet. *Nature*, 453:1023–1030, 2008.

[11] S. Wehner, D. Elkouss and R. Hanson. Quantum internet: A vision for the road ahead. *Science*, 362:6412, 2018.

Efficient learning of mixed-state tomography for photonic quantum walk

Xiao-Ye Xu[1 2 3 *]

[1]*CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China*
[2]*CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China*
[3]*Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*

Abstract. Noise-enhanced applications in open quantum walk (QW) has recently seen a surge due to their ability to improve performance. However, verifying the success of open QW is challenging, as mixed-state tomography is a resource-intensive process, and implementing all required measurements is almost impossible due to various physical constraints. To address this challenge, we present a neural-network–based method for reconstructing mixed states with a high fidelity (∼97.5%) while costing only 50% of the number of measurements typically required for open discrete-time QW in one dimension. Our method uses a neural density operator that models the system and environment, followed by a generalized natural gradient descent procedure that significantly speeds up the training process. Moreover, we introduce a compact interferometric measurement device, improving the scalability of our photonic QW setup that enables experimental learning of mixed states. Our results demonstrate that highly expressive neural networks can serve as powerful alternatives to traditional state tomography.

Keywords: Photonic Quantum Walks,Open System,Quantum State Tomography,Machine Learning, Neural Density Operator,Neural Network

QW can not only simulate complex many-body physical phenomena such as quantum thermalization and localization but also provide a basic framework for developing efficient quantum algorithms such as quantum search and PageRank algorithms. Open QW with specific noise has been shown to significantly improve quantum transport efficiency and then enhance problem-solving efficiency (for example, a maze escape problem) by several orders of magnitude compared to noise-free scenarios. Moreover, by adding controlled noise to quantum evolution, QW can be dynamically initialized in any high-dimensional form and generate the Haar random unitary operators required for quantum computation. To leverage the computational and simulated power of such open QW, it inevitably requires mixed-state characterization. However, the characterization of open quantum systems is a resource-intensive process, and implementing all required measurements is almost impossible due to various physical constraints, so it is still a significant challenge to address mixed-state tomography.

Recently, neural network approaches for efficient learning of open quantum systems have been theoretically proposed. Such an efficient method enables high-fidelity reconstructions of mixed quantum states using only partial measurements that are experimentally accessible. However, the high expressive ability of neural networks for mixed states demands a complex deep network structure, and directly applying this method faces complicated gradient-based training as the system size increases.

In this work, by establishing a mapping between the open QW and the restricted Boltzmann machine, we realized the effective learning of mixed quantum states for the open QW, in terms of reconstruction fidelity, the number of measurements, and the number of training iterations. To increase the network training data, we innovatively introduced an unequal-arm interferometer in the time domain based on the previously constructed large-scale photonic QW, thus significantly increasing the number of measurement bases. The neural network method was tested, and it showed that using only half the number of complete measurements (i.e., partial measurements), the trained neural network can well reconstruct arbitrary mixed states with an average fidelity of up to 97.5%. Moreover, we also introduced a generalized version of the natural gradient descent procedure to accelerate training efficiency, which enables about one order of magnitude fewer training iterations than the one using traditional gradient descent procedures. The efficient mixed-state learning method sheds new light on previous tomographic methods with a pure-state hypothesis and would inspire further research and discoveries in noise-assisted quantum computing and simulation.

References

[1] X.-Y. Xu, *et al.* Measuring the Winding Number in a Large-Scale Chiral Quantum Walk. *Physical Review Letters* **120**, 260501(2018).

[2] Q.-Q. Wang, *et al.* Efficient learning of mixed-state tomography for photonic quantum walk. *Science Advances* **10**, eadl4871(2024).

[3] G. Torlai, *et al.* Latent space purification via neural density operators. *Physical Review Letters* **120**, 240503(2018).

[4] S. Dong, *et al.* Generalization to the natural gradient descent. *arXiv*:2210.02764 [math.OC].

*xuxiaoye@ustc.edu.cn

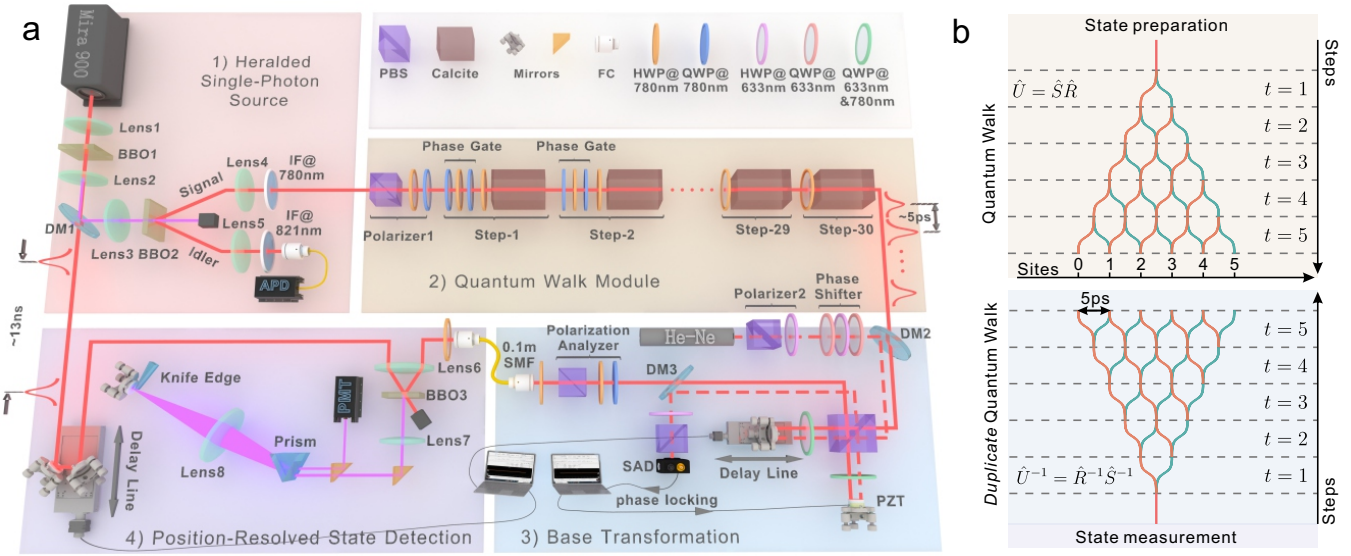


Figure 1: Photonic open QW. (a) The experimental setup mainly has four central parts: 1) Spontaneous parameter down-conversion generates the time-correlated photon pairs, where signal photons as the walker and the idler photons serve to herald; 2) The open QW, as reported in the upper panel of (b); 3) A Michelson interferometer together with a polarization analyzer implements the measurement base on position and coin; 4) A single-photon frequency up-conversion implements the position-resolved detection of the walker in each base. (b) The schematic diagrams show the QW dynamics of a localized initial state (top) and the duplicate QW with time inversion constructed for the state measurements in different bases (bottom), respectively. A list of abbreviations: $\beta - BaB_2O_4$ (BBO); dichroic mirror (DM); interference filter (IF); polarization-dependent beam splitter (PBS); half-wave plate (HWP); quarter-wave plate (QWP); piezoelectric ceramic (PZT); fiber collimator (FC); single-mode fiber (SMF); Si amplified detector (SAD); photomultiplier tube (PMT); avalanche photodiode detector (APD).

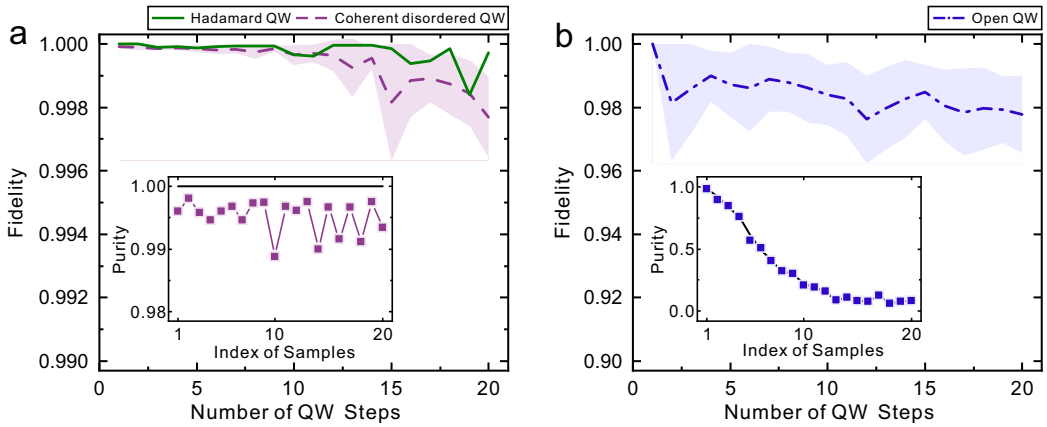


Figure 2: Benchmarking neural density operator tomography using partial measurements. Neural density operator reconstruction fidelity as a function of the number of time steps for (a) Hadamard QW(green solid line), coherent disordered QW (red dashed line), and (b) open QW with arbitrary mixing (blue dash-dotted line). The shaded regions for coherent disordered and open QW are the standard errors of neural density operator reconstruction with 20 random samples for each step, and the lines are the averaging results. The insets in (a) and (b) show the purity of reconstructed states of the 20 samples for a 20-step coherent disordered (red squares) and open QW (blue squares), respectively. The black solid lines in the insets are the theoretical values of the purity for target density matrix.

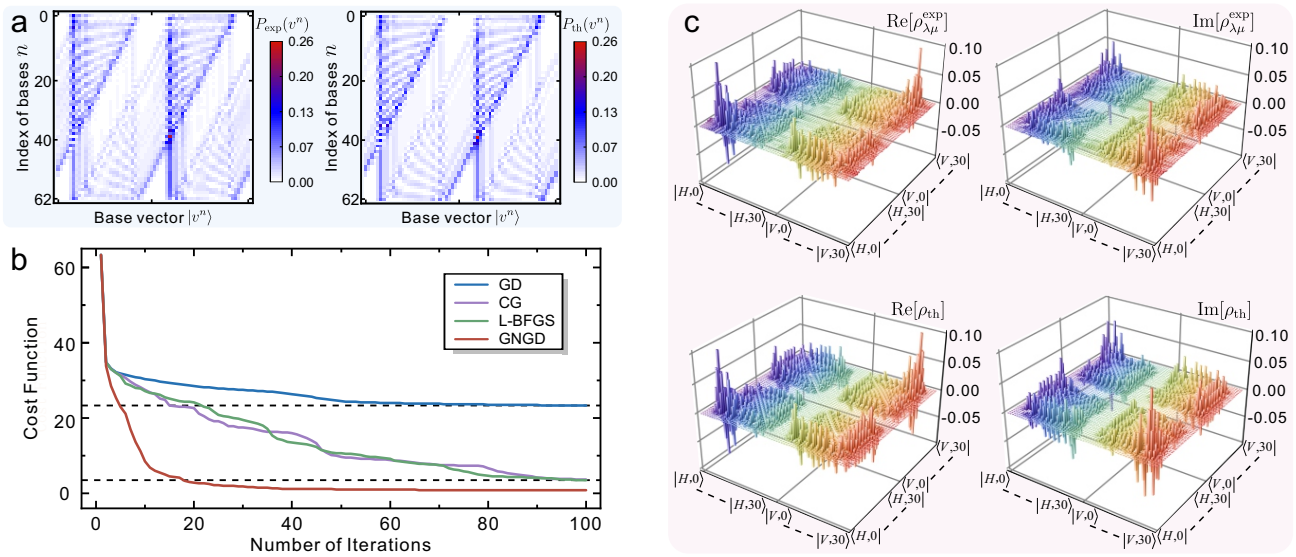


Figure 3: Experimental generalized-natural-gradient-descent-enhanced neural density operator tomography of QW in a real environment. (a) Measured probability distributions on 63 sets of bases for a 30-step Hadamard walk. The right panel is the theoretical expectation. (b) Cost function versus the number of training iterations for neural density operator tomography using the gradient descent, the conjugate gradient, the limited-memory Broyden-Fletcher-Goldfarb-Shanno, and the generalized natural gradient descent optimizer. (c) Real and imaginary parts of the neural density operator reconstructed state, and the theoretical expectations are shown in the lower panels.

Scalability enhancement of quantum computing under limited connectivity through distributed quantum computing

Shao-Hua Hu[1] George Biswas[1 2] Jun-Yi Wu[1 2 3 *]

[1]*Department of Physics, Tamkang University, New Taipei 25137, Taiwan, ROC*
[2]*Center for Advanced Quantum Computing, Tamkang University, New Taipei 25137, Taiwan, ROC*
[3]*Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan, ROC*

Abstract. We employ quantum-volume random-circuit sampling to benchmark two-QPU entanglement-assisted distributed quantum computing (DQC) and compare it with single-QPU quantum computing. Based on our error model, we derive an analytical approximation of the average gate fidelity and show the one-to-one correspondence of three figures of merits, namely average gate fidelity, heavy output probability, and linear cross-entropy, which is shown to align with numerical simulations. The approximation is calculated based on an allocation matrix obtained from the extended connectivity graph of a DQC device. Furthermore, we provide a simple formula to estimate the average gate fidelity, which also provides us with a heuristic method to evaluate the scalability enhancement in DQC, and unveils the scalability enhancement in DQC for the QPUs with limited connectivity. The full version is available on [1].

Keywords: Distributed quantum computing, Quantum volume, Randomized benchmarking, Entanglement-assisted LOCC, Average gate fidelity

In noisy intermediate-scale quantum (NISQ) computing, the imperfection on quantum processing units (QPUs) can not yet be corrected. The defects on a single QPU are much more difficult to suppress as the number of qubits increases. A solution for scaling up quantum computers is distributed quantum computing (DQC) [2, 3], in which one implements global quantum circuits over multiple high-quality small-size QPUs, assisted by classical and quantum communication across different QPUs, as shown in Figure 1.

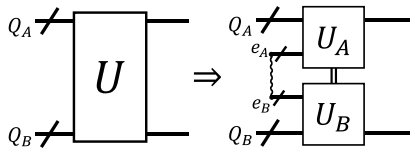


Figure 1: Implementation of a global unitary U over two QPUs Q_A and Q_B through local unitaries U_A and U_B, with the help of pre-shared entangled pairs e_{AB} (wiggly line) and classical communications (double line).

To verify the enhancement of computation power in DQC, one needs to benchmark and compare the quantum computing power of multi-QPU DQC devices and single-QPU devices. The corresponding figures of merits include average gate fidelity (AGF)[4], heavy output probability (HOP)[5], and linear cross-entropy (LXE)[6]. The quality of the quantum computing on a quantum processor can be then quantified by averaging these figures of merits over a random sampling of quantum circuits. We adopt different figures of merits (AGF, HOP, and LXE) to the QV random-circuit sampling, and extend them to two QPU DQC.

Error model. Establishing a noise model is essential for the characterization of errors in quantum devices. In the QV random-circuit sampling, the two-qubit gates are sampled from Haar random unitaries, hence we adopt the depolarizing channel as our fundamental noise model. The single-qubit depolarizing noise that affects each qubit individually is applicable with a preserving factor ranging from 1 to 0 for the identity channel and the complete depolarizing channel, respectively. Furthermore, such a random sampling allows us to assume that the unitary channels and the channel of the error are commutable. We can therefore rearrange the order of operations by moving all depolarizing channels to the end of the circuit, as shown in Figure 2. With this rearrangement, the outcome distribution $p_U(x)$ can be obtained from the ideal outcome distribution $q_U(x')$ transformed by a Markov matrix $D_{\vec{p}}(x|x')$,

*junyiwuphysics@gmail.com

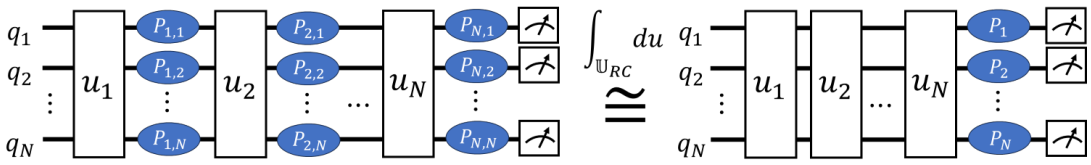


Figure 2: The single-qubit depolarizing channels (blue ellipse) come after every layer of the random gates u_i and act on each qubit individually.

i.e.

$$p_U(x) = \sum_{x' \in \{0,1\}^N} D_{\vec{P}}(x|x') q_U(x'), \tag{1}$$

where $\vec{P} = (P_1, P_2, \dots)$ are the preserving factor of the effective depolarizing channel on each qubit that is moved to the end of a circuit. By averaging over all random circuits, our three figures of merit are given by:

$$\bar{F}(\vec{P}) = \prod_{q \in \mathbb{Q}} \frac{1 + P_q}{2}, \tag{2}$$

$$\bar{H}(\vec{P}) = \bar{H}_{ideal} \frac{2^N \bar{F} - 1}{2^N - 1} + (1 - \bar{F}) \frac{2^{N-1}}{2^N - 1}, \tag{3}$$

$$\bar{\chi}(\vec{P}) = \frac{2^N \bar{F} - 1}{2^N - 1} \bar{\chi}_{ideal}. \tag{4}$$

These three equations show the one-to-one correspondence among average AGF $\bar{F}$, HOP $\bar{H}$, and LXE $\bar{\chi}$ in the QV random-circuit benchmarking under an approximated error model

$$\bar{F} \underset{\text{QV random}}{\longleftrightarrow} \bar{H} \underset{\text{QV random}}{\longleftrightarrow} \bar{\chi}. \tag{5}$$

Furthermore, in the large size limit($2^N \gg 1$), we had equality between the HOP and LXE given by

$$\bar{H}(\vec{P}) = \frac{1}{2} + \frac{1}{2} \ln 2 \frac{\bar{\chi}(\vec{P})}{\bar{\chi}_{ideal}}. \tag{6}$$

This result allows us to estimate average gate fidelity from heavy output probability or linear cross-entropy in numerical simulation, and vice versa.

Extended connectivity and allocation matrix. To assess the average AGF of multi-QPU DQC with limited connectivity one needs to incorporate the noises introduced by swapping gates and the telegating processes into the effective preserving factors $\bar{P}_q$. We achieve this by introducing an allocation matrix $\{A_{q,q'}\}_{q,q' \in \mathbb{Q}}$ modeled with the single-qubit depolarizing noises. Each element $A_{q,q'}$ of the allocation matrix describes the average number of single-qubit depolarizing channels propagating from the qubit q' to the qubit q when one implements a random two-qubit SU(4) gate on q. Hence we derive an approximated formula of $\bar{F}$ for the QV random-circuit benchmarking of multi-QPU DQC devices under arbitrary connectivity described by extended connectivity graphs, which is given as follows,

$$\bar{F} = \prod_{q \in \mathbb{Q}_w} \frac{1 + (\prod_{q' \in \mathbb{Q}} P_{q'}^{A_{q,q'}})^{2\lfloor \frac{N}{2} \rfloor}}{2}, \tag{7}$$

where $\mathbb{Q}_w$ is the set of the working qubits of the QPUs. To provide an efficient analytical tool, we introduce further approximation of average AGF for large-size and high-fidelity QPUs,

$$\bar{F}_{\mathbb{Q}}(\epsilon) \approx \exp(-\frac{N \mathcal{A}_{\mathbb{Q}}}{2} \epsilon), \tag{8}$$

where the error rate is assumed to be uniform for every qubit given by a constant ϵ, and the *characteristic cost* $\mathcal{A}_{\mathbb{Q}}$ of a connectivity configuration $\mathbb{Q}$ is the sum of all elements of the allocation matrix $A_{q,q'}$

$$\mathcal{A}_{\mathbb{Q}} \equiv \sum_{q \in \mathbb{Q}_w, q' \in \mathbb{Q}} A_{q,q'}. \tag{9}$$

The theory of error model approximation is supported by numerical simulations on qiskit. In our simulation, we select six types of connectivity, which can be categorized into two groups. The first group comprises devices utilizing a single QPU, while the second group consists of DQC-composited devices equipped with two local QPUs. as shown in Figure 3.

Numerical simulation. We employ linear cross-entropy for measuring average gate fidelity and set 5 different constant error rates from $\epsilon = 0.05\%$ to $\epsilon = 0.5\%(\epsilon \equiv 1 - P)$ across all qubits in qiskit. We sampled over 1000 random circuits for each connectivity graph and error rate and executed $10,000$ shots for each configuration. Then we plot the relation between the error rate in the qiskit and the effective error rate calculated by the simulation date through Eq. 2. The case of size 8 is shown in Figure 4.

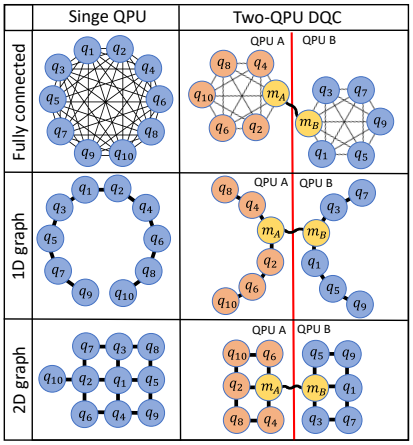


Figure 3: The blue and orange vertices represent the working qubits and the black edges identify the direct coupling between qubits. The yellow vertices denote the auxiliary memory qubits, which can either store the shared entangled state or facilitate the passage of additional swapping gates.

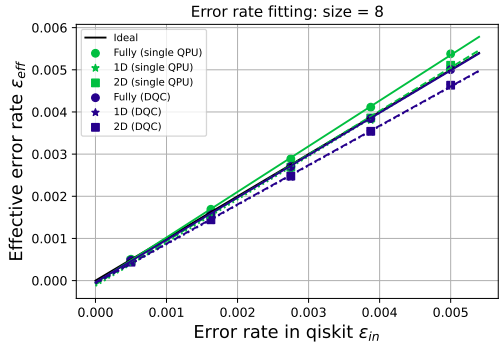
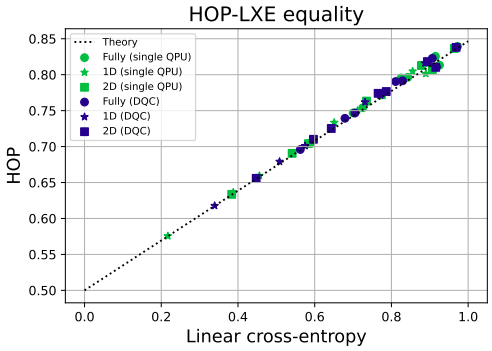
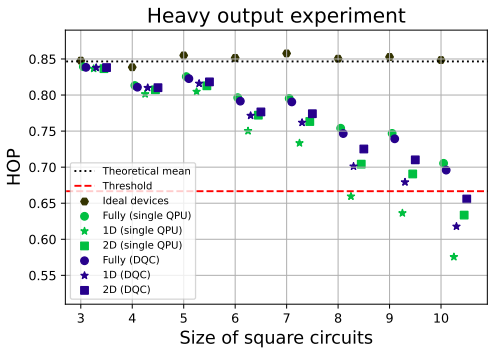


Figure 4: The scatter indicates the data point, here we use different shapes to denote the connectivity of the devices. The linear fitting gives the blue and green lines, where the ratio is close to 1 (1 ± 0.05) and the intercept is around $10^{-5} \sim 10^{-6}$.

To further support the AGF-HOP-LXE correspondence and scalability enhancement of DQC, we conduct additional simulations, in which we fix the error rate $\epsilon_{\text{in}} = 0.15\%$ for all qubits, and benchmark both single-QPU and two-QPU devices from 3-qubit to 10-qubit. Under this setting, we sample over 2000 random circuits for all sizes and take 10000 shots for each random circuit. For each configuration of connectivity graph G and qubit number n, we evaluate the measurement outputs with both average HOP and LXE and plot them in Fig. 5(a). One can clearly verify the HOP-LXE correspondence derived in Eq. (6) with these simulated data as a general property independent of the size and connectivity of a quantum computing device.



(a) HOP-LXE correspondence



(b) HOP for different sizes and connectivity

Figure 5: QV random-circuit benchmarking for $3-10$ qubit devices with different types of connectivity.

The theoretical calculation according to our theory agrees with the simulation results, which show the scalability enhancement in DQC for QPUs with limited connectivity, such as 1D and 2D graphs shown in Figure 5. The results suggest that the condition for scaling up quantum computing using DQC is to choose a good qubit such that the connectivity of the extended graph is improved. For given local QPUs, we can find the best position for the auxiliary memory qubit, with the help of the allocation matrix. The allocation matrix also provides a heuristic method of finding the best position through the minimization of the characteristic value $\mathcal{A}_{\mathbf{Q}}$.

The scalability enhancement is unveiled under the assumption of the same error rate for all devices. However, in practice, the error rate of a small-size QPU is usually smaller than the one of a large-size QPU. In this regard, the DQC may still enhance scalability, even without improving the connectivity.

References

[1]S.-H. Hu et al., "Scalability enhancement of quantum computing under limited connectivity through

distributed quantum computing", 10 . 48550 / ARXIV.2405.10942 (2024).

[2]M. Caleffi et al., *Distributed quantum computing: a survey*, 2022.

[3]D. Barral et al., *Review of distributed quantum computing. from single qpu to high performance quantum computing*, 2024.

[4]K. Życzkowski et al., "Average fidelity between random quantum states", Physical Review A **71**, 032313 (2005).

[5]A. W. Cross et al., "Validating quantum computers using randomized model circuits", Physical Review A **100**, 032328 (2019).

[6]F. Arute et al., "Quantum supremacy using a programmable superconducting processor", Nature **574**, 505–510 (2019).

Extended abstract : Robust fault-tolerant compilation of quantum error correction circuits based on SWAP gates

Shao-Hen Chiew[1] * Ezequiel Ignacio Rodríguez Chiacchio[1] † Vishal Sharma[1] ‡
Jing Hao Chai[1] § Hui Khoon Ng[1 2 3] ¶

[1] *Entropica Labs, 186b Telok Ayer Street, Singapore 068632*
[2] *Yale-NUS College, Singapore*
[2] *Centre for Quantum Technologies, National University of Singapore, Singapore*

Abstract. We propose a compilation protocol that enables fault-tolerant quantum error correction circuits to be implemented on devices with incompatible topologies, in such a way that preserves their fault-tolerance. By restricting the appearance of error mechanisms that break fault-tolerance, our protocol preserves the scaling of logical error rates with code distance and the presence of thresholds, while resulting in deteriorations that can be estimated using a simple effective noise model. As an example, we apply our protocol to the planar surface code, showing that it can be embedded onto the heavy-hexagonal and hexagonal lattices with only a constant overhead in additional timesteps, and deteriorations in threshold values by less than an order of magnitude. Simulations of logical error rates under a full-circuit noise model further verify our predictions. Our results are fully generalizable to any quantum circuit and device topologies, and relaxes the dependence of statements regarding fault-tolerance on device details.

Keywords: Fault-tolerant quantum error-correction, compilation, qubit routing

1 Introduction

Fault-tolerant quantum error correction (FTQEC) is an essential component for quantum computation at scale. Implementations of FTQEC protocols are designed with strict error propagation properties in mind that ensure their fault-tolerance, and are thus extremely sensitive to architectural aspects of the quantum device such as their topology.

In the broader context of implementing general quantum circuits under such device-topological constraints, a well-known approach to solve this compilation problem is via *qubit routing*, where a series of SWAP gates are employed to transform the non-local quantum circuit (which we call the *abstract circuit*) to one that obeys the locality constraints of the device (which we call the *routed circuit*) [1]. Crucially, qubit routing generally alters the manner in which errors are introduced and propagated. This is especially undesirable in the context of FTQEC, where abstract circuits (e.g. corresponding to syndrome extraction circuits or logical operations) are usually designed with error propagation behaviours in mind that ensure their fault-tolerance, such as transversality. It therefore appears that there is little flexibility when implementing FTQEC circuits, prohibiting execution on devices that do not meet the circuit's connectivity requirements.

With these considerations in mind, our results demonstrate that it is in fact possible to perform routing in a way that preserves the fault-tolerance of the underlying abstract circuit, thereby relaxing constraints on FTQEC set by device topologies. Using the surface code – a prime candidate for QEC at scale – routed on the heavy-hexagonal lattice as an example, we further show that our routing process results only in tolerable deteriorations in logical error rates and threshold values ($<$ an order of magnitude), rendering it a highly viable strategy for the implementation of FTQEC protocols under device-topological constraints.

*shaohen@entropicalabs.com
†ezequiel@entropicalabs.com
‡vishal@entropicalabs.com
§jhchai@entropicalabs.com
¶huikhoon.ng@yale-nus.edu.sg

2 Main results

We define a *routing schedule* to be the representation of a sequence of single-timestep/layer circuits consisting of either gates implementing the abstract quantum circuit (referred to as *interaction layers*), or SWAP gates which carry out qubit routing (referred to as *SWAP layers*).

Our goal is to perform the compilation by inserting SWAP gates in a manner that does not introduce and propagate errors in an uncontrollable way, relative to the underlying abstract circuit. For this purpose, we introduce the notion of *error-pattern-preserving (EPP) routing schedules*, which are routing schedules with SWAP gates that obey one of the following constraints:

1. its targets consists of a routing qubit and an abstract qubit (which we call *type-1 SWAP gates*),
2. its targets consists exclusively of abstract qubits (which we call *type-2 SWAP gates*) which furthermore must be connected in the interaction graph, and in a manner consistent with the ordering of the interactions.

The main result of our work concerns the fault-tolerance of abstract circuits routed with EPP schedules:

Theorem 1 *Consider an abstract circuit routed by a EPP schedule. If the abstract circuit is fault-tolerant, the routed circuit is also fault-tolerant.*

Here, the fault-tolerance of a circuit is defined rigorously by conditions that guarantee that the circuit does

not uncontrollably spread and convert correctable errors into uncorrectable ones, even when its components are faulty [2, 3].

The intuition behind the proof of Theorem 1 is that routing schedules preserve the structure and error mechanisms of the underlying faulty abstract circuit, while potentially introducing new error mechanisms that can break its fault-tolerance, depending on the form of the routing schedule. The constraints of EPP routing schedules prevent the appearance of such error mechanisms, and therefore allows routed circuits to inherit the fault-tolerance of the abstract circuit. Consequently, abstract circuits that have been designed to be fault-tolerant can be implemented on a device of a different topology in a similarly fault-tolerant manner, as long as a solution to the routing problem under the constraints of EPP schedules can be found.

The above result can also be seen as a consequence of the more general fact that fault-tolerance remains robust under weak, persistent coupling to an environment in a non-Markovian manner [2, 4]. The routing qubits act as a bath that is coupled to the abstract qubits via type-1 SWAP gates, with the possibility that multiple abstract qubits can share the same bath, and at different times over the course of the computation. Nonetheless, the structure of routed circuits ensures that spatio-temporally correlated errors involving s qubits occur with a probability that is exponentially suppressed as s, which can be interpreted as the noise effectively satisfying a locality condition [2] that preserves fault-tolerance.

When applied to the compilation of abstract circuits implementing quantum memories, the absence of new error mechanisms directly implies the preservation of the scaling of their logical error rates (LERs), including the exponential suppression of errors with increasing code distance and the presence of thresholds. Furthermore, it implies that the behaviour of the noisy routed circuit can be understood in terms of the abstract circuit subjected to an averaged, effective noise model. This effective noise model takes the same form as the noise model of the underlying abstract circuit, but is instead parametrized by an effective physical error rate:

$$p_{\text{eff}} = p + \frac{n_{\text{swap}}}{|E|} p_{\text{swap}}, \tag{1}$$

where p is the physical error rate of original noise model, E is the set of all error locations of the original quantum circuit, n_{swap} counts the total number of times qubits in the abstract circuit were involved in SWAP gates, and p_{swap} the physical error rate of faulty SWAP gates. The LERs and threshold value of the routed circuit can thus be approximately related to those of the abstract circuit via:

$$p'_L(p) \approx p_L(p_{\text{eff}}), \tag{2}$$

$$p'_{\text{threshold}} \approx p_{\text{threshold}} - \frac{n_{\text{swap}}}{|E|} p_{\text{swap}}. \tag{3}$$

3 Application to the surface code

Using our compilation protocol, we demonstrate how the surface code can be robustly implemented in a fault-tolerant manner on devices with incompatible topologies, such as the heavy-hexagonal lattice.

3.1 EPP routing schedules for the surface code

Firstly, to obtain valid EPP routing schedules, we extend existing routing algorithms based on a distance-minimizing greedy graph search [1, 5] to impose the constraints of EPP schedules. Furthermore, exploiting the fact that both the square lattice of the planar surface code and heavy-hexagonal lattices possess the same discrete translational symmetry, the search algorithm can be modified to be distance/scale-independent, by solely working with a primitive unit cell of the surface code. Applied to the surface code embedded onto the heavy-hexagonal lattice, we obtain the depth-minimal EPP routing schedule visualized in Fig. 1 (a)-(f).

This schedule consists of two SWAP layers between each pair of interaction layers, resulting in six SWAP layers and four interaction layers in total. Decomposing SWAPs in terms of CNOTs and parallelizing entangling gates whenever possible, we find that it contains 19 layers of two-qubit gates (instead of 4 layers, when executed natively on a square lattice). Each data/ancilla qubit experiences 3 SWAP gates in each cycle, albeit in an inhomogenous manner in time.

Notably, after each cycle, the entire routed surface code is translated diagonally across one cell of the heavy-hexagonal lattice. Reversing the syndrome extraction circuit at alternating cycles returns it to its original location in a periodic manner, and retains the prevention of aligned hook errors built into the circuit [6].

3.2 Approximate LERs and threshold values

For simplicity, we first assume that faulty CNOTs are the dominant contributors of errors. In this case, developing Eq. (1) to linear order in p for the surface code yields $p_{\text{eff}} = cp$, where a dimensionless constant:

$$c \equiv 1 + \frac{3}{4} \bar{n}_{\text{swap}} \tag{4}$$

appears, with $\bar{n}_{\text{swap}}$ the average number of times each qubit in the surface code has been involved in a SWAP gate throughout the routing schedule. Eqs. (2) and (3) then take the simple forms:

$$p'_L(p) \approx p_L(cp), \tag{5}$$

$$p'_{\text{threshold}} \approx p_{\text{threshold}}/c. \tag{6}$$

The calculation above allows us to estimate the deterioration in LERs and thresholds of a routed circuit solely based on the structure of the routing schedule, modulo the effects of single-qubit errors, and inhomogeneities of the SWAP schedule. For the schedule of Fig. **??**, where $\bar{n}_{\text{swap}} = 3$, we therefore expect the threshold of the routed surface code to approximately deteriorate by a multiplicative factor of $c = 3.25$ in the limit of large code distances.

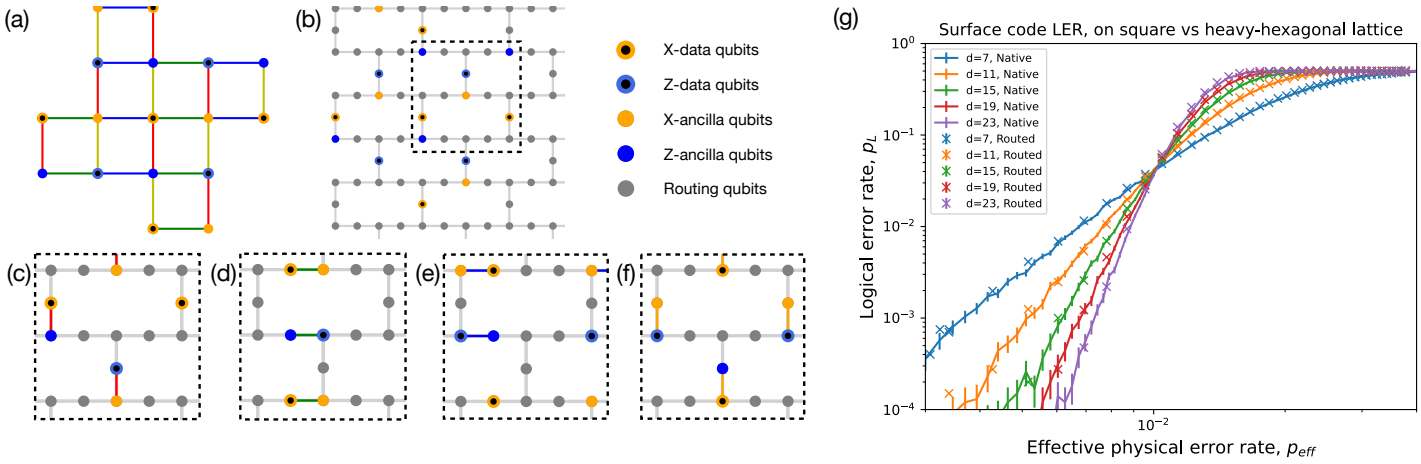


Figure 1: Schematic representation and LERs of the EPP routing schedule to embed the planar surface code onto the heavy-hexagonal lattice. (a) Visualization of the syndrome extraction circuit for the distance 3 planar surface code, showing CNOT gates (colored edges) between the data and ancilla qubits (17 colored circles, separated into four 'species' of qubits), arranged on a square lattice. (b) Initial layout of the 17 qubits in the heavy-hexagonal lattice. The dotted box denotes a tileable unit cell of the circuit. (c)-(f) Intermediate layouts during the four interaction layers of the surface code. Between each pair of interaction layers are two SWAP layers, consisting of only type-1 SWAP gates that permute data/ancilla qubits to reach the next layout. (g) p_L as a function of p_{eff} for the surface code executed natively on a square grid without routing (solid lines), and on the heavy-hexagonal lattice with the EPP schedule (crosses) for increasing code distances. $p_{\text{eff}} = p$ for the natively executed surface code, while $p_{\text{eff}} = 3.7p$ for the routed surface code. The threshold of the routed circuit has also deteriorated by a multiplicative factor of 3.7.

Notably, the value of $\bar{n}_{\text{swap}}$ (and thus c) depends explicitly on the structure of the routing schedule, which is in turn restricted by purely graph-theoretic constraints arising from the mismatch between the interaction graph of the QEC circuit and the topology of the quantum device. For instance, a naive lower bound of $\bar{n}_{\text{swap}} \geq 2$ can be obtained by considering the difference in minimum degrees between the heavy-hexagonal and the square lattices, putting the schedule of Fig. 1 (a)-(f) close to optimal depth.

3.3 Simulations of LERs and threshold values

We numerically simulate the LERs of the native and routed circuit codes for increasing distances under a full-circuit depolarizing noise model (errors during qubit initialization/reset, measurements, single-qubit gates, and two-qubit gates) parametrized by a common physical error rate p. MWPM is used for decoding, alongside a matching graph with edge weights modified according to the structure of the routing schedule.

To plot the two sets of p_L curves as a function of p_{eff}, we perform a numerical fitting procedure to empirically obtain the value of c for the routed circuit that minimizes their mean squared error, yielding $c = 3.7$, which is slightly higher than the predicted value of 3.25 due to the aforementioned approximations present in Eq. (4).

Fig. 1 (g) shows the LERs p_L of the native and routed surface codes, plotted against the effective physical error rates p_{eff}, where $p_{\text{eff}} = p$ for the native surface code by definition, and $p_{\text{eff}} = 3.7p$ for the routed surface code. The close agreement between the two sets of curves, including the position of the threshold, confirms our descriptions of the LERs of routed schedules.

4 Conclusion

In summary, we have proposed a general way to fault-tolerantly implement QEC circuits on devices with incompatible topologies.

Our approach is directly compatible with any device topology and quantum circuit, and opens up the possibility of implementing and testing QEC circuits on different devices, which is particularly relevant on the majority of current scalable quantum hardware architectures featuring geometrically local connectivities.

Furthermore, our approach yields relatively simple descriptions of LERs and threshold values, and does not significantly modify the classical decoding process, which facilitates implementation and analysis. Applied to the surface code, this simply translates to performing MWPM on the same matching graph, with locally modified edge weights.

From a broader theoretical perspective, Theorem 1 relaxes the dependence of statements regarding fault-tolerance on device details. It implies that FTQEC protocols can remain fault-tolerant when executed on a device of incompatible topology as long as a solution to the routing problem under the constraints of EPP schedules exists, which is a purely graph-theoretical problem concerning the connectivity of the device and the QEC code.

References

[1] Andrew M Childs, Eddie Schoute, and Cem M Unsal. "Circuit transformations for quantum architectures". In: *arXiv preprint arXiv:1902.09102* (2019).

[2] Panos Aliferis, Daniel Gottesman, and John Preskill. "Quantum accuracy threshold for concatenated distance-3 codes". In: *arXiv preprint quant-ph/0504218* (2005).

[3] Daniel Gottesman. "Fault-tolerant quantum computation with constant overhead". In: *arXiv preprint arXiv:1310.2984* (2013).

[4] Barbara M Terhal and Guido Burkard. "Fault-tolerant quantum computation for local non-Markovian noise". In: *Physical Review A* 71.1 (2005), p. 012336.

[5] Alexander Cowtan et al. "On the qubit routing problem". In: *arXiv preprint arXiv:1902.08091* (2019).

[6] Yu Tomita and Krysta M Svore. "Low-distance surface codes under realistic quantum noise". In: *Physical Review A* 90.6 (2014), p. 062320.

Accelerated decay rate due to operator spreading in bulk-dissipated many-body quantum systems

Tatsuhiko Shirai[1] * Takashi Mori[2]

[1] *Waseda Institute for Advanced Study, Waseda University*
[2] *Department of physics, Keio University*

Abstract. Many-body dissipative quantum systems exhibit complex relaxation dynamics towards stationality. While the Liouvillian gap describes the asymptotic decay rate, it does not always accurately give the relaxation time due to the existence of long transient regime. This study demonstrates that bulk-dissipated quantum systems exhibit the acceleration of decay rates in the transient regime. We introduce the instantaneous decay rate to study the transient dynamics and explain the accelerated decay rate from the viewpoint of operator spreading. Additionally, we discuss the implications of this accelerated decay rate in quantum computation using gate-based quantum computers.

Keywords: Many-body quantum dynamics, Quantum open systems, Quantum information scrambling, Quantum computation

1 Introduction

The relaxation dynamics of a quantum system coupled to a dissipative environment is a longstanding fundamental issue in non-equilibrium statistical physics [1, 2, 3]. Previous studies have mainly focused on relatively small systems, whereas our understanding of many-body open quantum systems remains limited. Recent experimental progress with superconducting transmons and trapped ions has enabled us to tackle problems using quantum computers. Current quantum computers are considered quantum many-body open systems because they consist of many qubits and are susceptible to errors due to the coupling with dissipative environment [4]. This also motivates us to study many-body open quantum systems.

It has been pointed out that many-body open quantum systems exhibit counter-intuitive dynamical features in the relaxation process. The Markovian dynamics, where the timescale of the environment is much shorter than others, is generated by the Lindblad operator [5, 6]. The spectral gap of the Lindblad operator, called the Liouvillian gap, indicates the asymptotic decay [3], but does not necessarily provide a correct estimate of the relaxation time. Indeed, the relaxation time is much longer than the estimate given by the Liouvillian gap in boundary-dissipated many-body quantum systems, where a conserved current flows in the bulk [7, 8, 9, 10]. This discrepancy is due to the existence of a long-time transient regime. Therefore, clarifying the relaxation dynamics in the transient regime is important.

In this work, we study the autocorrelation function in the steady state of bulk-dissipated quantum systems [11]. We find that bulk-dissipated quantum systems display a much shorter relaxation time than the inverse of the Liouvillian gap, which is different from boundary-dissipated systems. To study the transient dynamics, we introduce the instantaneous decay rate to establish a rigorous bound on the autocorrelation functions. We identify three distinct dynamic regimes (see Fig. 1): the acceleration regime, the plateau regime, and the asymptotic regime. The growth of the instantaneous decay rate in the acceleration regime implies an acceleration of the decay rate. In the following, we explain that the relaxation dynamics is universal in bulk-dissipated many-body systems from the viewpoint of the operator spreading. Additionally, we demonstrate the implications of the accelerated decay rate in quantum computation using noisy-intermediate-quantum (NISQ) computers.

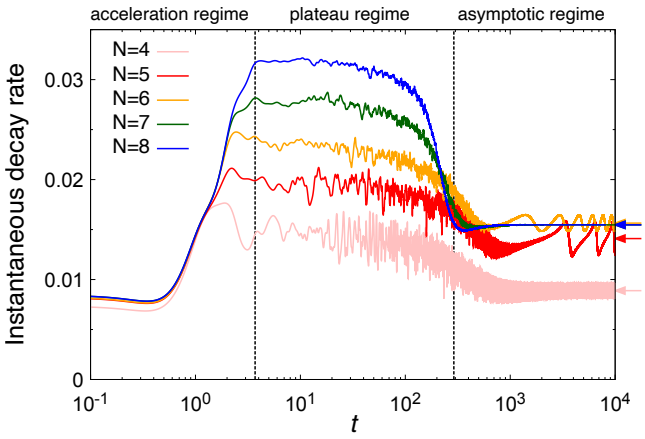


Figure 1: Dynamics of the instantaneous decay rate $\kappa_A(t)$ with $\hat{A} = \sigma_1^z - \langle \sigma_1^z \rangle_{\rm ss}$ for various system sizes in the bulk-dissipated system (see the model in Eq. (1)). Lines show $\kappa_A(t)$ for $N = 4, 5, 6, 7, 8$ from bottom to top. The arrows indicate the values of the Liouvillian gap.

2 Accelerated decay rate of autocorrelation function

We study the decay of autocorrelation functions in steady states. For clarity, we consider a bulk-dissipated one-dimensional N-spin system, which is governed by the

*tatsuhiko.shirai@aoni.waseda.jp

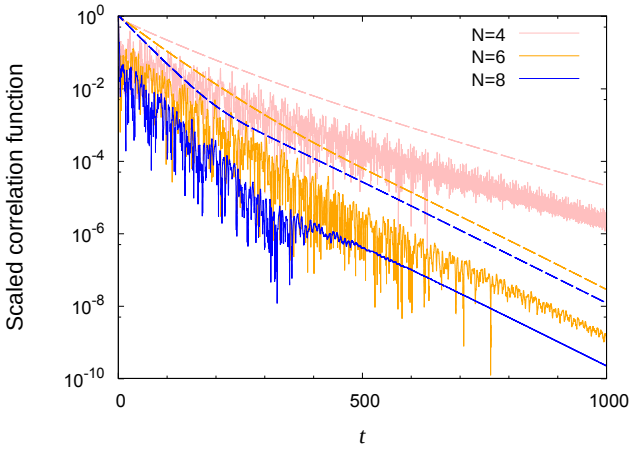


Figure 2: Dynamics of the autocorrelation function $C_A(t)$ with $\hat{A} = \hat{\sigma}_1^z - \langle\hat{\sigma}_1^z\rangle_{\rm ss}$ for various system sizes in the bulk-dissipated system. Scaled correlation function $|C_A(t)/C_A(0)|$ for $N = 4, 6, 8$ from top to bottom (solid lines) and the upper bounds in Eq. (3) (Dashed lines) are depicted.

following Lindblad equation:

$$\frac{d}{dt}\rho = -i[\hat{H},\rho] + \gamma\sum_{i=1}^{N}\left(\hat{L}_i\rho\hat{L}_i^\dagger - \frac{1}{2}\left\{\hat{L}_i^\dagger\hat{L}_i,\rho\right\}\right) = \mathcal{L}\rho,$$
$$\begin{cases}\hat{H} = \sum_{i=1}^{N}\left(h^z\hat{\sigma}_i^z + h^x\hat{\sigma}_i^x + J\hat{\sigma}_i^z\hat{\sigma}_{i+1}^z\right), \\ \hat{L}_i = \hat{\sigma}_i^- = \frac{1}{2}(\hat{\sigma}_i^x - i\hat{\sigma}_i^y),\end{cases} \tag{1}$$

where $\mathcal{L}$ is a superoperator called Liouvillian acting on density operator ρ, and $\{\hat{\sigma}_i^\alpha\}_{\alpha\in\{x,y,z\}}$ denote the Pauli operators acting on site i. We set the parameters to $h^z = 0.9045$, $h^x = 0.809$, and $J = 1$, for which the eigenstate thermalization hypothesis is numerically shown to be satisfied [13], and fix the dissipation strength as $\gamma = 0.01$. The steady state is given by $\mathcal{L}\rho_{\rm s} = 0$.

Figure 2 shows the dynamics of autocorrelation functions for various system sizes by solid lines. The autocorrelation function is defined as [12]

$$C_A(t) = \langle A(t), A\rangle_{\rm s} = \langle A(t), A\rho_{\rm s}\rangle \tag{2}$$

where $\langle A, B\rangle = {\rm Tr}(A^\dagger B)$ and $\langle A, B\rangle_{\rm s} = {\rm Tr}(A^\dagger B\rho_{\rm s})$. Here, $A(t) = \exp(\tilde{L}t)A$ where $\tilde{\mathcal{L}}$ is a conjugate superoperator defined as $\langle\tilde{\mathcal{L}}A, B\rangle = \langle A, \mathcal{L}B\rangle$. The autocorrelation function decays at the Liouvillian gap in the long-time asymptotic regime. However, we found a transient regime where the decay rate is larger than the decay rate of the Liouvillian gap.

In order to study the relaxation dynamics in the transient regime, we derive a rigorous upper bound on the autocorrelation functions as

$$|C_A(t)| \le \exp\left(-\int_0^t \kappa_A(\tau)d\tau\right)C_A(0), \tag{3}$$

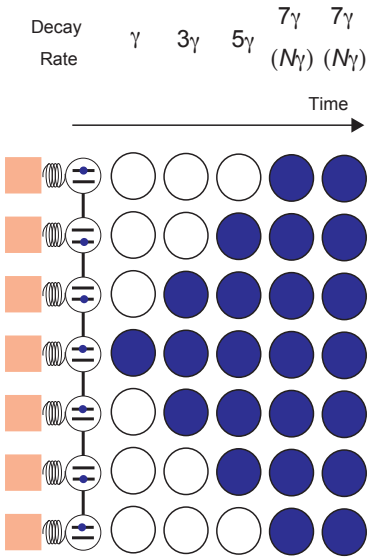


Figure 3: Schematic picture of accelerated decay in a bulk-dissipated quantum chain with $N = 7$ spins. Filled circles represent sites where an operator $\hat{A}(t)$ nontrivially acts.

where $\kappa_A(t)$ is called the instantaneous decay rate and is given by

$$\kappa_A(t) = -\frac{1}{2}\frac{\langle A(t), (\tilde{\mathcal{L}} + \tilde{\mathcal{L}}^*)A(t)\rangle_{\rm s}}{\langle A(t), A(t)\rangle_{\rm s}}. \tag{4}$$

Here, $\tilde{\mathcal{L}}^*$ is defined as $\langle\tilde{\mathcal{L}}^*A, B\rangle_{\rm s} = \langle A, \tilde{\mathcal{L}}B\rangle_{\rm s}$. The instantaneous decay rate satisfies (i) $\kappa_A(t) \ge 0$ and (ii) $\kappa_A(t)$ approaches the Liouvillian gap in the long-time limit.

Figure 1 shows the dynamics of the instantaneous decay rate for various system sizes. There are three dynamic regimes: the acceleration regime, the plateau regime, and the asymptotic regime. The instantaneous decay rate increases over time in the acceleration regime, takes a constant value in the plateau regime, and decays to the value of the Liouvillian gap in the asymptotic regime.

We explain the three dynamic regimes from the viewpoint of operator spreading. Figure 3 provides a schematic picture of the relaxation dynamics in one-dimensional bulk-dissipated systems. Here, we consider $A = O_i$ as an operator acting on a single site i. The filled circles in the figure denote sites where the operator $A(t)$ nontrivially acts. The number of filled circles represents the operator size. The operator is initially localized to site i, making the operator size one. Then, the operator size increases with time, resulting in the operator spreading [14, 15]. Let us study the effect of bulk dissipation. The single-site operator O_i decays at the rate of γ due to the coupling with the dissipative environment. As a result, operator $O_{i-1}O_iO_{i+1}$ decays at the rate of 3γ. In this way, the decay rate is accelerated in the acceleration regime as the operator size increases with time.

In the plateau regime, the instantaneous decay rate takes a constant value proportional to the system size. This observation can also be understood from the viewpoint of operator spreading. Namely, since the operator spreads over the entire system in this regime, the decay rate is proportional to the system size.

In the asymptotic regime, the instantaneous decay rate approaches the decay rate of the Liouvillian gap. The oscillation of the instantaneous decay rate for small-sized systems implies the degeneracy of the Liouvillian gap.

In Fig. 2, we compare the dynamics of the autocorrelation function with the upper bound in Eq. (3). The autocorrelation function exhibits a rapid decay in the initial stage of the dynamics, which is not captured by the bound. The instantaneous decay rate fails to describe this initial decay due to the Hamilton dynamics. However, It is worth noting that the upper bound accurately describes the decay rate in the long-time transient regime. The decay rate in the transient regime is well reproduced by the plateau value of the instantaneous decay rate.

3 Demonstration of accelerated decay rates in quantum computation

Here, we demonstrate the implications of the accelerated decay rate in quantum computation on noisy quantum computers.

Let us consider a two-point measurement on an N-qubit quantum system. First, we prepare an initial state $|\psi\rangle$ and perform a projection measurement of observable A. A post-measurement state $|a\rangle$ (i.e., $A|a\rangle = a|a\rangle$) is obtained with a probability of $|\langle a|\psi\rangle|^2$. Then, the system evolves over a time $2T$ in the presence of noise. The unitary part of the time evolution for the period $t \in [0, T]$ is given by U, and that for the period $t \in [T, 2T]$, it is given by $U^\dagger$. The state $|a\rangle\langle a|$ evolves as

$$\varepsilon_{U^\dagger}\varepsilon_U(|a\rangle\langle a|), \tag{5}$$

where ε_U and $\varepsilon_{U^\dagger}$ are noise maps corresponding to U and $U^\dagger$, respectively. Finally, a projection measurement of A gives an output state $|a'\rangle$ with a probability of $\langle a'|\varepsilon_{U^\dagger}\varepsilon_U(|a\rangle\langle a|)|a'\rangle$.

The correlation function resulting from the two-point measurement of observable A at times $t = 0$ and $t = 2T$ is given by

$$C_A^{(2)}(U) = \sum_{a,a'} aa' |\langle a|\psi\rangle|^2 \langle a'|\varepsilon_{U^\dagger}\varepsilon_U(|a\rangle\langle a|)|a'\rangle . \tag{6}$$

Randomly sampling the initial state $|\psi\rangle$ to satisfy $\overline{[|\psi\rangle\langle\psi|]} = I/2^N$ provides the ensemble average of $C_A^{(2)}(U)$ as

$$\overline{C_A^{(2)}(U)} = \frac{1}{2^N}\langle A_U, A_U\rangle , \tag{7}$$

$A_U = \varepsilon_U A$. When the dissipative dynamics ε_U is generated by the Lindblad equation as $\varepsilon_U = \mathrm{T}\exp\left(\int_0^T \mathcal{L}(t)dt\right)$, where T is a time-ordering operator and $\mathcal{L}(t)$ is a time-dependent Liouvillian at t, we obtain an equality:

$$\overline{C_A^{(2)}(U)} = \frac{1}{2^N}\exp\left(-\int_0^T 2\kappa_{As}(t)dt\right)\langle A, A\rangle . \tag{8}$$

Here, the instantaneous decay rate is given as

$$\kappa_{As}(t) = -\frac{1}{2}\frac{\langle A_s(t), (\mathcal{L}(t) + \tilde{\mathcal{L}}(t))A_s(t)\rangle}{\langle A_s(t), A_s(t)\rangle}, \tag{9}$$

where $A_s(t) = \mathrm{T}\exp(\int_0^t \mathcal{L}(\tau)d\tau)A$. The acceleration of the decay rate should be observed for $\overline{C_A^{(2)}(U)}$ since $\kappa_{As}(t)$ is proportional to the operator size of $A_s(t)$, and $A_s(t)$ spreads to the entire system with time.

We calculate the correlation function using a simulator of gate-typed quantum computers on the one-dimensional architecture. In Fig. 4 (upper), we present a quantum circuit representing ϵ_U with a circuit depth ℓ and $N = 4$. $U^{(2)}$ denotes the random two-qubit gate from Ref. [16], and $\mathcal{N}_z$ denotes the dephasing noise channel with a dissipation strength $\gamma = 0.001$. The initial state is set to $|\psi_i\rangle = |0\rangle^{\otimes N}$, where $\sigma_i^z|0\rangle = |0\rangle$. We average the result over 1000 samples of the set of $U^{(2)}$ and $\mathcal{N}_z$.

In Fig. 4 (lower), we depict the circuit-depth dependences of the two-point correlation functions of $A = \sum_{i=1}^N \sigma_i^z$ for various system sizes. C_A shows a crossover between the acceleration regime and the plateau regime. In the acceleration regime, $C_A = \exp(-O(\ell^2))$, whereas in the plateau regime, $C_A = \exp(-O(N\ell))$.

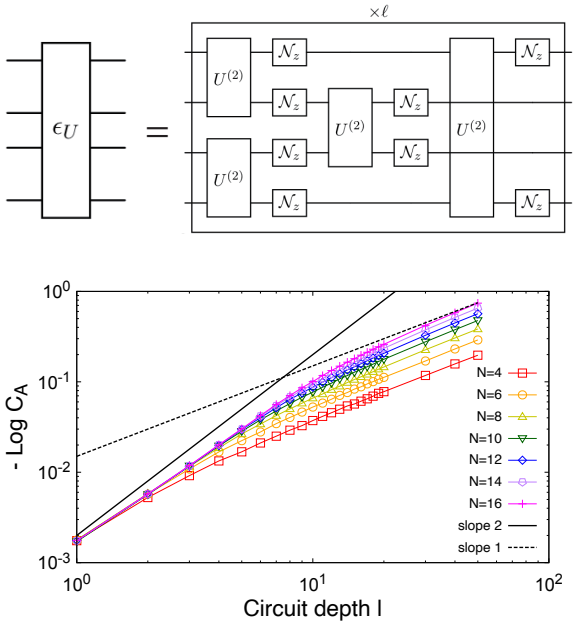


Figure 4: (Upper) Noise map ε_U on quantum circuit with 4 qubits. (Lower) Circuit-depth dependence of the two-point correlation functions of $A = \sum_{i=1}^N \sigma_i^z$ for various system sizes. The two guiding lines clearly illustrate the crossover between the acceleration regime and the plateau regime.

References

[1] H-P. P. Breuer. The theory of open quantum systems Oxford Univ. Press, 2002.

[2] H. Sphon. Kinetic equations from Hamiltonian dynamics: Markovian limits *Rev. Mod. Phys.*, 52:569–615, 1980.

[3] M. Žnidarič. Relaxation times of dissipative many-body quantum systems *Phys. Rev. E*, 92:042143, 2015.

[4] J. Preskill. Quantum Computing in the NISQ era and beyond *Quantum*, 2:79, 2018.

[5] G. Lindblad. On the generators of quantum dynamical semigroups *Comm. Math. Phys.*, 48:119–130, 1976.

[6] V. Gorini, A. Kossakowski, E. C. G. Sudarshan. Completely positive dynamical semigroups of N-level systems *Journal of Mathematical Physics*, 17:821–825, 1976.

[7] T. Mori, T. Shirai. Resolving a Discrepancy between Liouvillian Gap and Relaxation Time in Boundary-Dissipated Quantum Many-Body Systems *Phys. Rev. Lett.*, 125:230604, 2020.

[8] T. Haga, M. Nakagawa, R. Hamazaki, M. Ueda. Liouvillian Skin Effect: Slowing Down of Relaxation Processes without Gap Closing *Phys. Rev. Lett.*, 127:070402, 2021.

[9] G. Lee, A. McDonald, A. Clerk. Anomalously large relaxation times in dissipative lattice models beyond the non-Hermitian skin effect *Phys. Rev. B*, 108:064311, 2023.

[10] T. Sawada, K. Sone, R. Hamazaki, Y. Ashida, T. Sagawa. Role of Topology in Relaxation of One-Dimensional Stochastic Processes *Phys. Rev. Lett.*, 132:046602, 2024.

[11] T. Shirai, T. Mori. Accelerated Decay due to Operator Spreading in Bulk-Dissipated Quantum Systems *Phys. Rev. Lett.*, 133:040201, 2024.

[12] H. Carmichael. Statistical Methods in Quantum Optics 1: Master Equations and Fokker-Planck equations Springer Berlin, Heidelberg, 1998.

[13] H. Kim, T. N. Ikeda, D. A. Huse. Testing whether all eigenstates obey the eigenstate thermalization hypothesis *Phys. Rev. E*, 90:052105, 2014.

[14] C. W. von Keyserlingk, T. Rakovszky, P. Tibor, F. Pollmann, S. L. Sondhi. Operator Hydrodynamics, OTOCs, and Entanglement Growth in Systems without Conservation Laws *Phys. Rev. X*, 8:021013, 2018.

[15] A. Nahum, S. Vijay, J. Haah. Operator Spreading in Random Unitary Circuits *Phys. Rev. X*, 8:021014, 2018.

[16] D. Hanneke, J. P. Home, J. D. Jost, J. M. Amini, D. Leibfried, D. J. Wineland. Realization of a programmable two-qubit quantum processor *Nature Physics*, 6:13-16, 2010.

Quantum metrology performance with proper resource accounting

Yink Loong Len[1 2 3 *] Tejas Acharya[2 †] Alexia Auffèves[4 2 3 ‡] Hui Khoon Ng[1 2 3 §]

[1] *Yale-NUS College, Singapore*
[2] *Centre for Quantum Technologies, National University of Singapore, Singapore*
[3] *MajuLab, CNRS-UCA-SU-NUS-NTU International Joint Research Laboratory, Singapore*
[4] *Université Grenoble Alpes, CNRS, Grenoble INP, Grenoble, France*

Abstract. We quantify the full resource costs of quantum metrology protocols, in particular, taking into account the practical preparation costs of exotic quantum states. We benchmark performance according to actual resource costs, rather than a standard comparison based on the number of probes used. This leads us to a new figure of merit that unveils surprising results: Nonclassical states that offer estimation enhancements in the conventional characterization using probe numbers (e.g. give the Heisenberg scaling) can perform worse than classical states in practice, due to their high resource costs.

Keywords: Quantum metrology, resource, quantum technology, quantum estimation

1 Motivation & Introduction

Quantum metrology, with the utilization of nonclassical properties of the sensor states, allows for enhanced sensitivity in estimation of physical parameters beyond classical limits. Conventionally, the performance of quantum metrology schemes is quantified by the estimation accuracy, as defined by the quantum Fisher information (QFI), for a given number of probes n involved in the sensing task. In this picture then, nonclassical states such as squeezed states and NOON states are shown to be superior than classical states like coherent states: The former can attain estimation precision that increases quadratically with n (i.e., the Heisenberg scaling), whereas the estimation precision for the latter can at most scale linearly with n [1].

Such a characterization of metrological performance according to n disregards important details of the practical protocol, including the resource costs of the sensor state preparation, the application of controlled operations, the measurement stage, as well as experimental implementation constraints that may incur costs invisible to fundamental considerations. Indeed, there is no reason to expect a simple relation between the QFI for a metrological scheme and its actual resource cost; such a relation will depend on the actual implementation. In particular, one can anticipate that, while exotic nonclassical states may promise better QFI and precision scaling, the states in question may be so difficult to prepare in practice that the resource costs overwhelm any advantage achieved over conventional classical states. In addition, practical limitations that result in imperfect operations can affect the performance of different metrology schemes in different ways; a proper comparison must take into account also the robustness of a scheme against imperfections.

The main objective of this work is then **to identify and properly evaluate the relevant resources used in quantum metrology schemes, and to provide performance assessments based on resource costs, beyond simplistic probe-number comparisons.**

2 Results

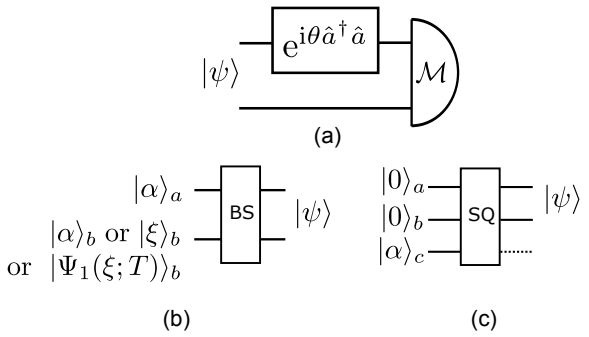


Figure 1: Phase sensing in a two-mode interferometer. (a) We prepare the input sensor state $|\psi\rangle$, which undergoes the transformation $|\psi\rangle \to e^{i\theta\hat{a}^\dagger\hat{a}}|\psi\rangle$, and is then measured by $\mathcal{M}$ such that information about θ can be obtained. For the choice of input state $|\psi\rangle$, we consider either (b) a coherent state in mode a and either a coherent state $|\alpha\rangle_b$, squeezed vacuum $|\xi\rangle_b$, or "cat" state $|\Psi_1(\xi;T)\rangle_b$ in mode b, then followed by a mixing at a 50:50 beam splitter (BS), or (c) a two-mode squeezed vacuum state, which is obtained by post selecting the signal and idler photons from the interaction between a coherent pump $|\alpha\rangle_c$ and a "squeezer" (SQ), which is essentially a non-linear crystal.

We focus on the standard metrology task of sensing the phase θ in an two-mode optical interferometer; see Fig. 1(a). A relevant practical resource cost is the average energy cost, or "work credit" W_c [2], of preparing the input sensor states. We compare the performance of four well-studied schemes, based on four different input states: (i) coherent states, (ii) NOON states, (iii) entangled coherent states (ECS) [3], and (iv) two-mode squeezed vacuum (TMSV) states. The coherent state case, regarded as the classical situation, shows QFI that scales linearly

*yinkloong@quantumlah.org
†tejas.acharya.221@gmail.com
‡alexia.auffeves@cnrs.fr
§huikhoon.ng@nus.edu.sg

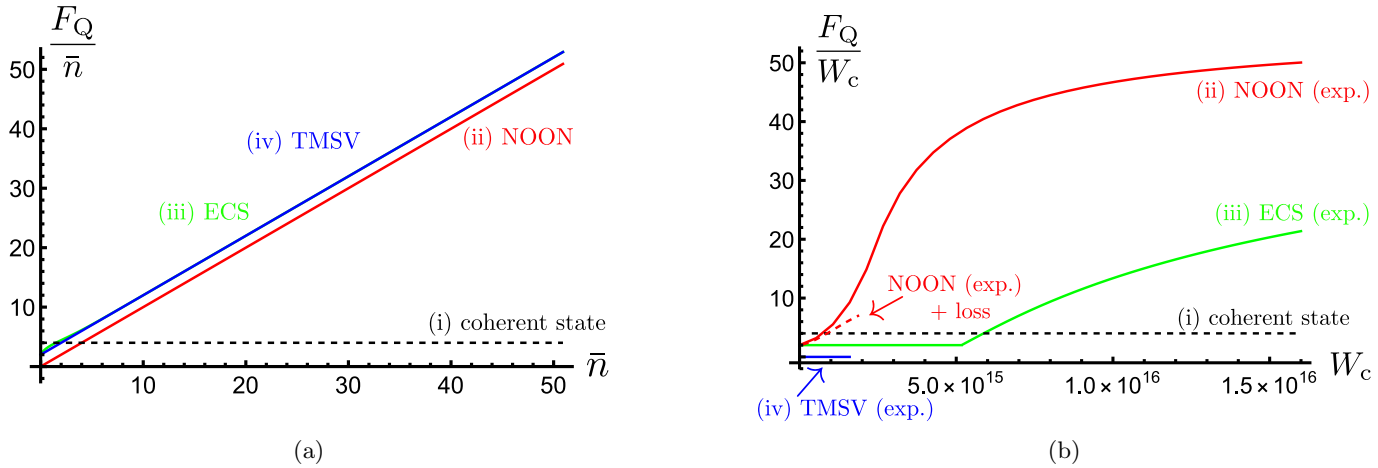


Figure 2: (a) Conventional plot for the performance of phase sensing in optical quantum metrology based on probe numbers: QFI per mean photon number $\bar{n}$ as a function of $\bar{n}$. All the three ideal nonclassical states exhibit the Heisenberg scaling, in contrast with the coherent state situation which serves as the reference for classical scaling. The curves for the ECS and TMSV coincide, except at small $\bar{n}$ values. (b) Resource-based figure of merit for the estimation performance, viz QFI per W_c as function of the energy cost W_c. W_c is evaluated for real experimental state-preparation schemes, and is quantified by the total average energy (units: $\hbar\omega$) used in the state generation. In contrast to the Heisenberg scaling of Fig. 2(a), none of the nonclassical states exhibit a quadratic scaling of QFI with W_c. Moreover, finite squeezing capabilities in experiments restricts TMSV to be produced only up to about the range of $W_c \approx 1.6 \times 10^{15}$; in that range, it performs worse than the coherent-state scheme. For the ECS scheme, it is more beneficial to not have the cat-state component in the input at all when W_c is less than about 5.2×10^{15}. We also perform a quick robustness check on the effect of weak loss on the NOON state scheme, as shown in the red dashed line.

as the mean photon number $\bar{n}$, whereas the other three nonclassical states, ideally, follow the Heisenberg scaling as shown in Fig. 2(a). For each state, we compute the work credit for the state preparation in practical experimental settings, and then the QFI for the encoded state.

To properly evaluate the resource costs of quantum phase sensing using these states, we have to consider experimental implementations. We select implementations that have actually been carried out, or are feasible for real experiments. For coherent states [see Fig. 1(b)], we consider two individual lasers ($|\alpha\rangle_a$ and $|\alpha\rangle_b$) that mix at a 50:50 beam splitter (BS). As beam splitter is energy-conserving, the work credit W_c is the sum of the average energy stored in the two laser fields ($2\hbar\omega|\alpha|^2$, with ω as the laser frequency). The NOON state and the ECS [see again Fig. 1(b)] are obtained by mixing a (single-mode) squeezed state $|\xi\rangle_b$ and a cat state $|\Psi_1(\xi;T)\rangle_b$, respectively, with a coherent state at a BS [5, 3]. The respective W_c is the sum of the cost to generate the squeezed state and the cat state, plus the energy of the coherent-state component. For the cat-state generation, we consider the method as proposed by [6], whereby a single-mode squeezed state is passed through a beam splitter, and is only post-selected upon detecting a photon in the reflected port of the beam splitter (reflectivity T). The cost for the cat state is thus the cost of a squeezed state multiplied by a factor that compensates for the non-unity post-selection probability. Lastly, TMSV state [see Fig. 1(c)] comprises the signal and idler states of the (type-II) spontaneous parametric down-conversion (SPDC) process [4], and its W_c is hence the cost for the SPDC.

The cost of generating a squeezed state (single- and two-mode) is the energy stored in the pump laser in the SPDC process. While one can argue that a careful pump-recycling protocol can recover the energy from the pump beam to be reused in a subsequent sensing round, one nevertheless has to come up with the "energy credit" needed to power the pump beam in the first place. We find that, to obtain a squeezed state with squeezing parameter ξ, we require the energy input of $\kappa^2|\xi|^2\hbar\omega$, where ω is the frequency of the down-converted photons, and the factor κ is a conversion factor that characterizes the efficiency of generating a squeezed state from the pump laser via the SPDC process. It captures the expensive nature of SPDC, and hence an important overhead in generating nonclassical states, as we find that $\kappa \sim \frac{1}{\chi^{(2)}} \sim 10^7$, where $\chi^{(2)}$ is the second-order nonlinear susceptibility of the nonlinear crystal needed for the SPDC.

Fig. 2(b) shows our results, in which the QFI per W_c is plotted against the energy cost W_c for the schemes based on the different input states, where for the NOON state and ECS, for the given W_c, the QFI plotted is obtained with further optimization over the distribution of the resource cost over their constituent components, i.e., the squeezed state or cat state component respectively with the coherent state component. The performance of the different schemes are notably distinct from the conventional picture Fig. 2(a). Not only are their relative performance altogether different, under this energetic comparison, the Heisenberg scaling is lost alto-

gether, and nonclassical states need not be energetically better, compared with classical states, for sensing tasks. Note that, to acknowledge current experimental capabilities in squeezed-state generation, we limit $|\xi| \leq 2$. This restricts the experimental TMSV scheme to a limited range of $W_{\rm c}$, in which it actually performs worse than the coherent-state scheme. This is also the reason the Heisenberg scaling is lost for the nonclassical states, as the energy is directed into the coherent state component when the squeezing reaches its practical limit. Finally, we perform a quick check on the effect of imperfect implementation in the form of weak lossy interferometer, on the NOON state scheme—calculate using expansion over the leading power of the weak loss parameter, and hence a limited range of $W_{\rm c}$ is plotted. As shown by the red dashed line, the overall performance is worsen, though it suggests still a performance that is better than an ideal scheme with coherent state.

Our plots tell the story for the specific experimental schemes we have studied, and for the specific way we have accounted for resource costs. Nevertheless, the general conclusion holds: **Whether a metrology scheme performs well has to be judged based on the total resource costs incurred in the whole protocol**, from the initial state preparation to the final measurement, and including any additional costs arising from experimental constraints. Naive comparisons based on abstract quantities separate from experimental considerations cannot tell the full story.

References

[1] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum metrology," Phys. Rev. Lett. **96**, 010401 (2006).

[2] P. Lipka-Bartosik and R. Demkowicz-Dobrzański, "Thermodynamic work cost of quantum estimation protocols," J. Phys. A: Math. Theor. **51**, 474001 (2018).

[3] J. Joo, W. J. Munro, and T. P. Spiller, "Quantum metrology with entangled coherent states," Phys. Rev. Lett. **107**, 083601 (2011).

[4] J. Qin, et al., "Unconditional and robust quantum metrological advantage beyond N00N states," Phys. Rev. Lett. **130**, 070801 (2023).

[5] I. Afek, O. Ambar, and Y. Silberberg, "High-noon states by mixing quantum and classical light," Science **328**, 879 (2010).

[6] M. Dakna, et al., "Generating Schrödinger-cat-like states by means of conditional measurements on a beam splitter," Phys. Rev. A **55**, 3184 (1997).

Limitations of Noisy Quantum Devices in Computational and Entangling Power

Yuxuan Yan[1] * Zhenyu Du[1] Junjie Chen[1] Xiongfeng Ma[1] †

[1] *Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, P. R. China*

Abstract. This work studies the limitations of noisy quantum devices with the help of general classical processing, showing that noisy quantum devices with a circuit depth exceeding $O(\log n)$ provide no computational advantage for any quantum algorithm, rigorously rules out the possibility of well-known quantum algorithms, including Shor's, Grover's, Harrow-Hassidim-Lloyd, and linear-depth variational algorithms. Moreover, we show one-dimensional noisy quantum devices have no super-polynomial computational advantage. Then, we study the entangling power of one- and two-dimensional noisy quantum devices, establishing a maximum entanglement growth bound of $O(\log n)$ for one-dimensional chains, underscoring significant constraints on quantum simulation and scalability. (Preprints: arXiv 2306.02836.)

Keywords: noisy intermediate-scale quantum devices, quantum computational advantage, quantum entanglement

Finding solid and practical quantum advantages via noisy quantum devices without error correction is a critical but challenging problem. Conversely, comprehending the fundamental limitations of the state-of-the-art is equally crucial. In this work, we observe the polynomial-time indistinguishability of n-qubit devices from random coins when circuit depths exceed $\omega(\log(n))$ under single-qubit depolarizing noise. Even with classical processing, we can demonstrate the absence of computational advantage in polynomial-time algorithms with noisy quantum circuits of super-logarithmic depths. This finding decisively negates the feasibility of executing prominent quantum algorithms such as Shor's[1], Grover's[2], and the Harrow-Hassidim-Lloyd algorithm[3]. In addition, our results apply to variational quantum algorithms [4, 5], error mitigation [6, 7, 8], and quantum simulation with polynomial depths. Furthermore, we consider noisy quantum devices with restraint gate topology. We rule out super-polynomial quantum advantages for a one-dimensional noisy qubit array in all-depth regimes. We also establish upper limits on entanglement generation: $O(\log(n))$ for one-dimensional arrays and $O(\sqrt{n}\log(n))$ for two-dimensional arrays. Our findings underscore the entanglement scalability constraints in noisy quantum devices. Our findings are summarised in Fig. 1.

Model of noisy quantum devices.—In the NISQ era, we may describe a noisy quantum device by employing layers of unitary gates followed by independent single-qubit depolarizing noise channels after each layer. The noise acts on a single qubit as

$$\Lambda_1(\rho) = (1-p)\rho + p\frac{I}{2}, \tag{1}$$

where p is the strength of noise. After applying all layers of gates and noise, we perform computational-basis measurements at the end of the circuit to obtain the classical output.

*yanyx21@mails.tsinghua.edu.cn
†xma@tsinghua.edu.cn

Definition 1 (Noisy quantum devices)**.** *A noisy quantum device, with n qubits, produces a quantum state at a depth of t,*

$$\rho(t) = \Lambda \circ \mathcal{U}_t \circ \Lambda \circ \cdots \circ \Lambda \circ \mathcal{U}_2 \circ \Lambda \circ \mathcal{U}_1(|0\rangle\langle 0|^{\otimes n}), \tag{2}$$

where $\mathcal{U}$'s are layers of gates and $\Lambda = \Lambda_1^{\otimes n}$ is the noise channel. In each layer, a qubit can be manipulated at most by one quantum gate. The classical output $C_n(\rho(t))$ from measurements obeys the distribution $\Pr[C_n(\rho(t)) = X] = \langle X|\rho(t)|X\rangle$, where X is a n-bit string and $|X\rangle$ is the corresponding computational basis.

Importantly, our model does not allow mid-circuit measurements and resets, i.e., replacing a qubit with a known pure state, such as $|0\rangle$. Because both of these operations will enable fault-tolerant quantum computing, which goes beyond the scope of the NISQ era [10, 11, 12]. As a result, we lose the information of the quantum state layer by layer without the opportunity of retrieval. The information loss is exponentially fast regarding the circuit depth[13, 14].

Limitations of depths for computational advantages.—By analyzing the entropy of measurement outcomes, we find that if an algorithm queries noisy quantum devices with $\omega(nq)$ depth, where q is the circuit execution times, the output of the noisy quantum device will be too noisy to provide any quantum advantages for computation, as depicted in Fig. 2. Therefore, if we replace them with random coins, the answer to decision problems will not be influenced. The observation leads to the following theorem.

Theorem 2 (informal version)**.** *Consider a hybrid algorithm operating within time $T(n)$ with nq queried measurement bits from noisy quantum devices. If circuit depth $t \geq \omega(\frac{\log(nq)}{-\log(1-p)})$, noisy devices will yield no quantum advantages. In this case, a classical algorithm, running within $T(n)$ time, can be used instead by replacing the noisy device queries with random coins from a uniform distribution.*

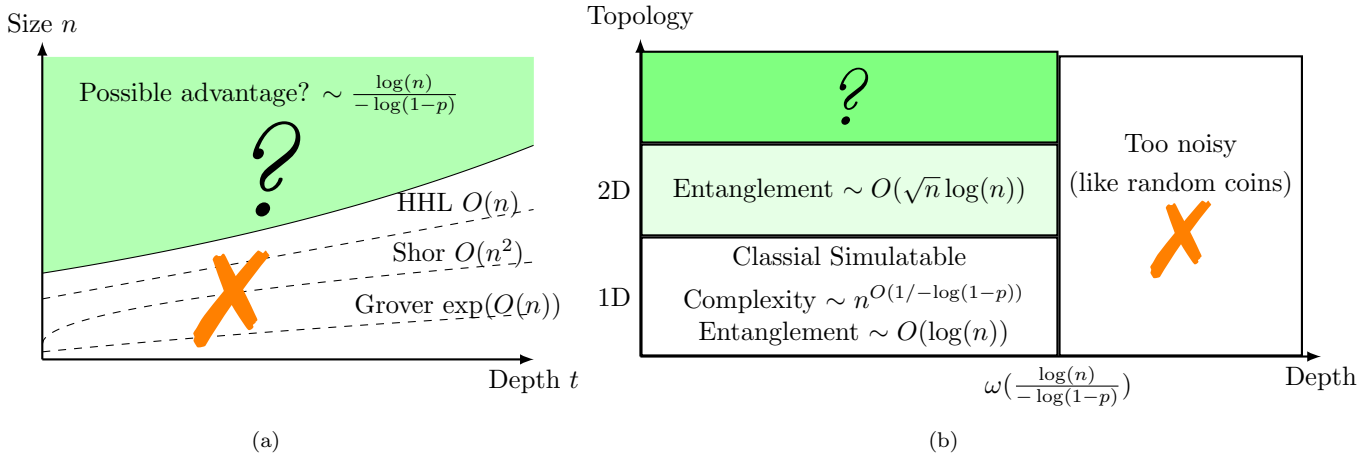


Figure 1: **Summary of the limitations for noisy quantum devices without error correction. a** For algorithms with generic classical processing control, we prove that devices with circuit depths beyond $\omega(\log(n))$ under single-qubit depolarizing noise are too noisy to offer any computational advantage in a polynomial running time, including well-known quantum algorithms, such as Shor's, Grover's, and the Harrow-Hassidim-Lloyd algorithm. The depth needed in the table is the best implementation, as far as we know, without additional space overhead. [2] The scaling of depth with the qubit number, n, is shown by dashed curves for each algorithm, with a solid curve showing the logarithmic upper limit. The scaling is shown in an asymptotical limit, i.e., when n is large. **b** In the regime where the circuit depth is below the logarithmic scaling, potential quantum advantages depend on the gate connection topology of noisy quantum devices. For the one-dimensional case, we prove classical simulatability for noisy devices and give an entanglement upper bound of $O(\log(n))$. For the two-dimensional case, the entanglement generation upper bound scales as $O(\sqrt{n}\log(n))$. Therefore, super-polynomial advantages without error correction are only possible when gate connectivity is higher than one dimension, and the circuit depth is below logarithmic scaling for noisy devices. Such a regime is colored in green.

An immediate consequence of Theorem 2 is that noisy quantum devices with super-logarithmic depth do not provide any quantum advantage for polynomial-time quantum algorithms, regardless of the classical processing that controls quantum devices depend on previous measurement outcomes. This result helps us explicitly eliminate the advantages of implementing a broad class of quantum algorithms on noisy quantum devices that only query super-logarithmic-depth quantum circuits. Examples include but are not limited to Shor's [1], Grover's [2], and Harrow-Hassidim-Lloyd (HHL) algorithm [3].

Moreover, our results have significant implications for NISQ algorithms, such as variational quantum algorithms and quantum error mitigation. Our information-theoretical limitations on circuit depth unify and generalize previous research in this area. Variational quantum algorithms aim to solve optimization problems in hybrid schemes [4, 5]. For linear growing circuit depth, our results directly lead to the exponential complexity required for quantum devices to impact the optimization result and to provide advantages. This has been perceived as the noise-induced barren plateau problem [15], which involves exponentially vanishing gradients with n due to noise. Quantum error mitigation techniques are introduced to decrease errors in expectation values at the cost of increased sampling overhead [6, 7, 16]. For single-copy mitigation schemes, our results strengthen the previous findings of exponential sampling overhead [17, 18] by considering the dependence of circuits on previous measurement outcomes.

Limitations of computational advantages in one-dimensional local circuits.—In real-world quantum devices, gates are often subjected to certain topologies, further restraining their computational power. Here, we consider a one-dimensional qubit array, where gates are restrained between the nearest neighbor qubits on a linear chain. We show that one-dimensional noisy devices do not possess any super-polynomial computational advantages, regardless of their depth. Our finding is based on a state-vector simulator created for constant-depth one-dimensional circuits [19]. We have demonstrated that the simulator can be extended to one-dimensional noisy devices with a depth of $O(\log(n))$, thus ruling out the possibility of super-polynomial advantages.

Theorem 3 (informal version)**.** *One-dimensional noisy quantum devices that run in computational time $T(n)$ can be simulated by a classical algorithm with $T(n)^{1+\frac{1}{-\log(1-p)}}$ computational time, thus having no super-polynomial quantum advantages.*

Our results stress that the connectivity of quantum devices is particularly important in the NISQ diagram. Higher than two dimensions in qubit connection are required to obtain super-polynomial advantages with noise.

Limitations of quantum entanglement production.—After ruling out quantum advantages of noisy quantum devices above $O(\log n)$ depth, we consider the

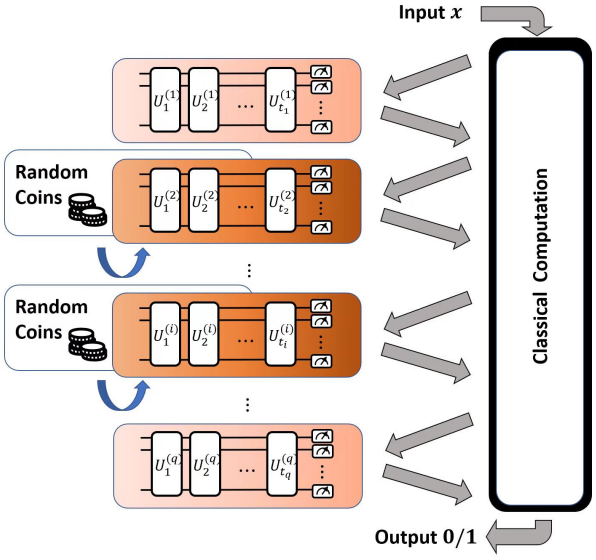


Figure 2: In a hybrid algorithm, the orange boxes on the left represent noisy quantum devices, in which a darker color represents noisy quantum devices with super-logarithmic depths, $\omega(\log(nq))$. The classical computer on the right can control noisy quantum devices based on previous measurement outputs. After obtaining data from quantum devices and classical computation, the classical computer outputs 0 or 1. We use the boxes with coins to show that random coins can replace noisy quantum devices with super-logarithmic depths. After the replacement, the new classical algorithm can still give the same output bit. Thus, the noisy quantum devices with super-logarithmic depth do not provide any quantum advantages.

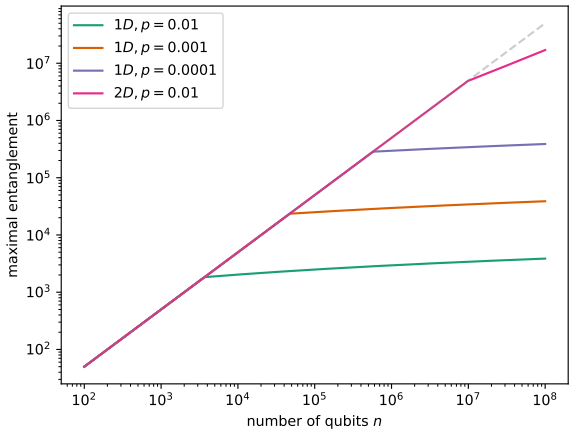


Figure 3: Limitations of entanglement production, with varied number of qubits n and noise strength p, which are taken as the minimal values between $n/2$ and the upper bounds in Theorem 4. Lines in different colors correspond to different noise strengths p, and one of the lines is for the case of a two-dimensional qubit connection.

case of arbitrary depths and study the entangling power. In this part, our results depend on the topology of qubit connections. We study one- and two-dimensional connections, which are two typical designs for quantum devices.

Theorem 4 (Upper bound on entanglement in one dimension for noisy quantum devices)**.** *For a contiguous half A and the complement half $\bar{A}$ in an n-qubit noisy quantum device with a one-dimensional connection topology, the quantum mutual information and hence the quantum relative entropy of entanglement are upper bounded by*

$$E_R(A:\bar{A}) \leq I(A:\bar{A}) \leq \frac{\log(n)}{-2\log(1-p)}, \qquad (3)$$

where p is the noise strength defined in Eq. (1).

Our findings have significant implications for quantum simulation, a crucial application of quantum computing devices [20, 21, 22]. The role of entanglement in quantum simulation is critical and was explored in recent experiments [23]. Our results imply any quantum systems with a super-logarithmic entanglement scaling will not be able to be efficiently prepared and simulated in a noisy quantum device. This limitation extends to a wide range of quantum systems, including highly excited states at the mid-spectrum of local Hamiltonian [24] and thermalized quantum states in quantum dynamics [25, 26], such as quantum many-body thermalization and black hole dynamics. We also investigate entanglement between distant regions in qubit chains and show that entanglement decay to $\frac{4p}{1-p}$ exponentially with the distance of the two regions.

For two-dimensional lattices, we consider qubits arranged in a square of side length $\sqrt{n}$ and show that maximal entanglement is $O(\sqrt{n}\log(n))$. For one- and two- dimensional lattices, we present the numerical upper bounds of entanglement for different values of noise strength p in Fig. 3. After the number of qubits reaches a certain number related to the noise strength, the further growth of quantum entanglement in the noisy quantum device will be suppressed. For the one-dimensional case, this will lead to an exponential cost of qubits required to scale up the system's entanglement further due to the logarithmic scaling of the upper bounds of entanglement. In the two-dimensional case, a polynomial cost is also required.

References

[1] Peter W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Journal on Computing*, 26(5):1484–1509, October 1997.

[2] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing - STOC '96*, pages 212–219, Philadelphia, Pennsylvania, United States, 1996.

[3] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum Algorithm for Linear Systems of Equations. *Physical Review Letters*, 103(15):150502, October 2009.

[4] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, 92(1):015003, March 2020.

[5] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, September 2021.

[6] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error Mitigation for Short-Depth Quantum Circuits. *Physical Review Letters*, 119(18):180509, November 2017.

[7] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical Quantum Error Mitigation for Near-Future Applications. *Physical Review X*, 8(3):031027, July 2018.

[8] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, 95(4):045005, December 2023.

[9] R. Cleve and J. Watrous. Fast parallel circuits for the quantum Fourier transform. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 526–536, November 2000.

[10] Emanuel Knill, Raymond Laflamme, and Wojciech H. Zurek. Resilient Quantum Computation. *Science*, 279(5349):342–345, January 1998.

[11] Dorit Aharonov and Michael Ben-Or. Fault-Tolerant Quantum Computation with Constant Error Rate. *SIAM Journal on Computing*, 38(4):1207–1282, January 2008.

[12] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The Complexity of NISQ, October 2022.

[13] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan. Limitations of Noisy Reversible Computation, November 1996.

[14] Michael J. Kastoryano and Fernando G. S. L. Brandão. Quantum Gibbs Samplers: The Commuting Case. *Communications in Mathematical Physics*, 344(3):915–957, June 2016.

[15] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12(1):6961, November 2021.

[16] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. Quantum Error Mitigation, October 2022.

[17] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. Fundamental limits of quantum error mitigation. *npj Quantum Information*, 8(1):114, September 2022.

[18] Ryuji Takagi, Hiroyasu Tajima, and Mile Gu. Universal Sampling Lower Bounds for Quantum Error Mitigation. *Physical Review Letters*, 131(21):210602, November 2023.

[19] Sergey Bravyi, David Gosset, and Ramis Movassagh. Classical algorithms for quantum mean values. *Nature Physics*, 17(3):337–341, March 2021.

[20] I. M. Georgescu, S. Ashhab, and Franco Nori. Quantum simulation. *Reviews of Modern Physics*, 86(1):153–185, March 2014.

[21] Xiao Yuan, Suguru Endo, Qi Zhao, Ying Li, and Simon C. Benjamin. Theory of variational quantum simulation. *Quantum*, 3:191, October 2019.

[22] Andrew J. Daley, Immanuel Bloch, Christian Kokail, Stuart Flannigan, Natalie Pearson, Matthias Troyer, and Peter Zoller. Practical quantum advantage in quantum simulation. *Nature*, 607(7920):667–676, July 2022.

[23] Manoj K. Joshi, Christian Kokail, Rick Van Bijnen, Florian Kranzl, Torsten V. Zache, Rainer Blatt, Christian F. Roos, and Peter Zoller. Exploring large-scale entanglement in quantum simulation. *Nature*, 624(7992):539–544, December 2023.

[24] Yichen Huang. Universal entanglement of mid-spectrum eigenstates of chaotic local Hamiltonians. *Nuclear Physics B*, 966:115373, May 2021.

[25] J. M. Deutsch. Thermodynamic entropy of a many-body energy eigenstate. *New Journal of Physics*, 12(7):075021, July 2010.

[26] Joshua M. Deutsch. Eigenstate thermalization hypothesis. *Reports on Progress in Physics*, 81(8):082001, July 2018.

Limitations of Noisy Quantum Devices in Computing and Entangling Power

Yuxuan Yan, Zhenyu Du, Junjie Chen, and Xiongfeng Ma*
Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, P. R. China

Finding solid and practical quantum advantages via noisy quantum devices without error correction is a critical but challenging problem. Conversely, comprehending the fundamental limitations of the state-of-the-art is equally crucial. In this work, we observe the polynomial-time indistinguishability of n-qubit devices from random coins when circuit depths exceed $\omega(\log(n))$ under single-qubit depolarizing noise. Even with classical processing, we can demonstrate the absence of computational advantage in polynomial-time algorithms with noisy quantum circuits of super-logarithmic depths. This finding decisively negates the feasibility of executing prominent quantum algorithms such as Shor's, Grover's, and the Harrow-Hassidim-Lloyd algorithm. In addition, our results apply to variational quantum algorithms, error mitigation, and quantum simulation with polynomial depths. Furthermore, we consider noisy quantum devices with restraint gate topology. We rule out super-polynomial quantum advantages for a one-dimensional noisy qubit array in all-depth regimes. We also establish upper limits on entanglement generation: $O(\log(n))$ for one-dimensional arrays and $O(\sqrt{n}\log(n))$ for two-dimensional arrays. Our findings underscore the entanglement scalability constraints in noisy quantum devices.

CONTENTS

* xma@tsinghua.edu.cn

Recent advancements in quantum computing have notably enhanced the scale and fidelity of quantum devices, surpassing classical brute-force simulation capabilities [1, 2]. Meanwhile, scalable quantum error correction remains out of reach, primarily due to high noise levels and the limited qubit count. Consequently, current quantum devices, ranging from dozens to several hundred noisy qubits, are situated between classical computing and fault-tolerant quantum computing, a regime known as Noisy Intermediate-Scale Quantum (NISQ) era [3]. Despite ongoing progress towards quantum error correction [4–11], transitioning from NISQ to fault-tolerant quantum computation presents a formidable challenge, which is expected to take years or even decades.

Within the NISQ paradigm, many important previous works have been devoted to seeking advantages over classical computers via noisy quantum devices without error correction. Here, "advantages" refer to the quantum device's capability to accelerate computational tasks beyond purely classical means. In this line of research, specific problems demonstrate theoretical quantum advantages of noisy devices in shallow circuit regimes [12] or when oracles are introduced [13]. Experimentally, noisy quantum circuit sampling has also challenged the ability of most powerful classical supercomputers [1, 2]. Yet, in more practical applications, the noise in quantum devices often undermines potential advantages. Notably, as important practical advantage candidates, variational quantum algorithms exhibit fragility to noise [14, 15]. These findings raise an important question: what is the limit of the noisy device's power in the NISQ era?

When assessing the power of NISQ devices, classical computers should be considered due to their ability to assist noisy quantum devices. When viewed in itself, a noisy quantum device experiences rapid loss of information due to noise [16, 17]. However, quantum algorithms can use classical inputs to control quantum devices and process measurement outcomes, thus enhancing noisy quantum devices. Classical processing is particularly crucial in the NISQ era as it mitigates noise and amplifies the capabilities of quantum hardware [18, 19]. The problem of assessing NISQ advantages, with classical computer enhancements considered, is complicated and requires systematic understanding.

The classical simulation complexity of noisy quantum devices is also a crucial factor in NISQ advantages, as easier simulation often implies weaker advantages. Existing work has developed algorithms [20–25] much more efficient than the brute-force way that calculates the state vector. Theoretical classical simulatability can be proven under certain conditions, such as having small underlying graph treewidth [20] or being in the anti-concentration regime under noise [25]. Under these conditions, polynomial-time classical algorithms exist, thereby excluding super-polynomial quantum advantages. More generally, the advantages of noisy quantum devices remain an open question.

In this work, we establish a clear boundary for the limitations of noisy quantum devices. Our analysis is based on a model where devices are affected by independent single-qubit depolarizing noise without error correction. We show the statistical indistinguishability of noisy quantum device outputs from a uniform distribution when circuit depth exceeds the logarithmic of the running time. Incorporating any form of classical processing and controls, we show that devices with super-logarithmic circuit depths $\omega(\log(n))$ fail to deliver any quantum advantage for polynomial-time quantum algorithms. Famous no-go examples include Shor's [26], Grover's [27], and the Harrow-Hassidim-Lloyd algorithm [28].

On quantum computing platforms like superconducting qubits, gates are restrained in specific topologies [1, 2], further challenging potential quantum advantages. We demonstrate that noisy quantum circuits for one-dimensional qubit arrays are simulatable with any depth and thereby rule out super-polynomial advantages. To further characterize the physical limitations, we investigate the maximal entanglement generation in noisy quantum devices. Our findings indicate that for a one-dimensional qubit array, the capacity to generate quantum entanglement is capped at $O(\log(n))$. This boundary rules out the efficient creation of highly entangled states, such as highly excited or thermalized states of almost all quantum systems. Our findings are summarized in Fig. 1.

I. RESULTS

A. Model of noisy quantum devices

In the NISQ era, we represent operations on noisy devices as a sequence of unitary gates interspersed with single-qubit depolarizing noise that acts independently. The depolarizing noise impacts each qubit as

$$\Lambda_1(\rho) = (1-p)\rho + p\frac{I}{2}. \tag{1}$$

After gate operations with associated noise, computational-basis measurements yield the output. The noisy device model is formally defined as follows, with a visual representation in Fig. 2(a).

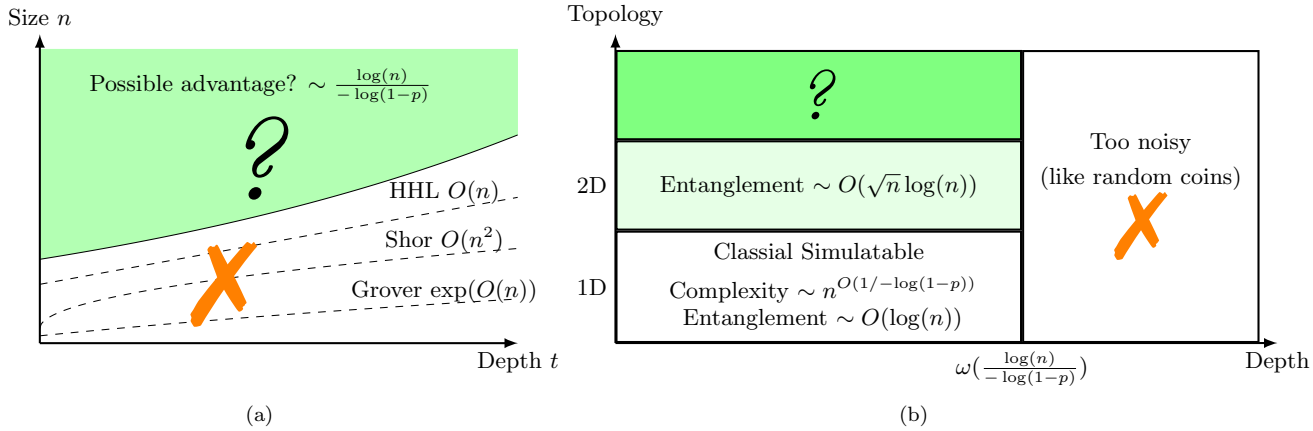


FIG. 1. **Summary of the limitations for noisy quantum devices without error correction. a** For algorithms with generic classical processing control, we prove that devices with circuit depths beyond $\omega(\log(n))$ under single-qubit depolarizing noise are too noisy to offer any computational advantage in a polynomial running time, including well-known quantum algorithms, such as Shor's, Grover's, and the Harrow-Hassidim-Lloyd algorithm. The scaling of depth with the qubit number, n, is shown by dashed curves for each algorithm, with a solid curve showing the logarithmic upper limit. The scaling is shown in an asymptotical limit, i.e., when n is large. **b** In the regime where the circuit depth is below the logarithmic scaling, potential quantum advantages depend on the gate connection topology of noisy quantum devices. For the one-dimensional case, we prove classical simulatability for noisy devices and give an entanglement upper bound of $O(\log(n))$. For the two-dimensional case, the entanglement generation upper bound scales as $O(\sqrt{n}\log(n))$. Therefore, super-polynomial advantages without error correction are only possible when gate connectivity is higher than one dimension, and the circuit depth is below logarithmic scaling for noisy devices. Such a regime is colored in green.

Definition 1 (Noisy quantum devices)**.** *A noisy quantum device of n qubits produces a quantum state at a depth of t,*

$$\rho(t) = \Lambda \circ \mathcal{U}_t \circ \Lambda \circ \cdots \circ \Lambda \circ \mathcal{U}_2 \circ \Lambda \circ \mathcal{U}_1(|0\rangle\langle 0|^{\otimes n}), \tag{2}$$

where $\mathcal{U}_i$ denotes a gate layer and $\Lambda = \Lambda_1^{\otimes n}$ represents the noise channel. Each layer allows at most one gate operation per qubit. The measurement output $C_n(\rho(t))$ will follow the distribution

$$\Pr[C_n(\rho(t)) = X] = \langle X|\, \rho(t)\, |X\rangle\,, \tag{3}$$

with X being an n-bit string and $|X\rangle$ its computational basis state.

Our NISQ model prohibits using mid-circuit measurements and refreshing qubits. Such restrictions imply that the system's entropy cannot be reduced and will rise following each noise layer. We make the assumption intentionally for the NISQ era, as allowing these operations could facilitate fault-tolerant quantum computing below a noise threshold [29, 30], a capability beyond the scope of the NISQ era [13].

As a result, all stored information will be inevitably lost due to the transition to the maximally mixed state with growing noisy circuit depth, regardless of the noise strength p. This convergence to the maximal mixed state is exponentially rapid as a function of circuit depth t. Quantitatively, the relative entropy between the state $\rho(t)$ and the maximally mixed state diminishes as

$$D(\rho(t)\|\sigma_0) \leq n(1-p)^{2t}, \tag{4}$$

where $\sigma_0 = I/2^n$ is the maximally mixed state [16, 17]. The relative entropy $D(\rho(t)\|\sigma_0)$ can also be seen as the residual knowledge within the state $\rho(t)$, since $D(\rho(t)\|\sigma_0) = n - S(\rho(t))$, with $S(\rho(t))$ being the von Neumann entropy of $\rho(t)$. In addition to the qubit case, we derive similar results for qudit systems and provide proofs in Supplementary Information [31].

Importantly, apart from our choice of single-qubit depolarizing channels, many other types of noise also have exponential decaying behaviors with decay rates in different forms [32–37]. Thus, the results of this work can be easily extended to those cases, with quantitive results straightforwardly modified according to the different decay rates.

B. Entropy analysis for a hybrid algorithm

As highlighted above, this work focuses on quantum algorithms with the assistance of classical processing. We use the term "hybrid algorithms" to emphasize the integration of quantum and classical processing. Within a hybrid algorithm, a classical computer calls noisy quantum devices for measurement outcome bits. In each query, the quantum device responds by returning one bit of measurement results to the classical computer. Here, the classical computer is assumed to be noiseless with persistent memory. This process is depicted in Fig. 2(b). Following the intuition, our Methods section will provide strict formulations using a probabilistic Turing machine.

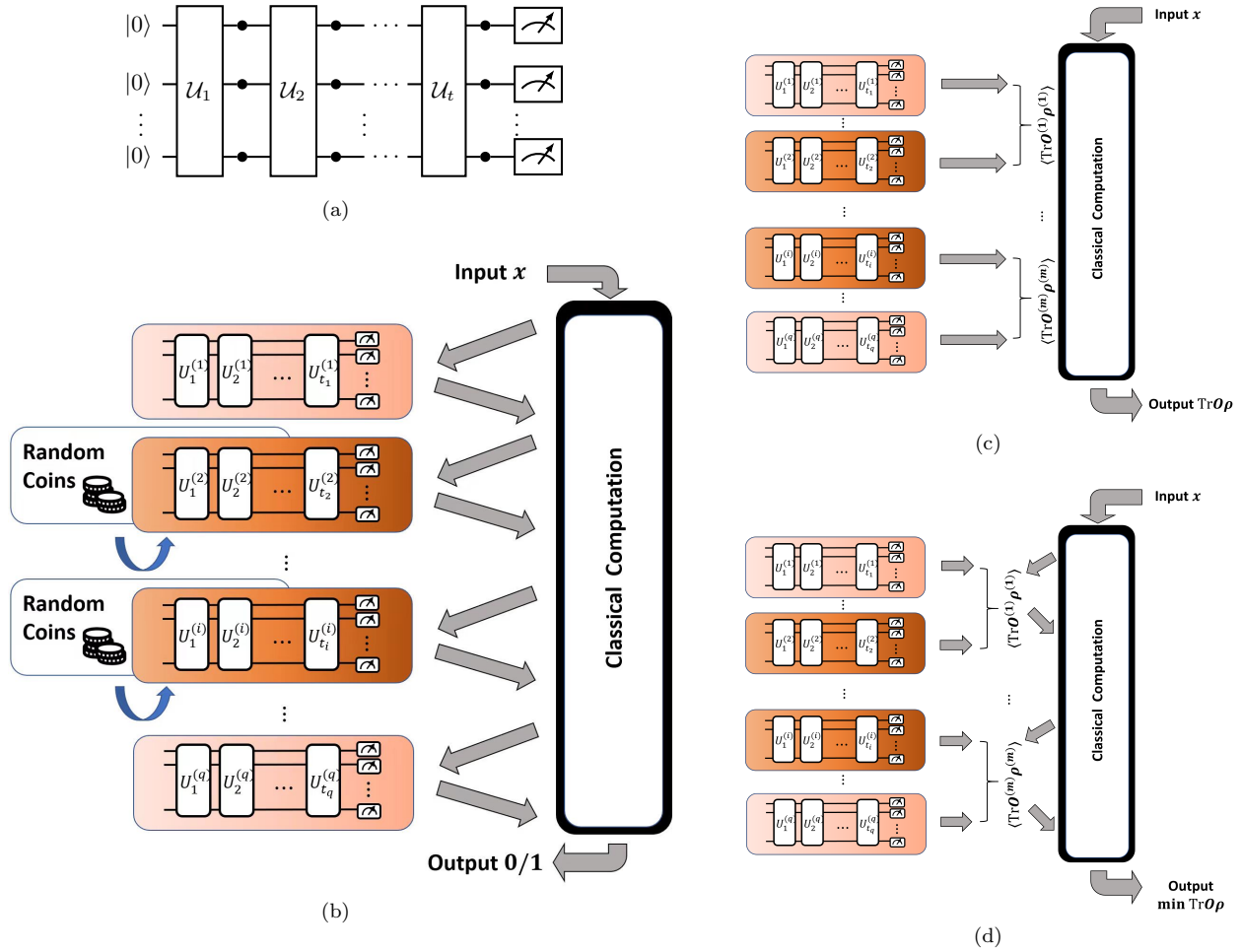


FIG. 2. **Illustration of a hybrid algorithm with noisy quantum devices and the replacement process from noisy quantum devices to random coins. a** Noisy quantum devices defined in Definition 1. Blocks of $\mathcal{U}_i$'s are gate layers, and black dots are single-qubit depolarizing noise channel Λ_1. Computational basis measurements are performed at the end of the circuit to provide classical information output. **b** In a hybrid algorithm, the orange boxes on the left represent noisy quantum devices, in which a darker color represents noisy quantum devices with super-logarithmic depths, $\omega(\log(nq))$. The classical computer on the right can control noisy quantum devices based on previous measurement outputs. After obtaining data from quantum devices and classical computation, the classical computer outputs 0 or 1. We use the boxes with coins to show that random coins can replace noisy quantum devices with super-logarithmic depths. After the replacement, the new classical algorithm can still give the same output bit. Thus, the noisy quantum devices with super-logarithmic depth do not provide any quantum advantages, as stated in Theorem 1. **c** Our results apply to single-copy quantum error mitigation, designed to suppress errors in expectation values, usually albeit with increased running time when single copies are fed. **d** Our results also apply to variational quantum algorithms that adapt circuits to update parameters during optimization. Our results suggest their limitations regarding noisy circuit depth, independent of classical processing designs.

Under our hybrid computing framework, we analyze the relation between the entropy of measurement results from all correlated queries $(X_1, X_2, \cdots, X_q)$ and the circuit depth t. Each query of a noisy quantum device produces outcome X_i from computational measurements on the final gate layer, with exponentially diminishing information regarding the

circuit depth t [16, 17]. However, extending this to multiple queries within a hybrid algorithm is not straightforward and necessitates carefully considering the correlations among the different queries. These correlations emerge from classical computers that can control subsequent quantum device operations based on previous measurements. A naive entropy analysis might suggest that such correlations could reduce the total entropy, potentially enabling hybrid algorithms to aggregate information and amplify quantum advantages.

Contrary to this presumption, our theoretical analysis proves that even if we consider query correlations, the aggregate information obtained from all measurement outputs $(X_1, X_2, \ldots, X_q)$ becomes exponentially small with increasing circuit depth t, as stated in the following lemma. We break down the total entropy into a sequential sum of conditional entropies based on preceding queries to obtain the result. The detailed proof is provided in Supplementary Information [31].

Lemma 1. *Suppose the hybrid algorithm calls for q times of circuit execution with a minimum circuit depth t. Denote X_i as the measurement output from the i-th circuit execution on noisy n-qubit devices. The collection of outcomes $X_1, \ldots, X_q$, containing nq queried measurement bits, yields*

$$S(X_1, \ldots, X_q) \geq \left(1 - (1-p)^{2t}\right) nq, \tag{5}$$

where p denotes the depolarizing channel's magnitude as specified in Eq. (1).

This lemma considers correlations among queries to quantum devices, which arise from arbitrary classical processing and controls, thereby going beyond isolated quantum device analysis in existing studies [16, 17]. Furthermore, this lemma makes our results applicable to generic hybrid algorithms.

C. Limitations of depths for computational advantages

Based on the above entropy analysis, we have identified the limitations of circuit depth for hybrid algorithms that use noisy quantum devices. Our study reveals that the maximum depth limit to provide advantages scales as the logarithm of the number of queries to the noisy quantum devices, nq. We clarify this argument with the following theorem. The formal version of the theorem is available in the Methods section.

Theorem 1 (informal version)**.** *Consider a hybrid algorithm operating within time $T(n)$ with nq queried measurement bits from noisy quantum devices. If circuit depth $t \geq \omega(\frac{\log(nq)}{-\log(1-p)})$, noisy devices will yield no quantum advantages. In this case, a classical algorithm, running within $T(n)$ time, can be used instead by replacing the noisy device queries with random coins from a uniform distribution.*

The running time of an algorithm can be broken down into three parts: quantum circuit execution, queries to measurement outcomes, and classical processing. We can represent this as the following equation:

$$T(n) = c_1 tq + c_2 nq + T_c(n). \tag{6}$$

Here, t represents the circuit depth, q represents the number of circuit executions, $T_c(n)$ represents classical processing time, and c_1 and c_2 are the constant time for executing a gate layer and querying a measurement outcome bit respectively. We use functions of the qubit number n to represent $T(n)$ and $T_c(n)$ as they typically depend on the number of qubits n, representing algorithm input size. We assume that one query returns one bit of measurement outcomes, taking c_2 time for the classical computer. Therefore, we always have $nq \leq \frac{1}{c_2} T(n)$ from the above equation.

In previous studies on isolated noisy devices [16, 33], the maximal circuit depth depends on the number of qubits. In contrast, we highlight query times, i.e., nq, and the running time $T(n)$ as an important factor in hybrid quantum algorithms. The difference is attributed to the role of classical computing in hybrid algorithms. For the same noisy device, the longer the duration of a hybrid algorithm, the greater the opportunities it has to query the device and receive quantum advantages via the assistance of classical processing, e.g., error mitigation protocols [18, 38, 39]. Following this intuition, we present the proof in the Supplementary Information [31].

The running time of efficient quantum algorithms is required to be at most polynomial growth with respect to the number of qubits. In this case, where $nq \leq \frac{1}{c_2} T(n) \leq O(\text{poly}(n))$, noisy quantum devices with super-logarithmic depth cannot provide any advantage for such polynomial-time quantum algorithms. This limitation is unavoidable via the sophistication of classical processing or adaptive operations based on measurement outcomes in previous queries. As a result, we have established strict no-go results on algorithms that necessitate a super-logarithmic circuit depth [1], such as Shor's [26], Grover's [27], and the Harrow-Hassidim-Lloyd (HHL) algorithm [28]. In quantum simulation,

[1] We should note that we do not consider implementations that reduce the depth of the algorithm at the cost of significantly increasing the number of qubits required. For instance, a modified version of Shor's algorithm that factors an n-bit number in $O(\log(n))$ depth but demands $O(n^5)$ qubits [40] falls under this category and hence, is excluded.

our findings establish an upper bound of $O(\log(n))$ for the allowable evolution time. The reason behind this is that, generally, simulating a quantum system with an evolution time of τ requires a circuit depth proportional to τ, which is known as the no-fast-forwarding theorem [41–43]. The examples mentioned are summarized in Table I.

TABLE I. Quantum algorithms that are shown to offer no advantage on noisy devices, where necessary depths exceed the super-logarithmic scale in qubits, n. Listed depths for Grover's algorithm and quantum simulations are theoretically optimal [44], while those for Shor's and HHL represent the most efficient known configurations without significant qubit overhead.

Algorithm	Depth	Advantages
Shor's algorithm [26]	$O(n^2)$ [45, 46]	No
Grover's algorithm [27]	$O(\exp(\frac{n}{2}))$	No
HHL algorithm [28]	$O(n)$	No
Quantum simulation	$O(\tau),\ \tau < O(\log(n))$	No

Our findings apply to NISQ algorithms, where the running time and circuit depths are usually variable. In particular, we consider the algorithms that calculate the expectation values of observables, such as in variational quantum algorithms [47, 48] and quantum error mitigation [18, 38, 39]. To estimate m expectation values, we collect Q_i computational-basis measurement outcomes for the i-th. Then, we use classical processing to estimate the statistical value of $f^{(i)} = \mathrm{Tr}\left[\hat{O}^{(i)}\rho^{(i)}\right]$. Here, $\hat{O}^{(i)}$ represents the i-th observable, and $\rho^{(i)}$ corresponds to the quantum state associated with it. Note that $\rho^{(i)}$'s may differ from each other, which are generated by various noisy circuits. Specifically, the structures of single-copy error mitigation and variational quantum algorithms are illustrated in Fig. 2(c) and 2(d), respectively, which are special cases of our hybrid computing framework in Definition 1. Therefore, we can apply Theorem 1 and derive the following corollary.

Corollary 1. *For variational quantum algorithms and single-copy quantum error mitigation algorithms, if the circuit depth of noisy quantum devices is t, then the algorithm must have running time $T(n) = \Omega(2^{|\log(1-p)|t})$ to provide quantum advantage exists.*

The corollary strengthens and unifies important findings in NISQ algorithm limitations. For variational quantum algorithms, our results imply the issue of noise-induced barren plateaus. This issue arises at linear circuit depth due to noise, causing the gradients to exponentially vanish as the number of qubits n increases and the optimization to fail [37]. For quantum error mitigation, our results imply the exponential sampling costs associated with single-copy mitigation schemes [49, 50]. For both aforementioned applications, we strengthen the previous results by considering the general dependence of circuits on previous measurements and classical processing, which may even go beyond the diagram in Fig. 2(c) and 2(d). Note that a higher sampling overhead for error mitigation has been demonstrated in a different noisy circuit model, where multiple layers of gates are executed between local depolarizing channels [51].

The implications of our research further extend to sampling algorithms. Following the idea of Theorem 1, we show that the $\omega(\log(nq))$ samples from noisy devices are statistically indistinguishable from those of uniformly distributed random coins if circuit depth exceeds $\omega(\log(nq))$. When the running time $T(n)$ is within $\omega(\mathrm{poly}(n))$, it suggests that sampling advantages cannot be demonstrated in polynomial time for noisy circuit depth exceeding $\omega(\log(n))$. The details are available in the Method section.

D. Limitations of computational advantages in one-dimensional local circuits

In real-world quantum devices, gates are often subjected to certain topologies, resulting in deeper circuits and further restraining their computational power under noise. For example, we can consider a one-dimensional qubit array, where gates are restrained between the nearest neighbor qubits on a linear chain.

We show that one-dimensional noisy devices do not possess any super-polynomial computational advantages, regardless of their depth. Our finding is based on a state-vector simulator designed for constant-depth one-dimensional circuits [52]. We have demonstrated that the simulator can be extended to one-dimensional noisy devices with a depth of $O(\log(n))$, thus ruling out the possibility of super-polynomial advantages. Further details are available in the Methods section.

Lemma 2. *The output distribution of one-dimensional quantum devices, without noise or with any single-qubit noise, can be sampled with $O(2^{2t}nt)$ computational times on a classical computer, where n is the qubit number, and t is circuit depth.*

Note that this lemma applies to both noiseless and noisy cases. In the noisy case, noise models can be chosen as any single-qubit noise, even beyond the single-qubit depolarizing model discussed in the sections above.

Suppose we assume depolarizing noise, as in Definition 1, super-polynomial advantages will be excluded in all depth regimes. To see this, we consider two depth regimes divided by $O(\log(n))$, respectively. Below $O(\log(n))$, Lemma 2 exclues super-polynomial advantages. Otherwise, if the circuit depth $t = \omega(\log(n))$, Theorem 1 suggests the absence of quantum advantages. Combining the two aspects, we exclude the super-polynomial advantages of one-dimensional noisy quantum devices, regardless of circuit depth, under the singe-qubit depolarizing noise model.

Theorem 2. *One-dimensional noisy quantum devices that run in computational time $T(n)$ can be simulated by a classical algorithm with $T(n)^{1+\frac{1}{-\log(1-p)}}$ computational time, thus having no super-polynomial quantum advantages.*

Our results stress the importance of connectivity in the NISQ diagram. With depolarizing noise and without error correction, super-polynomial advantages require stronger connectivity than a one-dimensional qubit array.

E. Limitations of entanglement generation

In the above sections, we have shown the computational limitations of noisy quantum devices. We aim to investigate further how noise physically impacts quantum devices with gate locality constraints. To this end, we analyze the entangling power in one-dimensional and two-dimensional qubit arrays.

For the one-dimensional qubit chain, we consider bipartite entanglement between two contiguous halves of the chain, denoted as A and $\bar{A}$. A key observation is that the interaction of a qubit is localized in a region whose radius grows with the depth t. The physical picture is that the qubits interact within a light cone. We generalize this observation to entanglement spreading and derive an upper bound of the entanglement monotone E between halves of the chain,

$$E(A:\bar{A}) \leq t. \tag{7}$$

The generation of entanglement requires sufficient circuit depth. Similar bounds have been derived for the dynamics of pure states without noise [53, 54]. Using local operation monotones and induction, we extend the entangling upper bound to the mixed-state case. This result is essential for analyzing noisy quantum devices.

On the other hand, noisy quantum devices suffer from an exponential loss of information with increasing depth t, which also leads to exponentially rapid decay of maximal entanglement. Jointly, the two effects result in a logarithmic upper bound of entanglement generation in noisy quantum devices at arbitrary circuit depths, as stated in the following theorem. The proof of our results, including the following theorem, can be found in Supplementary Information [31].

Theorem 3 (Entanglement upper bound on one dimension array)**.** *For a contiguous half A and the complement half $\bar{A}$ in an n-qubit noisy quantum device with a one-dimensional connection topology, the quantum mutual information and hence the quantum relative entropy of entanglement are upper bounded by*

$$E_R(A:\bar{A}) \leq I(A:\bar{A}) \leq \frac{\log(n)}{-2\log(1-p)}, \tag{8}$$

where p is the noise strength defined in Eq. (1).

Our findings have significant implications for preparing quantum states with large-scale entanglement by excluding the possibility of efficiently preparing any quantum state with a super-logarithmic entanglement scaling. This limitation extends to a wide range of scenarios, including high excitation [55] and thermalization in most quantum dynamics [56, 57]. This is also related to limitations in quantum simulation, where entanglement plays an important role [58]. We also investigate entanglement between distant regions in qubit chains and show that the upper bound of entanglement decays to a constant $\frac{4(1-p)}{p(2-p)}$, exponentially with the distance between the two regions.

For two-dimensional lattices, we consider qubits arranged in a square of side length $\sqrt{n}$ and show that maximal entanglement is $O(\sqrt{n}\log(n))$. For one- and two- dimensional lattices, we present the numerical upper bounds of entanglement for different values of noise strength p in Fig. 3. After the number of qubits reaches a certain number related to the noise strength, the further growth of quantum entanglement in the noisy quantum device will be suppressed. For the one-dimensional case, this will lead to an exponential cost of qubits required to scale up the system's entanglement further due to the logarithmic scaling of the upper bounds of entanglement. In the two-dimensional case, a polynomial cost is also required.

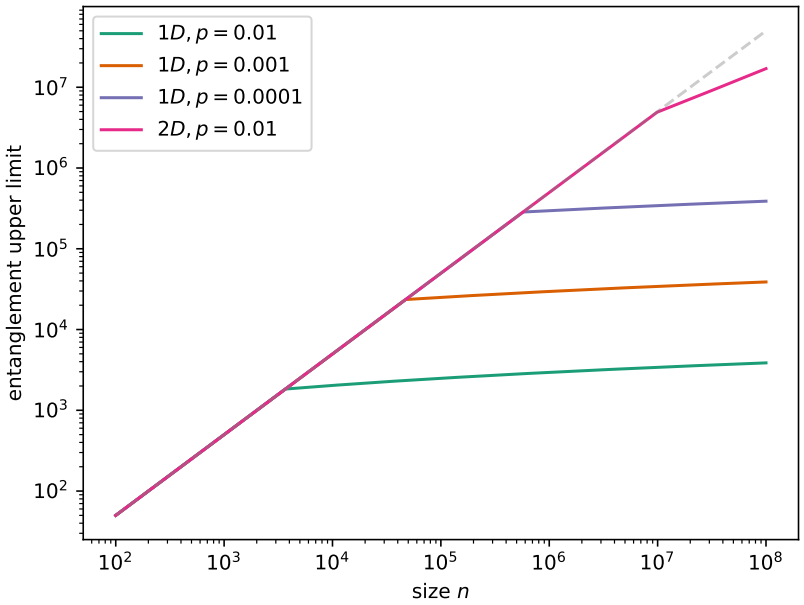


FIG. 3. Limitations of entanglement generation, with varied number of qubits n and noise strength p, which are taken as the minimal values between $n/2$ and the upper bounds in Theorem 3. Lines in different colors correspond to different noise strengths p, and one of the lines is for the case of a two-dimensional qubit connection.

II. DISCUSSION

The limitations on computational and entangling capabilities are inherent in our NISQ model. Future advancements in quantum computing hardware must strive to transcend the assumptions of this model and, therefore, be able to avoid the rapid convergence to the maximally mixed state. In the interim, strategies such as delaying the introduction of qubits into a noisy environment until the last possible moment or transforming the noise into more manageable forms may mitigate some of the issues [59]. The ultimate solution to quantum errors can be achieved through resetting qubits or using mid-circuit measurements with feedforward actions to purify the system. These techniques are important steps toward quantum error correction, but they require a sufficient supply of qubits and low error rates below the error tolerance threshold. Therefore, exceeding the limitations outlined in our study is necessary to achieve fault-tolerant quantum computation during the current NISQ era.

Future work will involve expanding our results to other types of noise. In this work, we adopt single-qubit depolarizing channels as an idealized noise model. As mentioned earlier, our results also apply to many other noise channels [32–37]. For Theorem 1-3 to stay valid, noise models will be required to result in exponentially fast convergence of relative entropy, as depolarizing noise does in Eq. (4). Yet, the situation will be very different for the noise that does not drive the state to the maximally mixed state, such as dephasing channels and the amplitude damping channel [60]. For these noise models, as long as they are single-qubit noise, Lemma 2 will still hold for any one-dimensional qubit array with noise, excluding superpolynomial advantages. For higher dimensional qubit arrays, these noise models need further investigation.

Our methods may also help establish limitations for other properties of noisy quantum devices, including quantum state complexity, topological order, magic, and quantum chaos. It will also be interesting to apply our results to more quantum computer applications.

III. METHODS

A. Formalism of hybrid quantum algorithms

Here, we present the formalism of hybrid quantum algorithms, which combine quantum computing with classical processing, as illustrated in Fig. 2(b). The scope of algorithms we discuss includes decision problems and sampling problems. We adopt the notion of languages and the probabilistic Turing machine (PTM), which are standard terms in theoretical computer science. For decision problems, the term *language* refers to a general problem where a determined answer is required as an answer to the problem.

Definition 2 (Language)**.** *A language L is a subset of $\{0,1\}^*$. For $x \in \{0,1\}^*$, $L(x)$ is defined as $L(x) = [x \in L]$.*

A probabilistic Turing machine (PTM) is a classical algorithm that can compute a language with a probability of giving the correct answer greater than $\frac{2}{3}$. It is important to note that the classical algorithm may be randomized. Regarding the randomness in algorithms, a random variable Y is introduced to represent the choice of Turing machines.

Definition 3 (Probabilistic Turing machine)**.** *A probabilistic Turing machine, denoted by M, decides a language L in time $T(n)$ if, for any string x, M halts within $T(|x|)$ steps and the probability of M outputting the correct answer for x is at least $\frac{2}{3}$ when given a random string Y.*

Here, Y is a random choice of Turing machines, which follows a uniform distribution over bit strings of length $T(|x|)$. When the input x and random choices Y are given, $M(x, Y)$ is the output of the chosen Turing machine.

Hybrid algorithms can interact with noisy quantum devices, as shown in Fig. 2(b). We formulate a hybrid algorithm as a PTM that queries noisy quantum devices, receives measurement outcomes in the form of bit strings, and performs classical processing on these outcomes. As mentioned in the main text, $T(n)$ denotes its running time, t denotes its circuit depth, and q denotes the times of circuit execution. Based on this formalism, we analyze the limitations of hybrid algorithms with noisy quantum devices, leading to Theorem 1, as formally stated below.

Theorem 1 (formal version)**.** *Consider a hybrid algorithm $\mathcal{A}$ that decides a decision problem L with nq queried measurement outcome bits in running time $T(n)$. A classical algorithm M exists that decides L in time $O(T(n))$ if $t \geq \frac{1}{2|\log(1-p)|}(\log(nq) + 5)$, where p is the depolarizing noise strength in Eq. (1). Here, M can be constructed by replacing noisy quantum devices with random coins from a uniform distribution.*

The intuition for the proof follows Lemma 1. In Supplementary Information [31], we provide detailed proof for the theorem and show how to establish equivalence between decision problems defined in Definition 2 and algorithms mentioned in the main text, such as Shor's algorithm.

Furthermore, we extend our result to quantum sampling, as demonstrated in experiments [1, 2]. In sampling problems, the quest is to obtain nq measurement output bits, i.e., samples, from noisy quantum devices. Here, we consider a distinguished who tries to tell whether samples are from noisy quantum devices or a uniform distribution. We show that such effort will fail if circuit depths exceed the same depth upper limits in Theorem 1. Formally, we have the following theorem, with proof available in Supplementary Information [31].

Theorem 4. *Consider nq samples generated from noisy quantum devices. If circuit depth $t \geq \frac{1}{2|\log(1-p)|}(\log(nq)+5)$, then the obtained samples are statistically indistinguishable from those from a uniform distribution. Namely, any distinguisher cannot tell the difference.*

Here, the distinguishment is formalized as a decision problem. By Definition 3, this requires the distinguisher to succeed with probability at least $\frac{2}{3}$. Our results highlight the difficulty of distinguishing between the outcomes of deep, noisy quantum circuits and random coin flips within a reasonable time. Our findings imply that any attempt to demonstrate a sampling advantage using noisy quantum devices with super-logarithmic circuit depth would require a super-polynomial number of samples nq, which implies a super-polynomial running time for the sampling algorithm.

B. Simulatability of one-dimensional qubit array with noise

We use the algorithm proposed in [52] to sample from a one-dimensional local circuit. We then analyze the complexity of the classical algorithm, proving Lemma 2 and Theorem 2.

Consider a one-dimensional noisy circuit with a depth of t. The intuition is that the light cone of each site only contains at most $2t$ qubits. Thus, we can calculate the conditional distribution by straightforwardly maintaining the

state of $O(t)$ qubits on a classical computer, which consumes $2^{O(t)}$ computational time. The entire one-dimensional noisy circuit as a channel is decomposed as

$$\mathcal{U} = \Lambda \circ \mathcal{U}_t \circ \Lambda \circ \cdots \circ \Lambda \circ \mathcal{U}_2 \circ \Lambda \circ \mathcal{U}_1, \tag{9}$$

where Λ and $\mathcal{U}_i$ are noise channel and unitary gate layers appeared in Eq. (2), respectively.

Consider the unitaries and noise channels within the light cone of the i-th qubit, denoted as $\mathcal{L}_i$ and shown in Fig. 4(a). Specifically, $\mathcal{L}_i$ is obtained by removing the gates and noise channel in $\mathcal{U}$ that do not impact the expectation value of any local observable O_i on the site i. For any observable O_i and density matrix ρ, we have $\mathrm{Tr}[\mathcal{U}(\rho)O_i] = \mathrm{Tr}[\mathcal{L}_i(\rho)O_i]$. Note that $\mathcal{L}_i$ only act non-trivially on $O(t)$ qubits.

To sample from each qubit, we should consider all gates and noise within its light cone. Conditioned on previous measurement outcomes, we introduce the effective channel for each site, denoted as $\mathcal{V}_i$, following these steps: Let $\mathcal{V}_1 = \mathcal{L}_1$. For $i > 1$, define $\mathcal{V}_i$ by removing the overlapped gates and noisy channels in $\mathcal{L}_i$ and $\mathcal{L}_{i-1}$ from $\mathcal{L}_i$, as illustrated in Fig. 4(b). This leads to another decomposition of the circuit channel in Eq. (9),

$$\mathcal{U} = \mathcal{V}_n \circ \mathcal{V}_{n-1} \circ \cdots \circ \mathcal{V}_1. \tag{10}$$

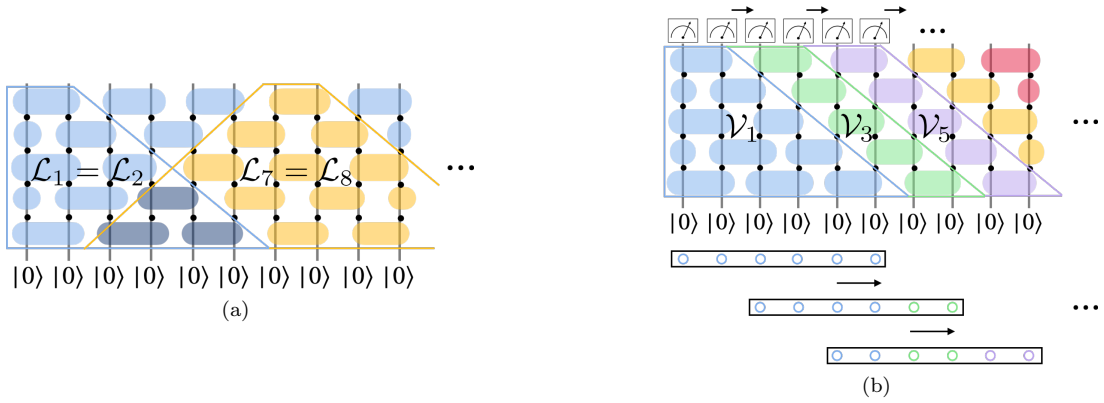


FIG. 4. **Illustration of the light cone and effective channels for each site and the sampling algorithm. a** For each site, only quantum gates and noise channels within the light cone of the i-th qubit impact the expectation value of local observables at site i. The number of qubits in each light cone is at most $2t$. **b** Due to the limited range of light cones, we can define an effective channel $\mathcal{V}_i$ on each site and the dynamic sweeping process in the algorithm. In the classical computer, we store the density matrix $t+1$ qubits, where t represents circuit depth. The procedure involves sweeping over each site to perform a measurement and subsequently discarding the measured qubit. Following this, we introduce the next qubit into the classical computer, applying effective channels on the classical memory as per Eq. (12). This iterative process continues as we alternatively calculate the conditional probability on each site and sample accordingly. After finishing the whole process, we will obtain a sample string $\mathbf{x}$ from $p(\mathbf{x}) = |\langle x|U|0\rangle^{\otimes n}|^2$. The algorithm operates with a time complexity of $2^{O(t)}$.

Based on the notion of effective channels, we show the sampling procedure for $p(\mathbf{x})$ by successively sampling from the conditional distribution $p(x_i|x_1 = a_1, \ldots, x_{i-1} = a_{i-1})$. For convenience, we define $[n] = \{1, 2, \cdots, n\}$, $M_j = |a_j\rangle\langle a_j| \otimes I_{[n]-j}$ for $a_j \in \{0, 1\}$. When measuring the j-th qubit of ρ, the probability of obtaining result a_j is $\mathrm{Tr}[M_j\rho]$, and the post-measurement state is $S_j(\rho) = \frac{M_j\rho M_j}{\mathrm{Tr}[M_j\rho]}$. With previous measurement outcomes fixed, the conditional probability can be written as

$$\begin{aligned} &p(x_i = 0|x_1 = a_1, \cdots, x_{i-1} = a_{i-1}) \\ &= \mathrm{Tr}[|0\rangle\langle 0|_i \, S_{i-1} \circ S_{i-2} \circ \cdots \circ S_1 \circ \mathcal{U}(\rho_0)] \\ &= \mathrm{Tr}[|0\rangle\langle 0|_i \, S_{i-1} \circ S_{i-2} \circ \cdots S_1 \circ \mathcal{V}_n \circ \mathcal{V}_{n-1} \circ \cdots \circ \mathcal{V}_1(\rho_0)], \end{aligned} \tag{11}$$

where $\rho_0 = |0\rangle\langle 0|^{\otimes n}$ is the initial state. We observe that S_i and $\mathcal{V}_{i+1}$ commute since they act on different qubits. Therefore, we re-arrange the operators in the equation above so that $\mathcal{V}_i$ and S_i are applied alternatively,

$$\begin{aligned} &p(x_i = 0|x_1 = a_1, \cdots, x_{i-1} = a_{i-1}) \\ &= \mathrm{Tr}[|0\rangle\langle 0|_i \, \mathcal{V}_i \circ S_{i-1} \circ \mathcal{V}_{i-1} \circ S_{i-2} \circ \mathcal{V}_{i-2} \circ \cdots S_1 \circ \mathcal{V}_1(\rho_0)]. \end{aligned} \tag{12}$$

The conditional probability in Eq. (12) can be estimated by simulating $O(t)$ qubits in a classical computer. Specifically, $\mathcal{V}_1(\rho_0)$ is supported on $(t+1)$ qubits, necessitating $2^{O(t)}$ time to compute the density matrix. As we have a

complete description of $\mathcal{V}_1(\rho_0)$ in the classical computer, we can calculate both $p(x_1 = a_1)$ and the post-measurement state $S_1(\mathcal{V}(\rho_0))$. The subsequent measurement process for the second qubit follows a similar procedure. Subsequently, we introduce new qubits in $\mathcal{V}_3$ and apply S_3 to measure the third qubit. Similarly, following Eq. (12), we can alternate between applying $\mathcal{V}_i$ and processing S_i, calculate the conditional probability, and sample from each qubit accordingly. We have outlined the algorithm in Box 1, which is also depicted in Fig. 4(b) for better understanding.

Box 1: Classical Algorithm for Simulating One-Dimensional Noisy Quantum Devices

1. Calculate the reduced density matrix of the first $t+1$ qubits, denoted as ρ.
2. Apply the effective channel, namely the unitaries and noisy channels that will affect the marginal probability of the first qubit, to ρ.
3. Sample according to the marginal probability of the first qubit.
4. Update ρ to the post-measurement state and remove the first qubit. After removal, the next qubit will be assigned as the first qubit.
5. Unless all qubits are traced out, add the qubits that will affect the marginal probability of the first qubit in ρ and go to Step 2.

We examine the space and time complexity as follows: The algorithm requires simulating $(t+1)$ qubits simultaneously, which takes up $O(2^{2t})$ space complexity on a classical computer. On $(t+1)$ qubits, simulating each two-qubit gate, each single-qubit noise channel, or each measurement takes up to $O(2^{2t})$ time. Given that the total number of gates with noise is within $O(nt)$, the time complexity to calculate all conditional probabilities, given previous sampling outcomes, is $O(2^{2t}nt)$. Our algorithm operates in $O(2^{2t}nt)$ computational time, with space complexity $O(2^{2t})$, thereby validating Lemma 2.

Note that our analysis and the resulting Lemma 2 is not restricted to any specific type of noise channel. Our results apply to noisy quantum devices with $O(\log(n))$ depth under various forms of local noise, including the noiseless case, in contrast to previous works [13, 15, 16]. When focusing on the single-qubit depolarizing noise channel, we can then prove Theorem 2 by combining Theorem 1 with Lemma 2 and substituting t in the latter with $\frac{1}{2|\log(1-p)|}(\log(T(n))+5)$.

Note added.—After posting our work on arXiv, we realized that [61] had also obtained an upper bound on entanglement growth regarding noisy circuit depth, similar to our Eq. (7), in their Lemma 14 using a different approach.

ACKNOWLEDGMENTS

The authors acknowledge Dorit Aharonov and Michael Ben-Or for their helpful discussion about the convergence of quantum states. The authors acknowledge Ryuji Takagi for his helpful discussion about the convergence bound in qubit systems [17] and about error mitigation [49, 50]. The authors also thank Libor Caha, Honghao Fu, Yizhi Huang, Chenxu Li, Weikang Li, Guoding Liu, Chendi Yang, Xiao Yuan, and Xingjian Zhang for their helpful discussion. This work was supported by the National Natural Science Foundation of China Grants No. 12174216 and the Innovation Program for Quantum Science and Technology Grant No. 2021ZD0300804.

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[2] Y. Wu, W.-S. Bao, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, *et al.*, Strong Quantum Computational Advantage Using a Superconducting Quantum Processor, Physical Review Letters **127**, 180501 (2021).

[3] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[4] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, and C. Monroe, Fault-tolerant control of an error-corrected qubit, Nature **598**, 281 (2021).

[5] M. Gong, X. Yuan, S. Wang, Y. Wu, Y. Zhao, C. Zha, S. Li, Z. Zhang, Q. Zhao, Y. Liu, F. Liang, J. Lin, Y. Xu, H. Deng, H. Rong, *et al.*, Experimental exploration of five-qubit quantum error-correcting code with superconducting qubits, National Science Review **9**, nwab011 (2022).

[6] L. Postler, S. Heuβen, I. Pogorelov, M. Rispler, T. Feldker, M. Meth, C. D. Marciniak, R. Stricker, M. Ringbauer, R. Blatt,

P. Schindler, M. Müller, and T. Monz, Demonstration of fault-tolerant universal quantum gate operations, Nature **605**, 675 (2022).
[7] M. H. Abobeih, Y. Wang, J. Randall, S. J. H. Loenen, C. E. Bradley, M. Markham, D. J. Twitchen, B. M. Terhal, and T. H. Taminiau, Fault-tolerant operation of a logical qubit in a diamond quantum processor, Nature **606**, 884 (2022).
[8] Y. Zhao, Y. Ye, H.-L. Huang, Y. Zhang, D. Wu, H. Guan, Q. Zhu, Z. Wei, T. He, S. Cao, F. Chen, T.-H. Chung, H. Deng, D. Fan, M. Gong, *et al.*, Realization of an Error-Correcting Surface Code with Superconducting Qubits, Physical Review Letters **129**, 030501 (2022).
[9] Google Quantum AI, R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, Nature **614**, 676 (2023).
[10] Z. Ni, S. Li, X. Deng, Y. Cai, L. Zhang, W. Wang, Z.-B. Yang, H. Yu, F. Yan, S. Liu, C.-L. Zou, L. Sun, S.-B. Zheng, Y. Xu, and D. Yu, Beating the break-even point with a discrete-variable-encoded logical qubit, Nature **616**, 56 (2023).
[11] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. B. Ataides, N. Maskara, I. Cong, X. Gao, P. S. Rodriguez, *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature , 1 (2023).
[12] S. Bravyi, D. Gosset, R. König, and M. Tomamichel, Quantum advantage with noisy shallow circuits, Nature Physics **16**, 1040 (2020).
[13] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, The complexity of NISQ, Nature Communications **14**, 6001 (2023).
[14] D. Stilck França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, Nature Physics **17**, 1221 (2021).
[15] G. De Palma, M. Marvian, C. Rouzé, and D. S. França, Limitations of Variational Quantum Algorithms: A Quantum Optimal Transport Approach, PRX Quantum **4**, 010309 (2023).
[16] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of Noisy Reversible Computation (1996), arxiv:quant-ph/9611028.
[17] M. J. Kastoryano and K. Temme, Quantum logarithmic Sobolev inequalities and rapid mixing, Journal of Mathematical Physics **54**, 052202 (2013).
[18] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Physical Review X **8**, 031027 (2018).
[19] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid Quantum-Classical Algorithms and Quantum Error Mitigation, Journal of the Physical Society of Japan **90**, 032001 (2021).
[20] I. L. Markov and Y. Shi, Simulating Quantum Computation by Contracting Tensor Networks, SIAM Journal on Computing **38**, 963 (2008).
[21] K. Noh, L. Jiang, and B. Fefferman, Efficient classical simulation of noisy random quantum circuits in one dimension, Quantum **4**, 318 (2020), arxiv:2003.13163 [quant-ph].
[22] S. Cheng, C. Cao, C. Zhang, Y. Liu, S.-Y. Hou, P. Xu, and B. Zeng, Simulating noisy quantum circuits with matrix product density operators, Physical Review Research **3**, 023005 (2021).
[23] J. C. Napp, R. L. La Placa, A. M. Dalzell, F. G. S. L. Brandão, and A. W. Harrow, Efficient Classical Simulation of Random Shallow 2D Quantum Circuits, Physical Review X **12**, 021021 (2022).
[24] Z. Cheng and M. Ippoliti, Efficient Sampling of Noisy Shallow Circuits Via Monitored Unraveling, PRX Quantum **4**, 040326 (2023).
[25] D. Aharonov, X. Gao, Z. Landau, Y. Liu, and U. Vazirani, A Polynomial-Time Classical Algorithm for Noisy Random Circuit Sampling, in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing* (Orlando FL USA, 2023) pp. 945–957.
[26] P. W. Shor, Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer, SIAM Journal on Computing **26**, 1484 (1997).
[27] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing - STOC '96* (Philadelphia, Pennsylvania, United States, 1996) pp. 212–219.
[28] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, Physical Review Letters **103**, 150502 (2009).
[29] E. Knill, R. Laflamme, and W. H. Zurek, Resilient Quantum Computation, Science **279**, 342 (1998).
[30] D. Aharonov and M. Ben-Or, Fault-Tolerant Quantum Computation with Constant Error Rate, SIAM Journal on Computing **38**, 1207 (2008).
[31] See Supplementary Information for detailed proofs of Lemma 1, Theorem 1, Lemma 2, Theorem 2, and Theorem 3, which includes [62–69].
[32] R. Carbone and A. Martinelli, Logarithmic Sobolev inequalities in non-commutative algebras, Infinite Dimensional Analysis, Quantum Probability and Related Topics **18**, 1550011 (2015).
[33] M. J. Kastoryano and F. G. S. L. Brandão, Quantum Gibbs Samplers: The Commuting Case, Communications in Mathematical Physics **344**, 915 (2016).
[34] S. Beigi, N. Datta, and C. Rouzé, Quantum Reverse Hypercontractivity: Its Tensorization and Application to Strong Converses, Communications in Mathematical Physics **376**, 753 (2020).
[35] I. Bardet, Á. Capel, A. Lucia, D. Pérez-García, and C. Rouzé, On the modified logarithmic Sobolev inequality for the heat-bath dynamics for 1D systems, Journal of Mathematical Physics **62**, 061901 (2021).
[36] Á. Capel, C. Rouzé, and D. S. França, The modified logarithmic Sobolev inequality for quantum spin systems: Classical

and commuting nearest neighbour interactions (2021), arxiv:2009.11817 [math-ph, physics:quant-ph].
[37] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nature Communications **12**, 6961 (2021).
[38] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Physical Review Letters **119**, 180509 (2017).
[39] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Reviews of Modern Physics **95**, 045005 (2023).
[40] R. Cleve and J. Watrous, Fast parallel circuits for the quantum Fourier transform, in *Proceedings 41st Annual Symposium on Foundations of Computer Science* (2000) pp. 526–536.
[41] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, Efficient Quantum Algorithms for Simulating Sparse Hamiltonians, Communications in Mathematical Physics **270**, 359 (2007).
[42] Y. Atia and D. Aharonov, Fast-forwarding of Hamiltonians and exponentially precise measurements, Nature Communications **8**, 1572 (2017).
[43] J. Haah, M. Hastings, R. Kothari, and G. H. Low, Quantum Algorithm for Simulating Real Time Evolution of Lattice Hamiltonians, in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (Paris, 2018) pp. 350–360.
[44] C. Zalka, Grover's quantum searching algorithm is optimal, Physical Review A **60**, 2746 (1999).
[45] C. Zalka, Fast versions of Shor's quantum factoring algorithm (1998), arxiv:quant-ph/9806084.
[46] S. A. Kutin, Shor's algorithm on a nearest-neighbor machine (2006), arxiv:quant-ph/0609001.
[47] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Reviews of Modern Physics **92**, 015003 (2020).
[48] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nature Reviews Physics **3**, 625 (2021).
[49] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, npj Quantum Information **8**, 114 (2022).
[50] R. Takagi, H. Tajima, and M. Gu, Universal Sampling Lower Bounds for Quantum Error Mitigation, Physical Review Letters **131**, 210602 (2023).
[51] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation (2023), arxiv:2210.11505 [math-ph, physics:quant-ph].
[52] S. Bravyi, D. Gosset, and R. Movassagh, Classical algorithms for quantum mean values, Nature Physics **17**, 337 (2021).
[53] J. Eisert, Entangling Power and Quantum Circuit Complexity, Physical Review Letters **127**, 020501 (2021).
[54] A. W. Harrow, L. Kong, Z.-W. Liu, S. Mehraban, and P. W. Shor, Separation of Out-Of-Time-Ordered Correlation and Entanglement, PRX Quantum **2**, 020339 (2021).
[55] Y. Huang, Universal entanglement of mid-spectrum eigenstates of chaotic local Hamiltonians, Nuclear Physics B **966**, 115373 (2021).
[56] J. M. Deutsch, Thermodynamic entropy of a many-body energy eigenstate, New Journal of Physics **12**, 075021 (2010).
[57] J. M. Deutsch, Eigenstate thermalization hypothesis, Reports on Progress in Physics **81**, 082001 (2018).
[58] M. K. Joshi, C. Kokail, R. Van Bijnen, F. Kranzl, T. V. Zache, R. Blatt, C. F. Roos, and P. Zoller, Exploring large-scale entanglement in quantum simulation, Nature **624**, 539 (2023).
[59] Y. Wu, S. Kolkowitz, S. Puri, and J. D. Thompson, Erasure conversion for fault-tolerant quantum computing in alkaline earth Rydberg atom arrays, Nature Communications **13**, 4657 (2022).
[60] M. Ben-Or, D. Gottesman, and A. Hassidim, Quantum Refrigerator (2013), arxiv:1301.1995 [quant-ph].
[61] N. Baspin, O. Fawzi, and A. Shayeghi, A lower bound on the overhead of quantum error correction in low dimensions (2023), arxiv:2302.04317 [quant-ph].
[62] A. Müller-Hermes, D. Stilck França, and M. M. Wolf, Relative entropy convergence for depolarizing channels, Journal of Mathematical Physics **57**, 022202 (2016).
[63] E. A. Carlen and E. H. Lieb, Brascamp–Lieb inequalities for non-commutative integration, Documenta Mathematica **13**, 553 (2008).
[64] S. Bravyi, M. B. Hastings, and F. Verstraete, Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order, Physical Review Letters **97**, 050401 (2006).
[65] K. Van Acoleyen, M. Mariën, and F. Verstraete, Entanglement Rates and Area Laws, Physical Review Letters **111**, 170501 (2013).
[66] A. Vershynina, Entanglement rates for Rényi, Tsallis, and other entropies, Journal of Mathematical Physics **60**, 022201 (2019).
[67] R. Trivedi and J. I. Cirac, Transitions in Computational Complexity of Continuous-Time Local Open Quantum Dynamics, Physical Review Letters **129**, 260405 (2022).
[68] Z. Li, S. Sang, and T. H. Hsieh, Entanglement dynamics of noisy random circuits, Physical Review B **107**, 014307 (2023).
[69] D. Aharonov, Quantum to classical phase transition in noisy quantum computers, Physical Review A **62**, 062311 (2000).

Supplementary Information: "Limitations of Noisy Quantum Devices in Computing and Entangling Power"

In Section I, we review the convergence of noisy quantum devices and provide a new proof. In Section II, we prove the results of computational limitations and provide additional lemmas and corollaries. In Section III, we prove the results of entangling power limitations.

CONTENTS

I. CONVERGENCE OF NOISY QUANTUM DEVICES

For qubit systems with depolarizing noise channels, the following lemma gives an exponential decay of relative entropy to the maximally mixed state.

Lemma 3 (Convergence of quantum qubit systems to the maximally mixed state [1], Eq. (4) in the main text)**.** *In a n-qubit noisy quantum device, after t layers of quantum gates, the relative entropy between the state $\rho(t)$ and the maximally mixed state $\sigma_0 = \frac{I}{2^n}$ decays as*

$$D(\rho(t)\|\sigma_0) \leq n(1-p)^{2t}. \tag{1}$$

For qudit systems with depolarizing channels, we also have an exponential decay, but the decay rate will be $(1-p)$. We state this result in the following lemma.

Lemma 4 (Convergence of quantum qudit systems to the maximally mixed state [2])**.** *In a n-qudit noisy quantum device, after t layers of quantum gates, the relative entropy between the state $\rho(t)$ and the maximally mixed state $\sigma_0 = \frac{I}{d^n}$ decays as*

$$D(\rho(t)\|\sigma_0) \leq n\log(d)(1-p)^{t}. \tag{2}$$

This suggests that the limitations discussed in this work can directly apply to qudit systems, with the decay rate changed to $(1-p)$.

Here, we also provide a new proof of the lemma. The proof is divided into two parts as follows. First, we provide a new proof of quantum Shearer's inequality based on the strong subadditivity of quantum entropy.

Lemma 5 (Quantum Shearer's inequality [3])**.** *Consider $t \in \mathbb{N}$ and a family $\mathcal{F} \subset 2^{\{1,2,\cdots,n\}}$ of subsets of $\{1,2,\cdots,n\}$ such that each i is included in more than t elements of $\mathcal{F}$. For any state $\rho \in \mathcal{D}(\mathbb{C}^{d_1}\otimes\mathbb{C}^{d_2}\otimes\cdots\otimes\mathbb{C}^{d_n})$ we have*

$$\sum_{F\in\mathcal{F}} S(\rho_F) \geq tS(\rho) \tag{3}$$

in which ρ_F is the reduced density matrix of ρ on subsystems F.

Proof. We will prove the theorem using mathematical induction.

For $t = 0$, the lemma holds because $S(\rho_F) \geq 0$ for any subsystem F.

For $t \geq 1$, let S denotes $\{1, 2, \cdots, n\}$. If $S \in \mathcal{F}$, then $\mathcal{F}' := \mathcal{F} \setminus \{S\}$ is a family in which each i is included in more than $t-1$ times. By induction,

$$\sum_{F \in \mathcal{F}'} S(\rho_F) \geq (t-1) S(\rho) \tag{4}$$

Thus,

$$\begin{aligned} \sum_{F \in \mathcal{F}} S(\rho_F) &= S(\rho) + \sum_{F \in \mathcal{F}'} S(\rho_F) \\ &\geq t S(\rho) \end{aligned} \tag{5}$$

If $S \notin \mathcal{F}$, find a maximal set S_1 in $\mathcal{F}$. Because $t \geq 1$, there must exists another set S_2, such that $S_2 \setminus S_1 \neq \emptyset$. Let $S_1' = S_1 \cup S_2, S_2' = S_1 \cap S_2, \mathcal{F}' = (\mathcal{F} \setminus \{S_1, S_2\}) \cup \{S_1', S_2'\}$, by the strong subadditivity of quantum entropy,

$$\begin{aligned} S(\rho_{S_1}) + S(\rho_{S_2}) &\geq S(\rho_{S_1'}) + S(\rho_{S_2'}) \\ \sum_{F \in \mathcal{F}} S(\rho_F) &\geq \sum_{F \in \mathcal{F}'} S(\rho_F) \end{aligned} \tag{6}$$

If $S_1' = S$, then combine equation (5) and (6), we get equation (4). If $S_1' \neq S$, we repeat this process on $\mathcal{F}'$. Because n is finite, this process will terminate in finite steps and finally get

$$\sum_{F \in \mathcal{F}} S(\rho_F) \geq \sum_{F \in \mathcal{F}'} S(\rho_F) \geq \cdots \geq t S(\rho) \tag{7}$$

□

Second, based on quantum Shearer's inequality, the following lemma characterizes the entropy increase caused by the noise channel.

Lemma 6. *For any n-qudit state ρ,*

$$S(\Lambda(\rho)) \geq (1-p) S(\rho) + pn \log(d), \tag{8}$$

where Λ is the noise channel and p is the strength of the depolarizing channel in noisy quantum devices.

Proof. By definition of the noise channel,

$$\Lambda(\rho) = \sum_{i=0}^{n} p^{n-i} (1-p)^i \sum_{F \subseteq [1,2,\cdots,n], |F|=i} \rho_F \otimes \frac{I}{d^{n-i}} \tag{9}$$

Take von Neumann entropy on both sides and use the subadditivity of entropy,

$$\begin{aligned} S(\Lambda(\rho)) &\geq \sum_{i=0}^{n} p^{n-i} (1-p)^i \sum_{F \subseteq [1,2,\cdots,n], |F|=i} S(\rho_F \otimes \frac{I}{d^{n-i}}) \\ &= \sum_{i=0}^{n} p^{n-i} (1-p)^i \sum_{F \subseteq [1,2,\cdots,n], |F|=i} [S(\rho_F) + (n-i)\log(d)]. \end{aligned} \tag{10}$$

Note that $\mathcal{F} = \{F \subseteq [1, 2, \cdots, n], |F| = i\}$ is a family that each i is included in more than $\binom{n-1}{i-1}$ times. Then

$$\begin{aligned} S(\Lambda(\rho)) &\geq \sum_{i=0}^{n} p^{n-i} (1-p)^i \binom{n}{i} [\frac{i}{n} S(\rho) + (n-i)\log(d)] \\ &= n\log(d) - \frac{n\log(d) - S(\rho)}{n} \sum_{i=0}^{n} p^{n-i} (1-p)^i \binom{n}{i} i \\ &= n\log(d) - (n\log(d) - S(\rho)) \sum_{i=1}^{n} p^{n-i} (1-p)^i \binom{n-1}{i-1} \\ &= n\log(d) - (n\log(d) - S(\rho))(1-p) \\ &= (1-p) S(\rho) + pn\log(d). \end{aligned} \tag{11}$$

The first inequality is from Lemma 5, and other equalities are from the direct calculation and combinatorial identity. □

Finally, we return to the proof of Lemma 4 and derive the exponential information loss.

Lemma 4's proof. For any n-qudit state ρ and the maximally mixed state $\sigma_0 = \frac{I}{d^n}$,

$$D(\rho\|\sigma_0) = n\log(d) - S(\rho). \tag{12}$$

By Lemma 6,

$$\begin{aligned} D(\Lambda(\rho)\|\sigma_0) &= n\log(d) - S(\Lambda\rho) \\ &\leq n\log(d) - (pn + (1-p)S(\rho)) \\ &= (1-p)(n\log(d) - S(\rho)) \\ &= (1-p)D(\rho\|\sigma_0) \end{aligned} \tag{13}$$

Entropy is invariant after passing through a unitary channel $\{\mathcal{U}_i\}$,

$$D(\mathcal{U}_i(\rho)\|\sigma_0) = D(\rho\|\sigma_0). \tag{14}$$

Note that the state after t layers is written as

$$\rho(t) = \Lambda \circ \mathcal{U}_t \circ \Lambda \circ \cdots \circ \Lambda \circ \mathcal{U}_2 \circ \Lambda \circ \mathcal{U}_1(|0\rangle\langle 0|^{\otimes n}), \tag{15}$$

where $\mathcal{U}(\rho) = U\rho U^\dagger$ is the superoperator for each layer of gates. Therefore, we can finish the proof by alternatively using Eq. (13) and (14). □

II. PROOFS OF LIMITATIONS OF DEPTHS FOR COMPUTATIONAL ADVANTAGES

A. Proof of lower bounds of total entropy in hybrid algorithms

This section provides proof of Lemma 1 in the main text. We recall the definition of conditional entropy.

Definition 5 (Conditional Entropy)**.** *For two discrete random variables X and Y, the conditional entropy is defined as*

$$S(X|Y) = -\sum_{x,y} P(X=x, Y=y)\log P(X=x|Y=y) \tag{16}$$

The conditional entropy is useful in our proof. We briefly prove two entropy equalities with direct calculation for clarity and completeness. Firstly,

$$\begin{aligned} S(X|Y) &= -\sum_y P(Y=y)\sum_x P(X=x|Y=y)\log P(X=x|Y=y) \\ &= \sum_y P(Y=y)S(X|Y=y) \end{aligned} \tag{17}$$

in which $S(X|Y=y)$ is the entropy of marginal distribution of (X,Y) when Y is fixed value y. Secondly,

$$\begin{aligned} S(XY) &= -\sum_{x,y} P(X=x,Y=y)\log P(X=x,Y=y) \\ &= -\sum_{x,y} P(X=x,Y=y)[\log P(X=x|Y=y) + \log P(Y=y)] \\ &= -\sum_{x,y} P(X=x,Y=y)\log P(Y=y) - \sum_{x,y} P(X=x,Y=y)\log P(X=x|Y=y) \\ &= -\sum_y P(Y=y)\log P(Y=y) - \sum_{x,y} P(X=x,Y=y)\log P(X=x|Y=y) \\ &= S(Y) + S(X|Y) \end{aligned} \tag{18}$$

After introducing the notion of conditional entropy, we provide proof of Lemma 1 in the main text. Here, we consider a slightly more general case, where the hybrid algorithm can query quantum devices with different numbers of qubits n_i.

Lemma 7 (A generalised version of Lemma 1 in the main text)**.** *Suppose we use noisy quantum devices of depth at least t for q times. Let $X_i = C_{n_i}(\rho_i)$ denotes the measurement result of the i-th quantum circuit. Here, $\rho_i = \Phi^{(i)}(|0\rangle\langle 0|^{\otimes n_i})$, $\Phi^{(i)} = \Lambda \circ \mathcal{U}_{t_i}^{(i)} \circ \Lambda \circ \cdots \circ \Lambda \circ \mathcal{U}_2^{(i)} \circ \Lambda \circ \mathcal{U}_1^{(i)}$ denotes the quantum channel as a whole process combining all gates and noise in sequential order, $t_i \geq t$, $\mathcal{U}_j^{(i)}$ is an arbitrary quantum channel. We obtain q random variables $X_1, \ldots, X_q$. Then $S(X_1, \cdots, X_q) \geq (\sum_{i=1}^q n_i)(1-\zeta)$, in which $\zeta = (1-p)^{2t}$.*

Proof. By Lemma 3, for any quantum channel $\mathcal{U}_j^{(i)}$,

$$S(\Phi^{(i)}(|0\rangle\langle 0|^{\otimes n_i})) \geq n_i(1-\zeta), \tag{19}$$

with $\zeta = (1-p)^{2t}$.

According to our assumption, measurements are noiseless and performed on a computational basis. The resulting distribution of measurement outcome X_i is on the diagonal of $\Delta(\Phi^{(i)}(\rho))$, where Δ denotes the dephasing channel for the computational basis. The entropy of the dephased state is the Shannon entropy of random strings X_i,

$$S(X_i) = S(\Delta(\Phi_t(\rho))) \tag{20}$$

Note that $S(\Delta(\rho)) \geq S(\rho)$ for arbitrary quantum state ρ from quantum data processing inequality. Thus,

$$S(X_i) = S(\Delta(\Phi_t(\rho))) \geq S(\Phi^{(i)}(\rho)) \geq n_i(1-\zeta) \tag{21}$$

Note that equation Eq. (21) holds regardless of channels $\{\mathcal{U}_j^{(i)}\}_{j=1,2,\cdots,t_i}$ implemented in the noisy quantum device. Although X_i may depend on $X_1, \ldots, X_{i-1}$, Eq. (21) holds regardless of previous measurement outcome. In other words, for all $1 \leq i \leq q$, $1 \leq j < i$ and $x_j \in \{0,1\}^{n_j}$, we have:

$$S(X_i | X_1 = x_1, X_2 = x_2, \cdots, X_{i-1} = x_{i-1}) \geq n_i(1-\zeta) \tag{22}$$

Then we sum over all previous measurement outcomes $x_1, \cdots, x_{i-1}$ and use Eq. (17),

$$\begin{aligned} S(X_i | X_1, X_2, \cdots, X_{i-1}) &= \sum_{x_1,\cdots,x_{i-1}} P(X_1 = x_1, X_2 = x_2, \cdots, X_{i-1} = x_{i-1}) S(X | X_1 = x_1, X_2 = x_2, \cdots, X_{i-1} = x_{i-1}) \\ &\geq \sum_{x_1,\cdots,x_{i-1}} P(X_1 = x_1, X_2 = x_2, \cdots, X_{i-1} = x_{i-1}) n_i(1-\zeta) \\ &\geq n_i(1-\zeta) \end{aligned} \tag{23}$$

By Eq. (18),

$$\begin{aligned} S(X_1, X_2, \cdots, X_q) &= S(X_q | X_1, X_2, \cdots, X_{q-1}) + S(X_1, X_2, \cdots, X_{q-1}) \\ &= S(X_q | X_1, X_2, \cdots, X_{q-1}) + S(X_{q-1} | X_1, X_2, \cdots, X_{q-2}) + S(X_1, X_2, \cdots, X_{q-2}) \\ &= \sum_{i=1}^{q} S(X_i | X_1, X_2 \cdots, X_{i-1}) \end{aligned} \tag{24}$$

Combined with (23),

$$S(X_1, X_2, \cdots, X_q) \geq \sum_{i=1}^{q} n_i(1-\zeta) \tag{25}$$

□

Following this generalized form, if considering all devices to have the same number of qubits $n_i = n$ and the number of queried measurement bits $Q = \sum_{i=1}^q n_i = qn$, we will get Lemma 1 in the main text.

B. Proofs of the limitations of hybrid algorithm

This section analyzes the limitations of hybrid algorithms with noisy quantum devices. First, we focus on decision problems. Then, we will extend the result of decision problems to other problems. The notion of languages, probabilistic Turing machines, and hybrid algorithms with noisy quantum devices are introduced in the Methods section of the main text.

Here, we provide proof of Theorem 1 in the main text.

Theorem 1's proof. For convenience, denote the depth upper limit in Theorem 1 by $t^\star = \frac{1}{2|\log(1-p)|}(\log(Q)+5)$. As stated in the theorem, we will consider the case where $t \geq t^\star$.

Let Y denote the random strings in $\mathcal{A}$, including the string from noisy quantum devices and random inputs for the probabilistic Turing machines. Let W_1 denote Y's substring from noisy quantum devices, namely, $W_1 = Y_{\mathcal{S}}$, where $\mathcal{S}$ is the indices that correspond to noisy quantum devices queries.

Now, consider a construction of a classical algorithm, i.e., PTM M. Replace W_1 in Y with a random string W_2 from a uniform distribution, and denote the new string as Z, such that $W_2 = Z_{\mathcal{S}}$. By Lemma 1 in the main text,

$$D(W_1 \| W_2) \leq Q\zeta, \tag{26}$$

in which $\zeta = (1-p)^{2t}$. Here, Y, Z can be regarded as generated by passing W_1, W_2 through the same channel Γ, that is, $Y = \Gamma(W_1), Z = \Gamma(W_2)$. The channel Γ represents the transfer from quantum measurement outcomes to the rest of classical random strings, determined by classical processing and controls. Then, by data processing inequality, we have

$$D(Y\|Z) = D(\Gamma(W_1)\|\Gamma(W_2)) \leq D(W_1\|W_2) \leq Q\zeta \leq \frac{1}{32}. \tag{27}$$

The first inequality is data processing inequality; the second is Eq. (26); the third is from the assumption $t \geq t^\star$. By construction, $\mathcal{A}$ and M has same classical processing structure, that is, $\forall x \in \{0,1\}^n, y \in \{0,1\}^{T(n)}, M(x,y) = M'(x,y)$. Then we have the following equation:

$$D(\mathcal{A}(x,Y)\|M(x,Z)) \leq D(Y\|Z) \leq \frac{1}{32}. \tag{28}$$

The first inequality is from data processing inequality; the second is from Eq. (27). Pinsker's inequality suggests

$$\|\mathcal{A}(x,Y) - M(x,Z)\|_1 \leq \sqrt{2D(\mathcal{A}(x,Y)\|M(x,Z))} \leq \frac{1}{4}. \tag{29}$$

We have the probability that the classical algorithm solves the decision problem L,

$$\Pr[M(x,Z) = L(x)] = \Pr[M(x,Z) = \mathcal{A}(x,Y)] \geq \frac{2}{3} - \frac{1}{8} = \frac{13}{24}. \tag{30}$$

Then, we can repeat the PTM M a constant number of times and then take a majority vote on the output results to obtain the final decision. By doing so, we can ensure that the probability of the output being equal to $L(x)$ is at least $\frac{2}{3}$. Therefore, L can be decided by a PTM without noisy quantum devices in $O(T(n))$ time. □

This leads to the absence of quantum advantages of many existing algorithms, including the examples mentioned in the main text, such as Shor's, Grover's, and HHL algorithms. For Shor's and HHL algorithms, the algorithm's output is not one bit as assumed in our decision problem formalism. Yet, in the following section, we will show that they can still be reduced to a decision problem, therefore, within the scope of Theorem 1.

For sampling problems, we use Theorem 2 to suggest statistical indistinguishability of noisy quantum samples from the uniform distribution for $t \geq \omega(\log(Q))$, where Q refers to the number of samples. We can easily adapt the previous proof to get this new result.

Theorem 2's proof. Consider the distinguisher, mentioned in the theorem, as a hybrid algorithm $\mathcal{A}$ with no input x and only queries noisy quantum devices. We require $\mathcal{A}(\emptyset, Y)$ to tell whether it actually queries noisy quantum devices or just random coins from a uniform distribution. Specifically, when given quantum devices, it should return a "true" with probability at least $\frac{2}{3}$; when given random coins, it should always return a "false."

Suppose it can indeed solve this decision problem. Now, we replace noisy quantum devices with random coins, equivalent to $M(\emptyset, Z)$. From Eq. (30), with probability at least $\frac{13}{24}$, the distinguisher will return the same answer in the two cases, $\mathcal{A}(\emptyset, Y)$ and $M(\emptyset, Z)$. Note that when given random coins, it should return a "false." Therefore, when given noisy quantum devices, it will answer "false" with probability at least $\frac{13}{24}$, which is a wrong answer. Therefore, it cannot succeed with at least $\frac{2}{3}$ probability, i.e., failing the decision problem. □

C. Equivalence of factorizing and solving linear systems to decision problems

In this part, we present some problems that are *equivalent* to a decision problem so that our result can pose limitations of these problems on noisy quantum devices. Here, *equivalent* means that solving one problem in polynomial time implies being able to solve the other problem in polynomial time.

Proposition 1 (Factorizing)**.** *The following two problems are equivalent:*

1. *Given n, output the smallest non-trivial factor of n.*

2. *Given (n, k), determine if there exist a non-trivial factor of n that is less than k.*

Proof. Suppose we solve problem 1 by algorithm $\mathcal{A}$ in polynomial time. Then for any input (n, k), we use $\mathcal{A}$ to factorize n and get its minimal non-trivial factor m. Then we compare m and k. Thus, we solve problem 2 in polynomial time.
Suppose we solve problem 2 by algorithm $\mathcal{B}$ in polynomial time. Then, we can use binary search to find the smallest non-trivial factor of n in polynomial time. □

Proposition 2 (Linear systems of equations)**.** *The following two problems are equivalent:*

1. *Given a classical description of the $N \times N$ matrix A, a unit vector $|b\rangle$, a quantum operator M, and a precision ϵ. Output $\langle x|M|x\rangle$ with precision ϵ, $|x\rangle$ is the solution of $A|x\rangle = |b\rangle$.*

2. *Given a classical description of the $N \times N$ matrix A, a unit vector $|b\rangle$, a quantum operator M, a real number a , and a precision ϵ. Determine if $\langle x|M|x\rangle$ is less than $a + \epsilon$ (output 0), or is larger than $a + \epsilon/2$(output 1). $|x\rangle$ is the solution of $A|x\rangle = |b\rangle$.*

Proof. Suppose we solve problem 1 by algorithm $\mathcal{A}$ in polynomial time. For an input $(A, |b\rangle, M, a, \epsilon)$ of problem 2, we query a with input $(A, |b\rangle, M, a, \epsilon/10)$ to get an output a'. If $a' < a + \frac{9\epsilon}{10}$, this means $\langle x| M |x\rangle \leq a' + \frac{\epsilon}{10} < a + \epsilon$, then output 0. Otherwise, $\langle x| M |x\rangle \geq a' - \frac{\epsilon}{10} \geq a + \frac{4\epsilon}{5}$ output 1. Then, we solve problem 2 in polynomial time.
Suppose we solve problem 2 by algorithm $\mathcal{B}$ in polynomial time. For an $(A, |b\rangle, M, \epsilon)$ of problem 1, we can solve problem 1 by performing a binary search in $O(\log(\frac{1}{\epsilon}))$ time on algorithm $\mathcal{B}$ with decision $\epsilon' = \frac{\epsilon}{10}$ inputted to $\mathcal{B}$. □

Proposition 1 and Proposition 2 suggest the applicability of Theorem 1 to problems of factoring and solving linear systems, corresponding to Shor's and HHL algorithms, respectively.

III. PROOFS OF LIMITATIONS OF ENTANGLEMENT PRODUCTION

A. Entanglement in a one-dimensional chain

A key observation is that gates act on the qubits in a spatial locality, respecting the one-dimensional chain topology. Consider a qubit in the chain within t layers. Its interaction should be restrained within a distance of t. The restraint can be understood through a physical picture of a light cone, illustrated in Fig. 1.
This will limit entanglement spreading and the bipartite entanglement between halves of the chain. We develop the intuition into the following lemma.

Lemma 8 (Eq. (6) in the main text)**.** *For a one-dimensional lattice, namely a chain, at a fixed number of layers t, bipartite entanglement $E(\rho(t))$ is upper bounded by $2t$. If A contains one end of the chain, the upper bound will become t. The upper bounds hold for any quantities E that do not increase under local operations and are upper bounded by t.*

Proof. Recall the output state is given by Eq. (15). Now, we divide gates in $\mathcal{U}_t$ into two kinds: $\mathcal{U}_t^{(1)}$ contains gates that act across A and $\bar{A}$; while $\mathcal{U}_t^{(2)}$ contains other gates, which act on two qubits both in a same subsystem. Importantly, $\mathcal{U}_t^{(2)}$ is a local operation concerning the partition and has no contribution to entanglement. We construct the following state by reducing Λ and $\mathcal{U}_t^{(2)}$,

$$\rho'(t) = \mathcal{U}_t^{(1)}(\rho(t-1)), \tag{31}$$

satisfying $\rho(t) = \Lambda \circ \mathcal{U}_t^{(2)}(\rho'(t))$. It can be converted to $\rho(t)$ by local operations, thus $E(\rho(t)) \leq E(\rho'(t))$. And by the definition of $\mathcal{U}_t^{(1)}$, the size of its support is at most four.
Similarly, to reduce the $(t-1)$th layer, we again divided $\mathcal{U}_{t-1}$ into two kinds. This time the gates and noise to be

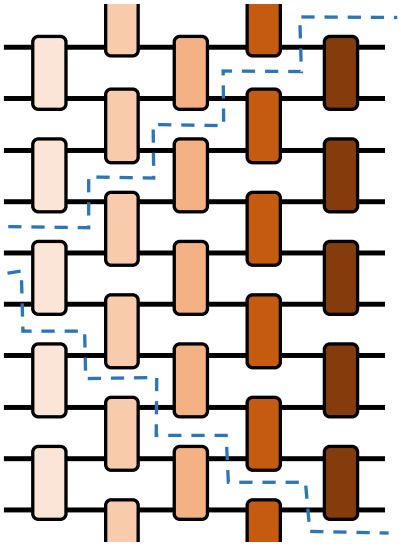


FIG. 1. The physical picture of the light cone in one dimension. Dot lines are boundaries of the light cone. We use a specific brickwise architecture for illustration, which is not required. The color of the layers gets darker with increasing depth, representing the state is approaching σ_0 according to Lemma 3.

reduced must satisfy an additional condition that they should commute with $\mathcal{U}_t^{(1)}$, which requires that they do not overlap with $\sup \mathcal{U}_t^{(1)}$ in general.

$$\rho'(t-1) = \mathcal{U}_t^{(1)} \circ \Lambda_{\sup(\mathcal{U}_t^{(1)})} \circ \mathcal{U}_{t-1}^{(1)}(\rho(t-2)) = \mathcal{U}'_{t-1}(\rho(t-2)), \tag{32}$$

where the combined channel $\mathcal{U}'_{t-1}$ satisfying $|\sup(\mathcal{U}'_{t-1})| \leq 4$. Local operations can again convert it to $\rho'(t)$.

We follow the procedure and iteratively reduce all layers of gates. The final resulting state is $\rho'(1) = \mathcal{U}'_1(|0\rangle\langle 0|^{\otimes n})$ with $|\sup(\mathcal{U}'_1)| \leq 4t$. Considering bipartite entanglement is at most half of the system size, we conclude that $E(\rho(t)) \leq E(\rho'(1)) \leq 2t$. The whole procedure is illustrated in Fig. 2.

The support can only grow to one side when A contains one end of the chain. Following the same procedure, the support size has an upper bound of t. Thus, $E(\rho(t)) \leq t$. □

Our result is a generalization of previous work that establishes a similar bound only for the entanglement entropy of pure states [4–7]. As far as we know, no such bounds have been derived for the mixed-state case, which is essentially the case for analyzing noisy quantum devices.

Then, by combining the results with Lemma 3, we derive the upper bounds of quantum relative entropy of entanglement and quantum mutual information as Theorem 3 in the main text.

Theorem 3's proof. First, we show both the quantum relative entropy of entanglement and quantum mutual information are upper bounded by $D(\rho\|\sigma_0)$. For quantum relative entropy of entanglement, this is because

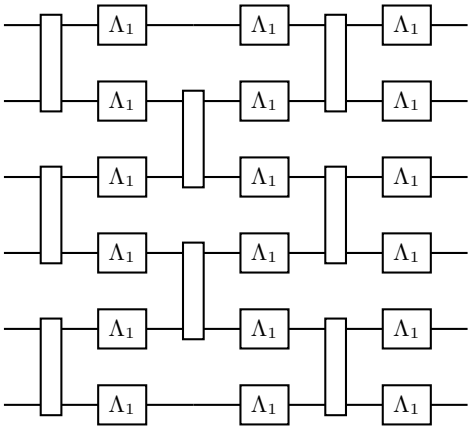
$$E_R(A:\bar{A}) = \min_{\sigma \in \mathrm{SEP}} D(\rho\|\sigma) \leq D(\rho\|\sigma_0), \tag{33}$$

where SEP denotes the set of separable states over A and $\bar{A}$ partition. For quantum mutual information, this is because

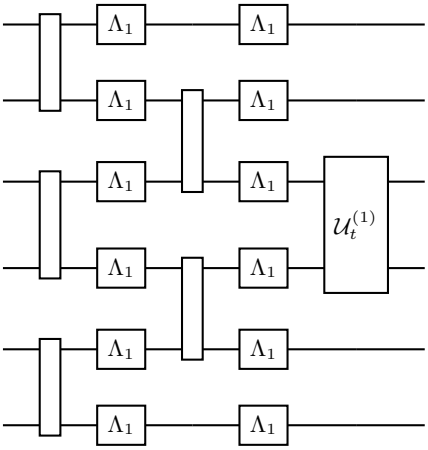
$$I(A:\bar{A}) = S(A) + S(\bar{A}) - S(\rho) \leq n - S(\rho) = D(\rho\|\sigma_0). \tag{34}$$

And also we know $E_R(A:\bar{A})$ is upper bounded by $I(A:\bar{A})$, because $E_R(A:\bar{A}) = \min_{\sigma\in\mathrm{SEP}} D(\rho\|\sigma) \leq D(\rho\|\rho_A \otimes \rho_{\bar{A}}) = I(A:\bar{A})$. Combined with Lemma 3,

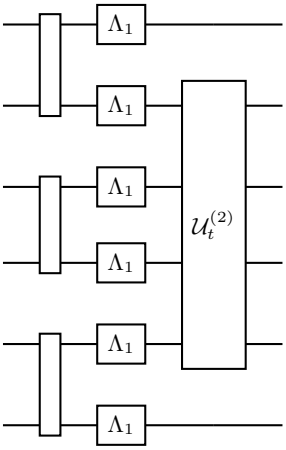
$$I(A:\bar{A}) \leq D(\rho\|\sigma_0) \leq n(1-p)^{2t}. \tag{35}$$



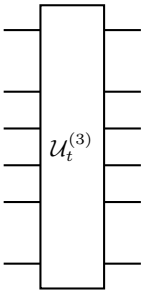
(a) The original circuits without reduction.



(b) After one iteration of reduction.



(c) After two iterations of reduction.



(d) Ater three iterations of reduction.

FIG. 2. The iterative reduction of layers of gates. We show a case of six qubits across one boundary between A and $\bar{A}$ as a part of a larger quantum device for illustration. The half A contains the upper three qubits, while $\bar{A}$ contains the rest. We show the iterative reduction by the series of subfigures.

From Lemma 8, we have another upper bounds

$$I(A:\bar{A}) \leq t. \tag{36}$$

By combining the two upper bounds, we have

$$I(A:\bar{A}) \leq \min\{n(1-p)^{2t}, t\} \leq t^\star, \tag{37}$$

where $t^\star$ is the greatest integer that satisfied $n(1-p)^{2t^\star} \geq t^\star$.

Here, we first consider the case where $p < \frac{2}{3}$. Then, If $n \geq 3$ so that $n > \frac{1}{2(1-p)}$, we will have $t^\star \geq 1$ and $t^\star \leq \frac{\log(n)}{-\log(1-p)}$. Therefore, the upper bound follows as

$$I(A:\bar{A}) \leq t^\star \leq \frac{\log(n)}{-\log(1-p)}. \tag{38}$$

When $n = 2$, the maximal entanglement is upper bounded by 1, and $1 < \frac{\log(n)}{-2\log(1-p)}$. We need at least two qubits to discuss entanglement.

Otherwise, $p \geq \frac{2}{3}$, we will show that the state will always be fully separable. To see this, we need only to show that any two-qubit quantum states ρ will become separable after the single-qubit depolarizing noise channel if $p \geq \frac{2}{3}$ so

that entanglement cannot be produced from a product initial state by such noisy circuits. Without loss of generality, we suppose ρ to be a pure state $|\psi\rangle\langle\psi|$. After the single-qubit depolarizing noise channel, its purity will be less than $\left(\frac{1+(1-p)^2}{2}\right)^2 < \frac{1}{3}$ by Lemma 9. Note that a two-qudit state with purity less than or equal to $\frac{1}{d+1}$ is separable. In our qubit case, the two-qubit state is less than $\frac{1}{3}$, therefore separable. □

Lemma 9 (Purity loss of a pure state)**.** *Suppose the pure state is $|\psi\rangle$, which can be arbitrarily entangled. After a layer of single-qubit depolarizing channels, its purity is less than $\left(\frac{1+(1-p)^2}{2}\right)^n$.*

Proof. After a layer of noise channel, the state becomes

$$\rho' = \sum_{a\in\{0,1\}^n} p^{W[a]}(1-p)^{n-W[a]} \frac{I_a}{2^{W[a]}} \otimes \mathrm{Tr}_a |\psi\rangle\langle\psi| . \tag{39}$$

Here a is a n-bit binary vector to denote a subsystem and $a_i = 1$ i.f.f. ith qubit is in the subsystem. $W[a] = \sum_{i=1}^n a_i$. Its purity is

$$\mathrm{Tr}\left[\rho'^2\right] = \sum_{a,b} p^{W[a]+W[b]}(1-p)^{2n-(W[a]+W[b])} \mathrm{Tr}\left((\frac{I_a}{2^{W[a]}} \otimes \mathrm{Tr}_a |\psi\rangle\langle\psi|)(\frac{I_b}{2^{W[b]}} \otimes \mathrm{Tr}_b |\psi\rangle\langle\psi|) \right) \tag{40}$$

Let $\tilde{W}[a,b] = p^{W[a]+W[b]}(1-p)^{2n-(W[a]+W[b])}$:

$$\begin{aligned} \mathrm{Tr}\left[\rho'^2\right] &= \sum_{a,b} \tilde{W}[a,b] \,\mathrm{Tr}\left((\mathrm{Tr}_{a\cup b} |\psi\rangle\langle\psi|)^2\right) \\ &\le \sum_{a,b} \tilde{W}[a,b] = (\frac{1+(1-p)^2}{2})^n. \end{aligned} \tag{41}$$

Here $a\cup b$ denotes qubits that are either in a or in b. The last equality is because all "trace" terms in the summation are equal to 1 if $|\psi\rangle = |0\rangle^{\otimes n}$. The physical meaning of the lemma is interesting: entangled states will lose more purity than product states. □

B. Entanglement in a two-dimensional lattice

For general dimensional topologies, if the entanglement between subsystem A and its complement $\bar{A}$ scales with the area of the boundary, the system follows the area-law scaling. Alternatively, if the entanglement scales with the volume of A, the system follows the volume-law scaling.

(a) One-dimensional lattice (b) Two-dimensional lattice

FIG. 3. Two typical topology setups for noisy quantum devices are studied in this work. The filled circles denote qubits. The lines are connections of those qubits where quantum gates can be placed.

For two-dimensional lattices, we take A as a square, whose size is $(n^{1/2}, n^{1/2})$. And one of A's vertex is on a vertex of the whole lattice.

Theorem 5. *In a two-dimensional lattice with the number of qubits $n > \frac{9}{(1-p)^4}$, for both quantum mutual information and quantum relative entropy of entanglement, we have upper bounds*

$$E(\rho(t)) < \frac{\frac{1}{2}\log(n) - 1}{-2\log(1-p)} 2n^{\frac{1}{2}} + \left(\frac{\frac{1}{2}\log(n) - 1}{-2\log(1-p)}\right)^2 . \tag{42}$$

Proof. In two-dimensional lattices, a light cone lemma can be established similarly. Now, the interaction of qubits can expand in two different dimensions. So the size of the support of the reduced channel, at last, will be upper bounded

$$|\sup(\mathcal{U}_1')| \leq (n^{\frac{1}{2}}+t)(n^{\frac{1}{2}}+t) - n = 2tn^{\frac{1}{2}} + t^2. \tag{43}$$

For both quantum mutual information and quantum relative entropy of entanglement, denoted by $E(A:\bar{A})$, we have the following two upper bounds,

$$\begin{aligned} E(A:\bar{A}) &\leq n(1-p)^{2t}, \\ E(A:\bar{A}) &\leq 2tn^{\frac{1}{2}} + t^2. \end{aligned} \tag{44}$$

Combining the two bounds, we have

$$E(A:\bar{A}) \leq \min\{n(1-p)^{2t}, 2tn^{\frac{1}{2}} + t^2\} \leq 2t^\star n^{\frac{1}{2}} + t^{\star 2}, \tag{45}$$

where $t^\star$ is the greatest integer that satisfies $n(1-p)^{2t} \geq 2n^{\frac{1}{2}}t + t^2$. With the requirement of $n > \frac{9}{(1-p)^4}$, we have

$$n(1-p)^2 > 3n^{\frac{1}{2}} \geq 2n^{\frac{1}{2}} + 1. \tag{46}$$

Then we assure that $t^\star \geq 1$ and can further derive

$$n(1-p)^{2t^\star} \geq 2n^{\frac{1}{2}} + 1 > 2n^{\frac{1}{2}}. \tag{47}$$

This will give $t^\star$ an upper bound,

$$t^\star \leq \frac{\frac{1}{2}\log(n) - 1}{-2\log(1-p)}. \tag{48}$$

Finally, we have the upper bound

$$E(A:\bar{A}) < \frac{\frac{1}{2}\log(n) - 1}{-2\log(1-p)} 2n^{\frac{1}{2}} + \left(\frac{\frac{1}{2}\log(n) - 1}{-2\log(1-p)}\right)^2. \tag{49}$$

□

When n is sufficiently large, the upper bound scales as $n^{\frac{1}{2}}\log(n)$, exhibiting an area-law scaling with an additional logarithmic factor. However, unlike the one-dimensional case, this entanglement scaling has no known implications for classical simulatability in the two-dimensional lattice. This is because no known efficient classical simulation algorithm exists for even area-law systems in two dimensions. As a two-dimensional extension of MPS, projected Entangled Pair States (PEPS) have exponential contraction complexity.

As in the one-dimensional case, the proposed entanglement scaling limits the simulation of two-dimensional quantum systems. However, it is feasible to use a two-dimensional noisy quantum device to simulate a one-dimensional quantum system [8].

We summarize our results in different lattice dimensions and compare them with the numerical results from the particular case of random noisy circuits [9, 10] in Table I.

TABLE I. Bounds of quantum mutual information and relative entropy of entanglement in a noisy quantum device. For two-dimensional lattices, A is a square with the size of $(n^{\frac{1}{2}}, n^{\frac{1}{2}})$. Previous works numerically study the entanglement production of noisy circuits with random two-qubit gates arranged in a brick-wise architecture, which is a special case of our model. We list their results for comparison.

Topology	Bounds	Entanglement scaling	Random circuits [9, 10]
1D Chain	$\frac{\log(n)}{-2\log(1-p)}$	$O(\log(n))$	O(1)
2D Lattice	$\frac{\frac{1}{2}\log(n)-1}{-2\log(1-p)}2n^{\frac{1}{2}} + \left(\frac{\frac{1}{2}\log(n)-1}{-2\log(1-p)}\right)^2$	$O(n^{\frac{1}{2}}\log(n))$	$O(n^{\frac{1}{2}})$

C. Entanglement between distant regions in a one-dimensional chain

Besides the entanglement between adjacent regions, we also investigate the entanglement between two distant parts of the devices. Consider two distant contiguous regions, A and B. We define their distance $d(A,B)$ as the shortest path connecting them in a given qubit connection topology.

If no noise exists, the entanglement between the two arbitrarily far parts can reach the optimal value of $\min(|A|,|B|)$ after a depth more than their distance. However, when the device suffers from noise and their distance is far, such a depth cannot be reached before the system gets too noisy. Formally, we have the following theorem.

Theorem 6. *The entanglement between the two distant regions A and B in a one-dimensional chain, with distance $d(A,B)$ is upper bounded by*

$$E(A:B) \leq (|A|+|B|)(1-p)^{2d(A,B)} + \frac{4(1-p)}{p(2-p)}. \tag{50}$$

Proof. Following the previously used method, we consider the loss of information in the system AB. Unlike in the previous problem, AB may gain information by interacting with the outside systems. AB can only interact with the four qubits adjacent to their boundaries for each layer in a one-dimensional chain. Therefore, four bits of information can be regained at most before depolarizing noise comes. The information loss in a layer will be

$$D(\rho(t+1)_{AB}\|(\sigma_0)_{AB}) \leq (1-p)^2\left[D(\rho(t)_{AB}\|(\sigma_0)_{AB})+4\right]. \tag{51}$$

We rewrite it into

$$D(\rho(t+1)_{AB}\|(\sigma_0)_{AB}) - \frac{4(1-p)}{p(2-p)} \leq (1-p)^2\left[D(\rho(t)_{AB}\|(\sigma_0)_{AB}) - \frac{4(1-p)}{p(2-p)}\right]. \tag{52}$$

This suggest that $D(\rho(t)_{AB}\|(\sigma_0)_{AB}) - \frac{4(1-p)}{p(2-p)}$ undergoes an exponential decay. Note that $D(\rho(0)_{AB}\|(\sigma_0)_{AB}) = |A|+|B|$ takes the maximal value. We will have an exponentially decaying upper bounds of entanglement with an extra term $\frac{4(1-p)}{p(2-p)}$.

$$E(A:B) \leq D(\rho(t)_{AB}\|(\sigma_0)_{AB}) \leq (|A|+|B|)(1-p)^{2t} + \frac{4(1-p)}{p(2-p)}. \tag{53}$$

From gate locality, we know that when $t < d(A,B)$, the entanglement will be strictly zero. To see this, we can still use the reduction techniques shown in Fig. 2 and do the same tricks. Combining this with the bounds we just obtained, we will eventually have

$$E(A:B) \leq (|A|+|B|)(1-p)^{2d(A,B)} + \frac{4(1-p)}{p(2-p)}. \tag{54}$$

□

The result we obtain is exponentially vanishing with the existence of an additional term $\frac{4(1-p)}{p(2-p)}$ related solely to the noise strength. Therefore, we show that the entanglement between far regions has an upper limit regardless of how large the sizes of the two regions are.

Previous work also studied the entanglement of distant regions, with more flexible consideration of noise model and error correction [11]. Their result depends on p being beyond or below a threshold p_c, which is related to percolation theory. When $p > p_c$, they found the same exponentially vanishing entanglement regarding $d(A,B)$, but without additional terms as in our result. Their result is better than ours for the $p > p_c$ case. Our result still applies for the $p < p_c$ case. This suggests that the exponential decay behavior of entanglement between distant regions is threshold-free when depolarizing noise is considered, and error correction is not applicable.

It is also important to point out that the upper bounds we derived could be strengthened. This is because when deriving Eq. (51), we suppose the surrounding environment cools the four boundary qubits in each layer throughout the process. However, considering that depolarizing noise acts on all qubits, such cooling could be difficult. Previous work has suggested that quantum cooling in the context of quantum computing is impossible when depolarizing noise is considered [12]. Although it cannot be used directly in our setup, the possibility of cooling to increase entanglement between distant regions is also small.

[1] M. J. Kastoryano and K. Temme, Quantum logarithmic Sobolev inequalities and rapid mixing, Journal of Mathematical Physics **54**, 052202 (2013).

[2] A. Müller-Hermes, D. Stilck França, and M. M. Wolf, Relative entropy convergence for depolarizing channels, Journal of Mathematical Physics **57**, 022202 (2016).

[3] E. A. Carlen and E. H. Lieb, Brascamp–Lieb inequalities for non-commutative integration, Documenta Mathematica **13**, 553 (2008).

[4] S. Bravyi, M. B. Hastings, and F. Verstraete, Lieb-Robinson Bounds and the Generation of Correlations and Topological Quantum Order, Physical Review Letters **97**, 050401 (2006).

[5] K. Van Acoleyen, M. Mariën, and F. Verstraete, Entanglement Rates and Area Laws, Physical Review Letters **111**, 170501 (2013).

[6] A. Vershynina, Entanglement rates for Rényi, Tsallis, and other entropies, Journal of Mathematical Physics **60**, 022201 (2019).

[7] J. Eisert, Entangling Power and Quantum Circuit Complexity, Physical Review Letters **127**, 020501 (2021).

[8] R. Trivedi and J. I. Cirac, Transitions in Computational Complexity of Continuous-Time Local Open Quantum Dynamics, Physical Review Letters **129**, 260405 (2022).

[9] K. Noh, L. Jiang, and B. Fefferman, Efficient classical simulation of noisy random quantum circuits in one dimension, Quantum **4**, 318 (2020), arxiv:2003.13163 [quant-ph].

[10] Z. Li, S. Sang, and T. H. Hsieh, Entanglement dynamics of noisy random circuits, Physical Review B **107**, 014307 (2023).

[11] D. Aharonov, Quantum to classical phase transition in noisy quantum computers, Physical Review A **62**, 062311 (2000).

[12] M. Ben-Or, D. Gottesman, and A. Hassidim, Quantum Refrigerator (2013), arxiv:1301.1995 [quant-ph].

Quantum circuits for diagonal unitary matrices with reflection symmetry

Xinchi Huang[1 2 *] Taichi Kosugi[1 2] Hirofumi Nishi[1 2] Yu-ichiro Matsushita[1 2 3]

[1] *Quemix Inc., Taiyo Life Nihombashi Building, 2-11-2, Nihombashi Chuo-ku, Tokyo, 103-0027, Japan*
[2] *School of Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan*
[3] *Quantum Materials and Applications Research Center, National Institutes for Quantum Science and Technology (QST), 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8550, Japan*

Abstract. During the noisy intermediate-scale quantum (NISQ) era, it is necessary to optimize the quantum circuits in circuit depth and gate count, especially entanglement gates, including the CNOT gate. Based on a natural gate set {CNOT, R_z}, we simplify the quantum circuits for diagonal unitary matrices in a specific case of reflection symmetry. Compared to the existing results for general diagonal unitary matrices, our proposed circuit, in this case, achieves nearly half reduction in both the CNOT count and the circuit depth. Moreover, we show that our result has practical applications, including the part of symmetric potential in the first-quantized Hamiltonian simulation.

Keywords: quantum circuit, diagonal unitary matrix, reflection symmetry, first-quantized Hamiltonian simulation

1 Problem description and our target

We consider a diagonal unitary matrix, which is given by

$$\mathrm{D}(\boldsymbol{\theta}) := \begin{pmatrix} e^{\mathrm{i}\theta_0} & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & e^{\mathrm{i}\theta_1} & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & e^{\mathrm{i}\theta_{2^{n-1}-1}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{\mathrm{i}\theta_{2^{n-1}}} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{\mathrm{i}\theta_{2^n-1}} \end{pmatrix}, \quad (1)$$

for $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{2^{n-1}-1}, \theta_{2^{n-1}}, \ldots, \theta_{2^n-1})^{\mathrm{T}} \in \mathbb{R}^{2^n}$ where $n \in \mathbb{N} := \{1, 2, \ldots\}$. This diagonal unitary matrix directly corresponds to a unitary operation:

$$U_{\mathrm{D}(\boldsymbol{\theta})} = \sum_{k=0}^{2^n-1} e^{\mathrm{i}\theta_k} |k\rangle \langle k| ,$$

and we are interested in its quantum circuit. Moreover, we introduce a vector $\boldsymbol{\varphi}$ defined by

$$\boldsymbol{\varphi} \equiv (\varphi_0, \ldots, \varphi_{2^n-1})^{\mathrm{T}} := -\frac{2}{\sqrt{2^n}} \mathbf{H}\boldsymbol{\theta} \in \mathbb{R}^{2^n}, \quad (2)$$

where $\mathbf{H}$ is the Hadamard transform.

Previous results: Based on a natural gate set {CNOT, R_z}, $2^n \times 2^n$ diagonal unitary matrices can be precisely implemented using 2^n (multifold) z rotation gates whose rotation angles are given by $\boldsymbol{\varphi}$ (see [1, 2]). In a suitable order of the (multifold) z rotation gates, some CNOT gates can be canceled, and a quantum circuit of $2^n - 2$ CNOT gates and $2^n - 1$ phase rotation gates with a depth of at most $2^{n+1} - 3$ was obtained in [1, 3, 4]. Although the gate count remained the same, Zhang et al. [5] reduced the circuit depth to 2^n by commuting and parallelizing the quantum gates.

Our target: The previous results seem optimal if the components of $\boldsymbol{\theta}$ are completely independent. By introducing a so-called reflection symmetry, we aim at a nearly half reduction of the gate count, as well as the circuit depth.

2 Preliminary and main result

Definition 1 *We call that a unitary diagonal matrix* (1) *admits* **reflection symmetry** *if the given vector* $\boldsymbol{\theta} \in \mathbb{R}^{2^n}$ *satisfies the following condition:*

$$\theta_k = \theta_{2^n-1-k} \quad \textit{for all } k = 0, 1, \ldots, 2^n - 1. \quad (3)$$

Under the above symmetric structure of $\boldsymbol{\theta}$, the Hadamard transform extracts the necessary information, so that half of the rotation angles vanish in the case of reflection symmetry.

Lemma 2 *Let* $n \in \mathbb{N}$. *Assume that* $\boldsymbol{\theta} \in \mathbb{R}^{2^n}$ *admits reflection symmetry, and a vector* $\boldsymbol{\varphi}$ *is defined by Eq.* (2). *Then, we have* $\varphi_k = 0$ *for any* $k = 0, 1, \ldots, 2^n - 1$ *such that*

$$\sum_{m=0}^{n-1} k_m \mod 2 = 1.$$

Here, $[k_0 k_1 \cdots k_{2^n-1}]$ is the binary representation of k. As for detailed proof, we refer to Theorem 1 in Appendix A of [6]. Equipped with the above lemma, we can prove the main result.

* kkou@quemix.com

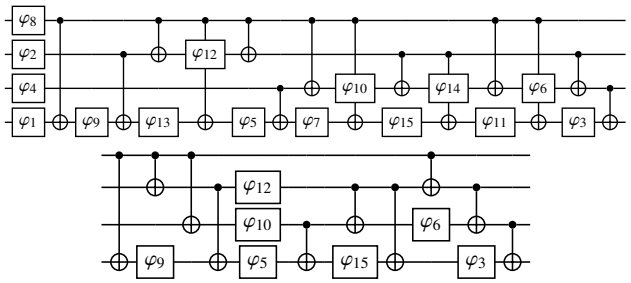


Figure 1: The comparison between the general algorithm in [5] (upper circuit) and our specifically designed algorithm for reflection symmetry (lower circuit) in the case of $n = 4$. Each square denotes a phase rotation gate with its rotation angle inside.

Theorem 3 *Any unitary operation corresponding to a diagonal unitary matrix that admits reflection symmetry can be implemented, up to a global phase, by a quantum circuit with* $2^{n-1} - 1$ *phase rotation gates and* $2^{n-1} + n - 2$ CNOT *gates with a depth of at most* $2^{n-1} + 2^{n-3}$. *Here, the depth is considered based on the gate set* $\{\text{CNOT}, R_z\}$.

The proof follows from a constructive algorithm, and we refer the technical details to Algorithm 1 and Appendix C of [6].

For a demonstration, we compare the quantum circuits using the general algorithm in [5] and our specifically designed algorithm (Algorithm 1 in [6]) for reflection symmetry in Fig. 1. We find that the CNOT count and the depth are reduced to 10 in the case of $n = 4$.

3 Application to real-time evolution in first-quantized Hamiltonian simulation

For the first-quantized Hamiltonian simulation, there is a well-known grid-based method that uses the so-called centered/shifted quantum Fourier transform U_{CQFT} to diagonalize the kinetic energy operator $\hat{T}$. Since the potential energy operator $\hat{V}$ is already diagonal in the real-space representation, one combines this with the Trotter-Suzuki formula to obtain an approximation scheme [7, 8, 9, 10]. For example, the first-order Trotter-Suzuki formula gives the following approximation for K steps:

$$\begin{aligned} e^{-\mathrm{i}\mathcal{H}K\Delta t} &= (e^{-\mathrm{i}\mathcal{H}\Delta t})^K \\ &\approx (e^{-\mathrm{i}\hat{T}\Delta t} e^{-\mathrm{i}\hat{V}\Delta t})^K \\ &= \left(U_{\text{CQFT}} U_{\text{kin}}(\Delta t) U_{\text{CQFT}}^{\dagger} U_{\text{pot}}(\Delta t)\right)^K, \end{aligned}$$

where $U_{\text{kin}}(\Delta t), U_{\text{pot}}(\Delta t)$ are two unitary operations corresponding to two diagonal unitary matrices with the parameter Δt.

We provide two illustrative examples of first-quantized Hamiltonian simulations in one dimension, for which the proposed quantum circuit can be applied.

Example 1 *One-particle simulation with Eckart barrier potential*

In this example, we have $\hat{V}|\psi\rangle = v_{\text{eck}}(|x - r_0|)|\psi\rangle$, where $v_{\text{eck}}(x) = A\,\text{sech}(ax)$ with two parameters A and a. Let L be the length of the simulation cell and $x_k = k\Delta x - L/2$, $k = 0, 1, \ldots, 2^n - 1$ be the grid points, where $\Delta x = L/2^n$. Then, we find that the diagonal unitary matrix $U_{\text{pot}}(\Delta t)$ subjects to a vector $\boldsymbol{\theta}$ with

$$\theta_k = -v_{\text{eck}}(|x_k - r_0|)\,\Delta t = -v_{\text{eck}}\left(\left|k - 2^{n-1} + 1/2\right|\Delta x\right)\Delta t,$$

for $k = 0, 1, \ldots, 2^n - 1$. Here, we have chosen $r_0 = (-1/2)\Delta x = -L/2^{n+1}$ so that $\boldsymbol{\theta}$ satisfies Eq.(3). Therefore, $U_{\text{pot}}(\Delta t)$ admits reflection symmetry.

Example 2 *Two-particle simulation in an electric field*

In this example, we consider a LiH molecule simulation where the two electrons are bound to the corresponding Li or H nucleus separately at the beginning. From $t = 0$, we impose an electric field and simulate the evolution of the two-electron system. $\hat{V}$ is composed of four parts: electron-nucleus, electron-electron, nucleus-nucleus, and external potentials, among which the particle-particle interactions are modeled by modified Coulomb potentials, and the external potential of a static electric field is described by $v_{\text{ext}}(x) = -\omega_0 x$. Here, we focus on the electron-electron potential v_{ee}. Using the same notations L, Δx, and x_k, we find that the unitary operation $e^{-\mathrm{i}\hat{V}_{\text{ee}}\Delta t}$ subjects to a vector $\boldsymbol{\theta} \in \mathbb{R}^{2^{2n}}$ with

$$\theta_k \equiv \theta_{2^n j + j'} = -v_{\text{ee}}(|x_j - x_{j'}|)\Delta t = -v_{\text{ee}}(|j - j'|\Delta x)\Delta t,$$

for $k = 0, 1, \ldots, 2^{2n} - 1$. By noting that $k \leftrightarrow (j, j')$ implies $2^{2n} - 1 - k \leftrightarrow (2^n - 1 - j, 2^n - 1 - j')$, again we conclude that the diagonal unitary matrix regarding the electron-electron interaction admits reflection symmetry.

Moreover, we provide the simulation results in Figs. 2,3 for the above examples using quantum emulator Qiskit [11]. For Example 1, we choose $n = 10$ and $\Delta t = 0.1$, and we find that the infidelity between the simulated wave function using Qiskit and the "exact" one (calculated by diagonalizing the discretized Hamiltonian matrix) is less than 0.05. Compared to the circuit using the previous work, we also find about 39% reduction in the CNOT count and 32% reduction in the depth by applying Algorithm 1. For Example 2, we choose $n = 7$ and $\Delta t = 0.1$ for the simulation, and the infidelity between the simulated wave function and the "exact" one is less than 0.001 for the case of a strong electric field,

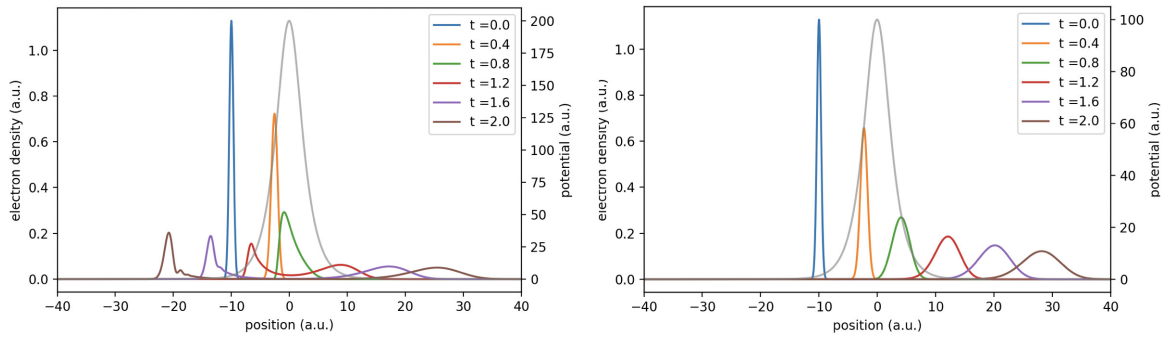


Figure 2: Hamiltonian simulation with different strength of the barrier at time points $t = 0, 0.4, 0.8, 1.2, 1.6, 2.0$ with grid size $n = 10$ and time step $\Delta t = 0.1$. In each panel, the left vertical axis represents the electron density, and the right vertical axis represents the Eckart barrier potential in gray. The left panel shows the case with a larger strength $A = 200$ where one observes that a partial wave penetrates the barrier while a partial wave returns. The right panel shows the case with a weaker strength $A = 100$ in which only penetration is observed.

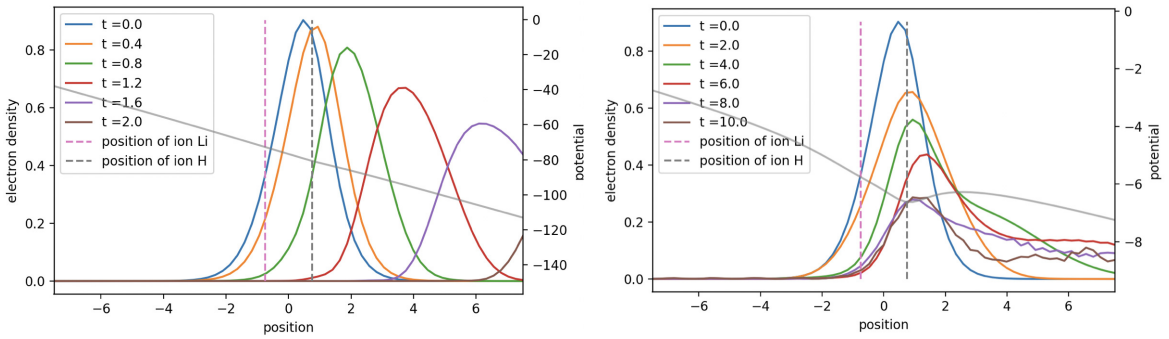


Figure 3: Hamiltonian simulation with grid size $n = 7$, time step $\Delta t = 0.1$, and different electric fields. In each panel, the left vertical axis represents the electron density of the two-electron system, and the right vertical axis represents the potential due to nuclei in gray. The left panel shows the fast decay of the electron density in the virtual domain of the molecule with a strong electric field of parameter $\omega_0 = 5$, and the probability of electrons remaining in the molecule is very small after $t > 2.0$. The right panel demonstrates the case with a weak electric field of parameter $\omega_0 = 0.3$, in which we observe that the electrons remain in the molecule with high probability even at $t = 10.0$.

and is less than 0.05 for the case of a weak electric field. Compared to the previous work, applying Algorithm 1 achieves about 48% reduction in the CNOT count and 37% reduction in the depth. The details on the numerical simulation can be found in Sect. 5 of [6].

4 Concluding remarks

The quantum circuit for an arbitrary $2^n \times 2^n$ diagonal unitary matrix is known to have $2^n - 2$ CNOT gates and $2^n - 1$ phase rotation gates with a depth of at most 2^n. Under the assumption of reflection symmetry, that is, Eq. (3), we provide a specifically designed circuit that is constructed by $2^{n-1} + n - 2$ CNOT gates and $2^{n-1} - 1$ phase rotation gates with a depth of at most $2^{n-1} + 2^{n-3}$, based on the gate set $\{\text{CNOT}, R_z\}$.

Diagonal unitary matrices with reflection symmetry serve as subroutines in applications, one of which is the quantum circuit for the potential part (either a symmetric or an interaction potential) in the Hamiltonian simulation illustrated above. Although the given examples consider only the one-dimensional case, our specifically designed circuit achieves a nearly half reduction in gate count and depth compared to the general one even in a high-dimensional case.

References

[1] N. Schuch. Implementation of quantum algorithms with Josephson charge qubits. PhD Thesis, Universität Regensburg, Germany (2002).

[2] N. Schuch and J. Siewert. Programmable networks for quantum algorithms. Phys. Rev. Lett. **91**, 027902 (2003).

[3] S. S. Bullock and I. L. Markov. Asymptotically optimal circuits for arbitrary n-qubit diagonal computations. Quantum Inf. Comput. **4**, 27 (2004).

[4] J. Welch, D. Greenbaum, S. Mostame, and A. Aspuru-Guzik. Efficient quantum circuits for diagonal unitaries without ancillas. New J. Phys. **16**, 033040 (2014).

[5] S. Zhang, K. Huang, and L. Li. Depth-optimized quantum circuit synthesis for diagonal unitary operators with asymptotically optimal gate count. Phys. Rev. A. **109**, 042601 (2024).

[6] X. Huang, T. Kosugi, H. Nishi, and Y. Matsushita. Optimized synthesis of circuits for diagonal unitary matrices with reflection symmetry. J. Phys. Soc. Jpn. **93**, 054002 (2024). arXiv:2310.06676

[7] I. Kassal, S. P. Jordan, P. J. Love, M. Mohseni, and A. Aspuru-Guzik. Polynomial-time quantum algorithm for the simulation of chemical dynamics. Proc. Natl. Acad. Sci. U.S.A. **105**, 18681 (2008).

[8] A. M. Childs, J. Leng, T. Li, J.-P. Liu, and C. Zhang. Quantum simulation of real-space dynamics. Quantum **6**, 860 (2022).

[9] H. H. S. Chan, R. Meister, T. Jones, D. P. Tew, and S. C. Benjamin. Grid-based methods for chemistry simulations on a quantum computer. Sci. Adv. **9**, eabo7484 (2023).

[10] T. Kosugi, H. Nishi, and Y. Matsushita. Exhaustive search for optimal molecular geometries using imaginary-time evolution on a quantum computer. npj Quantum Inf. **9**, 112 (2023).

[11] Qiskit contributors. Qiskit: An Open-source Framework for Quantum Computing. 2019. https://doi.org/10.5281/zenodo.2562111 (Accessed: 2024-03-07).

Potentials and Limitations of Analog Quantum Simulators in Variational Quantum Algorithms

Kasidit Srimahajariyapong * Supanut Thanasilp † Thiparat Chotibut ‡

1 Summary

Significant progress has been made in studying variational quantum algorithms (VQAs) [1, 2] using the digital-gate-based approach, where quantum dynamics are constructed through a sequence of quantum gates. However, this method requires precise control, posing practical challenges for near-term quantum device implementation. Alternatively, analog quantum simulators leverage the inherent quantum dynamics of the system to mimic the quantum dynamics of interest, requiring much less control and being more resilient to noise [3, 4]. Consequently, analog quantum simulation is one of the promising candidates for exploring quantum advantage in near-term practical quantum devices, yet it is relatively less explored in the context of VQAs [5, 6, 7, 8].

This work bridges the gap between digital-gate-based and analog quantum simulators by investigating the scalability of VQAs implemented with analog quantum simulators from the fundamental aspects of universality, expressivity, and untrainability. Specifically, we consider quench dynamics generated by a disordered Ising spin chain, which can operate in different quantum phases, as a key case study. We compare the VQA ansatz initialized in two distinct phases of matter—the thermalized and many-body localized (MBL) phases [9]. Our findings suggest a regime where initializing the ansatz in the MBL phase offers improved trainability and a sufficient number of parameters for achieving the desired unitary during the optimization process.

2 Framework

Variational quantum algorithms. We compute the cost function, defined as the expectation value of an observable O which defines a problem of interest: $C(\boldsymbol{\Theta}) = \mathrm{Tr}(OU(\boldsymbol{\Theta})\rho_0 U^\dagger(\boldsymbol{\Theta}))$ where ρ_0 is some initial state, $U(\boldsymbol{\Theta})$ is a parametrized unitary (also called an ansatze) on n qubits with $\boldsymbol{\Theta}$ as a set of trainable parameters. The training procedure typically consists of estimating the loss with quantum computers and classically optimize in a variational manner, ultimately aiming to obtain parameter values such that the loss function is minimized $\boldsymbol{\Theta}_{\min} = \mathrm{argmin}_{\boldsymbol{\Theta}} C(\boldsymbol{\Theta})$.

Analog quantum simulator. The explicit form of the ansatze significantly contributes to the success of VQAs. While much advancements have been made for implementing digital-gate-based quantum circuits, here we consider an alternative whereby an analog quantum is used as an ansatze. In particular, we consider a quantum quench dynamic of the form

$$U(\boldsymbol{\Theta}) = \prod_{m=1}^{M} e^{-i\theta_m^{(0)} H(\boldsymbol{\theta}_m)} , \tag{1}$$

where M is the total number of quenches, $\theta_m^{(0)}$ is an evolution time of the m^{th} quench (which can also be made trainable) and $H(\boldsymbol{\theta}_m)$ is a native Hamiltonian with coefficients $\boldsymbol{\theta}_m$ as trainable parameters.

In our study, we consider a disordered Ising spin chain with nearest-neighbor interactions

$$H(\boldsymbol{\theta}) = J\sum_{\langle i,j\rangle} Z_i Z_j + h\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \theta_i Z_i , \tag{2}$$

where J is a coupling strength, h is a uniform transverse field strength, $\{\theta_i\}$ are on-site disordered energies in which each is uniformly drawn from the interval $[-W/2, W/2]$ with the disorder strength W, and X_i and Z_i are the Pauli X and Z matrices for the i^{th} site, respectively. This Hamiltonian serves as a simplified version without long-range interactions of the trapped ion Hamiltonian [10] and hence constitutes an ideal playground for studying fundamental aspects of VQAs with analog quantum simulators.

Depending on the parameter regime, this Hamiltonian is capable of exhibiting different quantum phases of matter. For a weak disorder limit, the system is in a thermalized phase, being chaotic and enjoying an ergodicity for a sufficiently long evolution time. On the other hand, strong disorder can prevent the system from thermalization, leading to the many-body localized (MBL) phase. To identify the regime where the system falls into which phase under the given Hamiltonian, we numerically probe its level-spacing statistics which is a standard diagnostic tool in many-body quantum physics [11] (see Appendix A for further details).

3 Main Results

Universality. The first fundamental aspect that we investigate is whether the quantum quench dynamics described in Eq. (1) with the Hamiltonian in Eq. (2) can realize an arbitrary unitary. This is of particular interest when we employ the analog simulators to solve a problem that may not align with the native quantum evolution. In other words, in this scenario, $U(\boldsymbol{\Theta})$ in Eq. (1) acts as a problem-agnostic ansatze.

*kasidit.quantum@gmail.com
†supanut.thanasilp@gmail.com
‡thiparatc@gmail.com

By using a constructive method described in Ref. [12], we argue that, upon appropriately adjusting the on-site transverse and longitudinal fields and evolution time, the quench dynamics can be engineered to approximate any quantum evolution. This result is also in agreement with Dynamical Lie Algebra perspective in Ref [13].

Expressivity. Intuitively, expressivity measures how uniformly an unitary ensemble generated by varying the trainable parameters covers the whole unitary group. That is, it indicates how close the unitary distribution form the ensemble is to the Haar distribution. One approach to quantify expressivity is to compute the distribution of probabilities in computational basis (for a given intial state) and see how far it is from Porter-Thomas distribution (which is a theoretical prediction in the case of a Haar random state [1]). Here, we compute this difference using Kullback-Leibler Divergence (KLD).

Figure 1 shows KLD for both the thermalized and many-body localized (MBL) phases as the number of quenches increases. For both phases, the KLD eventually saturates, indicating that the expressivity of the ansatz reaches its maximum for a given number of qubits. The decay rate towards saturation is faster in the thermalized phase compared to the MBL phase, suggesting that the thermalized phase more rapidly approximates a uniform distribution over the possible output states.

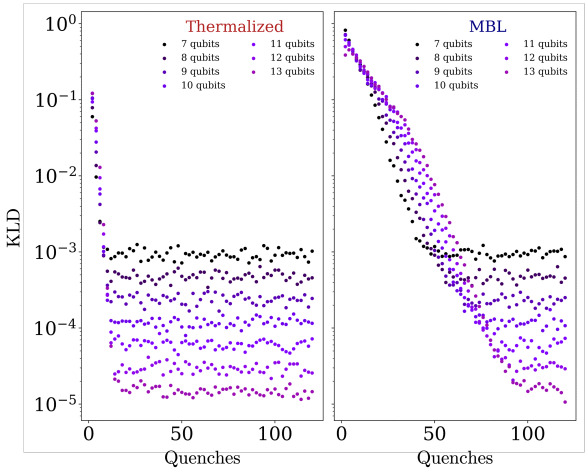


Figure 1: **The KL divergence saturates as the number of quenches increases in both the thermalized (left) and the MBL (right) initialization.** The parameters in the thermalized and the MBL phase are, respectively, $W = 5J$ and $W = 50J$. For each number of quenches, the KLD are averaged over 400 random initializations. Importantly, the MBL initialization takes many more quenches to achieve its maximal expressivity, given a fixed number of qubits.

Untrainability. Barren plataeu (BP) is a well-known trainability issue in VQAs where the gradients of cost function vanish exponentially with the number of qubits [14, 15, 16], causing an untrainable cost landscape as one scales up the system size. Since the variance of the cost function can also serve as an indicator for the BP [17], we explore this issue in our analog ansatz initialized in either the thermalized or the MBL phase by computing the variance of a local observable, $\langle Z_1 Z_2 \rangle$.

Figure 2 compares the behaviors of vanishing variances for the thermalized (left) and the MBL (right) phases. The thermalized phase exhibits a faster decay in variance as the number of quenches increases. However, with a sufficient number of quenches, both phases eventually reach the same saturated variance. This saturated variance is further investigated in Fig. 10, revealing an exponential decay in the number of qubits for both phases, a key signature of BP. Thus, both phases ultimately face the same trainability challenge: both suffer from BP. Additionally, as shown in Fig. 9, the saturation of KLD correlates with the onset of the barren plateau (BP) in the trainability analysis of Fig. 2, in agreement with the observation in Ref. [18]. Next, we further study the initialization strategy in the MBL phase, as the saturated-variance-regime arises later. We shall exploit this non-saturated regime with higher expressivity to mitigate BP issues.

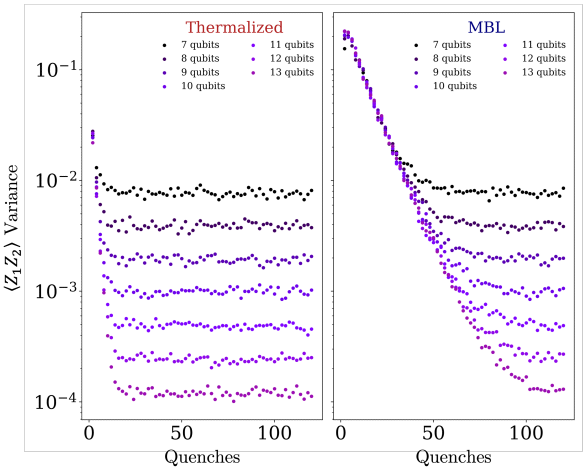


Figure 2: **The onset of barren plateaus in the thermalized (left) and the MBL (right) initialization.** The variance of $\langle Z_1 Z_2 \rangle$ of the thermalized ($W = 5J$) and MBL ($W = 50J$) initialization, averaged over 400 random initializations, is plotted against the number of quenches. The onset of barren plateaus arises more rapidly in the thermalized initialization than in the MBL initialization. This onset also correlates with the saturation point of KLD in Fig.1, see a detailed comparison in Fig. 9.

Initialization strategy. Building on the results in the previous sections, we conclude that for a given number of quenches (i.e., fixed M) there exists three different regimes for initialization

- Regime I (*small* number of quenches): Both thermalized and MBL phases are barren plateau free but are not maximally expressive.

[1] More precisely, the Porter-Thomas distribution serves as a benchmark, representing the distribution of the state $|0\rangle\langle 0|$ after evolution under a random Haar unitary.

- Regime II (*intermediate* number of quenches): The thermalized phase becomes maximally expressive but suffers from a barren plateau. On the other hand, the MBL phase has large loss differences at the initial training step and becomes more and more expressive (compared to Regime I).

- Regime III (*large* number of quenches): Both phases are maximally expressive but untrainable.

These three different regimes can be quantitatively determined by the ratio Q of a variance of a local observable at quench M to the saturated variance (i.e., variance at very large M), as shown in Fig. 3.

In practice, employing the analog ansatze generally requires us to predetermine the number of quenches (i.e., M is treated a hyperparameter) unless the adaptive strategy is used (which is not considered in this work). From the observation mentioned above, it motivates us to propose an initialization strategy where the number of quenches in the ansatze is chosen such that $U(\boldsymbol{\Theta})$ is in Regime II within the MBL phase. This allows the ansatze to be trainable at the initial step while having large enough expressivity during later training iterations.

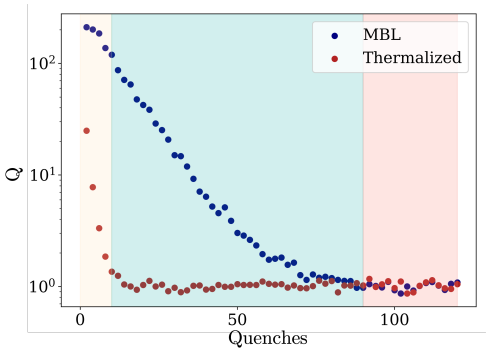


Figure 3: **Three initialization regimes** The ratio Q of variance to the saturated variance is plotted against the number of quenches for 10 qubits. The blue and red correspond to initialization in the MBL and thermalized phase, respectively. We categorize the initialization strategy into three regimes: regime I (yellow) in which both phases do not suffer from BP, regime II (cyan) in which the thermalized phase is untrainable (BP) while MBL phase still does not encounter MP, and regime III (red) in which both phases are untrainable. Thus, initialization in regime II of MBL phase can avoid BP while attaining relatively high expressivity at a larger number of quenches.

Benchmark on a toy example. To show that the proposed MBL initialization scheme can be trained to also express states that are not in an MBL phase, we benchmark the scheme on a variational quantum eigensolver task: finding a ground state of an instance of a long-range transverse field Ising model Hamiltonian

$$H_{\text{target}} = \sum_{i,j} J_{ij} Z_i Z_j + B \sum_i X_i + \sum_i h_i X_i, \quad (3)$$

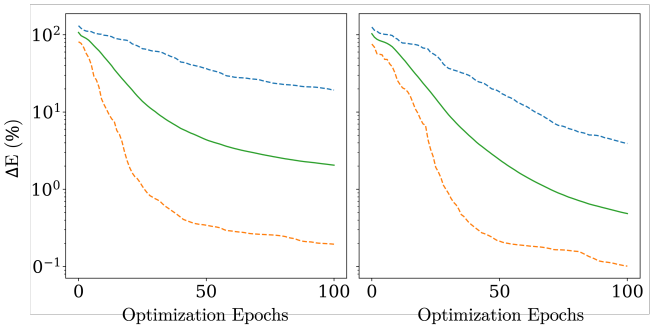


Figure 4: **VQE benchmark results for MBL initializations** The average relative error, comparing between the estimated ground state energy $\langle H_{target} \rangle_{\boldsymbol{\Theta}}$ and the actual ground state energy of H_{target}, is plotted in solid green lines on a logarithmic scale against the optimization epochs for quench dynamics initialized in the MBL phase over 100 realizations, using (left) 2 and (right) 6 quenches. The blue and orange dashed lines are the maximum and minimum relative error among all the realizations respectively. The standard derivation is 2.80% and 0.32% for 2-quench and 6-quench ansatze, respectively, which are unnoticeable on the logarithmic scale.

where $J_{ij} = \frac{1}{|i-j|}$, in a parameter regime such that the system is in the thermalized phase with 7 qubits. We employ a parameterized quench dynamics governed by the Hamiltonian in Eq. (2) initialized in the MBL phase as an ansatze and calculate the relative error between our trained model prediction and the true ground state energy ($\Delta E = \frac{E-E_0}{E_0}$). As shown in Fig. 4, after 100 training (optimization) epochs, we obtain an average relative error of 2.05% with a standard deviation of 2.80% using a 2-quench ansatze and an average relative error of 0.27% with a standard deviation of 0.32% using a 6-quench ansatze. Another example application of our analog VQA initialization strategy is to solve a Max-Cut problem, see the results in Appendix B.

4 Discussion

We study fundamental aspects of VQAs with analog quantum simulators. By considering a disordered Ising spin chain as an example, we study how expressivity and barren plateaus are fundamentally linked with the phases of matter. While maximal expressivity can be achieved in both phases through quench dynamics, the expressivity-induced barren plateaus have also arisen in this analog setting, similar to the situation in the digital-circuit-based VQAs counterpart. From the study of our analog ansatze, we propose a novel initialization strategy based on the MBL phase whereby the ansatze is trainable at the initial step and has sufficiently large expressivity to solve interesting problems. Our work bridges the gap between fundamental aspects of VQAs and quantum phases of matter.

References

[1] Dave Wecker, Matthew B. Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Physical Review A*, 92:042303, Oct 2015.

[2] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(1):625–644, 2021.

[3] Andrew J Daley, Immanuel Bloch, Christian Kokail, Stuart Flannigan, Natalie Pearson, Matthias Troyer, and Peter Zoller. Practical quantum advantage in quantum simulation. *Nature*, 607(7920):667–676, 2022.

[4] Alicia B Magann, Christian Arenz, Matthew D Grace, Tak-San Ho, Robert L Kosut, Jarrod R McClean, Herschel A Rabitz, and Mohan Sarovar. From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms. *PRX Quantum*, 2(1):010101, 2021.

[5] Jirawat Tangpanitanon, Supanut Thanasilp, Ninnat Dangniam, Marc-Antoine Lemonde, and Dimitris G Angelakis. Expressibility and trainability of parametrized analog quantum systems for machine learning applications. *Physical Review Research*, 2(4):043364, 2020.

[6] Oinam Romesh Meitei, Bryan T Gard, George S Barron, David P Pappas, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. Gate-free state preparation for fast variational quantum eigensolver simulations. *npj Quantum Information*, 7(1):155, 2021.

[7] Antoine Michel, Sebastian Grijalva, Loïc Henriet, Christophe Domain, and Antoine Browaeys. Blueprint for a digital-analog variational quantum eigensolver using rydberg atom arrays. *Physical Review A*, 107(4):042602, 2023.

[8] Robert De Keijzer, Oliver Tse, and Servaas Kokkelmans. Pulse based variational quantum optimal control for hybrid quantum computing. *Quantum*, 7:908, 2023.

[9] Dmitry A Abanin, Ehud Altman, Immanuel Bloch, and Maksym Serbyn. Colloquium: Many-body localization, thermalization, and entanglement. *Reviews of Modern Physics*, 91(2):021001, 2019.

[10] Jacob Smith, Aaron Lee, Philip Richerme, Brian Neyenhuis, Paul W Hess, Philipp Hauke, Markus Heyl, David A Huse, and Christopher Monroe. Many-body localization in a quantum simulator with programmable random disorder. *Nature Physics*, 12(10):907–911, 2016.

[11] Vadim Oganesyan and David A Huse. Localization of interacting fermions at high temperature. *Physical review b*, 75(15):155111, 2007.

[12] Adrian Parra-Rodriguez, Pavel Lougovski, Lucas Lamata, Enrique Solano, and Mikel Sanz. Digital-analog quantum computation. *Physical Review A*, 101(2):022305, 2020.

[13] Roeland Wiersema, Efekan Kökcü, Alexander F Kemper, and Bojko N Bakalov. Classification of dynamical lie algebras for translation-invariant 2-local spin systems in one dimension. *arXiv preprint arXiv:2309.05690*, 2023.

[14] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812, 2018.

[15] Michael Ragone, Bojko N Bakalov, Frédéric Sauvage, Alexander F Kemper, Carlos Ortiz Marrero, Martin Larocca, and M Cerezo. A unified theory of barren plateaus for deep parametrized quantum circuits. *arXiv preprint arXiv:2309.09342*, 2023.

[16] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J Coles, Lukasz Cincio, Jarrod R McClean, Zoë Holmes, and M Cerezo. A review of barren plateaus in variational quantum computing. *arXiv preprint arXiv:2405.00781*, 2024.

[17] Andrew Arrasmith, Zoë Holmes, Marco Cerezo, and Patrick J Coles. Equivalence of quantum barren plateaus to cost concentration and narrow gorges. *Quantum Science and Technology*, 7(4):045015, 2022.

[18] Zoë Holmes, Kunal Sharma, Marco Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3(1):010313, 2022.

Appendix

A Level-spacing Statistics

Here we provide the parameter setting for our model and more detailed description of the level-spacing statistics that characterize the thermalized and MBL phases in the main text. We consider the statistics of the energy level spacing r_i, defined as

$$r_i = \frac{\min(\Delta_i, \Delta_{i+1})}{\max(\Delta_i, \Delta_{i+1})}, \tag{4}$$

where Δ_i is the difference between the i^{th} and the $(i-1)^{th}$ energy level.

Specifically, we focus on the Hamiltonian in Eq. (2) using a uniform transverse field strength of $-2J$. The system governed by this Hamiltonian is in the thermalzied phase when the disorder strength $W = 5J$, characterized by the level-spacing statistics that follows that of the Gaussian Orthogonal Ensemble (GOE) shown in the right panel of Fig. 5. The GOE signifies the level repulsion, a characteristic feature of the thermalized phase. In contrast, at strong disorder, with $W = 50J$, the level spacing follows a Poisson distribution (left panel of Fig. 5), a characteristic of the MBL phase.

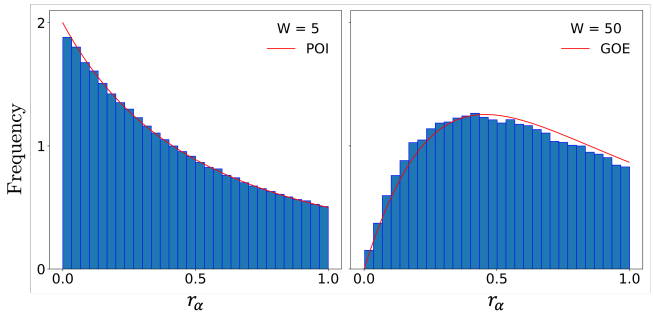


Figure 5: **Level-spacing statistics of disordered Ising model in a weak and a strong disorder strength** The histograms represent the level-spacing statistics for 10-qubits systems governed by the Hamiltonian in Eq. (2) (Right) with $W = 5$, the histogram follows the GOE statistic, indicative of the system being in the thermalized phase. (Left) Conversely, with $W = 50$, the histogram follows the Poisson statistics, indicative of the system being in the MBL phase. These different disordered strengths are used for initializing the parameters in our ansatze for the thermalized and the MBL initialization.

B Quadratic Unconstrained Binary Optimization: Max-Cut

We present another example application of our ansatze by solving a Max-Cut problem, a prototypical Quadratic Unconstrained Binary Optimization (QUBO) problem. This problem involves finding the maximum cut in a graph, which entails dividing the graph's vertices into two sets such that the total weight of the edges connecting vertices across these two sets is maximized. The QUBO Hamiltonian associated with such problem can be expressed as

$$H = \sum_{i>j} w_{ij} Z_i Z_j, \tag{5}$$

where w_{ij} is the weight of the edge connecting vertices i and j. As a demonstration, we consider a specific graph instance represented in Fig. 6. This graph has multiple Max-Cut solutions where vertices can be partitioned into sets {2,5} and {1,3,4}, or {1,4} and {2,3,5}. The ground state configuration of this problem Hamiltonian thus has four-fold degeneracy. As shown in Fig 7, we achieve the average approximation ratio of 99.59% with a standard deviation of 2.30% for the 2-quench case. This result identifies three configurations above the selection probability threshold ($p > 0.01$), which correspond to the degenerate ground state configuration. The performance is slightly better in the 6-quench case, with an average approximation ratio of 99.61% and a standard deviation of 2.27%. Additionally, it successfully identifies all four degenerate ground states given the same threshold.

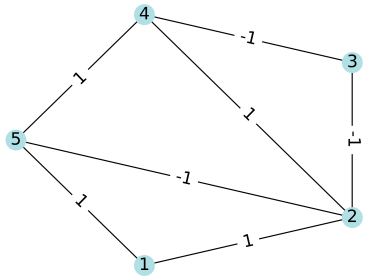


Figure 6: **Weighted graph for a max-cut problem** An instance of a graph for a max-cut problem with 5 vertices and 7 edges whose weights are either -1 and 1.

C More Realistic Model Hamiltonian

We explore the trainability of an analog VQAs in a more realistic model with long-range interactions captured by the Hamiltonian

$$H_{long-range} = J \sum_{i>j} \frac{Z_i Z_j}{|i-j|^\alpha} + B \sum_i X_i + \sum_i g_i X_i, \tag{6}$$

where J is the coupling strength, α controls the strength of long-range interaction, B represents a uniform effective transverse field, and g_i is an on-site disorder potential. In this model, for $\alpha = 1$ and $B = 0$, the system is in the thermalized phase when the on-site disorder is drawn uniformly from the interval $[-0.3J, 0.3J]$, and in the MBL phase when drawn uniformly from $[-7.5J, 7.5J]$. As illustrated in Fig. 8, we observe that the variances of the cost function saturate at the same number of quenches in both phases. Unfortunately, in this model, there does not appear to be an advantageous regime where MBL initialization is more preferable, unlike in the model studied

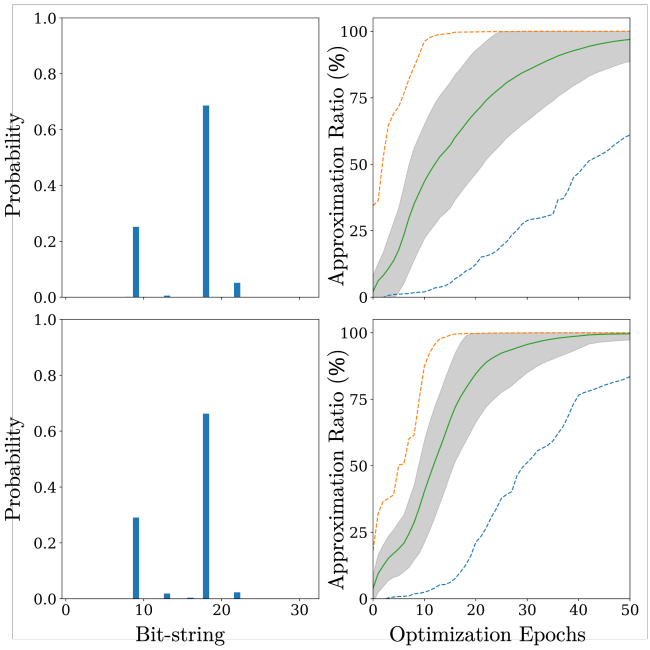


Figure 7: **Max-Cut optimization results** Probability distribution of computational bases from the output state after optimization with 50 epochs and the average approximation ratio plotted against optimization epoch, averaged over 100 realizations, are shown for a 2-quenches ansatze (top) and a 6-quench ansatze (bottom). For approximation ratio plot, the green solid line represents the average value, surrounded by a grey shade representing the error bars upper bounded by the maximum approximation ratio. The blue and orange dashed lines indicate the minimum and maximum approximation ratios among all the realizations, respectively.

in the main text. Thus, a preferable 'Regime II' for the MBL initialization strategy may not generally apply to other analog models that contain both MBL and thermalized phases.

D Additional Figures

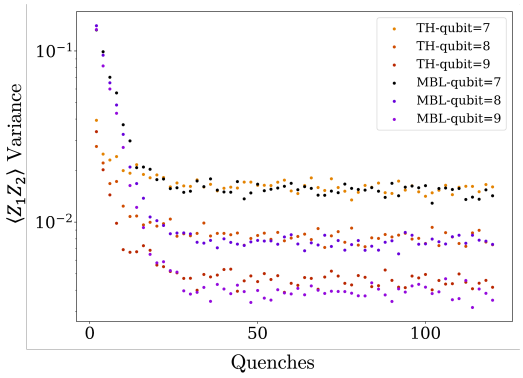


Figure 8: **The emergence of barren plateaus as the number of quenches increase in the long-range Hamiltonian of Eq. (6) in the thermalized and the MBL initialization** The variance of $\langle Z_1 Z_2 \rangle$ of the (left) thermalized and (right) MBL initialization is plotted against the number of quenches for 7-9 qubits averaged over 400 realizations. In this more realistic Hamiltonian, initialization in both phases requires approximately the same number of quenches to reach the onset of the barren plateaus.

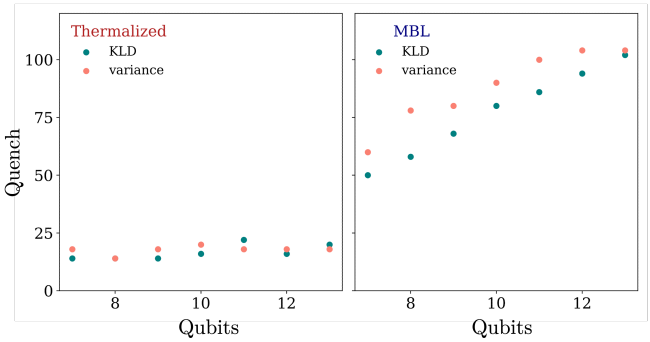


Figure 9: **The correspondence between the minimum number of quenches to saturate KLD (attaining ansatz maximal expressivity) and to saturate $\langle Z_1 Z_2 \rangle$ variance (the onset of BP)** For each number of qubits, The number of quenches when the variance saturates (orange) and when the KL divergence saturates (blue) are plotted. (Left) and (right) are for the thermalized and the MBL initialization, respectively.

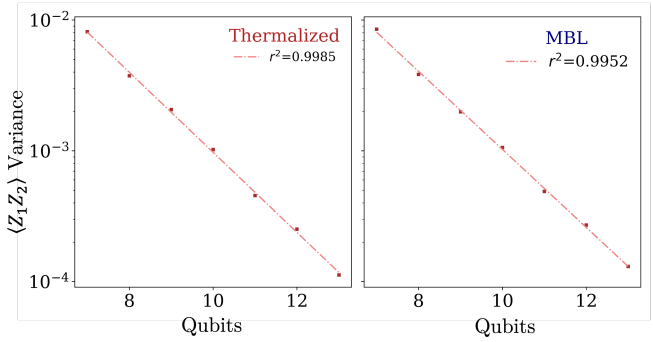


Figure 10: **Exponential decay of variance** The variance of $\langle Z_1 Z_2 \rangle$ is plotted against the number of qubits in the logarithmic scale in the y-axis for (left) thermalized ($W = 5J$) and (right) MBL ($W = 50J$) initializations. The number of quenches taken here is 120. This is the evidence of the barren plateaus exhibited in both phases in the long time limit.

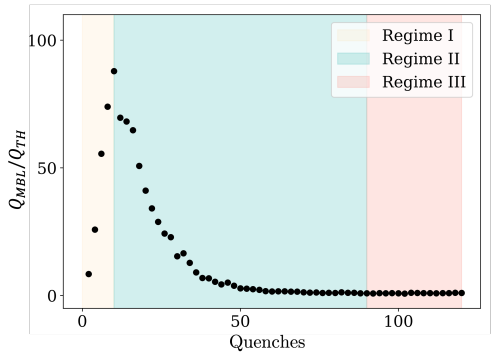


Figure 11: **Three initialization regimes** The ratio of Q in the MBL initialization to Q in the thermalized initialization is plotted against the number of quench for the 10-qubits system.

Robust Lindbladian Tomography with Error Amplification

Takanori Sugiyama[1 2 *]

[1] *Quantum Laboratory, Fujitsu Limited. Nakahara-ku, Kawasaki, Kanagawa 211-8588, Japan.*
[2] *RIKEN RQC-FUJITSU Collaboration Center, RIKEN. Wako, Saitama 351-0198, Japan.*

Abstract. Precise characterization of noisy quantum operations implemented plays an important role for realizing further accurate operations. Quantum tomography is a popular class of characterization methods, and its advanced methods like Gate-set tomography (GST) use error amplification circuit (EAC), a repetition of a sequence of quantum gates, for increasing their estimation precision on gates. Although GST has high precision, it suffers from highly nonlinear numerical optimization due to nonlinearity of EAC, which increases numerical cost and instability of GST. In order to overcome the GST's numerical problem, here we propose a new tomographic method for Lindbladian error of quantum gates with EAC. First, we develop new theoretical tools for analyzing effects of EAC on the Lindbladian error in arbitrary finite-dimensional system, which takes non-commutativity between different gates or between ideal and error parts of a gate, periodic properties of ideal gates, and repetition of gate sequence into consideration within a first order approximation. With the approximation, the numerical optimization of the proposed method reduces to positive semi-definite program (SDP). Therefore, compared to GST, the optimization problem is solvable more efficiently and stably, although its numerical cost grows exponentially with respect to the number of qubits, which is the same as other tomographic methods including GST. Second, we evaluate the performance of our method by numerical experiments on 1-qubit system. The result clearly shows an improvement of estimation precision by use of EAC. These results indicate that, even though we used a first order approximation, our method practically works well with numerical efficiency and stability better than GST.

Keywords: Quantum Tomography, Error Amplification, Lindbladian

1 Introduction

Further improvement of elemental quantum operations' accuracy is an inevitable task for realizing practical quantum computer. Characterization methods such as quantum tomography and randomized benchmarking are used for improving the accuracies, and take a role to obtain information of errors of the operations. Tomographic methods are suitable for obtaining detailed information of the errors, but its standard protocols suffer from not-negligible systematic errors originated from mismatch of our model values on states and measurements, which is called state-preparation-and-measurement (SPAM) error. Error amplification circuit (EAC) consists of a repetition of a sequence of quantum gates (Fig. 1). It is used to suppress effects of such SPAM error on estimation result in advanced tomographic methods such as Gate-set tomography (GST) [1], idle tomography (IT) [2], and Hamiltonian-Error Amplifying tomography (HEAT) [3].

GST has high precision, but it suffers from highly nonlinear numerical optimization due to nonlinearity of EAC, which increases numerical cost and instability. There are several approaches to make the numerical optimization simpler. We focus on characterization of gates, although GST treats all of states, measurements, and gates. In IT estimation object is limited to identity gates, and in HEAT it is limited to cross resonance gates with a specific error model. However, there are other elemental quantum gates, and actual errors are not necessarily included in the specific error model. So we need another tomographic method that is applicable to wider class of gates and error models compared to IT and HEAT and whose numerical efficiency and stability are better than GST.

2 Notation and Settings

Let us consider arbitrary finite d dimensional system. Let $\rho \in \mathbb{C}^{d\times d}$ and $\mathbf{\Pi} = \{\Pi_x\}_x$ denote density

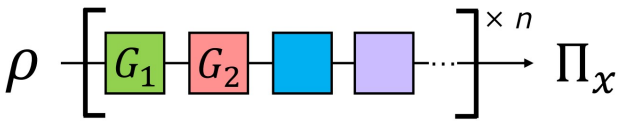


Figure 1: Quantum circuit diagram of an error amplification circuit. The superscript, "$\times n$", means n times repetition of the braketed gate sequence.

*sugiyama-taka@fujitsu.com

matrix and POVM, respectively. Let $\mathcal{G}$ denote a linear trace-preserving and completely-positive map representing action of a quantum gate. Let $\boldsymbol{B}$ denote an orthonormal matrix basis on $\mathbb{C}^{d\times d}$. Let $|\rho\rangle\rangle$ and G denote the matrix vectorization of ρ and matrix representation of $\mathcal{G}$ w.r.t. $\boldsymbol{B}$. When the gate $\mathcal{G}$ is implemented by dynamics of Lindblad master equation, G is represented in the following form,

$$G = e^{L^{\text{ideal}}+\delta L}, \tag{1}$$

where L^{ideal} is the matrix representation of the ideal Lindbladian (accumulated over finite time) of the gate, and δL is an Lindbladian error.

For a given ideal gate $\mathcal{G}^{\text{ideal}}$, if there exists a positive integer k satisfying

$$\left[\mathcal{G}^{\text{ideal}}\right]^k = \mathcal{I}, \tag{2}$$

we call the smallest k the period of the gate. For example, typical quantum gates like X90 and ZX90 has period $k = 4$. We assume that the all ideal gates in the EAC has periods. Then any gate sequence consisting the gates has a period.

In the EAC experiment depicted by Fig. 1, let ρ, $\boldsymbol{\Pi}$, G denote actual (implemented) quantum operations that are possibly noisy. When the superscript, ideal is put on an object, it corresponds to the ideal counterpart, and we assume the ideal value is known. The difference between the actual and ideal values is its error. In our method, the Lindbladian errors of some gates specified by an user is the estimation object, and errors of the other objects (gates, states, measurements) are not.

The probability that we observe an outcome x at the experiment depicted by Fig. 1 is given as

$$p_x(\boldsymbol{\delta L}, n) = \langle\langle \Pi_x | \left[\cdots G_2 G_1\right]^n |\rho\rangle\rangle, \tag{3}$$

where $\boldsymbol{\delta L}$ denote a set of Lindbladian errors for gates in the EAC. The functionality of p_x w.r.t. $\boldsymbol{\delta L}$ is highly non-linear for large n. This non-linearity makes the numerical optimization for tomographic data-fitting of the model to data hard and unstable.

3 Ideas and Theoretical Results

We briefly explain ideas and theoretical results [4]. Suppose that we perform experiments with common EAC and different repetition numbers $\boldsymbol{n} = \{\boldsymbol{n}_j\}_j$. We choose the numbers satisfying

$$n_j = k \cdot m_j + r,\ j = 1, 2, \ldots, \tag{4}$$

where k is the period of the repetition unit of the EAC, and r is the residual independent of j. When Eq. (4) holds, the difference between actual and ideal values of probability observing an outcome x can be expanded as the following form,

$$p_x(\boldsymbol{\delta L}, n_j) - p_x^{\text{ideal}} = \sum_{t=0}^{\infty} (n_j)^t \cdot q_x^{(t)}(\boldsymbol{\delta L}), \tag{5}$$

where the ideal value

$$p_x^{\text{ideal}} := p_x(\boldsymbol{\delta L} = \boldsymbol{O}, n_j) \tag{6}$$

depends on r but is independent of j. So, its takes a common value over $n_j \in \boldsymbol{n}$. The zero-th term ($t = 0$) is the not-amplified term, and the others are amplified terms. We use Eqs. (5) and (6) for deriving our fitting model for an amplified term and calculating its corresponding counterpart from data.

3.1 Fitting Model

We introduce two linear approximations for deriving our fitting model: one is w.r.t. n, and the other is w.r.t. $\boldsymbol{\delta L}$. First, focus on the linearly amplified term $q_x^{(t=1)}(\boldsymbol{\delta L})$ in Eq. (5). Next, we expand $q_x^{(t=1)}$ w.r.t. $\boldsymbol{\delta L}$. The lowest order term depends linearly on $\boldsymbol{\delta L}$. Let $\mathcal{F}_x^{(1)}$ denote the linear function, which represents a (doubly) linearized action of amplification of EAC on Lindbladian errors. We derive mathematical tools for arbitrary finite d to analyze such linearized action of gate composition and repetition by combining an integral formula of matrix exponential derivative [5, 6], matrix perturbation theory [7], and matrix diagonalization (spectral decomposition) [8]. With the mathematical tools derived, we also give an algorithm for calculating the action of $\mathcal{F}_x^{(1)}$ for a given EAC with arbitrary finite depth. We use $\mathcal{F}_x^{(1)}$ as the fitting model to data.

The algorithm is applicable to a class of gates, which is wider than IT and HEAT, but is not applicable to some gates with singularity. Expansion of the applicability is a future work.

3.2 Data Pre-processing

Let $f_x(n)$ denote relative frequencies for an outcome x when we choose n repetitions for amplification. The law of large number guarantees that $f_x(n)$ converges to $p_x(\boldsymbol{\delta L}, n)$ as the amount of data goes to infinity. We assume that the amount of data is sufficiently large, and f_x is sufficiently close to p_x. Since our fitting model is not for p_x but for the approximated $q_x^{(1)}$, we have to calculate the corresponding counterpart from p_x. In order to do that, we use polynomial fitting w.r.t. $n_j \in \boldsymbol{n}$ and calculate an estimate, say $h_x^{(1)}$, of $q_x^{(1)}$. This calculation can be done by a linear inversion w.r.t. $f_x(n)$.

3.3 Tomographic Experiment and Numerical Optimization

For obtaining multiple information of δL, we need multiple EACs, ρs, and $\mathbf{\Pi}$s. Let a denote an index for the multiple settings. We consider a constraint (weighted) least-squares problem with notation update from $\mathcal{F}_x^{(1)}$ to $\mathcal{F}_x^{(a,1)}$ and $h_x^{(1)}$ to $h_x^{(a,1)}$,

$$\begin{aligned} \delta L^{\mathrm{RLT}} \quad := \quad & \mathrm{argmin}_{\boldsymbol{\delta L}} \sum_a w^{(a)} \left\{ \mathcal{F}_x^{(a,1)}(\boldsymbol{\delta L}) - h_x^{(a,1)} \right\}^2 \\ & s.t.\ \delta L \text{ are physical.} \end{aligned} \tag{7}$$

We call this robust Lindbladian tomography (RLT). By linearity of $\mathcal{F}^{(a,1)}$, this optimization reduces to SDP as the standard tomographic protocols [9].

4 Numerical Results

We implemented the RLT estimator in Eq. (7) in Python. We used SCS [10] as the numerical SDP solver and CVXPY [11] as a parser. We performed numerical experiments on 1-qubit system ($d = 2$) for evaluating the performance of RLT as the first step (theoretical results hold for any finite d).

The estimation object is the Lindbladian error dL of X90 gate, which has 12 ($= d^4 - d^2$) degrees of freedom. We assume that the ideal Z90 gate is available by the virtual-Z gate (vZ) protocol [12]. We chose two EACs, $[X90]^{\times n}$ ($n = 4, 8, 16, 32$) and $[X90 \cdot vZ90]^{\times n}$ ($n = 3, 6, 12, 24$). The ideal states are four Pauli eigenstates, $|X+\rangle, |Y+\rangle, |Z+\rangle, |Z-\rangle$, and the ideal measurements are X-, Y-, and Z-projective measurements. We modeled noisy states and measurements with their Lindbladian errors as

$$|\rho\rangle\rangle = e^{\delta L_s} |\rho^{\mathrm{ideal}}\rangle\rangle, \tag{8}$$

$$\langle\langle \Pi_x | = \langle\langle \Pi_x^{\mathrm{ideal}} | e^{\delta L_p}, \tag{9}$$

and chose all 12 ($= 4 \times 3$) combinations of ρ and $\mathbf{\Pi}$ for each EAC. An analysis of the support of the maps $\mathcal{F}_x^1$ reveals that 10 of 12 degrees of freedom of δL is amplified in this setting. We used ideal values ρ^{ideal} and $\mathbf{\Pi}^{\mathrm{ideal}}$ in Eq. (7), leading to SPAM error.

A Lindbladian error can be decomposed into two parts: One is for Hamiltonian error (subscripted with H) and the other is for dissipation (subscripted with D). We set same values for the size of the H-parts ($\|\delta L_H\| = \|\delta L_{s,H}\| = \|\delta L_{p,H}\|$ w.r.t. the Frobenius norm), which is because 1-qubit gates are used for state preparation and measurement in practice, and sweeped from 10^{-3} to 10^{-1}. So, in this setting, the size of the H-part of the SPAM errors is compatible with the size of the H-part of the estimation object. We fixed $\|\delta L_{s,D}\| = \|\delta L_{p,D}\| = 10^{-2}$ (1% error) and $\|\delta L_D\| = 10^{-3}$ (0.1% error). These values are chosen along with the recent experiments on superconducting quantum circuits. Directions of the Lindbladian errors were randomly chosen. In order to distinguish the effect of amplification by the EACs on SPAM errors from statistical errors, we calculated the true probability distributions and used them as the relative frequencies, which corresponds to the case of the infinite amount of data.

Figure 2 is a numerical result of the setting. The vertical axis is the estimation error $\epsilon := \|\delta L^{\mathrm{RLT}} - \delta L\|$.The horisontal axs is the size of the Hamiltonian part of Lindbladian errors, $\|\delta L_H\| (= \|\delta L_{s,H}\| = \|\delta L_{p,H}\|)$. The blue dots are for all 12 error components, and the orange dots are for amplified 10 error components. The blue dots contain 2 not-amplified error components, and it scales as $\epsilon \sim \|\delta L_H\|$, implying that the size of the systematic error on the not-amplified components can be the same as $\|\delta L_H\|$ itself and the estimate is not reliable. On the other hand, the 10 amplified components (orange dots) is about ten times smaller than the black line ($y = x$) if $\|\delta L_H\| < 5 \times 10^{-2}$, which implies that the amplified components of the estimate are reliable in the region. This numerical result indicates that RLT captures the amplification effect of EACs even though it based on the first order approximations, and that RLT works reliably in a practical setting.

Acknowlegments

This work was supported by JST PRESTO (JPMJPR1915), JST ERATO (JPMJER1601), and MEXT Q-LEAP (JPMXS0118068682), Japan.

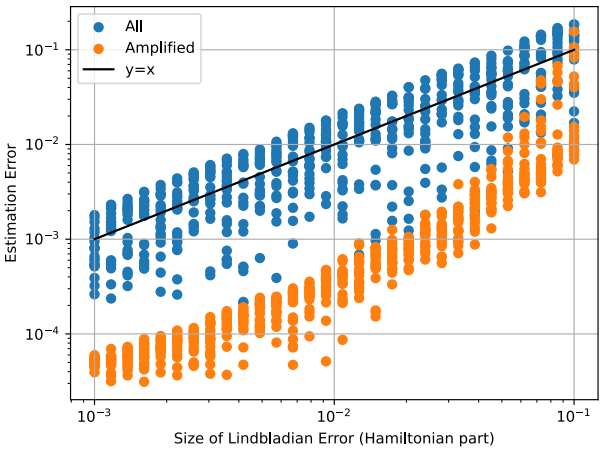


Figure 2: Performance of the proposed method (RLT) in a 1-qubit case (X90 gate) with SPAM error. See the main texts for the details.

References

[1] R. Blume-Kohout et al., Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nature Commun.* 8, 14485 (2017).

[2] R. Blume-Kohout, Idle Tomography. United States (2019) https://www.osti.gov/servlets/purl/1581878

[3] N. Sundaresan et al., Reducing Unitary and Spectator Errors in Cross Resonance with Optimized Rotary Echoes. *PRX Quantum* 1, 020318 (2020).

[4] T. Sugiyama (in preparation).

[5] R. M. Wilcox, Exponential Operators and Parameter Differentiation in Quantum Physics. *J. Math. Phys.* 8, 962 (1967).

[6] M. Hayashi, Quantum Information Theory. (2nd eds.), Springer (2016).

[7] G. W. Stewart and J. Sun, Matrix Perturbation Theory. Academic Press (1990).

[8] N. J. Higham, Functions of Matrices: Theory and Computation. SIAM (2008).

[9] R. L. Kosut, Quantum Process Tomography via L1-norm Minimization, arXiv:0812.4323 [quant-ph].

[10] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, SCS: Splitting Conic Solver. `https://github.com/cvxgrp/scs`

[11] CVXPY: A Python-embedded modeling language for convex optimization problems. `https://www.cvxpy.org`

[12] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient Z gates for quantum computing, *Phys. Rev. A* 96, 022330 (2017).

Unambiguous discrimination of sequences of quantum states

Tathagata Gupta[1] * Shayeef Murshid[2] † Somshubhro Bandyopadhyay[3] ‡

[1] *Physics and Applied Mathematics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India.*
[2] *Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India.*
[3] *Department of Physical Sciences, Bose Institute, EN 80, Bidhannagar, Kolkata 700091, India.*

Abstract. We consider the problem of determining the state of an unknown quantum sequence without error. The elements of the given sequence are drawn with equal probability from a known set of linearly independent pure quantum states with the property that their mutual inner products are all real and equal. This problem can be posed as an instance of unambiguous state discrimination where the states correspond to that of all possible sequences having the same length as the given one. We calculate the optimum probability by solving the optimality conditions of a semidefinite program. The optimum value is achievable by measuring individual members of the sequence, and no collective measurement is necessary.

Keywords: https://arxiv.org/pdf/2402.06365, state discrimination, quantum information

The task of state discrimination is one of the most fundamental primitives in any physical theory. The rules of quantum mechanics make this task a highly non-trivial one in the quantum domain. While any set of orthogonal states can be distinguished with certainty, once the states become non-orthogonal, distinct states do not imply distinguishability [1].

Consequently different variants of the quantum state discrimination problem have been formulated and extensively studied along with the potential scenarios for application [3]. The most common variants of the state discrimination problem are the minimum-error and the unambiguous discrimination problem. In the minimum-error paradigm, the aim is to minimize the average probability of error while remaining completely oblivious to the accuracy of the guess made in a given run [7]. In contrast to this, the unambiguous discrimination task demands an answer only when an error-free identification of the state has been made. In other words, the unambiguous discrimination task allows the discriminator to output an inconclusive "don't know" answer in a given run when the discriminator is not sure of its guess, but a mistaken identification is not allowed [8].

In this work, we exclusively deal with the unambiguous state discrimination paradigm, so it's worthwhile reviewing its basic results. First of all, unlike minimum-error discrimination, which can be done for any set of quantum states, a necessary and sufficient condition for a set of states to be unambiguously distinguishable is that it has to be linearly independent [4]. Secondly, while a closed form expression for the optimum success probability of unambiguous discrimination is known for two states, no such solutions are known for three or more states. The optimal success probability for unambiguous discrimination of two linearly independent states $|\psi_1\rangle$ and $|\psi_2\rangle$ occurring with prior probabilities η_1 and η_2 respectively, is given by $1 - p_?$ [2] where

$$p_? = 2\sqrt{\eta_1\eta_2}|\langle\psi_1|\psi_2\rangle| \tag{1}$$

is the minimum probability for an inconclusive result. For more than two states, the optimal probability of success is known only for some special cases, like three states with pairwise equal and positive inner product [9]. Although a general closed form solution has been elusive, the unambiguous state discrimination problem can be cast as a semi-definite program [6], which means that given a set of states we can find the optimum probability of success by some efficient algorithm.

In this paper, we study unambiguous discrimination of sequences of pure quantum states. Consider a set of quantum states $S_N = \{|\psi_i\rangle : i = 1, \ldots, N\}$, where $N \geq 2$. Suppose a sequence of k quantum states is formed by drawing states from S_N with replacement and given to an identifier whose task is to unambiguously identify the received sequence. This may arise in some information processing task where an identifier receives a sequence of quantum states from another party or a source, which emits states from S_N according to some probability distribution. Let $[n] = \{1, 2, \ldots, n : n \in \mathbb{N}\}$ denote the set of natural numbers from 1 to n and $\mathcal{F}(k, N)$ be the set of all functions from $[k]$ to $[N]$, where $k, N \in \mathbb{N}$. Given S_N, define the set $S_{N,k}$ by

$$S_{N,k} = \left\{ \left|\psi_{\sigma(1)}\right\rangle \otimes \cdots \otimes \left|\psi_{\sigma(k)}\right\rangle : \sigma \in \mathcal{F}(k, N) \right\}.$$

The sequence that the identifier receives is the tensor product of all the states that are chosen from S_N and is an element of the set $S_{N,k}$. Its unambiguous identification can be cast as the unambiguous discrimination of the set $S_{N,k}$.

It is already known that the set of states $S_{N,k}$ is linearly independent if and only if the set S_N is [5]. This means that, given the set $S_{N,k}$ of multi-particle states, we can unambiguously distinguish between them if and only if we can unambiguously distinguish between the single particle states that make up the set S_N. Since the states of $S_{N,k}$ are multipartite, one would expect that their optimal discrimination would entail collective measurements on all the particles. The main contribution of this paper is that we show that this is not the case for a large class of states. In particular, we show that local measurements on individual particles achieve the optimal

*tathagatagupta@gmail.com
†shayeef.murshid91@gmail.com
‡som@jcbose.ac.in

discrimination strategy if the states in S_N are chosen with equal probability and they satisfy the following condition

$$\langle\psi_i|\psi_j\rangle = s,\ \forall\ i,j,\ i \neq j \text{ and } s \in \mathbb{R}. \tag{2}$$

Let us state our result more formally. Let us denote by p and p_k the optimal probability of success of unambiguously discriminating the set S_N and $S_{N,k}$ respectively. Then, we show that, if S_N is a set of equiprobable linearly independent pure states whose pairwise inner products are real and equal then the optimal probabilities of success follow the relation

$$p_k = p^k,$$

where $S_{N,k}$ is the set of sequences formed by drawing states from S_N uniform randomly. Note that unambiguously identifying a given sequence is equivalent to unambiguously identifying all its component states. So a local protocol of discriminating between the sequences could always be carried out by performing the optimal measurement for the set S_N on each state of the sequence. Since the optimal probability of success for S_N is p, a successful identification of all the states, and equivalently of the sequence, is p^k. Our result implies that, although quantum mechanics allows more general measurements on the states taken together collectively, we cannot improve on this probability of success for the sequence; individual measurements on the component states is optimal, and no joint measurement will give us any advantage.

We now briefly discuss our proof method. Since the quantum sequences are quantum states, we can use SDP for unambiguous state discrimination. Given a set of N linearly independent pure states $|\chi_i\rangle$ with prior probabilities η_i, the SDP for their unambiguous discrimination is as follows [10]

$$\begin{aligned} \underset{\vec{p}}{\text{maximize}} \quad & \vec{\eta}\cdot\vec{p} \\ \text{subject to} \quad & \Gamma - P \succeq 0 \\ & \vec{p} \succeq 0. \end{aligned}$$

Here $\vec{\eta} = (\eta_1, \ldots, \eta_N)$ represents the prior probabilities of the states and $\vec{p} = (p_1, \ldots, p_N)$ where p_i is the SDP variable representing the probability of successfully detecting the i-th state $|\chi_i\rangle$; Γ is the Gram matrix whose elements are $\Gamma_{ij} = \langle\chi_i|\chi_j\rangle$ and $P = \text{diag}(p_1, \ldots, p_N)$. The first constraint says that the matrix $\Gamma - P$ should be positive semi-definite and the second constraint is simply the positive-semidefiniteness of the individual probabilities p_i's. First, we note that the primal problem is convex, and it is strictly feasible. Under these conditions Slater's theorem guarantees that strong duality holds, and the duality gap is zero. Therefore, we first present an ansatz probability of success and obtain a feasible solution for the primal problem, which is not necessarily optimal. Then, we present candidates for the dual variables and show that this makes the dual value equal to the primal. Since strong duality holds, this implies that our ansatz must be the optimal solution for the primal problem.

A number of questions arise from our work. The first one is the general case where the mutual inner products between the states of S_N are arbitrary. In this case numerical experiments on short sequences of length $k = 2, 3$ ran with $N = 3$ seem to suggest that the optimal probability is once again achievable by local protocol (assuming uniform prior probabilities).

The second problem deals with sequences without repetition. Restrict $k < N$ and by $S'_{N,k}$ denote the set of length k sequences where there is no repetition of any quantum state. If we let $\mathcal{G}(k, N)$ be the set of injective functions from $[k]$ to $[N]$, then

$$S'_{N,k} = \left\{ \left|\psi_{\tau(1)}\right\rangle \otimes \cdots \otimes \left|\psi_{\tau(k)}\right\rangle : \tau \in \mathcal{G}(k, N) \right\}.$$

The cardinality of this set is ${}^N P_k = \dfrac{N!}{(N-k)!}$. Now assume that the inner products between the states $|\psi_i\rangle$ are equal and positive, say $s > 0$. Then $S'_{N,k} \subset S_{N,k}$, where $S_{N,k}$ is the set of sequences considered here. Under these restricted conditions numerical experiments still suggest that the optimal probability for discriminating the set $S'_{N,k}$ unambiguously once again obeys $(1-s)^k$.

The question of whether optimal unambiguous sequence discrimination requires collective measurements, in general, is an interesting one. Our result showed that if the members of a sequence are uniformly drawn from a linearly independent set with a specific property (inner products are all real and equal), measuring the individual members will suffice, and collective measurements are not required. But the question in general scenarios (without assumptions on inner products) remains open, and our numerical attempts with a limited number of states and sequences have, so far, failed to yield a counter-example.

References

[1] J.A. Bergou. Discrimination of quantum states. *Journal of Modern Optics 57.3*: 160-180 (2010).

[2] G. Jaeger, A. Shimony. Optimal distinction between two non-orthogonal quantum states. *Physics Letters A* 197.2 (1995): 83-87.

[3] J. Bae, L.C. Kwek. Quantum state discrimination and its applications. *Journal of Physics A: Mathematical and Theoretical* 48.8 (2015): 083001.

[4] A. Chefles. Unambiguous discrimination between linearly independent quantum states. *Physics Letters A* 239.6 (1998): 339-347.

[5] A. Chefles, E. Andersson, I. Jex. Unambiguous comparison of the states of multiple quantum systems. *Journal of Physics A: Mathematical and General* 37.29 (2004): 7315.

[6] Y.C. Eldar. A semidefinite programming approach to optimal unambiguous discrimination of quantum states. *IEEE Transactions on information theory*49.2 (2003): 446-456.

[7] C.W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics* 1 (1969): 231-252.

[8] I.D. Ivanovic . How to differentiate between non-orthogonal states. *Physics Letters A* 123.6 (1987): 257-259.

[9] Y. Sun, M. Hillery, J.A. Bergou. Optimum unambiguous discrimination between linearly independent nonorthogonal quantum states and its optical realization. *Physical Review A* 64.2 (2001): 022311.

[10] H. Sugimoto, T. Hashimoto, M. Horibe, A. Hayashi. Complete solution for unambiguous discrimination of three pure states with real inner products. *Physical Review A* 82.3 (2010): 032338.

Parallel Gating of Noisy Silicon Flip-flop Qubits Arranged in Small Arrays with Various Geometries

Marco De Michielis[1] * , Elena Ferraro [1]

[1] *CNR-IMM, Unit of Agrate Brianza, Via C. Olivetti 2, Agrate Brianza (MB), 20864, Italy*

Abstract. Successfully implementing a quantum algorithm involves maintaining a low logical error rate by ensuring the validity of the quantum fault-tolerance theorem. The required number of physical qubits arranged in an array depends on the chosen Quantum Error Correction code and achievable physical qubit error rate. As the qubit count in the array increases, parallel gating —simultaneously manipulating many qubits— becomes a crucial ingredient for a successful computation.

In this study, small arrays of a type of donor- and quantum dot-based qubits, known as flip-flop qubits, are investigated. The flip-flop qubit utilizes antiparallel electron-nuclear spin states of a ^{31}P donor atom embedded in ^{28}Si and controlled by an applied electric field. Simulation results of gate fidelities in linear, square, and star arrays of four flip-flop qubits are presented to examine the effect of parallel gating, as well as that of charge noise and idling qubits. The obtained gate fidelities are compared to determine the optimal array and the reasons behind the selection.

Keywords: semiconductor qubits, flip-flop qubit, gate fidelity, qubit arrays

1 Introduction

In the realm of semiconductor qubit holders based on donor atoms and quantum dots in silicon [1, 2, 3, 4], the flip-flop (FF) qubit exploits antiparallel electron-nuclear spin states of a ^{31}P donor atom embedded in ^{28}Si and it is controlled by an applied electric field. It was designed by Tosi *et al.* in 2017 [5], gained increasing attention in the following years [6, 7, 8, 9, 10] and was experimentally demonstrated in 2023 [11].

The FF qubits have drawn interest for their potential utilization of long-range electric dipole-dipole interactions, typically spanning 200-500 nm [5], which can alleviate the stringent requirements for precise qubit placement and inter-qubit spacing in qubit arrays based only on electrostatic quantum dots.

As the number of qubits in an array increases, the demand for (high-fidelity) parallel gates becomes essential for successfully implementing Quantum Error Correction (QEC). A reliable evaluation of gate fidelity reduction induced by parallel gating is necessary for accurately estimating physical qubit error rates. While previous studies have simulated parallel gating, unwanted interactions with idling qubits, and noise effects on gate fidelities for FF qubits arranged in linear and square arrays [12], this study extends the investigation to a different type of array geometry.

2 Model

The scheme of the FF qubit is presented in Figure 1. The FF qubit comprises a ^{31}P donor atom situated within a ^{28}Si bulk, positioned at a distance from the Si/SiO_2 interface. An electric field, generated by a metal gate on top of the SiO_2 layer, regulates the movement of the donor-bound electron between the donor site and the Si/SiO_2 interface.

*marco.demichielis@cnr.it

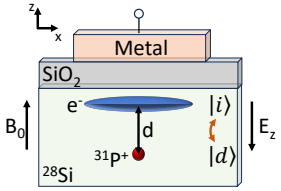


Figure 1: FF qubit scheme where a ^{31}P donor atom is embedded in a ^{28}Si bulk, positioned below a Si/SiO_2 interface and under the application of a constant magnetic field B_0. An electric field E_z controls the system states by moving the electron position between the nucleus (state $|d\rangle$) and the quantum dot at the interface (state $|i\rangle$).

The Hamiltonian $\hat{H}^i$ of the single FF qubit is [5, 8]:

$$\hat{H}^i = \hat{H}_{B_0} + \hat{H}_A + \hat{H}_{Orb} \tag{1}$$

where

$$\hat{H}_{B_0} = \gamma_e B_0 \left[\hat{\mathbb{I}} + \left(\frac{\hat{\mathbb{I}}}{2} + \frac{d\, e\Delta E_z}{2h\epsilon_0}\hat{\sigma}_z + \frac{V_t}{2\epsilon_0}\hat{\sigma}_x \right) \Delta_\gamma \right] \hat{S}_z + - \gamma_n B_0 \hat{I}_z, \tag{2}$$

$$\hat{H}_A = A \left(\frac{\hat{\mathbb{I}}}{2} - \frac{d\, e\Delta E_z}{2h\epsilon_0}\hat{\sigma}_z - \frac{V_t}{2\epsilon_0}\hat{\sigma}_x \right) \mathbf{S}\cdot\mathbf{I}, \tag{3}$$

describe the Zeeman splitting caused by a constant magnetic field B_0 (Eq. 2) and the hyperfine interaction (Eq. 3), respectively. In particular, in Eq. 2, Δ_γ takes into account the variation of the electron gyromagnetic ratio γ_e between the nucleus and the interface, while γ_n is the constant nuclear gyromagnetic ratio. $\mathbf{S}$ ($\mathbf{I}$) are the electron (nuclear) spin operators with $\hat{z}$ component S_z (I_z) and B_0 is a constant magnetic field. Moreover, V_t is the tunnel coupling between the donor and the interface potential well, $\Delta E_z = E_z - E_z^0$ where E_z^0 is the vertical electric field at the ionization point, i.e. the point in which the electron is shared halfway between the donor

and the interface, $\varepsilon_0 = \sqrt{V_t^2 + (de\Delta E_z/h)^2}$ is the energy difference between the orbital eigenstates, where h is the Planck's constant, e is the elementary charge and d is the distance between the positive charge of the ^{31}P nucleus and the negative charge of the bounded electron. In Eq. 3, the hyperfine coupling A decreases as the control electric field E_z is increased, following a function reported in Ref. [8, 12].

The orbital part $\hat{H}_{Orb}$, which gives a treatment of the electron position between the interface and the donor as a two level system is given by

$$\hat{H}_{Orb} = -\frac{\epsilon_0}{2}\hat{\sigma}_z + \\ -\frac{d\, eE_{ac}(t)\cos(\omega_E t+\phi)}{2h}\left(\frac{d\, e\Delta E_z}{h\epsilon_0}\hat{\sigma}_z + \frac{V_t}{\epsilon_0}\hat{\sigma}_x\right),$$

where $E_{ac}(t)$ is the time dependent amplitude of an applied oscillating electric field with pulsation ω_E and phase ϕ.

The interaction Hamiltonian H_{int}^{ij} for two FF qubits with indexes i,j is defined as [8, 12]:

$$\hat{H}_{int}^{i,j} = \frac{1}{4\pi\varphi_0\varphi_r r_{ij}^3}\left[\mathbf{p}_i\cdot\mathbf{p}_j - \frac{3(\mathbf{p}_i\cdot\mathbf{r}_{ij})(\mathbf{p}_j\cdot\mathbf{r}_{ij})}{r_{ij}^2}\right] \quad (4)$$

where φ_0 is the vacuum permittivity, φ_r is the material dielectric constant , $\mathbf{r}_{ij}$ is the vector distance between the two qubits and $\mathbf{p}_{i(j)} = \frac{ed}{2}\left(\hat{\mathbb{I}}_{i(j)} + \hat{\sigma}_{z,i(j)}^{id}\right)$ is the dipole operator of the qubit to whom is associated the position operator $\hat{\sigma}_z^{id} = \frac{d\, e\Delta E_z}{h\epsilon_0}\hat{\sigma}_z + \frac{V_t}{\epsilon_0}\hat{\sigma}_x$, whose eigenstates $|i\rangle$ and $|d\rangle$ indicate if the electron is localized near the interface or the donor, respectively, and $\hat{\sigma}_z = |g\rangle\langle g| - |e\rangle\langle e|$, $\hat{\sigma}_x = |g\rangle\langle e| + |e\rangle\langle g|$ are the Pauli matrices with $|g\rangle(|e\rangle)$ the electron ground (excited) state. The qubit logical basis is $|0\rangle = |g\downarrow\Uparrow\rangle$ and $|1\rangle = |g\uparrow\Downarrow\rangle$, with $|\uparrow\rangle, |\downarrow\rangle$ and $|\Uparrow\rangle, |\Downarrow\rangle$ being the electron and nuclear spin states, respectively.

The quantum gates under investigation here, namely $R_z(-\frac{\pi}{2})$, $R_x(-\frac{\pi}{2})$ and $\sqrt{iSWAP}$, constitute a universal set of quantum gates through a total electrical manipulation [5] with details of the control signal waveforms ΔE_z and E_{ac} reported in Ref. [12].

Three types of arrays made up of four FF qubits are simulated: one with qubits displaced in a linear array (LA), the second one in a square array (SA) and the last one in a star array (STA) as sketched in Figure 2. In STA a central qubit is equidistant from the other ones placed at the vertices of an equilateral triangle. The shortest inter-qubit distance is r_0.

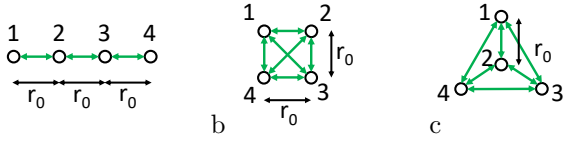


Figure 2: Schemes of the arrays of four qubits with shortest inter-qubit distance r_0. a) Scheme of a LA composed by equally displaced qubits. b) SA scheme. c) STA scheme.

The Hamiltonian $\hat{H}^{4FFQ}$ describing an array of four FF qubits is:

$$\hat{H}^{4FFQ} = \sum_{i=1}^{4}\hat{H}^i(\Delta E_z^i, E_{ac}^i) + \sum_{i=1}^{3}\sum_{j=i+1}^{4}\hat{H}_{int}^{i,j}(\Delta E_z^i, \Delta E_z^j) \quad (5)$$

where the dependencies from the control inputs ΔE_z and E_{ac} of each qubit are explicitly displayed.

The gate unitary matrix $U(t)$ obtained at the end of the application of specific gate control signals is calculated as $U(t) = \exp\left(-it\hat{H}^{4FFQ}(\Delta E_z(t), E_{ac}(t))/\hbar\right)$.

The selected figure of merit to compare the different arrays is the entanglement fidelity F that does not depend on the qubit initial state and is given by [13, 12]

$$F = tr[\rho^{RS}\mathbb{I}_R \otimes (U_i^{-1}U_d)_S\rho^{RS}\mathbb{I}_R \otimes (U_d^{-1}U_i)_S], \quad (6)$$

where U_i (U_d) is the ideal (disturbed) quantum gate matrix and $\rho^{RS} = |\psi\rangle\langle\psi|$, where $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes 2n} + |1\rangle^{\otimes 2n})$ for a n-qubit gate (n=1, 2, 4), represents a maximally entangled state in a double state space generated by two identical Hilbert spaces R and S. For each gate under study, N_{rep} instances of the 1/f charge noise in the time domain are generated with an amplitude $\alpha_{\Delta E_z}$ and added to the ideal sequence signals performing the operation for each qubit. Each qubit noise signal is considered uncorrelated to other qubit noises. Finally the average over the resulting entanglement infidelities $(1-F)$ is calculated.

3 Results and Discussion

In this section the results of simulated infidelities, affected by charge noise and unwanted qubit interactions, for one-qubit operations, for two parallel one-qubit gates, for a single two-qubit operation and finally for two parallel two-qubit gates in the three considered arrays are presented when the noise amplitude $\alpha_{\Delta E_z}$ spans a range from 0 to 100 V/m. r_0 is set to 360 nm in all the simulations and resulting infidelities are averaged over N_{rep}=100 repetitions.

3.1 One-qubit operations

Figure 3 reports $1-F$ as a function of $\alpha_{\Delta E_z}$ when a one-qubit gate, $R_z(-\frac{\pi}{2})$ and $R_x(-\frac{\pi}{2})$, is performed to one qubit while the others are in an idle state. Regardless of the type of performed operation, the set of indexes of the qubit(s) under manipulation, hereafter named configurations "c" followed by the corresponding index(es), are the c1 and c2 configurations for the LA (a), the c1 for the SA (b) and the c1 and c2 for the STA (c). All the others possible configurations are from a geometrical point of view equivalent to the one presented and then give the same $1-F$ results.

Gate infidelities in the three arrays considered increase as $\alpha_{\Delta E_z}$ raises and $1-F$ remain below the $5\cdot10^{-3}$ value in the whole range of noise amplitude studied, with $R_z(-\frac{\pi}{2})$ showing lower $1-F$ than $R_x(-\frac{\pi}{2})$. $1-F$ in c1 configurations are always smaller than c2 ones (c2=c1 for a SA thus it is not plotted). Qubit arranged in a LA, due to lower inter-qubit connectivity, stably shows lower $1-F$ than in SA and STA ones.

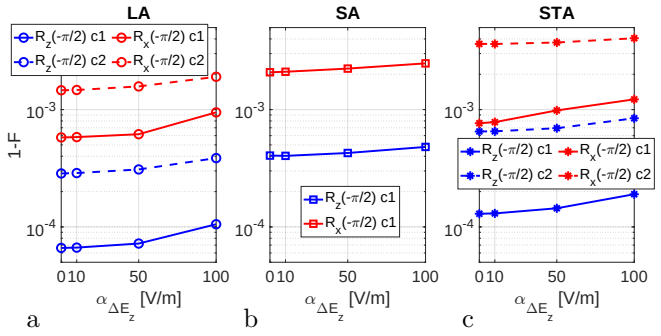


Figure 3: $R_z(-\frac{\pi}{2})$ and $R_x(-\frac{\pi}{2})$ infidelities vs $\alpha_{\Delta E_z}$. The quantum gates are disturbed by the noise and by the idle qubits for a) c1 and c2 configurations in LA, b) c1 in SA and c) c1 and c2 in STA.

3.2 Parallel one-qubit operations

Figure 4 illustrates simulated $1-F$ as a function of $\alpha_{\Delta E_z}$ when two parallel one-qubit gate are applied to a couple of qubits while the other ones are in an idle state. The configurations analyzed depend on the array geometry and they are c12, c13, c14, c23 configurations in LA (a), c12 and c13 in SA (b) and c12 and c13 in STA (c). Parallel gating reasonably produces an overall

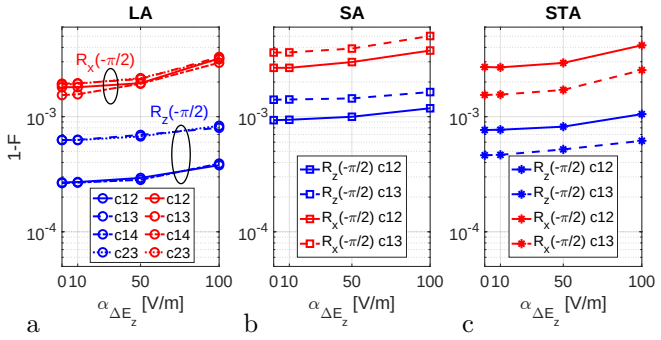


Figure 4: Two parallel $R_z(-\frac{\pi}{2})$ and $R_x(-\frac{\pi}{2})$ infidelities vs the noise amplitude $\alpha_{\Delta E_z}$ for a) the c12, c13, c14 and c23 configurations in the LA, b) the c12 and c13 configurations in the SA and c) the c12 and c13 configurations in the STA.

increased gate infidelity with respect to the one-qubit case. Two parallel $R_z(-\frac{\pi}{2})$ shows lower $1-F$ than two parallel $R_x(-\frac{\pi}{2})$ ones in all the arrays. In the LA $1-F(R_x(-\frac{\pi}{2}))$ in all configurations are almost overlapped whereas $1-F(R_z(-\frac{\pi}{2}))$ in c13 and c23 stay close to each other and the same happens for c12 and c14. In the SA infidelity of c13 is higher than $1-F$ of c12 for both gates and conversely the STA has $1-F$ of c12 larger than $1-F$ of c13. Configuration c12 in the STA has an higher $1-F$ due to the higher connectivity of qubit 2. The LA performs better than others arrays due to lower inter-qubit connectivity.

3.3 Two-qubit operations

Figure 5 illustrates $1-F$ of the two-qubit operation $\sqrt{iSWAP}$ as a function of $\alpha_{\Delta E_z}$ while the other two qubits are in an idle state, for the c12 and c23 configurations in a LA (a) and the c12 configuration in a SA (b) and in a STA (c). The two-qubit operation is applied only between qubits at distance r_0.

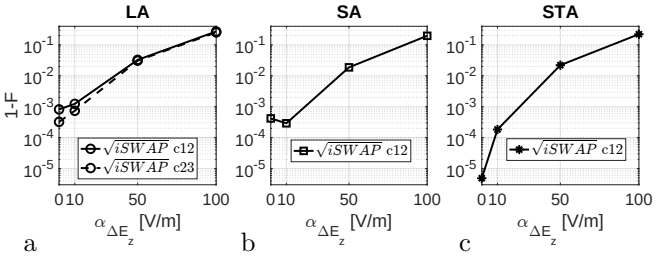


Figure 5: $\sqrt{iSWAP}$ infidelity vs the noise amplitude $\alpha_{\Delta E_z}$ for a) the c12 and c23 configurations in the LA, b) the c12 configuration in the SA and c) the c12 configuration in the STA.

$\sqrt{iSWAP}$ of c23 shows a lower infidelity than that one of c12 in a LA. Only the c12 configuration has been studied for SA and STA showing almost the same $1-F$ values. All the infidelities show an high sensitivity to the noise.

3.4 Parallel two-qubit operations

Figure 6 illustrates the $1-F$ of two parallel $\sqrt{iSWAP}$ gates vs $\alpha_{\Delta E_z}$ for the c12-34 configuration in the LA (a) and in the SA (b). No results are reported for the STA because in this array it is not possible to apply two parallel two-qubit operations with an inter-qubit distance equal to r_0.

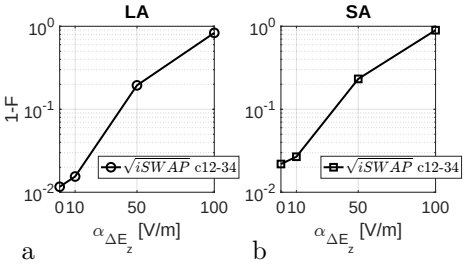


Figure 6: Two parallel $\sqrt{iSWAP}$ c12-34 infidelities vs the noise amplitude $\alpha_{\Delta E_z}$ for a) the LA and b) the SA.

Two parallel $\sqrt{iSWAP}$ infidelity of c12-34 configuration in a LA is lower than that one of c12-34 in a SA. This behaviour is related to the reduced connectivity in the LA with respect to the SA. Both the infidelities are very high even at very low noise values.

4 Conclusion

Good values of infidelity for parallel one-qubit gates in linear, square and star arrays of four flip-flop qubit are obtained. Despite a reduced qubit density of the linear array that reasonably limits the maximum number of arrangeable qubits, results point out that it can achieve lower infidelities than the square and star arrays. This performance advantage of the linear array with respect to the square one vanishes for parallel two-qubit operations. In this case high infidelity results for both the linear and square array are obtained even at very low noise values.

Acknowledgements

The work was partially funded by PNRR MUR projects PE0000023-NQSTI and CN0000013-HPC financed by the European Union – Next Generation EU

References

[1] L. M. K. Vandersypen, H. Bluhm, J. S. Clarke, A. S. Dzurak, R. Ishihara, A. Morello, D. J. Reilly, L. R. Schreiber, M. Veldhorst, Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent, *Npj Quantum Information* **2017**, *3* 34.

[2] J. C. McCallum, B. C. Johnson, T. Botzem, Donor-based qubits for quantum computing in silicon, *Applied Physics Reviews* **2021**, *8* 031314.

[3] G. Burkard, T. D. Ladd, A. Pan, J. M. Nichol, J. R. Petta, Semiconductor spin qubits, *Rev. Mod. Phys.* **2023**, *95* 025003.

[4] M. De Michielis, E. Ferraro, E. Prati, L. Hutin, B. Bertrand, E. Charbon, D. J. Ibberson, M. F. Gonzalez-Zalba, Silicon spin qubits from laboratory to industry, *Journal of Physics D: Applied Physics* **2023**, *56* 363001.

[5] G. Tosi, F. A. Mohiyaddin, V. Schmitt, S. Tenberg, R. Rahman, G. Klimeck, A. Morello, Silicon quantum processor with robust long-distance qubit couplings, *Nature Communications* **2017**, *8* 450.

[6] G. Tosi, F. A. Mohiyaddin, S. Tenberg, A. Laucht, A. Morello, Robust electric dipole transition at microwave frequencies for nuclear spin qubits in silicon, *Physical Review B* **2018**, *98* 075313.

[7] A. Morello, J. J. Pla, P. Bertet, D. N. Jamieson, Donor spins in silicon for quantum technologies, *Advanced Quantum Technologies* **2020**, *3* 2000005.

[8] E. Ferraro, D. Rei, M. Paris, M. De Michielis, Universal set of quantum gates for the flip-flop qubit with 1/f noise, *EPJ Quantum Technology* **2022**, *9*, 2.

[9] D. Rei, E. Ferraro, M. De Michielis, Parallel gate operations fidelity in a linear array of flip-flop qubits, *Advanced Quantum Technologies* **2022**, *5*, 4 2100133.

[10] F. A. Calderon-Vargas, E. Barnes, S. E. Economou, Fast high-fidelity single-qubit gates for flip-flop qubits in silicon, *Phys. Rev. B* **2022**, *106* 165302.

[11] R. Savytskyy, T. Botzem, I. F. de Fuentes, B. Joecker, J. J. Pla, F. E. Hudson, K. M. Itoh, A. M. Jakob, B. C. Johnson, D. N. Jamieson, A. S. Dzurak, A. Morello, An electrically driven single-atom flip-flop qubit, *Science Advances* **2023**, *9*, 6 eadd9408.

[12] M. De Michielis, D. Rei, E. Ferraro, Parallel gate fidelity of flip-flop qubits in small 1d- and 2d-arrays in a noisy environment, *Advanced Quantum Technologies* **2024**, 2300455.

[13] M. A. Nielsen, The entanglement fidelity and quantum error correction, *arXiv:quant-ph/9606012* **1996**.

On fundamental aspects of Quantum Extreme Learning Machines and Reservoir Computing

Weijie Xiong,[1,*]
Giorgio Facelli,[1,*] Mehrad Sahebi,[2,*] Owen Agnel,[3]
Thiparat Chotibut,[4] Supanut Thanasilp,[1] and Zoë Holmes[1]
[1] *Institute of Physics, EPFL, Switzerland*
[2] *Institute of Electrical and Micro Engineering, EPFL, Switzerland*
[3] *Department of Computer Science, University of Oxford, UK*
[4] *Department of Physics, Chulalongkorn University, Bangkok, Thailand*

* The first three authors contributed equally to this work.
weijie.xiong@epfl.ch

Keywords: Quantum algorithms, quantum machine learning, reservoir computing.

Abstract

Quantum Extreme Learning Machines (QELMs) have emerged as a promising framework for quantum machine learning. Their appeal lies in the rich feature map induced by the dynamics of a quantum substrate -the quantum reservoir- and the efficient post-measurement training via linear regression. Here we study the expressivity of QELMs by decomposing the prediction of QELMs into a Fourier series. We show that the achievable Fourier frequencies are determined by the data encoding scheme, while Fourier coefficients depend on both the reservoir and the measurement. Notably, the expressivity of QELMs is fundamentally limited by the number of Fourier frequencies and the number of observables, while the complexity of the prediction hinges on the reservoir. As a cautionary note on scalability, we identify four sources that can lead to the exponential concentration of the measurement outcomes as the system size grows (Haar-expressivity of both reservoir and encoding, hardware noise, entanglement, and global measurements) and show how this can turn QELMs into useless input-agnostic oracles.

We further generalize our analytics to Quantum Reservoir Computing (QRC), which is typically used for time series prediction. We obtain fundamental upper bounds on the total information processing capacity and show that the above four sources can also induce the exponential concentration of observables for QRC with finite length inputs. Our analysis elucidates the potential and fundamental limitations of QELM and QRC, and lays the groundwork for systematically exploring quantum reservoir systems for other machine learning tasks.

References

[1] Xiong, Weijie, et al. *arXiv preprint arXiv:2312.15124 (2023).*

Figures

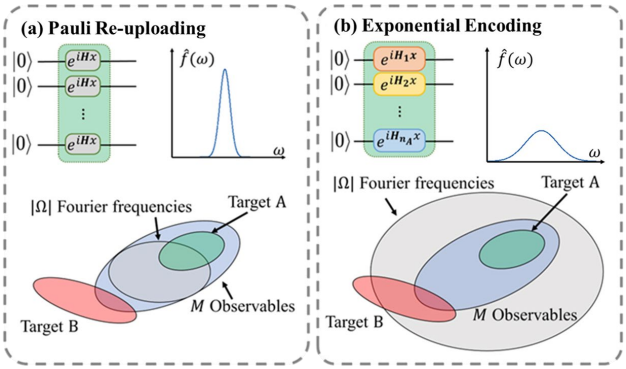


Figure 1: Pauli re-uploading and the exponential encoding lead to polynomially and exponentially many frequencies, resp. Hence, the prediction of a QELM with Pauli encoding has a more concentrated Fourier spectrum, which is more efficiently simulated classically. The exponential encoding, corresponding to the partial control regime M < |Ω|, allows for a wider range of target functions compared to the Pauli re-uploading, where M > |Ω| and might offer a quantum advantage.

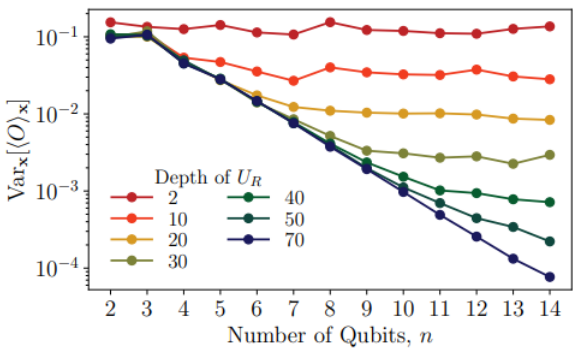


Figure 2: Reservoir Haar-expressivity-induced concentration. Variance of the expectation value of observable Z_1Z_2 over a set of inputs uniformly sampled from [-π, π], as a function of the number of total qubits n and for different depths of the reservoir defined in Eq. (47) of Ref. [1].

Fast computation of magic monotones

Hiroki Hamaguchi[1] * Kou Hamada[1] † Nobuyuki Yoshioka[2] ‡

[1] *Graduate School of Information Science and Technology, University of Tokyo, Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[2] *Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

Abstract. The nonstabilizerness, or magic, is an essential quantum resource to perform universal quantum computation. While the mathematical formalism of nonstabilizerness can be given in a concise manner, it is in general extremely challenging to determine the exact value in practice, in particular when we must deal with superexponentially many pure stabilizer states. In this work, we present fast novel algorithms to compute nonstabilizerness such as the robustness of magic (RoM), stabilizer extent, and stabilizer fidelity. The crucial techniques are subroutines for overlap calculation between the target and all pure stabilizer states that achieve (i) *exponential* reduction of time complexity per stabilizer state, (ii) *superexponential* reduction in the space complexity. Based on these subroutines, we develop algorithms based on the Column Generation method which iteratively updates the subset of stabilizer states necessary to solve the optimization problem. The proposed algorithms allow us to compute the exact values of the RoM and stabilizer extent of arbitrary states up to $n = 8$ and 9 qubits, respectively, while the naive method requires a memory size of at least 86 PiB and 305 EiB, which cannot be executed on any state-of-the-art classical computer. Considering that the proposed iterative algorithms rely on the strong duality of the monotones, which are common for various other resource measures, we envision that the algorithm readily generalizes beyond the targets discussed in this work. Our work paves a novel avenue for quantum computing architecture design based on a resource theoretic approach.
(Remark: Our submission combines two works provided by technical manuscript (TM) 1, 2)

Keywords: Resource theory of magic, Robustness of Magic, Stabilizer extent, Pauli decomposition, Classical simulation

1 Motivation and background

Universal fault-tolerant quantum computation is often formulated such that the elementary gates consist of both classically simulatable gates and costful gates, such as in the most well-known Clifford+T formalism [1–6]. Since the non-Clifford gates are indispensable for any quantum advantage [7–10], there is a surging need to evaluate the complexity of quantum circuits using the framework of resource theory, in order to explore the boundary of quantum and classical computers [11–19].

One of the earliest attempts is the proposal of a quantity called the robustness of magic (RoM) by Howard and Campbell [20–23], which characterizes the quasiprobability-based classical simulation overhead or complexity of a given quantum state. The computation of the RoM can be reduced to a simple L^1 norm minimization problem, and hence can be done with polynomial time to the problem size. However, the number of pure stabilizer states grows superexponentially with the number of qubits n, and hence it is extremely challenging to perform computation beyond $n > 5$ qubits. Such a bottleneck is also common in other resource measures [15, 24], such as the stabilizer extent which is known to characterize the classical simulation cost of rank-based simulators [11, 15, 24, 25]. While besides the work by Heinrich and Gross that utilizes the permutation symmetry of disentangled state for the RoM [22], there is no strategy to offload such a heavy burden to compute such magic monotones.

2 Preliminary on magic monotones

In order to understand the bottleneck more in detail, we provide some basic notations and then the definitions of magic monotones. Let $\mathcal{P}_n = \{\pm 1, \pm i\} \times \{I, X, Y, Z\}^{\otimes n}$ be the n-qubit Pauli group. When there exist a set of n independent Pauli operators $\{P_i\}$ such that $P_i|\psi\rangle = |\psi\rangle$, the state $|\psi\rangle$ is an n-qubit stabilizer state. We denote the entire set of n-qubit stabilizer states as $\mathcal{S}_n$, whose size scales superexponentially as $|\mathcal{S}_n| = 2^n \prod_{k=0}^{n-1}(2^{n-k}+1) = 2^{\mathcal{O}(n^2)}$ [26, 27].

The *Robustness of magic* (RoM) of an n-qubit mixed state ρ is defined as

$$\mathcal{R}(\rho) = \min_{x \in \mathbb{R}^{|\mathcal{S}_n|}} \left\{ \|x\|_1 \;\middle|\; \rho = \sum_{\sigma_i \in \mathcal{S}_n} x_i \sigma_i \right\}. \quad (1)$$

This can be rewritten into a standard form of Linear Program (LP), and thus we can formulate ei-

*hamaguchi-hiroki0510@g.ecc.u-tokyo.ac.jp
†zkouaaa@g.ecc.u-tokyo.ac.jp
‡nyoshioka@ap.t.u-tokyo.ac.jp

Target	Application	Formulation	Subroutine time complexity		Memory	
			Naive	Ours	Naive	Ours
Robustness of magic [20]	Clifford+T sim. Circuit synthesis	LP	$\mathcal{O}(\lvert\mathcal{S}_n\rvert 2^n)$	$\mathcal{O}(\lvert\mathcal{S}_n\rvert n)$	$\mathcal{O}(\lvert\mathcal{S}_n\rvert 2^n)$	$\mathcal{O}(2^n)$
Stabilizer extent [24]	Clifford+T sim.	SOCP	$\mathcal{O}(\lvert\mathcal{S}_n\rvert 2^n n^2)$	$\mathcal{O}(\lvert\mathcal{S}_n\rvert)$	$\mathcal{O}(\lvert\mathcal{S}_n\rvert 2^n)$	$\mathcal{O}(2^n)$
Stabilizer fidelity [24]	Bound for RoM	Overlap calculation	$\mathcal{O}\big(\lvert\mathcal{S}_n\rvert 2^n n^2\big)$	$\mathcal{O}(\lvert\mathcal{S}_n\rvert)$	$\mathcal{O}(2^n)$	$\mathcal{O}(2^n)$
Pauli decomposition	Circuit simulation Noise analysis Quantum benchmark	Matrix-vector multiplication	$\mathcal{O}(16^n)$	$\mathcal{O}(4^n n)$	$\mathcal{O}(4^n)$	$\mathcal{O}(4^n)$

Table 1: Summary for computational complexity of algorithms for magic resource measures and related subroutines. Details for RoM and stabilizer extent are provided in TM 1 and 2, respectively.

Magic Monotone	Qubit count n	5	6	7	8	9
RoM	$\lvert\mathcal{S}_n\rvert$	2.42e+06	3.15e+08	8.13e+10	4.18e+13	4.29e+16
	size of A_n^{RoM}	379 MiB	95 GiB	86 TiB	86 PiB	172 EiB
	Runtime (naive)	2 min	×	×	×	×
	Runtime (ours)	2.3 s	7.0 min	1.6 h	2.0 d	×
Stabilizer extent	$\lvert\mathcal{S}_n\rvert$	2.42e+06	3.15e+08	8.13e+10	4.18e+13	4.29e+16
	size of A_n^{SE}	1011 MiB	254 GiB	153 TiB	153 PiB	305 EiB
	Runtime (naive)	7.7 min	×	×	×	×
	Runtime (ours)	1.5 s	3.8 s	12.9 s	8.8 min	19.2 h

Table 2: Numerical demonstration of fast magic monotone calculations.

ther via the primal or dual formalism as

$$\mathcal{R}(\rho) = \begin{cases} \min\limits_{x\in\mathbb{R}^{|\mathcal{S}_n|}} \left\{ \|x\|_1 \;\middle|\; A_n^{\mathrm{RoM}} x = b \right\} & \text{(Primal)}, \\ \max\limits_{y\in\mathbb{R}^{4^n}} \left\{ b^\top y \;\middle|\; \left\| A_n^{\mathrm{RoM}\top} y \right\|_\infty \le 1 \right\} & \text{(Dual)}. \end{cases} \tag{2}$$

Here, $b_j = \mathrm{Tr}[\rho P_j]$ gives the unique Pauli decomposition that stores the information of the state ρ, and $(A_n^{\mathrm{RoM}})_{j,i} = \mathrm{Tr}[\sigma_i P_j]$ encapsulates the information of the entire pure stabilizer states. We also introduce primal and dual variables x and y.

The *stabilizer extent* of an n-qubit pure state $|\psi\rangle$ is defined as

$$\xi(\psi) \coloneqq \min_{c\in\mathbb{C}^{|\mathcal{S}_n|}} \left\{ \|c\|_1^2 \;\middle|\; |\psi\rangle = \sum_{j=1}^{|\mathcal{S}_n|} c_j |\phi_j\rangle \right\}. \tag{3}$$

This is pointed out to be a Second-Order Cone Program (SOCP) [28], and thus we can further simplify as the complex L^1-norm minimization problem as

$$\sqrt{\xi(\psi)} = \begin{cases} \min\limits_{x\in\mathbb{C}^{|\mathcal{S}_n|}} \left\{ \|x\|_1 \;\middle|\; A_n^{\mathrm{SE}} x = b \right\} & \text{(Primal)}, \\ \max\limits_{y\in\mathbb{C}^{2^n}} \left\{ \mathrm{Re}(b^\dagger y) \;\middle|\; \left\| A_n^{\mathrm{SE}\dagger} y \right\|_\infty \le 1 \right\} & \text{(Dual)}. \end{cases} \tag{4}$$

Here, we define $A_n^{\mathrm{SE}} \in \mathbb{C}^{2^n\times|\mathcal{S}_n|}$ as $(A_n^{\mathrm{SE}})_{ij} \coloneqq \langle i|\phi_j\rangle$ and $b \in \mathbb{C}^{2^n}$ as $b_i \coloneqq \langle i|\psi\rangle$ using the computational basis $\{|i\rangle\}_{i=0}^{2^n-1}$.

It is known that both LP and SOCP can be solved with polynomial time with respect to the problem size, while the matrices A_n^{RoM} and A_n^{SE} are superexponentially large. This renders naive solvers intractable for $n > 5$. However, to our knowledge, none of existing algorithms have exploited the dual formalism of the problem nor the mathematical structures of A_n^{RoM} and A_n^{SE}. In this regard, our goal is to combine the newly introduced canonical forms of stabilizer states and the art of optimization techniques to push the state-of-the-art computation of resource monotones.

3 Summary of results

Our primal findings are the fast computation algorithms for magic resource measures such as the RoM, stabilizer extent, and stabilizer fidelity, as summarized in Table 1. We find that the strong duality of magic monotones can be exploited. Namely, it is extremely beneficial to employ a technique called the Column Generation (CG) method in the dual formalism, which takes a subset of stabilizer states and iteratively updates it until one finds the subset necessary to obtain the exact solution (see Algorithm 1). The initial guess and updates are based on overlap calculation be-

tween the target state and all the stabilizer states, for which we propose two distinct algorithms for mixed states (see Sec. 3 in TM1) and pure states (see Sec. 3 in TM2) that reduces the time complexity *exponentially* per stabilizer state. Furthermore, compared to the naive method of solving the primal problems in Eqs. (2) and (4), our proposal is *superexponentially* more efficient in terms of memory consumption.

We have performed numerical demonstrations to show the significant improvement achieved by our proposal. Regarding the calculation of the RoM, we have successfully applied Algorithm 1 to random mixed states up to $n = 8$ qubits within 2 days using a cluster computer, while naive computation consumes memory of 86 PiB. As a byproduct, we have also proposed an algorithm that computes the Pauli decomposition b of quantum states with nearly quadratically improved time complexity from $\mathcal{O}(16^n)$ to $\mathcal{O}(4^n n)$ (see Appendix B in TM1). It is worth mentioning that there are other works with improved time complexity of $\mathcal{O}(8^n)$ [29, 30] and the one with similar complexity [31]; we have performed an exhaustive comparison to find that our proposal is the fastest among state-of-the-art algorithms (see Appendix B in TM1).

We find that there is an even more drastic improvement in the computation of the stabilizer extent. Owing to our pruning technique, dubbed as the *stabilizer pruning*, we can avoid unnecessary overlap computation by leveraging the newly proposed canonical form of stabilizers (see Sec. 3.1 in TM2). This enabled us to compute the stabilizer extent $\xi(\psi)$ of random pure quantum state up to $n = 9$ qubits, which naively requires memory of 305 EiB. In similar to the calculation of RoM, we have utilized the CG method to obtain the subset of stabilizer states required to obtain the exact solution with significantly reduced problem size.

While the idea of applying resource theory to quantum computing has attracted great amount of interest, the barrier of computational hardness (in particular memory consumption) has prevented us from gaining further benefits for circuit design and optimization. Our work provides a comprehensive methodology that is not limited to the resource measures considered above, but is also expected to generalize to other monotones such as the dyadic negativity [15].

4 Key technical contributions

The key to the significant speed up of the overlap calculation is modified canonical forms of stabilizer states which allow us to perform the overlap calculation of expontentially many stabilizer states with reduced time complexity.

Algorithm 1: Exact Magic Monotone Calculation by Column Generation

Input: Vector b encoding state information
Column set $\mathcal{A}_n = \{a_j\}$ of A_n
Output: Exact value $\mathcal{M}$ ($= \mathcal{R}$ or ξ)

$\mathcal{C}_0 \leftarrow$ Partial set of $\mathcal{A}_n$
`/* Initialize using overlap values */`
for $k = 0, 1, 2, \ldots$ **do**
  Primal x_k, dual $y_k \leftarrow \texttt{Solve}(\mathcal{C}_k, b)$
  `/* Solve problem restricted to` $\mathcal{C}_k$ `*/`
  Compute $\hat{\mathcal{M}}_k$ from x_k
  $\mathcal{C}' \leftarrow \left\{ a_j \in \mathcal{A}_n \;\middle|\; \left| a_j^\dagger y_k \right| > 1 \right\}$
  `/* Use of core subroutine */`
  if $\mathcal{C}' = \emptyset$ **then**
    return $\mathcal{M} = \hat{\mathcal{M}}_k$
  $\mathcal{C}_{k+1} \leftarrow \mathcal{C}_k \cup \mathcal{C}'$

For the case of the RoM, our findings are summarized by the following (see Lemma 2 in TM1):

Lemma 1 *(Decomposition of A_n^{RoM} into Walsh–Hadamard matrix) For all $n \in \mathbb{N}$, there exists a constructive and efficient way of enumerating properly sparsified Walsh–Hadamard matrices $\{W_j\}_{j=1}^{|\mathcal{S}_n|/2^n}$ such that*

$$A_n^{\mathrm{RoM}} = \left[W_1 \cdots W_{|\mathcal{S}_n|/2^n} \right]. \tag{5}$$

Once such a canonical form is obtained, we can utilize an in-place calculation algorithm called the Fast Walsh-Hadamard Transform to compute the matrix-vector multiplication for each W_j with time complexity of $\mathcal{O}(n2^n)$ and space complexity of $\mathcal{O}(2^n)$.

For the calculation of the stabilizer extent, first note that there exists a canonical form for every stabilizer state as [32–34]

$$|\phi\rangle = \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x} i^{c^\top x} |Rx + t\rangle, \tag{6}$$

where $k \in \{0, ..., n\}$, $R \in \mathbb{F}_2^{n\times k}$, $Q \in \mathbb{F}_2^{k\times k}$, $t \in \mathbb{F}_2^n$. In our work, we find that we can construct a modified and equivalent canonical form even if we impose the following conditions (see Theorem 1 in the TM2):

(1) Q is an upper triangular matrix

(2) R is a reduced row echelon form matrix with rank k

(3) t is a representative of element in quotient space $\mathbb{F}_2^n/\mathrm{Im}(R)$

By using these properties, we can show that there exists a recursive procedure that allows us to prune the unnecessary calculations to a significant amount (see Appendix A in TM2).

References

[1] D. Gottesman, *The Heisenberg Representation of Quantum Computers*, 1998.

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[3] S. Bravyi and A. Kitaev, "Universal Quantum Computation with ideal Clifford gates and noisy ancillas", Physical Review A **71**, 022316 (2005).

[4] D. Litinski, "A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery", Quantum **3**, 128 (2019).

[5] D. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, "Surface code quantum computing by lattice surgery", New Journal of Physics **14**, 123011 (2012).

[6] A. G. Fowler and C. Gidney, *Low overhead quantum computation using lattice surgery*, 2019.

[7] C. Gidney and M. Ekerå, "How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits", Quantum **5**, 433 (2021).

[8] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, "Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction", PRX Quantum **2**, 030305 (2021).

[9] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, "Quantum computing enhanced computational catalysis", Physical Review Research **3**, 033055 (2021).

[10] N. Yoshioka, T. Okubo, Y. Suzuki, Y. Koizumi, and W. Mizukami, "Hunting for quantum-classical crossover in condensed matter problems", npj Quantum Information **10**, 45 (2024).

[11] S. Bravyi, G. Smith, and J. Smolin, "Trading classical and quantum computational resources", Physical Review X **6**, 021043 (2016).

[12] E. Tirrito, P. S. Tarabunga, G. Lami, T. Chanda, L. Leone, S. F. E. Oliviero, M. Dalmonte, M. Collura, and A. Hamma, "Quantifying nonstabilizerness through entanglement spectrum flatness", Physical Review A: Atomic, Molecular, and Optical Physics **109**, L040401 (2024).

[13] O. Hahn, A. Ferraro, L. Hultquist, G. Ferrini, and L. García-Álvarez, "Quantifying Qubit Magic Resource with Gottesman-Kitaev-Preskill Encoding", Physical Review Letters **128**, 210502 (2022).

[14] T. Haug, S. Lee, and M. S. Kim, *Efficient quantum algorithms for stabilizer entropies*, 2023.

[15] J. R. Seddon, B. Regula, H. Pashayan, Y. Ouyang, and E. T. Campbell, "Quantifying Quantum Speedups: Improved Classical Simulation From Tighter Magic Monotones", PRX Quantum **2**, 010345 (2021).

[16] Z.-W. Liu and A. Winter, "Many-body quantum magic", PRX Quantum **3**, 020333 (2022).

[17] L. Leone, S. F. E. Oliviero, and A. Hamma, "Stabilizer Rényi Entropy", Physical Review Letters **128**, 050402 (2022).

[18] M. Beverland, E. Campbell, M. Howard, and V. Kliuchnikov, "Lower bounds on the non-clifford resources for quantum computations", Quantum Science and Technology **5**, 035009 (2020).

[19] L. Leone, S. F. E. Oliviero, and A. Hamma, "Stabilizer rényi entropy", Phys. Rev. Lett. **128**, 050402 (2022).

[20] M. Howard and E. T. Campbell, "Application of a resource theory for magic states to fault-tolerant quantum computing", Physical Review Letters **118**, 090501 (2017).

[21] H. Pashayan, J. J. Wallman, and S. D. Bartlett, "Estimating outcome probabilities of quantum circuits using quasiprobabilities", Physical Review Letters **115**, 070501 (2015).

[22] M. Heinrich and D. Gross, "Robustness of Magic and Symmetries of the Stabiliser Polytope", Quantum **3**, 132 (2019).

[23] H. Pashayan, S. D. Bartlett, and D. Gross, "From estimation of quantum probabilities to simulation of quantum circuits", Quantum **4**, 223 (2020).

[24] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, "Simulation of quantum circuits by low-rank stabilizer decompositions", Quantum **3**, 181 (2019).

[25] L. Kocia and M. Sarovar, "Classical simulation of quantum circuits using fewer gaussian eliminations", Phys. Rev. A **103**, 022603 (2021).

[26] S. Aaronson and D. Gottesman, "Improved Simulation of Stabilizer Circuits", Physical Review A **70**, 052328 (2004).

[27] H. J. García, I. L. Markov, and A. W. Cross, "On the geometry of stabilizer states", Quantum Info. Comput. **14**, 683 (2014).

[28] A. Heimendahl, F. Montealegre-Mora, F. Vallentin, and D. Gross, "Stabilizer extent is not multiplicative", Quantum **5**, 400 (2021).

[29] S. Vidal Romero and J. Santos-Suárez, "Paulicomposer: compute tensor products of pauli matrices efficiently", Quantum Information Processing **22**, 449 (2023).

[30] T. Jones, "Decomposing dense matrices into dense pauli tensors", arXiv preprint arXiv:2401.16378 (2024).

[31] L. Hantzko, L. Binkowski, and S. Gupta, "Tensorized pauli decomposition algorithm", arXiv preprint arXiv:2310.13421 (2023).

[32] G. Struchalin, Y. A. Zagorovskii, E. Kovlakov, S. Straupe, and S. Kulik, "Experimental Estimation of Quantum State Properties from Classical Shadows", PRX Quantum **2**, 010307 (2021).

[33] M. V. den Nest, "Classical simulation of quantum computation, the gottesman-Knill theorem, and slightly beyond", Quantum Inf. Comput. **10**, 258 (2010).

[34] J. Dehaene and B. De Moor, "Clifford group, stabilizer states, and linear and quadratic operations over GF(2)", Physical Review A **68**, 042318 (2003).

TM1: Handbook for Efficiently Quantifying Robustness of Magic

Hiroki Hamaguchi[1], Kou Hamada[1], and Nobuyuki Yoshioka[2,3,4]

[1]Department of Mathematical Engineering and Information Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[2]Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[3]Theoretical Quantum Physics Laboratory, RIKEN Cluster for Pioneering Research (CPR), Wako-shi, Saitama 351-0198, Japan
[4]JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

The nonstabilizerness, or magic, is an essential quantum resource to perform universal quantum computation. Robustness of Magic (RoM) in particular characterizes the degree of usefulness of a given quantum state for non-Clifford operation. While the mathematical formalism of RoM can be given in a concise manner, it is extremely challenging to determine the RoM in practice, since it requires dealing with superexponentially many pure stabilizer states. In this work, we present efficient novel algorithms to compute the RoM. The crucial technique is a subroutine that achieves the remarkable features in the calculation of overlaps between pure stabilizer states: (i) the time complexity per state is reduced *exponentially*, (ii) the space complexity in total is reduced *superexponentially*. Based on this subroutine, we present algorithms to compute the RoM for arbitrary states up to $n = 8$ qubits, while the naive method requires a memory size of at least 86 PiB, which cannot be executed on any state-of-the-art classical computer. We find as a byproduct that, the proposed subroutine allows us to simulate the stabilizer fidelity up to $n = 8$ qubits. We further propose novel algorithms that utilize the preknowledge of the structure of the target quantum state such as the permutation symmetry or disentanglement, and numerically demonstrate our state-of-the-art results for copies of magic states and partially disentangled quantum states. The series of algorithms constitutes a comprehensive "handbook" to scale up the computation of the RoM, and we envision that the proposed technique applies to the computation of other quantum resource measures as well.

1 Introduction

Universal fault-tolerant quantum computation is often formulated such that the elementary gates consist of both classically simulatable gates and costful gates, such as in the most well-known Clifford+T formalism of the magic state model [1, 2, 3, 4, 5, 6]. Since the

Hiroki Hamaguchi: hamaguchi-hiroki0510@g.ecc.u-tokyo.ac.jp
Kou Hamada: zkouaaa@g.ecc.u-tokyo.ac.jp
Nobuyuki Yoshioka: nyoshioka@ap.t.u-tokyo.ac.jp

consumption of the non-Clifford gates is indispensable for any quantum advantage [7, 8, 9, 10], there is a surging need to evaluate the complexity of quantum circuits using the framework of resource theory, in order to explore the boundary of quantum and classical computers [11, 12, 13, 14, 15, 16, 17]. One such attempt is the proposal of a quantity called the Robustness of Magic (RoM) by Howard and Campbell [18, 19, 20]; the RoM characterizes the classical simulation overhead or complexity of a given quantum state based on its effective amount of magic, and it geometrically quantifies the distance from the convex set of stabilizer states.

The computation of the RoM can be reduced to a simple L^1 norm minimization problem. Meanwhile, since the size of the set of pure stabilizer states grows superexponentially with the number of qubits, both the time and space complexity are prohibitively large, so it is extremely challenging to perform computation beyond $n > 5$ qubits. Concretely, the memory consumption blows up to 86 PiB even for $n = 8$ qubit system. Existing works have been done to offload such a heavy burden; for instance, Ref. [20] proposed to utilize the symmetry of the target state. By exploiting the permutation symmetry between copies of identical states and also the internal (or local) symmetry, it has been shown that, if one considers copies of symmetric magic states such as $|H\rangle^{\otimes n}$ used for T-gates, one can simulate up to $n = 26$ qubits. However, when one is interested in the magic resource of noisy states, for instance, there is no valid way to scale up the characterization of the resource.

In this work, we propose a systematic procedure as shown in Fig. 1 to compute the RoM value that overcomes the barrier in the existing works. Central to our work is the subroutine that computes the overlaps of a given quantum state between stabilizer states with (i) exponentially faster time complexity per state and (ii) superexponentially smaller space complexity in total. Using this efficient subroutine, we propose algorithms that surpass the state-of-the-art results of the RoM calculation for arbitrary states up to $n = 8$ qubits. We also extend the capability of methods that utilize the preknowledge of the structure of the target quantum state such as the permutation symmetry and decoupled structure, and show that we may compute the RoM for multiple copies of arbitrary single qubit states up to $n = 17$ qubits.

The remainder of this work is organized as follows. In Sec. 2, we present the preliminaries regarding the formalism of RoM. In Sec. 3, we first give the main subroutine on the overlap calculation (Theorem 1) and then present the algorithms that compute the RoM value with the reduced computational resource by utilizing the information of all the overlap values between the target state and stabilizer states. In Sec. 4, we present algorithms for practical target states that are decoupled from each other, such as the multiple copies of single-qubit states or tensor products over subsystems. Finally, in Sec. 5, we provide the discussion and future perspective of our work.

2 Preliminaries

2.1 Robustness of Magic

Let $\mathcal{P}_n = \{\pm 1, \pm i\} \times \{I, X, Y, Z\}^{\otimes n}$ be the n-qubit Pauli group. For any n-qubit stabilizer state $|\psi\rangle$, we denote by $\mathrm{Stab}(|\psi\rangle) = \langle P_1, \ldots, P_n\rangle$ the stabilizer group of $|\psi\rangle$, i.e., the group generated by the set of n independent Pauli operators such that $P_i |\psi\rangle = |\psi\rangle$ for each generator P_i. We denote the entire set of n-qubit stabilizer states as $\mathcal{S}_n$, whose size scales superexponentially as $|\mathcal{S}_n| = 2^n \prod_{k=0}^{n-1}(2^{n-k} + 1) = 2^{\mathcal{O}(n^2)}$ [21, 22], and also denote the convex hull of them as $\mathrm{STAB}_n = \{\sum_i p_i \sigma_i \mid \sigma_i \in \mathcal{S}_n, p_i \geq 0, \sum_i p_i = 1\}$.

The Robustness of Magic (RoM) of a given n-qubit quantum state ρ can be interpreted

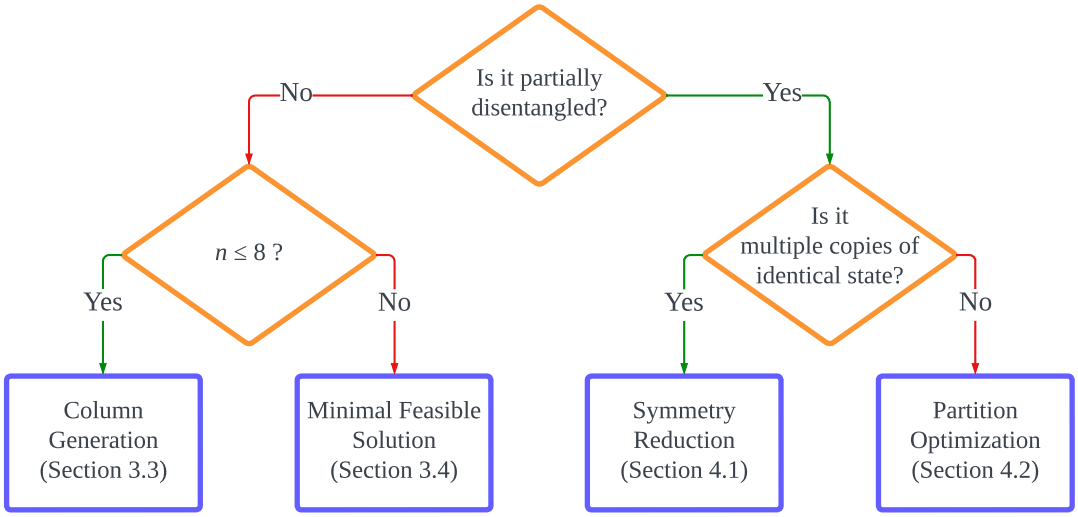


Figure 1: Flow chart of RoM computation.

Method	Target	Qubit count	Exact/Approximate
Naive LP [18]	Arbitrary	$n \leq 5$	Exact
Top-overlap	Arbitrary	$n \leq 8$	Exact*
Column Generation (CG)	Arbitrary	$n \leq 8$	Exact
Minimal Feasible Solution	Arbitrary	$n \leq 14$	$2^{n/2}$-approximation, if random
Symmetry Reduction	$\rho^{\otimes n}$	$n \leq 17$	Exact up to $n \leq 7$
Partition Optimization	$\bigotimes_i \rho_i$	$n \leq 15$	Approximation
Symmetry Reduction [20]	$\rho_{H,F}^{\otimes n}$	$n \leq 26$	Exact up to $n \leq 9, 10$

Table 1: Methods for calculating RoM. The top four methods are applicable to arbitrary n-qubit states, while the latter three assume certain structures such as permutation symmetry, decoupled structure, and local symmetry. The expression "Exact*" is intended to denote that there is a hyperparameter that controls the accuracy of the solution, while we have performed numerical demonstrations that successfully find the exact solution up to $n = 5$ qubits.

as distance from the polytope of free states identified with STAB_n and is defined as [18, 23]

$$\mathcal{R}(\rho) := \min_{\sigma_+, \sigma_- \in \mathrm{STAB}_n} \left\{ 2p+1 \;\middle|\; \rho = (p+1)\sigma_+ - p\sigma_-,\ p \geq 0 \right\}. \tag{1}$$

It is straightforward to show that this yields an equivalent expression as

$$\mathcal{R}(\rho) = \min_{\boldsymbol{x}} \left\{ \sum_{i=1}^{|\mathcal{S}_n|} |x_i| \;\middle|\; \rho = \sum_{\sigma_i \in \mathcal{S}_n} x_i \sigma_i,\ x_i \in \mathbb{R} \right\}, \tag{2}$$

which can be further simplified as

$$\mathcal{R}(\rho) = \min_{\boldsymbol{x}} \left\{ \sum_{i=1}^{|\mathcal{S}_n|} |x_i| \;\middle|\; \boldsymbol{A}_n \boldsymbol{x} = \boldsymbol{b} \right\}. \tag{3}$$

Here, we have utilized the unique decomposition of the quantum state into n-qubit Pauli operators to define $b_j = \mathrm{Tr}[\rho P_j]$ and $(\boldsymbol{A}_n)_{j,i} = \mathrm{Tr}[\sigma_i P_j]$ where P_j $(1 \leq j \leq 4^n)$ is the j-th

Pauli operator in lexical order. Here, $\boldsymbol{A}_n$ encapsulates the information of the entire pure stabilizer states, whereas $\boldsymbol{x}$ stores that of the target quantum state ρ. It is also convenient to define $\mathcal{A}_n = \{\boldsymbol{a}\}$ as the entire set of columns in $\boldsymbol{A}_n$ such that $\boldsymbol{A}_n = (\boldsymbol{a})_{\boldsymbol{a}\in\mathcal{A}_n}$. In practice, one may solve Eq. (3) as follows:

$$\begin{aligned} \underset{\boldsymbol{u}}{\text{minimize}} \quad & \sum_i u_i \\ \text{subject to} \quad & \begin{pmatrix} \boldsymbol{A}_n & -\boldsymbol{A}_n \end{pmatrix} \boldsymbol{u} = \boldsymbol{b}, \\ & \boldsymbol{u} \geq \boldsymbol{0}, \end{aligned} \tag{4}$$

where the inequality denotes the element-wise inequality. This is a standard form of linear programming problem, and thus can be solved by Linear Programming (LP). For the sake of convenience for later discussion, we denote a function $\texttt{SolveLP}(\mathcal{A}, \boldsymbol{b})$ that returns $\mathcal{R}(\rho)$ and $\boldsymbol{x}$ by solving Eq. (4) given a set of columns $\mathcal{A}$, and also denote the minimization problem itself as $\texttt{Prob}(\boldsymbol{A}, \boldsymbol{b})$ using the matrix $\boldsymbol{A}$ determined from the set $\mathcal{A}$.

2.2 Dualized Robustness of Magic

Since the RoM can be formalized via the standard form of linear programming problem, the strong duality holds, which implies that the dual problem gives an equivalent definition. Concretely, the value of the RoM can be computed via the following:

$$\mathcal{R}(\rho) = \max_{\boldsymbol{y}} \left\{ \boldsymbol{b}^\top \boldsymbol{y} \;\middle|\; -\boldsymbol{1} \leq \boldsymbol{A}_n^\top \boldsymbol{y} \leq \boldsymbol{1} \right\}, \tag{5}$$

where $\boldsymbol{1}$ is a length-4^n vector with all the elements given by unity. By the nature of the dual problem, any feasible solution yields a lower bound on the RoM. For instance, it can be shown by taking $\boldsymbol{y} = (\dots, \operatorname{sgn}(\operatorname{Tr}[\rho P_j])/2^n, \dots)$ that, the RoM can be lower-bounded by the st-norm as $\|\rho\|_{\mathrm{st}} = \frac{1}{2^n}\|\boldsymbol{b}\|_1$ [20].

3 Scaling up RoM calculation for arbitrary states

It is well known that linear programming problems are solvable in polynomial time with respect to the matrix size. However, it must be noted that the matrix size of $\boldsymbol{A}_n$ itself is $4^n \times |\mathcal{S}_n|$ where $|\mathcal{S}_n| = 2^{\mathcal{O}(n^2)}$, and hence it is impractical to use the entire $\boldsymbol{A}_n$ to tackle $n > 5$ qubit systems [14].

Motivated by such a problematic situation, we propose two numerical algorithms that compute the RoM for arbitrary quantum states beyond the state-of-the-art system size. The key technique is to utilize a subroutine that achieves the following two remarkable features in overlap calculation: (i) total time complexity is drastically improved from $\mathcal{O}(2^n|\mathcal{S}_n|)$ to $\mathcal{O}(n|\mathcal{S}_n|)$, which is due to the exponential reduction from $\mathcal{O}(2^n)$ to $\mathcal{O}(n)$ per stabilizer state, (ii) the space complexity is reduced superexponentialy from $\mathcal{O}(2^n|\mathcal{S}_n|)$ to $\mathcal{O}(2^n)$ since we do not explicitly construct the entire $\boldsymbol{A}_n$. Based on this subroutine, the first algorithm referred to as the top-overlap method (Algorithm 1) solves the primal problem (3) with a limited set of stabilizers whose overlaps with the target quantum states are taken from largest or smallest ones. The second algorithm referred to as the Column Generation method (Algorithm 2), on the other hand, iteratively adds stabilizer states for the decomposition until all the inequality constraints in the dual problem (5) are satisfied, such that one has a guarantee for the exact solution.

In the following, we first present the fast overlap computation algorithm in Sec. 3.1, and then proceed to introduce two novel algorithms in Sec. 3.2 and 3.3, respectively. We show that these algorithms enable us to compute the exact RoM value up to $n = 8$ qubit system. In this case, the memory consumption for the main subroutine is suppressed by a factor of 10^8; compared to the entire $\boldsymbol{A}_n$ size 86 PiB, we can run the efficient subroutine with only 512 MiB.

3.1 Core subroutine: fast computation of stabilizer overlaps

First, we introduce the core subroutine in our work that computes the overlaps between the target state and pure stabilizer states, or the stabilizer overlaps in short. When we resort to a naive calculation, it requires time complexity of $\mathcal{O}(2^n|\mathcal{S}_n|)$ to compute all the stabilizer overlaps via the matrix-vector product of $\boldsymbol{A}_n^\top \boldsymbol{y}$, even if we utilize the sparsity of $\boldsymbol{A}_n$. We show this can be done exponentially faster per stabilizer state:

Theorem 1. *(Complexity of computing all stabilizer overlaps) Computation of $\boldsymbol{A}_n^\top \boldsymbol{y}$ can be done in time complexity of $\mathcal{O}(n|\mathcal{S}_n|)$ and space complexity of $\mathcal{O}(2^n)$.*

One of the most practical applications of this theorem is to apply to the computation of overlaps $\boldsymbol{A}_n^\top \boldsymbol{b}$ for the Pauli vector $\boldsymbol{b}$ of a given quantum state; as we later detail in Sec. 3.2, this technique is essential to scale up the RoM calculation to larger systems.

In order to show Theorem 1, first we introduce the Fast Walsh–Hadamard Transform (FWHT) algorithm that efficiently performs matrix-vector product operation, when there is a tensor product structure in the matrix [24]. Here we use the unnormalized Walsh–Hadamard matrix as $H_n := H^{\otimes n}$ where $H := \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and refer to the matrix-vector multiplication of H_n as the FWHT algorithm. As is evident from the well-known pseudocode provided in Appendix A, we can show that the computational cost is given as in the following lemma:

Lemma 1. *(Complexity of FWHT algorithm) Matrix-vector multiplication of unnormalized Walsh–Hadamard matrix can be done by in-place computation with time complexity of $\mathcal{O}(n2^n)$ and space complexity of $\mathcal{O}(2^n)$.*

Next, we show that $\boldsymbol{A}_n$ is essentially constructed by concatenating unnormalized Walsh–Hadamard matrices (see also Fig. 2). Let $\mathcal{W}_n$ denote a set of all matrices that can be expressed as sparsified form of $\begin{bmatrix} H_n \\ O \end{bmatrix}$ by reordering and flipping the signs of the rows appropriately, where O denotes a null matrix of size $(4^n - 2^n) \times 2^n$. We can state the following lemma (see Appendix D for the proof):

Lemma 2. *(Decomposition of $\boldsymbol{A}_n$ into Walsh–Hadamard matrix) For all $n \in \mathbb{N}$, there exists a constructive and efficient way of enumerating properly sparsified Walsh–Hadamard matrices $\{W_j\}_{j=1}^{|\mathcal{S}_n|/2^n}$ $(W_j \in \mathcal{W}_n)$ such that*

$$\boldsymbol{A}_n = \left[W_1 \cdots W_{|\mathcal{S}_n|/2^n} \right]. \tag{6}$$

By combining Lemma 1 and 2, we complete the proof of Theorem 1. As a direct corollary, we also obtain that the stabilizer fidelity defined as $F_{\mathrm{STAB}}(|\psi\rangle) = \max_{|\phi\rangle \in \mathcal{S}_n} |\langle\psi|\phi\rangle|^2$ [25] can be shown to be computed exponentially faster as well (see Appendix B for details):

Corollary 1. *(Complexity of computing stabilizer fidelity) Stabilizer fidelity can be computed with time complexity of $\mathcal{O}(n|\mathcal{S}_n|)$ and space complexity of $\mathcal{O}(2^n)$.*

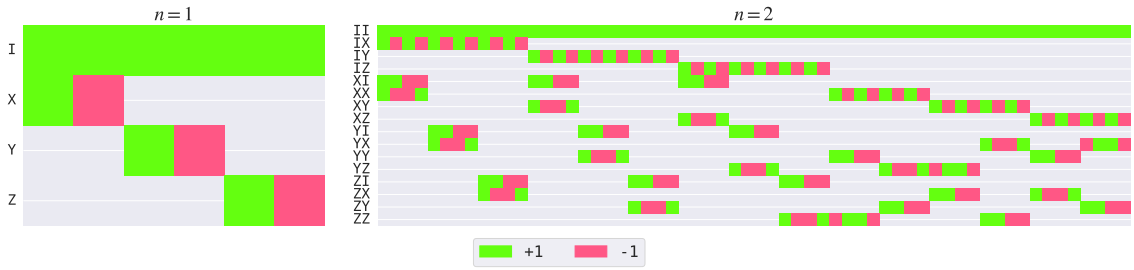


Figure 2: Visualization of $\boldsymbol{A}_n$ for $n = 1$ and $n = 2$.

It has been recognized that the stabilizer fidelity cannot be computed with moderate computational cost for $n > 5$ [14]. Meanwhile, our algorithm allows us to compute up to $n = 8$ in 4 hours. Note that this numerical experiment was conducted using C++17 compiled by GCC 9.4.0 and a cluster computer powered by Intel(R) Xeon(R) CPU E5-2640 v4 with 270 GB of RAM using 40 threads. Even if we use a *laptop* powered by Intel(R) Core(TM) i7-10510U CPU with 16 GB RAM, we can compute $n = 7$ in 2 minutes using 8 threads.

3.2 Top-overlap method for primal RoM

Using the efficient overlap calculation subroutine presented in Sec. 3.1, we propose a novel algorithm that computes the exact/approximate value of the RoM by utilizing the following properties: (i) by the nature of L^1 norm minimization problem, the solution of the optimal stabilizer decomposition is sparse [26], (ii) the stabilizer overlap is closely related to the optimal stabilizer decomposition (see Fig. 3 (a)). Concretely, as we provide the detail in Algorithm 1, we restrict the number of columns in $\boldsymbol{A}_n$ and consider only a fraction K of pure stabilizer states with the largest or smallest overlaps; the fraction of $1-K$ is neglected.

As a useful visualization to confirm the observation (ii), we show in Fig. 3(a) the distribution of stabilizer overlaps $\{\mathrm{Tr}[\rho\sigma_i]\}_i$ and their weights $\{x_i\}_i$ for random 4-qubit mixed state $\rho = \sum_{\sigma_i \in \mathcal{S}_n} x_i \sigma_i$. Indeed, we find strong correspondence between the stabilizer overlaps and weights. A similar property can be seen in various random instances in larger systems as well.

It is natural that stabilizer states with large overlaps have large weights. On the other hand, it seems quite counterintuitive that stabilizer states with small overlaps contribute non-negligibly. In this regard, we mention that, in the field of operations research, there is an approximation method called Orthogonal Matching Pursuit for L^1 norm minimization problem that greedily takes near-orthogonal basis to improve the solution [27, 28]. It can be understood that the orthogonal bases contribute to extending the effective dimension of the space spanned by the chosen basis, and thus are essential to enhance the quality of the approximation for random states.

As we highlight in Fig. 3(b), we only need a small fraction from the entire $\mathcal{S}_n$ in order to obtain a nearly-exact solution. We find that, for a random mixed state of $n = 4$ qubit system, it is sufficient to use a fraction of $K \sim 0.05$ in order to achieve an absolute error of 0.023. The fraction required to achieve similar accuracy for larger systems are orders of magnitude smaller as $K = 10^{-2}, 10^{-3}, 10^{-5}$ for $n = 5, 6, 7$, respectively. Meanwhile, we remark that we must take a significantly larger column set to assure the exact RoM value; the fraction is $K \sim 0.32$ for $n = 4$ qubit case. We provide further numerical details in Appendix H.1.

Algorithm 1: Top-overlap method for primal RoM

Input: Pauli vector $\boldsymbol{b}$ for quantum state ρ,
Fraction K $(0 < K \leq 1)$

Output: Approximate RoM

Compute overlap $\boldsymbol{a}^\top \boldsymbol{b}$ for each $\boldsymbol{a} \in \mathcal{A}_n$ using FWHT algorithm

$\mathcal{C} \leftarrow$Partial column set $\{\boldsymbol{a}\}$ with $K|\mathcal{S}_n|$ largest and smallest overlaps

return `SolveLP`$(\mathcal{C}, \boldsymbol{b})$

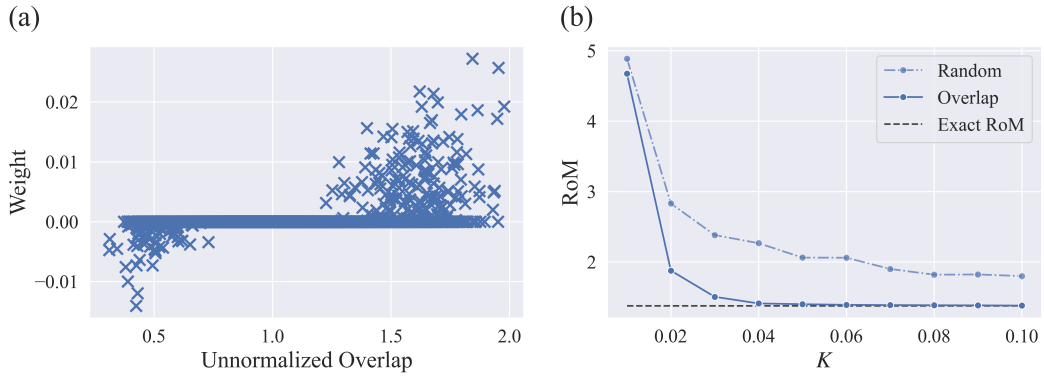


Figure 3: (a) Stabilizer overlaps $\mathrm{Tr}[\rho\sigma_i]$ and the weights x_i for random mixed state of $n = 4$ qubits. (b) RoM value computed under a restricted set of stabilizers of ratio $0 < K \leq 1$ for random mixed state of $n = 4$ qubits.

3.3 Column generation method for dualized RoM

Despite the significant improvement over naive methods, there are three major issues in Algorithm 1: (i) we cannot tell whether the partial column set $\mathcal{C} \subset \mathcal{A}_n$ is sufficient to yield the exact solution, (ii) there is no quantitative measure to judge the quality of the approximate solution, (iii) there is a large gap in the computational resource between "highly approximate" and "exact" solutions. While we do not address these problems explicitly for the primal formulation of RoM, we find that these issues are well addressed when we consider the dualized formalism instead.

Recall that the dualized formulation of RoM is given as

$$\mathcal{R}(\rho) = \max_{\boldsymbol{y}} \left\{ \boldsymbol{b}^\top \boldsymbol{y} \;\middle|\; -\mathbf{1} \leq \boldsymbol{A}_n^\top \boldsymbol{y} \leq \mathbf{1} \right\}. \tag{7}$$

Let us assume that we have computed the approximate value of the RoM under a restricted column set $\mathcal{C}$, and that the solver has returned a dual variable $\hat{\boldsymbol{y}}$, which is often the case for practical implementations. If $\hat{\boldsymbol{y}}$ does not obey all the constraints in Eq. (7), i.e., if there exists $\boldsymbol{a} \in \mathcal{A}_n$ such that $|\boldsymbol{a}^\top \hat{\boldsymbol{y}}| > 1$, then we must increase the size of the reduced column set $\mathcal{C}$ to include violated $\boldsymbol{a}$'s if we wish to obtain the exact value of the RoM (see also Fig. 4). Conversely, when there is no violation, then the solution is exact, due to the strong duality of the problem.

This discussion naturally motivates us to employ an iterative method that gradually takes constraints into account until there is no violation of the constraints at all. Such a strategy is known as the Column Generation technique in the field of operations research [31], and here we propose a method that unifies the knowledge of such technique

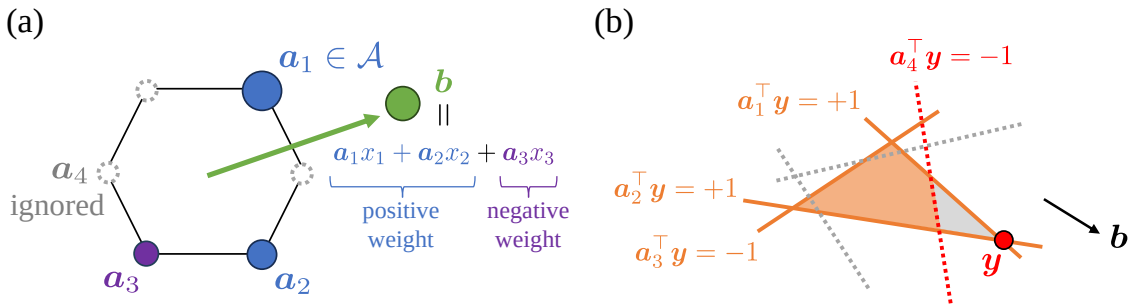


Figure 4: (a) Graphical description of the primal formalism for the RoM. The Pauli vector $\boldsymbol{b}$ of the target state is decomposed into a sum over those of pure stabilizer states $\boldsymbol{a}$ denoted by vertices on the stabilizer polytope, so that the L^1 norm of primal variables $\|\boldsymbol{x}\|_1$ is minimized. The restriction on columns of $\boldsymbol{A}_n$ is expressed by gray vertices that are eliminated from the decomposition. (b) Graphical description of the dual formalism for the RoM. The equality constraints in primal problem are now given as inequalities $-\mathbf{1} \leq \boldsymbol{A}_n^\top \boldsymbol{y} \leq \mathbf{1}$ for the dual variables $\boldsymbol{y}$, while the objective function is the inner product with $\boldsymbol{b}$. A solution $\boldsymbol{y}$ obtained from reduced column set $\mathcal{C}$ may violate some constraints, denoted by the red dotted line in the figure, so that one shall add more columns to improve the solution, while there are some columns denoted by gray dotted lines that do not affect the result.

Algorithm 2: Exact dual RoM calculation by Column Generation

Input: Pauli vector $\boldsymbol{b}$ of target state ρ
Output: Exact RoM $\mathcal{R}(\rho)$

$\mathcal{C} \leftarrow$ Partial set of $\mathcal{A}_n$ `/* Initialize using top and least overlaps */`
while *true* **do**
  $R, \boldsymbol{y} \leftarrow \texttt{SolveLP}(\mathcal{C}, \boldsymbol{b})$
  $\mathcal{C}' \leftarrow \left\{ \boldsymbol{a} \in \mathcal{A}_n \;\middle|\; \left| \boldsymbol{a}^\top \boldsymbol{y} \right| > 1 \right\}$ `/* Use of FWHT */`
  if $\mathcal{C}' = \emptyset$ **then**
    break
  $\mathcal{C} \leftarrow \mathcal{C} \cup \mathcal{C}'$
return R

and the FWHT algorithm introduced in Sec. 3.2. In particular, we compute the overlap $\boldsymbol{a}^\top \hat{\boldsymbol{y}}$ for every $\boldsymbol{a} \in \mathcal{A}_n$ at each iteration in order to improve the quality of the solution; the columns that violate the constraint is added to the reduced column set $\mathcal{C}$. By iteratively updating $\mathcal{C}$ until there is no violation, we obtain the exact RoM value. See also Algorithm 2 for the pseudocode.

In Fig. 5, we present the results of the numerical demonstration of Algorithm 2. We have initialized the partial column sets with $K = 10^{-5}, 10^{-8}$ for random mixed states of $n = 7, 8$ qubit systems, respectively. We can see that the numbers of violated columns decrease rapidly so that the exact solutions can be obtained after small number of iterations. Using the machines described as in Sec. 3.1, the total run time was 2 hours using the laptop for $n = 7$ and 2 days using the cluster computer for $n = 8$. Note that there is no nontrivial bound on the time complexity, and thus the algorithm is not guaranteed to yield exact solutions within realistic run time. However, we strongly expect that the exact value of RoM can be obtained with similar run time for average case, unless a malicious input state is provided. See Appendix H.1 for details.

We remark that there are two practical tricks to suppress $|\mathcal{C}|$ which results in drastic enhancement of the efficiency of the computation. First, we introduce a threshold d ($0 \leq$

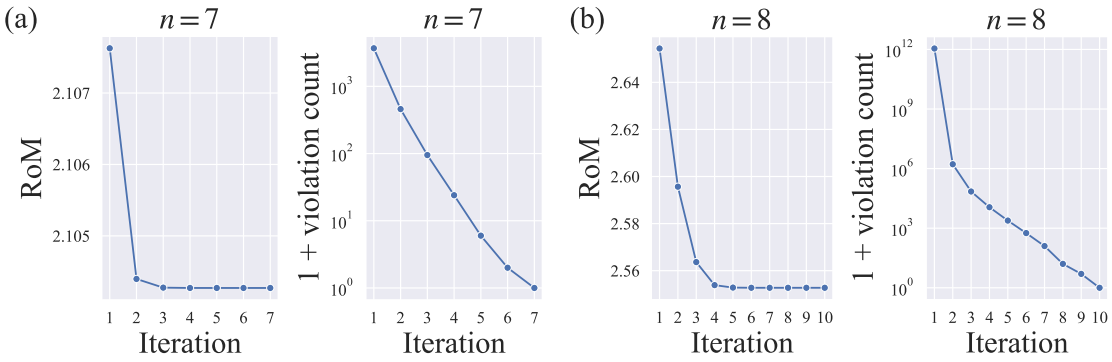


Figure 5: Demonstration of Algorithm 2 for random mixed state of (a) $n = 7$ qubits and (b) $n = 8$ qubits with the fraction K chosen as in Table 2. Here we display the value of approximate RoM and the number of columns that violate the inequality condition in Eq. (7). The number of violating columns quickly reduces to zero, which assures that the exact value of the RoM is obtained. The run time is 2 hours using the laptop for $n = 7$ and 2 days using the cluster computer for $n = 8$, whose specifications are provided in Sec. 3.1.

Qubit count n	Full-size $\boldsymbol{A}_n$	Our work	K
4	3 MiB	301 KiB	10^{-1}
5	379 MiB	4 MiB	10^{-2}
6	95 GiB	97 MiB	10^{-3}
7	86 TiB	499 MiB	10^{-5}
8	86 PiB	512 MiB	10^{-8}

Table 2: Memory size required to store the $\boldsymbol{A}$ matrix. The full-size $\boldsymbol{A}_n$ takes the entire columns into account, while the reduced $\boldsymbol{A}$ takes only the fraction K from the column set, as in the initialization step in Algorithm 2. We empirically find that the maximal memory consumption during the iterations are approximately 1.5 times larger than the values shown here. The data is based on sparse matrix format in SciPy [29]. The code for generating $\boldsymbol{A}_n$ and proposed algorithms are available via GitHub [30].

$d \leq 1$) to discard columns $\boldsymbol{a}_i$ that satisfy both $\left|\boldsymbol{a}_i^\top \boldsymbol{y}\right| < d$ and $x_i = 0$, where x_i is the corresponding primal variable. Second, instead of adding the entire violating columns to the reduced column set $\mathcal{C}$ at once, we set an upper bound on the number of the columns so that the memory consumption does not become infeasible. In particular, for the calculation given in Fig. 5, we have added only $K|\mathcal{S}_n|$ columns with either large or small overlaps at each iteration. As a consequence of these two tricks, we can significantly suppress the memory consumption. Concretely, we empirically find that $|\mathcal{C}|$ is no more than $2K|\mathcal{S}_n|$ in our numerical experiments.

3.4 Minimal feasible solution with accuracy guarantee

Exact solutions for large-scale systems may require prohibitively large computational resources, while we may still wish to compute a feasible solution. Here, we propose a method with minimal computational resources that is always guaranteed to yield a feasible solution. By utilizing the LP for the cover matrix that is later introduced in Theorem 2, we have obtained the approximate RoM value with its corresponding feasible solution $\boldsymbol{x}$ for random state of $n = 14$ qubits within a minute.

A technical contribution of this section is the construction of a reduced $\boldsymbol{A}$ matrix that

always guarantees a feasible decomposition of ρ into stabilizer states. While we guide readers for the proof to Appendix E, here we simply provide the core proposition:

Proposition 1. *(Existence of cover stabilizer set) Let S_n be a subset of $\mathcal{S}_n$ such that, for any $P \in \mathcal{P}_n$ there exists $|\psi\rangle \in S_n$ that satisfies $\{P, -P\} \cap \mathrm{Stab}(|\psi\rangle) \neq \emptyset$. Then, for all $n \in \mathbb{N}$, the size of the set is bounded as $|S_n| \geq 2^n + 1$, and one can construct S_n such that $|S_n| = 2^n + 1$.*

Using this proposition, it is straightforward to see that the following theorem holds:

Theorem 2. *(Existence of cover matrix) Let $S_n = \{|\psi_j\rangle\}_j$ be a minimal cover stabilizer set that is obtained from Proposition 1. Let the stabilizer group denoted as $\mathrm{Stab}(|\psi_j\rangle) = \langle Q_1^{(j)}, \ldots, Q_n^{(j)} \rangle$, and let $|\psi_{j,\chi}\rangle$ be defined for $\chi = (\chi_1, \ldots, \chi_n) \in \{0,1\}^n$ so that $\mathrm{Stab}(|\psi_{j,\chi}\rangle) = \langle (-1)^{\chi_1} Q_1^{(j)}, \ldots, (-1)^{\chi_n} Q_n^{(j)} \rangle$, and $\boldsymbol{a}_{j,\chi}$ be its Pauli vector. Then, by defining $M := (\ldots, \boldsymbol{a}_{j,\chi}, \ldots)$, it is guaranteed that there exists a feasible solution to $\mathtt{Prob}(M, \boldsymbol{b})$ that can be solved with time complexity of $\mathcal{O}(n4^n)$.*

Proof. The cover stabilizer set S_n is constructed so that, for any $P \in \{I, X, Y, Z\}^{\otimes n} \setminus \{I^{\otimes n}\}$, there is a unique $|\psi_j\rangle$ that satisfies $P|\psi_j\rangle \in \{|\psi_j\rangle, -|\psi_j\rangle\}$. Consequently, for all $|\psi_j\rangle \in S_n$ $(1 \leq j \leq 2^n + 1)$, one may enumerate the set of indices $\mathcal{I}_j = \{\nu_j \mid P_{\nu_j}|\psi_j\rangle \in \{|\psi_j\rangle, -|\psi_j\rangle\}\}$. By extracting the elements from $\boldsymbol{b}$ as $\boldsymbol{b}_j = (b_{\nu_j})_{\nu_j \in \mathcal{I}_j}$, the primal problem in the reduced bases is equivalent to $H_n \boldsymbol{x}_j = \boldsymbol{b}_j$ where H_n is the unnormalized Walsh–Hadamard matrix. Therefore, for each j we can simply compute by the FWHT algorithm as

$$\boldsymbol{x}_j = (1/2^n) H_n \boldsymbol{b}_j, \tag{8}$$

and then combine all the solutions to construct $\boldsymbol{x}$ that is assured to be feasible by the nature of the cover stabilizer set S_n (see Fig. 6 for example in $n = 2$). Since each FWHT algorithm can be performed with time complexity of $\mathcal{O}(n2^n)$, the total time complexity is $\mathcal{O}(n4^n)$. □

By using the cover matrix, we have We remark that we can not only ensure feasibility by Theorem 2 but also provide an accuracy bound for the algorithm. Let R_{FWHT} be the approximate RoM value obtained from the minimal feasible solution. We can show that the following holds (see Appendix F):

Lemma 3. *(Accuracy bound of minimal feasible solution) Let $\boldsymbol{b}_j$ drawn uniformly from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ for all j. Then,*

$$\mathbb{E}_{\boldsymbol{b}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})} \left[\frac{R_{\mathrm{FWHT}}}{\|\rho\|_{\mathrm{st}}} \right] \approx 2^{n/2}. \tag{9}$$

Note that this lemma allows us to expect that $R_{\mathrm{FWHT}} \leq 2^{n/2} \mathcal{R}(\rho)$ for a random state with high probability at the asymptotic limit of large n, while we practically observe convergence already at $n \leq 10$. While R_{FWHT} significantly deviates from the exact RoM value, we envision that the feasible solution can be utilized as, e.g., approximate initialization.

4 Quantum resource of multiple magic states

One of the most important applications of RoM is to measure the total nonstabilizerness of multiple magic states that are decoupled from each other. For instance, one may wish

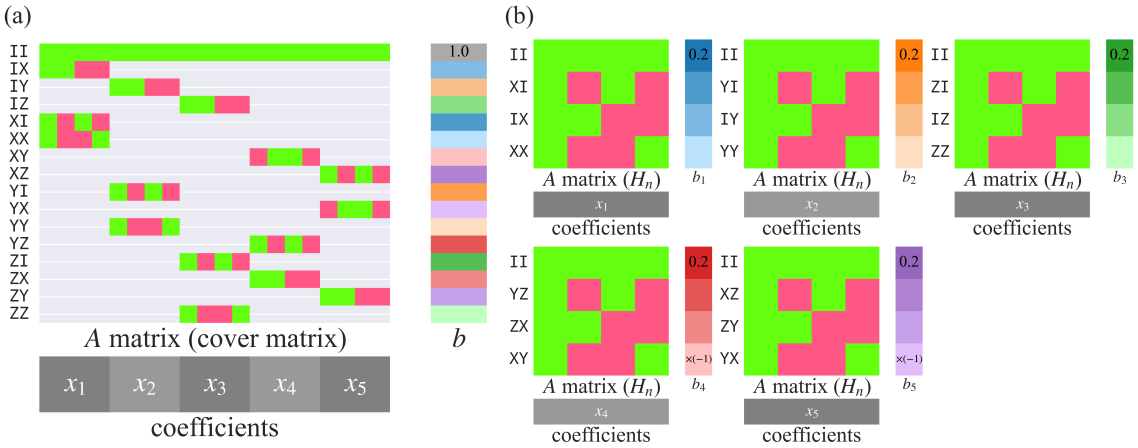


Figure 6: Graphical description of (a) the cover matrix and (b) application of FWHT algorithm for individual segments. Since the identity operator is present in any stabilizer group, we divide the vector element equally by the size of the cover stabilizer set $|S_n| = 2^n + 1$, which yields the vector element 0.2 for each submatrix for $n = 2$. Also, note that the signs of some vector elements in b_4 and b_5 are flipped (denoted by $\times(-1)$).

to evaluate the amount of magic resources for copies of multiple states $\rho^{\otimes n}$ to estimate the upper bound on the number of generatable clean magic states. In general, this could even include situations where the quantum states are nonequivalent, such as partially decoupled states $\bigotimes_i \rho_i$. Note that there exists a previous work by Heinrich and Gross [20] that has utilized the symmetry of some pure magic states such as $|H\rangle\langle H| = \frac{1}{2}\left(I + \frac{1}{\sqrt{2}}(X+Y)\right)$ and $|F\rangle\langle F| = \frac{1}{2}\left(I + \frac{1}{\sqrt{3}}(X+Y+Z)\right)$ to scale up the simulation up to $n = 26$ qubits, we still lack a method to investigate general quantum states with partial disentangled structure (see Fig. 7).

In this section, we apply the algorithms proposed in Sec. 3 to practical problems; copies of identical quantum states $\rho^{\otimes n}$ and partially disentangled quantum states $\bigotimes_i \rho_i$. In particular, we first discuss the case of permutation symmetric state $\rho^{\otimes n}$ in Sec. 4.1, and then also consider general partially disentangled states in Sec. 4.2.

4.1 Copies of general quantum states

When the target quantum state is given as an identical copies of a quantum state as $\rho^{\otimes n}$, we may compress the size of $\boldsymbol{A}_n$ by utilizing the permutation symmetry to combine multiple columns of $\mathcal{A}_n$. In this work, we have employed the compression method for $\mathcal{A}_n$ proposed in Ref. [20] to define a set of permutation symmetric columns $\mathcal{Q}_n$. As in Ref. [20], we also make use of the data by Danielsen [32]. This enables us to run the algorithm to obtain the exact solutions for $n \leq 7$ qubits.

Beyond $n = 7, 8$ qubits, it is not realistic to obtain the exact solution even when we use $\boldsymbol{Q}_n$ instead of $\boldsymbol{A}_n$. Note that $\boldsymbol{Q}_n$ is the matrix given by ordering all the columns in $\mathcal{Q}_n$ whose number of rows (and correspondingly that of $\boldsymbol{b}$) is reduced by permutation symmetry as well. Therefore, here we propose an approximate method that performs divide-and-conquer computation. As we present the details in Algorithm 3, we consider all possible decomposition of m qubits into two groups with j and k qubits ($j + k = m$), and compute the optimal stabilizer decomposition. If one has stored the solution of the LP for $\texttt{Prob}(\boldsymbol{Q}_i, \boldsymbol{b}_i)$ for $i < m$, one can simply load the result. We take the tensor product of

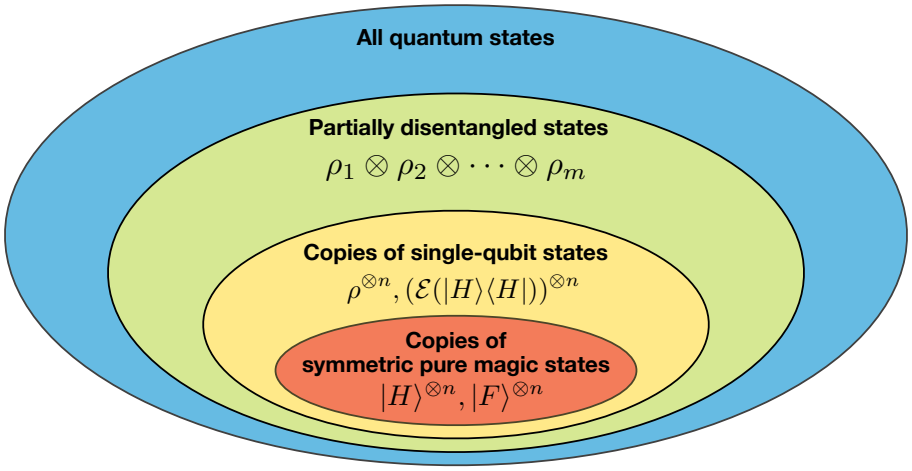


Figure 7: Hierarchy of partially disentangled quantum states.

Algorithm 3: Approximate RoM for permutation symmetric states

Data: Compressed column set $\mathcal{Q}_n$ of matrix $\boldsymbol{Q}_n$
Input: Positive integer n, k $(n \geq k)$, Pauli vector $\boldsymbol{b}$ for target state ρ
Output: Approximate RoM value R_i of $\rho^{\otimes i}$ $(i = 1, \ldots, n)$

for $i \leftarrow 1$ **to** k **do**
  $R_i, \mathcal{C}_i \leftarrow \texttt{SolveLP}(\mathcal{Q}_n, \boldsymbol{b})$
for $i \leftarrow k+1$ **to** n **do**
  $\mathcal{C}' \leftarrow \emptyset$
  for $l \leftarrow 1$ **to** $\lfloor i/2 \rfloor$ **do**
    $m \leftarrow i - l$
    $\mathcal{C}' \leftarrow \mathcal{C}' \cup \{\rho_l \otimes \rho_m \mid \rho_l \in \mathcal{C}_l, \rho_m \in \mathcal{C}_m\}$
  $R_i, \mathcal{C}_i \leftarrow \texttt{SolveLP}(\mathcal{C}', \boldsymbol{b})$
return $(R_1, \ldots, R_n)$

stabilizers with nonzero weights as $\{\rho_j \otimes \rho_k \mid \rho_j \in \mathcal{C}_j, \rho_k \in \mathcal{C}_k\}$. By taking the union over all states to construct $\mathcal{C}_m$, we compute the approximate value of RoM, which is assured to be less than the product of RoMs computed for subsystems.

Figure 8 shows the results of a numerical demonstration of the proposed algorithm applied to copies of pure magic state $|H\rangle$, pure random state, and mixed random state. Using the exact stabilizer decomposition up to $k = 7$ qubits, we have successfully computed the approximate RoM value up to $n = 17$ for the pure and mixed random state, while the compressed column set size $|\mathcal{C}_n|$ is significantly smaller for $|H\rangle$ so that we have reached $n = 21$. While this is not as large as $n = 26$ reported in Ref. [20], we emphasize that the present work is based on an algorithm that is agnostic to the internal symmetry of the single-qubit state. As a remark, we mention that the approximate RoM values for copies of pure magic states are almost identical to those presented in Ref. [20]; the value was at most 1.007 times larger.

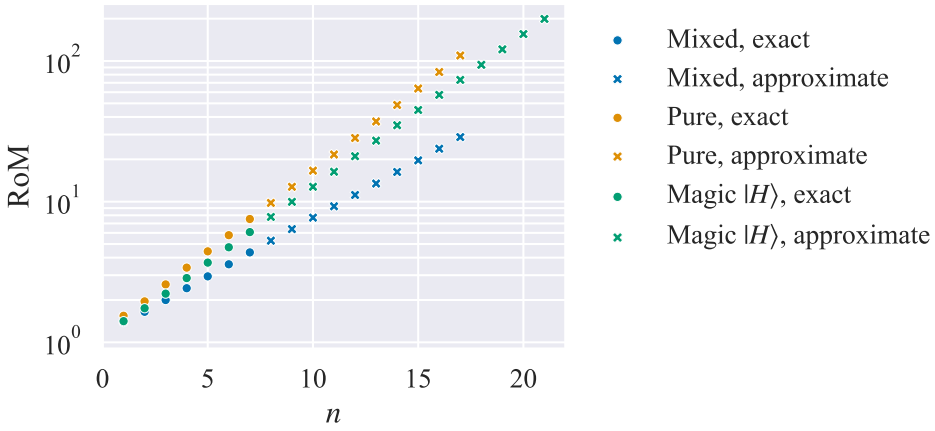


Figure 8: Exact and approximate RoM values computed for n copies of single-qubit state $\rho^{\otimes n}$. The blue, orange, and green data denote the RoM value for random mixed state, random pure state, and the magic state $|H\rangle$. The circle (crossed) points indicate that the solution is exact (approximate). The approximate values are computed from the exact solutions for $n \leq 7$ qubits.

4.2 Partially disentangled states

Let us assume that we are interested in a general partially decoupled state of m subsystems as $\rho = \bigotimes_{i=1}^{m} \rho_i$, where $\sum_{i=1}^{m} n_i = n$ with n_i being the qubit count of i-th subsystem and ρ_i corresponding to the local quantum state. In similar to the previous section, we may first compute the optimal decomposition for each subsystem, and then take tensor products over non-zero weight stabilizers to construct reduced basis for the total system. In this regard, we first show that assures an upper bound with small computational effort:

Proposition 2. *(Approximate RoM from multiplicativity) Let ρ_i be given for the i-th subsystem, $\mathcal{R}(\rho_i)$ and $\boldsymbol{x}_i$ be the exact solution for* $\texttt{Prob}(\boldsymbol{A}_{n_i}, \boldsymbol{b}_i)$*. Then,*

$$R = \prod_{i=1}^{m} \mathcal{R}(\rho_i), \quad \boldsymbol{x} = \bigotimes_{i=1}^{m} \boldsymbol{x}_i, \tag{10}$$

is one of the exact solutions for $\texttt{Prob}(\bigotimes_{i=1}^{m} \boldsymbol{A}_{n_i}, \boldsymbol{b})$ *where* $\bigotimes_{i=1}^{m} \boldsymbol{A}_{n_i} \subset \boldsymbol{A}_n$

Proof. See Appendix G. □

By noting the submultiplicativity of exact RoM [18], i.e., $\mathcal{R}(\bigotimes_i \rho_i) \leq \prod_{i=1}^{m} \mathcal{R}(\rho_i)$, it is natural to expect that one can further improve the approximate RoM value by extending the column set from those used in $\bigotimes_{i=1}^{m} \boldsymbol{A}_{n_i}$. For instance, one may group several subsystems so that each partial LP consists of 6 or 7-qubit systems.

We find that it is more effective to consider multiple variations to divide subsystems. As we show the pseudocode in Algorithm 4, we may divide subsystems into groups so that the exact (or highly accurate approximate) value of the RoM can be computed for each group. By comparing the product of those values for various decompositions, we take the minimal value as the approximate RoM. For instance, in the case of a 15-qubit system that is decoupled into 5 subsystems as $\rho = \rho_1 \otimes \rho_2 \otimes \rho_3 \otimes \rho_4 \otimes \rho_5$, where each ρ_i is a 3-qubit state. One may compute the approximate value as $\mathcal{R}(\rho_1 \otimes \rho_2) \times \mathcal{R}(\rho_3 \otimes \rho_4) \times \mathcal{R}(\rho_5)$ or $\mathcal{R}(\rho_1) \times \mathcal{R}(\rho_2 \otimes \rho_3) \times \mathcal{R}(\rho_4 \otimes \rho_5)$, for instance, and take the minimal value as the approximate output. While there could be combinatorially many possibilities for such

Algorithm 4: Optimization of subsystem division

Input: Target state $\rho = \bigotimes_{i=1}^{m} \rho_i$
Subsystem decomposition

Output: Approximate RoM value

foreach *Decomposition of* ρ_i **do**
  $R_j \leftarrow \prod \mathcal{R}(\bigotimes \rho_i)$ `/* Approximate RoM value for each decomposition */`
return $\min_j R_j$

groupings in general, we expect that the difference is not significant when ρ_i resemble each other, e.g., when we simulate the RoM of noisy magic states. One may also speed up the computation by brute-force parallelization if needed.

5 Discussion

In this work, we have proposed a systematic procedure to compute the RoM value to surpass the state-of-the-art results for random arbitrary states, multiple copies of single-qubit magic states, and partially disentangled quantum states. We have presented the core subroutine that is capable of computing the overlap between the target state and a pure stabilizer state with exponentially improved time complexity per state and superexponentially improved space complexity in total. Based on the efficient overlap simulation subroutine, we have proposed algorithms for arbitrary quantum states that significantly reduce the computational cost by reducing the number of stabilizer states based on the overlap values, so that RoM for $n = 8$ qubit state can be computed exactly with approximately 10^8-fold reduction in memory consumption. We have also proposed algorithms to incorporate the nature of the target quantum state, such as the permutation symmetry between multiple copies of states and the partially decoupled structure for inhomogeneous magic resources, and have numerically shown that we can scale the approximate RoM calculation up to $n = 17$ qubits.

Numerous future directions can be envisioned. First, it is intriguing to seek generalization to other quantum resource measures. Since the core subroutine in this work only assumes that the total system is composed of local systems with discrete degrees of freedom, we envision that our work can be applied to other resource monotones that are formulated with L^p norm optimization (in particular $p = 0, 1$) such as stabilizer extent [25], channel robustness [23], negativity [19]. In particular, it is nontrivial if we can extend the framework when the pure free states constitute a continuous set, such as in the case of fermionic non-Gaussianity [33, 34, 35]. Second, it is interesting to investigate whether it is possible to further scale up computations for weakly decoupled states such as tensor network states. While exact computation may require as costly calculation as in the generic case, we may perform approximate computation with an accuracy bound that depends on the entanglement.

Acknowledgements.— We would like to thank T. Oki for valuable comments on the manuscript. N.Y. wishes to thank JST PRESTO No. JPMJPR2119 and the support from IBM Quantum. This work was supported by JST Grant Number JPMJPF2221. This work was supported by JST ERATO Grant Number JPMJER2302 and JST CREST Grant Number JPMJCR23I4, Japan.

References

[1] Daniel Gottesman. "The Heisenberg Representation of Quantum Computers" (1998). arxiv:quant-ph/9807006.

[2] Michael A. Nielsen and Isaac L. Chuang. "Quantum Computation and Quantum Information: 10th Anniversary Edition". Cambridge University Press. (2010).

[3] Sergei Bravyi and Alexei Kitaev. "Universal Quantum Computation with ideal Clifford gates and noisy ancillas". Physical Review A **71**, 022316 (2005).

[4] Daniel Litinski. "A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery". Quantum **3**, 128 (2019).

[5] Dominic Horsman, Austin G Fowler, Simon Devitt, and Rodney Van Meter. "Surface code quantum computing by lattice surgery". New Journal of Physics **14**, 123011 (2012).

[6] Austin G. Fowler and Craig Gidney. "Low overhead quantum computation using lattice surgery" (2019). arxiv:1808.06709.

[7] Craig Gidney and Martin Ekerå. "How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits". Quantum **5**, 433 (2021).

[8] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush. "Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction". PRX Quantum **2**, 030305 (2021).

[9] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. "Quantum computing enhanced computational catalysis". Physical Review Research **3**, 033055 (2021).

[10] Nobuyuki Yoshioka, Tsuyoshi Okubo, Yasunari Suzuki, Yuki Koizumi, and Wataru Mizukami. "Hunting for quantum-classical crossover in condensed matter problems". npj Quantum Information **10**, 45 (2024).

[11] Sergey Bravyi, Graeme Smith, and John Smolin. "Trading classical and quantum computational resources". Physical Review X **6**, 021043 (2016).

[12] Emanuele Tirrito, Poetri Sonya Tarabunga, Gugliemo Lami, Titas Chanda, Lorenzo Leone, Salvatore F. E. Oliviero, Marcello Dalmonte, Mario Collura, and Alioscia Hamma. "Quantifying nonstabilizerness through entanglement spectrum flatness". Physical Review A: Atomic, Molecular, and Optical Physics **109**, L040401 (2024).

[13] Oliver Hahn, Alessandro Ferraro, Lina Hultquist, Giulia Ferrini, and Laura García-Álvarez. "Quantifying Qubit Magic Resource with Gottesman-Kitaev-Preskill Encoding". Physical Review Letters **128**, 210502 (2022).

[14] Tobias Haug, Soovin Lee, and M. S. Kim. "Efficient quantum algorithms for stabilizer entropies" (2023). arxiv:2305.19152.

[15] James R. Seddon, Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl T. Campbell. "Quantifying Quantum Speedups: Improved Classical Simulation From Tighter Magic Monotones". PRX Quantum **2**, 010345 (2021).

[16] Zi-Wen Liu and Andreas Winter. "Many-body quantum magic". PRX Quantum **3**, 020333 (2022).

[17] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. "Stabilizer Rényi Entropy". Physical Review Letters **128**, 050402 (2022).

[18] Mark Howard and Earl T. Campbell. "Application of a resource theory for magic states to fault-tolerant quantum computing". Physical Review Letters **118**, 090501 (2017).

[19] Hakop Pashayan, Joel J. Wallman, and Stephen D. Bartlett. "Estimating outcome probabilities of quantum circuits using quasiprobabilities". Physical Review Letters **115**, 070501 (2015). arxiv:1503.07525.

[20] Markus Heinrich and David Gross. "Robustness of Magic and Symmetries of the Stabiliser Polytope". Quantum **3**, 132 (2019). arxiv:1807.10296.

[21] Scott Aaronson and Daniel Gottesman. "Improved Simulation of Stabilizer Circuits". Physical Review A **70**, 052328 (2004). arxiv:quant-ph/0406196.

[22] Héctor J. García, Igor L. Markov, and Andrew W. Cross. "On the geometry of stabilizer states". Quantum Info. Comput. **14**, 683–720 (2014).

[23] James Seddon and Earl Campbell. "Quantifying magic for multi-qubit operations". Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **475**, 20190251 (2019).

[24] Fino and Algazi. "Unified Matrix Treatment of the Fast Walsh-Hadamard Transform". IEEE Transactions on Computers **C-25**, 1142–1146 (1976).

[25] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. "Simulation of quantum circuits by low-rank stabilizer decompositions". Quantum **3**, 181 (2019).

[26] Gilbert Strang. "Linear algebra and learning from data". Wellesley-Cambridge Press. (2019).

[27] Michael Elad. "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing". Springer. New York, NY (2010).

[28] David L. Donoho and Yaakov Tsaig. "Fast Solution of ℓ_1-Norm Minimization Problems When the Solution May Be Sparse". IEEE Transactions on Information Theory **54**, 4789–4812 (2008).

[29] SciPy. "scipy.sparse.csc_matrix". SciPy.

[30] Hiroki Hamaguchi, Kou Hamada, and Nobuyuki Yoshioka. "RoM-handbook". GitHub Repository (2024).

[31] Guy Desaulniers, Jacques Desrosiers, and Marius M. Solomon Solomon, editors. "Column Generation". Springer US. (2005).

[32] Danielsen Lars Eirik. "Database of self-dual quantum codes". link.

[33] Beatriz Dias and Robert Koenig. "Classical simulation of non-Gaussian fermionic circuits" (2023). arxiv:2307.12912.

[34] Oliver Reardon-Smith, Michał Oszmaniec, and Kamil Korzekwa. "Improved simulation of quantum circuits dominated by free fermionic operations" (2024). arxiv:2307.12702.

[35] Joshua Cudby and Sergii Strelchuk. "Gaussian decomposition of magic states for matchgate computations" (2023). arxiv:2307.12654.

[36] Lukas Hantzko, Lennart Binkowski, and Sabhyata Gupta. "Tensorized Pauli decomposition algorithm" (2023) arXiv:2310.13421.

[37] Tyson Jones. "Decomposing dense matrices into dense Pauli tensors" (2024). arxiv:2401.16378.

Algorithm 5: Fast Walsh–Hadamard Transform (FWHT) algorithm

Input: $v \in \mathbb{R}^{2^n}$
Output: In-place computation result of $v^\top H_n$

```
Function FWHT(v)
    h ← 1
    while h < len(v) do
        for i ← 0 to len(v) − 2h by 2h do
            for j ← i to i + h − 1 do
                x ← v[j]
                y ← v[j + h]
                v[j] ← x + y
                v[j + h] ← x − y
            end
        end
        // if normalize, v ← v/2
        h ← 2h
    end
```


[38] Sebastián Vidal Romero and Juan Santos-Suárez. "PauliComposer: Compute tensor products of Pauli matrices efficiently". Quantum Information Processing **22**, 449 (2023).

[39] Tyson Jones. "Densepaulidecomposer". GitHub Repository (2024).

[40] G.I. Struchalin, Ya. A. Zagorovskii, E.V. Kovlakov, S.S. Straupe, and S.P. Kulik. "Experimental Estimation of Quantum State Properties from Classical Shadows". PRX Quantum **2**, 010307 (2021).

[41] Regina C. Elandt-Johnson and Norman L. Johnson. "Survival Models and Data Analysis". Wiley Series in Probability and Statistics. Wiley. (1999).

[42] Arno Jaeger and Bertram Mond. "On direct sums and tensor products of linear programs". Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **3**, 19–31 (1964).

A Pseudocode of fast Walsh–Hadamard transform algorithm

Here, we provide the pseudocode of the Fast Walsh–Hadamard Transform (FWHT) algorithm. Note that we omitted the normalization factor in the pseudocode since the H_n itself is unnormalized. Clearly, the in-place computation allows the time complexity of $\mathcal{O}(n2^n)$ and the space complexity of $\mathcal{O}(2^n)$.

B Basic properties of inner product in Pauli vector representation

In this section, we provide a brief review of the basic properties of inner products in Pauli vector representation.

B.1 Complexity of computing Pauli vector representation

Let ρ be an n-qubit quantum state whose Pauli vector representation is given by $b_j = \mathrm{Tr}[\rho P_j]$ where P_j is the j-th Pauli operator. In order to compute all the elements, naively the computational complexity scales as $\mathcal{O}(8^n)$ even if we use the sparse structure of each Pauli matrix. In the following, we show that we can perform an in-place computation that exponentially reduces the time complexity:

Lemma 4. *(Complexity of computing Pauli vector) Given the full density matrix representation of n-qubit quantum state ρ, its Pauli vector representation can be computed with time complexity of $\mathcal{O}(n4^n)$.*

Proof. First let us introduce a map from an n-qubit density matrix to a $2n$-qubit statevector as follows:

$$\rho = \sum_{i,j} \rho_{i_1 \cdots i_n, j_1 \cdots j_n} |i_1 \cdots i_n\rangle\langle j_1 \cdots j_n| \mapsto \sum_{i,j} \rho_{i_1 \cdots i_n, j_1 \cdots j_n} |i_1, j_1, \ldots, i_n, j_n\rangle. \tag{11}$$

Note that this is different from the well-known Choi map $\rho \mapsto \sum_{ij} \rho_{ij} |i\rangle |j\rangle$, and thus we refer to it as modified Choi vectorization. We introduce a modified Choi vector $\boldsymbol{c}$ such that c_k denotes the k-th element, which can be obtained practically via extracting the matrix elements of ρ in the Z-order curve. Then, we find that $\boldsymbol{c}$ is related with the Pauli vector representation $\boldsymbol{b}$ as

$$\boldsymbol{b} = M_n \boldsymbol{c}, \tag{12}$$

where the transformation matrix M_n is defined as

$$M_n := M^{\otimes n}, \quad M := \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & i & -i & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}. \tag{13}$$

Similar to the FWHT algorithm as provided in Algorithm 5, in-place computation for such a tensor-product structure can be done with time complexity of $\mathcal{O}(n4^n)$ and space complexity of $\mathcal{O}(4^n)$, which completes the proof. □

We describe the numerical comparison of our Pauli decomposition algorithm with other previous studies [36, 37, 38]. The algorithms are implemented with C++ and run on the laptop. Our C++ implementation is mostly based on the Python implementation by Jones [39], except for the iterative algorithm by Hantzko et al. since it is not included in the Python implementation. The algorithms run on 50 random density matrices; both real and imaginary parts of each entry are independently and uniformly sampled from $[0, 1)$.

Fig. 9 shows the benchmark of Pauli decomposition. It shows that our Pauli decomposition algorithm is the fastest. Although our algorithm and the algorithms by Hantzko et al. both have the best time complexity $\mathcal{O}(n4^n)$ among the algorithms, our algorithm is faster by a constant factor. We consider it is because our algorithm uses in-place computation, which is cache efficient.

B.2 Pauli vector overlap and quantum state fidelity

The fidelity between quantum states ρ_1 and ρ_2 is defined as

$$F(\rho_1, \rho_2) := \mathrm{Tr}\left(\sqrt{\rho_1^{1/2} \rho_2 \rho_1^{1/2}}\right). \tag{14}$$

We can show that the fidelity is closely related to the Pauli vector as follows:

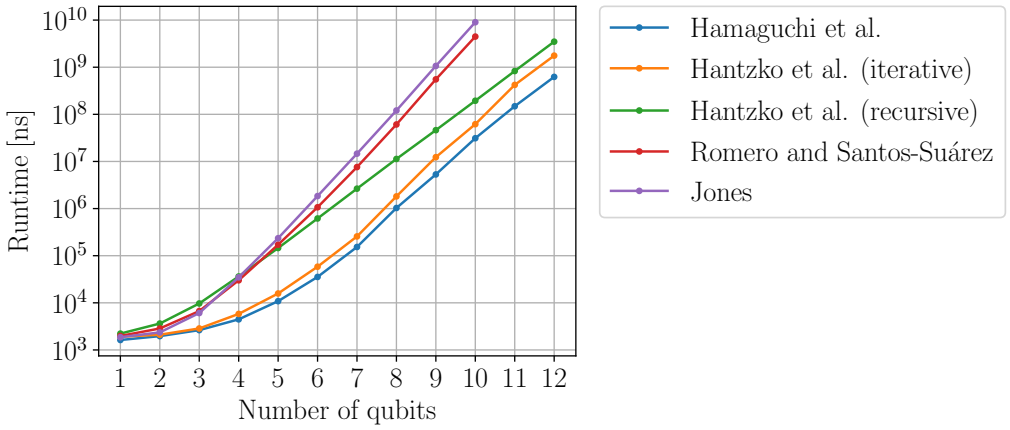


Figure 9: Comparison of Pauli decomposition algorithms. The algorithms are implemented with C++ and run on the laptop. The first three methods in the legend have a time complexity of $\mathcal{O}(n4^n)$, while the next two methods have a time complexity of $\mathcal{O}(8^n)$.

Lemma 5. *Let the Pauli vector representation of quantum states ρ_1 and ρ_2 be $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$, respectively. If at least one of ρ_1 and ρ_2 is a pure state, then the overlap $\frac{1}{2^n}\boldsymbol{b}_1^\top \boldsymbol{b}_2$ coincides with $F^2(\rho_1, \rho_2)$.*

Proof. It follows directly from the orthogonality of Pauli operators that $\frac{1}{2^n}\boldsymbol{b}_1^\top \boldsymbol{b}_2 = \mathrm{Tr}[\rho_1\rho_2]$. By taking $\rho_1 = |\psi\rangle\langle\psi|$ to be a pure state, we can show that $F^2(\rho_1, \rho_2) = \langle\psi|\rho_2|\psi\rangle = \mathrm{Tr}[\rho_1\rho_2]$. □

It follows directly that the stabilizer fidelity can also be computed from the overlap computation between Pauli vectors. By denoting the Pauli vector of a pure stabilizer state $|\phi\rangle \in \mathcal{S}_n$ as $\boldsymbol{b}_\phi$, we can show the following:

Corollary 2. *(Stabilizer fidelity as Pauli vector inner product) Let $\rho = |\psi\rangle\langle\psi|$ be a pure state with its Pauli vector given as $\boldsymbol{b}_\rho$. Then,*

$$\frac{1}{2^n}\left(\max_{|\phi\rangle\in\mathcal{S}_n} \boldsymbol{b}_\rho^\top \boldsymbol{b}_\phi\right) = \max_{|\phi\rangle\in\mathcal{S}_n} |\langle\phi|\psi\rangle|^2 = F_{\mathrm{STAB}}(|\psi\rangle). \tag{15}$$

Finally, we provide a fact that is useful as a preknowledge regarding the distribution of overlaps in addition to the fact that $0 \leq \boldsymbol{b}_\rho^\top \boldsymbol{b}_\phi \leq 2^n$.

Lemma 6. *(Bound on overlap counts) Let ρ be an arbitrary n-qubit quantum state. Then, for all $n \in \mathbb{N}$, the count on the pure stabilizer states satisfies the following:*

$$\#\{\phi_- \in \mathcal{S}_n \mid \boldsymbol{b}_\rho^\top \boldsymbol{b}_{\phi_-} \in [0, 1]\} \geq |\mathcal{S}_n|/2^n, \tag{16}$$

$$\#\{\phi_+ \in \mathcal{S}_n \mid \boldsymbol{b}_\rho^\top \boldsymbol{b}_{\phi_+} \in [1, 2^n]\} \geq |\mathcal{S}_n|/2^n. \tag{17}$$

Proof. Let us take an arbitrary $\phi \in \mathcal{S}_n$, and consider a set Φ of 2^n stabilizer states whose stabilizer generators are equivalent to ϕ except for their signs. Then, it holds that

$$\sum_{\hat{\phi}\in\Phi} \boldsymbol{b}_\rho^\top \boldsymbol{b}_{\hat{\phi}} = 2^n. \tag{18}$$

Therefore, if we assume that the overlap is either all $\boldsymbol{b}_\rho^\top \boldsymbol{b}_{\hat{\phi}} < 1$ or all $\boldsymbol{b}_\rho^\top \boldsymbol{b}_{\hat{\phi}} > 1$, then this contradicts with Eq. (18), which completes the proof. □

C Check matrix representation

One of the most well-known concise representations of a stabilizer state is the stabilizer tableau [21], which uses binary representation and the sign of each stabilizer generator. Meanwhile, in order to certify if a given set of Pauli operators suffices as an n-qubit stabilizer group generators, we may neglect the sign information and focus only on the commutativity and linear independence of generators. Here, we briefly introduce the check matrix representation [2] for the sake of convenience in discussion of Appendix D and E.

Let an n-qubit stabilizer group be given as $\langle g_1, \ldots, g_n \rangle$ so that $-I^{\otimes n}$ is not included as an element. Let each generator be expressed in the binary symplectic form as

$$g_i = (-1)^{\chi_i} X_i^{\boldsymbol{\alpha}_i} Z^{\boldsymbol{\beta}_i}, \tag{19}$$

where $X^{\boldsymbol{\alpha}_i} := X^{\alpha_{i,1}} \otimes \cdots \otimes X^{\alpha_{i,n}}$ $(\alpha_{i,j} \in \{0,1\})$, $Z^{\boldsymbol{\beta}_i} := Z^{\beta_{i,1}} \otimes \cdots \otimes Z^{\beta_{i,n}}$ $(\beta_j \in \{0,1\})$, and $\chi_i \in \{0,1\}$. The check matrix representation of the stabilizer group is given as $n \times 2n$ matrix as

$$\boldsymbol{C} = [\boldsymbol{X}\ \boldsymbol{Z}], \tag{20}$$

where $(\boldsymbol{X})_{i,j} = \alpha_{i,j}$ and $(\boldsymbol{Z})_{i,j} = \beta_{i,j}$ denote the (i,j) elements of the left and right half of the check matrix, respectively. Note that such a representation is unique except for the degrees of signs. Using the check matrix representation, we may confirm the linear independence and the commutativity of stabilizer generators. Such useful properties can be summarized as follows:

Lemma 7. *(Linear independence of stabilizer generators, Proposition 10.3 of Ref. [2]) The generators of a stabilizer group are mutually independent if and only if the rows of the corresponding check matrix $\boldsymbol{C}$ are linearly independent, i.e., full row rank.*

Lemma 8. *(Commutativity of stabilizer generators) Let $G = \{g_1, \ldots, g_n\}$ be a set of n-qubit stabilizer generators and $\boldsymbol{C}$ be the check matrix corresponding to G. Then, the following two conditions are equivalent:*

(i) $[g_i, g_j] = 0$ *for all* $g_i, g_j \in G$.

(ii) $\boldsymbol{C}\begin{pmatrix} O & I_n \\ I_n & O \end{pmatrix}\boldsymbol{C}^\top = 0.$

Note that the operations for the check matrix are done modulo two.

D Proof of Lemma 2

Here we prove Lemma 2 in the main text that states that $\boldsymbol{A}_n$ can be decomposed into sparsified Walsh–Hadamard matrices.

Lemma 9. *(Restatement of Lemma 2 in the main text) For all $n \in \mathbb{N}$, there exists a constructive and efficient way of enumerating properly sparsified Walsh–Hadamard matrices $\{W_j\}_{j=1}^{|\mathcal{S}_n|/2^n}$ $(W_j \in \mathcal{W}_n)$ such that*

$$\boldsymbol{A}_n = \left[W_1 \cdots W_{|\mathcal{S}_n|/2^n}\right]. \tag{21}$$

Proof. Recall that, any set of n-qubit stabilizer generators $\{g_i\}_{i=1}^n$ can be related to other 2^n-1 stabilizer states by considering a state corresponding to $\{(-1)^{\chi_i} g_i\}_{i=1}^n$. This implies that a single check matrix corresponds to 2^n stabilizer states; thus, a check matrix yields a sparsified Walsh–Hadamard matrix.

Next, we show that there is a constructive and efficient way to enumerate all the check matrices using a standard form that allows the unique description of a check matrix. While some standard forms suffice for such a purpose, here we employ the following standard form:

$$\left(\begin{array}{cc|cc} I_k & X_1 & Z_1 & O \\ O & O & X_1^\top & I_{n-k} \end{array}\right), \tag{22}$$

where $X_1 \in \mathbb{F}_2^{k\times(n-k)}$ is given by reduced row echelon form of rank k and $Z_1 = Z_1^\top \in \mathbb{F}_2^{k\times k}$ is a symmetric matrix. Since all the choices of X_1 and Z_1 give rise to mutually different check matrices that satisfy both Lemmas 7 and 8, we can assure that Eq. (22) yields a unique and complete construction of the entire set of check matrix. □

Two remarks are in order. First, we can check the validity of the above construction by computing the total number of check matrices. The number of choices for Z_1 is $2^{k(k+1)/2}$ while the number of choices for X_1 is given by the q-binomial coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_2$, which is defined for general $q(\neq 1)$ as

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(1-q^n)(1-q^{n-1})\dots(1-q^{n-k+1})}{(1-q)(1-q^2)\dots(1-q^k)}. \tag{23}$$

Therefore, by using $\sum_{k=0}^n \begin{bmatrix} n \\ k \end{bmatrix}_2 2^{k(k+1)/2} = |\mathcal{S}_n|/2^n$ which can be derived from the q-binomial theorem [40], we can certify the validity. Second, in the actual numerical implementation, we have utilized the gray code so that operations on matrix representations and group elements can be done efficiently. See the codes available via GitHub [30] for details.

E Proof of Proposition 1

In this section, we provide the proof for Proposition 1 regarding the existence of the cover stabilizer set.

Proposition 3. *(Restatement of Proposition 1 in the main text) Let S_n be a subset of $\mathcal{S}_n$ such that, for any $P \in \mathcal{P}_n$ there exists $|\psi\rangle \in S_n$ that satisfies $\{P, -P\} \cap \mathrm{Stab}(|\psi\rangle) \neq \emptyset$. Then, for all $n \in \mathbb{N}$, the size of the set is bounded as $|S_n| \geq 2^n + 1$, and one can construct S_n such that $|S_n| = 2^n + 1$.*

E.1 Main proof

E.1.1 Proof on size bound

First we show that $|S_n| \geq 2^n + 1$. Note that $I^{\otimes n} \in \mathrm{Stab}(|\psi\rangle)$ holds for any $|\psi\rangle \in \mathcal{S}_n$, and therefore $\mathrm{Stab}(|\psi\rangle) \setminus \{I^{\otimes n}\}$ consists of $2^n - 1$ elements. Therefore, in order to cover the remaining $4^n - 1$ elements of $\{I, X, Y, Z\}^{\otimes n} \setminus \{I^{\otimes n}\}$, the number of stabilizers satisfies

$$|S_n| \geq (4^n - 1)/(2^n - 1) = 2^n + 1. \tag{24}$$

E.1.2 Proof on the existence of minimum cover stabilizer set

Next, we prove the latter half of Proposition 3 that, there exists a cover stabilizer set S_n such that $|S_n| = 2^n+1$. In the following, we explicitly construct S_n with $|S_n| = 2^n+1$, and show that S_n satisfies the desired properties. Note that the overall sign of the stabilizer generators in the following argument is not relevant, and hence not discussed explicitly.

Let us take $|+\rangle^{\otimes n}$ as the 0-th element for S_n. For the k-th element ($1 \leq k \leq 2^n$), we take a state whose check matrix $\boldsymbol{C}_k = [\boldsymbol{X}_k \ \boldsymbol{Z}_k]$ is given as follows:

1. $\boldsymbol{Z}_k$ is an $n \times n$ identity matrix.
2. $\boldsymbol{X}_k$ is given such that Lemma 10 is satisfied.

Lemma 10. *There exists an explicit construction for a set of symmetric matrices* $\{\boldsymbol{X}_k \mid \boldsymbol{X}_k \in \mathbb{F}_2^{n\times n}\}_{k=1}^{2^n}$ *such that the following is satisfied.*

$$\forall \boldsymbol{v} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\},\ \{\boldsymbol{X}_k\boldsymbol{v}\}_{k=1}^{2^n} = \mathbb{F}_2^n. \tag{25}$$

From the fact that $\boldsymbol{X}_k$ are all symmetric matrices, it follows that the condition in Lemma 8 is satisfied, and therefore $\boldsymbol{C}_k = [\boldsymbol{X}_k \ \boldsymbol{Z}_k]$ indeed yields a valid representation of some pure stabilizer state.

Let us show that the constructed S_n indeed satisfies the condition in Proposition 3. Since it is obvious that $P \in \{I, X\}^{\otimes n}$ is covered by $|+\rangle^{\otimes n}$, we focus on other Pauli operators that are denoted as $P = (-1)^\chi X^{\boldsymbol{\alpha}} Z^{\boldsymbol{\beta}}$ with $\beta \neq 0$.

Consider some state $|\psi_k\rangle$ whose check matrix is given by $\boldsymbol{X}_k$ and $\boldsymbol{Z}_k$. For any $\boldsymbol{f} \in \mathbb{F}_2^n$, we can take $P_{k,\boldsymbol{f}} \in \{I, X, Y, Z\}^{\otimes n}$ so that its binary symplectic form yields $P_{k,\boldsymbol{f}} \approx X^{\boldsymbol{X}_k\boldsymbol{f}} Z^{\boldsymbol{Z}_k\boldsymbol{f}}$, and thus $\{P_{k,\boldsymbol{f}}, -P_{k,\boldsymbol{f}}\} \cap \mathrm{Stab}(|\psi_k\rangle) \neq \emptyset$.

Now let us take $\boldsymbol{f} = \boldsymbol{\beta}$. Due to Lemma 10, we can take k such that $\boldsymbol{X}_k\boldsymbol{\beta} = \boldsymbol{\alpha}$. This implies that the X and Z exponents of binary symplectic form of $P_{k,\boldsymbol{f}}$ can be given as $\boldsymbol{X}_k\boldsymbol{f} = \boldsymbol{\alpha}$ and $\boldsymbol{Z}_k\boldsymbol{f} = \boldsymbol{\beta}$, respectively. This implies $P_{k,\boldsymbol{f}} = P$ so that $\{P, -P\} \cap \mathrm{Stab}(|\psi_k\rangle) \neq \emptyset$, and therefore satisfies the condition of Proposition 3.

E.2 Proof of Technical Lemma 10

Now the remaining work is to prove Lemma 10. We first provide the explicit construction of $\{\boldsymbol{X}_k\}_{k=1}^{2^n}$, and then prove that it indeed satisfies Eq. (25).

E.2.1 Construction of $\boldsymbol{X}_k$

We first introduce some algebraic concepts necessary for the discussion. We denote the polynomial ring over $\mathbb{F}_2$ by $\mathbb{F}_2[x]$. Let f be an arbitrary irreducible polynomial of degree n. We consider a quotient ring $\mathbb{F}_2[x]/(f)$, where (f) denotes the ideal generated by f. Then, $\mathbb{F}_2[x]/(f)$ is a field because f is irreducible. It is also noteworthy that $\mathbb{F}_2[x]/(f)$ is a vector space over $\mathbb{F}_2$ and $\{x^0, \ldots, x^{n-1}\}$ can be taken as a basis.

In what follows, we discuss the construction of $\{\boldsymbol{X}_k\}_{k=1}^{2^n}$. We define a symmetric matrix $C(x) \in (\mathbb{F}_2[x]/(f))^{n\times n}$ as a matrix whose (i,j) entry equals x^{i+j-2}. Namely, $C(x)$ can be represented as follows:

$$C(x) = \begin{pmatrix} x^0 & x^1 & x^2 & \cdots & x^{n-1} \\ x^1 & x^2 & & & \\ x^2 & & \ddots & & \vdots \\ \vdots & & & & \\ x^{n-1} & & \cdots & & x^{2n-2} \end{pmatrix}.$$

Every entry of $C(x)$ can be represented as a linear combination of a basis $\{x^0, \dots, x^{n-1}\}$, and we define C_i be a matrix consisting of such x^i coefficients. In other words, $C_i \in \mathbb{F}_2^{n\times n}$ is defined so that $C(x) = C_0x^0 + \cdots + C_{n-1}x^{n-1}$ holds. Note that C_i is also symmetric.

We give a concrete example below. We take $n = 3$ and $f = 1 + x + x^3$, which is irreducible. Then, C_0, C_1, C_2 can be derived in the following way:

$$\begin{aligned} C(x) &= \begin{pmatrix} x^0 & x^1 & x^2 \\ x^1 & x^2 & x^3 \\ x^2 & x^3 & x^4 \end{pmatrix} \\ &= \begin{pmatrix} 1 & x & x^2 \\ x & x^2 & 1+x \\ x^2 & 1+x & x+x^2 \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} 1 & & \\ & & 1 \\ & 1 & \end{pmatrix}}_{C_0} x^0 + \underbrace{\begin{pmatrix} & 1 & \\ 1 & & 1 \\ & 1 & 1 \end{pmatrix}}_{C_1} x^1 + \underbrace{\begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & 1 \end{pmatrix}}_{C_2} x^2. \end{aligned}$$

Next, we consider the following set of symmetric matrices:

$$\left\{ \sum_{i=0}^{n-1} a_i C_i \;\middle|\; a_i \in \mathbb{F}_2 \right\}.$$

Since the elements of the set are distinct, the set has 2^n elements. We take this set as the set $\{\boldsymbol{X}_k\}_{k=1}^{2^n}$.

E.2.2 Proof that $\boldsymbol{X}_k$ satisfies Lemma 10

Next, we prove that $\{\boldsymbol{X}_k\}_{k=1}^{2^n}$ given in the previous subsection satisfies Eq. (25).

By the definition of $\boldsymbol{X}_k$, one can show that Eq. (25) holds if and only if $\{C_i\boldsymbol{v}\}_{i=0}^{n-1}$ is linearly independent for any $\boldsymbol{v} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}$. From here, we show the linear independence of $\{C_i\boldsymbol{v}\}_{i=0}^{n-1}$ by proving several lemmas.

Lemma 11. *For any vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}$, $\boldsymbol{u}^\top C(x)\boldsymbol{v} \neq 0$.*

Proof. Using a vector $\boldsymbol{x} = (x^0, \dots, x^{n-1})^\top$, we have $C(x) = \boldsymbol{x}\boldsymbol{x}^\top$. Thus, by defining two polynomials $u(x) = \boldsymbol{u}^\top\boldsymbol{x} = \sum_{i=0}^{n-1} u_i x^i$ and $v(x) = \boldsymbol{v}^\top\boldsymbol{x} = \sum_{i=0}^{n-1} v_i x^i$, $\boldsymbol{u}^\top C(x)\boldsymbol{v}$ can be represented as $u(x)v(x)$. Hence, it suffices to show that $u(x)v(x) \neq 0$. Because $\boldsymbol{u}$ and $\boldsymbol{v}$ are nonzero, $u(x)$ and $v(x)$ are nonzero as well. Noting that $\mathbb{F}_2/(f)$ is a field, the product $u(x)v(x)$ is also nonzero. □

Lemma 12. *For any vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}$, there exists i such that $\boldsymbol{u}^\top C_i\boldsymbol{v} \neq 0$.*

Proof. By multiplying $C(x) = C_0x^0 + \cdots + C_{n-1}x^{n-1}$ by $\boldsymbol{u}$ from the left and $\boldsymbol{v}$ from the right, we obtain

$$\boldsymbol{u}^\top C(x)\boldsymbol{v} = (\boldsymbol{u}^\top C_0\boldsymbol{v})x^0 + \cdots + (\boldsymbol{u}^\top C_{n-1}\boldsymbol{v})x^{n-1}. \tag{26}$$

Eq. (26) expresses $\boldsymbol{u}^\top C(x)\boldsymbol{v}$ as a polynomial with coefficients $\boldsymbol{u}^\top C_i\boldsymbol{v} \in \mathbb{F}_2$. Since $\boldsymbol{u}^\top C(x)\boldsymbol{v}$ is nonzero from Lemma 11, there exists a nonzero coefficient, i.e, $\boldsymbol{u}^\top C_i\boldsymbol{v} \neq 0$ for some i. □

Lemma 13. *For any vector $\boldsymbol{v} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}$, $\{C_i\boldsymbol{v}\}_{i=0}^{n-1}$ is linearly independent.*

Proof. We consider a matrix $[C_0\boldsymbol{v}, \dots, C_{n-1}\boldsymbol{v}]$ with $C_i\boldsymbol{v}$ as the column vectors. By multiplying an arbitrary nonzero vector $\boldsymbol{u} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}$ from the left, we obtain a vector $(\boldsymbol{u}^\top C_0\boldsymbol{v}, \dots, \boldsymbol{u}^\top C_{n-1}\boldsymbol{v})$, which is nonzero by Lemma 12. Therefore, we can confirm that the matrix $[C_0\boldsymbol{v}, \dots, C_{n-1}\boldsymbol{v}]$ is non-singular, which implies the linear independence of $\{C_i\boldsymbol{v}\}_{i=0}^{n-1}$. □

Having shown Lemma 13, it is also proved that $\{\boldsymbol{X}_k\}_{k=1}^{2^n}$ given in the previous subsection satisfies Eq. (25), i.e., Lemma 10 is proved.

F Proof of Lemma 3

In this section, we prove Lemma 3 in the main text, which provides the accuracy bound on the minimal feasible solution for RoM calculation. Let us first recall that the Pauli vector $\boldsymbol{b}$ of length-4^n is decomposed in correspondence with the minimal cover matrix as $\{\boldsymbol{b}_i\}_{i=1}^{2^n+1}$ where each $\boldsymbol{b}_i$ is a length-2^n vector. The Lemma 3 is based on the observation that $\boldsymbol{b}_i$ seems to obey normal distribution for random states. This motivates us to show the random average of the approximate value R_{FWHT} as follows.

Lemma 14. *(Restatement of Lemma 3) Let $\boldsymbol{b}_i$ drawn uniformly from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ for all j. Then,*

$$\mathbb{E}_{\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{I})} \left[\frac{R_{\mathrm{FWHT}}}{\|\rho\|_{\mathrm{st}}} \right] \approx 2^{n/2}. \tag{27}$$

Proof. Recall that the FWHT algorithm for i-th segment $\boldsymbol{b}_i$ yields the segment of minimal feasible solution $\boldsymbol{x}_i$, from which the approximate RoM value is obtained as

$$R_{\mathrm{FWHT}} = \sum_{i=1}^{2^n+1} \|\boldsymbol{x}_i\|_1 = \frac{1}{2^n} \sum_{i=1}^{2^n+1} \|H_n \boldsymbol{b}_i\|_1. \tag{28}$$

By combining with the definition of the st-norm $\|\rho\|_{\mathrm{st}} = \frac{1}{2^n}\|\boldsymbol{b}\|_1 = \frac{1}{2^n}\sum_{i=1}^{2^n+1}\|\boldsymbol{b}_i\|_1$, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{I})} \left[\frac{R_{\mathrm{FWHT}}}{\|\rho\|_{\mathrm{st}}} \right] &= \mathbb{E}_{\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{I})} \left[\frac{\sum_{i=1}^{2^n+1} \|H_n \boldsymbol{b}_i\|_1}{\sum_{i=1}^{2^n+1} \|\boldsymbol{b}_i\|_1} \right] \\
&= \sum_{i=1}^{2^n+1} 2^n \mathbb{E}_{\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{I})} \left[\frac{\left|\boldsymbol{1}^\top \boldsymbol{b}_i\right|}{\sum_{i=1}^{2^n+1} \|\boldsymbol{b}_i\|_1} \right] \\
&= 2^n \mathbb{E}_{b_{ij} \sim \mathcal{N}(0,1)} \left[\frac{\sum_{i=1}^{2^n+1} \left|\sum_{j=1}^{2^n} b_{ij}\right|}{\sum_{i=1}^{2^n+1} \sum_{j=1}^{2^n} |b_{ij}|} \right],
\end{aligned} \tag{29}$$

where, to derive the second equation, we have used the fact the matrix elements of the Walsh–Hadamard matrix H_n only consist of ± 1 and that the elements of $\boldsymbol{b}_i$ are distributed symmetrically with respect to sign change. We have also denoted the elements of the vector as $\boldsymbol{b}_i = (\dots, b_{ij}, \dots)$.

In order to evaluate Eq. (29), we utilize the second order formula for random variable as [41]

$$\mathbb{E}\left[\frac{X}{Y}\right] \approx \frac{\mathbb{E}[X]}{\mathbb{E}[Y]} - \frac{\mathrm{Cov}(X,Y)}{\mathbb{E}[Y]^2} + \frac{\mathrm{Var}(Y)\mathbb{E}[X]}{\mathbb{E}[Y]^3}. \tag{30}$$

It can be shown that $X = \sum_{i=1}^{2^n+1} \left| \sum_{j=1}^{2^n} b_{ij} \right|$, $Y = \sum_{i=1}^{2^n+1} \sum_{j=1}^{2^n} |b_{ij}|$ can be evaluated as

$$\frac{\mathbb{E}[X]}{\mathbb{E}[Y]} = \frac{1}{2^{n/2}}, \quad \frac{\mathrm{Cov}(X,Y)}{\mathbb{E}[Y]^2} = o\left(\frac{1}{2^n}\right), \quad \frac{\mathrm{Var}(Y)\mathbb{E}[X]}{\mathbb{E}[Y]^3} = o\left(\frac{1}{2^n}\right). \tag{31}$$

Therefore, by substituting this into Eq. (30), we finally obtain

$$2^n \mathbb{E}\left[\frac{X}{Y}\right] \approx 2^{n/2} + o(1),$$

which completes the proof. □

While the Lemma holds for asymptotically large n, we numerically find that the convergence is observed for moderate n. Namely, we have numerically computed the values for 100 random 10-qubit mixed states to show that

$$\left(\frac{R_{\mathrm{FWHT}}}{\|\rho\|_{\mathrm{st}}}\right)^{\frac{1}{n}} = 1.41412 \pm 0.00006. \tag{32}$$

This implies that we can readily expect the accuracy bound to hold in moderate-size systems.

G Proof of Proposition 2

In this section, we prove Proposition 2 which assures that when one restricts the pure stabilizer sets so that the target problem is described by tensor product as $\bigotimes_i \boldsymbol{A}_{n_i} \subset \boldsymbol{A}_n$, the solution is simply a tensor product of individual solution. The proof can be similarly done by following arguments provided in Ref. [42].

Let us consider a general L^1 norm minimization problem with a tensor product structure over M subsystems as

$$\begin{aligned} \text{(P)} \ \underset{\boldsymbol{x}}{\text{minimize}} \quad & \|\boldsymbol{x}\|_1 \\ \text{subject to} \quad & \left(\bigotimes_{i=1}^{M} \boldsymbol{A}_i\right)\boldsymbol{x} = \bigotimes_{i=1}^{M} \boldsymbol{b}_i, \end{aligned}$$

where the i-th subsystem can be formulated as

$$\begin{aligned} (\mathrm{P}_i) \ \underset{\boldsymbol{x}_i}{\text{minimize}} \quad & \|\boldsymbol{x}_i\|_1 \\ \text{subject to} \quad & \boldsymbol{A}_i\boldsymbol{x}_i = \boldsymbol{b}_i. \end{aligned}$$

Assuming that optimal solution exists for every (P_i), the following holds.

Lemma 15. *Let $\boldsymbol{x}_i^*$ be the optimal solution for (P_i). Then, $\boldsymbol{x}^* = \bigotimes_{i=1}^{M} \boldsymbol{x}_i^*$ is the optimal solution for (P).*

Proof. It is obvious that $\boldsymbol{x}^*$ is a feasible solution, and therefore we show the optimality of $\boldsymbol{x}^*$ via duality. The dual problem for the total system is given as

$$\begin{aligned} \text{(D)} \ \underset{\boldsymbol{y}}{\text{maximize}} \quad & \left(\bigotimes_{i=1}^{M} \boldsymbol{b}_i^\top\right)\boldsymbol{y} \\ \text{subject to} \quad & \left\| \left(\bigotimes_{i=1}^{M} \boldsymbol{A}_i^\top\right)\boldsymbol{y} \right\|_\infty \leq 1. \end{aligned}$$

where the dual problem for each subsystem can also be provided as

$$(\mathrm{D}_i)\ \underset{\boldsymbol{y}_i}{\text{maximize}} \quad \boldsymbol{b}_i^\top \boldsymbol{y}_i \qquad \text{subject to} \quad \left\| \boldsymbol{A}_i^\top \boldsymbol{y}_i \right\|_\infty \leq 1.$$

Now the strong duality of the problem assures that optimal solution $\boldsymbol{y}_i^*$ exists for any (D_i) with the optimal value given as $\|\boldsymbol{x}_i^*\|_1 = \boldsymbol{b}_i^\top \boldsymbol{y}_i^*$. By taking $\boldsymbol{y}^* := \bigotimes_{i=1}^M \boldsymbol{y}_i^*$, then it follows from the property of L^∞ norm that

$$\left\| \left(\bigotimes_{i=1}^M \boldsymbol{A}_i^\top \right) \boldsymbol{y}^* \right\|_\infty = \left\| \bigotimes_{i=1}^M \left(\boldsymbol{A}_i^\top \boldsymbol{y}_i^* \right) \right\|_\infty = \prod_{i=1}^M \left\| \boldsymbol{A}_i^\top \boldsymbol{y}_i^* \right\|_\infty \leq 1,$$

which guarantees that $\boldsymbol{y}^*$ is a feasible solution of (D). Now the objective function for the dual problem (D) satisfies

$$\left(\bigotimes_{i=1}^M \boldsymbol{b}_i^\top \right) \boldsymbol{y}^* = \prod_{i=1}^M \left(\boldsymbol{b}_i^\top \boldsymbol{y}_i^* \right) = \prod_{i=1}^M \|\boldsymbol{x}_i^*\|_1 = \left\| \bigotimes_{i=1}^M \boldsymbol{x}_i^* \right\|_1 = \|\boldsymbol{x}^*\|_1. \tag{33}$$

Therefore, the objective functions of feasible solutions $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ for the primal and dual problems coincide with each other and hence are optimal due to the strong duality. □

Given Lemma 15, Proposition 2 in the main text follows directly by applying to the calculation of RoM.

H Numerical details on RoM calculation

In this section, we provide details on the numerical results on RoM calculation.

H.1 More results on top-overlap method

Here, we provide the results on numerical demonstration regarding the top-overlap method introduced in Sec. 3.2 in the main text. Concretely, we present results for random mixed, pure, tensor product states of $n = 4, 5, 6, 7$ qubit system.

As is shown in Fig. 10, we can see that the top-overlap method significantly outperforms naive random selection methods in all cases. We can see that the improvement becomes more evident when we simulate larger systems; in particular, for $n = 7$ qubit case it suffices to take only $K = 10^{-5}$ to reach near-optimal value for any target.

Two remarks are in order. First, we note that we have added the column set of cover matrix in the case of tensor product states, in order to assure the existence of a feasible solution. In other words, the restriction of the column set solely using the information of overlaps may lead to rank deficient matrix. Second, the run time of our algorithm for the $n = 6$ qubit system was approximately 5 seconds for computing all the overlaps and 3 minutes for solving the LP. For the case of the $n = 7$ qubit system, the overlap computation consumes 15 minutes at most and 15 minutes for solving the LP. The simulation was done on a single laptop, and we envision that the use of MPI or GPU shall further speed up the computation.

H.2 Assuring feasibility

As mentioned in the previous subsection, we cannot obtain any feasible solution if the set of stabilizer states are not appropriately restricted.

One of the most robust ways to assure feasibility for arbitrary quantum states is to utilize an efficiently computed approximate solution with relatively low precision. Formally, we can understand this as modifying the problem as

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \|\boldsymbol{x}\|_1 + c\|\boldsymbol{e}\|_1 \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{e}, \end{aligned} \tag{34}$$

where $\boldsymbol{e}$ is introduced to absorb the numerical error due to rank deficiency, and c is a hyperparameter that determines the penalty due to such an error. In practice, one may add the cover matrix to the set of stabilizers for such a purpose.

The second method that is applicable in the case of tensor product states is to utilize the solution obtained from small scale systems. As shown in Lemma 15 in Appendix G, we can always obtain a feasible solution by taking tensor products and hence can be used as an initialization. Therefore, one may extend the set of stabilizers so that the quality of the solution is improved. Such a technique is also valid for the Column Generation method presented in Sec. 4.

H.3 Overlap computation using singular value decompomsition

Here, we discuss how to efficiently compute the overlap between the Pauli vector $\boldsymbol{b}$ of $(n+m)$-qubit system and a stabilizer state that can be decomposed into a tensor product of n and m-qubit stabilizer state.

First, let us consider singular value decomposition of the Pauli vector as $\boldsymbol{b} = \sum_{k=1}^{r} \sigma_k (u_k \otimes v_k)$ where r is the number of nonzero singular values, u_k and v_k is the k-th vector of n and m-qubit system with singular value of σ_k. Then, the overlap with all stabilizer states in $\mathcal{S}_n \otimes \mathcal{S}_m$ $(\subset \mathcal{S}_{n+m})$ can be computed efficiently by noting that

$$\begin{aligned} \boldsymbol{b}^\top (\boldsymbol{A}_n \otimes \boldsymbol{A}_m) &= \sum_{k=1}^{r} \sigma_k ((u_k^\top \boldsymbol{A}_n) \otimes (v_k^\top \boldsymbol{A}_m)) && (35) \\ &= \mathrm{vec}(U^\top V), && (36) \end{aligned}$$

where "vec" denotes the vectorization of a matrix, U and V are matrices whose k-th row are given as $\sqrt{\sigma_k} u_k^\top \boldsymbol{A}_n$ and $\sqrt{\sigma_k} v_k^\top \boldsymbol{A}_m$, respectively. From Eq. (35) to Eq. (36), we have utilized the tensor product structure as

$$\left(\sum_k \boldsymbol{\lambda}^{(k)} \otimes \boldsymbol{\xi}^{(k)} \right)_{(l,m)} = \sum_k \lambda_l^{(k)} \xi_m^{(k)} = \left(\Lambda^\top \Xi \right)_{l,m} = \mathrm{vec}(\Lambda^\top \Xi), \tag{37}$$

where $\Lambda = (\dots, \boldsymbol{\lambda}^{(k)}, \dots)$ and $\Xi = (\dots, \boldsymbol{\xi}^{(k)}, \dots)$.

The matrix-matrix product in Eq. (36) can be computed with time complexity of $\mathcal{O}(r(n|\mathcal{S}_n| + m|\mathcal{S}_m| + |\mathcal{S}_n||\mathcal{S}_m|))$. In particular, when the target state is completely decoupled into two subsystems as $r = 1$, then the maximal value of all the overlaps between $|\mathcal{S}_n||\mathcal{S}_m|$ states can be computed with time complexity of $\mathcal{O}(n|\mathcal{S}_n| + m|\mathcal{S}_m|)$.

Note that, the above technique can be utilized as an initialization technique. Namely, one may discard small singular values so that we can perform data compression as long as the truncated state well-reproduces the original target state. Given the compressed data,

we have used the Khatri-Rao product to regenerate the columns in the $n+m$-qubit system. In this sense, one may further extend the approximation algorithm to low-entangled states that can be represented efficiently by, e.g., tensor networks.

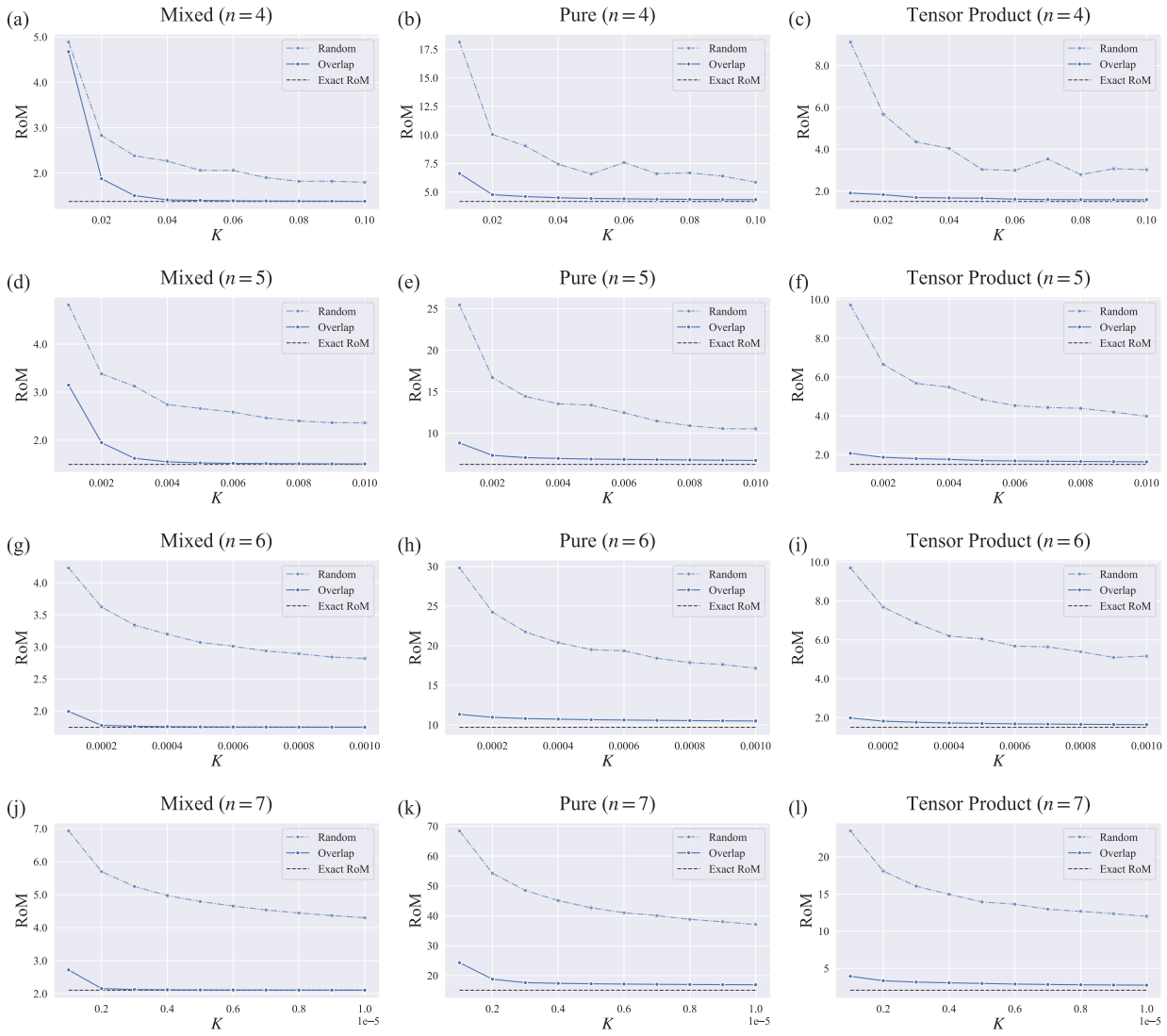


Figure 10: Numerical demonstration of top-overlap method introduced in Sec. 3.2 in the main text. The column set is restricted to $K|\mathcal{S}_n|$ $(0 < K \leq 1)$. The cyan and blue lines denote the approximate RoM values computed by restricting the column set of stabilizers at random and by taking the largest and smallest overlaps, respectively. The black dotted lines indicate the exact RoM values which is computed with Column Generation method.

TM2: Faster Computation of Stabilizer Extent

Hiroki Hamaguchi[1], Kou Hamada[1], and Nobuyuki Yoshioka[2,3,4]

[1]Graduate School of Information Science and Technology, University of Tokyo, Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[2]Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[3]Theoretical Quantum Physics Laboratory, RIKEN Cluster for Pioneering Research (CPR), Wako-shi, Saitama 351-0198, Japan
[4]JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

Characterization of nonstabilizerness is fruitful due to its application in gate synthesis and classical simulation. In particular, the resource monotone called the *stabilizer extent* is indispensable to estimate the simulation cost using the rank-based simulators, which are one of the state-of-the-art simulators of Clifford+T circuits. In this work, we propose fast numerical algorithms to compute the stabilizer extent. Our algorithm utilizes the column generation technique, which iteratively updates the subset of pure stabilizer states used for calculating the value of the monotone, based on the information of the state overlap between the target state and all stabilizer states. Upon updating the subset, we make use of a newly proposed subroutine for overlap calculation that (i) prunes unnecessary computation that does not contribute to stabilizer extent, (ii) reduces the time complexity of necessary stabilizer states exponentially per state, (iii) reduces the space complexity superexponentially by in-place calculation. As a result, we have demonstrated our algorithm for random pure states up to $n = 9$ qubits, which naively requires memory of 305 EiB. We also show that our algorithm runs even more efficient when the target state vector is real; the size of optimization problem is reduced exponentially so that we can even simulate the case of $n = 10$ qubits in 7.4 hours.

1 Introduction

In the domain of universal fault-tolerant quantum computation, elementary gates are often formulated to include both classically simulatable gates and more resource-intensive gates, as exemplified by the prominent Clifford+T formalism of the magic state model [1, 2, 3, 4, 5, 6]. Since Clifford circuits are classically simulatable [1], non-Clifford gates are essential for achieving quantum advantage [7, 8, 9, 10], and naturally it is crucial to improve and characterize classical simulation algorithms to quantitatively understand the computational speedups in quantum circuits [11, 12, 13, 14, 15, 16, 17].

When we address optimization problems involving the entire set of stabilizer states, such as those related to Robustness of Magic (RoM) or stabilizer extent, the task becomes

Hiroki Hamaguchi: hamaguchi-hiroki0510@g.ecc.u-tokyo.ac.jp
Kou Hamada: zkouaaa@g.ecc.u-tokyo.ac.jp
Nobuyuki Yoshioka: nyoshioka@ap.t.u-tokyo.ac.jp

Table 1: The size of $\mathcal{S}_n$, the data size of A_n in sparse matrix format [19], and the time to numerically compute the stabilizer extent for Haar random pure state by the naive method and our proposed method in Section 3, or the method in Section 4 for $n = 10$.

n	5	6	7	8	9	10
$\lvert\mathcal{S}_n\rvert$	2.42e+06	3.15e+08	8.13e+10	4.18e+13	4.29e+16	8.79e+19
size of A_n	1011 MiB	254 GiB	153 TiB	153 PiB	305 EiB	1 YiB
naive	7.7 min	×	×	×	×	×
proposed	1.5 s	3.8 s	12.9 s	8.8 min	19.2 h	(Real Case)

exceedingly difficult due to the superexponential number of states involved. In the case of RoM, it was shown by authors [18] that we can dramatically push the simulatable system size by combining the column generation method and fast overlap calculations. Naturally, this raises the question of whether the computation of stabilizer extent, which is utilized as the state-of-the-art in classical simulators, can also be accelerated. However, this is not straightforward for the following reasons: (i) the Fast Walsh-Hadamard Transform cannot be used for inner product calculations, and (ii) the problem class is not given by Linear Program (LP) but by a more challenging class of Second-Order Cone Program (SOCP), reflecting the fact that variables are complex instead of real. Due to such a complication, it has remained unclear whether stabilizer extent can be computed faster for larger qubit counts.

In this work, we show that the computation of stabilizer extent can actually be accelerated even further compared to the one for RoM. We find a novel canonical form of pure stabilizer states that allows us to both enumerate them efficiently and perform fast overlap computation that can be done exponentially faster per stabilizer state. We further find that, when we search for the subset of stabilizer states during the Column Generation (CG) method, we can prune the computation based on the bounds of overlap values, resulting in skipping nearly half of the overlap calculation. We numerically demonstrate that our proposed algorithm allows us to compute the stabilizer extent of random pure state up to $n = 9$ qubits, which naively requires memory of 305 EiB. Furthermore, with the scope of application to entanglement resource states such as GHZ or W states and eigenstates of physically important Hamiltonians, we show that real-coefficient state vectors can be computed with even less computational cost. Concretely, the size of relevant stabilizer states is reduced exponentially, so that we can compute the stabilizer extent of random state up to $n = 10$ qubits.

The remainder of this paper is organized as follows. In Section 2, we present the preliminaries on the formalism of stabilizer extent. In Section 3, we first introduce how to calculate the overlap with all stabilizer states efficiently in Theorem 3, which serves as the main subroutine for our algorithm. Then, we describe the algorithm that computes stabilizer extent up to 9-qubit states with reduced computational resources by utilizing these overlap values. In Section 4, we demonstrate that by specializing the algorithm for states with certain properties, we can further compute the stabilizer extent for 10-qubit states. Finally, in Section 5, we discuss our findings and provide future perspectives on our work.

2 Preliminaries

Let $\mathcal{S}_n := \{|\phi_j\rangle\}$ be the entire set of n-qubit stabilizer states. We also define the density matrix for $|\phi_j\rangle$ as $\sigma_j := |\phi_j\rangle\langle\phi_j|$. The size of $\mathcal{S}_n$ scales superexponentially as $|\mathcal{S}_n| = 2^n \prod_{k=0}^{n-1}(2^{n-k}+1) = 2^{\mathcal{O}(n^2)}$ [20, Proposition 1]. See also Table 1 for the size of $\mathcal{S}_n$.

The *Robustness of Magic* (RoM) is introduced in [21] to quantify an n-qubit state ρ, represented by a density matrix. RoM is defined as follows:

$$\mathcal{R}(\rho) := \min_{c\in\mathbb{R}^{|\mathcal{S}_n|}} \left\{ \|c\|_1 \;\middle|\; \rho = \sum_{j=1}^{|\mathcal{S}_n|} c_j \sigma_j \right\}.$$

The *stabilizer extent* is introduced in [22, Definition 3] to quantify an n-qubit state ψ, represented by a state vector. Stabilizer extent is defined as follows:

$$\xi(\psi) := \min_{c\in\mathbb{C}^{|\mathcal{S}_n|}} \left\{ \|c\|_1^2 \;\middle|\; |\psi\rangle = \sum_{j=1}^{|\mathcal{S}_n|} c_j |\phi_j\rangle \right\}. \tag{1}$$

In this paper, we focus on numerical computation of stabilizer extent. This definition (1) can be simplified as the complex L^1-norm minimization problem in the following.

$$\sqrt{\xi(\psi)} = \min_{x\in\mathbb{C}^{|\mathcal{S}_n|}} \{\|x\|_1 \mid A_n x = b\} \tag{2}$$

Here, we define $A_n \in \mathbb{C}^{2^n\times|\mathcal{S}_n|}$ as $(A_n)_{ij} := \langle i|\phi_j\rangle$ and $b \in \mathbb{C}^{2^n}$ as $b_i := \langle i|\psi\rangle$ using the computational basis $\{|i\rangle\}_{i=0}^{2^n-1}$. In reference to [23], problem (2) is a Second-Order Cone Program (SOCP). Thus, by defining $\mathcal{A}_n$ as the columns set $\{a_j\}$ of A_n, its dual problem can be derived as [23, Appendix A][24, Section 5.1.6]

$$\sqrt{\xi(\psi)} = \max_{y\in\mathbb{C}^{2^n}} \left\{ \mathrm{Re}(b^\dagger y) \;\middle|\; \left|a_j^\dagger y\right| \le 1 \text{ for all } a_j \in \mathcal{A}_n \right\} \tag{3}$$

where $\dagger$ denotes the conjugate transpose.

Further, we introduce a function SolveSOCP$(\mathcal{C}, b)$ to describe our algorithm in later sections. The $\mathcal{C} \subseteq \mathcal{A}_n$ represents a column subset, and this function solves problem (3) restricting $\mathcal{A}_n$ to $\mathcal{C}$, and returns the solution x for the corresponding restricted primal problem of (2), as well as the solution y for the restricted dual problem of (3). In actual implementation, this function can be realized by just solving the corresponding primal problem (2) with SOCP solver, such as MOSEK [25] or CVXPY [26, 27].

3 Scaling Up the Exact Stabilizer Extent Calculation

In the preceding sections, we introduced two similar quantum resource measures: Robustness of Magic and stabilizer extent. Despite both being efficiently quantifiable through convex optimization problems, solving them naively for $n > 5$ qubit systems becomes impractical due to the superexponential growth of the number of stabilizer states $|\mathcal{S}_n|$ as shown in Table 1. Moreover, the solver requires at least twice the memory size of A_n. To address this challenge, in [18] we proposed employing a classical optimization technique known as the Column Generation (CG) method [28] for RoM calculation. However, it remained unclear whether the same approach could be applied to stabilizer extent, since the structure of the matrix we use for the calculation, A_n, is largely different, and SOCP

is more general and difficult than Linear Program (LP) [24, Section 4.4.2], which is used in RoM calculation. Here, we demonstrate that leveraging the specific structure of stabilizer states enables us to use the branch and bound method for the size reduction of A_n and a similar method to work effectively for calculating stabilizer extent as well.

3.1 Core Subroutine: Calculating Overlap

Before considering the stabilizer extent, we define *stabilizer fidelity* of a pure quantum state $|\psi\rangle$ using its state vector $b \in \mathbb{C}^{2^n}$ as

$$\sqrt{F(|\psi\rangle)} \coloneqq \max_{\phi \in \mathcal{S}_n} |\langle\phi|\psi\rangle| = \max_{a_j \in \mathcal{A}_n} \left|a_j^\dagger b\right|,$$

which is the maximal overlap between the target state and the stabilizer states. The importance of the stabilizer fidelity is highlighted by its role in computing the RoM [18] and its direct relationship with stabilizer extent, as demonstrated in [22, Definition 4][23]. In later sections, we will show that the stabilizer fidelity is crucial in our proposed approach. Here, we demonstrate how to efficiently compute the stabilizer fidelity up to 9-qubit systems.

To this end, we introduce the following theorem which is useful for enumerating all the stabilizer states. This theorem is a variant of previous works [29, Theorem 2], [30, Section 5], [31, Theorem 5.(ii)]. The proof is given in Appendix A.1.

Theorem 1. *For all $k \in \{1, \dots, n\}$, define the following set:*

$$\begin{aligned}
\mathcal{Q}_k &\coloneqq \left\{ Q \;\middle|\; Q \in \mathbb{F}_2^{k\times k} \text{ is a upper triangular matrix} \right\}, \\
\mathcal{R}_k &\coloneqq \left\{ R \;\middle|\; R \in \mathbb{F}_2^{n\times k} \text{ is a reduced row echelon form matrix with } \operatorname{rank}(R) = k \right\}, \\
\mathcal{T}_k(R) &\coloneqq \left\{ t \;\middle|\; t \in \mathbb{F}_2^n \text{ is a representative of element in the quotient space } \mathbb{F}_2^n / \operatorname{Im}(R) \right\}.
\end{aligned}$$

Define the set of states $\mathcal{S}_{n,k}$ as

$$\mathcal{S}_{n,k} \coloneqq \left\{ \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x} i^{c^\top x} |Rx + t\rangle \;\middle|\; Q \in \mathcal{Q}_k, c \in \mathbb{F}_2^n, R \in \mathcal{R}_k, t \in \mathcal{T}_k(R) \right\}, \tag{4}$$

and define $\mathcal{S}_{n,0} \coloneqq \{|t\rangle \mid t \in \mathbb{F}_2^n\}$. Then, we have $\bigcup_{k=0}^n \mathcal{S}_{n,k} = \mathcal{S}_n$.

Let $|\phi\rangle$ be one of the stabilizer states with $k > 0$ in Theorem 1, which means $|\phi\rangle = \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x} i^{c^\top x} |Rx + t\rangle$. Then, by denoting $a \in \mathcal{A}_n$ as the corresponding state vector of $|\phi\rangle$, the overlap between states $|\psi\rangle$ and $|\phi\rangle$ is given by a and b as

$$\left|a^\dagger b\right| = |\langle\phi|\psi\rangle| = \left| \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} \overline{(-1)^{x^\top Q x} i^{c^\top x}} \langle Rx + t|\psi\rangle \right| = \left| \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x} i^{c^\top x} \left(\frac{1}{2^{k/2}} b_{Rx+t}^\dagger \right) \right|.$$

In the following, we define $P_x \coloneqq \frac{1}{2^{k/2}} b_x^\dagger$, and for the simplicity, we fix $k = n, R = I_n, t = 0$. This assumption is not restrictive since the other cases can be easily reduced to this case. Recall that what we want is $\max_{a_j \in \mathcal{A}_n} \left|a_j^\dagger b\right|$. Owing to the equation above, this is basically equivalent to the following problem:

$$\max_{Q,c} \left\{ \left| \sum_{x=0}^{2^n-1} (-1)^{x^\top Q x} i^{c^\top x} P_x \right| \right\}. \tag{5}$$

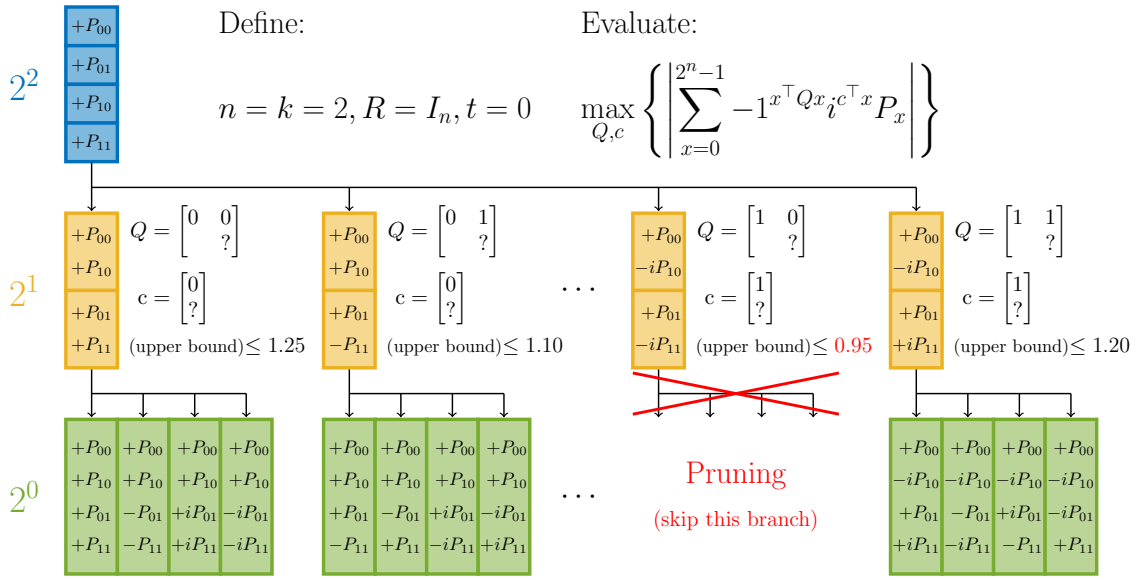


Figure 1: Visualization of stabilizer pruning for Theorem 2. Each cell stores the evaluated value of the expression, and the $2^{n+n(n+1)/2}$ leaf nodes correspond to the value $\sum_{x=0}^{2^n-1}(-1)^{x^\top Qx} i^{c^\top x} P_x$. Since we only need one set of cells per color during the procedure, we can do it in-place. This means the space complexity is $\mathcal{O}(\sum_{i=0}^{n} 2^i)$, i.e., $\mathcal{O}(2^n)$.

If we solve (5) naively, the time complexity is $\mathcal{O}\left(2^{n+n(n+1)/2} 2^n n^2\right)$, where $2^{n+n(n+1)/2}$ is the number of combinations for (Q, c), 2^n is the number of the terms in the summation, and n^2 is the computational cost per each term. However, we can compute this much more efficiently.

Theorem 2. *Problem (5) can be solved in $\mathcal{O}\left(2^{n+n(n+1)/2}\right)$ time complexity and $\mathcal{O}(2^n)$ space complexity.*

We refer to the algorithm used in Theorem 2 as *stabilizer pruning* in this paper. Details are provided in Figure 1 and Appendix A.2. Stabilizer pruning is based on the branch and bound method and is a recursive, in-place procedure. Since we are solving a maximization problem, solutions inferior to the current best solution (or to 1 for the case in Section 3.2.2) are unnecessary. Thus, we can terminate branches if the upper bound of the current branch is lower than these values. For more details on the pruning strategy, see Appendix A.3. By applying a similar argument for every k, R, and t, we can derive the next theorem as a consequence.

Theorem 3. *Stabilizer fidelity of a n-qubit state $|\psi\rangle$ can be computed in time complexity of $\mathcal{O}(|S_n|)$ and space complexity of $\mathcal{O}(2^n)$.*

Let us briefly discuss the numerical results of the stabilizer fidelity calculation. The time to compute stabilizer fidelity of a Haar random 8-qubit state was 5 seconds, and that of a 9-qubit state was 26 minutes, thanks to the pruning by the branch and bound method. We will use a slightly modified version of this algorithm as a subroutine to compute the stabilizer extent; namely, besides finding the maximum, we also identify other large overlaps in the CG method. All numerical experiments in this paper are conducted using C++17 compiled by GCC 9.4.0 and a cluster computer powered by Intel(R) Xeon(R) CPU E52640 v4 with 270 GB of RAM using 40 threads, and all the codes are available at GitHub [32].

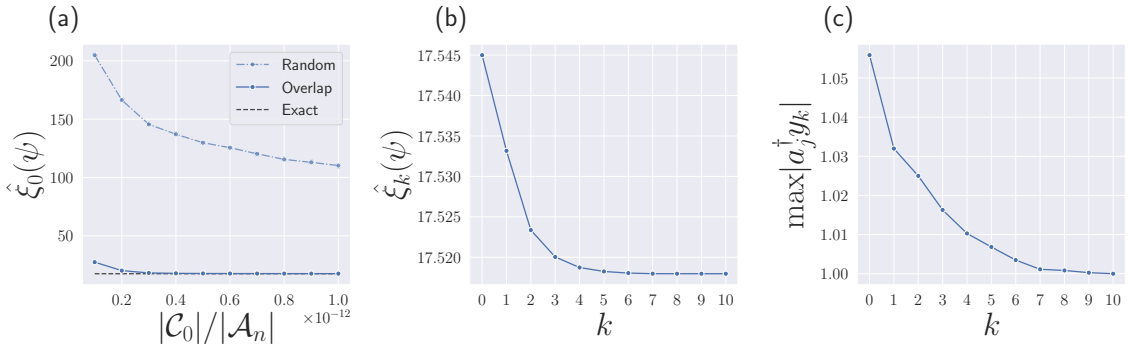


Figure 2: (a) Provisional value $\hat{\xi}_k(\psi)$ in the Algorithm 1 for a Haar random 9-qubit state. The ratio $|\mathcal{C}_0|/|\mathcal{A}_n|$ varies from 10^{-13} to 10^{-12}. We got much better results with the top overlap method compared to the randomly selected $\mathcal{C}_0$. The black dotted line labeled as "Exact" represents $\xi(\psi)$. (b) The convergence of the CG method for the same state. (c) $\max_{a_j \in \mathcal{A}_n} \left|a_j^\dagger y_k\right|$ reached 1.00 after 10 iterations, indicating that the optimal solution has been found.

3.2 CG Method for Stabilizer Extent Calculation

Algorithm 1: Exact Stabilizer Extent Calculation by Column Generation

Input: vector $b \in \mathbb{C}^{2^n}$ corresponding to the state ψ
Output: exact stabilizer extent $\xi(\psi)$

$\mathcal{C}_0 \leftarrow$ Partial set of $\mathcal{A}_n$ `/* Initialize using top overlap` $\left|a_j^\dagger b\right|$ `*/`
for $k = 0, 1, 2, \ldots$ **do**
  $x_k, y_k \leftarrow$ `SolveSOCP`$(\mathcal{C}_k, \boldsymbol{b})$
  $\hat{\xi}_k(\psi) \leftarrow \|x_k\|_1^2$
  $\mathcal{C}' \leftarrow \left\{a_j \in \mathcal{A}_n \;\middle|\; \left|a_j^\dagger y_k\right| > 1\right\}$ `/* Use of subroutine in Section 3.1 */`
  if $\mathcal{C}' = \emptyset$ **then**
    return $\xi(\psi) = \hat{\xi}_k(\psi)$
  $\mathcal{C}_{k+1} \leftarrow \mathcal{C}_k \cup \mathcal{C}'$

Next, we introduce the CG method, the algorithm to compute the exact stabilizer extent $\xi(\psi)$ up to 9-qubit systems. This is outlined in Algorithm 1, and is an iterative algorithm that solves a subproblem restricted to $\mathcal{C} \subseteq \mathcal{A}_n$ per each iteration. It begins with a small subset of columns $\mathcal{C}_0$ and progressively adds a set of columns $\mathcal{C}'$ that violate the constraints of the dual problem (3), and terminate if there are no more violated columns. For further implementation techniques, we direct the reader to [18]. There are two key aspects of this algorithm: the initialization process and the optimality of the solution. We will discuss these in subsequent sections.

3.2.1 Initialization

In the initial step of Algorithm 1, we select a subset $\mathcal{C}_0 \subseteq \mathcal{A}_n$ in descending order of $\left|a_j^\dagger b\right|$, which can be computed efficiently as stated in Theorem 3. The size of $\mathcal{C}_0$ is arbitrary. In the experiment shown in Table 1, we set it to 10,000 for $n \leq 8$ and 100,000 for $n = 9$. The use of $\left|a_j^\dagger b\right| = |\langle \phi_j | \psi \rangle|$ as the indicator can be justified with various interpretations. One

of them is to consider it as the "closeness" between the states $|\phi_j\rangle$ and $|\psi\rangle$, which means choosing states based on their overlaps is reasonable. The numerical experiment result in Figure 2 also supports the effectiveness of this indicator. For a Haar random pure 9-qubit state, even if we use as small subset as $|\mathcal{C}_0| = 10^{-12}|\mathcal{A}_n|$, the obtained value $\hat{\xi}_0(\psi)$ closely approximated the exact extent $\xi(\psi)$ and outperformed randomly selected $\mathcal{C}_0$.

3.2.2 Optimality of Solution

The terminate criterion for Algorithm 1 is the absence of columns that violate the dual constraints $\left|a_j^\dagger y_k\right| \leq 1$, which can be checked efficiently by Theorem 3 as well. The termination of the CG method indicates that the optimal dual solution for problem (3) has been found, and the primal solution x_k is also optimal for problem (2) thanks to the strong duality of the SOCP. Consequently, we can affirm that Algorithm 1 is certain to find the exact stabilizer extent for any state $|\psi\rangle$ once it terminates. The convergence of the CG method is also confirmed in numerical experiments. For the same 9-qubit state as in the initialization, $\max_{a_j \in \mathcal{A}_n} \left|a_j^\dagger y_k\right|$ reaches 1.00 after 10 iterations, indicating the discovery of the optimal solution.

4 Calculation for States With Special Properties

So far, we have explored the method for calculating the stabilizer extent applicable to the general case. However, the method is limited to around $n \leq 9$ due to the superexponential growth of $|\mathcal{S}_n|$. Nevertheless, for states possessing certain properties, computations can be extended to even larger quantum systems.

One of the examples is a product state, $\psi = \otimes_j \psi_j$. The multiplicative of stabilizer extent asserts that $\xi(\psi) = \prod_j \xi(\psi_j)$ holds true if all factors are at most 3-qubits state [22]. Unfortunately, if the factors contain a 4 or more qubits state, we cannot guarantee that the multiplicativity always holds [23]. However, we can still use the tensor product of each solution x_j of $\xi(\psi_j)$ as the initial guess for the solution of $\xi(\psi)$.

In this section, we will explore how to leverage another property of a state: the realness of expansion coefficients. To the best of our knowledge, this property has not been investigated in previous works, and it offers significant advantages for calculation.

4.1 Applications

Firstly, we will present examples of states with real coefficients. One of the well-known examples is the W-state and the GHZ-state, defined as follows:

$$|W\rangle \coloneqq \frac{1}{\sqrt{n}}(|100\ldots0\rangle + |010\ldots0\rangle + \cdots + |000\ldots1\rangle), \quad |GHZ\rangle \coloneqq \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}}$$

Other important applications include eigenstates of quantum many-body Hamiltonians with time-reversal symmetry, whose matrix elements are given by real components. For instance, physically important Hamiltonians such as in XXZ and Heisenberg spin models, transverse-field Ising model, Fermi-Hubbard model, and t-J model all preserve the time-reversal symmetry regardless of the underlying lattice. Beyond condensed matter systems, we may also consider first-principle quantum chemistry Hamiltonians or lattice gauge theory Hamiltonians such as the 1d Schwinger model. Note that the abundance of examples reflects the fact that the microscopic equation of motion is time-reversal symmetric unless there is a spontaneous symmetry breaking or the weak interaction.

In this section, we will compute the stabilizer extent for uniformly random quantum states with real expansion coefficients up to 10-qubit systems. Targeting random states allows us to avoid assuming unnecessary specificity, thereby demonstrating the broad computational potential.

4.2 Reduction of Problem Size

Now, we will state how to efficiently compute the stabilizer extent for the state with real coefficients. Firstly, we define $\mathcal{S}'_n \subset \mathcal{S}_n$ as follows:

$$\mathcal{S}'_n = \{|\phi_j\rangle \in \mathcal{S}_n \mid \langle i|\phi_j\rangle \in \mathbb{R} \text{ for all } i\}.$$

This means that $\mathcal{S}'_n$ is the union of $\mathcal{S}_{n,0}$ and the set of states with $c = 0$ in Theorem 1. Let $\mathcal{A}'_n$ denote the corresponding subset of the columns in $\mathcal{A}_n$. Also refer to Figure 3 for the definition of $\mathcal{A}'_n$. Then, the next lemma holds.

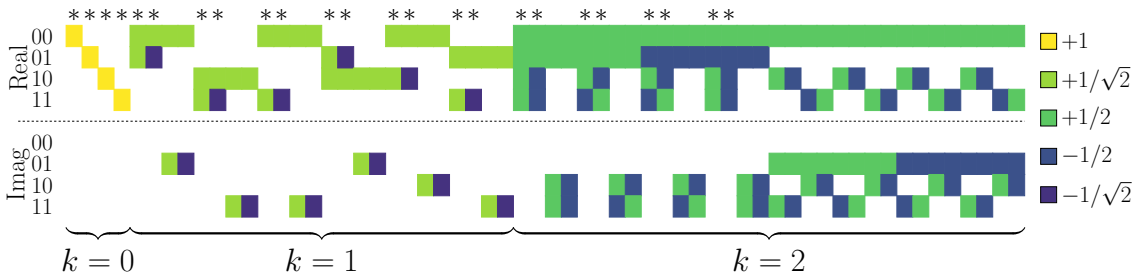


Figure 3: Visualization of the matrix A_n, i.e., the column set $\mathcal{A}_n$, with $n = 2$. The upper half corresponds to the real part, and the lower half corresponds to the imaginary part. The j-th column of this represents a column a_j for a stabilizer state $|\phi_j\rangle$. The k below the matrix corresponds to the k in Theorem 1. By restricting the column set $\mathcal{A}_n$ to the starred columns which are real vectors, we can obtain $\mathcal{A}'_n$.

Lemma 1. *Suppose y is a real vector and satisfies $\left|a^\dagger y\right| \leq 1$ for all $a \in \mathcal{A}'_n$. Then, y satisfies $\left|a^\dagger y\right| \leq 1$ for all $a \in \mathcal{A}_n$.*

Proof. Fix $a \in \mathcal{A}_n$ and let $|\phi\rangle$ denote the corresponding state. We will check that $\left|a^\dagger y\right| \leq 1$. Now, consider $|\phi\rangle$ in the form in Theorem 1. The case $k = 0$ is trivial since then $a \in \mathcal{A}'_n$. Suppose that $k > 0$ and $|\phi\rangle = \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x} i^{c^\top x} |Rx + t\rangle$, and $a^\dagger y = \alpha + i\beta \quad (\alpha, \beta \in \mathbb{R})$. The following two states

$$|\phi_+\rangle := \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x} |Rx + t\rangle, \quad |\phi_-\rangle := \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x + c^\top x} |Rx + t\rangle$$

belong to $\mathcal{S}'_n$. Let define a_+ and a_- as the columns in $\mathcal{A}'_n$ of $|\phi_+\rangle$ and $|\phi_-\rangle$, respectively. Then, we have $a_+^\dagger y = \alpha + \beta, a_-^\dagger y = \alpha - \beta$, and

$$\left|a^\dagger y\right| = \sqrt{\alpha^2 + \beta^2} \leq |\alpha| + |\beta| = \max\{|\alpha + \beta|, |\alpha - \beta|\} \leq 1.$$

The last inequality follows from the assumption, completing the proof. □

As a result, we can derive the following corollary of Thorem 3.

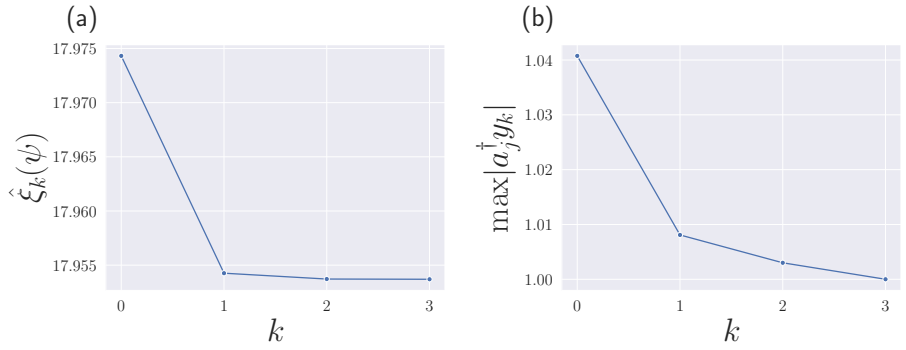


Figure 4: The convergence of the CG method for a uniformly random state with real coefficients. It shows a similar behavior to Figure 2 and indicates the optimality of the solution.

Corollary 1. *Stabilizer fidelity of a n-qubit state $|\psi\rangle$ with real coefficients can be computed in time complexity of $\mathcal{O}(|S_n|/2^n)$ and space complexity of $\mathcal{O}(2^n)$.*

Now, we are ready to prove the following theorem.

Theorem 4. *Suppose that $|\psi\rangle$ is a state with real coefficients. The optimal solution for problem (3) is also optimal for a restricted problem where $\mathcal{A}_n$ is substituted by $\mathcal{A}'_n$.*

Proof. The main idea is the same as in [18, Proposition 2]. Let x^* and y^* be the optimal solutions of the restricted primal and dual problems, namely, problem (2) and problem (3) with the column set $\mathcal{A}'_n$ instead of $\mathcal{A}_n$. We can assure such solutions always exist. Now, we show that the x^*, y^* are optimal not only for the restricted problems but also for the original problems.

Let OPT be the optimal value for the original problems. Since x^* can be a feasible solution for the original primal problem, it is clear that $\mathrm{OPT} \leq \|x^*\|_1$. By the strong duality theorem, OPT is also the optimal value for the original dual problem. From Lemma 1, we can see that y^* is a feasible solution for the original dual problem and $\mathrm{OPT} \geq \mathrm{Re}(b^\dagger y^*)$. Again, by applying the strong duality theorem to the restricted problems, we have $\|x^*\|_1 = \mathrm{Re}(b^\dagger y^*)$, which means that $\mathrm{OPT} = \|x^*\|_1 = \mathrm{Re}(b^\dagger y^*)$. Therefore, x^* and y^* are also optimal solutions for the original problems. □

Thanks to Theorem 4, we can reduce the size of the column set size by a factor of $\mathcal{O}(2^n)$. Figure 4 presents the outcomes of our numerical experiments conducted on a uniformly random 10-qubit state. The results demonstrate our success in computing its stabilizer extent, achieved within a time frame of 7.4 hours.

5 Discussion

In this paper, we have shown that the stabilizer fidelity and stabilizer extent can be efficiently calculated by leveraging the specific structure of stabilizer states. We proposed an algorithm based on the branch and bound method and the CG method to compute the exact stabilizer extent, and demonstrated its applicability to sufficiently large systems. Additionally, we proposed a specialized algorithm for states with real coefficients.

While the idea of applying resource theory to quantum computing has attracted a great amount of interest, the barrier of computational hardness (in particular memory

consumption) has prevented us from gaining further benefits for circuit design and optimization. We envision that the methodology proposed in this work shall not be limited to the stabilizer extent but also expected to generalize to other monotones such as the dyadic negativity [15].

Acknowledgements

We express our sincere gratitude to Naoki Marumo for invaluable comments on the manuscript. We also thank Shigeo Hakkaku, Bartosz Regula, and Ryuji Takagi for helpful discussions. N.Y. wishes to thank JST PRESTO No. JPMJPR2119 and the support from IBM Quantum. This work was supported by JST Grant Number JPMJPF2221. This work was supported by JST ERATO Grant Number JPMJER2302 and JST CREST Grant Number JPMJCR23I4, Japan.

References

[1] Daniel Gottesman. "The Heisenberg Representation of Quantum Computers" (1998). arxiv:quant-ph/9807006.

[2] Michael A. Nielsen and Isaac L. Chuang. "Quantum Computation and Quantum Information: 10th Anniversary Edition". Cambridge University Press. (2010).

[3] Sergei Bravyi and Alexei Kitaev. "Universal Quantum Computation with ideal Clifford gates and noisy ancillas". Physical Review A **71**, 022316 (2005).

[4] Daniel Litinski. "A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery". Quantum **3**, 128 (2019).

[5] Dominic Horsman, Austin G Fowler, Simon Devitt, and Rodney Van Meter. "Surface code quantum computing by lattice surgery". New Journal of Physics **14**, 123011 (2012).

[6] Austin G. Fowler and Craig Gidney. "Low overhead quantum computation using lattice surgery" (2019). arxiv:1808.06709.

[7] Craig Gidney and Martin Ekerå. "How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits". Quantum **5**, 433 (2021).

[8] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush. "Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction". PRX Quantum **2**, 030305 (2021).

[9] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. "Quantum computing enhanced computational catalysis". Physical Review Research **3**, 033055 (2021).

[10] Nobuyuki Yoshioka, Tsuyoshi Okubo, Yasunari Suzuki, Yuki Koizumi, and Wataru Mizukami. "Hunting for quantum-classical crossover in condensed matter problems". npj Quantum Information **10**, 45 (2024).

[11] Sergey Bravyi, Graeme Smith, and John Smolin. "Trading classical and quantum computational resources". Physical Review X **6**, 021043 (2016).

[12] Emanuele Tirrito, Poetri Sonya Tarabunga, Gugliemo Lami, Titas Chanda, Lorenzo Leone, Salvatore F. E. Oliviero, Marcello Dalmonte, Mario Collura, and Alioscia Hamma. "Quantifying nonstabilizerness through entanglement spectrum flatness". Physical Review A: Atomic, Molecular, and Optical Physics **109**, L040401 (2024).

[13] Oliver Hahn, Alessandro Ferraro, Lina Hultquist, Giulia Ferrini, and Laura García-Álvarez. "Quantifying Qubit Magic Resource with Gottesman-Kitaev-Preskill Encoding". Physical Review Letters **128**, 210502 (2022).
[14] Tobias Haug, Soovin Lee, and M. S. Kim. "Efficient quantum algorithms for stabilizer entropies" (2023). arxiv:2305.19152.
[15] James R. Seddon, Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl T. Campbell. "Quantifying Quantum Speedups: Improved Classical Simulation From Tighter Magic Monotones". PRX Quantum **2**, 010345 (2021).
[16] Zi-Wen Liu and Andreas Winter. "Many-body quantum magic". PRX Quantum **3**, 020333 (2022).
[17] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. "Stabilizer Rényi Entropy". Physical Review Letters **128**, 050402 (2022).
[18] Hiroki Hamaguchi, Kou Hamada, and Nobuyuki Yoshioka. "Handbook for Efficiently Quantifying Robustness of Magic" (2023). arxiv:2311.01362.
[19] SciPy. "scipy.sparse.csc_matrix". SciPy.
[20] Scott Aaronson and Daniel Gottesman. "Improved simulation of stabilizer circuits". Physical Review A: Atomic, Molecular, and Optical Physics **70**, 052328 (2004).
[21] Mark Howard and Earl Campbell. "Application of a resource theory for magic states to fault-tolerant quantum computing". Physical Review Letters **118**, 090501 (2017).
[22] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. "Simulation of quantum circuits by low-rank stabilizer decompositions". Quantum **3**, 181 (2019).
[23] Arne Heimendahl, Felipe Montealegre-Mora, Frank Vallentin, and David Gross. "Stabilizer extent is not multiplicative". Quantum **5**, 400 (2021).
[24] Stephen Boyd and Lieven Vandenberghe. "Convex optimization". Cambridge University Press. (2004).
[25] MOSEK ApS. "Mosek modeling cookbook". MOSEK (2019).
[26] Steven Diamond and Stephen Boyd. "CVXPY: A python-embedded modeling language for convex optimization". Journal of Machine Learning Research **17**, 2909–2913 (2016).
[27] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. "A rewriting system for convex optimization problems". Journal of Control and Decision **5**, 42–60 (2018).
[28] Guy Desaulniers, Jacques Desrosiers, and Marius M. Solomon Solomon, editors. "Column Generation". Springer US. (2005).
[29] G.I. Struchalin, Ya. A. Zagorovskii, E.V. Kovlakov, S.S. Straupe, and S.P. Kulik. "Experimental Estimation of Quantum State Properties from Classical Shadows". PRX Quantum **2**, 010307 (2021).
[30] Maarten Van den Nest. "Classical simulation of quantum computation, the gottesman-Knill theorem, and slightly beyond". Quantum Inf. Comput. **10**, 258–271 (2010).
[31] Jeroen Dehaene and Bart De Moor. "Clifford group, stabilizer states, and linear and quadratic operations over GF(2)". Physical Review A **68**, 042318 (2003).
[32] Hiroki Hamaguchi, Kou Hamada, and Nobuyuki Yoshioka. "stabilizer_extent". GitHub Repository (2024).

A Fast Algorithm for Overlap

In this section, we will explain the details of the stabilizer pruning in Section 3.1 and introduce some heuristics to improve efficiency.

A.1 Efficient Enumeration of Stabilizer States

In this section, we prove Theorem 1.

Theorem 1. *For all $k \in \{1, \dots, n\}$, define the following set:*

$$\begin{aligned}
\mathcal{Q}_k &:= \left\{ Q \;\middle|\; Q \in \mathbb{F}_2^{k\times k} \text{ is a upper triangular matrix} \right\}, \\
\mathcal{R}_k &:= \left\{ R \;\middle|\; R \in \mathbb{F}_2^{n\times k} \text{ is a reduced row echelon form matrix with } \operatorname{rank}(R) = k \right\}, \\
\mathcal{T}_k(R) &:= \left\{ t \;\middle|\; t \in \mathbb{F}_2^n \text{ is a representative of element in the quotient space } \mathbb{F}_2^n / \operatorname{Im}(R) \right\}.
\end{aligned}$$

Define the set of states $\mathcal{S}_{n,k}$ as

$$\mathcal{S}_{n,k} := \left\{ \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top Q x} i^{c^\top x} |Rx + t\rangle \;\middle|\; Q \in \mathcal{Q}_k, c \in \mathbb{F}_2^n, R \in \mathcal{R}_k, t \in \mathcal{T}_k(R) \right\}, \tag{4}$$

and define $\mathcal{S}_{n,0} := \{|t\rangle \mid t \in \mathbb{F}_2^n\}$. Then, we have $\bigcup_{k=0}^n \mathcal{S}_{n,k} = \mathcal{S}_n$.

Proof. The main idea comes from [29]. From previous works [29, Theorem 2], [30, Section 5], and [31, Theorem 5.(ii)], we know that any state in $\bigcup_{k=0}^n \mathcal{S}_{n,k}$ is a stabilizer state. Thus, we can construct a inclusion map from $\bigcup_{k=0}^n \mathcal{S}_{n,k}$ to $\mathcal{S}_n$. In this proof, we will show that this map is bijective, which means this map is an identity mapping. The assertion is trivial for the case $k = 0$ with 2^n instances. We will only consider the case $k > 0$. Define $f : (Q, c, R, t) \mapsto |\phi\rangle$ as a map from (Q, c, R, t) to the corresponding stabilizer state $|\phi\rangle$. We will confirm that f is bijective. First, we show that f is injective. We can say that

$$\begin{aligned}
&\left\{ R_1 x + t_1 \;\middle|\; x \in \mathbb{F}_2^{n-k} \right\} = \left\{ R_2 x + t_2 \;\middle|\; x \in \mathbb{F}_2^{n-k} \right\} \\
\iff &(\operatorname{Im}(R_1) = \operatorname{Im}(R_2)) \wedge (t_1 - t_2 \in \operatorname{Im}(R_1)) \\
\iff &R_1 = R_2 \wedge t_1 = t_2.
\end{aligned}$$

The last equivalence is due to the property of the reduced row echelon form and the quotient space. Given that Q is an upper triangular matrix, both Q and c can be uniquely reconstructed from the expansion coefficients of the state. Consequently, the f is injective.

Next, we show that f is surjective. Since f is injective, we only have to show that the cardinality of the domain is equal to that of the codomain, i.e., $-2^n + |\mathcal{S}_n|$. It is known that the number of $\mathbb{F}_2^{n\times k}$ reduced row echelon form matrices R with $\operatorname{rank}(R) = k$ is $\begin{bmatrix} n \\ k \end{bmatrix}_2$, which is a q-binomial coefficient with $q = 2$. Therefore, the number of Q, c, R, t is $2^{k(k+1)/2}, 2^k, \begin{bmatrix} n \\ k \end{bmatrix}_2, 2^{n-k}$, respectively, and the total number of states is

$$\sum_{k=1}^{n} 2^{k(k+1)/2} 2^k \begin{bmatrix} n \\ k \end{bmatrix}_2 2^{n-k} = -2^n + 2^n \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_2 2^{k(k+1)/2} = -2^n + 2^n \prod_{k=1}^{n} (2^k + 1) = -2^n + |\mathcal{S}_n|.$$

In the second to last equation, we used the q-binomial theorem. Therefore, the mapping is surjective, which concludes the proof. □

In Theorem 1, we used $\mathbb{F}_2$. However, from the perspective of the branch and bound method, it is more practical to use $\{0,1\} \subset \mathbb{Z}$ and permit the term $c^\top x$ to be any integer value. Otherwise, $-1 = i^{1+1} \neq i^0 = 1$ although $1+1=0$ in $\mathbb{F}_2$, which makes the algorithm more complicated. Hence, the subsequent corollary is valuable.

Corollary 2. *In Theorem 1, We can substitute $\mathbb{F}_2$ with $\{0,1\} \subset \mathbb{Z}$.*

Proof. By substituting $\mathbb{F}_2$ with $\{0,1\} \subset \mathbb{Z}$, the term $(-1)^{x^\top Qx}$ is invariant, and the term $i^{c^\top x}$ is multiplied by -1 iff $p \equiv 2,3 \pmod 4$, where p is the number of i such that $c_i = 1$ and $x_i = 1$. Now, we consider the following form for $k > 0$:

$$|\phi\rangle \coloneqq \frac{1}{2^{k/2}} \sum_{x=0}^{2^k-1} (-1)^{x^\top (Q+Q')x} i^{c^\top x} |Rx+t\rangle \tag{6}$$

where $Q'_{ij} = 1$ iff $(i<j) \wedge (c_i = c_j = 1)$. Now, if the pair (Q,c,R,t) in (6) is the same as that of the original form (4), then the two states represent the exactly same state since

$$(-1)^{x^\top Q'x} = (-1)^{\binom{p}{2}} = \begin{cases} 1 & \text{if } p \equiv 0,1 \pmod 4, \\ -1 & \text{if } p \equiv 2,3 \pmod 4. \end{cases}$$

Therefore, by identifying the $Q+Q'$ in $\mathbb{Z}$ with the Q in $\mathbb{F}_2$, we can conclude the proof. □

A.2 Branching in Branch and Bound Method

In this section, we prove Theorem 2. Note that problem (5) is equivalent to the following problem thank to Corollary 2:

$$\max_{Q \in \{0,1\}^{n\times n}, c \in \{0,1\}^n} \left\{ \left| \sum_{x=0}^{2^n-1} (-1)^{x^\top Qx} i^{c^\top x} P_x \right| \right\}.$$

Theorem 2. *Problem (5) can be solved in $\mathcal{O}\left(2^{n+n(n+1)/2}\right)$ time complexity and $\mathcal{O}(2^n)$ space complexity.*

Proof. Define $x \coloneqq \begin{bmatrix} x_0 \\ \overline{x} \end{bmatrix}$ $(x_0 \in \{0,1\}, \overline{x} \in \{0,1\}^{n-1})$, $Q \coloneqq \begin{bmatrix} Q_{00} & Q_0^\top \\ 0 & \overline{Q} \end{bmatrix}$ $(Q_{00} \in \{0,1\}, Q_0 \in \{0,1\}^{n-1}, \overline{Q} \in \{0,1\}^{(n-1)\times(n-1)})$, and $c \coloneqq \begin{bmatrix} c_0 \\ \overline{c} \end{bmatrix}$ $(c_0 \in \{0,1\}, \overline{c} \in \{0,1\}^{n-1})$. Since $x^\top Qx = x_0(Q_{00} + Q_0^\top \overline{x}) + \overline{x}^\top \overline{Q} \overline{x}$ and $c^\top x = c_0 x_0 + \overline{c}^\top \overline{x}$, we can derive that

$$\begin{aligned} \sum_{x=0}^{2^n-1} (-1)^{x^\top Qx} i^{c^\top x} P_x &= \sum_{\overline{x}=0}^{2^{n-1}-1} (-1)^{\overline{x}^\top \overline{Q} \overline{x}} i^{\overline{c}^\top \overline{x}} \left(P_{2\overline{x}} + (-1)^{Q_{00}+Q_0^\top \overline{x}} i^{c_0} P_{2\overline{x}+1} \right) \\ &= \sum_{\overline{x}=0}^{2^{n-1}-1} (-1)^{\overline{x}^\top \overline{Q} \overline{x}} i^{\overline{c}^\top \overline{x}} \overline{P}_{\overline{x}} \end{aligned} \tag{7}$$

where $\overline{P}_{\overline{x}} \coloneqq P_{2\overline{x}} + (-1)^{Q_{00}+Q_0^\top \overline{x}} i^{c_0} P_{2\overline{x}+1}$, and we identify a vector $\begin{bmatrix} x_0 & x_1 & \cdots & x_{n-1} \end{bmatrix}^\top$ as an integer $\sum_{i=0}^{n-1} x_i 2^i$. Since (7) is the same form as the original one, this problem can be solved recursively by fixing the value Q_{00}, Q_0 and c_0.

We now analyze the time complexity of this recursive algorithm. There are 2^{n+1} possible combinations of Q_{00}, Q_0, and c_0. For each such combination, $\overline{P}_{\overline{x}}$ can be computed

in $\mathcal{O}(n2^{n-1})$ time. Hence, we establish the following recurrence relation for the time complexity $T(n)$:

$$T(n) = 2^{n+1}(T(n-1) + n2^{n-1}), \quad T(1) = 4.$$

Solving this recurrence relation yields

$$T(n) = 2^{n+\frac{n(n+1)}{2}} + \sum_{d=2}^{n} 2^{n+\frac{n(n+1)}{2}-\frac{d(d-1)}{2}} d,$$

$$\frac{T(n)}{2^{n+\frac{n(n+1)}{2}}} = 1 + \sum_{d=2}^{n} 2^{-\frac{d(d-1)}{2}} d \leq 1 + \sum_{d=2}^{n} 2^{-d+1} d = 4 - (n+2)2^{-n+1} \to 4 \quad (n \to \infty).$$

Therefore, the time complexity is $\mathcal{O}\left(2^{n+n(n+1)/2}\right)$. The statement of the space complexity is apparent from Figure 1. □

While the algorithm and proof presented above may seem somehow rough, our actual implementation is significantly more precise and efficient. You can access it at GitHub [32]. Moreover, we can enhance efficiency further by employing branch-cut heuristics, as we will explain in the next section.

A.3 Pruning for the Branch and Bound Method

In the previous section, we explained the stabilizer pruning. This algorithm can be much faster by using the branch-cut heuristics we will introduce in this section. Firstly, recall that we are maximizing the following:

$$\max_{Q,c} \left\{ \left| \sum_{x=0}^{2^n-1} (-1)^{x^\top Q x} i^{c^\top x} P_x \right| \right\}.$$

This can be easily bounded by

$$\max_{Q,c} \left\{ \left| \sum_{x=0}^{2^n-1} (-1)^{x^\top Q x} i^{c^\top x} P_x \right| \right\} \leq \max_{Q,c} \left\{ \sum_{x=0}^{2^n-1} \left| (-1)^{x^\top Q x} i^{c^\top x} P_x \right| \right\} = \sum_{x=0}^{2^n-1} |P_x|.$$

Such a bound is important for the branch and bound method, because it allows us to terminate the branch if the current value is inferior to the bound. However, this bound can be more refined. Since each coefficient takes only $1, -1, i$ or $-i$, we can bound as follows

$$\max_{Q,c} \left\{ \left| \sum_{x=0}^{2^n-1} (-1)^{x^\top Q x} i^{c^\top x} P_x \right| \right\} \leq \max_{c_x} \left\{ \left| \sum_{x=0}^{2^n-1} i^{c_x} P_x \right| \right\} \tag{8}$$

where c_x takes values independently from the set $\{0, 1, 2, 3\}$. Let $P \coloneqq \sum_{x=0}^{2^n-1} i^{c_x} P_x$, and define $\theta_x \coloneqq \arg(i^{c_x} P_x)$. We have

$$\max_{c_x} \{|P|\} = \max_{c_x,\theta} \left\{ \langle P, e^{i\theta} \rangle \right\} = \max_{c_x,\theta} \left\{ \sum_{x=0}^{2^n-1} \langle i^{c_x} P_x, e^{i\theta} \rangle \right\} = \max_{c_x,\theta} \left\{ \sum_{x=0}^{2^n-1} |P_x| \cos(\theta_x - \theta) \right\} \tag{9}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of complex values. Then, we can confirm that Algorithm 2 is certain to return the optimal solution for (8) as follows. If we fix the value of θ in (9), the optimal values of c_x can be determined so that $\cos(\theta_x - \theta)$ is maximized, i.e., $\theta_x \in [\theta - \pi/4, \theta + \pi/4)$. Then, instead of trying all the possible values of θ, we can run Algorithm 2 to cover all the possible optimal solutions of c_x, which are sufficient to

Algorithm 2: Bounding for the Branch and Bound Method

Input: Coefficients P_x for $x = 0, 1, \ldots, 2^n - 1$

Output: The answer for problem (8)

Modify the c_x and sort so that $0 \leq \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{2^n-1} < \pi/2$.

ans $\leftarrow 0$

for $x \leftarrow 0$ **to** $2^n - 1$ **do**

  ans $\leftarrow \max\left(\text{ans}, \left|\sum_{x=0}^{2^n-1} i^{c_x} P_x\right|\right)$

  $c_x \leftarrow c_x + 1$

return ans

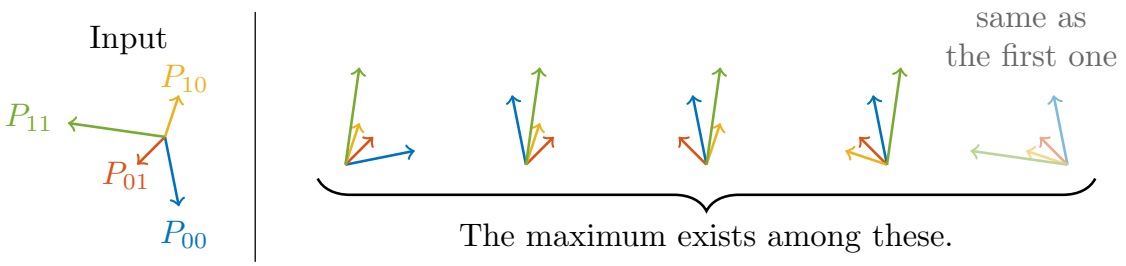


Figure 5: Visualization of Algorithm 2. Suppose that $n = 2$ and P_x are represented as the vectors in the complex plane (e.g., $P_{00} = 1 - 5i$) in the left figure. Iterating the loop in Algorithm 2 yields 2^n patterns of the coefficients c_x, as depicted in the right figure. The maximum of problem (8) exists among these 2^n patterns.

calculate $\max_{c_x}\{|P|\}$. Refer to Figure 5 for a visual representation of this algorithm. The time complexity of this approach is $\mathcal{O}(n2^n)$ owing to the sorting of 2^n elements.

As the end of this section, we evaluate the performance of this bound. We can obtain the lower bound of (9) by taking the expected value with respect to θ as follows:

$$\begin{aligned}\max_{c_x}\{|P|\} &= \max_{c_x,\theta}\left\{\sum_{x=0}^{2^n-1}|P_x|\cos(\theta_x-\theta)\right\}\\ &\geq \mathbb{E}\left[\max_{c_x}\left\{\sum_{x=0}^{2^n-1}|P_x|\cos(\theta_x-\theta)\right\}\right] = \sum_{x=0}^{2^n-1}|P_x|\cdot\mathbb{E}\left[\max_{c_x}\{\cos(\theta_x-\theta)\}\right]. \quad (10)\end{aligned}$$

Here, we assume θ is drawn from the uniform distribution over $[0, 2\pi)$. Then, we can replace each term $\mathbb{E}[\max_{c_x}\{\cos(\theta_x - \theta)\}]$ with $\mathbb{E}[\cos(\theta'_x)]$ where θ'_x follows the uniform distribution over the interval $[-\pi/4, +\pi/4)$. Then, we can derive that

$$\frac{\max_{c_x}\{|P|\}}{\sum_{x=0}^{2^n-1}|P_x|} \geq \frac{\sum_{x=0}^{2^n-1}|P_x|\cdot\mathbb{E}[\cos\theta'_x]}{\sum_{x=0}^{2^n-1}|P_x|} = \frac{\int_{-\frac{\pi}{4}}^{+\frac{\pi}{4}}\cos(\theta)\,\mathrm{d}\theta}{\pi/2} = \frac{2\sqrt{2}}{\pi} = 0.900316\cdots.$$

The result of a numerical experiment suggests that this lower bound serves as a rough approximation of the ratio. We independently sampled P_x from the standard normal distribution and θ_x from the uniform distribution over $[0, 2\pi)$. The numerical experiment results obtained are as follows. Firstly, the average of

$$\frac{\max_{c_x}\{|P|\}}{\sum_{x=0}^{2^n-1}|P_x|} = \frac{\max_{c_x}\left\{\left|\sum_{x=0}^{2^n-1} i^{c_x}P_x\right|\right\}}{\sum_{x=0}^{2^n-1}|P_x|}$$

over 100 runs yields 0.935624 for $n = 4$. This confirms that Algorithm 2 provides a better bound compared to $\sum_{x=0}^{2^n-1} |P_x|$. However, in the same setting, it turned out that the average of

$$\frac{\max_{Q,c} \left\{ \left| \sum_{x=0}^{2^n-1} (-1)^{x^\top Q x} i^{c^\top x} P_x \right| \right\}}{\sum_{x=0}^{2^n-1} |P_x|}$$

yields 0.824056, implying the bound (8) may not necessarily be optimal. Whether a better bound can be obtained with fewer computational cost is left for an open problem.

One-shot and asymptotic classical capacity in general physical theories

Shintaro Minagawa[1] * Hayato Arai[2] †

[1] *Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan*
[2] *RIKEN, Center for Quantum Computing, Mathematical Quantum Information RIKEN Hakubi Research Team, 2-1 Hirosawa, Wako 351-0198, Japan*

Abstract. We extend classical information transmission via quantum channels to general physical theories where states and measurements are operationally defined. By generalizing the method invented by Wang and Renner [Phys. Rev. Lett. 108, 200501 (2012)], we obtain the upper bound of the one-shot classical capacity, the optimal rate of classical information transmitted using a single channel constrained by a certain error probability, in general physical theories. We also derive its lower bound by showing the existence of a good code. Then we demonstrate the asymptotic equivalence between classical capacity and hypothesis testing relative entropy even in any general physical theory.

Keywords: General probabilistic theories, Classical capacity, Hypothesis testing

1 Introduction

Since Shannon invented the information theory [48], it has been increasingly important (see e.g., Ref. [14]). The goal of information theory is basically to express the optimal efficiency for some tasks, and the optimal efficiencies for different tasks are sometimes equivalent or directly related through some information quantities like mutual information. A typical example is the asymptotic equivalence between the exponent rate of hypothesis testing and classical information transmission capacity [52].

Recently, as quantum information theory (see e.g., Refs. [56, 55]) has flourished, similar relations are known in quantum theory. In particular, the same relationship between hypothesis testing and channel capacity also holds in quantum theory [20, 37, 44, 45, 21, 7]. Such facts imply that an information theory should possess such relations between the optimal efficiencies for some tasks independently of the mathematical structure of its background physical systems.

However, when we establish an information theory standing by the operationally minimum principles, possible models of background physical systems are not restricted to classical and quantum theory. Such theories are called *General Probabilistic Theories* [22, 15, 42, 35, 30, 31, 16, 18, 38, 43, 19, 6, 39, 27, 49, 3, 28, 36, 10, 11, 34, 5, 12, 13, 26, 29, 25, 1, 57, 47, 46, 53, 40, 2, 24, 33, 41] (for a review, see, e.g., Refs. [15, 22, 42, 35]). The framework of GPTs is a kind of generalization of classical and quantum theory whose states and measurements are operationally defined, and studies of GPTs have been widespread recently.

Even in such general models, some properties of information theory also hold similarly to quantum theory. One of such results of preceding studies of GPTs is the no-cloning theorem in GPTs [4]. It is clarified that any model except for classical theory cannot copy any information freely, similar to the no-cloning theorem in quantum theory, which means that quantum theory is not a special theory with no-cloning, but the classical theory is a special theory with cloning.

On the other hand, some properties of information theory are drastically changed in GPTs. A typical example is *entropy*. Because entropy is not generalized straightforwardly in GPTs [3, 49, 28, 40, 41], we cannot easily obtain a similar result of optimal efficiency for certain information tasks. Therefore, whether there are the same relations between optimal efficiencies for different tasks as the relations in classical and quantum theory is a difficult problem.

In this work [32], we discuss hypothesis testing and classical information transmission in GPTs in the same way as classical and quantum theory following Ref. [54]. Next, we estimate the upper and lower bound of one-shot classical capacity by hypothesis testing relative entropy[1] in GPTs. As a result, we obtain upper and lower bounds similar to that of quantum theory. Moreover, due to the construction of the achievable case of our bound, our result of the one-shot case can be applied to the asymptotic case even though the asymptotic scenario is complicated in GPTs. Consequently, we show the asymptotic equivalence between the above two efficiencies even in GPTs.

2 General probabilistic theories

Here we introduce the mathematical basics of GPTs following Refs. [57, 22, 42, 35]. Let V be a finite-dimensional real vector space and the subset $K \subset V$ be a positive cone, i.e., a set satisfying the following three conditions: (i) $\lambda x \in K$ holds for any $x \in K$ and any $\lambda \geq 0$. (ii) K is convex and has a non-empty interior. (iii) $K \cap (-K) = \{0\}$. The dual cone of K, denoting K^* is defined as follows:

$$K^* := \{y \in V^* \mid \langle y, x\rangle \geq 0 \ \forall x \in K\} \tag{1}$$

where $\langle , \rangle$ is the inner product of the vector space V. Besides, an inner point $u \in K^*$, called *unit effect*, is fixed for a model. Then, a state in this model is defined as an element $\rho \in K$ satisfying $\langle \rho, u\rangle = 1$. The state space,

*minagawa.shintaro@nagoya-u.jp
†hayato.arai@riken.jp

[1] This quantity was first introduced by Ref. [9], and the relationship to one-shot classical capacity was derived by Ref. [54], but we generalize it to GPTs in this paper.

i.e., the set of all states, is denoted as $S(K)$. Due to the convexity of K, the state space $S(K)$ is also convex.

Also, a measurement is defined as a family $\boldsymbol{e} := \{e_j\}_{j\in J}$ satisfying $e_j \in K^*$ for any $j \in J$ and $\sum_{j\in J} e_j = u$. The measurement space, i.e., the set of all measurements with finite outcomes, is denoted as $\mathcal{M}(K)$. Here, $\langle e_j, \rho\rangle$ corresponds to the probability of obtaining an outcome $j \in J$ when we perform a measurement $\mathbf{e}$ to a state $\rho \in \mathcal{S}(K)$. Next, we define an order relation $\geq$ on K^*. We say that $f \geq e$ if $f - e \in K^*$. This means that for any element $x \in K$, $\langle f, x\rangle \geq \langle e, x\rangle$. Here, we remark that a family $\{e, u - e\}$ is a measurement in $\mathcal{M}(K)$ if and only if the element $e \in K^*$ satisfies $0 \leq e \leq u$ by using the above order relation.

Next, we define a *measurement channel* associated with a measurement $\mathbf{e}$ as the following map $\mathcal{E}_{\mathbf{e}}$ from $\mathcal{S}(K)$ to $\mathcal{S}(\mathbb{R}^n_+)$ [42]:

$$\mathcal{E}_{\mathbf{e}}(\rho) := \sum_{j\in J} \langle e_j, \rho\rangle \, |j\rangle\langle j| \ . \tag{2}$$

We also define an adjoint map of a measurement channel $\mathcal{E}_{\mathbf{e}}$ as $\mathcal{E}^\dagger_{\mathbf{e}}(f) := \sum_{j\in J}\langle f, |j\rangle\langle j|\rangle e_j$. Note that the data-processing inequality also holds for measurement channels even in the framework of GPTs.

3 Hypothesis testing relative entropy in GPTs

Next, we introduce hypothesis testing relative entropy in general models. In quantum theory, hypothesis testing relative entropy is defined for $0 \leq \epsilon \leq 1$ as follows [9, 8, 54, 17]:

$$D^\epsilon_{\mathrm{H}}(\rho||\sigma) := -\log_2 \min_{\substack{E:0\leq E\leq \mathbb{1},\\ \mathrm{Tr}\{E\rho\}\geq 1-\epsilon}} \mathrm{Tr}\{E\sigma\} \ , \tag{3}$$

where $\mathbb{1}$ is an identity operator.

As a generalization of this definition, we can introduce hypothesis testing relative entropy in GPTs as follows:

Definition 1 *Let $\rho, \sigma \in \Omega$ be states and q be an effect where $0 \leq \langle q, \rho\rangle \leq 1$ holds for any state $\rho \in \Omega$. Let $0 \leq \epsilon \leq 1$ be a real value. We define hypothesis testing relative entropy in the considering GPT as follows:*

$$D^\epsilon_{\mathrm{H,G}}(\rho||\sigma) := -\log_2 \min_{\substack{q:\ 0\leq q\leq u,\\ \langle q,\rho\rangle\geq 1-\epsilon}} \langle q, \sigma\rangle \ . \tag{4}$$

4 One-shot classical capacity in GPTs

Here we consider one-shot classical capacity in GPTs based on the setup given by Ref. [54]. First, we describe our setup of one-shot classical information transmission from the sender in the system A to the receiver in the system B.

The sender and receiver share a channel Φ from $\mathcal{X}$ to $\mathcal{S}(K)$ defined as $\Phi(|x\rangle\langle x|) = \sigma^{\mathrm{B}}_x$, where $\mathcal{X}$ is an alphabet. The sender encodes an n-length bit string $j \in \Gamma := \{0, 1, 2, \cdots, 2^n - 1\}$ to $x \in \mathcal{X}$ by using a function $g(j) = x$ called *encoder*. Also, the set $\mathcal{G} = g(\Gamma)$ and the element $g(j)$ are called *codebook* and *codeword*, respectively. The

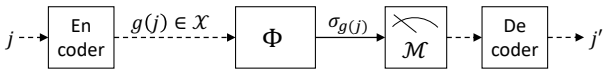


Figure 1: The setup of sending classical information. Reprinted from [32]. The dotted arrows mean transmissions of classical information. The solid line means a transmission of a state in the general theory. The sender chooses the massage $j \in \Gamma$ and encodes it by a function $g : \Gamma \to \mathcal{X}$. Classical information $g(j)$ is transformed into the state $\sigma_{g(j)}$ by a channel Φ whose input is the classical information and whose output is the state of the GPT of the receiver's system. Then the receiver performs a measurement $\mathcal{M}$ to $\sigma_{g(j)}$ and decodes that result to obtain classical information j'. We say that the measurement decodes the message j correctly when $j' = j$.

receiver performs a measurement $\mathbf{m}^{\mathrm{B}} := \{m^{\mathrm{B}}_j\}_{j\in\Gamma}$ to the arrived state $\sigma^{\mathrm{B}}_{g(j)}$, where $m^{\mathrm{B}}_j \geq 0$ and $\sum_{j\in\Gamma} m^{\mathrm{B}}_j = u$. The error probability for a given message $j \in \Gamma$, encoder g and measurement $\mathbf{m}^{\mathrm{B}}$ is defined as

$$\Pr(\mathrm{error}|j, g, \mathbf{m}^{\mathrm{B}}) = \langle u - m^{\mathrm{B}}_j, \sigma^{\mathrm{B}}_{g(j)}\rangle \ . \tag{5}$$

The sender and receiver aim to maximize the size of bit strings under the condition that the average error is small enough. In order to define the rate of this task and the capacity of the channel, we define a $(2^n, \epsilon)$-*code* that fulfills this aim.

Definition 2 *Let $\Gamma = \{0, 1, \ldots, 2^n - 1\}$ be a n-length bit string. A $(2^n, \epsilon)$-code for a map $\Phi : |x\rangle\langle x| \mapsto \sigma^{\mathrm{B}}_x$ consists of an encoder $g : \Gamma \to \mathcal{X}$ and decoding measurement $\mathbf{m}^{\mathrm{B}} := \{m^{\mathrm{B}}_j\}_{j\in\Gamma}$ whose average error probability when the messages $j \in \Gamma$ is chosen uniformly at random is bounded from the above by ϵ, in a formula,*

$$\Pr(\mathrm{error}|g, \mathbf{m}^{\mathrm{B}}) := \frac{1}{2^n}\sum_{j\in\Gamma} \Pr(\mathrm{error}|j, g, \mathbf{m}^{\mathrm{B}}) \leq \epsilon \ . \tag{6}$$

Then, we define the rate and capacity as follows.

Definition 3 *A real number $R \geq 0$ is a one-shot ϵ-achievable rate for one-shot classical information transmission through Φ if there is a $(2^R, \epsilon)$-code.*

Definition 4 *The one-shot ϵ-classical capacity of a map Φ, $\mathrm{C}^\epsilon(\Phi)$ is defined as*

$$\mathrm{C}^\epsilon(\Phi) := \sup\{R \mid R \text{ is a one-shot } \epsilon\text{-achievable rate}\} \ . \tag{7}$$

Now, we define the following ensemble:

$$\pi^{\mathrm{AB}}_{P_X} := \sum_{x\in\mathcal{X}} P_X(x) \, |x\rangle\langle x|^{\mathrm{A}} \otimes \sigma^{\mathrm{B}}_x \ , \tag{8}$$

where $P_X(x)$ is a probability distribution of a random variable associated with the alphabet $\mathcal{X}$. The marginal states with respect to A and B are the followings, respectively:

$$\pi^{\mathrm{A}}_{P_X} = \sum_{x\in\mathcal{X}} P_X(x) \, |x\rangle\langle x|^{\mathrm{A}} \ , \quad \pi^{\mathrm{B}}_{P_X} = \sum_{x\in\mathcal{X}} P_X(x)\sigma^{\mathrm{B}}_x \ . \tag{9}$$

In quantum theory, the ϵ-one-shot classical capacity is asymptotically equivalent to the optimal hypothesis testing relative entropy between the above ensemble π^{AB} and the product of its marginal states. This paper shows that the equivalence also holds even in GPTs.

Firstly, we show the converse part, i.e., the upper bound of $\mathrm{C}^{\epsilon}(\Phi)$ by applying the generalization of data-processing inequality.

Theorem 5 *The ϵ-one-shot classical capacity of a map $\Phi : |x\rangle\langle x| \mapsto \sigma_x^{\mathrm{B}}$ is bounded as follows:*

$$\mathrm{C}^{\epsilon}(\Phi) \le \sup_{P_X} D^{\epsilon}_{\mathrm{H,G}}(\pi^{\mathrm{AB}}_{P_X} || \pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X}) \,, \tag{10}$$

where the supremum is taken over all probability distribution P_X, and where $\pi^{\mathrm{A}}_{P_X}$ and $\pi^{\mathrm{B}}_{P_X}$ is a marginal state of $\pi^{\mathrm{AB}}_{P_X}$ with regard to system A *and* B*, respectively.*

Next, we show the achievable part, which gives the lower bound of $\mathrm{C}^{\epsilon}(\Phi)$.

Theorem 6 *The ϵ-one-shot classical capacity of a map $\Phi : x \in \mathcal{X} \to \sigma_x^{\mathrm{B}}$ satisfies the following inequality for any $\epsilon' \in (0, \epsilon)$, any $s > 1$, and any $t > s$ satisfying $\epsilon > s\epsilon'$:*

$$\mathrm{C}^{\epsilon}(\Phi) \ge \sup_{P_X} D^{\epsilon'}_{\mathrm{H,G}}(\pi^{\mathrm{AB}}_{P_X} || \pi^{\mathrm{A}}_{P_X} \otimes \pi^{\mathrm{B}}_{P_X}) - \log_2 \frac{t}{\epsilon - s\epsilon'} \,. \tag{11}$$

5 Asymptotic i.i.d. case

In this section, we consider how the capacity is expressed when a channel is used m (a positive integer) times in an independent and identical distribution (i.i.d.).

We express m-length sequence consists of alphabet $\mathcal{X}$ as $x_1 \dots x_m$. Let us fix the probability of occurrence for each symbol as $P_X(x)$. Then, when a channel Φ is used m times, the sender and receiver can share the following state:

$$\pi^{AB}_{P_X{}^m} := \sum_{x_1 \dots x_m \in \mathcal{X}^n} P_X{}^m(x_1 \dots x_m) \, |x_1 \dots x_m\rangle\langle x_1 \dots x_m|^{\mathrm{A}} \otimes \sigma^{\mathrm{B}}_{x_1 \dots x_m} \,. \tag{12}$$

Here, $\mathcal{X}^m$ means the set of all m-length sequences consist of the alphabet $\mathcal{X}$ and $\sigma^{\mathrm{B}}_{x_1 \dots x_m}$ is the abbreviation of $\sigma^{\mathrm{B}}_{x_1} \otimes \sigma^{\mathrm{B}}_{x_2} \otimes \cdots \otimes \sigma^{\mathrm{B}}_{x_m}$. We denote the above map from $\sigma^{\mathrm{B}}_{x_1 \dots x_m}$ to $\sigma^{\mathrm{B}}_{x_1} \otimes \sigma^{\mathrm{B}}_{x_2} \otimes \cdots \otimes \sigma^{\mathrm{B}}_{x_m}$ as $\Phi^{\otimes m}$.

Here, we remark on the composition of the model of GPTs. In the standard setting of GPTs, i.e., the case when we assume no-signaling and local tomography [6, 23], an n-composite model of a model defined by K is defined by a positive cone K_n satisfying

$$\bigotimes_{i=1}^{n} K_i \subset K_n \subset \left(\bigotimes_{i=1}^{n} K_i^* \right)^* \,, \tag{13}$$

where the set $\bigotimes_{i=1}^{n} K_i$ is defined as $\bigotimes_{i=1}^{n} K_i := \mathrm{Conv}\{\otimes \rho_i | \rho_i \in K_i\}$. In other words, an n-composite model of a single model is not uniquely determined in GPTs. Therefore, we need to be more careful in GPTs when we consider an asymptotic scenario. However, in the above asymptotic scenario, we only need to consider m-uses of a channel Φ, which is a channel from a classical m-length bit to an m-partite product state $\sigma^{\mathrm{B}}_{x_1 \dots x_m}$. Due to the inclusion relation (13), an m-partite product state can be regarded as a state in any composite model of a single system, and therefore, the map $\Phi^{\otimes n}$ is well-defined even in GPTs. Hence, we can apply the results in the single-shot scenario to the asymptotic scenario.

Now we consider the situation where we encode a message $j \in \Gamma$ to an m-length sequence $x_1 \dots x_m$ by an encoder g_m, that is, $g_m(j) = x_1 \dots x_m$. Here, notice that the size of the set of all messages Γ depends on m and we denote it as $|g_m|$. Also, let us denote the decoding error ϵ_m when the message j appears uniformly at random. Then, similar to the single-shot scenario, we define an ϵ-asymptotic achievable rate as a real number $R \ge 0$ if there exists a sequence of $(m, |g_m|, \epsilon)$-codes satisfying $\liminf_{m\to\infty} \frac{1}{m} \log |g_m| = R$. Finally, we define ϵ-asymptotic classical capacity following [52].

Definition 7 *The ϵ-asymptotic classical capacity of Φ is defined as follows:*

$$\tilde{\mathrm{C}}^{\epsilon}(\Phi) := \sup \{ R \mid R \text{ is } \epsilon\text{-achievable rate for } \Phi \} \,. \tag{14}$$

By definitions of one-shot ϵ-classical capacity and ϵ-classical capacity, we have

$$\tilde{\mathrm{C}}^{\epsilon}(\Phi) = \liminf_{m\to\infty} \frac{1}{m} \mathrm{C}^{\epsilon}(\Phi^{\otimes m}) \,. \tag{15}$$

Then, we show the asymptotic equivalence between the upper and lower bounds of $\tilde{\mathrm{C}}^{\epsilon}(\Phi)$ as the following theorem.

Theorem 8 *In any model of GPTs and any $\epsilon \in (0, 1)$, a channel Φ satisfies*

$$\tilde{\mathrm{C}}^{\epsilon}(\Phi) = \lim_{m\to\infty} \frac{1}{m} \sup_{P_X{}^m} D^{\epsilon}_{\mathrm{H,G}}(\pi^{\mathrm{AB}}_{P_X{}^m} \| \pi^{\mathrm{A}}_{P_X{}^m} \otimes \pi^{\mathrm{B}}_{P_X{}^m}) \,. \tag{16}$$

In other words, the classical capacity and the hypothesis testing relative entropy are asymptotically equivalent even in GPTs.

Here, we remark on the dependence of ϵ. In quantum theory, because of quantum Stein's lemma [20, 37], we have the following inequality [54]:

$$\forall \epsilon \in (0, 1), \quad \lim_{n\to\infty} \frac{1}{n} D^{\epsilon}_{\mathrm{H}}(\rho^{\otimes n} || \sigma^{\otimes n}) = D(\rho || \sigma) \,.$$

In other words, there is no ϵ-dependence in quantum theory, and the rates are equal to Umegaki relative entropy [50, 51], which leads to Holevo–Schumacher–Westmoreland theorem [21, 45] as explained in Ref. [54]. Both of the above problems are still open in GPTs, i.e., (1) Are asymptotic classical capacity and asymptotic hypothesis testing relative entropy independent with ϵ even in GPTs? (2) Are asymptotic classical capacity and asymptotic hypothesis testing relative entropy related to standard relative entropy even in GPTs? The answer to both problems should give an important new operational perspective of entropies and information rates.

Acknowledgments

The authors thank Mark M. Wilde for pointing out errors in the previous version and for helpful comments and discussions. Helpful comments and discussions from Francesco Buscemi are gratefully acknowledged. Comments from Tan Van Vu on the previous version of the manuscript are gratefully acknowledged. S. M. would like to take this opportunity to thank the "Nagoya University Interdisciplinary Frontier Fellowship" supported by Nagoya University and JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2120 and "THERS Make New Standards Program for the Next Generation Researchers" supported by JST SPRING, Grant Number JPMJSP2125.

References

[1] H. Arai, Y. Yoshida, and M. Hayashi. Perfect discrimination of non-orthogonal separable pure states on bipartite system in general probabilistic theory. *J. Phys. A: Math. Theor.*, 52(46):465304, 2019.

[2] G. Aubrun, L. Lami, C. Palazuelos, and M. Plávala. Entanglement and superposition are equivalent concepts in any physical theory. *Phys. Rev. Lett.*, 128:160402, Apr 2022.

[3] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke. Entropy and information causality in general probabilistic theories. *New J. Phys.*, 12(3):033024, mar 2010.

[4] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Cloning and broadcasting in generic probabilistic theories. *arXiv preprint quant-ph/0611295*, 2006.

[5] H. Barnum, M. P. Müller, and C. Ududec. Higher-order interference and single-system postulates characterizing quantum theory. *New J. Phys.*, 16(12):123029, 2014.

[6] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, Mar 2007.

[7] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem. *IEEE transactions on Information Theory*, 48(10):2637–2655, 2002.

[8] F. Buscemi and N. Datta. Distilling entanglement from arbitrary resources. *J. Math. Phys.*, 51(10):102201, 2010.

[9] F. Buscemi and N. Datta. The quantum capacity of channels with arbitrarily correlated noise. *IEEE Trans. Inf. Theory*, 56(3):1447–1460, 2010.

[10] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Probabilistic theories with purification. *Phys. Rev. A*, 81:062348, Jun 2010.

[11] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, Jul 2011.

[12] G. Chiribella and C. M. Scandolo. Entanglement and thermodynamics in general probabilistic theories. *New J. Phys.*, 17(10):103027, 2015.

[13] G. Chiribella and C. M. Scandolo. Operational axioms for diagonalizing states. *EPTCS*, 195:96–115, 2015.

[14] T. M. Cover and J. A. Thomas. *Elements of Information Theory, 2nd edition (Wiley Series in Telecommunications and Signal Processing)*. John Wiley & Sons, New York, 2006.

[15] G. M. D'Ariano, G. Chiribella, and P. Perinotti. *Quantum Theory from First Principles: An Informational Approach*. Cambridge University Press, Cambridge, England, 2017.

[16] E. B. Davies and J. T. Lewis. An operational approach to quantum probability. *Commun. Math. Phys.*, 17(3):239–260, 1970.

[17] F. Dupuis, L. Kraemer, P. Faist, J. M. Renes, and R. Renner. Generalized entropies. In *XVIIth international congress on mathematical physics*, pages 134–153. World Scientific, 2014.

[18] S. Gudder. Convex structures and operational quantum mechanics. *Commun. Math. Phys.*, 29(3):249 – 264, 1973.

[19] L. Hardy. Quantum theory from five reasonable axioms. *arXiv: quant-ph/0101012*, 2001.

[20] F. Hiai and D. Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.*, 143(1):99–114, 1991.

[21] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998.

[22] P. Janotta and H. Hinrichsen. Generalized probability theories: what determines the structure of quantum theory? *J. Phys. A: Math. Theor.*, 47(32):323001, 2014.

[23] P. Janotta and R. Lal. Generalized probabilistic theories without the no-restriction hypothesis. *Phys. Rev. A*, 87:052131, May 2013.

[24] A. Jenčová. Assemblages and steering in general probabilistic theories. *J. Phys. A: Math. Theor.*, 55(43):434001, 2022.

[25] A. Jenčová. Incompatible measurements in a class of general probabilistic theories. *Phys. Rev. A*, 98:012133, Jul 2018.

[26] G. Kimura, J. Ishiguro, and M. Fukui. Entropies in general probabilistic theories and their application to the holevo bound. *Phys. Rev. A*, 94:042113, Oct 2016.

[27] G. Kimura, T. Miyadera, and H. Imai. Optimal state discrimination in general probabilistic theories. *Phys. Rev. A*, 79:062306, Jun 2009.

[28] G. Kimura, K. Nuida, and H. Imai. Distinguishability measures and entropies for general probabilistic theories. *Rep. Math. Phys.*, 66(2):175–206, 2010.

[29] M. Krumm, H. Barnum, J. Barrett, and M. P. Müller. Thermodynamics and the structure of quantum theory. *New J. Phys.*, 19(4):043025, 2017.

[30] G. Ludwig. Versuch einer axiomatischen grundlegung der quantenmechanik und allgemeinerer physikalischer theorien. *Zeitschrift für Physik*, 181(3):233–260, 1964.

[31] G. Ludwig. Attempt of an axiomatic foundation of quantum mechanics and more general theories, ii. *Communications in Mathematical Physics*, 4(5):331–348, 1967.

[32] S. Minagawa and H. Arai. One-shot and asymptotic classical capacity in general physical theories. *Phys. Rev. A*, 109:062416, Jun 2024.

[33] S. Minagawa, H. Arai, and F. Buscemi. Von neumann's information engine without the spectral theorem. *Phys. Rev. Research*, 4:033091, Aug 2022.

[34] M. P. Müller and C. Ududec. Structure of reversible computation determines the self-duality of quantum theory. *Phys. Rev. Lett.*, 108:130401, Mar 2012.

[35] M. P. Müller. Probabilistic theories and reconstructions of quantum theory. *SciPost Phys. Lect. Notes*, page 28, 2021.

[36] K. Nuida, G. Kimura, and T. Miyadera. Optimal observables for minimum-error state discrimination in general probabilistic theories. *J. Math. Phys.*, 51(9):093505, 2010.

[37] T. Ogawa and H. Nagaoka. Strong converse and stein's lemma in quantum hypothesis testing. In *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, pages 28–42. World Scientific, 2005.

[38] M. Ozawa. Optimal measurements for general quantum systems. *Rep. Math. Phys.*, 18(1):11–28, 1980.

[39] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, 2009.

[40] P. Perinotti, A. Tosini, and L. Vaglini. Shannon theory beyond quantum: Information content of a source. *Phys. Rev. A*, 105:052222, May 2022.

[41] P. Perinotti, A. Tosini, and L. Vaglini. Which entropy for general physical theories? *arXiv:2302.01651*, 2023.

[42] M. Plávala. General probabilistic theories: An introduction. *Phys. Rep.*, 1033:1–64, 2023. General probabilistic theories: An introduction.

[43] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, 24(3):379–385, 1994.

[44] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51:2738–2747, Apr 1995.

[45] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, Jul 1997.

[46] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens. Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality. *Phys. Rev. A*, 107:062203, Jun 2023.

[47] F. Shahandeh. Contextuality of general probabilistic theories. *PRX Quantum*, 2:010330, Feb 2021.

[48] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, 1948.

[49] A. J. Short and S. Wehner. Entropy in general physical theories. *New J. Phys.*, 12(3):033023, 2010.

[50] H. Umegaki. On information in operator algebras. *Proc. Japan Acad.*, 37(8):459–461, 1961.

[51] H. Umegaki. Conditional expectation in an operator algebra, iv (entropy and information). *Kodai Math. Sem. Rep.*, 14(2):59–85, 1962.

[52] S. Verdú and T. S. Han. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157, 1994.

[53] E. Wakakuwa. Gentle measurement as a principle of quantum theory. *arXiv preprint arXiv:2103.15110*, 2021.

[54] L. Wang and R. Renner. One-shot classical-quantum capacity and hypothesis testing. *Phys. Rev. Lett.*, 108:200501, May 2012.

[55] J. Watrous. *The theory of quantum information*. Cambridge university press, Cambridge, England, 2018.

[56] M. M. Wilde. *Quantum Information Theory, 2nd edition*. Cambridge University Press, Cambridge, England, 2017.

[57] Y. Yoshida, H. Arai, and M. Hayashi. Perfect discrimination in approximate quantum theory of general probabilistic theories. *Phys. Rev. Lett.*, 125:150402, Oct 2020.

Black box work extraction and composite hypothesis testing

Kaito Watanabe[1] * Ryuji Takagi[1] †

[1] *Department of Basic Science, The University of Tokyo, Tokyo 153-8902, Japan*

Abstract. We introduce a general framework for work extraction, an essential process in quantum thermodynamics, which addresses the inaccessibility of information on the initial state. We show that the optimal extractable work in the black box setting is completely characterized by the performance of a composite hypothesis testing task, a fundamental problem in information theory. We employ this general relation to reduce the asymptotic black box work extraction to the quantum Stein's lemma in composite hypothesis testing, exhibiting the difficulty in this task in comparison to the standard setting. We also show a new quantum Stein's lemma motivated in this physical setting, where a composite hypothesis contains a certain correlation. Our work exhibits the importance of information about the initial state and gives a new interpretation of the quantities in the composite quantum hypothesis testing, encouraging the interplay between the physical settings and the information theory.

Keywords: Quantum thermodynamics, Work extraction, Composite hypothesis testing, Quantum Stein's lemma

1 Overview

One of the major goals in thermodynamics is to characterize the ultimate efficiency of work extraction. In particular, provided the recent technological developments in accurately controlling nanoscale systems, it is of fundamental importance to obtain a precise understanding of the extractable work from small systems where quantum properties are not negligible. Recently, there has been much progress in characterizing extractable work in quantum systems employing quantum information-theoretic approaches [1–4]. These results not only reveal that the optimal single-copy (one-shot) extractable work is represented as the optimal performance of the standard information-theoretic task known as *quantum hypothesis testing*, where one aims to distinguish two quantum states, but also offers a smooth connection to the many-copy (asymptotic, thermodynamic) limit via the well-known result in quantum hypothesis testing called quantum Stein's lemma [5, 6], where the Helmholtz free energy arises as an emergent quantity [7].

Although these results entail fundamental insights into the problem of work extraction, they do not represent natural operational settings. Crucially, the optimal work characterized so far assumes that the description of the initial state is provided, allowing the experimenters to tailor the work extraction protocol depending on the given state. However, in many settings—such as the scenarios where the state is obtained by a complicated quantum process that cannot be efficiently simulated classically, or the state experiences unknown noise—**we are not in possession of the complete information about the initial state**. To run the "state-aware" protocol in these settings, one would first attempt to learn the state description by quantum state tomography [8, 9], which is also hard because of the large number of the required systems which can affect the performance of the task, and the physical limitation inherent in thermodynamic processes. To encompass this large class of "state-agnostic" scenarios, new techniques are required.

Here, we establish the fundamental relation between these two—state-agnostic work extraction and *composite* hypothesis testing, a more general setting of hypothesis testing where one aims to distinguish two sets of states by measuring a state picked from either of the sets. We show that the optimal guaranteed extractable work from a black box can be exactly characterized by the performance of composite hypothesis testing between the black box and thermal Gibbs state. This not only extends the result of state-aware work extraction to much more general and operational settings but also provides the first operational interpretation of composite hypothesis testing in terms of quantum thermodynamics.

We further employ this relation to obtain the asymptotic work extraction rate in the black box setting. Employing the composite quantum Stein's lemmas, which include our new result, we characterize the asymptotic work extraction rate from a general black box with several standard classes of thermodynamic processes [2] and find that it can be smaller than the minimum Helmholtz free energy of the state in the black box. This indicates the fundamental difficulty in the state-agnostic setting compared to the standard setting. Additionally, we show that a similar characterization can be extended to a class of thermodynamic operations amenable to easier physical implementation [1].

Work extraction protocol in quantum thermodynamics is an example of *quantum resource distillation*. Potential and limitations of resource distillation with unknown input states were discussed for some specific cases of entanglement and magic states by different approaches [10–12]. Our results complement these findings, offering a platform that allows mutual developments in state-agnostic resource distillation and composite hypothesis testing, boosting the interplay between physically motivated tasks and information-theoretic problems.

2 Preliminaries

Thermodynamic operations We consider a system associated with a finite-dimensional Hilbert space with some Hamiltonian H, which is in contact with the heat bath whose inverse temperature is β. Here, we employ a resource-theoretic approach to quantum thermodynamics to formalize the set of thermodynamic operations available for work extraction. The idea of quantum resource theory is to consider the set of states that can easily be prepared in the given physical setting (often called *free states*) and the ones preserving

* watanabe715@g.ecc.u-tokyo.ac.jp
† ryujitakagi.pat@gmail.com

the set of free states as accessible operations (often called *free operations*).

In quantum thermodynamics, it is standard to consider the thermal Gibbs state $\tau := \exp(-\beta H)/\mathrm{Tr}[\exp(-\beta H)]$ as the only free state that can be prepared without any cost accessible. Therefore, thermodynamically "free" operations are the ones that preserve the Gibbs state [13]. There are several classes of operations which satisfy this condition.

Composite hypothesis testing The composite hypothesis testing aims to distinguish the set of states, in which the hypotheses $\mathcal{S}$ and $\mathcal{T}$ are subsets of the set $\mathcal{D}(\mathcal{H})$ of all quantum states in $\mathcal{H}$, and one performs the binary POVM $\{M, I-M\}$ and guesses which set the measured state came from. The performance of this task is represented as the quantity called the composite hypothesis testing divergence defined as

$$D_H^\varepsilon(\mathcal{S}||\mathcal{T}) = -\log \inf_{\substack{0\le M\le I \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}[\rho(I-M)]\le\varepsilon}} \sup_{\sigma\in\mathcal{T}} \mathrm{Tr}[\sigma M] \tag{1}$$

When a composite hypothesis $\mathcal{T}$ consists of a single element τ, we simply write τ to represent $\mathcal{T} = \{\tau\}$.

3 Black box work extraction

Framework We now introduce the framework of black box work extraction. Consider a system associated with a finite-dimensional Hilbert space $\mathcal{H}$ and the Hamiltonian H, and another system called a 'battery' associated with a 2-dimensional Hilbert space and its Hamiltonian. We represent the extracted work by the difference between the two energy eigenvalues of the battery system, i.e., we prepare the battery system in the thermal state, and if one can convert this system to the excited state by the free operations and the initial state, we say that we can extract work.

We represent the inaccessibility to the information of the given state as a black box, a subset $\mathcal{S} \subset \mathcal{D}(\mathcal{H})$ of states acting on $\mathcal{H}$. The experimenters are informed about the description of $\mathcal{S}$ and that the initial state is an element of the black box $\mathcal{S}$ but are not told which state is actually given, preventing them from tailoring work extraction protocols depending on the initial state.

The problem is to find the maximum energy gap of the battery system which can be charged for every choice of the state from the black box and the allowed operation which is independent of the initial state. We denote the one-shot extractable work of the black box $\mathcal{S} \subset \mathcal{D}(\mathcal{H})$ under allowed operations $\mathbb{O}$ permitting some error ε with respect to the fidelity as $\beta W_{\mathbb{O}}^\varepsilon(\mathcal{S})$. Note that we define the extractable work as the maximum value of the work drawn regardless of the states picked from the black box; in other words, it is the worst-case extractable work with respect to the states in the black box.

One can also consider the asymptotic limit of the extractable work from the black box by considering a sequence of the black boxes. Consider the situation where there are n systems with the same Hamiltonians H. To take the limit $n \to \infty$, we consider a family $\{\mathcal{S}_n\}_{n=1}^\infty$ of black boxes with $\mathcal{S}_n \subset \mathcal{D}(\mathcal{H}^{\otimes n})$. We define the asymptotic black box extractable work as follows.

$$\beta W_{\mathbb{O}}(\{\mathcal{S}_n\}_{n=1}^\infty) := \lim_{\varepsilon\to+0} \limsup_{n\to\infty} \frac{1}{n} \beta W_{\mathbb{O}}^\varepsilon(\mathcal{S}_n). \tag{2}$$

Namely, the asymptotic black box extractable work of the sequence of the black boxes is the work drawn from the whole system per the number of subsystems.

Unless stated otherwise, in the following discussion we focus on the family with a tensor-product structure $\mathcal{S}_n(S) := \left\{\bigotimes_{i=1}^n \rho_i \mid \rho_i \in S\ \forall i\right\}$ generated by an arbitrary set $S \subset \mathcal{D}(\mathcal{H})$.

Black box work extraction with Gibbs-preserving operations We are in the position to characterize the performance of black box work extraction. We first consider Gibbs-preserving operations, the largest class of operations which include all the operations which map the thermal state of the input state to the thermal state of the output state, as available thermodynamic processes. The following result provides the general characterization of one-shot extractable work in terms of composite hypothesis testing divergence.

Theorem 1. *One-shot extractable work from an arbitrary black box $\mathcal{S}$ under Gibbs-preserving operations satisfy*

$$\beta W_{\mathrm{GPO}}^\varepsilon(\mathcal{S}) = D_H^\varepsilon(\mathcal{S}||\tau). \tag{3}$$

Theorem 1 establishes a tight connection between the composite hypothesis testing and the work extraction task and provides a physical meaning of the composite hypothesis testing divergence in the context of thermodynamics.

We remark that if two black boxes $\mathcal{S}$ and $\mathcal{T}$ satisfy $\mathcal{S} \subset \mathcal{T}$, due to the definition of the composite hypothesis testing divergence, it holds that $\beta W_{\mathrm{GPO}}^\varepsilon(\mathcal{S}) \ge \beta W_{\mathrm{GPO}}^\varepsilon(\mathcal{T})$, which means that the more detailed information about the initial state increases the extractable work.

Let us now extend this to asymptotic work extraction via the composite hypothesis version of the quantum Stein's lemma, a central question in information theory which investigates whether hypothesis testing divergence connects to the standard relative entropy, the quantity which corresponds to the Helmholtz free energy in the context of quantum thermodynamics.

Our general characterization in Theorem 1 allows us to obtain the following simple expression for the asymptotic work extraction.

Theorem 2. *The asymptotic black box extractable work of the sequence $\{\mathcal{S}_n(S)\}_{n=1}^\infty$ of the black boxes under Gibbs-preserving operations is given by*

$$\beta W_{\mathrm{GPO}}(\{\mathcal{S}_n(S)\}_{n=1}^\infty) = \lim_{n\to\infty} \frac{1}{n} \min_{\rho_n\in C(\mathcal{S}_n(S))} D(\rho_n||\tau^{\otimes n}). \tag{4}$$

where $C(\mathcal{S}_n(S))$ denotes the convex hull of the set $\mathcal{S}_n(S)$.

Theorem 2 clarifies the fundamental restriction imposed by not knowing the input state. To see this, consider $S = \{|\phi_1\rangle\langle\phi_1|, \ldots, |\phi_d\rangle\langle\phi_d|\}$, where $|\phi_1\rangle, \ldots, |\phi_d\rangle$ are the eigenstates of the Hamiltonian of a single subsystem H. If we have information about the initial state, we can extract the

nonzero work from it since the free energy of any state in the black box is strictly larger than that of the thermal states. However, for any positive integer n, the thermal state τ is included in $C(\mathcal{S}_n)$, which implies that one cannot extract any work from this sequence of the black boxes asymptotically. This example reveals the underlying difference between the standard state-aware work extraction task and the black box work extraction task.

Black box work extraction under Gibbs-preserving covariant operations Although Gibbs-preserving operations admit relatively simple mathematical analysis, there is also doubt in its operational justification. Notably, they can create quantum coherence from scratch [14], and some Gibbs-preserving operations require even unbounded quantum coherence to implement [15]. This motivates us to impose additional constraints that operations should be time-translation covariant as the thermal operations—which prohibits the creation and detection of quantum coherence—and this is precisely the class of Gibbs-preserving covariant operations, which includes all the Gibbs-preserving, time-tanslationally covariant operations.

We show that the one-shot extractable work with this class of operations is also related to the composite hypothesis testing divergence as

$$\beta W^{\varepsilon}_{\mathrm{GPC}}(\mathcal{S}) = D^{\varepsilon}_{H}(\mathcal{P}(\mathcal{S})||\tau), \tag{5}$$

where $\mathcal{P}(\cdot) \coloneqq \sum_{E_i} \Pi_{E_i}(\cdot)\Pi_{E_i}$ is the pinching operator with respect to the Hamiltonian of the whole system, and $\mathcal{P}(\mathcal{S})$ is the pinched black box. Here, Π_{E_i} is the projector onto the eigenspace of the Hamiltonian of the whole system corresponding to the eigenvalue E_i.

We would also like to understand the behavior in the asymptotic limit via quantum Stein's lemma, as we did in the case of Gibbs-preserving operations. However, the structure of the composite hypothesis is more involved in this case because of the correlation between different subsystems generated by the pinching channel. Namely, $\mathcal{P}(\otimes_i \rho_i)$ is not a product state in general. Nevertheless, we show that quantum Stein's lemma holds even in such a case.

Lemma 3. *For an arbitrary set S of states,*

$$\begin{aligned}&\lim_{\varepsilon\to+0}\lim_{n\to\infty}\frac{1}{n}D^{\varepsilon}_{H}(\mathcal{P}(\mathcal{S}_n(S))||\tau^{\otimes n})\\ &\quad=\lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)}D(\rho_n||\tau^{\otimes n})\end{aligned} \tag{6}$$

Let us remark on the relation between Lemma 3 and the generalized quantum Stein's lemma [16, 17]. Recent studies have revealed that the relation between (state-aware) resource distillation and composite hypothesis testing holds at the high level of generality [18, 19]. The major open question along this line, when trying to characterize the asymptotic state-aware resource distillation, is the generalized quantum Stein's lemma, whose difficulty rests on the correlation between the subsystems in the family of composite hypotheses. In this sense, Lemma 3, which involves correlation in the composite hypothesis, might be found useful in this context.

In the setting of black box work extraction, Lemma 3 is precisely the one that brings one-shot result (Eq. (5)) to the asymptotic setting, which is characterized as follows.

Theorem 4. *The asymptotic black box extractable work of the sequence of the black boxes $\{\mathcal{S}_n(S)\}_{n=1}^{\infty}$ under Gibbs-preserving covariant operations is given by*

$$\beta W_{\mathrm{GPC}}(\{\mathcal{S}_n(S)\}_{n=1}^{\infty}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n(S))}D(\rho_n||\tau^{\otimes n}), \tag{7}$$

Theorem 4 shows that although Gibbs-preserving covariant operations come with restrictions compared to Gibbs-preserving operations in one-shot level (as can be seen in Theorem 1 and Eq. (5)), their performance coincides in the asymptotic limit, both of which are characterized by the standard free energy. This result, therefore, extends the similar observation in the standard state-aware work extraction [7], in which the work extraction rate also agrees in the asymptotic limit.

Black box work extraction under thermal operations Since Gibbs-preserving operations and Gibbs-preserving covariant operations are axiomatic classes of the operations, they do not always reflect the physical implementability [15]. This motivates us to study thermal operations [1], which is an operationally well-motivated class of thermodynamic processes. Here, we focus on i.i.d. black boxes of the form $\mathcal{S}_n^{\mathrm{i.i.d.}}(S) \coloneqq \{\rho^{\otimes n} \mid \rho \in S\}$ generated by a set S of finite size, i.e., $|S| < \infty$.

In the technical manuscript, we devise an explicit work extraction protocol with *covariantly conditioned thermal operations*, which we newly introduce. We then show that this class coincides with thermal operations in the work extraction scenario, resulting in the following characterization.

Theorem 5. *The asymptotic black box extractable work of $\{\mathcal{S}_n^{\mathrm{i.i.d.}}(S)\}_n$ satisfying $|S| < \infty$ under thermal operations is represented as*

$$\beta W_{\mathrm{TO}}\left(\{\mathcal{S}^{\mathrm{i.i.d.}}(S)\}_{n=1}^{\infty}\right) = \min_{\rho\in S} D(\rho||\tau). \tag{8}$$

In Ref. [7], it is shown that the extractable work of the known i.i.d. state is equal to the quantum relative entropy under any of the three free operations mentioned in the discussion above. Our result indicates that the same holds true in the i.i.d. black box case. If there exists a sequence of the black box with which the work extraction rate differs between the closure of thermal operations and the Gibbs-preserving covariant operations, it would imply that the operational capabilities in state transformation of these two sets are distinct, resolving an important open problem in the field [13, 20].

4 Discussion

Our work clarifies when and how the lack of information about the initial state crucially affects the work extraction performance and what one can still do under such restricted scenarios. As our framework forms a new connection between the work extraction tasks in quantum thermodynamics and the quantities in the composite quantum hypothesis testing, the black box work extraction offers a richer landscape in general quantum resource theories, complementing and extending the state-aware asymptotic distillation tied to generalized quantum Stein's lemma.

References

[1] Micha l Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature communications*, 4(1):2059, 2013.

[2] Philippe Faist and Renato Renner. Fundamental work cost of quantum processes. *Physical Review X*, 8(2), apr 2018.

[3] Fernando Brandão, Micha l Horodecki, Nelly Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[4] Fernando G. S. L. Brandão, Micha l Horodecki, Jonathan Oppenheim, Joseph M. Renes, and Robert W. Spekkens. Resource theory of quantum states out of thermal equilibrium. *Phys. Rev. Lett.*, 111:250404, Dec 2013.

[5] Fumio Hiai and Dénes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in mathematical physics*, 143:99–114, 1991.

[6] T. Ogawa and H. Nagaoka. Strong converse and stein's lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory*, 46(7):2428–2433, 2000.

[7] Gilad Gour. Role of quantum coherence in thermodynamics. *PRX Quantum*, 3(4), nov 2022.

[8] K. Vogel and H. Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A*, 40:2847–2849, Sep 1989.

[9] K Banaszek, M Cramer, and D Gross. Focus on quantum tomography. *New Journal of Physics*, 15(12):125020, dec 2013.

[10] Keiji Matsumoto and Masahito Hayashi. Universal distortion-free entanglement concentration. *Phys. Rev. A*, 75:062338, Jun 2007.

[11] Scott Aaronson, Adam Bouland, Bill Fefferman, Soumik Ghosh, Umesh Vazirani, Chenyi Zhang, and Zixin Zhou. Quantum pseudoentanglement, 2023.

[12] Andi Gu, Lorenzo Leone, Soumik Ghosh, Jens Eisert, Susanne Yelin, and Yihui Quek. A little magic means a lot, 2023.

[13] Matteo Lostaglio. An introductory review of the resource theory approach to thermodynamics. *Rep. Prog. Phys.*, 82(11):114001, oct 2019.

[14] Philippe Faist, Jonathan Oppenheim, and Renato Renner. Gibbs-preserving maps outperform thermal operations in the quantum regime. *New J. Phys.*, 17(4):043003, 2015.

[15] Hiroyasu Tajima and Ryuji Takagi. Gibbs-preserving operations requiring infinite amount of quantum coherence, 2024.

[16] Fernando G. S. L. Brandão and Martin B. Plenio. A generalization of quantum stein's lemma. *Communications in Mathematical Physics*, 295(3):791–828, feb 2010.

[17] Mario Berta, Fernando G. S. L. Brandão, Gilad Gour, Ludovico Lami, Martin B. Plenio, Bartosz Regula, and Marco Tomamichel. On a gap in the proof of the generalised quantum Stein's lemma and its consequences for the reversibility of quantum resources. *Quantum*, 7:1103, September 2023.

[18] Zi-Wen Liu, Kaifeng Bu, and Ryuji Takagi. One-shot operational quantum resource theory. *Phys. Rev. Lett.*, 123:020401, Jul 2019.

[19] Bartosz Regula and Ryuji Takagi. One-shot manipulation of dynamical quantum resources. *Phys. Rev. Lett.*, 127:060402, Aug 2021.

[20] Piotr Ćwikliński, Micha l Studziński, Micha l Horodecki, and Jonathan Oppenheim. Limitations on the evolution of quantum coherences: Towards fully quantum second laws of thermodynamics. *Phys. Rev. Lett.*, 115:210403, Nov 2015.

Black box work extraction and composite hypothesis testing

Kaito Watanabe* and Ryuji Takagi†
Department of Basic Science, The University of Tokyo, Tokyo 153-8902, Japan

Work extraction is one of the most central processes in quantum thermodynamics. However, the prior analysis of optimal extractable work has been restricted to a limited operational scenario where complete information about the initial state is given. Here, we introduce a general framework of black box work extraction, which addresses the inaccessibility of information on the initial state. We show that the optimal extractable work in the black box setting is completely characterized by the performance of a composite hypothesis testing task, a fundamental problem in information theory. We employ this general relation to reduce the asymptotic black box work extraction to the quantum Stein's lemma in composite hypothesis testing, allowing us to provide their exact characterization in terms of the Helmholtz free energy. We also show a new quantum Stein's lemma motivated in this physical setting, where a composite hypothesis contains a certain correlation. Our work exhibits the importance of information about the initial state and gives a new interpretation of the quantities in the composite quantum hypothesis testing, encouraging the interplay between the physical settings and the information theory.

Introduction.— One of the major goals in thermodynamics is to characterize the ultimate efficiency of work extraction. In particular, provided the recent technological developments in accurately controlling nanoscale systems, it is of fundamental importance to obtain a precise understanding of the extractable work from small systems where quantum properties are not negligible. Recently, there has been much progress in characterizing extractable work in quantum systems employing quantum information-theoretic approaches [1–6]. These results not only provide an explicit form of the optimal single-copy (one-shot) extractable work, but also offer a smooth connection to the many-copy (asymptotic, thermodynamic) limit, where the Helmholtz free energy arises as an emergent quantity [4, 6].

Although these results entail fundamental insights into the problem of work extraction, they do not represent natural operational settings. Crucially, the optimal work characterized so far assumes that the description of the initial state is provided, allowing the experimenters to tailor the work extraction protocol depending on the given state. However, in many settings—such as the scenarios where the state is obtained by a complicated quantum process that cannot be efficiently simulated classically, or the state experiences unknown noise—we are not in possession of the complete information about the initial state. To run the "state-aware" protocol in these settings, one would first attempt to learn the state description by quantum state tomography [7, 8]. At this point, the characterization of the prior results becomes unclear because (1) state tomography requires multiple (indeed, many) copies of the initial state and thus could significantly change the effective performance of work extraction and (2) the full state tomography may not be possible due to the physical limitation inherent in thermodynamic processes. To encompass this large class of "state-agnostic" scenarios, new techniques are required.

A major observation from a series of works is that work extraction is closely related to the standard information-theoretic task known as *hypothesis testing*, where one aims to distinguish two quantum states. These entirely different-looking operational tasks turn out to be quantitatively connected via their performances—the maximum amount of work extractable from a single copy of the given known state is precisely characterized by *hypothesis-testing divergence* [6, 9, 10]—the standard quantifier for the asymmetric state discrimination—between the initial (known) state and the thermal Gibbs state [5].

Interestingly, hypothesis testing has been extended to a more general setting—instead of distinguishing two states, one aims to distinguish two *sets* of states by measuring a state picked from either of the sets. This task is known as *composite hypothesis testing* and has been an active investigation in classical [11] and quantum [12–14] information theory. In particular, there has been a rising interest in quantum Stein's lemma [15, 16], which connects composite hypothesis testing divergence to the optimized relative entropy in the asymptotic composite hypothesis testing setting. Nevertheless, unlike the case of standard hypothesis testing, the operational significance of composite hypothesis testing in the context of quantum thermodynamics has been unclear.

Here, we establish the fundamental relation between these two—state-agnostic work extraction and composite hypothesis testing. We introduce a general framework for state-agnostic work extraction by considering a "black box", from which a state is picked and an experimenter—who knows what states are contained in the box but does not know which state was actually picked—applies a work extraction protocol. We show that the optimal guaranteed extractable work from a black box can be exactly characterized by the performance of composite hypothesis testing between the black box and thermal Gibbs state. This not only extends the result of state-aware work extraction to much more general and operational settings, but also provides the first operational interpretation of composite hypothesis testing in terms of quantum thermodynamics.

We further employ this relation to obtain the asymptotic work extraction rate in the black box setting. Notably, we prove a new kind of Stein's lemma for composite hypothesis testing, where state copies from the composite hypothesis have a correlation generated by a pinching channel [15, 17, 18]. This—together with the general connection between black box work extraction and composite hypothesis testing—shows that the asymptotic work extraction rate from a black box with several standard classes of thermodynamic processes [2] and find that it can be smaller than the minimum Helmholtz free energy of the

* watanabe715@g.ecc.u-tokyo.ac.jp
† ryujitakagi.pat@gmail.com

state in the black box. This indicates the fundamental difficulty in the state-agnostic setting compared to the standard setting. Additionally, we show that a similar characterization can be extended to a class of thermodynamic operations amenable to easier physical implementation [1].

Although the main focus here is work extraction in quantum thermodynamics, our framework can readily be extended to other quantum resource theories, such as quantum entanglement [19], magic states [20, 21], and even a general class including those [22]. Indeed, work extraction protocol in quantum thermodynamics is an example of *quantum resource distillation*. Potential and limitations of resource distillation with unknown input states were discussed for some specific cases of entanglement and magic states by different approaches [23–25]. Our results complement these findings, offering a platform that allows mutual developments in state-agnostic resource distillation and composite hypothesis testing, boosting the interplay between physically motivated tasks and information-theoretic problems.

Thermodynamic operations.— We consider a system associated with a finite-dimensional Hilbert space with some Hamiltonian H, which is in contact with the heat bath whose inverse temperature is β. Here, we employ a resource-theoretic approach to quantum thermodynamics to formalize the set of thermodynamic operations available for work extraction. The idea of quantum resource theory is to consider the set of states that can easily be prepared in the given physical setting (often called *free states*) and the ones preserving the set of free states as accessible operations (often called *free operations*).

In quantum thermodynamics, it is standard to consider the thermal Gibbs state $\tau := \exp(-\beta H)/\mathrm{Tr}[\exp(-\beta H)]$ as the only free state that can be prepared without any cost accessible. Therefore, thermodynamically "free" operations are the ones that preserve the Gibbs state [5]. However, this constraint does not single out the unique set of free operations, and indeed, several classes of operations have been investigated depending on the goal of the study. The largest class that satisfies the minimum requirement is the *Gibbs-preserving operations* [2, 26], which include all operations that map the thermal state of the input system to that of the output system. This class is mainly studied due to its simple mathematical structure, which led to a number of recent key progress in quantum thermodynamics [2, 10, 27–32]. On the other hand, the class that respects the physical implementability is known as *thermal operations*, in which appending the thermal state as an ancillary system, applying the energy-conserving unitary, and tracing out any subsystems are considered to be free, i.e., a completely positive trace-preserving (CPTP) map $\mathcal{E}$ from systems A to B is called a thermal operation if $\mathcal{E}$ can be written as $\mathcal{E}(\cdot) = \mathrm{Tr}_{(A+E)\setminus B}\left[U(\cdot \otimes \tau_E)U^\dagger\right]$, where $\tau_E := \exp(-\beta H_E)/Z_E$ is a thermal state of the ancillary system, and U is a energy-conserving unitary satisfying $[U, H_A \otimes I_E + I_A \otimes H_E] = 0$.

It is easily checked that the thermal operation maps the thermal state τ_A to the thermal state τ_B, which implies that the thermal operation is Gibbs-preserving. Another important property of the thermal operation is the time translation covariance, i.e., any thermal operation $\mathcal{E}$ from systems A to B satisfies

$$\mathcal{E}\left(e^{-iH_A t}\rho_A e^{iH_A t}\right) = e^{-iH_B t}\mathcal{E}(\rho_A)e^{iH_B t}, \quad \forall t \in \mathbb{R}. \tag{1}$$

This property prohibits the operation from creating energetic coherence—which serves another important thermodynamic resource [6, 33–36]—from scratch. One can find a Gibbs-preserving operation that does not satisfy this property by constructing a map that is Gibbs-preserving and creates coherence from the incoherent state [37] or detects coherence [36]. The class of operations that is mathematically easy to handle and closer to thermal operations is called *Gibbs-preserving covariant operations*, which are Gibbs-preserving and time-translation covariant. In Ref. [38], it was shown that in the qubit case, the set of thermal operations and the set of Gibbs-preserving covariant operations are the same, but in the qutrit case, there exists an operation that can be done in Gibbs-preserving covariant operations but not in thermal operations. Note that these three classes of operations satisfy the following relation.

$$\text{Thermal} \subset \text{Gibbs-preserving covariant} \subsetneq \text{Gibbs-preserving}. \tag{2}$$

Whether the closure of the set of thermal operations and the set of Gibbs-preserving covariant operations agree is an open problem.

Composite hypothesis testing.— The standard quantum hypothesis testing aims to distinguish two different states ρ and σ called a null hypothesis and an alternative hypothesis respectively, with a binary measurement with the POVM elements $\{M, I - M\}$ in which the outcome corresponding to M means that one guesses the given state is ρ, and the outcome that corresponds to $I - M$ means that one guesses the state is σ. The performance of the distinguishing task is rephrased as how much one can minimize the probability of making mistakes in the guess. In this work, we mainly focus on the minimum probability of guessing σ as ρ known as type II error, $\mathrm{Tr}[\sigma M]$ under the constraint in which type I error $\mathrm{Tr}[\rho(I - M)]$—the probability of guessing ρ as σ—is at most ε. The performance of this task is represented as the quantity called the hypothesis testing divergence defined as [9]

$$D_H^\varepsilon(\rho||\sigma) = -\log \inf_{\substack{0\le M\le I \\ \mathrm{Tr}[\rho(I-M)]\le\varepsilon}} \mathrm{Tr}[\sigma M]. \tag{3}$$

This framework can be extended to the composite case, where the hypotheses are the set of states [12, 14]. In the composite hypothesis testing, the hypotheses $\mathcal{S}$ and $\mathcal{T}$ are subsets of the set $\mathcal{D}(\mathcal{H})$ of all quantum states in $\mathcal{H}$, and one performs the binary POVM and guesses which set the measured state came from. The hypothesis testing divergence can also be extended to the composite case by focusing on the worst-case error probability.

$$D_H^\varepsilon(\mathcal{S}||\mathcal{T}) = -\log \inf_{\substack{0\le M\le I \\ \sup_{\rho\in\mathcal{S}}\mathrm{Tr}[\rho(I-M)]\le\varepsilon}} \sup_{\sigma\in\mathcal{T}} \mathrm{Tr}[\sigma M] \tag{4}$$

Note that this quantity does not change even if we take the convex hull of either the composite hypothesis.

Black box work extraction.— We now introduce the framework of black box work extraction. Consider a system associated with a finite-dimensional Hilbert space $\mathcal{H}$ and the Hamiltonian H, and another system called a 'battery' associated with a 2-dimensional Hilbert space $\mathcal{H}_X = \mathrm{Span}\{|0\rangle, |1\rangle\}$. We take the Hamiltonian for the battery system as $H_X = E_{X,0}|0\rangle\langle 0| + E_{X,1}|1\rangle\langle 1|$ with $E_{X,1} - E_{X,0} = \beta^{-1}\log(m-1)$ so that the thermal state of the battery system μ_m is $\mu_m = \frac{m-1}{m}|0\rangle\langle 0| + \frac{1}{m}|1\rangle\langle 1|$ for $m \geq 1$. If an allowed operation can transform an initial state and the equilibrium state μ_m in the battery system to the state $|1\rangle\langle 1|_X$ of the battery, we say that we can "charge" the battery.

We represent the inaccessibility to the information of the given state as a black box, a subset $\mathcal{S} \subset \mathcal{D}(\mathcal{H})$ of states acting on $\mathcal{H}$. The experimenters are informed about the description of $\mathcal{S}$ and that the initial state is an element of the black box $\mathcal{S}$ but are not told which state is actually given, preventing them from tailoring work extraction protocols depending on the initial state.

The problem is to find the maximum m such that the battery with the thermal state μ_m can be charged with the unknown initial state picked up from the black box and allowed operations $\mathbb{O}$, i.e., to find the largest m such that $\rho \otimes \mu_m \xrightarrow{\mathbb{O}} |1\rangle\langle 1|_X$ is possible for every choice of the state from the black box and the allowed operation which is independent of the initial state $\rho \in \mathcal{S}$. Here, we formulate the optimal performance of the black box work extraction.

Definition 1. The one-shot extractable work of the black box $\mathcal{S} \subset \mathcal{D}(\mathcal{H})$ with error ε is defined as

$$\begin{aligned}&\beta W^{\varepsilon}_{\mathbb{O}}(\mathcal{S})\\ &:= \log \max\left\{ m \in \mathbb{R} \;\middle|\; F_{\mathbb{O}}((\mathcal{S},\tau),(|1\rangle\langle 1|_X,\mu_m)) \geq 1-\varepsilon \right\},\end{aligned} \tag{5}$$

where $F_{\mathbb{O}}$ is the conversion fidelity defined as

$$\begin{aligned}&F_{\mathbb{O}}((\mathcal{S},\tau),(|1\rangle\langle 1|_X,\mu_m)) = \max_{\mathcal{E}\in\mathbb{O}} \min_{\rho\in\mathcal{S}} F(\mathcal{E}(\rho),|1\rangle\langle 1|_X),\\ &F(\rho,\sigma) := \left(\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1\right)^2 = \left(\mathrm{Tr}\left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right]\right)^2.\end{aligned} \tag{6}$$

Note that we define the extractable work as the maximum value of the work drawn regardless of the states picked from the black box, in other words, the worst-case extractable work with respect to the states in the black box. For the justification of this definition, see Appendix A. Also, note that when we take $\mathcal{S} = \{\rho\}$, the problem is reduced to the ordinary work extraction setting.

One can also consider the asymptotic limit of the extractable work from the black box by considering a sequence of the black boxes. Consider the situation where there are n systems with the same Hamiltonians H. Note that the Hamiltonian of the whole system is represented as $H^{\times n} := H\otimes I\otimes\cdots\otimes I + I\otimes H\otimes\cdots\otimes I + \cdots + I\otimes I\otimes\cdots\otimes H$. To take the limit $n\to\infty$, we consider a family $\{\mathcal{S}_n\}_{n=1}^{\infty}$ of black boxes with $\mathcal{S}_n \subset \mathcal{D}(\mathcal{H}^{\otimes n})$. We define the asymptotic black box extractable work as follows.

Definition 2. The asymptotic black box extractable work of the sequence $\{\mathcal{S}_n\}_{n=1}^{\infty}$ of the black boxes is

$$\beta W_{\mathbb{O}}(\{\mathcal{S}_n\}_{n=1}^{\infty}) := \lim_{\varepsilon\to+0}\limsup_{n\to\infty}\frac{1}{n}\beta W^{\varepsilon}_{\mathbb{O}}(\mathcal{S}_n). \tag{7}$$

Namely, the asymptotic black box extractable work of the sequence of the black boxes is the work drawn from the whole system per the number of subsystems.

Unless stated otherwise, in the following discussion we focus on the family with tensor-product structure

$$\mathcal{S}_n(S) := \left\{ \bigotimes_{i=1}^{n}\rho_i \mid \rho_i \in S\ \forall i \right\} \tag{8}$$

generated by an arbitrary set $S \subset \mathcal{D}(\mathcal{H})$.

Black box work extraction with Gibbs-preserving operations.— We are in the position to characterize the performance of black box work extraction. We first consider Gibbs-preserving operations as available thermodynamic processes. The following result provides the general characterization of one-shot extractable work in terms of composite hypothesis testing divergence (Proof in Appendix B 1).

Theorem 3. *One-shot extractable work from an arbitrary black box $\mathcal{S}$ under Gibbs-preserving operations satisfy*

$$\beta W^{\varepsilon}_{\mathrm{GPO}}(\mathcal{S}) = D^{\varepsilon}_{H}(\mathcal{S}||\tau). \tag{9}$$

Theorem 3 establishes a tight connection between the composite hypothesis testing and the work extraction task and provides a physical meaning of the composite hypothesis testing divergence in the context of thermodynamics. We stress that Theorem 3 holds for an arbitrary black box $\mathcal{S}$, which may be non-convex and could contain an uncountably infinite number of states. In the case of a singleton set $\mathcal{S} = \{\rho\}$, our result recovers the known result for state-aware work extraction [6, 10]. We also remark that if two black boxes $\mathcal{S}$ and $\mathcal{T}$ satisfy $\mathcal{S}\subset\mathcal{T}$, due to the definition of the composite hypothesis testing divergence, it holds that $\beta W^{\varepsilon}_{\mathrm{GPO}}(\mathcal{S}) \geq \beta W^{\varepsilon}_{\mathrm{GPO}}(\mathcal{T})$, which means that the more detailed information about the initial state increases the extractable work.

Let us now extend this to asymptotic work extraction. Theorem 3 allows us to focus our attention on analyzing how the composite hypothesis testing divergence behaves under the asymptotic limit. This is a central question in information theory known as Stein's lemma, which investigates whether hypothesis testing divergence connects to the standard relative entropy. This comes with a further physical significance in the context of quantum thermodynamics because relative entropy precisely corresponds to the free energy playing a central role in the second law of thermodynamics.

When a composite hypothesis is involved, it is typically a formidable task to establish Stein's lemma. Nevertheless, previous works found that there are several settings in which Stein's lemma can be established [11, 13, 14, 39–41]. In particular, an extension of quantum Sanov's theorem [40], together with our general characterization in Theorem 3, implies the following simple expression for the asymptotic work extraction.

Theorem 4. *The asymptotic black box extractable work of the sequence $\{\mathcal{S}_n(S)\}_{n=1}^{\infty}$ of the black boxes under Gibbs-preserving operations is given by*

$$\beta W_{\mathrm{GPO}}(\{\mathcal{S}_n(S)\}_{n=1}^{\infty}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n(S))} D(\rho_n||\tau^{\otimes n}), \tag{10}$$

where $C(\mathcal{S}_n(S))$ is the convex hull of $\mathcal{S}_n(S)$. Furthermore, the convex hull can be removed if every element in the black box is permutationally invariant.

In Appendix C 1, we prove a more general result that implies Theorem 4 by employing the recent result in Ref. [14]. We remark that if the sequence of the black boxes is composed of i.i.d. states, i.e., $\rho_i = \rho_j\ \forall i, j$ in (8), the right-hand side of Eq. (10) is reduced to $\min_{\rho\in S} D(\rho||\tau)$, which can also be seen as a consequence of the quantum Sanov's theorem [39].

Theorem 4 clarifies the fundamental restriction imposed by not knowing the input state. To see this, consider $S = \{|\phi_1\rangle\langle\phi_1|, \ldots, |\phi_d\rangle\langle\phi_d|\}$, where $|\phi_1\rangle, \ldots, |\phi_d\rangle$ are the eigenstates of the Hamiltonian of a single subsystem H. If we have information about the initial state, we can extract the nonzero work from it since the free energy of any state in the black box is strictly larger than that of the thermal states. However, for any positive integer n, the thermal state τ is included in $C(\mathcal{S}_n)$, which implies that one cannot extract any work from this sequence of the black boxes asymptotically. This example reveals the underlying difference between the standard state-aware work extraction task and the black box work extraction task.

Black box work extraction under Gibbs-preserving covariant operations.— Although Gibbs-preserving operations admit relatively simple mathematical analysis, there is also doubt in its operational justification. Notably, they can create quantum coherence from scratch [37], and some Gibbs-preserving operations require even unbounded quantum coherence to implement [36]. This motivates us to impose additional constraints described in (1) that operations should be time-translation covariant—which prohibits the creation and detection of quantum coherence—and this is precisely the class of Gibbs-preserving covariant operations.

The time-translation covariant condition, which restricts an operation to utilize the time information, naturally introduces a channel that acts as the time average. This is a special form of *pinching channels* often employed as an important analytical tool in information theory [15, 17, 18]. In our setting, the relevant pinching channel is the one with respect to the Hamiltonian of the whole system defined as

$$\mathcal{P}(\cdot) = \lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T} dt e^{-iHt}\cdot e^{iHt} = \sum_{E_i}\Pi_{E_i}(\cdot)\Pi_{E_i}, \tag{11}$$

where Π_{E_i} is the projector onto the eigenspace of the Hamiltonian of the whole system corresponding to the eigenvalue E_i. We then define the pinched black box

$$\mathcal{P}(\mathcal{S}) \coloneqq \left\{\mathcal{P}(\rho)\ \middle|\ \rho\in\mathcal{S}\right\}. \tag{12}$$

The following result shows that the black box work extraction with the time-tranlation covariant condition can be characterized by the composite hypothesis divergence for a pinched black box. (Proof in Appendix B 2.)

Theorem 5. *One-shot extractable work from an arbitrary black box $\mathcal{S}$ under Gibbs-preserving covariant operations satisfy*

$$\beta W_{\mathrm{GPC}}^{\varepsilon}(\mathcal{S}) = D_H^{\varepsilon}(\mathcal{P}(\mathcal{S})||\tau). \tag{13}$$

We would also like to understand the behavior in the asymptotic limit via quantum Stein's lemma, as we did in the case of Gibbs-preserving operations. However, the structure of the composite hypothesis is more involved in this case because of the correlation between different subsystems generated by the pinching channel. Namely, $\mathcal{P}(\otimes_i\rho_i)$ is not a product state in general. This prevents us from directly applying the prior results on composite quantum Stein's lemma [11, 13, 14]. Indeed, when correlation is present in a composite hypothesis, Stein's lemma can become extremely difficult to handle [12, 42]. Nevertheless, we show that quantum Stein's lemma holds in our setting.

Lemma 6. *For an arbitrary set S of states,*

$$\begin{aligned}&\lim_{\varepsilon\to+0}\lim_{n\to\infty}\frac{1}{n}D_H^{\varepsilon}(\mathcal{P}(\mathcal{S}_n(S))||\tau^{\otimes n})\\ &\quad=\lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)} D(\rho_n||\tau^{\otimes n})\end{aligned} \tag{14}$$

In Appendix C 2, we prove a slightly more general result that includes Lemma 6. We remark that non-composite version of this was previously shown in Ref. [43].

Let us remark on the relation between Lemma 6 and the generalized quantum Stein's lemma. Recent studies have revealed that the relation between (state-aware) resource distillation and hypothesis testing holds at the high level of generality [31, 44]. In fact, this is generally characterized by composite hypothesis testing divergence—but in a different way where the *second* argument of the divergence is a composite hypothesis (while our black box setting contains a composite hypothesis in the *first* argument). The major open question along this line, when trying to characterize the asymptotic state-aware resource distillation, is the generalized quantum Stein's lemma. The difficulty of the generalized quantum Stein's lemma rests on the fact that the family of composite hypotheses generally has a correlation between different subsystems. Therefore, more insights into the asymptotic behavior of composite hypotheses with correlation will be helpful. In this sense, Lemma 6, which involves correlation in the composite hypothesis, might be found useful in this context, although it does not appear to directly contribute to the resolution of the problem at the moment.

In the setting of black box work extraction, Lemma 6 is precisely the one that brings one-shot result (Theorem 5) to the asymptotic setting, which is characterized as follows.

Theorem 7. *The asymptotic black box extractable work of the sequence of the black boxes $\{\mathcal{S}_n(S)\}_{n=1}^{\infty}$ under Gibbs-preserving covariant operations is given by*

$$\beta W_{\mathrm{GPC}}(\{\mathcal{S}_n(S)\}_{n=1}^{\infty}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n(S))} D(\rho_n||\tau^{\otimes n}), \tag{15}$$

where $C(\mathcal{S}_n(S))$ is the convex hull of $\mathcal{S}_n(S)$. Furthermore, the convex hull can be removed if every element in the black box is permutationally invariant.

Theorem 7 shows that although Gibbs-preserving covariant operations come with restrictions compared to Gibbs-preserving operations in one-shot level (as can be seen in Theorems 3 and 5), their performance coincides in the asymptotic limit, both of which are characterized by the standard free energy. This result, therefore, extends the similar observation in the standard state-aware work extraction [6], in which the work extraction rate also agrees in the asymptotic limit.

Asymptotic black box work extraction under thermal operations.— Since the Gibbs-preserving operations and Gibbs-preserving covariant operations are axiomatic classes of the operations, they do not always reflect the physical implementability [36]. This motivates us to study thermal operations, which is an operationally well-motivated class of thermodynamic processes [1]. Here, we focus on i.i.d. black boxes of the form $\mathcal{S}_n^{\mathrm{i.i.d.}}(S) := \{\rho^{\otimes n} \mid \rho \in S\}$ generated by a set S of finite size, i.e., $|S| < \infty$. In the standard state-aware setting, the work extraction from i.i.d. state is discussed in Ref. [4], which constructed a protocol that extracts work whose rate asymptotically converges to $D(\rho||\tau)$, where ρ is the known initial state.

Toward characterizing the asymptotic black box work extraction with thermal operations, we first introduce a new class of thermodynamic processes, which contains thermal operations.

Definition 8. Let $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C$ be Hilbert spaces, and $\mathcal{E} : \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_C)$ be a CPTP map. We call $\mathcal{E}$ a *covariantly conditioned thermal operation* if $\mathcal{E}$ has the form

$$\mathcal{E} = \sum_a \mathcal{E}_a^{\mathrm{TO}} \circ \Lambda_a^{\mathrm{meas}}. \tag{16}$$

Here, each $\mathcal{E}_a^{\mathrm{TO}} : \mathcal{D}(\mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_C)$ is a thermal operation and

$$\Lambda_a^{\mathrm{meas}}(\rho_{ABC}) := \mathrm{Tr}_A \left[(M_a^{\mathrm{cov}} \otimes I_{BC}) \rho_{ABC} (M_a^{\mathrm{cov}} \otimes I_{BC})^\dagger \right] \tag{17}$$

is an instrument representing a covariant measurement, where M_a^{cov} is a POVM element satisfying $\mathcal{P}(M_a^{\mathrm{cov}}) = M_a^{\mathrm{cov}}$.

We remark that covariantly conditioned thermal operations clearly contain thermal operations, while this is a subset of the class called conditioned thermal operations introduced in Ref. [45], in which all measurements are allowed to be performed.

The following result shows that covariantly conditioned thermal operations perform as well as Gibbs-preserving operations and Gibbs-preserving covariant operations in the asymptotic setting.

Proposition 9. *The asymptotic black box extractable work of $\left\{\mathcal{S}_n^{\mathrm{i.i.d.}}(S)\right\}_n$ satisfying $|S| < \infty$ under covariantly conditioned thermal operations is given by*

$$\beta W_{\mathrm{CCTO}}\left(\left\{\mathcal{S}^{\mathrm{i.i.d.}}(S)\right\}_{n=1}^{\infty}\right) = \min_{\rho \in S} D(\rho||\tau). \tag{18}$$

Here, we sketch the main idea of the protocol to achieve this rate, while we defer a detailed proof in Appendix D 1. Our strategy is to first learn the given state using some copies and run the state-aware protocol by Ref. [4]. An apparent restriction here is that we are only allowed to use covariant measurement, which may not give us the full information about the given state. For example, a covariant measurement can never distinguish $|+\rangle\langle+|$ from $|-\rangle\langle-|$, even when one has infinitely many copies. Nevertheless, we observe that the protocol in Ref. [4] does not require the full information about the initial state ρ—what only matters is $\mathcal{P}(\rho^{\otimes n})$ for extracting work from n copies.

However, this raises another potential issue. Since $\{\mathcal{P}(\rho^{\otimes n})\}_n$ is not an i.i.d series anymore, it is not clear whether one could learn the description of $\mathcal{P}(\rho^{\otimes n})$ for an arbitrary large n—which is generally required for asymptotic work extraction—only using the sublinear number of copies so that it would not affect the work extraction rate. We show that it is indeed possible. We prove that the structure of the energy blocks ensures that it suffices to identify pinched d states $\mathcal{P}(\rho^{\otimes d})$ by performing the quantum state tomography with the covariant measurement. Since the number of the systems for this tomography does not depend on n, the cost for the tomography becomes negligible asymptotically.

We now extend Proposition 9 to thermal operations. To this end, we show that in the situation where the dimensions of the input and output systems are the same and the measurements are projective, covariantly conditioned thermal operations coincide with the thermal operations. By showing that our work extraction protocol can be modified to satisfy these conditions, we obtain the following result (Proof in Appendix D 2).

Theorem 10. *The asymptotic black box extractable work of $\left\{\mathcal{S}_n^{\mathrm{i.i.d.}}(S)\right\}_n$ satisfying $|S| < \infty$ under thermal operations is represented as*

$$\beta W_{\mathrm{TO}}\left(\left\{\mathcal{S}^{\mathrm{i.i.d.}}(S)\right\}_{n=1}^{\infty}\right) = \min_{\rho \in S} D(\rho||\tau). \tag{19}$$

In Ref. [6], it is shown that the extractable work of the known i.i.d. state is equal to the quantum relative entropy under any of the three free operations mentioned in the discussion above. Our result indicates that the same holds true in the i.i.d. black box case. Whether this holds true in the more general setting is not known. If there exists a sequence of the black box with which the work extraction rate differs between the closure of thermal operations and the Gibbs-preserving covariant operations, it would imply that the operational capabilities in state transformation of these two sets are distinct, resolving an important open problem in the field [5, 38]. We leave a further investigation along this line as a future work.

Conclusion.— We introduced a framework of black box work extraction, which represents the scenarios where one is to extract work from an unknown quantum state. We presented the optimal guaranteed extractable work in various settings by establishing the connection between one-shot black box work extraction and composite hypothesis testing. We utilized this general relation to characterize the asymptotic work distillation rate by employing and extending quantum Stein's lemma for composite hypothesis testing. Besides composite hypothesis

testing, we also devised an explicit protocol for asymptotic black box work extraction for physically motivated classes of thermodynamic processes, which is shown to perform as well as much larger classes of operations.

Our work clarifies when and how the lack of information about the initial state crucially affects the work extraction performance and what one can still do under such restricted scenarios. The state-agnostic setting discussed in this work has not been investigated well despite its operational significance and still has much room to explore. Potential future directions include an extension of our results to a more general family of black boxes without a tensor product structure. Another important extension is to other quantum resource theories beyond one-shot distillation with the maximal set of free operations discussed in this work. As our framework forms a new connection between the resource distillation tasks in the quantum resource theory and the quantities in the composite quantum hypothesis testing, the black box resource distillation offers a richer landscape in general quantum resource theories, complementing and extending the state-aware asymptotic distillation tied to generalized quantum Stein's lemma.

Acknowledgments.— We thank Gilad Gour, Bartosz Regula, Christoph Hirsch, and Seth Lloyd for helpful discussions. This work is supported by JSPS KAKENHI Grant Number JP23K19028, JP24K16975, JST, CREST Grant Number JPMJCR23I3, Japan, and MEXT KAKENHI Grant-in-Aid for Transformative Research Areas A "Extreme Universe" Grant Number JP24H00943.

[1] M. Horodecki and J. Oppenheim, *Fundamental limitations for quantum and nanoscale thermodynamics,* Nature communications **4**, 2059 (2013).
[2] P. Faist and R. Renner, *Fundamental work cost of quantum processes,* Physical Review X **8** (2018), 10.1103/physrevx.8.021011.
[3] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, *The second laws of quantum thermodynamics,* Proceedings of the National Academy of Sciences **112**, 3275 (2015), https://www.pnas.org/doi/pdf/10.1073/pnas.1411728112.
[4] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, *Resource theory of quantum states out of thermal equilibrium,* Phys. Rev. Lett. **111**, 250404 (2013).
[5] M. Lostaglio, *An introductory review of the resource theory approach to thermodynamics,* Rep. Prog. Phys. **82**, 114001 (2019).
[6] G. Gour, *Role of quantum coherence in thermodynamics,* PRX Quantum **3** (2022), 10.1103/prxquantum.3.040323.
[7] K. Vogel and H. Risken, *Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase,* Phys. Rev. A **40**, 2847 (1989).
[8] K. Banaszek, M. Cramer, and D. Gross, *Focus on quantum tomography,* New Journal of Physics **15**, 125020 (2013).
[9] L. Wang and R. Renner, *One-shot classical-quantum capacity and hypothesis testing,* Phys. Rev. Lett. **108**, 200501 (2012).
[10] X. Wang and M. M. Wilde, *Resource theory of asymmetric distinguishability,* Phys. Rev. Res. **1**, 033170 (2019).
[11] F. G. S. L. Brandão, A. W. Harrow, J. R. Lee, and Y. Peres, *Adversarial hypothesis testing and a quantum stein's lemma for restricted measurements,* IEEE Transactions on Information Theory **66**, 5037 (2020).
[12] F. G. S. L. Brandão and M. B. Plenio, *A generalization of quantum stein's lemma,* Communications in Mathematical Physics **295**, 791 (2010).
[13] M. Berta, F. G. S. L. Brandão, and C. Hirche, *On composite quantum hypothesis testing,* Communications in Mathematical Physics **385**, 55–77 (2021).
[14] B. Bergh, N. Datta, and R. Salzmann, *Composite Classical and Quantum Channel Discrimination,* arXiv e-prints , arXiv:2303.02016 (2023), arXiv:2303.02016.
[15] F. Hiai and D. Petz, *The proper formula for relative entropy and its asymptotics in quantum probability,* Communications in mathematical physics **143**, 99 (1991).
[16] T. Ogawa and H. Nagaoka, *Strong converse and stein's lemma in quantum hypothesis testing,* IEEE Transactions on Information Theory **46**, 2428 (2000).
[17] M. Hayashi, *Optimal sequence of quantum measurements in the sense of stein's lemma in quantum hypothesis testing,* Journal of Physics A: Mathematical and General **35**, 10759 (2002).
[18] M. Tomamichel, *Quantum Information Processing with Finite Resources* (Springer International Publishing, 2016).
[19] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Quantum entanglement,* Rev. Mod. Phys. **81**, 865 (2009).
[20] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, *The resource theory of stabilizer quantum computation,* New Journal of Physics **16**, 013009 (2014).
[21] M. Howard and E. Campbell, *Application of a resource theory for magic states to fault-tolerant quantum computing,* Phys. Rev. Lett. **118**, 090501 (2017).
[22] E. Chitambar and G. Gour, *Quantum resource theories,* Rev. Mod. Phys. **91**, 025001 (2019).
[23] K. Matsumoto and M. Hayashi, *Universal distortion-free entanglement concentration,* Phys. Rev. A **75**, 062338 (2007).
[24] S. Aaronson, A. Bouland, B. Fefferman, S. Ghosh, U. Vazirani, C. Zhang, and Z. Zhou, *Quantum pseudoentanglement,* (2023), arXiv:2211.00747 [quant-ph].
[25] A. Gu, L. Leone, S. Ghosh, J. Eisert, S. Yelin, and Y. Quek, *A little magic means a lot,* (2023), arXiv:2308.16228 [quant-ph].
[26] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, *Thermodynamic Cost of Reliability and Low Temperatures: Tightening Landauer's Principle and the Second Law,* Int. J. Theor. Phys. **39**, 2717 (2000).
[27] P. Faist, M. Berta, and F. Brandão, *Thermodynamic capacity of quantum processes,* Phys. Rev. Lett. **122**, 200601 (2019).
[28] N. Shiraishi and T. Sagawa, *Quantum thermodynamics of correlated-catalytic state conversion at small scale,* Phys. Rev. Lett. **126**, 150502 (2021).
[29] F. Buscemi, D. Sutter, and M. Tomamichel, *An information-theoretic treatment of quantum dichotomies,* Quantum **3**, 209 (2019).
[30] T. Sagawa, P. Faist, K. Kato, K. Matsumoto, H. Nagaoka, and F. G. S. L. Brandão, *Asymptotic reversibility of thermal operations for interacting quantum spin systems via generalized quantum stein's lemma,* Journal of Physics A: Mathematical and Theoretical **54**, 495303 (2021).
[31] Z.-W. Liu, K. Bu, and R. Takagi, *One-shot operational quantum resource theory,* Phys. Rev. Lett. **123**, 020401 (2019).
[32] B. Regula, K. Bu, R. Takagi, and Z.-W. Liu, *Benchmarking one-shot distillation in general quantum resource theories,* Phys. Rev. A **101**, 062315 (2020).

[33] R. Takagi and N. Shiraishi, *Correlation in catalysts enables arbitrary manipulation of quantum coherence,* Phys. Rev. Lett. **128**, 240501 (2022).

[34] N. Shiraishi and R. Takagi, *Arbitrary amplification of quantum coherence in asymptotic and catalytic transformation,* Phys. Rev. Lett. **132**, 180202 (2024).

[35] H. Tajima, R. Takagi, and Y. Kuramochi, *Universal trade-off structure between symmetry, irreversibility, and quantum coherence in quantum processes,* (2022), arXiv:2206.11086 [quant-ph].

[36] H. Tajima and R. Takagi, *Gibbs-preserving operations requiring infinite amount of quantum coherence,* (2024), arXiv:2404.03479 [quant-ph].

[37] P. Faist, J. Oppenheim, and R. Renner, *Gibbs-preserving maps outperform thermal operations in the quantum regime,* New J. Phys. **17**, 043003 (2015).

[38] P. Ćwikliński, M. Studziński, M. Horodecki, and J. Oppenheim, *Limitations on the evolution of quantum coherences: Towards fully quantum second laws of thermodynamics,* Phys. Rev. Lett. **115**, 210403 (2015).

[39] I. Bjelaković, J.-D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze, and A. Szko la, *A Quantum Version of Sanov's Theorem,* Commun. Math. Phys. **260**, 659 (2005).

[40] J. Nötzel, *Hypothesis testing on invariant subspaces of the symmetric group: part i. quantum sanov's theorem and arbitrarily varying sources,* J. Phys. A: Math. Theor. **47**, 235303 (2014).

[41] M. Mosonyi, Z. Szilágyi, and M. Weiner, *On the error exponents of binary state discrimination with composite hypotheses,* IEEE Trans. Inf. Theory **68**, 1032 (2022).

[42] M. Berta, F. G. S. L. Brandão, G. Gour, L. Lami, M. B. Plenio, B. Regula, and M. Tomamichel, *On a gap in the proof of the generalised quantum Stein's lemma and its consequences for the reversibility of quantum resources,* Quantum **7**, 1103 (2023), arXiv:2205.02813 [quant-ph].

[43] P. Lipka-Bartosik, C. T. Chubb, J. M. Renes, M. Tomamichel, and K. Korzekwa, *Quantum dichotomies and coherent thermodynamics beyond first-order asymptotics,* PRX Quantum **5**, 020335 (2024).

[44] B. Regula and R. Takagi, *One-shot manipulation of dynamical quantum resources,* Phys. Rev. Lett. **127**, 060402 (2021).

[45] V. Narasimhachar and G. Gour, *Resource theory under conditioned thermal operations,* Phys. Rev. A **95**, 012313 (2017).

[46] M. Sion, *On general minimax theorems.* (1958).

[47] G. Gour, *Resources of the quantum world,* (2024), arXiv:2402.05474 [quant-ph].

[48] M. Piani, *Relative entropy of entanglement and restricted measurements,* Physical Review Letters **103** (2009), 10.1103/physrevlett.103.160504.

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley-Interscience, USA, 2006).

[50] R. O'Donnell and J. Wright, *Efficient quantum tomography,* in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16 (Association for Computing Machinery, New York, NY, USA, 2016) p. 899–912.

[51] M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph, *Quantum coherence, time-translation symmetry, and thermodynamics,* Phys. Rev. X **5**, 021001 (2015).

Appendix A: Justification of the definition of the extractable work

The definition of extractable work in this letter is a little bit anomalous. In this section, we show that the definition is equivalent to the standard definition of work extraction.

Suppose that we have the charged state of the battery system $|1\rangle\langle 1|$, and the thermal state of the battery system is $\mu_m = 1/m\,|1\rangle\langle 1| + (m-1)/m\,|0\rangle\langle 0|$. Note that the charged state is incoherent, i.e., does not include any superposition of the energy eigenstates of different energy. In [1], the amount of work is expressed as the energy gaps between the two energy eigenvalues of the Hamiltonian of the two-dimensional system called work storage W. In such a setting, it is shown that the necessary amount of work to obtain $|1\rangle\langle 1|$ from the thermal state with thermal operations is [1]

$$\beta W_{\text{formation}}(|1\rangle\langle 1|) \geq D_{\max}(|1\rangle\langle 1|\,||\mu_m) = \log m, \tag{A1}$$

where $D_{\max}$ is the max-divergence defined as

$$D_{\max}(\rho||\sigma) = \log \min\{\lambda \mid \rho \leq \lambda\sigma\}. \tag{A2}$$

On the other hand, the work extracted from $|1\rangle\langle 1|$ with thermal operations is [1]

$$\beta W_{\text{extractable}} \leq D_{\min}(|1\rangle\langle 1|\,||\mu_m) = \log m, \tag{A3}$$

where $D_{\min}$ is the min-divergence defined as

$$D_{\min}(\rho||\sigma) = -\log \text{Tr}\left[\Pi_{\text{supp}(\rho)}\sigma\right] \quad (\Pi_{\text{supp}(\rho)} \text{ is the projector onto supp}(\rho).). \tag{A4}$$

These inequalities also hold when one can perform the Gibbs-preserving operations [2]. In our setting, the work necessary for the formation of the charged state and extractable work from the charged system is the same, and this is the reason why we consider this quantity.

Appendix B: One-shot black box work extraction

1. One-shot black box work extraction under Gibbs-preserving operations (Proof of Theorem 3)

Theorem S.1 (Theorem 3 in the main text)**.** *Let $\mathcal{S} \subset \mathcal{D}(\mathcal{H})$ be the black box. The one-shot extractable work under the Gibbs-preserving operations is represented as*

$$\beta W^{\varepsilon}_{\text{GPO}}(\mathcal{S}) = D^{\varepsilon}_{H}(\mathcal{S}||\tau). \tag{B1}$$

Proof. We first show the achievable part $\beta W_{\text{GPO}}(\mathcal{S}) \geq D^{\varepsilon}_{H}(\mathcal{S}||\tau)$. To show this, it suffices to show that some Gibbs-preserving operation achieves this extractable work yield. Consider the CPTP map $\mathcal{E} : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H}_X)$ which has the following form.

$$\mathcal{E}(\rho) = \text{Tr}[M\rho]\,|1\rangle\langle 1|_X + \text{Tr}[(I-M)\rho]\,|0\rangle\langle 0|_X\,, \quad 0 \leq M \leq I \tag{B2}$$

This map is Gibbs-preserving if and only if this map satisfies $\mathcal{E}(\tau) = \mu_m$, i.e.,

$$\begin{aligned} &\mathcal{E}(\tau) = \text{Tr}[M\tau]\,|1\rangle\langle 1|_X + \text{Tr}[(I-M)\tau]\,|0\rangle\langle 0|_X = \frac{1}{m}\,|1\rangle\langle 1|_X + \frac{m-1}{m}\,|0\rangle\langle 0|_X\,, \\ &\Leftrightarrow m = (\text{Tr}\,[M\tau])^{-1}. \end{aligned} \tag{B3}$$

If we take M such that M satisfies $\text{Tr}\,[M\rho] \geq 1-\varepsilon$ for every $\rho \in \mathcal{S}$, we can see that for any $\rho \in \mathcal{S}$

$$\begin{aligned} F(\mathcal{E}(\rho), |1\rangle\langle 1|_X) &= F(\text{Tr}[M\rho]\,|1\rangle\langle 1|_X + \text{Tr}[(I-M)\rho]\,|0\rangle\langle 0|_X\,, |1\rangle\langle 1|_X) \\ &\geq \text{Tr}[M\rho]F(|1\rangle\langle 1|_X\,, |1\rangle\langle 1|_X) + \text{Tr}\,[(I-M)\rho]F(|0\rangle\langle 0|_X\,, |1\rangle\langle 1|) \\ &\geq \text{Tr}[M\rho] \geq 1-\varepsilon. \end{aligned} \tag{B4}$$

In the second line, we used the concavity of the fidelity. Recalling the definition, the one-shot blackbox extractable work is calculated as

$$\begin{aligned} \beta W^{\varepsilon}_{\text{GPO}}(\mathcal{S}) &= \log \max\left\{m \in \mathbb{R} \mid \max_{\mathcal{E}\in\mathbb{O}} \min_{\rho\in\mathcal{S}} F(\mathcal{E}(\rho), |1\rangle\langle 1|_X) \geq 1-\varepsilon\right\} \\ &\geq \log \max\left\{(\text{Tr}\,[M\tau])^{-1} \mid \forall\rho \in \mathcal{S},\ \text{Tr}\,[M\rho] \geq 1-\varepsilon,\ 0 \leq M \leq I\right\}. \end{aligned} \tag{B5}$$

The last line is the composite hypothesis testing divergence $D_H^\varepsilon(\mathcal{S}||\tau)$. Therefore, we obtain $\beta W_{\mathrm{GPO}}^\varepsilon(\mathcal{S}) \geq D_H^\varepsilon(\mathcal{S}||\tau)$.

To show the converse part $\beta W_{\mathrm{GPO}}(\mathcal{S}) \leq D_H^\varepsilon(\mathcal{S}||\tau)$, we start by showing the data processing inequality of the composite hypothesis testing divergence, i.e., for any CPTP map $\mathcal{E} : \mathcal{D}(\mathcal{H}) \to \mathcal{D}(\mathcal{H}')$ and any composite hypotheses $\mathcal{S}, \mathcal{T} \subset \mathcal{D}(\mathcal{H})$,

$$D_H^\varepsilon(\mathcal{E}(\mathcal{S})||\mathcal{E}(\mathcal{T})) \leq D_H^\varepsilon(\mathcal{S}||\mathcal{T}) \tag{B6}$$

holds. Recalling the definition of the composite hypothesis testing divergence, $D_H^\varepsilon(\mathcal{E}(\mathcal{S})||\mathcal{E}(\mathcal{T}))$ can be written as

$$D_H^\varepsilon(\mathcal{E}(\mathcal{S})||\mathcal{E}(\mathcal{T})) = -\log \inf_{\substack{0\leq M'\leq I_{\mathcal{H}'} \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}[(I-M')\mathcal{E}(\rho)]\leq\varepsilon}} \sup_{\tau\in\mathcal{T}} \mathrm{Tr}[\mathcal{E}(\tau)M']. \tag{B7}$$

Here, we denote the conjugate of $\mathcal{E}$ as $\mathcal{E}^\dagger$, i.e., $\mathcal{E}^\dagger$ satisfies $\mathrm{Tr}[\mathcal{E}(A)B] = \mathrm{Tr}\left[A\mathcal{E}^\dagger(B)\right]$ for any matrices $A \in \mathcal{L}(\mathcal{H}_A)$ and $B \in \mathcal{L}(\mathcal{H}_B)$. Since $\mathcal{E}$ is a CPTP map, $\mathcal{E}^\dagger$ is a CP unital map, which maps $I_{\mathcal{H}'}$ to $I_{\mathcal{H}}$. This can be rewritten as

$$-\log \inf_{\substack{0\leq M'\leq I_{\mathcal{H}'} \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}[(I-M')\mathcal{E}(\rho)]\leq\varepsilon}} \sup_{\tau\in\mathcal{T}} \mathrm{Tr}[\mathcal{E}(\tau)M'] = -\log \inf_{\substack{0\leq M'\leq I_{\mathcal{H}'} \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}\left[(I-\mathcal{E}^\dagger(M'))\rho\right]\leq\varepsilon}} \sup_{\tau\in\mathcal{T}} \mathrm{Tr}\left[\tau\mathcal{E}^\dagger(M')\right]. \tag{B8}$$

Here, we used the definition and the unitality of $\mathcal{E}^\dagger$. Here, we can easily check that $\mathcal{E}^\dagger(M')$ satisfies $0 \leq \mathcal{E}^\dagger(M') \leq I_{\mathcal{H}}$, which follows from the completely positivity of $\mathcal{E}^\dagger$. From this, the following holds.

$$-\log \inf_{\substack{0\leq M'\leq I_{\mathcal{H}'} \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}\left[(I-\mathcal{E}^\dagger(M'))\rho\right]\leq\varepsilon}} \sup_{\tau\in\mathcal{T}} \mathrm{Tr}\left[\tau\mathcal{E}^\dagger(M')\right] \leq -\log \inf_{\substack{0\leq M\leq I_{\mathcal{H}} \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}[(I-M)\rho]\leq\varepsilon}} \sup_{\tau\in\mathcal{T}} \mathrm{Tr}[\tau M] = D_H^\varepsilon(\mathcal{S}||\mathcal{T}). \tag{B9}$$

Combining these, we obtain the data processing inequality of the composite hypothesis testing divergence.

If we take the composite alternative hypothesis $\mathcal{T}$ as $\mathcal{T} = \{\tau\}$, the situation is reduced to our original setting. Here, Let $\mathcal{E}^*$ be the Gibbs-preserving operation which achieves the optimal work extraction, and $m^* = 2^{\beta W_{\mathrm{GPO}}^\varepsilon(\mathcal{S})}$ be the optimal m. From the data processing inequality of the composite hypothesis testing divergence,

$$\begin{aligned} D_H^\varepsilon(\mathcal{S}||\tau) &\geq D_H^\varepsilon(\mathcal{E}^*(\mathcal{S})||\mu_{m^*}) \\ &= -\log \inf_{\substack{0\leq M\leq I \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}[(I-M)\mathcal{E}^*(\rho)]\leq\varepsilon}} \mathrm{Tr}[\mu_{m^*}M] \end{aligned} \tag{B10}$$

holds. Recalling that $\mathcal{E}^*$ satisfies $F(|1\rangle\langle 1|, \mathcal{E}^*(\rho)) = \mathrm{Tr}[|1\rangle\langle 1|\,\mathcal{E}^*(\rho)] \geq 1-\varepsilon,\ \forall\rho \in \mathcal{S}$ because of the definition of the extractable work, we can substitute $M = |1\rangle\langle 1|$ and obtain the following.

$$-\log \inf_{\substack{0\leq M\leq I \\ \sup_{\rho\in\mathcal{S}} \mathrm{Tr}[(I-M)\mathcal{E}^*(\rho)]\leq\varepsilon}} \mathrm{Tr}[\mu_m M] \geq -\log \mathrm{Tr}\left[\mu_{m^*}|1\rangle\langle 1|\right] = \log m^* = \beta W_{\mathrm{GPO}}^\varepsilon(\mathcal{S}). \tag{B11}$$

From these, we obtain the converse part. □

Since the ε- one shot extractable work of the state ρ is obtained as $D_H^\varepsilon(\rho||\tau)$([6]), we can show the direct part in a different way.

Proof. (Alternative proof for the direct part.) As one can see, if one takes the convex hull on the black box, the conversion fidelity decreases, i.e.,

$$\max_{\mathcal{E}\in\mathrm{GPO}} \min_{\rho\in\mathcal{S}} F(\mathcal{E}(\rho), |1\rangle\langle 1|_X) \geq \max_{\mathcal{E}\in\mathrm{GPO}} \min_{\rho\in C(\mathcal{S})} F(\mathcal{E}(\rho), |1\rangle\langle 1|_X). \tag{B12}$$

From this, one can see that

$$\beta W_{\mathrm{GPO}}^\varepsilon(\mathcal{S}) \geq \beta W_{\mathrm{GPO}}^\varepsilon(C(\mathcal{S})). \tag{B13}$$

In the following discussion, we focus on the RHS. $F(\mathcal{E}(\rho), |1\rangle\langle 1|_X) = \mathrm{Tr}\left[\mathcal{E}(\rho)\,|1\rangle\langle 1|_X\right]$ is linear with respect to the $\mathcal{E}$ and ρ. What is more, the set of the Gibbs-preserving operations and $C(\mathcal{S})$ are both convex. Since $\mathcal{S}$ is

closed, the convex hull $C(\mathcal{S})$ is also closed, we can see that $C(\mathcal{S})$ is bounded. From these, we can apply Sion's minimax theorem [46] as follows.

$$\max_{\mathcal{E}\in\mathrm{GPO}} \min_{\rho\in C(\mathcal{S})} F(\mathcal{E}(\rho), |1\rangle\langle 1|_X) = \min_{\rho\in C(\mathcal{S})} \max_{\mathcal{E}\in\mathrm{GPO}} F(\mathcal{E}(\rho), |1\rangle\langle 1|_X) \tag{B14}$$

Therefore, the $\beta W^{\varepsilon}_{\mathrm{GPO}}(C(\mathcal{S}))$ is reduced to the following expression.

$$\begin{aligned}\beta W^{\varepsilon}_{\mathrm{GPO}}(C(\mathcal{S})) &= \log \max\left\{ m\in\mathbb{R} \mid \min_{\rho\in C(\mathcal{S})} \max_{\mathcal{E}\in\mathrm{GPO}} F(\mathcal{E}(\rho), |1\rangle\langle 1|_X) \geq 1-\varepsilon \right\} \\ &= \min_{\rho\in C(\mathcal{S})} \beta W^{\varepsilon}_{\mathrm{GPO}}(\rho)\end{aligned} \tag{B15}$$

From the result in [6, 10], $\beta W^{\varepsilon}_{\mathrm{GPO}}(\rho) = D^{\varepsilon}_{H}(\rho||\tau)$, and the above can be rewritten as

$$\beta W^{\varepsilon}_{\mathrm{GPO}}(C(\mathcal{S})) = \min_{\rho\in C(\mathcal{S})} D^{\varepsilon}_{H}(\rho||\tau) \geq D^{\varepsilon}_{H}(C(\mathcal{S})||\tau). \tag{B16}$$

The last inequality follows due to [14, Lemma 16]. By the definition, the composite hypothesis testing divergence does not change when the convex hull is removed. Combining these discussions, we obtain

$$\beta W^{\varepsilon}_{\mathrm{GPO}}(\mathcal{S}) \geq \beta W^{\varepsilon}_{\mathrm{GPO}}(C(\mathcal{S})) \geq D^{\varepsilon}_{H}(C(\mathcal{S})||\tau) = D^{\varepsilon}_{H}(\mathcal{S}||\tau). \tag{B17}$$

□

While Sion's minimax is a very powerful tool in our setting, it is not the appropriate tool in the subsequent discussion in which we discuss the asymptotic limit of the extractable work from the black box.

2. One-shot black box work extraction under Gibbs-preserving covariant operations (Proof of Theorem 5)

To obtain the one-shot black box extractable work under Gibbs-preserving covariant operations, we show the following lemma without proof. For details, see [47, Exercise 3.5.20].

Lemma S.2. *Let $\mathcal{E} : \mathcal{D}(\mathcal{H}_A) \to \mathcal{D}(\mathcal{H}_B)$ be a time-translation covariant operation, i.e., $\mathcal{E}$ satisfies*

$$\mathcal{E}(e^{-iH_At}\rho e^{iH_At}) = e^{-iH_Bt}\mathcal{E}(\rho)e^{iH_Bt}, \quad \forall t\in\mathbb{R}, \tag{B18}$$

where H_A, H_B are the Hamiltonians of the input system A and the output system B respectively. Then,

$$\mathcal{P}_B \circ \mathcal{E} = \mathcal{E} \circ \mathcal{P}_A \tag{B19}$$

holds, where $\mathcal{P}_A, \mathcal{P}_B$ are the pinching channels with respect to the Hamiltonian H_A and H_B respectively.

From the lemma above, we can see that the conversion fidelity under the Gibbs-preserving covariant operations is connected to that under the Gibbs-preserving operations.

Lemma S.3.

$$F_{\mathrm{GPC}}((\mathcal{S},\tau) \to (|1\rangle\langle 1|, \mu_m)) = F_{\mathrm{GPO}}((\mathcal{P}(\mathcal{S}),\tau) \to (|1\rangle\langle 1|, \mu_m)), \tag{B20}$$

where $\mathcal{P}(\mathcal{S})$ is the set of pinched states of $\mathcal{S}$

$$\mathcal{P}(\mathcal{S}) := \left\{ \mathcal{P}(\rho) \;\middle|\; \rho\in\mathcal{S} \right\}, \tag{B21}$$

and $\mathcal{P}$ is a pinching map with respect to the Hamiltonian of the whole system.

Proof. The idea of the proof is taken from [6]. First, we show the ($\leq$) inequality. From the definition of the conversion fidelity,

$$F_{\mathrm{GPC}}((\mathcal{S},\tau) \to (|1\rangle\langle 1|, \mu_m)) = \max_{\mathcal{E}\in\mathrm{GPC}} \min_{\rho\in\mathcal{S}} F(\mathcal{E}(\rho), |1\rangle\langle 1|) \tag{B22}$$

holds. Due to the contractility of the fidelity,

$$\begin{aligned}\max_{\mathcal{E}\in\mathrm{GPC}}\min_{\rho\in\mathcal{S}} F(\mathcal{E}(\rho),|1\rangle\langle 1|) &\leq \max_{\mathcal{E}\in\mathrm{GPC}}\min_{\rho\in\mathcal{S}} F(\mathcal{P}\circ\mathcal{E}(\rho),\mathcal{P}(|1\rangle\langle 1|))\\ &= \max_{\mathcal{E}\in\mathrm{GPC}}\min_{\rho\in\mathcal{S}} F(\mathcal{P}\circ\mathcal{E}(\rho),|1\rangle\langle 1|)\end{aligned} \tag{B23}$$

holds. Here, from Lemma S.2,

$$\max_{\mathcal{E}\in\mathrm{GPC}}\min_{\rho\in\mathcal{S}} F(\mathcal{P}\circ\mathcal{E}(\rho),|1\rangle\langle 1|) = \max_{\mathcal{E}\in\mathrm{GPC}}\min_{\rho\in\mathcal{S}} F(\mathcal{E}\circ\mathcal{P}(\rho),|1\rangle\langle 1|). \tag{B24}$$

Finally, noting that Gibbs-preserving covariant operations are Gibbs-preserving,

$$\begin{aligned}\max_{\mathcal{E}\in\mathrm{GPC}}\min_{\rho\in\mathcal{S}} F(\mathcal{E}\circ\mathcal{P}(\rho),|1\rangle\langle 1|) &\leq \max_{\mathcal{E}\in\mathrm{GPO}}\min_{\rho\in\mathcal{S}} F(\mathcal{E}\circ\mathcal{P}(\rho),|1\rangle\langle 1|)\\ &= F_{\mathrm{GPO}}((\mathcal{P}(\mathcal{S}),\tau)\to(|1\rangle\langle 1|,\mu_m))\end{aligned} \tag{B25}$$

holds. From these, $F_{\mathrm{GPC}}((\mathcal{S},\tau)\to(|1\rangle\langle 1|,\mu_m))\leq F_{\mathrm{GPO}}((\mathcal{P}(\mathcal{S}),\tau)\to(|1\rangle\langle 1|,\mu_m))$ is shown.

Next, we show the opposite inequality. From the definition, we can see

$$\begin{aligned}F_{\mathrm{GPO}}((\mathcal{P}(\mathcal{S}),\tau)\to(|1\rangle\langle 1|,\mu_m)) &= \max_{\mathcal{E}\in\mathrm{GPO}}\min_{\rho\in\mathcal{S}} F(\mathcal{E}\circ\mathcal{P}(\rho),|1\rangle\langle 1|)\\ &\leq \max_{\mathcal{E}\in\mathrm{GPO}}\min_{\rho\in\mathcal{S}} F(\mathcal{P}\circ\mathcal{E}\circ\mathcal{P}(\rho),|1\rangle\langle 1|).\end{aligned} \tag{B26}$$

Again, we used the contractility of fidelity. As one can check easily, $\mathcal{P}\circ\mathcal{E}\circ\mathcal{P}$ is Gibbs preserving covariant for any Gibbs-preserving operation $\mathcal{E}$. From this,

$$\max_{\mathcal{E}\in\mathrm{GPO}}\min_{\rho\in\mathcal{S}} F(\mathcal{P}\circ\mathcal{E}\circ\mathcal{P}(\rho),|1\rangle\langle 1|)\leq \max_{\mathcal{E}\in\mathrm{GPC}}\min_{\rho\in\mathcal{S}} F(\mathcal{E}(\rho),|1\rangle\langle 1|) = F_{\mathrm{GPC}}((\mathcal{S},\tau)\to(|1\rangle\langle 1|,\mu_m)) \tag{B27}$$

follows. From these, we $F_{\mathrm{GPC}}((\mathcal{S},\tau)\to(|1\rangle\langle 1|,\mu_m))\geq F_{\mathrm{GPO}}((\mathcal{P}(\mathcal{S}),\tau)\to(|1\rangle\langle 1|,\mu_m))$ is shown. Combining the two inequalities, the proof is completed. □

This lemma provides us the expression of the one-shot black box extractable work using the composite hypothesis testing divergence.

Theorem S.4. *Let $\mathcal{S}\subset\mathcal{D}(\mathcal{H})$ be the black box. The one-shot extractable work under the Gibbs-preserving covariant operations is represented as*

$$\beta W^{\varepsilon}_{\mathrm{GPC}}(\mathcal{S}) = D^{\varepsilon}_{H}(\mathcal{P}(\mathcal{S})||\tau). \tag{B28}$$

Proof. From the definition of the one-shot extractable work and Lemma S.3,

$$\begin{aligned}\beta W^{\varepsilon}_{\mathrm{GPC}}(\mathcal{S}) &= \log\max\{m\in\mathbb{R}\mid F_{\mathrm{GPC}}((\mathcal{S},\tau)\to(|1\rangle\langle 1|,\mu_m))\geq 1-\varepsilon\}\\ &= \log\max\{m\in\mathbb{R}\mid F_{\mathrm{GPO}}((\mathcal{P}(\mathcal{S}),\tau)\to(|1\rangle\langle 1|,\mu_m))\geq 1-\varepsilon\}\\ &= \beta W^{\varepsilon}_{\mathrm{GPO}}(\mathcal{P}(\mathcal{S})) = D^{\varepsilon}_{H}(\mathcal{P}(\mathcal{S})||\tau).\end{aligned} \tag{B29}$$

holds. We used Lemma S.3. In the last equation, we used the result in Theorem S.1. □

Appendix C: Asymptotic black box work extraction and composite quantum Stein's lemmas

1. Asymptotic black box work extraction under Gibbs-preserving operations (Proof of Theorem 4)

In this section, we consider the asymptotic black box work extraction. To take the asymptotic limit, we have to consider the sequence of black boxes, i.e., the sequence of the subsets of the density matrices $\{\mathcal{S}_n\}_{n=1}^{\infty}$, $\mathcal{S}_n\subset\mathcal{D}(\mathcal{H}^{\otimes n})$. Additionally, we consider a specific sequence of black boxes that satisfies the following:

1. For any $n\in\mathbb{N}$, $\mathcal{S}_n$ is closed.
2. For any $n\in\mathbb{N}$, $\mathcal{S}_n$ is closed under the measurement on any subsystems and conditioning on the measurement outcome.

3. For any $n \in \mathbb{N}$, $\mathcal{S}_n$ is closed under taking partial trace on any subsystems.

4. For any $n \in \mathbb{N}$, $\mathcal{S}_n$ is closed under permutation of the subsystems.

When we impose these conditions on the sequence of the black boxes, the problem can be reduced to the classical adversarial hypothesis testing, which leads us to obtain the asymptotic limit [11, 14].

Some examples which satisfy these conditions are the following:

- The i.i.d. states black box $\mathcal{S}_n^{\text{i.i.d.}}(S) = \left\{\rho^{\otimes n} \mid \rho \in S\right\}$
- The tensor product states black box $\mathcal{S}_n^{\text{TP}}(S) = \left\{\bigotimes_{i=1}^n \rho_i \mid \rho_i \in S, \forall i\right\}$

Here, $S \subset \mathcal{D}(\mathcal{H})$ is a closed subset. When we take S as $S = \{\rho\}$, it is reduced to the trivial black box.

Furthermore, we assume that the thermal state is represented as $\tau^{\otimes n}$, which means that the Hamiltonian of the n systems are the same and have no correlation. Then, employing Theorem S.1, the one-shot asymptotic extractable work of the black box $\mathcal{S}_n$ under the Gibbs-preserving operations is represented as

$$\beta W_{\text{GPO}}^{\varepsilon}(\mathcal{S}_n) = D_H^{\varepsilon}(\mathcal{S}_n||\tau^{\otimes n}). \tag{C1}$$

To take the $n \to \infty$ limit, we employ a previous result of the composite hypothesis testing and related quantum Stein's lemma.

Proposition S.5. *([14, Theorem 5]) Let $\mathcal{S} = \{\mathcal{S}_n\}$ and $\mathcal{T} = \{\mathcal{T}_n\}$ be the sequence of black boxes satisfying the conditions above. Then,*

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}D_H^{\varepsilon}(\mathcal{S}_n||\mathcal{T}_n) = \lim_{n\to\infty}\frac{1}{n}\min_{\substack{\sigma_n\in C(\mathcal{S}_n)\\ \tau_n\in C(\mathcal{T}_n)}} D(\sigma_n||\tau_n), \tag{C2}$$

where $C(\cdot)$ denotes the convex hull of the set. Furthermore, when for any $n \in \mathbb{N}$, $\mathcal{S}_n$ is contained by the subspace of $\mathcal{D}(\mathcal{H})$ the dimension of which is polynomial to n, the convex hull on $\mathcal{S}_n$ can be removed.

When we take $\mathcal{T}_n = \left\{\tau^{\otimes n}\right\}$, we immediately obtain the following result.

Theorem S.6. *The asymptotic extractable work of the sequence of black boxes $\{\mathcal{S}_n\}_{n=1}^{\infty}$ which satisfies the conditions above under Gibbs-preserving operations is*

$$\begin{aligned}\beta W_{\text{GPO}}(\{\mathcal{S}_n\}_{n=1}^{\infty}) &= \lim_{\varepsilon\to+0}\lim_{n\to\infty}\frac{1}{n}D_H^{\varepsilon}(\mathcal{S}_n||\tau^{\otimes n})\\ &= \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)} D(\rho_n||\tau^{\otimes n}),\end{aligned} \tag{C3}$$

where $C(\mathcal{S}_n)$ is the convex hull of $\mathcal{S}_n$. Furthermore, when for any $n \in \mathbb{N}$, $\mathcal{S}_n$ is contained by the subspace of $\mathcal{D}(\mathcal{H})$ the dimension of which is polynomial to n, the convex hull on $\mathcal{S}_n$ can be removed.

When the black box is the i.i.d. states black box, for example, the convex hull can be removed. When the black box is the i.i.d. states black box, from the additivity of the relative entropy,

$$\begin{aligned}\beta W_{\text{GPO}}(\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}) &= \lim_{n\to\infty}\frac{1}{n}\min_{\rho^{\otimes n}\in\mathcal{S}_n} D(\rho^{\otimes n}||\tau^{\otimes n})\\ &= \min_{\rho\in S} D(\rho||\tau)\end{aligned} \tag{C4}$$

holds. In this case, the asymptotic black box extractable work equals the worst-case extractable work in the normal setting where the experimenters have complete information.

2. Asymptotic black box work extraction under Gibbs-preserving covariant operations (Proofs of Lemma 6 and Theorem 7)

In the same way as the previous discussion, for a given sequence of the black boxes $\{\mathcal{S}_n\}_{n=1}^{\infty}$, the asymptotic black box work extraction under Gibbs-preserving covariant operations is expressed as follows.

$$\beta W_{\text{GPC}}(\{\mathcal{S}_n\}_{n=1}^{\infty}) = \lim_{\varepsilon\to+0}\lim_{n\to\infty}\frac{1}{n}D_H^{\varepsilon}(\mathcal{P}(\mathcal{S}_n)||\tau^{\otimes n}) \tag{C5}$$

The RHS is more complicated than the LHS of Proposition S.5, since $\mathcal{P}(\mathcal{S}_n)$ no longer has the tensor-product structure, which means that $\mathcal{P}(\mathcal{S}_n)$ is not closed under the measurement on any subsystems and conditioning on the measurement result. Here, we show that another type of composite quantum Stein's lemma holds even in this case.

Proposition S.7. *Let $\{\mathcal{S}_n\}_{n=1}^{\infty}$ be a sequence of black boxes which satisfies the conditions above. Here, the following holds.*

$$\lim_{\varepsilon\to+0}\lim_{n\to\infty}\frac{1}{n}D_H^{\varepsilon}(\mathcal{P}(\mathcal{S}_n)||\tau^{\otimes n}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)}D(\rho_n||\tau^{\otimes n}) \tag{C6}$$

Note that a similar type of the quantum Stein's lemma can be seen in [43, Lemma 16], and is included by Proposition S.7.

To show this, we start from showing the following lemma.

Lemma S.8. *Let $\rho,\tau\in\mathcal{D}(\mathcal{H})$ be arbitrary states, and $\mathcal{P}$ be the pinching channel with respect to τ. Then, the following holds.*

$$0 \le D(\rho||\tau) - D(\mathcal{P}(\rho)||\tau) \le \log|\mathrm{spec}(\tau)|, \tag{C7}$$

where $|\mathrm{spec}(\tau)|$ is the number of the different eigenvalues of τ.

Proof. $0 \le D(\rho||\tau) - D(\mathcal{P}(\rho)||\tau)$ is shown by using the data processing inequality of the relative entropy. To show the other inequality, we first employ Hayashi's pinching inequality [17]

$$\mathcal{P}(\rho) \ge \frac{\rho}{|\mathrm{spec}(\tau)|}. \tag{C8}$$

Due to this inequality and the properties of the relative entropy, the following holds.

$$\begin{aligned} D(\rho||\tau) &= D(\mathcal{P}(\rho)||\tau) + D(\rho||\mathcal{P}(\rho)) \\ &\le D(\mathcal{P}(\rho)||\tau) + D(\rho||\rho/|\mathrm{spec}(\tau)|) = D(\mathcal{P}(\rho)||\tau) + \log|\mathrm{spec}(\tau)|, \end{aligned} \tag{C9}$$

where in the inequality, we used the operator monotonicity of log. □

In the subsequent discussion, we denote the CPTP map which represent the measurement whose POVM elements are $\{E_a\}_a$ as

$$\mathcal{M}(\rho) := \sum_a \mathrm{Tr}[\rho E_a]\,|a\rangle\langle a|\,, \tag{C10}$$

where $\{|a\rangle\}_a$ is the orthogonal vectors in the classical system. We review the concepts called compatible pair.

Definition S.9. [11] Let $M=(M_1,M_2,\ldots)$ be the sequence of the measurements with M_n reprsenting the set of measurements on $\mathcal{D}(\mathcal{H}^{\otimes n})$. Furthermore, let $\mathcal{S}=(\mathcal{S}_1,\mathcal{S}_2,\ldots)$ be the sequence of the sets of state where $\mathcal{S}_n$ is the subset of $\mathcal{D}(\mathcal{H}^{\otimes n})$ for every $n\in\mathbb{N}$. We say that $(M,\mathcal{S})$ is the compatible pair when the sequence $\mathcal{S}$ is closed under the measurement in M on any subsystems and conditioning on the measurement outcome, i.e., for any state $\rho_{n+k}\in\mathcal{S}_{n+k}$, after performing any measurement in M_k and conditioning on the outcome, the post-measured state is the element of $\mathcal{S}_n$.

Furthermore, we consider the restricted set of measurements called covariant measurements, which guarantees that the probability distribution obtained by such a measurement is invariant under time translation.

Definition S.10. Let H be the Hamiltonian of the considered system, and $\{M_a\}_a$ be POVM elements of a measurement M. We say that M is a covariant measurement if and only if all the measurement operators M_a satisfy the following condition.

$$\forall a,\ \ \mathcal{P}(M_a) = M_a. \tag{C11}$$

We denote the sequence of the covariant measurement as M^{cov}.

We also denote the sequence of all measurements as M^{all}. Before showing Proposition S.7, we show the following two lemmas.

Lemma S.11. *Let $\mathcal{S} = \{\mathcal{S}_n\}_{n=1}^{\infty}$ be the sequence of the black boxes satisfying the condition mentioned above. Furthermore, let $\mathcal{P}(C(\mathcal{S}))$ be the sequence of the sets of states $\{\mathcal{P}(C(\mathcal{S}_n))\}_{n=1}^{\infty}$. Here, $(M^{\mathrm{cov}}, \mathcal{P}(C(\mathcal{S})))$ is the compatible pair.*

Note that $(M^{\mathrm{cov}}, \mathcal{P}(\mathcal{S}))$ is the compatible pair too, which is proven in the same way as the following proof. Here, we consider the convex hull of the black boxes to use this lemma to show Proposition S.7.

Proof. of Lemma S.11. We first note that due to the structure of any POVM elements of the covariant measurement M_a^{cov},

$$\mathcal{P}\left((M_a^{\mathrm{cov}})^{\dagger} M_a^{\mathrm{cov}}\right) = (M_a^{\mathrm{cov}})^{\dagger} M_a^{\mathrm{cov}} \tag{C12}$$

holds. Here, it suffices to show that for any measurement operators M_a^{cov} and the arbitrary state $\mathcal{P}(\rho_{n+k}) \in \mathcal{P}(C(\mathcal{S}_{n+k})), \forall n, k \in \mathbb{N}$

$$\mathrm{Tr}_{n+1,\ldots,n+k}\left[(I \otimes M_a^{\mathrm{cov}})\rho_{n+k}(I \otimes M_a^{\mathrm{cov}})^{\dagger}\right] \in \mathcal{P}(C(\mathcal{S}_n)) \tag{C13}$$

holds. This can be checked as follows.

$$\begin{aligned}
\mathrm{Tr}_{n+1,\ldots,n+k}\left[(I \otimes M_a^{\mathrm{cov}})\rho_{n+k}(I \otimes M_a^{\mathrm{cov}})^{\dagger}\right] &= \mathrm{Tr}\left[(I \otimes (M_a^{\mathrm{cov}})^{\dagger} M_a^{\mathrm{cov}})\mathcal{P}(\rho_{n+k})\right] \\
&= \mathrm{Tr}\left[\mathcal{P}(I \otimes (M_a^{\mathrm{cov}})^{\dagger} M_a^{\mathrm{cov}})\rho_{n+k}\right] \\
&= \mathrm{Tr}\left[(I \otimes (M_a^{\mathrm{cov}})^{\dagger} M_a^{\mathrm{cov}})\rho_{n+k}\right] \in \mathcal{P}(C(\mathcal{S}_n)).
\end{aligned} \tag{C14}$$

The first line is because the pinching channel satisfies $\mathcal{P}^{\dagger} = \mathcal{P}$, and the property mentioned at the beginning of this proof is used in the second line. Finally, the third line is due to the second property of the sequence of the black boxes $\mathcal{S} = \{\mathcal{S}_n\}_{n=1}^{\infty}$. □

Lemma S.12. *It holds that*

$$\sup_{\mathcal{M} \in M_n^{\mathrm{cov}}} \min_{\rho_n \in C(\mathcal{S}_n)} D(\mathcal{M}(\rho_n)||\mathcal{M}(\tau^{\otimes n})) = \sup_{\mathcal{M} \in M_n^{\mathrm{all}}} \min_{\rho_n \in C(\mathcal{S}_n)} D(\mathcal{M}(\mathcal{P}(\rho_n))||\mathcal{M}(\tau^{\otimes n})). \tag{C15}$$

Proof. First, we show the $(\leq)$ inequality. For any states ρ_n and any POVM element of the covariant measurement M_a^{cov}, Noting that $\mathcal{P}(M_a^{\mathrm{cov}}) = M_a^{\mathrm{cov}}$, it holds that

$$\mathrm{Tr}\left[\rho_n M_a^{\mathrm{cov}}\right] = \mathrm{Tr}\left[\rho_n \mathcal{P}\left(M_a^{\mathrm{cov}}\right)\right] = \mathrm{Tr}\left[\mathcal{P}(\rho_n) M_a^{\mathrm{cov}}\right]. \tag{C16}$$

This implies that all the probability distributions which are obtained by measuring the state ρ_n with the covariant measurement can be realized by measuring the pinched state $\mathcal{P}(\rho_n)$ with the covariant measurement. From this,

$$\sup_{\mathcal{M} \in M_n^{\mathrm{cov}}} \min_{\rho_n \in C(\mathcal{S}_n)} D(\mathcal{M}(\rho_n)||\mathcal{M}(\tau^{\otimes n})) \leq \sup_{\mathcal{M} \in M_n^{\mathrm{all}}} \min_{\rho_n \in C(\mathcal{S}_n)} D(\mathcal{M}(\mathcal{P}(\rho_n))||\mathcal{M}(\tau^{\otimes n})). \tag{C17}$$

holds. To show the $(\geq)$ inequality, note that

$$\mathrm{Tr}\left[\mathcal{P}(\rho_n) M_a^{\mathrm{all}}\right] = \mathrm{Tr}\left[\rho_n \mathcal{P}(M_a^{\mathrm{all}})\right] \tag{C18}$$

holds. Here, $\left\{\mathcal{P}(M_a^{\mathrm{all}})\right\}_a$ satisfies the conditions for POVM elements, i.e.,

$$\forall a,\ \mathcal{P}(M_a^{\mathrm{all}}) \geq 0, \quad \sum_a \mathcal{P}(M_a^{\mathrm{all}}) = \mathcal{P}\left(\sum_a M_a^{\mathrm{all}}\right) = \mathcal{P}(I) = I \tag{C19}$$

holds. Furthermore, due to the definition of the pinching channel, the measurement whose POVM elements are represented as $\left\{\mathcal{P}(M_a^{\mathrm{all}})\right\}_a$ is the covariant measurement. From this, one can see that all the probability distribution obtained by measuring the pinched state $\mathcal{P}(\rho_n)$ with any measurement can be realized by measuring ρ_n with covariant measurement which implies

$$\sup_{\mathcal{M} \in M_n^{\mathrm{cov}}} \min_{\rho_n \in C(\mathcal{S}_n)} D(\mathcal{M}(\rho_n)||\mathcal{M}(\tau^{\otimes n})) \geq \sup_{\mathcal{M} \in M_n^{\mathrm{all}}} \min_{\rho_n \in C(\mathcal{S}_n)} D(\mathcal{M}(\mathcal{P}(\rho_n))||\mathcal{M}(\tau^{\otimes n})), \tag{C20}$$

which completes the proof. □

Here, we are ready to show Proposition S.7.

Proof. of Proposition S.7. Note that

$$D_H^{\varepsilon}(\mathcal{P}(\mathcal{S}_n)||\tau^{\otimes n}) = D_H^{\varepsilon}(C(\mathcal{P}(\mathcal{S}_n))||\tau^{\otimes n}) \tag{C21}$$

holds and $D_H^{\varepsilon}(C(\mathcal{P}(\mathcal{S}_n))||\tau^{\otimes n})$ can be interpreted as the hypothesis testing divergence of $C(\mathcal{S}_n)$ with respect to $\tau^{\otimes n}$ when the allowed measurement is restricted to the covariant measurement. Due to Lemma S.11 and [11, Theorem 16], the following holds.

$$\begin{aligned}\lim_{\varepsilon\to+0}\lim_{n\to\infty}\frac{1}{n}D_H^{\varepsilon}(C(\mathcal{P}(\mathcal{S}_n))||\tau^{\otimes n}) &= \lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{M}\in M_n^{\mathrm{cov}}}\min_{\rho_n\in C(\mathcal{S}_n)}D(\mathcal{M}(\rho_n)||\mathcal{M}(\tau^{\otimes n}))\\ &= \lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{M}\in M_n^{\mathrm{all}}}\min_{\rho_n\in C(\mathcal{S}_n)}D(\mathcal{M}(\mathcal{P}(\rho_n))||\mathcal{M}(\tau^{\otimes n}))\end{aligned} \tag{C22}$$

Here, the second line follows from Lemma S.12 Employing [11, Lemma 13], we can exchange the sup and min, i.e.,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sup_{\mathcal{M}\in M_n^{\mathrm{all}}}\min_{\rho_n\in C(\mathcal{S}_n)}D(\mathcal{M}(\mathcal{P}(\rho_n))||\mathcal{M}(\tau^{\otimes n})) &= \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)}\sup_{\mathcal{M}\in M_n^{\mathrm{all}}}D(\mathcal{M}(\mathcal{P}(\rho^{\otimes n}))||\mathcal{M}(\tau^{\otimes n}))\\ &= \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)}D_{M_n^{\mathrm{all}}}(\mathcal{P}(\rho_n)||\tau^{\otimes n}),\end{aligned} \tag{C23}$$

where $D_{M^{\mathrm{all}}}(\mathcal{P}(\rho_n)||\tau^{\otimes n})$ is the M_n^{all}-measured relative entropy of $\mathcal{P}(\rho_n)$ with respect to $\tau^{\otimes n}$ [48].

We first note that the infimum is achieved at a permutation invariant state [14, Lemma 23]. Therefore, letting PI_n be the set of n-qudit permutation invariant states, we get

$$\lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)}D_{M_n^{\mathrm{all}}}(\mathcal{P}(\rho_n)||\tau^{\otimes n}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)\cap\mathrm{PI}_n}D_{M_n^{\mathrm{all}}}(\mathcal{P}(\rho_n)||\tau^{\otimes n}). \tag{C24}$$

We now recall [13, Lemma 2.4], showing that for all permutation invariant states η_n and σ_n, it holds that

$$D(\eta_n\|\sigma_n) - \log\mathrm{poly}(n) \leq D_{M_{\mathrm{all}}}(\eta_n\|\sigma_n) \leq D(\eta_n\|\sigma_n). \tag{C25}$$

This implies

$$\lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)\cap\mathrm{PI}_n}D_{M_{\mathrm{all}}}(\mathcal{P}(\rho_n)||\tau^{\otimes n}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)\cap\mathrm{PI}_n}D(\mathcal{P}(\rho_n)||\tau^{\otimes n}). \tag{C26}$$

Note that the number of the different eigenvalue of $H^{\times n}$ is upper bounded by the number of type classes of n length strings when the set of alphabets is $\{0,1,\ldots,d-1\}$. Since the number of type classes is upper-bounded by $(n+1)^d$ [49], due to Lemma S.8, it holds that for any states ρ_n,

$$D(\rho_n||\tau^{\otimes n}) - \log\mathrm{poly}(n) \leq D(\mathcal{P}(\rho_n)||\tau^{\otimes n}) \leq D(\rho_n||\tau^{\otimes n}), \tag{C27}$$

which implies

$$\lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)\cap\mathrm{PI}_n}D(\mathcal{P}(\rho_n)||\tau^{\otimes n}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)\cap\mathrm{PI}_n}D(\rho_n||\tau^{\otimes n}). \tag{C28}$$

Again, employing [14, Lemma 23], we obtain

$$\lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)\cap\mathrm{PI}_n}D(\rho_n||\tau^{\otimes n}) = \lim_{n\to\infty}\frac{1}{n}\min_{\rho_n\in C(\mathcal{S}_n)}D(\rho_n||\tau^{\otimes n}). \tag{C29}$$

Combining these, we complete the proof.

□

Appendix D: Asymptotic black box work extraction under thermal operations

1. Construction of work extraction protocol under covariantly conditioned thermal operations

In the following discussion, we consider the i.i.d. black boxes, which contain a finite number of states, i.e., the black box $\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}$ with $|S| < \infty$. Our goal in this section is to show that under thermal operations, one can extract the same amount of work asymptotically as the Gibbs-preserving operations and Gibbs-preserving covariant operations when the given state is picked from the i.i.d. black boxes, which contain a finite number of states.

First, we introduce a new class of operations called covariantly conditioned thermal operations, thermal operations conditioned by the outcome of the covariant measurement. In [45], the class of operations called conditioned thermal operations is introduced, in which one performs the measurement on one of the bipartite systems, and applies the thermal operation conditioned by the measurement outcome. The class called covariantly conditioned thermal operation restricts the measurement one can perform to the covariant measurement. The rigorous definition of the class is the following.

Definition S.13. Let $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C$ be Hilbert spaces, and $\mathcal{E} : \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_C)$ be a CPTP map. $\mathcal{E}$ is called a thermal operations + covariant measurements when $\mathcal{E}$ can be decomposed as follows.

$$\mathcal{E} = \sum_a \mathcal{E}_a^{\text{TO}} \circ \Lambda_a^{\text{meas}}. \tag{D1}$$

Here, $\mathcal{E}_a^{\text{TO}} : \mathcal{D}(\mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_C),\ a = 1, 2, \ldots, m$ is thermal operations and

$$\Lambda_a^{\text{meas}}(\rho_{ABC}) := \text{Tr}_A\left[(M_a^{\text{cov}} \otimes I_{BC})\rho_{ABC}(M_a^{\text{cov}} \otimes I_{BC})^\dagger\right],\ a = 1, \ldots, m \tag{D2}$$

be the instruments which represent the covariant measurement, where M_a^{cov} is the measurement operator on $\mathcal{D}(\mathcal{H}_A)$ which satisfies $\sum_a (M_a^{\text{cov}})^\dagger M_a^{\text{cov}} = I$. We denote the set of covariantly conditioned thermal operations as $\text{CCTO}(A, B \to C)$ Furthermore, when the measurements are restricted to the covariant and projective measurements, we say that the operation is thermal operation + covariant projective measurement. We denote the set of thermal operations + covariant projective measurements as $\text{CCPTO}(A, B \to C)$

Our first goal is to show the following proposition.

Proposition S.14. *Let $S \subset \mathcal{D}(\mathcal{H})$ be a subset of density matrices that contain a finite number of density matrices. The asymptotic extractable work of the sequence of the i.i.d. black boxes $\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}$ under covariantly conditioned thermal operations is represented as*

$$\beta W_{\text{CCTO}}(\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}) = \min_{\rho \in S} D(\rho||\tau). \tag{D3}$$

Proof. (of ($\leq$) part.) To show the ($\leq$) inequality, we note the hierarchy of the operations

$$\text{Thermal} \subset \text{covariantly conditioned Thermal} \subset \text{Gibbs} - \text{preserving covariant} \subset \text{Gibbs} - \text{preserving}, \tag{D4}$$

which implies

$$\beta W_{\text{CCTO}}(\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}) \leq \beta W_{\text{GPO}}(\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}) = \min_{\rho \in S} D(\rho||\tau). \tag{D5}$$

The last inequality is due to Eq. (C4). □

To show the other inequality, we construct the concrete protocol as follows.

1. Given n copies of some unknown state ρ, pick up k_{δ', p_e} copies of states and perform the covariant measurement. Here, k_{δ', p_e} is a natural number that depends on the necessary accuracy to identify the initial state.

2. Identify $\mathcal{P}(\rho^{\otimes d})$ from the measurement outcome.

3. Perform the protocol in [4] using the information obtained in Step 2.

One may see it weird that the goal of the second step is not to identify ρ. Actually, it is not possible to perform the state tomography with the covariant measurement even if the experimenters have an infinite number of copies. The simplest situation is where the experimenters are given a qubit system which is either $|+\rangle\langle+|$ or $|-\rangle\langle-|$ where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$, and are told to guess which is the state with covariant measurements. We assume that the Hamiltonian of the system is $H = |1\rangle\langle 1|$. Here, from the definition of the covariant measurement, for any $\rho \in \mathcal{D}(\mathcal{H})$ and the POVM elements of the covariant measurement E_a^{cov} on the system,

$$\mathrm{Tr}\left[E_a^{\mathrm{cov}}\rho\right] = \mathrm{Tr}\left[\mathcal{P}(E_a^{\mathrm{cov}})\rho\right] = \mathrm{Tr}\left[E_a^{\mathrm{cov}}\mathcal{P}(\rho)\right] \tag{D6}$$

holds. This implies that the probability distribution of the outcome of the covariant measurement does not change when the state is pinched. Here, we observe the two state $\mathcal{P}(|+\rangle\langle+|^{\otimes n})$ and $\mathcal{P}(|-\rangle\langle-|^{\otimes n})$ coincide for every $n \in \mathbb{N}$. Let $\boldsymbol{s}, \boldsymbol{t} \in \{0,1\}^n$ be the n-bit string. One can see that due to the definition of the pinching channel

$$\langle \boldsymbol{s}|\, \mathcal{P}(\rho^{\otimes n})\, |\boldsymbol{t}\rangle \neq 0 \Rightarrow \boldsymbol{s} \text{ and } \boldsymbol{t} \text{ belong to the same type class.} \tag{D7}$$

Furthermore, the direct calculation shows that

$$\langle \boldsymbol{s}|\, \rho^{\otimes n}\, |\boldsymbol{t}\rangle = \prod_{i=1}^{n} \langle s_i|\, \rho\, |t_i\rangle\,, \tag{D8}$$

where s_i and t_i are the i-th alphabets of $\boldsymbol{s}$ and $\boldsymbol{t}$ respectively. When $\boldsymbol{s}$ and $\boldsymbol{t}$ belong to the same type class,

$$|\{i \in \{1,\dots,d\} | s_i = 0,\ t_i = 1\}| = |\{i \in \{1,\dots,d\} | s_i = 1,\ t_i = 0\}| \tag{D9}$$

holds.

From this, we can see that all the elements of $\mathcal{P}(|-\rangle\langle-|^{\otimes n})$ in the energy subspaces are all $1/2^n$, and we can conclude that $\mathcal{P}(|+\rangle\langle+|^{\otimes n}) = \mathcal{P}(|-\rangle\langle-|^{\otimes n})$ for any n. Therefore, even if the experimenters perform any covariant measurement, they can never distinguish $|+\rangle\langle+|$ and $|-\rangle\langle-|$. Due to the discussion above, we can never estimate the unknown state ρ by covariant measurement.

Then, what information can one obtain by the covariant measurement? In the subsequent discussion, we give an answer to this question.

First of all, we restrict the Hamiltonian in consideration to what satisfies the following property.

Definition S.15. Let H be a Hamiltonian, and $E_1, \dots, E_d$ be the eigenvalues of H. We say that H is rationally independent when H satisfies the following.

$$\sum_i N_i E_i = 0,\ \sum_i N_i = 0,\ N_i \in \mathbb{Z},\ \forall i \Rightarrow N_i = 0,\ \forall i. \tag{D10}$$

In the following discussion, we assume that the Hamiltonian of each system satisfies this property. Rational independence prohibits any number of copies of the systems from having degenerate energy levels other than the degeneracy that comes from the permutation of the systems. A simple example is the qutrit system $\mathcal{H}_3 = \mathrm{Span}\{|0\rangle, |1\rangle, |2\rangle\}$ and the Hamiltonian of the system $H = E\,|1\rangle\langle 1| + 2E\,|2\rangle\langle 2|$. The Hamiltonian H itself does not have degeneracy. However, when we prepare two copies of this, the energy subspace which corresponds to the energy eigenvalue $2E$ of the Hamiltonian of the whole system $H^{\times 2}$ is $\mathrm{Span}\{|02\rangle\,|20\rangle, |11\rangle\}$. This additional degeneracy comes up because of the rational dependence of the Hamiltonian.

In the following discussion, we mainly focus on the Hamiltonian, which has this property. We note that the energy subspaces of the n copies of the system are spanned by the vectors that belong to the same type class, and there exists a one-to-one correspondence between the energy eigenvalue of the n copies of the system and the type of the eigenvectors. After that, we also extend the discussion to the case where the Hamiltonian is not rationally independent.

In the subsequent discussion, $\rho \in \mathcal{D}(\mathcal{H})$ is a density operator, and $\{|i\rangle\}_{i=1}^d$ are the eigenvectors of the Hamiltonian H. Furthermore, we denote $\rho_{ij} := \langle i|\, \rho\, |j\rangle$. We show that we can estimate the values called cyclic product defined below in any accuracy by the covariant measurement.

Definition S.16. Let $\boldsymbol{s} \in \{1,\dots,d\}^m$ be a string of length m, which is composed of m different alphabets, i.e., $s_i = s_j \Leftrightarrow i = j$. Here, $m \leq d$ holds. The cyclic product of ρ with respect to the string $\boldsymbol{s}$ is defined as

$$\prod_{i=1}^{m} \rho_{s_i s_{i+1}} \quad \left(= \langle s_1 s_2 \cdots s_m|\, \rho^{\otimes m}\, |s_2 \cdots s_m s_1\rangle\right), \tag{D11}$$

where we set $s_{m+1} = s_1$.

Note that the number of the different cyclic products is finite. Furthermore, any cyclic products are in the energy subspace, since $\boldsymbol{s} = s_1 s_2 \cdots s_m$ and $\boldsymbol{s}' = s_2 \cdots s_m s_1$ belong to the same type class.

Lemma S.17. *Suppose that every matrix element of $\mathcal{P}(\rho^{\otimes d})$ is given. Then, all the cyclic products of ρ can be calculated from the elements of $\mathcal{P}(\rho^{\otimes d})$.*

Proof. We first consider the diagonal elements of ρ. Since for any $i \in \{1, \ldots, d\}$, $\rho_{ii} \geq 0, \langle ii \ldots i | \mathcal{P}(\rho^{\otimes d}) | ii \ldots i \rangle \geq 0$ due to the positive semidefiniteness of ρ and $\mathcal{P}(\rho^{\otimes d})$, and $\langle ii \ldots i | \mathcal{P}(\rho^{\otimes d}) | ii \ldots i \rangle = (\rho_{ii})^d$ holds, we can calculate ρ_{ii} as $\rho_{ii} = (\langle ii \ldots i | \mathcal{P}(\rho^{\otimes d}) | ii \ldots i \rangle)^{1/d}$. One can calculate any cyclic products with respect to the string $\boldsymbol{s}$ by choosing $j \in \{1, \ldots, d\}$ such that $\rho_{jj} \neq 0$ and noting that

$$\begin{aligned} \langle j \ldots j \boldsymbol{s} | \mathcal{P}(\rho^{\otimes d}) | j \ldots j \boldsymbol{s}' \rangle &= (\rho_{jj})^{d-|\boldsymbol{s}|} \prod_{i=1}^{m} \rho_{s_i s_{i+1}}, \\ \prod_{i=1}^{m} \rho_{s_i s_{i+1}} &= \frac{\langle j \ldots j \boldsymbol{s} | \mathcal{P}(\rho^{\otimes d}) | j \ldots j \boldsymbol{s}' \rangle}{(\langle jj \ldots j | \mathcal{P}(\rho^{\otimes d}) | jj \ldots j \rangle)^{\frac{d-|\boldsymbol{s}|}{d}}}. \end{aligned} \tag{D12}$$

□

Once we obtain the list of the all cyclic products, under the assumption that the Hamiltonian is rationally independent, we can reconstruct $\mathcal{P}(\rho^{\otimes n})$ for any $n \in \mathbb{N}$.

Lemma S.18. *Any nonzero elements of $\mathcal{P}(\rho^{\otimes n})$ $\forall n \in \mathbb{N}$ can be represented as the product of cyclic products.*

Proof. Due to the definition of the pinching channel, the matrix elements of $\mathcal{P}(\rho^{\otimes n})$ that are not inside the energy block are 0. Therefore, it suffices to consider the matrix elements inside the energy blocks. Any nonzero elements of $\mathcal{P}(\rho^{\otimes n})$ can be written in the form of

$$\langle \boldsymbol{s} | \rho^{\otimes n} | \boldsymbol{t} \rangle = \prod_{i=1}^{n} \rho_{s_i t_i}, \tag{D13}$$

where $\boldsymbol{s}, \boldsymbol{t} \in \{1, \ldots, d\}^n$ are the strings of length n which belong to the same type class. One can rearrange $\rho_{s_1 t_1}, \ldots, \rho_{s_n t_n}$ to the following form.

$$\rho_{a_1 a_2} \rho_{a_2 a_3} \cdots \rho_{a_n a_1}, \quad a_1, \ldots, a_n \in \{1, \ldots, d\} \tag{D14}$$

Note that this does not mean that any matrix elements of $\mathcal{P}(\rho^{\otimes n})$ are the cyclic products, since the $a_1, \ldots, a_n$ can include the same alphabet. The existence of such a sequence is guaranteed by the assumption that $\boldsymbol{s}, \boldsymbol{t} \in \{1, \ldots, d\}^n$ are in the same type class. Now, we separate this sequence into the cyclic products in the following way. If $a_i = a_j (= \alpha)$, $i \neq j$, we divide the sequence above as

$$\rho_{a_1 a_2} \cdots \rho_{a_{i-1} \alpha} \rho_{\alpha a_{i+1}} \cdots \rho_{a_{j-1} \alpha} \rho_{\alpha a_{j+1}} \cdots \rho_{a_n a_1} \to \rho_{a_1 a_2} \cdots \rho_{a_{i-1} \alpha} \rho_{\alpha a_{j+1}} \cdots \rho_{a_n a_1}, \ \rho_{\alpha a_{i+1}} \cdots \rho_{a_{j-1} \alpha} \tag{D15}$$

These two terms also have the form in Eq. (D14). This procedure can be carried out until the divided terms have no overlaps in the alphabets, in other words, they are divided into the cyclic products. This decomposition can be done in any nonzero matrix elements in $\mathcal{P}(\rho^{\otimes n})$, which completes the proof. □

From these lemmas above, we can easily see the following.

Lemma S.19. *For any density matrix of the qudit system $\rho_1, \rho_2 \in \mathcal{D}(\mathcal{H})$,*

$$\mathcal{P}(\rho_1^{\otimes d}) = \mathcal{P}(\rho_2^{\otimes d}) \Leftrightarrow \mathcal{P}(\rho_1^{\otimes n}) = \mathcal{P}(\rho_2^{\otimes n}), \ \ \forall n \in \mathbb{N}. \tag{D16}$$

Proof. ($\Leftarrow$) is obvious, and ($\Rightarrow$) follows because the left condition implies that all cyclic products of ρ_1 and ρ_2 are the same, which means the condition of the right-hand side due to Lemma S.18. □

Using these lemmata, we can show the Proposition S.14.

Proof. (of (≥) part.) In [4], it is shown that one can perform the protocol Λ in which for any $\varepsilon' > 0$ and $\eta' > 0$ there exists $N \in \mathbb{N}$ such that

$$n \geq N \Rightarrow \exists m_n \in \mathbb{N} \text{ s.t. } F(\Lambda \circ \mathcal{P}(\rho^{\otimes n}), (|1\rangle\langle 1|, \mu_{m_n})) \geq 1 - \varepsilon', \left| D(\rho||\tau) - \frac{1}{n} \log m_n \right| < \eta'. \tag{D17}$$

Note that since in this protocol, one first applies the pinching channel, it suffices to obtain information about the pinched state by the covariant measurement for the work extraction protocol. Furthermore, due to Lemma S.19, in order to specify the input state, one just needs to perform the quantum state tomography on the pinched d copies of input state $\mathcal{P}(\rho^{\otimes d})$.

Note that the black box contains a finite number of states. We define $\delta > 0$ as

$$2\delta := \min_{\substack{\rho_i, \rho_j \in S \\ \mathcal{P}(\rho_i^{\otimes d}) \neq \mathcal{P}(\rho_j^{\otimes d})}} \|\mathcal{P}(\rho_i^{\otimes d}) - \mathcal{P}(\rho_j^{\otimes d})\|_1. \tag{D18}$$

To perform the quantum state tomography on $\mathcal{P}(\rho^{\otimes d})$ with accuracy δ' with respect to the trace distance and with success probability $1 - p_e$, it suffices to use

$$k_{\delta', p_e} = O\left(\frac{d^{2d}}{\delta'^2} \log \left(\frac{1}{p_e} \right) \right) \tag{D19}$$

copies of $\mathcal{P}(\rho^{\otimes d})$ [50, Corollary 1.4]. If we take δ' smaller than δ, the probability of judging ρ as other elements in black boxes p_e can be arbitrarily small.

Now, we perform the covariant measurement and the thermal operations conditioned by the measurement outcome as follows. We denote the appropriate operations for the input state $\rho_i^{\otimes n}$ as $\Lambda_{i,n}$. We perform state tomography with covariant measurement using k_{δ', p_e} copies of given unknown state ρ, and obtain an estimate of $\mathcal{P}(\rho^{\otimes d})$, which we call $\hat{E}$. We then choose $\tilde{i}$ such that $\mathcal{P}(\rho_{\tilde{i}}^{\otimes d})$ realizes the closest value to $\hat{E}$, i.e., $\tilde{i} = \mathrm{argmin}_i \|\hat{E} - \mathcal{P}(\rho_i^{\otimes d})\|_1$. We then apply $\Lambda_{\tilde{i},n}$ to the rest of the states $\rho^{\otimes n}$. Noting that the probability of successfully identifying the unknown state as $\rho_{\tilde{i}} = \rho_i$ is at least $1 - p_e$, this protocol ensures that when the given state is ρ_i, the operation applied to $\rho_i^{\otimes n}$ has the form $(1 - p_e)\Lambda_{i,n} + p_e \Xi$, where Ξ is some quantum channel. This guarantees that for any i and arbitrary $\varepsilon', \eta' > 0$, there exists a sufficiently large n such that

$$\begin{aligned} &F\left(\left((1 - p_e)\Lambda_{i,n} + p_e \Xi \right) \circ \mathcal{P}\left(\rho^{\otimes n}\right), (|1\rangle\langle 1|, \mu_{m_n}) \right) \geq (1 - p_e)(1 - \varepsilon') \\ &\left| D(\rho||\tau) - \frac{1}{n} \log m_n \right| < \eta' \end{aligned} \tag{D20}$$

Let us fix arbitrary $\varepsilon > 0$ and $\eta > 0$. We choose k_{δ', p_e} and n to satisfy

$$(1 - p_e)(1 - \varepsilon') \geq 1 - \varepsilon. \tag{D21}$$

Here, the point is that p_e does not depend on n but only on the accuracy of the state tomography δ'. Furthermore, with respect to the extractable work,

$$\begin{aligned} \left| D(\rho||\tau) - \frac{1}{n + k_{\delta', p_e}} \log m_n \right| &= \left| D(\rho||\tau) - \frac{1}{n} \log m_n + \left(\frac{1}{n} - \frac{1}{n + k_{\delta', p_e}} \right) \log m_n \right| \\ &\leq \left| D(\rho||\tau) - \frac{1}{n} \log m_n \right| + \frac{k_{\delta', p_e}}{n(n + k_{\delta', p_e})} \log m_n \\ &< \eta' + \frac{k_{\delta', p_e}}{n^2} (D(\rho||\tau) + \eta'). \end{aligned} \tag{D22}$$

This implies that if we take sufficiently large n, we can achieve

$$\left| D(\rho||\tau) - \frac{1}{n + k_{\delta', p_e}} \log m_n \right| < \eta. \tag{D23}$$

From these discussions, we can conclude this protocol can achieve the same work extraction as the protocol in [4]. Since we defined the black box extractable work as the worst-case work extraction, it holds that

$$\beta W_{\mathrm{CCTO}}(\{\mathcal{S}_n^{\mathrm{i.i.d.}}(S)\}_{n=1}^{\infty}) \geq \min_{\rho \in S} D(\rho||\tau). \tag{D24}$$

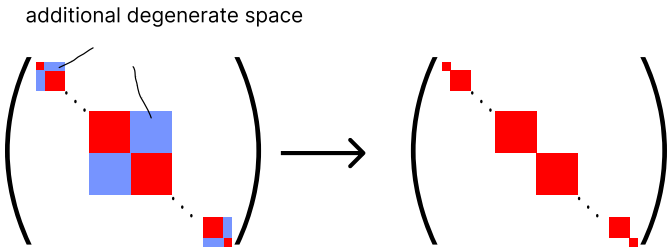


FIG. S.1. The procedure to erase the additional energy subspace due to the rational dependence of the Hamiltonian of each system.

Even when the Hamiltonian is rationally dependent, We can achieve the same performance. To see this, we consider the pinching channel with respect to the rationally independent Hamiltonian, not the Hamiltonian of the system itself. We denote this channel as $\tilde{\mathcal{P}}$. In a similar way as the proof in [6] one can see that $\tilde{\mathcal{P}}$ is a thermal operation. Due to the definition, $\tilde{\mathcal{P}}$ erases the additional degenerate spaces in the pinched density matrix with respect to the original Hamiltonian, and satisfies $\tilde{\mathcal{P}} \circ \mathcal{P} = \tilde{\mathcal{P}} = \mathcal{P} \circ \tilde{\mathcal{P}}$ (see FIG. S.1). Note that applying this pinching channel in advance does not increase the extractable work, i.e.,

$$\beta W_{\mathrm{CCTO}}(\{\mathcal{S}_n^{\mathrm{i.i.d.}}(S)\}_{n=1}^{\infty}) \geq \beta W_{\mathrm{CCTO}}(\{\tilde{\mathcal{P}}(\mathcal{S}_n^{\mathrm{i.i.d.}}(S))\}_{n=1}^{\infty}). \tag{D25}$$

After we apply this pinching channel, we can apply the same state tomography protocol and thermal operation which follows $\tilde{\mathcal{P}}$ conditioned by the result of the tomography. Applying the same protocol, we can achieve the same extractable work as the case where the Hamiltonian is rationally independent. Therefore, it holds that

$$\beta W_{\mathrm{CCTO}}(\{\mathcal{S}_n^{\mathrm{i.i.d.}}(S)\}_{n=1}^{\infty}) \geq \beta W_{\mathrm{CCTO}}(\{\tilde{\mathcal{P}}(\mathcal{S}_n^{\mathrm{i.i.d.}}(S))\}_{n=1}^{\infty}) \geq \min_{\rho \in S} D(\rho||\tau), \tag{D26}$$

which completes the proof. □

Note that the measurement is used to perform the quantum state tomography, we use only the projective measurements for the protocol above.

2. Equivalence of thermal operation and covariantly conditioned thermal operation with projective measurements

In this subsection, we show that one can perform the work extraction protocol in thermal operations. We denote the set of thermal operations from $\mathcal{D}(\mathcal{H}_A)$ to $\mathcal{D}(\mathcal{H}_B)$ as $\mathrm{TO}(A \to B)$. We start with the following proposition.

Proposition S.20. *Let A, B be the input systems and C be the output system. If* $\dim \mathcal{H}_B = \dim \mathcal{H}_C$,

$$\mathrm{TO(B \to C)} = \mathrm{CCPTO(A, B \to C)} \tag{D27}$$

holds.

Proof. The idea stems from [51, Appendix C]. Since $\mathcal{E}_i^{\mathrm{TO}}$ is a thermal operation for any $i \in \{1, \ldots, m\}$, $\mathcal{E}_i^{\mathrm{TO}}$ can be decomposed as follows.

$$\mathcal{E}_i^{\mathrm{TO}}(\rho_B) = \mathrm{Tr}_{E_i'}\left[U_i\left(\rho_B \otimes \tau_{E_i}\right)U_i^{\dagger}\right] \tag{D28}$$

Here, $\tau_{E_i} = \exp(-\beta H_{E_i})/Z_{E_i}$ is a thermal state of the ancillary system E_i, which is associated with the Hilbert space $\mathcal{H}_{E_i}$ and the Hamiltonian H_{E_i}. The unitary operator U_i conserves the energy of the total system, i.e., $\left[U_i, H_B + H_{E_i}\right] = 0$. The system E_i' satisfies $B + E_i = C + E_i'$. We fix an arbitrary $\mathcal{E} \in \mathrm{TO} + \mathrm{cov.\ proj.\ meas.}$, and $\mathcal{E}$ is written as follows.

$$\mathcal{E}(\rho_{AB}) = \sum_{i=1}^{m} \mathrm{Tr}_{E_i'}\left[U_i\Big(\mathrm{Tr}_A\left[(P_i^A \otimes I_B)\rho_{AB}\right] \otimes \tau_{E_i}\Big)U_i^{\dagger}\right] \tag{D29}$$

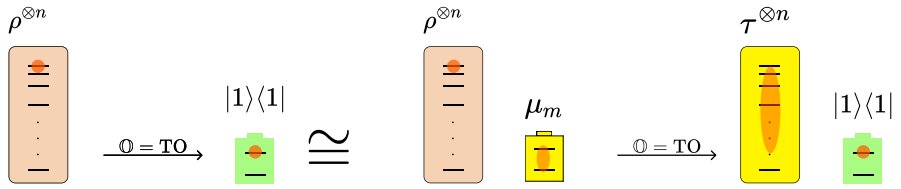


FIG. S.2. In our work extraction setting, we can append the thermal states to the initial and output states to make the dimensions of the input and output systems the same.

Consider another map $\tilde{\mathcal{E}} : \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{D}(\mathcal{H}_C)$, which has the following form.

$$\begin{aligned} \tilde{\mathcal{E}}(\rho_{AB}) &= \mathrm{Tr}_{A,E_1',\ldots,E_m'}\left[\tilde{U}\left(\rho_{AB}\otimes\left(\bigotimes_{i=1}^{m}\tau_{E_i}\right)\right)\tilde{U}^\dagger\right], \\ \tilde{U} &= \sum_{i=1}^{m} P_i^A \otimes U_i. \end{aligned} \tag{D30}$$

Note that $\tilde{U}$ is indeed a unitary operator and commutes with the Hamiltonian of the whole system $H_{\mathrm{all}} = H_A + H_B + \sum_i H_{E_i}$. These can be checked as follows.

$$\begin{aligned} \tilde{U}\tilde{U}^\dagger &= \sum_i \sum_j \left(P_i^A \otimes U_i\right)\left(P_j^A \otimes U_j^\dagger\right) \\ &= \sum_i \sum_j \delta_{ij} P_i^A \otimes U_i U_j^\dagger \\ &= \sum_i P_i^A \otimes I = I. \\ \left[\tilde{U}, H_{\mathrm{all}}\right] &= \sum_i [P^A \otimes U_i, H_A + H_B + H_{E_i}] \\ &= \sum_i [P^A \otimes U_i, H_A] + \sum_i [P^A \otimes U_i, H_B + H_{E_i}] = 0. \end{aligned} \tag{D31}$$

Therefore, the map defined in Eq. (D30) is a thermal operation. Eq. (D30) can be calculated as

$$\begin{aligned} \tilde{\mathcal{E}}(\rho_{AB}) &= \mathrm{Tr}_{A,E_1',\ldots,E_m'}\left[\tilde{U}\left(\rho_{AB}\otimes\left(\bigotimes_{i=1}^{m}\tau_{E_i}\right)\right)\tilde{U}^\dagger\right] \\ &= \mathrm{Tr}_{A,E_1',\ldots,E_m'}\left[\left(\sum_{i=1}^{m} P_i^A \otimes U_i(\rho_{AB})\otimes\left(\bigotimes_{i=1}^{m}\tau_{E_i}\right)\right)\left(\sum_{j=1}^{m} P_j^A \otimes U_j\right)^\dagger\right] \\ &= \mathrm{Tr}_{A,E_1',\ldots,E_m'}\left[\sum_{i=1}^{m} P_i^A \otimes U_i\left(\rho_{AB}\otimes\left(\bigotimes_{i=1}^{m}\tau_{E_i}\right)\right)\left(P_i^A \otimes U_i\right)^\dagger\right] \\ &= \sum_{i=1}^{m} \mathrm{Tr}_{A,E_i'}\left[P_i^A \otimes U_i\left(\rho_{AB}\otimes\tau_{E_i}\right)\left(P_i^A \otimes U_i\right)^\dagger\right] \\ &= \sum_{i=1}^{m} \mathrm{Tr}_{E_i'}\left[\mathrm{Tr}_A\left\{P_i^A \otimes U_i\left(\rho_{AB}\otimes\tau_{E_i}\right)\left(P_i^A \otimes U_i\right)^\dagger\right\}\right] \\ &= \sum_{i=1}^{m} \mathrm{Tr}_{E_i'}\left[U_i\left(\mathrm{Tr}_A\left[(P_i^A \otimes I_B)\rho_{AB}\right]\otimes\tau_{E_i}\right)U_i^\dagger\right], \end{aligned} \tag{D32}$$

and we obtain Eq. (D29). □

Since tracing out the subsystems and adding other thermal states are free operations, we can take the dimensions of the input and output systems equally, and we can undo this by tracing out the added systems, this procedure does not affect the extracted work (see FIG. S.2). Therefore, we can take the input and output system so that the

condition $\dim \mathcal{H}_B = \dim \mathcal{H}_C$ is satisfied, which implies that we can carry out the protocol mentioned above by a thermal operation.

Combining Proposition S.14 and Proposition S.20, we reached the following theorem.

Theorem S.21. *The asymptotic extractable work of the sequence of the i.i.d. black boxes* $\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}$ *satisfying* $|S| < \infty$ *under thermal operations is*

$$\beta W_{\text{TO}}\left(\left\{\mathcal{S}_n^{\text{i.i.d.}}(S)\right\}_{n=1}^{\infty}\right) = \min_{\rho \in S} D(\rho||\tau). \tag{D33}$$

Exploring long-range entangled states via variational LOCC-assisted circuits

Yuxuan Yan[1] * Muzhou Ma[2] * You Zhou[3] † Xiongfeng Ma[1] ‡

[1] *Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China*

[2] *Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

[3] *Key Laboratory for Information Science of Electromagnetic Waves (Ministry of Education), Fudan University, Shanghai 200433, China*

Abstract. Long-range entanglement is essential in topological orders and quantum error-correcting codes. Preparing long-range entangled states often requires polynomial depth unitary circuits, which pose significant experimental challenges. Circuits assisted by local operations and classical communication (LOCC) offer a promising avenue to reduce the required circuit depth substantially. But developing such short-depth circuits for general long-range entangled states quantum states is an open question. In this work, we address this challenge using a classical-quantum hybrid approach—the LOCC-assisted variational quantum eigensolver. The algorithm accurately solves the ground state for long-range entangled models in numerical experiments. We also establish the conditions for the absence of barren plateaus, ensuring the effectiveness of our approach.

Keywords: Long-range entanglement, LOCC-assisted circuits, variational quantum circuits

Quantum systems often exhibit entanglement, but the structures in which states become entangled can vary greatly. In the asymptotic limit, which means when the number of qubits n is infinite, some quantum states cannot be prepared using finite-depth unitary circuits. These states are long-range entangled states, including the Greenberger–Horne–Zeilinger state and the toric code state. On the other hand, short-range entangled states are the remaining quantum states. Long-range and short-range entangled states represent two different quantum phases that cannot be connected by finite-depth unitary circuits [1, 2].

Long-range entangled states are of great interest and importance. For example, certain long-range entangled states are topologically ordered and can be used as resources for topological quantum memory and computation [3]. Besides, the crucial parameters of quantum error correcting codes rely on long-range entanglement [4, 5]. Thereby, the significance of long-range entanglement in quantum information, quantum computation, and condensed matter physics is closely related.

Important as they are, the polynomial depth requirements severely challenge the experimental preparation of long-range entanglement under geometric locality constraints [6, 7, 8]. Fortunately, a promising method is found by introducing *mid-circuit measurements* to help information spread faster and significantly reduce the depth of the circuit. This process utilizes circuits assisted by local operations and classical communication (LOCC) [9], which are no longer unitary due to the existence of measurements.[1] There are various LOCC-assisted protocols for a great variety of models with topological order [10, 11, 12, 13, 14, 15]. And the feasibility of LOCC-assisted circuits is experimentally demonstrated [16, 17].

Beyond previous success in the aforementioned specific cases, how to prepare more general long-ranged entangled systems via shallow LOCC-assisted circuits is an open question. In essence, we need to perform optimization over various LOCC-assisted circuits. While variational quantum algorithms can solve problems without LOCC [18], the same cannot be said for LOCC-assisted circuits. This problem becomes much more challenging due to the added complexity of classical communications and controls involved in LOCC, which is a prerequisite for exploring general-structured long-range entanglement.

In this work, we propose a quantum-classical hybrid algorithm to tackle this challenge, named LOCC-assisted variational quantum eigensolver (LOCC-VQE). The algorithm calculates the gradients of classical control parameters and allows for a flexible design of classical control protocol. For example, this protocol can be in the form of look-up tables or neural networks. To demonstrate the accuracy of our algorithm, we have chosen long-range entangled models and numerically solved their ground state. We also theoretically discuss the condition for the absence of barren plateaus, which is a common trainability problem for variational quantum algorithms. Compared to other variational quantum algorithms, mid-circuit measurement is introduced for the first time. This merit opens new opportunities for exploring long-range entangled states in general forms and reducing depths in other variational quantum algorithm scenarios.

Variational LOCC-assisted quantum circuits.— Here, we first revisit the definition of LOCC-assisted circuits [9], and propose the variational LOCC-assisted

*These authors contributed equally to this work.

†you_zhou@fudan.edu.cn

‡xma@tsinghua.edu.cn

[1] The model of circuits assisted by LOCC is known by various names, such as adaptive circuits, dynamical circuits, and circuits with measurements and feed-forward. Although these terms are conceptually close, we will use the term "LOCC-assisted" in this work to highlight the significance of classical communications.

quantum circuits, illustrated in FIG. 1.

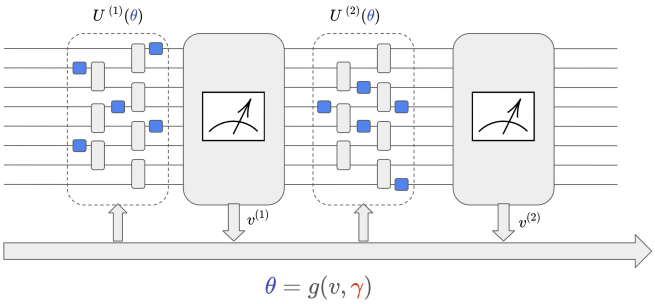


Figure 1: An illustration of a variational LOCC-assisted circuit. The circuit consists of gate layers, with blue blocks denoting Pauli rotation gates, and measurement layers, alternatively arranged. The arrows show the communication of measurement outcomes, which are processed by a classical computer.

Definition 1 (LOCC-assisted circuits)**.** *Starting from the initial state* $|\Psi_0\rangle$*, we alternatively apply unitaries or measurements. Assumed that the outcomes are* $\mathbf{v} = \{v_j\}$*, the unnormalized outcome state will be*

$$\left|\tilde{\Phi}_{\mathbf{v}}\right\rangle = U_{\mathbf{v}}^{(d)}\Pi_{\mathbf{v}}^{(d-1)}\cdots\Pi_{\mathbf{v}}^{(1)}U^{(1)}\left|\Psi_0\right\rangle. \tag{1}$$

Here, U *denotes unitaries, and* Π *denotes measurement projectors. Unitaries* $U_{\mathbf{v}}^{(i)}$ *may depend on earlier measurement outcomes corresponding to projectors* Π_j *for* $j < i$*. The LOCC-assisted, on average, will generate the following state:*

$$\Psi = \sum_{\mathbf{v}} \left|\tilde{\Phi}_{\mathbf{v}}\right\rangle\left\langle\tilde{\Phi}_{\mathbf{v}}\right|. \tag{2}$$

The variational LOCC-assisted circuits are constructed by introducing tunable parameters to the classical control protocols, which decide unitaries based on earlier measurement outcomes.

Definition 2 (Variational LOCC-assisted circuits)**.** *The variational LOCC-assisted circuits are defined by a classical control function* g *and classical control parameters* γ*. By feeding parameters* γ *and mid-circuit measurement outcome* $\mathbf{v}$ *into the function, we get the Pauli rotation angles in the circuit as* $\theta = g(\gamma, \mathbf{v})$*.*

The state generated from a variational LOCC-assisted circuit is denoted by Ψ_γ, which can be expressed as a mixture of different measurement outcomes,

$$\Psi_\gamma = \sum_{\mathbf{v}} P_\theta(\mathbf{v})\Phi_{\theta,\mathbf{v}}, \tag{3}$$

where $\Phi_{\theta,\mathbf{v}}$ are normalized post-selected states and $P_\theta(\mathbf{v})$ is the corresponding probability.

Gradient estimation protocol.— To efficiently find the optimal γ for the variational LOCC-assisted circuits, we need to calculate the gradients $\nabla_\gamma \operatorname{Tr}\left[\hat{O}\Psi_\gamma\right]$, where $\hat{O}$ is the problem-specified observable, e.g. the Hamiltonian, and employ gradient-based optimization. The following proposition expresses these gradients, which can be estimated in a classical-quantum hybrid way.

Proposition 3 (Quantum gradients for variational LOCC-assisted circuits)**.** *The gradients of a variational LOCC-assisted circuit can be obtained by*

$$\frac{\partial \operatorname{Tr}\left[\hat{O}\Psi_\gamma\right]}{\partial\gamma_j} = \sum_{i,\mathbf{v}} \frac{1}{2}\frac{\partial g_i(\gamma,\mathbf{v})}{\partial\gamma_j}\Big(\Big(P_\theta(\mathbf{v})\operatorname{Tr}\left[\hat{O}\Phi_{\theta,\mathbf{v}}\right]\Big)\Big|_{\theta=g_{i+}(\gamma,\mathbf{v})} - \Big(P_\theta(\mathbf{v})\operatorname{Tr}\left[\hat{O}\Phi_{\theta,\mathbf{v}}\right]\Big)\Big|_{\theta=g_{i-}(\gamma,\mathbf{v})}\Big), \tag{4}$$

where $\mathbf{e}_i$ *is the unit vector for single parameter shift on* θ_i *and* $g_{i\pm}(\gamma,\mathbf{v}) = g(\gamma,\mathbf{v}) \pm \frac{\pi}{2}\mathbf{e}_i$*.*

To estimating Eq. (4), we propose a classical-quantum hybrid approach:

1. On a quantum computer, for each θ_i, estimate contribution to the expectation value with parameter shifts from different mid-circuit measurement outcome $\mathbf{v}$.
2. On a classical computer, for each θ_i, γ_j, and sampled $\mathbf{v}$, calculate $\frac{\partial g_i(\gamma,\mathbf{v})}{\partial\gamma_j}$ from automatic differentiation.
3. Reweight contribution from different outcome $\mathbf{v}$ by $\frac{\partial g_i(\gamma,\mathbf{v})}{\partial\gamma_j}$.

The specific algorithm follows, where $|\cdot|$ denotes the parameter vector length, and $\mathbf{e}_i$ denotes the i-th unit vector:

Algorithm 1: Gradient estimation protocol

Data: Observable $\hat{O}$; ansatz Ψ_γ defined by $g(\gamma,\mathbf{v})$; estimation sample rounds M.
Result: Estimated gradient $\{G_j\}_{j=1,\cdots,|\gamma|}$.

```
for i ← 1 to |θ| do
    g_{i±}(γ, v) ← g(γ, v) ± (π/2) e_i;
    C_{i+} ← ∅;
    for k ← 1 to M do
        Run the LOCC-assisted circuit using g_+;
        /* quantum computer */
        v ← mid-circuit measurement results;
        c ← one-shot estimation of Ô using g_+;
        Add the pair, (v, c), to C_{i+};
    Do the same procedure to get C_{i−} from g_−;
for j ← 1 to |γ| do
    G_+ ← 0;
    G_− ← 0;
    for i ← 1 to |θ| do
        for (v, c) ∈ C_{i+} do
            G_+ ← G_+ + (1/2) ∂g_i(γ,v)/∂γ_j c;
        for (v, c) ∈ C_− do
            G_− ← G_+ + (1/2) ∂g_i(γ,v)/∂γ_j c;
    G_j ← 1/(M|θ|) (G_+ − G_−);
```

In our proposed algorithm, sample data $\mathcal{C}_{i\pm}$ are reused to estimate gradients for various γ_j. Therefore, the sample complexity of our algorithm is the same as variational

quantum algorithms with unitary circuits, which depends on the number of tunable Pauli rotations in the circuit.

Conditions for the absence of barren plateaus.—With LOCC-VQE, we can prepare long-range entanglement using short-depth quantum circuits. Further, we find that the short depth of the variational LOCC-assisted circuits is crucial for ensuring the trainability of the LOCC-VQE protocol. With an additional condition for the classical control protocols, we show the absence of barren plateaus [**?**], i.e., the aforementioned gradients will not vanish exponentially as the number of qubits scales.

Proposition 4. *The following conditions can ensure the absence of barren plateaus:*

- $\mathcal{A}1$. *Observables are local.—The support of $\hat{O}$ has constant size, i.e., $|\mathrm{supp}(\hat{O})| = \mathcal{O}(1)$.*
- $\mathcal{A}2$. *The circuit depth is constant.*
- $\mathcal{A}3$. *The gradient of the classical function g will not vanish as the size of its input increases.*
- $\mathcal{A}4$. *Each classical protocol parameter γ_j controls a constant number of Pauli angles.—Each γ_j in function g has a sparse support of size $\mathcal{O}(1)$.*

Here, the first two conditions imply the absence of barren plateaus in variational quantum circuits without LOCC assistance, which is considered a special case of LOCC-assisted circuits. The last two conditions concern the additional classical protocol introduced in this work. The third condition is a natural condition for most gradient-based optimization algorithms, and the last condition holds for many classical protocols, such as lookup tables or a fully connected or convolutional neural network layer. In this work, we ensure that the LOCC-assisted ansatzes satisfy the conditions necessary for trainability.

Numerical experiments.—To test the performance of LOCC-VQE, we numerically find the LOCC-assisted circuits to prepare ground states of various long-range entangled systems. Meanwhile, we perturbed these Hamiltonian with magnetic fields of varied perturbation strengths to test the robustness and flexibility of our algorithm.

Due to the limited length of this abstract, we only show the results of the perturbed rotated surface code, a quantum error-correcting code defined on a two-dimensional rectangular lattice with open boundary conditions. More numerical results will be provided in the arXiV version, which will be released soon. We denote the perturbation intensity as λ, and consequently, the perturbed surface code Hamiltonian is:

$$H_{sur}(\lambda) = -(1-\lambda)\sum_{v} A_v - (1-\lambda)\sum_{p} B_p - \lambda \sum_{i=1}^{2N_xN_y} Z_i. \quad (5)$$

where N_x and N_y are the numbers of rows and columns of independent the regular lattice, and A_v and B_p are stabilizers for the unperturbed surface code, shown in Fig. 2. The λ is the perturbation intensity, and Z_i's are the Pauli Z operators on each site for the magnetic field.

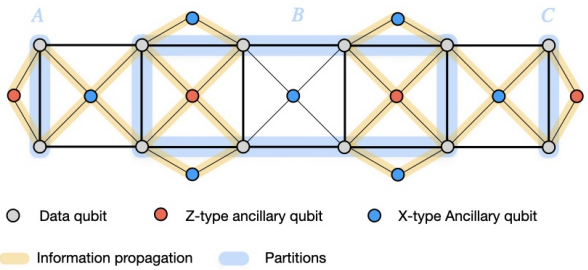


Figure 2: The rotated surface code. For every vertex v in the lattice, Z-type stabilizers, denoted as A_v, operate on the four incident edges with Z operators, while X-type stabilizers, denoted as B_p, act on the boundary of each plaquette p with X operators.

When the perturbation is small enough, ground states of the above Hamiltonian possess long-range entanglement. We adopt a finite size, $N_x = 6$, and $N_y = 2$, in our experiments. Theoretically, with four layers of local two-qubit unitary gates, creating long-range entanglement between the left and right sides is impossible [16]. Our LOCC-VQE finds four-layer LOCC-assisted circuits to approximately prepare the ground states with long-range entanglement, demonstrating energy accuracy advantages over unitary circuit ansatzes, shown in Fig. 3.

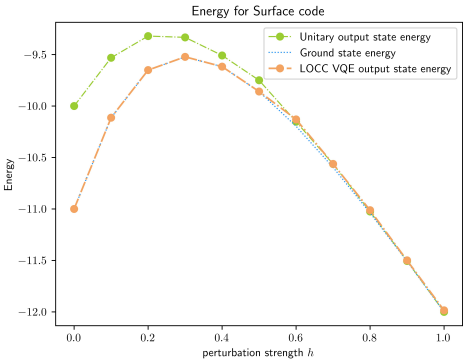


Figure 3: Optimization results for LOCC-VQE on perturbed surface code. Energy optimization results from our four-layer LOCC-VQE circuits and four-layer brick-wise unitary circuits. We use the same training setting, where iterations are sufficient for the optimization for unitary ansatz to converge. Our LOCC-VQE is significantly advantageous over unitary circuits, especially when the perturbation intensity is small and long-range entanglement dominates.

To summarize, we show the advantages of variational LOCC-assisted circuits over their unitary counterparts. Meanwhile, we find that the numerical simulation to demonstrate LOCC-VQE is expensive when the qubit number scales up. We have made efforts by using the current state-of-the-art simulation technique based on tensor networks [19] with massive parallelization. Currently, our results contain up to 20 qubits. We are still improving the numeric size to give a stronger demonstration of LOCC-VQE's depth advantages.

References

[1] Xiao-Gang Wen. Topological Order: From Long-Range Entangled Quantum Matter to a Unified Origin of Light and Electrons. 2013:e198710.

[2] Xie Chen, Zheng-Cheng Gu, and Xiao-Gang Wen. Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order. 82(15):155138.

[3] A.Yu. Kitaev. Fault-tolerant quantum computation by anyons. 303(1):2–30.

[4] Nouédyn Baspin, Omar Fawzi, and Ala Shayeghi. A lower bound on the overhead of quantum error correction in low dimensions.

[5] Jinmin Yi, Weicheng Ye, Daniel Gottesman, and Zi-Wen Liu. Complexity and order in approximate quantum error-correcting codes.

[6] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan. Limitations of Noisy Reversible Computation.

[7] Alexander Müller-Hermes, Daniel Stilck França, and Michael M. Wolf. Relative entropy convergence for depolarizing channels. 57(2):022202.

[8] Yuxuan Yan, Zhenyu Du, Junjie Chen, and Xiongfeng Ma. Limitations of Noisy Quantum Devices in Computational and Entangling Power.

[9] Lorenzo Piroli, Georgios Styliaris, and J. Ignacio Cirac. Quantum Circuits Assisted by Local Operations and Classical Communication: Transformations and Phases of Matter. 127(22):220503.

[10] Nathanan Tantivasadakarn, Ryan Thorngren, Ashvin Vishwanath, and Ruben Verresen. Long-range entanglement from measuring symmetry-protected topological phases.

[11] Tsung-Cheng Lu, Leonardo A. Lessa, Isaac H. Kim, and Timothy H. Hsieh. Measurement as a Shortcut to Long-Range Entangled Quantum Matter. 3(4):040337.

[12] Sergey Bravyi, Isaac Kim, Alexander Kliesch, and Robert Koenig. Adaptive constant-depth circuits for manipulating non-abelian anyons.

[13] Nathanan Tantivasadakarn, Ruben Verresen, and Ashvin Vishwanath. Shortest Route to Non-Abelian Topological Order on a Quantum Processor. 131(6):060405.

[14] Yabo Li, Hiroki Sukeno, Aswin Parayil Mana, Hendrik Poulsen Nautrup, and Tzu-Chieh Wei. Symmetry-enriched topological order from partially gauging symmetry-protected topologically ordered states assisted by measurements. 108(11):115144.

[15] Nathanan Tantivasadakarn, Ashvin Vishwanath, and Ruben Verresen. Hierarchy of Topological Order From Finite-Depth Unitaries, Measurement, and Feedforward. 4(2):020339.

[16] Michael Foss-Feig, Arkin Tikku, Tsung-Cheng Lu, Karl Mayer, Mohsin Iqbal, Thomas M. Gatterman, Justin A. Gerber, Kevin Gilmore, Dan Gresh, Aaron Hankin, Nathan Hewitt, Chandler V. Horst, Mitchell Matheny, Tanner Mengle, Brian Neyenhuis, Henrik Dreyer, David Hayes, Timothy H. Hsieh, and Isaac H. Kim. Experimental demonstration of the advantage of adaptive quantum circuits.

[17] Mohsin Iqbal, Nathanan Tantivasadakarn, Thomas M. Gatterman, Justin A. Gerber, Kevin Gilmore, Dan Gresh, Aaron Hankin, Nathan Hewitt, Chandler V. Horst, Mitchell Matheny, Tanner Mengle, Brian Neyenhuis, Ashvin Vishwanath, Michael Foss-Feig, Ruben Verresen, and Henrik Dreyer. Topological Order from Measurements and Feed-Forward on a Trapped Ion Quantum Computer.

[18] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. 3(9):625–644.

[19] Shi-Xin Zhang, Jonathan Allcock, Zhou-Quan Wan, Shuo Liu, Jiace Sun, Hao Yu, Xing-Han Yang, Jiezhong Qiu, Zhaofeng Ye, Yu-Qin Chen, Chee-Kong Lee, Yi-Cong Zheng, Shao-Kai Jian, Hong Yao, Chang-Yu Hsieh, and Shengyu Zhang. TensorCircuit: A Quantum Software Framework for the NISQ Era. 7:912.

Effect of Synchronization Errors on Coherent-State Qubits

Masaki Takekoshi[1] * Shion Kitamura[1] † Tiancheng Wang[1 2] ‡
Tsuyoshi Sasaki Usuda[1] §

[1] *Graduate School of Information Science and Technology, Aichi Prefectural University, Aichi, Japan.*
[2] *Faculty of Informatics, Kanagawa University, Kanagawa, Japan.*

Abstract. In recent years, research on errors in quantum deletion/insertion channels has made significant progress. Our previous study, focused on a synchronization error, which are hypothesized to induce quantum deletion/insertion errors when employing coherent-state qubits. We demonstrated that, in the context of fundamental qubits, these errors differ from those predicted by the conventional error model. Here, we provide a comprehensive analysis of synchronization errors within the framework of the conventional error model for general coherent-state qubits. Furthermore, we quantitatively assess the discrepancies between the conventional error model and our more realistic approach, highlighting the limitations of the former in practical applications.

Keywords: synchronization error, coherent-state qubits, quantum insertion error

1 Introduction

Currently, quantum computers predominantly utilize superconducting or ion-trap technologies to create qubits. However, the technology that will prevail in the future remains uncertain. In this context, optical qubits are anticipated to be a promising candidate for scalable quantum computing.

In general, errors are inevitable for qubits. Recently, a new channel model called the quantum deletion/insertion channel [1] has been proposed, and study on codes that correct quantum deletion/insertion errors has commenced [1–7]. Quantum deletion/insertion errors are believed to result from synchronization errors occurring in the channel. In quantum states with temporal spread, such as coherent-state qubits [8,9], synchronization errors are particularly significant, making the investigation of the applicability of conventional error models highly relevant.

In our previous study, we examined synchronization error in the context of coherent-state qubits and demonstrated that, for fundamental qubits, the error differ from those predicted by the conventional error model [10]. In this paper, we consider a realistic synchronization error in superposition states based on the coherent-state qubits and demonstrate the extent to which the realistic error can be approximated by the conventional error model.

*im241006@cis.aichi-pu.ac.jp
†im233003@cis.aichi-pu.ac.jp
‡wang@kanagawa-u.ac.jp
§usuda@ist.aichi-pu.ac.jp

2 Preliminary

2.1 Quantum mechanical description of light

Light propagates through space as waves composed of electric and magnetic fields. Its quantum mechanical treatment is based on the quantization of the electromagnetic field. In this paper, we focus on a finite region of space of volume V. For instance, in quantum communication, digital information can be encoded into a light beam for transmission. In this context, it is common to consider the region defined by the beam diameter and the length occupied by a single pulse carrying digital information.

2.2 Coherent state

The coherent state is described as the quantum state of light closest to a sine wave. In general, sine waves have finite power, but infinite energy. However, because the coherent state is defined within the finite region of space under consideration, the sine wave we consider is the part confined within this region, and for which the average energy of the coherent state is finite.

As mentioned above, the region of space we focus on based on the beam diameter and the length of a single pulse carrying digital information. The length of the pulse can be converted into a duration using the speed of light. Therefore, when the beam diameter is fixed, the duration T of the pulse's presence determines the volume V of this region. The synchronization errors, which are the focus of this study, can also be interpreted as errors in estimat-

ing the duration. Given these considerations, it is possible that the quantum state discussed in this paper is only partially read out. As a first step, we discuss instances of short readout time.

3 Quantum insertion channels

A quantum insertion error occurs when unknown quantum states are inserted at specific positions due to synchronization errors, increasing the number of output states. For instance, suppose we have an input quantum state $|\psi_1\rangle|\psi_2\rangle|\psi_3\rangle$, an unknown quantum state $|\phi\rangle$, and an error occurring when this state is inserted between the first and second quantum states. The input-output relationship of the channel can then be expressed as follows:

$$|\psi_1\rangle|\psi_2\rangle|\psi_3\rangle \mapsto |\psi_1\rangle|\phi\rangle|\psi_2\rangle|\psi_3\rangle. \quad (1)$$

4 Change in coherent-state qubits due to synchronization errors

To read out the temporally distributed coherent-state qubits, the qubits are sequentially and repeatedly read out one by one at a set time duration. The error model discussed here is based on a type of synchronization error in which the receiver misidentifies the temporal boundaries between states, leading to an incorrect time duration. Because coherent-state qubits are formed by temporally continuous waves, we must consider scenarios in which the receiver reads only part of a state.

In this study, we examine general qubits, which are superpositions of the fundamental qubits $\{|\alpha\rangle, |-\alpha\rangle\}$, as follows:

$$|\phi\rangle = k(c_0|\alpha\rangle + c_1|-\alpha\rangle), \quad (2)$$

where c_0 and c_1 denote superposition coefficients satisfying $|c_0|^2 + |c_1|^2 = 1$ and k denotes the normalization factor for normalizing the norm of quantum state to 1. For simplicity, we consider the state in which two fundamental qubits are equally superposed as the input state when performing specific numerical calculations.

4.1 Output for short readout time

First, consider the transmission of signal $|\phi_1\rangle|\phi_2\rangle\cdots|\phi_n\rangle$ with n instances of $|\phi\rangle$, each lasting a duration of 1. If the first quantum state is read with a duration T' shorter than 1, the state is divided into $\hat{\rho}_1^{err}$ and $\hat{\rho}_1^{err'}$ corresponding respectively to durations T' and $1-T'$, as follows:

$$|\phi_1\rangle \mapsto \hat{\rho}_1^{err} \otimes \hat{\rho}_1^{err'}. \quad (3)$$

Here, it is assumed that the synchronization, that is, the judgement of the boundaries after the second quantum state $|\phi_2\rangle$ is correct.

Note that the energy of each quantum state resulting from splitting diminishes compared with that of the original quantum state. Thus, the two output quantum states can be represented using a general attenuation model.

Denoting the original qubit of mode S by $|\phi\rangle_{\mathrm{S}}$ and the vacuum state of ancilla mode A by $|0\rangle_{\mathrm{A}}$, the state of the composite system is as follows:

$$\begin{aligned} U^{(\mathrm{S}\otimes\mathrm{A})}|\phi\rangle_{\mathrm{S}}|0\rangle_{\mathrm{A}} =& c_0|\sqrt{T'}\alpha\rangle_{\mathrm{S}}|\sqrt{1-T'}\alpha\rangle_{\mathrm{A}} \\ &+c_1|-\sqrt{T'}\alpha\rangle_{\mathrm{S}}|-\sqrt{1-T'}\alpha\rangle_{\mathrm{A}}. \end{aligned} \quad (4)$$

The quantum state $\hat{\rho}^{(\mathrm{S})}$ of the signal mode S can be obtained by performing a partial trace over the ancilla mode A.

Using this partial trace, the output $\hat{\rho}_1^{err}$ after splitting is expressed as follows:

$$\begin{aligned} \hat{\rho}_1^{err} =& c_0^2|\sqrt{T'}\alpha\rangle_{\mathrm{A}}\langle\sqrt{T'}\alpha| + 2c_0c_1L|-\sqrt{T'}\alpha\rangle_{\mathrm{A}}\langle\sqrt{T'}\alpha| \\ &+ c_1^2|-\sqrt{T'}\alpha\rangle_{\mathrm{A}}\langle-\sqrt{T'}\alpha|, \end{aligned} \quad (5)$$

where L is the inner product of $|\sqrt{1-T'}\alpha\rangle_{\mathrm{A}}$ and $|-\sqrt{1-T'}\alpha\rangle_{\mathrm{A}}$. Similarly, we can derive $\hat{\rho}_1^{err'}$.

Therefore, for a coherent-state qubit, we see that the number of outputs increases when the quantum state splits, and each output quantum state becomes a mixed state.

4.2 Calculation of the output fidelity

By calculating the fidelity of the output quantum states corresponding to the error described in the previous subsection and that given by the conventional error model of the quantum insertion channel, we demonstrate how well the realistic errors can be approximated by the conventional error model.

Here, as in the previous subsection, we consider n superposition states of coherent states, $|\phi_1\rangle|\phi_2\rangle\cdots|\phi_n\rangle$, and assume that the first state is split and synchronization becomes stable afterwards. This leads to an increase in the number of states corresponding to a quantum insertion error for which an unknown quantum state $|\phi'\rangle$ is inserted between the first and second states.

In this instance, the input/output of the former is

$$|\phi_1\rangle|\phi_2\rangle\cdots|\phi_n\rangle \mapsto \hat{\rho}_1^{err} \otimes \hat{\rho}_1^{err'} \otimes \hat{\rho}_2 \otimes \cdots \otimes \hat{\rho}_n, \quad (6)$$

where $\hat{\rho}_1^{err}$ and $\hat{\rho}_1^{err'}$ denote the states split due to the synchronization error, and $\hat{\rho}_i = |\phi_i\rangle\langle\phi_i|$ $(i =$

$2, \ldots, n$) denotes the density operator corresponding to the quantum state $|\phi_i\rangle$.

Moreover, the input/output of the latter is

$$|\phi_1\rangle|\phi_2\rangle \cdots |\phi_n\rangle \mapsto |\phi_1\rangle|\phi'\rangle|\phi_2\rangle \cdots |\phi_n\rangle. \quad (7)$$

When calculating the fidelity, only the first and second quantum states need to be considered, as the states from the third onwards are identical in both channels, with an inner product of 1.

In the conventional error model, the inserted quantum state $|\phi'\rangle$ is unknown, and various quantum states are assumed to be inserted. The quantum state considered here is described by equation (8), using coherent states as the fundamental qubits, with two degrees of freedom given by coefficients c_0' and c_1'; specifically

$$|\phi'\rangle = k(c_0'|\alpha\rangle + \exp[\mathbf{i}\theta]c_1'|-\alpha\rangle), \quad (8)$$

where k denotes the normalization factor that normalizes the quantum state $|\phi'\rangle$ to unity, and the coefficients c_0', c_1' are real numbers satisfying $c_0'^2 + c_1'^2 = 1$.

Based on the above, and as a simple example, we assume that $|\phi_1\rangle$ is split in half and calculate the fidelity of this output quantum state and $|\phi_1\rangle|\phi'\rangle$. Figure 1 shows fidelity results when $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi$. To observe the effect of varying T', Fig. 2 shows the dependence of fidelity with the coefficient of $|\phi'\rangle$ when $\theta = 0$ and $T' = \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}$. Furthermore, Fig. 3 focuses on $F(\hat{\rho}_1^{err'}, |\phi'\rangle)$ in the above results.

These results indicate that the fidelity increases when the energy difference between the two states we compared is small.

Thus, by introducing the degrees of freedom for the inserted quantum states and varying the values of T', we demonstrate that the fidelity values consistently remain low, suggesting a difficulty in approximating realistic errors using the conventional error model.

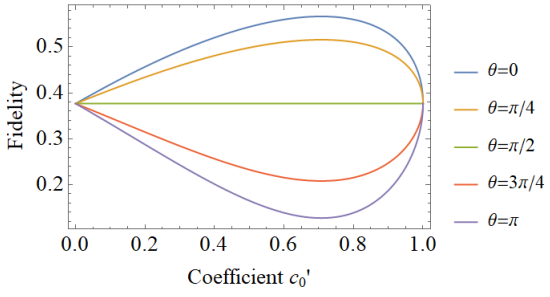


Figure 1: Fidelity with respect to the coefficient of $|\phi'\rangle$ when $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi$ and $T' = \frac{1}{2}$.

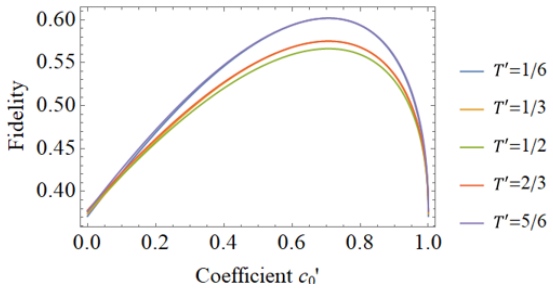


Figure 2: Fidelity with respect to the coefficient of $|\phi'\rangle$ when $T' = \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}$ and $\theta = 0$.

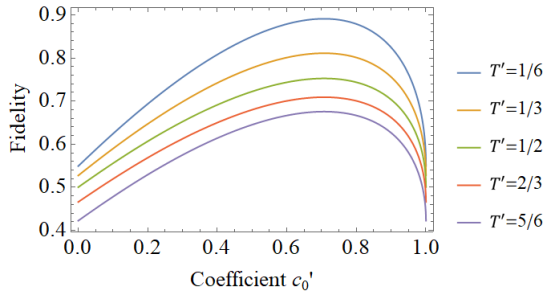


Figure 3: $F(\hat{\rho}_1^{err'}, |\phi'\rangle)$ with respect to the coefficient of $|\phi'\rangle$ when $T' = \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}$ and $\theta = 0$.

5 Conclusion

In this paper, we used coherent-state qubits as a fundamental qubit $\{|\alpha\rangle, |-\alpha\rangle\}$ and examined the synchronization error that misestimates the duration of the general qubits, created by equally superposing the coherent-state qubits, as being shorter than their actual duration. Our results indicate that the original state is split and represented by the tensor product of two mixed states. Furthermore, we investigated whether the conventional error model can approximate the errors considered in this paper by calculating the fidelity. We demonstrated that the fidelity results never approach unity regardless of the inserted states, indicating a difficulty in approximating errors using the conventional error model. Future work includes exploring more generalized expressions by investigating other misestimations of qubit duration.

Acknowledgments: This work has been supported in part by JSPS KAKENHI Grant Number JP20K20397, JP20H00581, JP21K04064, and the Hibi Science Foundation.

We thank Richard Haase, Ph.D, and Liam Exelby from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

References

[1] J. Leahy, D. Touchette, and P. Yao,"Quantum insertion deletion channels," arXiv:1901.00984, (2019).

[2] A. Nakayama and M. Hagiwara, "The first quantum error-correcting code for single deletion errors," IEICE Communications Express, 9(4), pp.100-104, (2020).

[3] M. Hagiwara, "The four qubits deletion code is the first quantum insertion code," IEICE Communications Express, 10(5), pp.243-247, (2021).

[4] Y. Ouyang, "Permutation-invariant quantum coding for quantum deletion channels," 2021 IEEE International Symposium on Information Theory (ISIT), pp.1499-1503, (2021).

[5] T. Shibayama and Y. Ouyang,"The equivalence between correctability of deletions and insertions of separable states in quantum codes," 2021 IEEE Information Theory Workshop (ITW), pp.1-6, (2021).

[6] R. Matsumoto and M. Hagiwara, "Constructions of l-adic t-deletion-correcting quantum codes," IEICE Trans. on Fundamentals. E104(12), pp.1654-1664, (2021).

[7] Y. Adachi, S. Usami, and T. S. Usuda, "Towards the construction of quantum error correcting codes for multiple insertion errors," 2022 International Symposium on Information Theory and Its Applications (ISITA2022), Proc. of ISITA2022, p.219, (2022).

[8] P. T. Cochrane, G. J. Milburn, and W. J. Munro, "Macroscopically distinct quantum-superposition states as a bosonic code for amplitude damping," Phys. Rev. A59, no.4, pp.2631-2634, (1999).

[9] S. Tatsuta, T. S. Usuda, I. Takumi, M. Hata, "A study on quantum computer with Schrödinger cat states," Proc. of SITA1999, pp.487-490, (1999). (in Japanese)

[10] M. Takekoshi, S. Kitamura, T. Wang, T. S. Usuda, "On Quantum Deletion/Insertion Errors Caused by Synchronization Error" Proc. of SITA2023, pp.570-575, (2023). (in Japanese)

[11] C. W. Helstrom, "The conversion of a pure state into a statistical mixture by the linear quantum channel," Optics Communications, Vol.37, pp.175-177, (1981).

Approximation accuracy of von Neumann entropy for M-ary ASK coherent-state signals

Keisuke Goto[1] * Shion Kitamura[1] † Tiancheng Wang[1 2] ‡ Tsuyoshi Sasaki Usuda[1] §

[1] *Graduate School of Information Science and Technology, Aichi Prefectural University, Aichi, Japan.*
[2] *Faculty of Informatics, Kanagawa University, Kanagawa, Japan.*

Abstract. Calculating the von Neumann entropy for a quantum information source is a critical issue in both quantum communication and quantum cryptography. However, as the number of signals increases, the calculation becomes increasingly challenging. This difficulty arises because the computation requires solving the eigenvalue problem of the Gram matrix, the size of which is proportional to the number of signals. In this study, we propose an approximation method for eigenvalues and the von Neumann entropy for amplitude-shift-keying coherent-state signals by approximating the Gram matrix as a tridiagonal matrix. Our results demonstrate the effectiveness and practicality of this approximation.

Keywords: Quantum cryptography, Gram matrix, von Neumann entropy, ASK coherent-state signals

1 Introduction

Estimating accurately the performance limit of quantum communication is crucial for both the design of quantum communication systems and the security evaluation of quantum cryptographic systems [1]. In quantum cryptography, systems such as Y-00 [2], which is anticipated to be an ultrafast quantum cryptosystem, have recently been tested with more than 4 billion signals [3]. However, estimating the performance limit of quantum communication becomes difficult as the number of signals increases. The quantum signals examined in [4] are amplitude-shift-keying (ASK) coherent-state signals, for which the Gram matrix is approximated as an n-diagonal matrix by exploiting the exponential decay of the Gram matrix components as one moves away from the diagonal. The Gram matrix of the ASK coherent-state signal is a symmetric Toeplitz matrix, and in this approximation, for $n = 3$, the approximated matrix is a tridiagonal matrix; analytical solutions for the eigenvalues and eigenvectors of these matrices are well known [5]. This paper compares the eigenvalue distributions of the Gram matrix and its tridiagonal approximation for the ASK coherent-state signal, and discusses the factors influencing the accuracy of the approximation.

*im243004@cis.aichi-pu.ac.jp
†im233003@cis.aichi-pu.ac.jp
‡wang@kanagawa-u.ac.jp
§usuda@ist.aichi-pu.ac.jp

2 Basic Theory

2.1 The von Neumann entropy and Gram matrix

Let $\{|\psi_i\rangle \mid i = 1, 2, \ldots, M\}$ be an M-ary pure-state signal system and let ξ_i be the *a priori* probability of each $|\psi_i\rangle$. The density operator of the quantum information source corresponding to this signal system is defined as

$$\rho = \sum_{i=1}^{M} \xi_i |\psi_i\rangle\langle\psi_i|. \tag{1}$$

The von Neumann entropy is defined as

$$\chi = -\mathrm{Tr}(\rho \log_2 \rho) \tag{2}$$

for the density operator of this source. The quantity that maximizes the von Neumann entropy with respect to $\{\xi_i\}$ is the quantum channel capacity, which defines the transmission limit of quantum communication. For simplicity, we assume in this paper that the *a priori* probabilities of the signals are uniform. The Gram matrix Γ for an M-ary pure-state signal system $\{|\psi_i\rangle \mid i = 1, 2, \ldots, M\}$ is the $M \times M$ matrix with the inner product $\langle\psi_i|\psi_j\rangle$ between signal quantum states as the (i, j)-th component:

$$\Gamma = \begin{bmatrix} \langle\psi_1|\psi_1\rangle & \langle\psi_1|\psi_2\rangle & \cdots & \langle\psi_1|\psi_M\rangle \\ \langle\psi_2|\psi_1\rangle & \langle\psi_2|\psi_2\rangle & \cdots & \langle\psi_2|\psi_M\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\psi_M|\psi_1\rangle & \langle\psi_M|\psi_2\rangle & \cdots & \langle\psi_M|\psi_M\rangle \end{bmatrix}. \tag{3}$$

The Gram matrix is a positive semi-definite Hermitian matrix and is closely related to the theory of square-root measurement (SRM), known as a quasi-optimal measurement. Furthermore, it is extremely important that the density operators ρ and $\frac{1}{M}\Gamma$ of the corresponding information source are isomorphic when the *a priori* probabilities of the signal quantum states are equal $\frac{1}{M}$.

Therefore, since the eigenvalues of ρ and $\frac{1}{M}\Gamma$ coincide, the von Neumann entropy is determined by the eigenvalue λ_j of $\frac{1}{M}\Gamma$ as

$$\chi = \sum_j \lambda_j \log_2 \lambda_j. \tag{4}$$

2.2 ASK coherent-state signals

Coherent states are known as quantum states of light that are very close to sinusoidal waves and are defined as

$$|\beta\rangle = \exp\left(-\frac{1}{2}|\beta|^2\right)\sum_{n=0}^{\infty}\frac{\beta^n}{\sqrt{n!}}|n\rangle, \tag{5}$$

where β denotes a complex amplitude, $|n\rangle$ a number state with exactly n photons, and the set $\{|n\rangle\}$ forms an orthonormal basis of the infinite-dimensional Hilbert space representing the optical system. ASK coherent-state signals are signals for which the amplitude is modulated by digital information. In this paper, we assume the complex amplitudes of coherent states are real. The complex amplitudes of M-ary ASK coherent-state signals are given by

$$\alpha_i = \left(i - \frac{M+1}{2}\right)\alpha \quad (i = 1, 2, \ldots, M), \tag{6}$$

where α is the amplitude difference from the adjacent signal and α_i takes values from $-\frac{1}{2}(M-1)\alpha$ to $\frac{1}{2}(M-1)\alpha$. The set of M-ary ASK coherent-state signals is represented as $\left\{|\alpha_i\rangle \;\middle|\; i = 1, 2, \ldots, M\right\}$. The inner product of the coherent-state signals $|\alpha_i\rangle$ and $|\alpha_j\rangle$ is given by

$$\langle\alpha_i|\alpha_j\rangle \;=\; \exp\left[-\frac{1}{2}\alpha^2(i-j)^2\right]. \tag{7}$$

2.3 Gram matrix and von Neumann entropy approximation

In this section, we first describe the properties of the Gram matrix of the ASK coherent-state signal. Then, we present the Gram matrix and the von Neumann entropy approximation. Finally, we introduce evaluation metrics for assessing the accuracy of the approximation quantitatively.

As in Eq. (7), the inner product of the M-ary ASK coherent-state signal depends on the difference $i-j$. Thus, the Gram matrix of the M-ary ASK coherent-state signal is a symmetric Toeplitz matrix,

$$\Gamma = \begin{bmatrix} g_1 & g_2 & \cdots & g_M \\ g_2 & g_1 & \cdots & g_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ g_M & g_{M-1} & \cdots & g_1 \end{bmatrix}. \tag{8}$$

Moreover, from the exponential decay property of the inner product, Eq. (7), we have

$$1 = g_1 > g_2 > \cdots > g_M > 0 \tag{9}$$

when $\alpha \neq 0$, implying the further away a component of the Gram matrix is from the diagonal the smaller its value. In this paper, we approximate the Gram matrix as a tridiagonal matrix for which all but the main diagonal and the two subdiagonals are approximated to zero. Let $\tilde{\Gamma}_3$ be the approximate Gram matrix. The eigenvalues $\tilde{\lambda}_j$ $(j = 1, \ldots, M)$ of $\frac{1}{M}\tilde{\Gamma}_3$ are known to have the following form;

$$\tilde{\lambda}_j = \frac{1}{M}\left\{1 + 2g_2\cos\left(\frac{j\pi}{M+1}\right)\right\}. \tag{10}$$

It is expected that $\tilde{\lambda}_j$ approximates the eigenvalue λ_j of $\frac{1}{M}\Gamma$—that is, $\tilde{\lambda}_j \approx \lambda_j$—thus providing an approximation of the von Neumann entropy (4).

2.4 Evaluation factor of eigenvalue approximations

Because the von Neumann entropy yields a single real value, its approximation can be easily evaluated by direct comparison. In this paper, we also provide approximate values of the original eigenvalues. Since the matrix $\frac{1}{M}\Gamma$ is isomorphic to the density operator of the source, its eigenvalues are non-negative real numbers that sum to 1. Therefore, the eigenvalue distribution $\{\lambda_j\}$ can be regarded as a probability distribution. Although there is no guarantee that the approximate eigenvalue distribution $\{\tilde{\lambda}_j\}$ can be regarded as a probability distribution, it is expected to be approximately so. For this reason, as an approximate evaluation of the eigenvalue distribution, we use the variation distance, which measures the closeness of two probability distributions,

$$d = \sum_{j=1}^{M}\left|\lambda_j - \tilde{\lambda}_j\right|. \tag{11}$$

From the definition, $d \geq 0$ and $d = 0$ if and only if the eigenvalue distributions coincide. If $\{\tilde{\lambda}_j\}$ satisfies properties of a probability distribution, then

$d \leq 1$. Therefore, when d is significantly less than 1 and close to 0, the approximation is considered good.

3 Result

First, we estimate approximate values for the eigenvalues. Since the true values of the eigenvalues are non-negative, we will consider instances for which Eq. (10) is always non-negative. This requires the condition

$$1 + 2g_2 \cos\left(\frac{j\pi}{M+1}\right) \geq 0$$

to hold. Noting $-1 \leq \cos\theta \leq 1$, it suffices that

$$g_2 = \exp\left[-\frac{1}{2}\alpha^2\right] \leq \frac{1}{2}.$$

Therefore, we require

$$\alpha \geq \sqrt{2\log_e 2} \approx 1.17741 \tag{12}$$

to be satisfied.

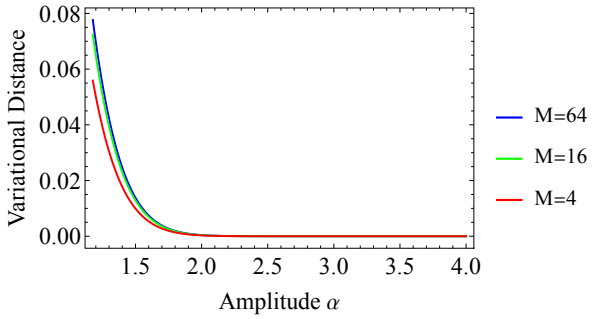


Figure 1: Variation distance of distribution of eigenvalues and their approximate values (linear scale).

Figure 1 plots the variation distance of the distribution of eigenvalues and their approximations against the amplitude difference α from the adjacent signals. The number of signals is set to $M = 4$,16, and 64. From Fig. 1, we observe that the variation distance monotonically decreases with increasing α, and the variation distance is nearly the same for any number of signals. This suggests that the variation distance, or approximation accuracy, is almost independent of the number of signals and depends primarily on the amplitude difference α from the adjacent signals. In other words, even with a very large number of signals, high approximation accuracy can be expected if α is appropriately chosen.

Figure 2 plots the von Neumann entropy and its approximation with respect to the amplitude difference α from the adjacent signal. The black line

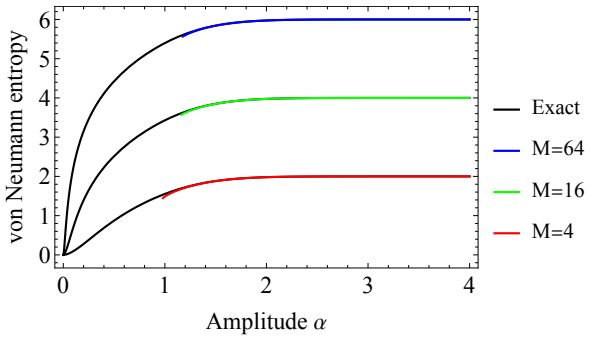


Figure 2: The von Neumann entropy and its approximation.

represents the von Neumann entropy, whereas the colored line indicates the approximate value. It is evident that the approximation of the von Neumann entropy is accurate for $\alpha \geq 1.17741$, where the approximate eigenvalue is always positive for each number of signals, $M = 4$,16, and 64.

4 Conclusion

In this paper, we considered approximating the Gram matrix as a tridiagonal Toeplitz matrix for M-ary ASK coherent-state signals and investigated the approximation accuracy of the eigenvalue distribution of the Gram matrix and that of the von Neumann entropy. The results show that the approximation accuracy is almost independent of the number of signals M and is determined primarily by the amplitude difference between adjacent signals. From this result, it is expected that high approximation accuracy can be obtained even when the number of signals is very large and α is obtained appropriately. Since analytical solutions of eigenvalues and eigenvectors are known for the tridiagonal Toeplitz matrix, it is straightforward to compute eigenvalues and the von Neumann entropy approximations even when the number of signals is very large. Therefore, the results of this paper demonstrate that the use of approximate values is effective for investigating the performance of large-scale quantum communications or quantum cryptography.

Acknowledgments: This work has been supported in part by JSPS KAKENHI Grant Number JP20K20397, JP20H00581, JP21K04064, and the Hibi Science Foundation.

We thank Richard Haase, Ph.D, and Liam Exelby from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

References

[1] C. W. Helstrom: *Quantum detection and estimation theory*, Academic Press, New York, (1976).

[2] H. P. Yuen: "KCQ: A new approach to quantum cryptography I. General principles and key generation," arXiv:quant-ph/0311061v6, (2004).

[3] X. Chen, K. Tanizawa, P. Winzer, P. Dong, J. Cho, F. Futami, K. Kato, A. Melikyan, and K. W. Kim: "Experimental demonstration of 4,294,967,296-QAM based Y-00 quantum stream cipher template carrying 160-Gb/s 16-QAM signals," Opt. Express, **29**, 5658, (2021).

[4] K. Goto, S. Kitamura, T. Wang, and T. S. Usuda, "Approximation of error probability using n-diagonal matrix similar to the Gram matrix for ASK coherent state-signals," Proc. of SITA2023, pp.409-413, (2023). (in Japanese)

[5] C. D. Meyer, *Matrix analysis and applied linear algebra*, SIAM, USA, (2004).

Symmetric and asymmetric strategies for Bell-inequality violation

Hsin-Yu Hsu[1] Gelo Noel M. Tabia[2] Kai-Siang Chen[1] Bo-An Tsai[1]
Yeong-Cherng Liang[1 3 *]

[1] *Department of Physics and Center for Quantum Frontiers of Research & Technology (QFort), National Cheng Kung University, Tainan 701, Taiwan*
[2] *Foxconn Quantum Computing Research Center, Taipei, Taiwan*
[3] *Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan*

Abstract. In quantum information, asymmetry, i.e., the lack of symmetry, is a resource allowing one to accomplish certain tasks that are otherwise impossible. In a Bell test using any given Bell inequality, the maximum violation achievable using quantum strategies that respect or disregard a certain symmetry can be different. When a gap is present, a quantum violation beyond the symmetric bound immediately witnesses the asymmetry in the underlying quantum strategies. Here, we focus on the symmetry of permutation invariance possessed by identical quantum particles. For Bell scenarios with binary inputs, we provide evidence showing that the family of symmetric Collins-Gisin-Linden-Massar-Popescu inequalities can always be maximally violated by symmetric quantum strategies minimal in the Hilbert space dimension.

Keywords: Quantum nonlocality, quantum correlation, symmetry, device-independent witness, Bell inequality, quantum bound

1 Introduction

Symmetry plays a fundamental role in physics; for instance, the indistinguishability of identical particles has many important consequences in quantum theory. Symmetry conditions are also useful in those cases where they are broken, as these have often been indicative of some interesting new phenomena.

From the violation of a Bell inequality, we know that correlations derived from entangled states may give nonlocal properties inconsistent with locally-causal theories [1, 2]. Consider a bipartite Bell experiment where two spatially separated parties, Alice and Bob, share a quantum state ρ_{AB} acting on the Hilbert space $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$. Moreover, let $\{M^{(A)}_{a|x}\}$ and $\{M^{(B)}_{b|y}\}$ be, respectively, the positive-operator-valued measure (POVM) corresponding to Alice and Bob's x-th ($x \in \{1, 2, \cdots, m_a\}$) and y-th ($y \in \{1, 2, \cdots, m_b\}$) measurement, with outcomes labeled by $a \in \{1, 2, \cdots, n_a\}$ and $b \in \{1, 2, \cdots, n_b\}$. According to Born's rule, the local measurement outcomes follow the joint conditional probability distribution: $P(a, b|x, y) = \mathrm{tr}(\rho_{AB} M^{(A)}_{a|x} \otimes M^{(B)}_{b|y})$. We refer to the collection of these distributions as a correlation $\vec{P} = \{P(a, b|x, y)\}_{a,b,x,y}$.

Below, we first provide some relevant definitions in the context of a Bell experiment. Throughout, we use the term symmetric to be mean party-permutation invariance.

Definition 1 *A symmetric quantum correlation $\vec{P}$ is one that satisfies Born's rule and*

$$P(a, b|x, y) = P(b, a|y, x), \quad \forall\ a, b, x, y. \tag{1}$$

Definition 2 *A symmetric quantum strategy is one where (i) Alice and Bob share a permutation-invariant state $\rho_{AB} = \rho_{BA}$, and (ii) they perform the same local measurements, i.e., $M^{(A)}_{a|x} = M^{(B)}_{a|x}$ for all a, x. Clearly, a symmetric quantum strategy entails a symmetric quantum correlation since*

$$\begin{aligned} P(a, b|x, y) &= \mathrm{tr}(\rho_{AB} M^{(B)}_{a|x} \otimes M^{(A)}_{b|y}) \\ &= \mathrm{tr}(\rho_{BA} M^{(B)}_{a|x} \otimes M^{(A)}_{b|y}) = P(b, a|y, x). \end{aligned} \tag{2}$$

By the same token, we say that a Bell inequality $I_{\vec{\beta}} : \sum_{a,b,x,y} \beta^{x,y}_{a,b} P(a, b|x, y) \overset{\mathcal{L}}{\leq} B$ is symmetric if $\beta^{x,y}_{a,b} = \beta^{y,x}_{b,a}$ for all a, b, x, y. Clearly, not all Bell inequalities take a symmetric form. However, via an appropriate relabeling [3] of the measurement settings x, y and/ or outcomes a, b, it may be possible to recast a Bell inequality in a symmetric form. It is known [4, 5] that the maximal violation of a symmetric Bell inequality is always achievable using a symmetric strategy.

2 Technical work

Next, we present a symmetric quantum strategy for the quantum violation of the Collins-Gisin-Linden-Massar-Popescu (CGLMP) Bell inequality I_d [6], relevant to a Bell scenario with two parties, each performing two d-outcome measurements. To facilitate subsequent discussions, we first write the CGLMP Bell inequality in a symmetric form (via appropriate relabeling) as follows:

$$\begin{aligned} I_d = \sum_{k=0}^{\lfloor \frac{d}{2} \rfloor - 1} \left(1 - \tfrac{2k}{d-1}\right) \Big\{ & \\ \big[P(A_1 = d - B_1 - k) + P(B_2 = d - A_1 + k) & \\ + P(B_1 = d - A_2 + k) + P(A_2 = d - B_2 - k - 1)\big] & \\ - \big[P(A_1 = d - B_1 + k + 1) + P(B_2 = d - A_1 - k - 1) & \\ + P(B_1 = d - A_2 - k - 1) + P(A_2 = d - B_2 + k)\big] \Big\} & \overset{\mathcal{L}}{\leq} 2 \end{aligned} \tag{3}$$

where $P(A_x = d - B_y - k) \equiv \sum_{a,b} P(a, b|x, y) \delta_{a,d-b-k}$ and all additions in the arguments of P are understood to be taken modulo d. In the simplest case of $d = 2$, inequality (3)

*ycliang@mail.ncku.edu.tw

is equivalent to the Clauser-Horne-Shimony-Holt (CHSH) [7] Bell inequality:

$$I_{\rm CHSH} = \langle A_1 B_1\rangle + \langle A_1 B_2\rangle + \langle A_2 B_1\rangle - \langle A_2 B_2\rangle \leq 2 \quad (4)$$

and whose optimal quantum strategy is not usually presented in a symmetrical form. Yet, we can easily verify that if Alice and Bob share

$$|\psi\rangle_{AB} = \frac{1}{2\sqrt{2-\sqrt{2}}}\left[|00\rangle + (1-\sqrt{2})(|01\rangle + |10\rangle) - |11\rangle\right] \quad (5)$$

and both measure σ_z for their first measurement and σ_x for their second measurement, then they do attain the maximal CHSH violation of $I_2 = 2\sqrt{2}$.

Next, we describe how to construct a symmetric strategy that attains the same CGLMP Bell violation as that given by the known optimal asymmetric strategy (see, e.g., [8]). To turn the asymmetric strategy into a symmetric one, we keep Alice's optimal measurement settings from [8] and set Bob's one to be the same, i.e., $M^{(B)}_{a|x} = M^{(A)}_{a|x}$. Then, we solve the largest eigenvalue of the corresponding Bell-operator [9]. The resulting optimal symmetric strategy may be specified as $M^{(A)}_{a|1} = |a\rangle\langle a|$, $M^{(A)}_{a|2} = |\tilde{a}\rangle\langle\tilde{a}|$ with $(|\tilde{a}\rangle)_i = (U)_{ia}$, $U = TW$, and

$$\begin{aligned}(T)_{ij} &= \begin{cases} (-1) & , 2 \leq i \leq j \\ 1 & \text{otherwise} \end{cases}, \\ (W)_{jk} &= \frac{1}{d}\left|\frac{1}{\sin\left[(j-k-\frac{1}{2})\frac{\pi}{d}\right]}\right|. \end{aligned} \quad (6)$$

As an explicit example, note that the optimal corresponding eigenstate optimizing the eigenvalue of the Bell operator, and hence the quantum violation of I_3 is:

$$\begin{aligned}|\psi_3\rangle = \frac{1}{3}\gamma\Bigg[&\frac{5-\sqrt{33}}{2}|00\rangle + (|01\rangle + |10\rangle) \\ &+ \frac{-7+\sqrt{33}}{\sqrt{2}}(|02\rangle + |20\rangle - |11\rangle) \\ &- \frac{5-\sqrt{33}}{2}(|12\rangle + |21\rangle) + |22\rangle\Bigg]. \end{aligned} \quad (7)$$

with $\gamma = 2\sqrt{\frac{2}{55-9\sqrt{33}}}$. The above computation suggests a general procedure for "symmetrizing" the known optimal CGLMP measurements and we have found that it recovers the best known quantum violation of CGLMP for $4 \leq d \leq 15$.

Given the above observation, one may wonder whether the maximal quantum violation of a Bell inequality can always be achieved using a symmetric strategy. For the Bell scenarios with two binary-outcome measurement (i.e., the CHSH scenario), two ternary-outcome measurements (i.e., the scenario for I_3), and three binary-output measurements [3], this is known to be *impossible* for facet-defining Bell inequalities — they can all be cast in a symmetric form via relabeling. Naturally, we thus explore the Bell scenario involving four measurement settings, each with binary measurement outcomes ($a_i, b_j \in \{+1, -1\}$) for which the facet-defining Bell inequalities are completely described in [10]:

$$B_{4422} = \delta + \sum_{i=1}^{4}\alpha_i\langle a_i\rangle + \sum_{j=1}^{4}\beta_j\langle b_j\rangle + \sum_{i,j=1}^{4}\gamma_{ij}\langle a_i b_j\rangle \overset{\mathcal{L}}{\leq} 0, \quad (8)$$

which one may also present equivalently using the table

$$B_{4422} = \begin{array}{c|cccc} \delta & \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \hline \alpha_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \alpha_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \alpha_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \\ \alpha_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{array} \overset{\mathcal{L}}{\leq} 0, \quad (9)$$

Note that in this four-setting two-outcome Bell scenario, hereafter abbreviated as the 4422 Bell scenario, the Bell polytope is completely specified by 175 classes of facet-defining Bell inequalities, one of which is the trivial positivity facet, taking the form of $P(a, b|x, y) \geq 0$. For the others, by systematically considering all possible relabelings, we came to the following observation.

Observation 1 *Among all 174 classes of nontrivial facet-defining Bell inequalities in the 4422 Bell scenario, 54 of them can be cast in a symmetric form.*

In general, we are interested to know whether the requirement of symmetry results in any nontrivial consequences on Bell violation. To this end, we can incorporate the constraints of Eq. (1) into any of the hierarchies of semidefinite programming (SDP) outer approximations for the quantum set of correlations. For instance, the symmetry constraints can be easily included in the SDP hierarchy due to Navascués, Pironio, and Acín (NPA) [11, 12] or the one due to Moroder *et al.* [13]. Using such a modified hierarchy, we can upper bound the maximal quantum violation of any Bell inequality by symmetric quantum violation. A systematic investigation has led to the following observation.

Observation 2 *For all 174 classes of nontrivial facet-defining Bell inequalities in the 4422 Bell scenario, there is at least one relabeled version from each class that cannot be violated by symmetric quantum correlations.*

In particular, for three of these, see Appendix C, we find that their symmetric quantum bound even coincides with the Bell-local bound, even though a different relabeling of them can again be violated by symmetric qubit strategies.

Returning to the symmetric Bell inequalities in the 4422 scenario, we have found that 35 of them can be maximally violated using symmetric qubit strategies (these are listed in Tables 3 to 5 among others that *cannot* be cast in the symmetric form, but whose symmetric bound matches the qubit bound). However, 11 other of these inequalities, which we know can be maximally violated using qubit strategies—based on our numerical investigations using Eq. (19)—apparently *cannot* be violated maximally using *symmetric* qubit strategies. Using the notations of [10], these are I^4_{4422}, I^{13}_{4422}, I^{14}_{4422}, I^{15}_{4422}, I^{16}_{4422}, J^{18}_{4422}, J^{29}_{4422}, J^{42}_{4422}, J^{86}_{4422}, J^{97}_{4422}, and J^{113}_{4422}. Interestingly, for the remaining 8 symmetric inequalities, which are

known to be violated maximally using only high-dimensional quantum strategies, we also do not find matching maximally-violating symmetric strategies of minimal dimension.

Finally, let us note that for all 175 classes of these facet-defining Bell inequalities, we have always found at least one (relabeled) version of it where symmetric quantum correlations (apparently) fail to give the maximal quantum violation, as indicated by a gap in the SDP bound obtained with and without imposing the symmetric constraint of Eq. (1). For some of these inequalities, see, e.g., Table 4, the gap between the general quantum bound and the symmetric quantum bound can be confirmed by noting that a matching lower bound of the latter can be given by an explicit symmetric two-qubit strategy. Apart from the trivial positivity facets, these findings suggest that their violation beyond the symmetric bound can be used as a device-independent witness for asymmetry.

3 Conclusion

In this work, we have presented a family of party-permutation-invariant quantum strategies that recover the best-known quantum violation of the CGLMP Bell inequality with d outcomes for $d \leq 15$. For larger values of d, we conjecture that our symmetric quantum strategy remains optimal. In contrast, in the bipartite, 4-input, 2-output Bell scenario, we can easily identify facet-defining Bell inequalities where the general quantum bound is strictly larger than those achievable using only symmetric quantum strategies.

References

[1] J. S. Bell, *Speakable and unspeakable in quantum mechanics: collected papers on quantum philosophy*, 2nd ed. (Cambridge University Press, 2004).

[2] T. Norsen, "John S. Bell's concept of local causality", Am. J. Phys. **79**, 1261–1275 (2011).

[3] D. Collins and N. Gisin, "A relevant two qubit Bell inequality inequivalent to the CHSH inequality", J. Phys. A: Math. Gen. **37**, 1775 (2004).

[4] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner, "The quantum moment problem and bounds on entangled multi-prover games", in 23rd Annu. IEEE Conf. on Comput. Comp, 2008, CCC'08 (2008), pp. 199–210.

[5] A. Tavakoli, M. Farkas, D. Rosset, J.-D. Bancal, and J. Kaniewski, "Mutually unbiased bases and symmetric informationally complete measurements in Bell experiments", Sci. Adv. **7**, eabc3847 (2021).

[6] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, "Bell inequalities for arbitrarily high-dimensional systems", Phys. Rev. Lett. **88**, 040404 (2002).

[7] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories", Phys. Rev. Lett. **23**, 880–884 (1969).

[8] A. Acín, T. Durt, N. Gisin, and J. I. Latorre, "Quantum nonlocality in two three-level systems", Phys. Rev. A **65**, 052325 (2002).

[9] S. L. Braunstein, A. Mann, and M. Revzen, "Maximal violation of Bell inequalities for mixed states", Phys. Rev. Lett. **68**, 3259–3261 (1992).

[10] E. Oudot, J.-D. Bancal, P. Sekatski, and N. Sangouard, "Bipartite nonlocality with a many-body system", New J. Phys. **21**, 103043 (2019).

[11] M. Navascués, S. Pironio, and A. Acín, "Bounding the set of quantum correlations", Phys. Rev. Lett. **98**, 010401 (2007).

[12] M. Navascués, S. Pironio, and A. Acín, "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations", New J. Phys. **10**, 073013 (2008).

[13] T. Moroder, J.-D. Bancal, Y.-C. Liang, M. Hofmann, and O. Gühne, "Device-independent entanglement quantification and related applications", Phys. Rev. Lett. **111**, 030501 (2013).

[14] R. F. Werner and M. M. Wolf, "Bell inequalities and entanglement", Quantum Info. Comput. **1**, 1 (2001).

[15] Y.-C. Liang, C.-W. Lim, and D.-L. Deng, "Reexamination of a multisetting Bell inequality for qudits", Phys. Rev. A **80**, 052116 (2009).

Appendix A Symmetric strategy

A.1 Symmetric strategy in CGLMP inequality

For the Collins-Gisin-Linden-Massar (CGLMP) inequality [6], which involves two parties who each perform d-two outcome measurements. We found a symmetry version of the quantum strategy that satisfy the maximal violation of following symmetric CGLMP inequality [6], which is $A_1 \leftrightarrow (d-A_1)$ and for $A_1,\ A_2 \leftrightarrow (d-A_2-1)$ and for A_2 compared to the general version that described in [6]:

$$\begin{aligned} I_d = \sum_{k=0}^{[d/2]-1} (1-\frac{2k}{d-1})\{[&P(A_1 = d-B_1-k) \\ &+ P(B_2 = d-A_1+k) \\ &+ P(B_1 = d-A_2+k) \\ &+ P(A_2 = d-B_2-k-1)] \\ &- [P(A_1 = d-B_1+k+1) \\ &+ P(B_2 = d-A_1-k-1) \\ &+ P(B_1 = d-A_2-k-1) \\ &+ P(A_2 = d-B_2+k)]\} \leq 2. \end{aligned} \tag{10}$$

as $P(A_x = d-B_y-k) \equiv \sum_{a,b} P(a,b|x,y)\delta_{a,d-b-k}$, where can get optimal Bell value from the symmetric quantum strategy for any $d \geq 2$ dimension.

The measurements are given by making some adjustments to [8], while the d output measurement can be written in the projectors:

$$M^A_{a|1} = |a\rangle\langle a|,\ M^A_{a|2} = |\tilde{a}\rangle\langle\tilde{a}| \tag{11}$$

where the first measurement are in the standard basis and $(|\tilde{a}\rangle)_i = (U)_{ia}$, and $U = TW$ with matrix $T_{d\times d}$

$$(T)_{ij} = \begin{cases} (-1) & ,2 \leq i \leq j \\ 1 & else \end{cases}$$

and W denote as the matrix obtained by taking the absolute value of each matrix element of $W = U_{FT}V_1U^\dagger_{FT}$ and it can also be simplified as:

$$(W)_{jk} = \frac{1}{d}\left|\frac{1}{\sin[(j-k-\frac{1}{2})\frac{\pi}{d}]}\right|. \tag{12}$$

From [8], we know that the optimal measurement bases consist of Alice applying a diagonal unitary followed by a discrete Fourier transform $U_{FT} = \frac{1}{\sqrt{d}}e^{i(j-1)(k-1)\frac{2\pi}{d}}$ before measuring in the standard basis. The W is the inner product of two measurement basis $M^i_a = (U_{FT}V_i)^\dagger$ with

$$V_0 = \mathbb{I}_d,\ \ V_1 = \sum_a e^{i\frac{a\pi}{d}}|a\rangle\langle a| \tag{13}$$

This approach is applied to the corresponding symmetry CGLMP inequality to get the optimal violation and also the optimal violation of equivalent symmetric I_{22dd} inequality [3] with the same correlation:

$$\begin{aligned} I_{22dd} = &\sum_{a=1}^{d-1}\sum_{b=1}^{d-a} P(a,b|1,1) \\ &+ \sum_{a=1}^{d-1}\sum_{b=d-a}^{d-1}[P(a,b|1,2) + P(a,b|2,1) - P(a,b|2,2)] \\ &- \sum_{a=1}^{d-1} P(a|1) - \sum_{b=1}^{d-1} P(b|1) \leq 0, \end{aligned} \tag{14}$$

We describe the $d = 3$ matrix as an example, we can compute the inner product W between two bases M^i_a from Alice's side:

$$\begin{aligned} W &= M^{0\dagger}_a M^1_a \\ &= \begin{pmatrix} \langle a_{11}|a_{21}\rangle & \langle a_{11}|a_{21}\rangle & \langle a_{11}|a_{21}\rangle \\ \langle a_{12}|a_{22}\rangle & \langle a_{12}|a_{22}\rangle & \langle a_{12}|a_{22}\rangle \\ \langle a_{13}|a_{23}\rangle & \langle a_{13}|a_{23}\rangle & \langle a_{13}|a_{23}\rangle \end{pmatrix}. \end{aligned} \tag{15}$$

The actual unitary constructed by measurement basis for the second measurement of the symmetric strategy are:

$$\begin{aligned} U &= \begin{pmatrix} |\tilde{0}\rangle & |\tilde{1}\rangle & |\tilde{2}\rangle \end{pmatrix}. \\ &= \frac{1}{3}\begin{pmatrix} 2 & 1 & 2 \\ 2 & -2 & -1 \\ 1 & 2 & -2 \end{pmatrix}. \end{aligned} \tag{16}$$

From the strategy $M^{(A)}_{a|x} = M^{(B)}_{a|x}$, the first measurement unitary $M^{(A)}_{1|x} = M^{(B)}_{1|x}$ is constructed by standard basis, that is an identity respect to dimension d. While adding the minus sign on the upper triangle from the second column and row of W, the second measurement basis is described as the column in U. The optimal asymmetric strategy consists of employing the partially entangled two-qutrit state:

$$\begin{aligned} |\psi_3\rangle = \frac{1}{3}\gamma\Bigg[&\frac{5-\sqrt{33}}{2}|00\rangle + (|01\rangle + |10\rangle) \\ &+ \frac{-7+\sqrt{33}}{\sqrt{2}}(|02\rangle + |20\rangle - |11\rangle) \\ &- \frac{5-\sqrt{33}}{2}(|12\rangle + |21\rangle) + |22\rangle\Bigg]. \end{aligned} \tag{17}$$

with $\gamma = 2\sqrt{\frac{2}{55-9\sqrt{33}}}$ and the state is given by the eigenstate of the maximal eigenvalue $\lambda_1 = 2.9149$ of corresponding Bell operator B [9]:

$$B = \begin{pmatrix} -\frac{4}{9} & -\frac{2}{9} & -\frac{2}{3} & -\frac{2}{9} & 0 & \frac{4}{9} & -\frac{2}{3} & \frac{4}{9} & \frac{4}{9} \\ -\frac{2}{9} & -\frac{2}{3} & \frac{4}{9} & 0 & -\frac{4}{9} & -\frac{4}{9} & \frac{4}{9} & \frac{2}{9} & 0 \\ -\frac{2}{3} & \frac{4}{9} & \frac{10}{9} & \frac{4}{9} & -\frac{4}{9} & \frac{2}{3} & \frac{4}{9} & 0 & \frac{4}{9} \\ -\frac{2}{9} & 0 & \frac{4}{9} & -\frac{2}{3} & -\frac{4}{9} & \frac{2}{9} & \frac{4}{9} & -\frac{4}{9} & 0 \\ 0 & -\frac{4}{9} & -\frac{4}{9} & -\frac{4}{9} & \frac{10}{9} & -\frac{2}{3} & -\frac{4}{9} & -\frac{2}{3} & -\frac{4}{9} \\ \frac{4}{9} & -\frac{4}{9} & \frac{2}{3} & \frac{2}{9} & -\frac{2}{3} & -\frac{4}{9} & 0 & -\frac{4}{9} & \frac{2}{9} \\ -\frac{2}{3} & \frac{4}{9} & \frac{4}{9} & \frac{4}{9} & -\frac{4}{9} & 0 & \frac{10}{9} & \frac{2}{3} & \frac{4}{9} \\ \frac{4}{9} & \frac{2}{9} & 0 & -\frac{4}{9} & -\frac{2}{3} & -\frac{4}{9} & \frac{2}{3} & -\frac{4}{9} & \frac{2}{9} \\ \frac{4}{9} & 0 & \frac{4}{9} & 0 & -\frac{4}{9} & \frac{2}{9} & \frac{4}{9} & \frac{2}{9} & -\frac{2}{3} \end{pmatrix}.$$

Appendix B Numerical optimization

B.1 Optimization with symmetric quantum strategy

We define a symmetric quantum strategy to be one where the same measurement settings are used by Alice and Bob in a bipartite Bell scenario. To get the optimal Bell value for a symmetric strategy, we use the fminunc function in MATLAB to help us maximize Bell value by searching for the best state in symmetric subspace and rank-1 projective measurements $\Pi_{a|x}$ that correspond to the set of projections onto the columns of some unitary matrix U^x_{ij} on dimension d:

$$\begin{aligned} &\Pi_{a|x} = U^x_{ia} U^{x\dagger}_{ia}, \\ &\Pi_{b|y} = \Pi_{a|x},\ \forall x,\ y \in \{0, \cdots, m\}. \end{aligned} \tag{18}$$

where i, j are the labels of columns and rows of the unitary matrix. This non-linear optimization happens to a proper result in qubits.

B.2 See-saw optimization with symmetry correlation

For finding the optimal quantum strategy, there is also a heuristic approach called the see-saw method [14, 15], where we exploit the fact that the objective function of a Bell value is a *bilinear* function of Alice and Bob's local measurement operators. Thus, by iteratively optimizes each party's measurements, we can converge to a local maximum of the maximal Bell violation. The iterations stop only when the change in the objective value is smaller than some chosen threshold.

This means that we can write each iteration as a semidefinite program. For example, our first iteration consists of finding the optimal $M_{a|x}$ given some randomly chosen initial values for Bob's measurements $M_{b|y}$ and state ρ_{AB}. Since we are interested in symmetry correlations $\vec{P}$, our optimization problem can be written as

$$\begin{aligned} \max\ & S = \vec{\beta} \cdot \vec{P} \\ \text{s.t. } & P(a,b|x,y) = P(b,a|y,x), \\ & \vec{P} = \mathrm{tr}(\rho_{AB} M_{a|x} \otimes M_{b|y}), \\ & M_{a|x} \succeq 0, \\ & \sum_a M_{a|x} = \mathbb{I},\ \forall\, x \in \{0, \cdots, n\}. \end{aligned} \tag{19}$$

After we find some optimal measurements $M_{a|x}$ for Alice, then we can execute the next iteration This time, we maximize the Bell value S using the optimal measurements we found in Eq. (19) for Alice and the initial value of the state. This time we want to find an optimal $M_{b|y}$ so we have a similar optimization as described in Eq. (19), except we replace the constraints for $M_{b|y}$. Finally, for the last iteration, we use the optimal values of $M_{a|x}$ and $M_{b|y}$ to find the optimal state using in Eq. (20), again with symmetry constraints on $\vec{P}$:

$$\begin{aligned} \max\ & S = \vec{\beta} \cdot \vec{P} \\ \text{s.t. } & P(a,b|x,y) = P(b,a|y,x), \\ & \vec{P} = \mathrm{tr}(\rho_{AB} M_{a|x} \otimes M_{b|y}), \\ & \rho_{AB} \geq 0, \\ & \mathrm{tr}(\rho_{AB}) = 1, \end{aligned} \tag{20}$$

Note that this process is repeated many times until the objective value converges to some value, where the convergence is defined up to some numerical precision. To obtain a correlation that violates the Bell inequality and even achieves maximal violation in symmetric correlations, it is necessary to consider an appropriate initial state. This can be a maximally entangled state or a partially entangled state in the symmetric subspace of the Hilbert space.

Appendix C Some numerical results

Ineq.	L	NPA-L3	NPA-Swap-L3	qubits
J^{45}_{4422}	0	3.0822	0.0000	-0.2245
J^{73}_{4422}	0	4.4902	0.0000	-0.7300
J^{126}_{4422}	0	3.2094	0.0000	0.0000

Table 1: Summary of the three Bell inequalities whose upper bound on their symmetric quantum bound reduce to the local bound. Here and in the other tables, NPA-L3 means an upper bound computed using the NPA hierarchy at level 3 whereas NPA-Swap-L3 means an upper bound on the Bell violation computed with the NPA hierarchy at level 3 and the symmetric constraint of Eq. (1). Also included in the best symmetric qubit bound that we have found.

Ineq.	L	NPA-L3	NPA-Swap-L3	qubits
J^{45}_{4422}	0	3.0822	2.9736	2.5451
J^{73}_{4422}	0	4.4902	4.0378	4.0030
J^{126}_{4422}	0	3.2094	2.7356	2.7356

Table 2: Summary of the (symmetric) quantum bound for the three inequalities of Table 1 after relabeling. Also included in the best symmetric qubit bound that we have found.

Ineq.	NPA-Swap-L3
CHSH	0.8284
I_{3322}	1.0035
AII_1	2.4222
AS_1	2.1650
A_5	1.7413
AS_2	3.5140
I^{9}_{4422}	1.8467
I^{10}_{4422}	2.4558
I^{11}_{4422}	0.8809
I^{12}_{4422}	2.4753
I^{17}_{4422}	2.6856
J^{12}_{4422}	1.5755
J^{17}_{4422}	0.1201
J^{19}_{4422}	2.6969
J^{22}_{4422}	3.2625
J^{26}_{4422}	2.5610
J^{27}_{4422}	1.1701
J^{28}_{4422}	3.0000
J^{32}_{4422}	2.3606
J^{41}_{4422}	1.5886
J^{58}_{4422}	3.5258
J^{60}_{4422}	2.3691
J^{61}_{4422}	3.2702
J^{85}_{4422}	3.9051
J^{90}_{4422}	3.3593
J^{91}_{4422}	5.1971
J^{92}_{4422}	4.2593
J^{102}_{4422}	2.6603
J^{105}_{4422}	4.2969
J^{108}_{4422}	3.8706
J^{109}_{4422}	6.9042
J^{110}_{4422}	3.7830
J^{125}_{4422}	4.0000
S^{51}_{242}	4.0541
S^{52}_{242}	3.4815

Table 3: Summary of a partial list of symmetric facet-defining Bell inequalities in the 4422 scenario and their quantum bound, which, except for I_{3322}, is achievable using a symmetric two-qubit strategy. Here, we follow the normalization of [10]; the local bounds, cf. Eq. (8), are thus 0.

Ineq.	NPA-L3	NPA-Swap-L3
I^{2}_{4422}	2.4855	2.0000
I^{3}_{4322}	1.7459	1.6167
I^{5}_{4422}	1.7459	1.7354
I^{7}_{4422}	1.8193	1.7097
I^{11}_{4422}	2.5534	0.8809
J^{2}_{4422}	2.4560	2.3660
J^{3}_{4422}	3.2425	2.4632
J^{6}_{4422}	1.7846	1.5303
J^{12}_{4422}	2.9047	1.5755
J^{16}_{4422}	2.2696	2.1900
J^{17}_{4422}	2.5518	0.1200
J^{23}_{4422}	2.4361	2.2273
J^{25}_{4422}	2.0591	1.5912
J^{27}_{4422}	3.8571	1.1701
J^{33}_{4422}	2.4603	2.2554
J^{35}_{4422}	2.7730	2.1187
J^{36}_{4422}	2.6826	1.2361
J^{39}_{4422}	3.3058	1.8902
J^{41}_{4422}	3.0384	1.5886
J^{43}_{4422}	1.8743	1.7105
J^{44}_{4422}	3.2094	2.7356
J^{46}_{4422}	3.8867	1.8439
J^{47}_{4422}	3.0579	1.2610
J^{48}_{4422}	3.0068	1.8356
J^{50}_{4422}	3.4224	1.3531
J^{51}_{4422}	2.7001	2.1674
J^{52}_{4422}	4.3996	4.3617
J^{53}_{4422}	3.2373	3.1025
J^{57}_{4422}	3.4439	2.6058
J^{56}_{4422}	3.2768	2.4926
J^{59}_{4422}	2.5518	1.4209
J^{63}_{4422}	2.4314	1.5833
J^{64}_{4422}	3.6668	2.2977
J^{65}_{4422}	4.4441	2.4510
J^{66}_{4422}	2.4745	0.8376
J^{67}_{4422}	2.8851	2.4108
J^{68}_{4422}	4.0714	2.8477
J^{71}_{4422}	2.3291	2.2046
J^{72}_{4422}	3.1513	1.7519
J^{74}_{4422}	1.7394	1.0342
J^{75}_{4422}	2.5081	2.3899
J^{76}_{4422}	3.3164	2.2137
J^{77}_{4422}	3.3624	3.1937
J^{78}_{4422}	3.5750	3.1361
J^{80}_{4422}	3.6205	3.3504
J^{81}_{4422}	2.3983	2.3310
J^{82}_{4422}	2.9232	2.8710
J^{83}_{4422}	2.6767	1.1386
J^{84}_{4422}	4.2678	1.6680

Table 4: Summary of a partial list of *asymmetric* facet-defining Bell inequalities in the 4422 scenario, their quantum bound, and their symmetric quantum bound, which is achievable using a symmetric two-qubit strategy. Again, the local upper bound is 0.

Ineq.	NPA-L3	NPA-Swap-L3
J_{4422}^{87}	2.7395	1.1650
J_{4422}^{88}	2.4640	2.3452
J_{4422}^{89}	4.0142	3.8477
J_{4422}^{93}	3.8509	3.7623
J_{4422}^{98}	4.5988	2.9244
J_{4422}^{99}	3.5962	3.1623
J_{4422}^{100}	5.0701	4.8958
J_{4422}^{101}	4.1184	1.8614
J_{4422}^{103}	4.2080	4.1362
J_{4422}^{104}	6.3503	6.3071
J_{4422}^{106}	3.3530	1.6759
J_{4422}^{107}	3.7359	1.4608
J_{4422}^{111}	3.0304	1.8046
J_{4422}^{112}	2.4990	2.4376
J_{4422}^{114}	4.2612	3.5249
J_{4422}^{116}	3.0607	2.4743
J_{4422}^{117}	3.8884	3.2446
J_{4422}^{119}	3.5944	3.2022
J_{4422}^{121}	2.3883	2.0440
J_{4422}^{124}	3.7639	2.6913
J_{4422}^{123}	4.2564	2.7188
J_{4422}^{126}	3.2094	2.7356
J_{4422}^{128}	4.0384	3.5783
J_{4422}^{129}	4.2090	1.1362
N_{4422}^{1}	4.0303	2.7188
N_{4422}^{2}	2.7127	2.1279
N_{4422}^{3}	3.0947	1.6173
N_{4422}^{5}	5.7507	3.4798
N_{4422}^{7}	2.4974	1.7370
N_{4422}^{8}	4.0755	0.4685
N_{4422}^{12}	5.2999	1.1686

Table 5: Summary of a partial list of *asymmetric* facet-defining Bell inequalities in the 4422 scenario, their quantum bound, and their symmetric quantum bound, which is achievable using a symmetric two-qubit strategy. Again, the local upper bound is 0.

Scalable surface-code quantum error correction based on cavity-QED network

Rui Asaoka[1 2 *] Yasunari Suzuki[1 2 †] Yuuki Tokunaga[1 2]

[1] *NTT Computer and Data Science Laboratories, NTT Corporation, Musashino 180-8585, Japan*
[2] *NTT Research Center for Theoretical Quantum Information, Atsugi, 243-0198, Japan*
[3] *JST, PRESTO, Kawaguchi, 332-0012, Japan*

Abstract. The exploration of an efficient and scalable architecture of fault-tolerant quantum computing (FTQC) is vital for the demonstration of useful quantum computing. Here, we propose a scalable, high-performance, and practical architecture with cavity-quantum-electrodynamics (CQED) network. Our architecture takes advantage of the stability of neutral atoms and the flexibility of a CQED network. We show a concrete framework of the implementation of surface codes and numerically analyze the logical error probability and threshold values for two extreme network architectures. Our results open up a new direction of FTQC with neutral atoms.

Keywords: quantum error correction, cavity quantum electrodynamics

1 Introduction

The recent development in quantum processing using neutral atoms has recently attracted much attention [1]. It has several advantages such as much longer lifetime than bulk qubits and negligible correlated errors between atoms thanks to no charge. Alternatively, the neutral-atom processor has a drawback, a weak interaction between qubits.

One strategy to compensate the drawback is cavity quantum electrodynamics (CQED). The recent development in CQED-network technology, such as the nanofiber-cavity network [2, 3], is realizing CQED-network systems with both seamless cavity connection and strong coupling between neutral atoms and cavity field. To utilize the itinerancy of photons opens the possibility for high designability in quantum processing. Nevertheless, there are few studies that investigate its fault tolerance and scalability compared with other physical platforms.

In our talk, we propose a fault-tolerant and scalable architecture with trapped neutral atoms with cavity networks. In this architecture, we utilize neutral atoms as data qubits. Neutral atoms are reset to a ground state and trapped with the magneto-optical trap. We adopt the standard two-dimensional surface code [4] to construct a logical qubit. As is well known, this code requires only the nearest-neighbor interaction between qubits; our architecture with itinerant photons has room for considering quantum codes which shows higher performance, such as efficient low-density parity check codes. Nevertheless, the surface code is an unavoidable stepping stone to exploit high performance codes. Thus, we first investigate its fault-tolerance and scalability.

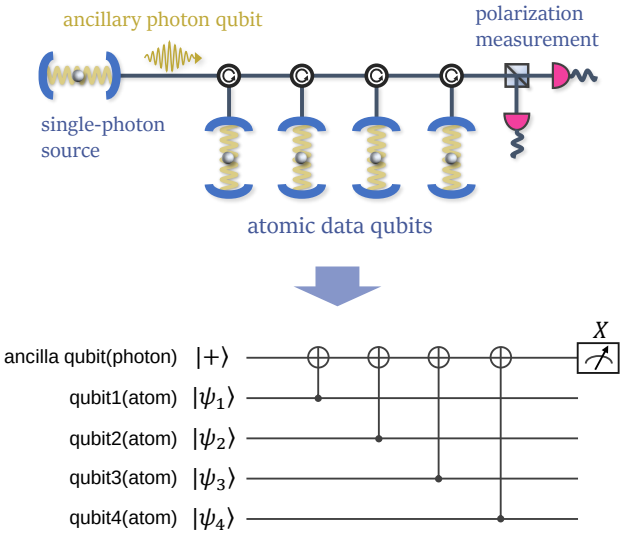


Figure 1: Stabilizer measurement with CQED network for constructing surface codes.

2 Method

2.1 Stabilizer measurement based on CQED

Thanks to the strong interaction of atoms with optical modes mentioned above, we can perform some important two-qubit gates, such as the controlled-Z gate between a photon and an atom [5, 6, 7]. Therefore, with appropriate basis changes, we can perform a multi-qubit Pauli measurement with single-photon inputs, which is enough for performing stabilizer measurements. Figure 1 shows a schematic picture of a stabilizer measurement for four atomic qubits. Ancillary photons pass through a CQED network, which indicates that stabilizer measurements in our architecture can be passively performed and that there is no need to arrange the atomic data qubits in two dimension in real space. Here we focus on the case of a single logical qubit for simplicity, that is, mainly discuss the fault-tolerance of the parity-check measurement

*rui.asaoka@ntt.com
†yasunari.suzuki@ntt.com

process; nevertheless, we would like to emphasize that our idea can be straightforwardly extended to the case of multiple logical qubits.

2.2 Error sources

We expect three types of errors in the procedure of syndrome-value readout. The first one is the T_2 decay of trapped atoms. This forces the period of each syndrome measurement cycle sufficiently shorter than the lifetime. We note that T_1 can be neglected here because it is very slow compared to other time scales characteristic of CQED systems, such as the decay of the excited state.

The second is the infidelity of the atom-photon gate. The reduction in the fidelity of the CZ gate comes from unbalanced photon loss between the computational bases and distortion of reflected pulse shapes. The former means that the photon loss probability depends on the states of atomic and photonic qubits. This causes the effective rotation of atomic qubits. However, it is known that the unbalanced photo loss can be canceled by designing cavity parameters appropriately [8], and we do not consider this error in this study. The latter is caused by frequency-dependent phase shift between the input (incident) and output (reflected) photon pulses. In the frequency-dependent phase shift, the reduction in the gate fidelity mainly caused by the first order term regarding frequency, or the delay in an output pulse. Thus, in this paper, we consider that the infidelity of the atom-photon gate is due to the pulse delay.

The last dominant error source is photon loss through dissipative channels, namely undesirable scattering and absorption inside cavities, atomic spontaneous emission, transmission loss, and losses in detectors, circulators, and switches. This photon loss error differ from the first two error sources in that this error can be detected, i.e., we can know that the photon is lost when the photodetectors do not click. While we can perform error correction by ignoring the heralded signal, we can achieve higher performance by utilizing the photon-loss information in the error estimation process.

2.3 CQED-network structures

Here we propose two particular CQED-network structures for constructing a logical qubit: N-cavity structure (Fig. 2(a)) and 4-cavity multi-atom structure (Fig. 2(b)), both of which are extensions of the fundamental structure in Fig. 1. The former is the most straightforward realization of the surface code; $N = 2d^2 + 2d + 1$ cavities (d is the code distance), each including a single atomic qubit, are allocated in the two-dimensional (2D) grid array with the nearest neighbors connected (we note that the cavities need not be arranged in the 2D grid in the real space). Switches can rearrange the path of photon pulses and even have the choices to connect polarimeters or photon sources. This structure requires the same number of cavities as atomic qubits, but instead enables a highly parallel syndrome measurement if the cavities are connected in two-dimensional grid where each node

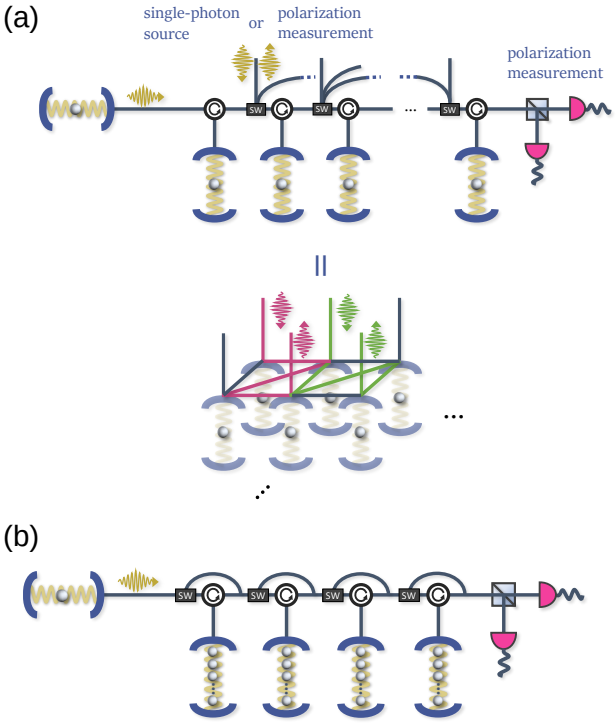


Figure 2: Cavity-network structures for constructing surface codes. (a) Optimal structure when d^2 cavities can be prepared. (b) Structure when only the minimum number of cavities, or 4 cavities, is available, where each cavity includes $d^2/4$ atomic qubits. Any one cavity is skipped with switches when three-qubit Pauli strings on the edge of a logical qubit are measured.

is connected to a single-photon source and a polarimeter. The latter has only four cavities, each including a 1D array of trapped atoms. We can somehow choose which atoms to couple to the cavities, such as resonance shift depending on atomic position by gradient electric or magnetic field with respect to a target atom, Stark shift by selective laser irradiation, or position shift of each atom by optical tweezer. This structure is of the lowest parallelism of the syndrome measurements, or the largest syndrome-readout depth, but instead requires the minimum number of cavities for constructing a logical qubit relying on the fundamental structure in Fig. 1.

Thus, these structures are the extremes with respect to a trade-off relation between the experimental resource and the period of each syndrome measurement cycle, or between the difficulty in implementing a logical qubit and the T_2 error. In this study, we investigate the fault-tolerance of these two extreme cases.

3 Results and Conclusion

3.1 Threshold for each CQED-network structure

Here, we shows the boundaries between the regions where $p_{\mathrm{L},5}/p_{\mathrm{L},3}$ or $p_{\mathrm{L},7}/p_{\mathrm{L},5}$ ($p_{\mathrm{L,d}}$ is defined as the logical error for code distance d) is less than unity and

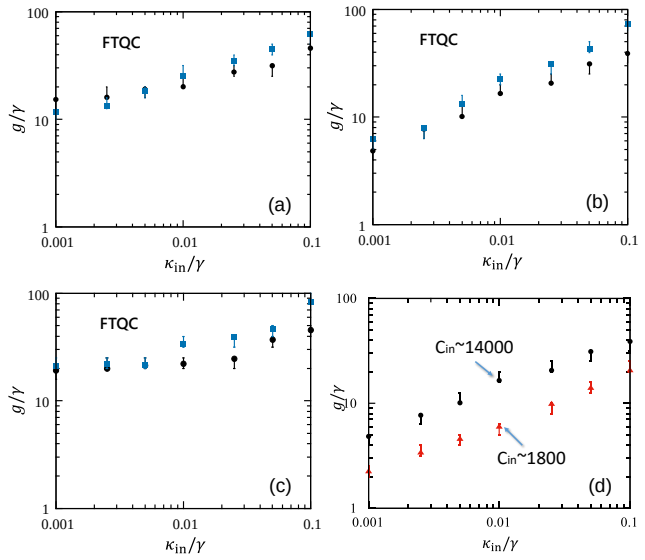


Figure 3: Boundaries between the regions where $p_{\mathrm{L},5}/p_{\mathrm{L},3}$ (circles) or $p_{\mathrm{L},7}/p_{\mathrm{L},5}$ (squares) is less than unity and greater than unity as a function of g/γ and $\kappa_{\mathrm{in}}/\gamma$. (a) N-cavity structure for $T_2\gamma = 10^4$. (b) N-cavity structure for $T_2\gamma = 10^6$. (c) 4-cavity structure for $T_2\gamma = 10^6$. Here we assume that the peripheral devices are ideal, namely, $p_{\mathrm{SW}} + p_{\mathrm{cir}} = 0$. (d) Boundary of $p_{\mathrm{L},5}/p_{\mathrm{L},3}$ for the N-cavity structure calculated based on the stabilizer simulation using the improved MWPMA utilizing the information of the loss event of an ancillary photonic qubit (triangles). We show $p_{\mathrm{L},5}/p_{\mathrm{L},3}$ boundary in (a) as a reference.

greater than unity as a function of the cavity parameters $(g, \kappa_{\mathrm{in}}, \gamma)$. Here g, κ_{in}, and γ are the coupling strength between cavity field and an atom, the undesirable cavity decay rate due to the imperfection of a cavity, and the atomic decay rate, respectively. In Figs. 3(a) and (b), we show the difference between different dephasing times $T_2\gamma = 10^4$ and 10^6 in the case of the N-cavity structure. The upper regions to the data points indicate that QC is fault-tolerant. When $\kappa_{\mathrm{in}}/\gamma$ is small, the requirement for the CQED parameters is notably relaxed for the longer dephasing time. This is because a longer dephasing time allows a longer input pulse, resulting in achieving high gate fidelity even for a small Rabi splitting, namely, a small g. This effect is emphasized for smaller κ_{in} because the cavity linewidth κ becomes sharp. The minimum value of the internal cooperativity $C_{\mathrm{in}} \equiv \frac{g^2}{2\kappa_{\mathrm{in}}\gamma}$, a fundamental characteristic of the cavity performance for QC, required for FTQC is a few tens of thousands in both cases. Figure 3(c) shows the error boundaries defined in the same way as Figs. 3(a) and (b) in the case of the 4-cavity structure. The dephasing time T_2 is the same as Fig. 3(a). In this case, the requirements of the cavity parameters are a little more demanding for small $\kappa_{\mathrm{in}}/\gamma$ compared to the N-cavity case (Fig. 3(a)). This is because the 4-cavity structure sacrifices the parallelism of the syndrome measurements instead of the ease of implementation; this is equivalent to experiencing a short dephasing time, which is an opposite case of Fig. 3(b) of a long dephasing time.

3.2 Threshold improvement utilizing loss information of ancillary qubits

So far, we have calculated the logical error rate using a error estimation protocol, the minimum weight perfect matching algorithm (MWPMA), in the stabilizer simulation. In this section, we investigate whether an advantage of our proposed system, being able to detect the loss event of an ancillary photonic qubit, makes the error estimation in MWPMA more efficient. The central idea in our improved MWPMA is that we set the error probability high around the places where ancillary photonic qubits are lost. Figure 3(d) shows the $p_{\mathrm{L},5}/p_{\mathrm{L},3}$ boundary for the N-cavity structure calculated based on the stabilizer simulation using the improved MWPMA utilizing the information of the loss event of an ancillary photonic qubit. It can be seen that the threshold is really improved by utilizing the loss information; for example, the value of the internal cooperativity C_{in}, which is defined by $g^2/2\kappa_{\mathrm{in}}\gamma$, required for the fault-tolerant quantum computation is about 1/10 compared to the stabilizer simulation which does not use the loss information when $\kappa_{\mathrm{in}}/\gamma = 0.01$.

3.3 Conclusion

We have estimated the error thresholds in a surface code for extreme two CQED-network structures. Our estimation is a reading study considering specific CQED-network structure and its scalability. Moreover, our error-decoding algorithm tailored for the proposed architecture greatly relaxes the required performance of CQED networks to achieve the error threshold. Our results open up a new direction of FTQC with neutral atoms.

References

[1] Loïc Henriet *et al.* Quantum computing with neutral atoms. Quantum **4**, 327 (2020).

[2] S. Kato and T. Aoki. Strong coupling between a trapped single atom and an all-fiber cavity. Phys. Rev. Lett. **115**, 093603 (2015).

[3] S. K. Ruddell, K. E. Webb, M. Takahata, S. Kato, and T. Aoki. Ultra-low-loss nanofiber Fabry 寰撒erot cavities optimized for cavity quantum electrodynamics. Opt. Lett. **45**, 4875 (2020).

[4] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland. Surface codes: Towards practical large-scale quantum computation Phys. Rev. A **86**, 032324 (2012).

[5] L.-M. Duan and H. J. Kimble Scalable photonic quantum computation through cavity-assisted interactions. Phys. Rev. Lett. **92**, 127902 (2004).

[6] L.-M. Duan, B. Wang, and H. J. Kimble. Robust quantum gates on neutral atoms with cavity-assisted photon scattering. Phys. Rev. A **72**, 032333 (2005).

[7] R. Asaoka, Y. Tokunaga, R. Kanamoto, H. Goto, and T. Aoki. Requirements for fault-tolerant quantum computation with cavity-QED-based atom-atom gates mediated by a photon with a finite pulse length. Phys. Rev. A **104**, 043702 (2021).

[8] T. Utsugi, R. Asaoka, Y. Tokunaga, and T. Aoki. Optimal cavity design for minimizing errors in cavity-QED-based atom-photon entangling gates with finite temporal duration. arXiv:2211.04151

Acknowledgement

This work is supported by PRESTO, JST, Grant No. JPMJPR1916; ERATO, JST, Grant No. JPMJER1601; CREST, JST, Grant No. JPMJCR1771; MEXT Q-LEAP Grant No. JPMXS0120319794 and JPMXS0118068682; Moonshot R&D, JST, Grant No. JPMJMS2061.

Efficient Verification of Genuinely Entangled Subspaces

Congcong Zheng[1 2] Xutao Yu[1 2 3] Ping Xu[4] Kun Wang[4 *]

[1] *State Key Lab of Millimeter Waves, Southeast University, Nanjing 211189, China*
[2] *Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, China*
[3] *Purple Mountain Laboratories, Nanjing 211111, China*
[4] *Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China*

1 Overview of the results

We initialize the study of efficient verification of genuinely entangled subspaces (GES). We establish a general verification framework and provide efficient verification strategies for various GESs of practical interests. Firstly, we construct an efficient verification strategy for the GES spanned by the 3-qubit GHZ state and W state. This strategy involves one-way adaptive local measurements, which we call the "Pauli+2" strategy. Then, we construct two strategies to verify GES induced by stabilizer codes of size k in an n-qubit system. Notably, *we present the first verification strategies for the genuinely entangled stabilizer subspaces induced by the prominent five-qubit code and toric code.* These strategies use a limited number of Pauli measurements and are non-adaptive, thus are experimentally feasible.

Our findings demonstrate that genuinely entangled subspaces (GES), including the stabilizer subspaces fundamental in quantum error correction as special case, can be efficiently verified using experimentally feasible measurements. We believe the results are beneficial to the broader audience of AQIS, *especially to those who are working in constructing and benchmarking the qualities of quantum error correction codes.*

A full technical version can be found in the attached technical PDF.

2 Quantum subspace verification

In this section, we first formally define the task of quantum subspace verification. Then, we discuss this task under local constraints, specifically focusing on verification strategies that can be implemented locally.

2.1 Task description

Suppose we have N copies n-qubit states $\sigma_1, \cdots, \sigma_N$ produced by a same quantum device $\mathcal{D}$. The quantum state verification task answers the question: "Are the states σ_i generated by $\mathcal{D}$ equal to a fixed state $|\psi\rangle\langle\psi|$?" Similarly, the quantum subspace verification aims to answer the question:

"Are the states σ_i generated by $\mathcal{D}$ contained in the subspace $\mathcal{V}$ spanned by the orthonormal basis $\{|\psi_j\rangle\}_j$?"

To mathematically verify whether a state is in the target subspace $\mathcal{V}$, we define the projector $\Pi := \sum_j |\psi_j\rangle\langle\psi_j|$ and provide the following lemma.

Lemma 1 *For a fixed quantum state σ,*

$$\mathrm{Tr}[\Pi\sigma] = \sum_j \langle\psi_j|\sigma|\psi_j\rangle = 1, \tag{1}$$

if and only if $\sigma \in \mathrm{span}\{|\psi_j\rangle\}$.

Thus, we can now formally define the quantum subspace verification task—Given a quantum device $\mathcal{D}$, distinguish between the following two cases:

1. **Good**: for all $i \in [N]$, $\mathrm{Tr}[\Pi\sigma_i] = 1$;
2. **Bad**: for all $i \in [N]$, $\mathrm{Tr}[\Pi\sigma_i] \leq 1 - \epsilon$ for some fixed ϵ.

A visual depiction of the quantum subspace verification task is given in Figure 1.

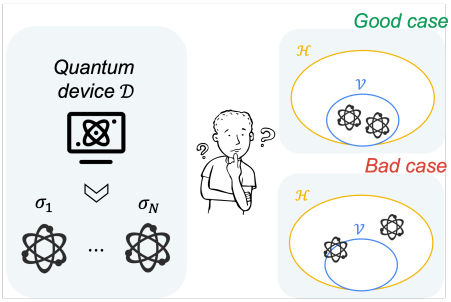


Figure 1: Quantum subspace verification. Given a quantum device $\mathcal{D}$, we aim to distinguish exclusively between two cases: **Good** case: all states prepared is in a target subspace $\mathcal{V} \subseteq \mathcal{H}$. **Bad** case: $\exists\, i$, $\sigma_i \notin \mathcal{V}$.

2.2 Practical verification with local constraints

Suppose that we have access to a set of POVM elements $\mathcal{M}$ and $\forall\, M \in \mathcal{M}$, M is a local projector (assisted

*nju.wangkun@gmail.com

by classical communication). Then for each state preparation, we pick a POVM element $M \in \mathcal{M}$ with some probability and consider the corresponding two-outcomes POVMs $\{M, \mathbb{1} - M\}$, where M has output "pass" and $\mathbb{1} - M$ has output "fail". Moreover, we define a probability mass $\mu : \mathcal{M} \to [0, 1]$, $\sum_{M\in\mathcal{M}} \mu(M) = 1$. The probability of a generated quantum state σ passing the test can be expressed as

$$\Pr\{\text{"pass"}|\sigma\} = \sum_{M\in\mathcal{M}} \mu(M)\,\mathrm{Tr}[M\sigma] \equiv \mathrm{Tr}[\Omega\sigma], \quad (2)$$

where the *verification operator* of this strategy is defined as

$$\Omega := \sum_{M\in\mathcal{M}} \mu(M) M. \quad (3)$$

To satisfy the requirement of the verification task, we impose two conditions on the verification operator Ω: *perfect completeness condition* and *soundness condition*. The perfect completeness condition requires that

$$\mathrm{Tr}[\Omega\sigma] = 1, \quad \forall\sigma \in \mathrm{span}\{|\psi_j\rangle\}. \quad (4)$$

This condition can be equivalently characterized using the projector Π associated with the target subspace $\mathcal{V}$ as follows.

Lemma 2 *The perfect completeness condition can be equivalently characterized as*

$$\mathrm{Tr}[\Omega\Pi] = \mathrm{rank}(\Pi), \quad (5)$$

where $\mathrm{rank}(\Pi)$ *is the rank of the projector.*

Now let's consider the soundness condition. We find the the worst-case passing probability $p(\Omega)$, defined as

$$p(\Omega) := \max_{\sigma:\mathrm{Tr}[\Pi\sigma]\le 1-\epsilon} \Pr\{\text{“pass”}|\sigma\}, \quad (6)$$

in the **Bad** case is uniquely determined by the largest eigenvalue of the projected effective verification operator, as elucidated in the following theorem.

Theorem 3 *It holds that*

$$p(\Omega) := \max_{\sigma:\mathrm{Tr}[\Pi\sigma]\le 1-\epsilon} \mathrm{Tr}[\Omega\sigma] = 1 - (1 - \lambda_{\max}(\widehat{\Omega}))\epsilon, \quad (7)$$

where $\widehat{\Omega} := (\mathbb{1} - \Pi)\Omega(\mathbb{1} - \Pi)$ *is the projected effective verification operator and* $\lambda_{\max}(X)$ *denotes the maximum eigenvalue of the Hermitian operator* X.

Therefore, the probability of accepting the **Bad** case is bounded as follows,

$$\Pr\{\text{"accept"}|\sigma_1, \cdots, \sigma_N\} \le (1 - \nu(\Omega)\epsilon)^N, \quad (8)$$

where $\nu(\Omega) := 1 - \lambda_{\max}(\widehat{\Omega})$ is the *spectral gap*. To achieve the bound δ, we have

$$N \ge \frac{1}{\nu(\Omega)} \times \frac{1}{\epsilon} \ln\frac{1}{\delta}. \quad (9)$$

3 Subspace spanned by GHZ state and W state

In this section, we propose an efficient verification protocol for the subspace spanned by the GHZ state and W state,

$$|\mathrm{GHZ}\rangle = (|000\rangle + |111\rangle)/\sqrt{2}, \quad (10a)$$

$$|\mathrm{W}\rangle = (|001\rangle + |010\rangle + |100\rangle)/\sqrt{3}. \quad (10b)$$

We accomplish this verification task using *Pauli+2* strategy and the framework of our strategy is outlined in Figure 2.

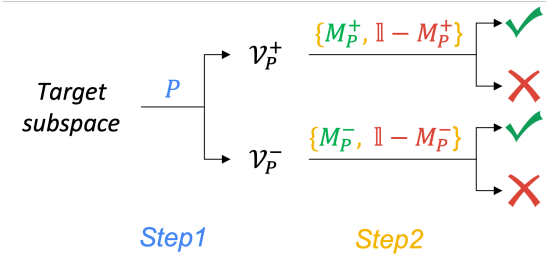


Figure 2: The framework of *Pauli+2* strategy, designed to verify the subspace spanned by the 3-qubit GHZ state and the W state. In the first step, we randomly perform Pauli measurement P_i. Then, we can obtain a post-measurement subspace $\mathcal{V}_{P_i}^o$ with measurement result $o \in \{+, -\}$. Subsequently, we perform the measurement $\{M_{P_i}^o, \mathbb{1} - M_{P_i}^o\}$ defined in the technical version. It the outcome is corresponding to the M_P^o, then we accept this state; otherwise, we reject it.

Firstly, we randomly choose a qubit index i and measure i-th qubit in a random Pauli X, Y, or Z basis. The corresponding Pauli measurement is represented as P_i. For the Pauli measurement P_i, there are two possible outcomes, $+1$ and -1. Conditioned on the measurement outcome, the remaining two qubits will live in a two-qubit subspace, spanned by two post-measurement states, which we term the *post-measurement subspace*. We denote the post-measurement subspace resulting from measurement P_i and outcome $o \in \{+1, -1\}$ as $\mathcal{V}_{P_i}^o$.

Secondly, we use the two-qubit subspace verification strategy, which is proposed in the technical version. We design two-qubit subspace verification strategies based on the outcome of the first measurement. For each post-measurement subspace $\mathcal{V}_{P_i}^o$, we can construct a measurement operator $M_{P_i}^o$ and perform POVM $\{M_{P_i}^o, \mathbb{1} - M_{P_i}^o\}$. We reject the state if the outcome is corresponding to $\mathbb{1} - M_{P_i}^o$; otherwise, we pass it.

Now, we consider the complexity of *Pauli+2* strategy. Numerical calculations show that the optimal spectral gap is about 0.358. Thus, the required number of copies N needs to satisfy

$$N \ge 2.79\frac{1}{\epsilon}\ln\frac{1}{\delta} \quad (11)$$

to achieve confidence level $1 - \delta$.

4 Stabilizer Subspace

In this section, we describe two efficient protocols for verifying the stabilizer subspaces. For a n-qubit system, a subspace $\mathcal{V}$ can be determined by a set of k stabilizer generators $\mathcal{G}_k$. We can construct a set of stabilizer operators $\mathcal{S}_k = \{P_{\boldsymbol{y}} : \boldsymbol{y} \in \mathbb{Z}_2^k\}$ as follows:

$$P_{\boldsymbol{y}} := \prod_{i=1}^{k} S_i^{y_i}. \tag{12}$$

Protocol I works by uniformly and randomly choosing a stabilizer operator $P_{\boldsymbol{y}}$ from $\mathcal{S}_k$ and measure the target state with $P_{\boldsymbol{y}}$. Mathematically, the verification operator of **Protocol I** reads

$$\Omega_{\mathrm{I}} := \frac{1}{2^k - 1} \sum_{P \in \mathcal{S}_k \setminus \{\mathbb{1}\}} P^+, \tag{13}$$

where $P^+ := (P + \mathbb{1})/2$ is the projector onto the positive eigenspace of stabilizer operator P. What's more, the verification efficiency, i.e. the spectral gap, of Ω_{I} satisfies

$$\nu(\Omega_{\mathrm{I}}) = \frac{2^{k-1}}{2^k - 1}. \tag{14}$$

To achieve a confidence level $1 - \delta$, it suffices to take

$$N(\Omega_{\mathrm{I}}) = \frac{(2^k - 1)}{2^{k-1}} \frac{1}{\epsilon} \ln \frac{1}{\delta} \approx 2 \frac{1}{\epsilon} \ln \frac{1}{\delta} \tag{15}$$

number of state copies, which is independent with the subspace size k. Notably, this strategy necessitates at most twice as many copies as the verification strategy without local constraints. The disadvantage of **Protocol I** is self-evident: the experimenters must be able to implement a total number $2^k - 1$ Pauli measurement settings, which increases exponentially in the subspace size k and is challenging. This disadvantage motivates the second protocol which requires far less number of measurement settings.

Protocol II works by randomly choosing a stabilizer generator S from $\mathcal{G}_k$, each with probability $1/k$. Mathematically, the verification operator of **Protocol II** reads

$$\Omega_{\mathrm{II}} := \frac{1}{k} \sum_{S \in \mathcal{G}_k} S^+, \tag{16}$$

where $S^+ := (S + \mathbb{1})/2$ is the projector onto the positive eigenspace of stabilizer generator S. Subsequently, we examine the verification efficiency of Ω_{II}, which is

$$\nu(\Omega_{\mathrm{II}}) = \frac{1}{k}. \tag{17}$$

Therefore, to achieve a confidence level $1 - \delta$, it suffices to take

$$N(\Omega_{\mathrm{II}}) = k \frac{1}{\epsilon} \ln \frac{1}{\delta} \tag{18}$$

number of state copies. Therefore, the drawback of **Protocol II** is obvious: it requires $k/2$ times more state copies than **Protocol I**. This indicates a fundamental trade-off between the total number of required state copies and the number of measurement settings, which deserves further investigation.

In the following, we present the *first* verification strategies for the genuinely entangled stabilizer subspaces induced by the prominent five-qubit code and the toric code.

Five-qubit code. Consider the GESS induced by the *five-qubit code* [1]. This subspace is generated by the following 4 generators,

$$\begin{aligned} S_1 &= X_1 Z_2 Z_3 X_4, \quad S_2 = X_2 Z_3 Z_4 X_5, \\ S_3 &= X_1 X_3 Z_4 Z_5, \quad S_4 = Z_1 X_2 X_4 Z_5. \end{aligned} \tag{19}$$

To verify such a subspace, **Protocol I** requires $2^4 - 1 = 15$ measurement settings, which are determined by 15 stabilizer operators (excluding $\mathbb{1}$) defined in Eq. (12). To achieve a confidence level $1 - \delta$, we need $15/(8\epsilon) \ln 1/\delta$ state copies. On the other hand, **Protocol II** only 4 measurement settings, determined by 4 generators defined in Eq. (19). However, **Protocol II** requires more state copies, specifically $4/\epsilon \ln 1/\delta$,to achieve the same confidence level $1 - \delta$.

Toric code. Consider the GESS induced by the *toric code* [1]. A toric code can be presented by a $L \times L$ lattice, where each edge represents a qubit. The corresponding stabilizer generators can be written as

$$S_v = \prod_{i \in v} X_i, \quad S_p = \prod_{i \in p} Z_i. \tag{20}$$

There are $2L^2 - 2$ stabilizer generators in total. Therefore, **Protocol I** requires $2^{2L^2-2} - 1$ measurement settings and $\approx 2/\epsilon \ln 1/\delta$ state copies to achieve a confidence level $1 - \delta$. **Protocol II** only requires $2L^2 - 2$ measurement settings but needs $(2L^2 - 2)/\epsilon \ln 1/\delta$ state copies to achieve the same confidence level $1 - \delta$.

Though we only provide two examples of GESSs, it should be noted that for any arbitrary GESSs, we can construct the corresponding verification strategies.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant Nos. 2019YFA0308700 and 2022YFF0712800), the Jiangsu Key R&D Program Project (Grant No. BE2023011-2), the Fundamental Research Funds for the Central Universities (Grant No. 2242022k60001), and the National Natural Science Foundation of China (Grant Nos. 61960206005 and 61871111).

References

[1] Flavio Baccari, Remigiusz Augusiak, Ivan Šupić, and Antonio Acín. Device-independent certification of genuinely entangled subspaces. *Physical Review Letters*, 125(26):260507, December 2020.

Efficient Verification of Genuinely Entangled Subspaces

Congcong Zheng,[1,2] Xutao Yu,[1,2,3,*] Ping Xu,[4] and Kun Wang[4,†]

[1]*State Key Lab of Millimeter Waves, Southeast University, Nanjing 211189, China*
[2]*Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, People's Republic of China*
[3]*Purple Mountain Laboratories, Nanjing 211111, People's Republic of China*
[4]*Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China*

(Dated: May 21, 2024)

We initialize the study of efficient verification of genuinely entangled subspaces (GES). We establish a general verification framework and provide efficient verification strategies for various GESs of practical interests. Firstly, we construct an efficient verification strategy for the GES spanned by the 3-qubit GHZ state and W state. This strategy involves one-way adaptive local measurements, which we call the "Pauli+2" strategy. Along the way, we categorize two-qubit subspaces with dimension 2 into three distinct types, which might be of independent interest. Then, we present two strategies to verify GES induced by stabilizer codes of size k in an n-qubit system. These strategies only use a limited number of Pauli measurements and are non-adaptive, thus are experimentally feasible. Notably, we present the first verification strategies for the genuinely entangled stabilizer subspaces induced by the prominent five-qubit code and the toric code.

I. INTRODUCTION

Current quantum systems often fail to work as desired due to the presence of quantum noise. Therefore, accurately describing the actual quantum system is a crucial task in quantum information. Quantum tomography is a standard method for characterizing the entire quantum system. However, it is resource-intensive, making it impractical for large quantum systems. On the other hand, in many practical scenarios, it is unnecessary to describe the entire quantum system. Consequently, many resource-efficient methods are developed to certify the quantum system [1, 2], such as fidelity estimation [3–7], entanglement detection [8–12]. Among these methods, quantum state verification [9, 13] is designed to verify whether quantum states are prepared as desired. Specifically, the verification strategies primarily focus on using local operators and classical communication (LOCC) to verify entangled states. Recently, resource-optimal verification protocols based on LOCC have been found for many groups of states, such as bipartite maximally entangled states [10], two-qubit pure states [14], GHZ states [15], stabilizer states [16, 17] and antisymmetric basis states [18].

Meanwhile, a particular line of research on entanglement in multipartite systems concerns the characterization of subspaces composed solely of entangled states, known as entangled subspaces. These special subspaces have proven useful in quantum error correction [19–22] and quantum cryptography [23]. An important type of entangled subspace is the *genuinely entangled subspace* (GES), which is defined to be a subspace of multipartite system that contains only genuinely multiparty entangled (GME) states [24–29]. Naturally, certifying GES is as important as certifying entangled states. Recently, Baccari *et al.* [25] partially addressed this certification problem by presenting the first self-testing protocols for two specific GESs, both of which are stabilizer subspaces.

In this work, we generalize quantum state verification to *quantum subspace verification*, aiming to determine whether a prepared state belongs to a GES using LOCC. Note that quantum subspace verification has been previously mentioned in [30, 31], where protocols are designed to verify ground states of local Hamiltonians. Here, we establish a general verification framework and provide efficient verification strategies for various GESs of practical interests. Firstly, we consider the GES spanned by the 3-qubit GHZ state and W state and construct an efficient strategy for it. This strategy involves one-way adaptive local measurements, which we call the "Pauli+2" strategy. Additionally, we study the verification of two-qubit subspaces with dimension 2 and categorize them into 3 distinct types. Subsequently, we investigate the subspaces determined by the stabilizer codes, referred to as stabilizer subspaces. We propose two non-adaptive verification strategies for these subspaces using only Pauli measurements. For a GES determined by k stabilizer generators, the first strategy requires $2^k - 1$ measurement settings, while the second requires only k measurement settings. However, the sample complexity of the first strategy is lower than that of the second and is independent of the size of the quantum system. Concretely, we propose the verification strategies for two genuinely entangled stabilizer subspaces determined by the five-qubit code and the toric code, respectively.

The rest of the paper is organized as follows. Section II introduces the concept of GES and stabilizer subspaces. Section III formally defines the quantum subspace verification task. Section IV presents our first result: the efficient verification of the subspace spanned by the 3-qubit GHZ state and W state. Section V devotes to constructing verification strategies for stabilizer subspaces.

* Corresponding author: yuxutao@seu.edu.cn
† Corresponding author: nju.wangkun@gmail.com

II. PRELIMINARIES

In this section, we introduce some necessary preliminaries for our work. Firstly, we define genuinely entangled states and subspaces. Then, we introduce stabilizer subspaces, which can be genuinely entangled.

A. Genuinely entangled subspace

We focus on finite-dimensional n-partite product Hilbert spaces $\mathcal{H}_{d_1,d_2,\cdots,d_n} = \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \cdots \otimes \mathbb{C}^{d_n}$, where d_i is the dimension of the local Hilbert space corresponding to the system A_i. An n-partite pure state $|\psi\rangle_{A_1\cdots A_n}$ is said to be *fully product* if it can be written as

$$|\psi\rangle_{A_1\cdots A_n} = |\varphi_1\rangle_{A_1} \otimes \cdots \otimes |\varphi_N\rangle_{A_n}. \tag{1}$$

Otherwise, it is *entangled*. Among such states, there is one distinguished class within which the quantum states are said to be *genuinely multiparty entangled* (GME).

Definition 1 ([27]). *A multipartite pure state is GME if*

$$|\psi\rangle_{A_1\cdots A_n} \neq |\varphi\rangle_S \otimes |\phi\rangle_{\bar{S}} \tag{2}$$

for any bipartite cut $S|\bar{S}$ of $A_1 \cdots A_n$.

Genuinely entangled subspace is defined to be a subspace that only contains genuinely multiparty entangled states.

Definition 2 ([27]). *A subspace $\mathcal{V} \subset \mathcal{H}_{d_1,\cdots,d_n}$ is called a GES if all pure states $|\psi\rangle \in \mathcal{V}$ are GME.*

A well-known example of GES in $\mathcal{H}_{2,2,2}$ is the one spanned by the GHZ state $(|000\rangle+|111\rangle)/\sqrt{2}$ and the W state $(|001\rangle+|010\rangle+|100\rangle)/\sqrt{3}$ [26].

B. Stabilizer subspace

Here, we consider the case where $d_1 = \cdots = d_n = 2$ and introduce the stabilizer subspace, which can be genuinely entangled. Let I, X, Y, Z be the Pauli matrices, and let G_n denotes the Pauli group on n qubits, consisting of n-fold tensor products of I, X, Y, Z with the overall factors ± 1 or $\pm i$. Consider a subset of G_n, represented as

$$\mathcal{G}_k = \{S_1, S_2, \cdots, S_k\}, \quad S_i \in G_n, 1 \leq i \leq k, \tag{3}$$

if it stabilizes a nontrivial subspace $\mathcal{V}$, that is, $S_i|\psi\rangle = |\psi\rangle$ for all $i \in [k]$ and $|\psi\rangle \in \mathcal{V}$, then, we call $\mathcal{G}_k$ a *stabilizer generator* and $\mathcal{V}$ a *stabilizer subspace* determined by $\mathcal{G}_k$. There are many stabilizer subspaces, while we are mainly concerned with genuinely entangled stabilizer subspace (GESS). A typical example of GESS is the GHZ state, which can be viewed as a subspace whose dimension is 1. On the other hand, not every stabilizer subspace is a GESS, e.g., the subspace determined by one generator Z_1Z_2, where P_i denotes the Pauli operator P on i-th qubit. Therefore, a natural problem is whether the subspace determined by $\mathcal{G}_k$ is a GESS. In [25, 26], the authors provided a simple sufficient criterion to decide whether a given stabilizer subspace is a GESS. In Section V, we will introduce a general framework to verify GESS.

III. QUANTUM SUBSPACE VERIFICATION

In this section, we first formally define the task of quantum subspace verification. Then, we introduce how to complete this task perfectly if there is no constraints. Finally, we discuss this task under local constraints, specifically focusing on verification strategies that can be implemented locally.

A. Task description

Suppose we have N copies n-qubit states $\sigma_1, \sigma_2, \cdots, \sigma_N$ produced by a same quantum device $\mathcal{D}$. The quantum state verification task answers the question: "Are the states σ_i generated by $\mathcal{D}$ equal to a fixed state $|\psi\rangle\langle\psi|$?" Similarly, the quantum subspace verification aims to answer the question:

"Are the states σ_i generated by $\mathcal{D}$ contained in the subspace $\mathcal{V}$ spanned by the orthonormal basis $\{|\psi_j\rangle\}_j$?"

To mathematically verify whether a state is in the target subspace $\mathcal{V}$ spanned by the orthonormal basis $\{|\psi_j\rangle\}_j$, we define the projector $\Pi := \sum_j |\psi_j\rangle\langle\psi_j|$ and provide the following lemma, whose proof can be found in Appendix A 1.

Lemma 3. *For a fixed quantum state σ,*

$$\mathrm{Tr}[\Pi\sigma] = \sum_j \langle\psi_j|\sigma|\psi_j\rangle = 1, \tag{4}$$

if and only if $\sigma \in \mathrm{span}\{|\psi_j\rangle\}$.

With the help of the above lemma, we can now formally define the quantum subspace verification task—Given a quantum device $\mathcal{D}$, distinguish between the following two cases:

1. **Good**: for all $i \in [N]$, $\mathrm{Tr}[\Pi\sigma_i] = 1$;
2. **Bad**: for all $i \in [N]$, $\mathrm{Tr}[\Pi\sigma_i] \leq 1 - \epsilon$ for some fixed ϵ.

A visual depiction of the quantum subspace verification task is given in Figure 1.

B. Optimal verification without constraints

Here we consider the verification task without constraints on the set of available measurements, thus entangled measurements are possible. The complexity of this "globally optimal" strategy serves as a reasonable benchmark for other resource constraint verification strategies.

We define the test POVM $\{\Omega, \mathbb{1} - \Omega\}$ where $\Omega = \Pi$. We call the outcome of Ω "pass" and the one of $\mathbb{1} - \Omega$ "fail". For an arbitrary state σ, the probability that it passes the test is

$$\Pr\{\text{"pass"}|\sigma\} = \mathrm{Tr}[\Omega\sigma] = \sum_j \langle\psi_j|\sigma|\psi_j\rangle. \tag{5}$$

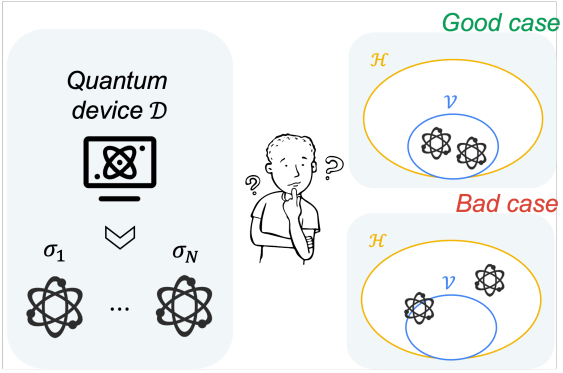


FIG. 1: Quantum subspace verification. Given a quantum device $\mathcal{D}$, we aim to distinguish exclusively between two cases: **Good** case: all states prepared is in a target subspace $\mathcal{V} \subseteq \mathcal{H}$. **Bad** case: $\exists\, i, \sigma_i \notin \mathcal{V}$.

Therefore, the state in the target subspace will be accepted with certainty. Now, we consider the **Bad** case. For states $\{\sigma_i\}_{i=1}^N$ with $\mathrm{Tr}[\Pi\sigma_i] \leq 1-\epsilon$ for all $i \in [N]$, the probability that all states pass is

$$\Pr\{\text{“pass”}|\sigma_1, \cdots, \sigma_N\} = \prod_i \mathrm{Tr}[\Pi\sigma_i] \leq (1-\epsilon)^N. \quad (6)$$

We want this probability to be bounded from above by $\delta > 0$, i.e.,

$$(1-\epsilon)^N \leq \delta \Rightarrow N \geq \frac{1}{\epsilon}\ln\frac{1}{\delta}. \quad (7)$$

This gives the least required number of states copies N.

It should be noted that the globally optimal verification strategy necessitates the use of entangled measurements if the target subspace is entangled (in which case there is at least one entangled basis state). Implementing entangled measurements is experimentally challenging. In the following, we discuss subspace verification under local constraints on measurements, yielding experimental friendly verification strategies.

C. Practical verification with local constraints

Suppose that we have access to a set of POVM elements $\mathcal{M}$ and $\forall\, M \in \mathcal{M}$, M is a local projector (assisted by classical communication). Then for each state preparation, we pick a POVM element $M \in \mathcal{M}$ with some probability and consider the corresponding two-outcomes POVMs $\{M, \mathbb{1}-M\}$, where M has output "pass" and $\mathbb{1}-M$ has output "fail". Moreover, we define a probability mass $\mu : \mathcal{M} \to [0,1]$, $\sum_{M\in\mathcal{M}} \mu(M) = 1$. The probability of a generated quantum state σ passing the test can be expressed as

$$\Pr\{\text{"pass"}|\sigma\} = \sum_{M\in\mathcal{M}} \mu(M)\,\mathrm{Tr}[M\sigma] \equiv \mathrm{Tr}[\Omega\sigma], \quad (8)$$

where the *verification operator* of this strategy is defined as

$$\Omega := \sum_{M\in\mathcal{M}} \mu(M) M. \quad (9)$$

To satisfy the requirement of the verification task, we impose two conditions on the verification operator Ω: *perfect completeness condition* and *soundness condition*. The perfect completeness condition requires that

$$\mathrm{Tr}[\Omega\sigma] = 1, \quad \forall\sigma \in \mathrm{span}\{|\psi_j\rangle\}. \quad (10)$$

This condition can be equivalently characterized using the projector Π associated with the target subspace $\mathcal{V}$ as follows; See Appendix A 2 for the proof.

Lemma 4. *The perfect completeness condition can be equivalently characterized as*

$$\mathrm{Tr}[\Omega\Pi] = \mathrm{rank}(\Pi), \quad (11)$$

where $\mathrm{rank}(\Pi)$ *is the rank of the projector.*

Now let's consider the soundness condition. We find the the worst-case passing probability $p(\Omega)$, defined as

$$p(\Omega) := \max_{\sigma:\mathrm{Tr}[\Pi\sigma]\leq 1-\epsilon} \Pr\{\text{“pass”}|\sigma\}, \quad (12)$$

in the **Bad** case is uniquely determined by the largest eigenvalue of the projected effective verification operator, as elucidated in the following theorem. The proof can be found in Appendix A 3.

Theorem 5. *It holds that*

$$p(\Omega) := \max_{\sigma:\mathrm{Tr}[\Pi\sigma]\leq 1-\epsilon} \mathrm{Tr}[\Omega\sigma] = 1 - (1 - \lambda_{\max}(\widehat{\Omega}))\epsilon, \quad (13)$$

where $\widehat{\Omega} := (\mathbb{1}-\Pi)\Omega(\mathbb{1}-\Pi)$ *is the projected effective verification operator and* $\lambda_{\max}(X)$ *denotes the maximum eigenvalue of the Hermitian operator* X.

Therefore, the probability of accepting the **Bad** case is bounded as follows,

$$\Pr\{\text{"accept"}|\sigma_1, \cdots, \sigma_N\} \leq (1 - \nu(\Omega)\epsilon)^N, \quad (14)$$

where $\nu(\Omega) := 1 - \lambda_{\max}(\widehat{\Omega})$ is the *spectral gap*. Similarly, to achieve the bound δ, we have

$$N \geq \frac{1}{\nu(\Omega)} \times \frac{1}{\epsilon}\ln\frac{1}{\delta}. \quad (15)$$

This inequality provides a guideline for constructing efficient verification by maximizing $\nu(\Omega)$.

In the following, we provide efficient verification strategies for various genuinely entangled subspaces of practical interests: the three-qubit GES spanned by the 3-qubit GHZ state and W state and the general stabilizer subspaces, including the genuinely entangled stabilizer subspaces induced by the prominent five-qubit code and the toric code as special cases.

IV. SUBSPACE SPANNED BY GHZ STATE AND W STATE

In this section, we propose an efficient verification protocol for the subspace spanned by the GHZ state and W state,

$$|\mathrm{GHZ}\rangle := (|000\rangle + |111\rangle)/\sqrt{2}, \quad (16a)$$

$$|\mathrm{W}\rangle := (|001\rangle + |010\rangle + |100\rangle)/\sqrt{3}. \quad (16b)$$

It has been proven in [26] that such a space is genuinely entangled. Our protocol is based on the result of the two-qubit subspace verification, which will be introduced first. Then, we introduce our verification strategy, termed as *Pauli+2* strategy. Lastly, we analyze the complexity of this strategy, i.e., the required number of copies of the states.

A. Verification of two-qubit subspace

For the three-qubit target space to be verified, measuring any qubit will naturally generate a two-qubit subspace on the other two qubits conditioned on the measurement outcome. Thus, it is necessary to discuss the verification of two-qubit subspaces first. Remarkably, we can categorize two-qubit subspaces with dimension 2 into three types, each with its own characteristics, detailed in Appendix B.

Firstly, we determine which kinds of subspaces can be verified. Intuitively, if the complementary subspace of the target subspace can be spanned using LOCC, then the subspace can be verified; otherwise, it cannot. Based on this principle, for a two-qubit subspace with only one product state in it, we can not verify this two-qubit subspace. Consequently, we named this kind of subspaces as *unverifiable* subspaces. On the other hand, if there are two different product states $|\tau_0\rangle, |\tau_1\rangle$ in the two-qubit subspace, then, we call it is a *verifiable* subspace. Specially, if $\langle\tau_0|\tau_1\rangle = 0$, then, we call it is a *perfectly verifiable* subspace.

With this classification, we define verification operators tailored to each type of subspace, as described in the following lemma. Further details can be found in Appendix B.

Lemma 6. *The verification operator of different kinds subspace is defined in the following form:*

1. ***unverifiable*** *subspace:*

$$\Omega_u = \mathbb{1} - |\tau\rangle\langle\tau|, \tag{17}$$

where $|\tau\rangle$ is the only one product state in the complementary subspace. And we have $\nu(\Omega_u) = 1$.

2. ***perfectly verifiable*** *subspace:*

$$\Omega_p = |\tau_0\rangle\langle\tau_0| + |\tau_1\rangle\langle\tau_1|, \tag{18}$$

where $|\tau_i\rangle (i = 0, 1)$ are product states in the target subspace. And we have $\nu(\Omega_p) = 0$.

3. ***verifiable*** *subspace:*

$$\Omega_v = \mathbb{1} - \frac{1}{2}(|\tau_2\rangle\langle\tau_2| + |\tau_3\rangle\langle\tau_3|), \tag{19}$$

where $|\tau_i\rangle (i = 2, 3)$ are the product states in the complementary subspace. And we have $\nu(\Omega_v) = \frac{1}{2}(1 + |\langle\tau_2|\tau_3\rangle|^2)$.

To practically verify a two-qubit subspace, we utilize different measurement strategies based on the types of verification operators defined in Lemma 6.

- For an unverifiable subspace, we construct a two-outcomes POVM $\{\mathbb{1} - |\tau\rangle\langle\tau|, |\tau\rangle\langle\tau|\}$, where $|\tau\rangle$ is defined in the Eq. (17). We reject states with outcomes corresponding to $|\tau\rangle\langle\tau|$. However, this strategy is inevitably fooled by an entangled state $|\tau'\rangle$, where $|\tau'\rangle$ in the complementary subspace and $\langle\tau|\tau'\rangle = 0$.

- For a perfectly verifiable subspace, we construct a two-outcomes POVM $\{|\tau_0\rangle\langle\tau_0| + |\tau_1\rangle\langle\tau_1|, \mathbb{1} - |\tau_0\rangle\langle\tau_0| - |\tau_1\rangle\langle\tau_1|\}$, where $|\tau_i\rangle (i = 0, 1)$ are defined in the Eq. (18). We pass the state with the result corresponding to the $|\tau_0\rangle\langle\tau_0| + |\tau_1\rangle\langle\tau_1|$. Notably, no states from the complementary subspace can pass this strategy.

- For a verifiable subspace, the strategy is a litter more complex than others. It involves two POVMs: $\{\mathbb{1} - |\tau_2\rangle\langle\tau_2|, |\tau_2\rangle\langle\tau_2|\}$ and $\{\mathbb{1} - |\tau_3\rangle\langle\tau_3|, |\tau_3\rangle\langle\tau_3|\}$, where $|\tau_i\rangle (i = 2, 3)$ are defined in the Eq. (19). Each POVM is performed with probability $\frac{1}{2}$ and we reject the states with the result corresponding to the $\mathbb{1} - |\tau_i\rangle\langle\tau_i| (i = 2, 3)$. Although the state in the complementary subspace can pass each test, it cannot pass with certainty.

B. One-way adaptive measurement

Based on the two-qubit subspace verification method proposed in Section IV A, we show in the following a general subroutine to construct one-way adaptive measurements that are applicable for verifying three-qubit subspaces. The construction is very intuitive: we first measure a qubit, then we verify the induced two-qubit subspace conditioned on the measurement outcome.

Assume now we choose a qubit index i and measure this qubit in Pauli operator P. The corresponding Pauli measurement is represented as P_i, where $i \in [3]$ and $P \in \{X, Y, Z\}$. For the Pauli measurement P_i, there are two possible outcomes, $+1$ and -1, corresponding to the positive and negative eigenspaces of P_i. Conditioned on the measurement outcome, the remaining two qubits will live in a two-qubit subspace, spanned by two post-measurement states, which we term the *post-measurement subspace*. We denote the post-measurement subspace resulting from measurement P_i and outcome $o \in \{+1, -1\}$ as $\mathcal{V}^o_{P_i}$. For example, if a Z_1 measurement is performed and the outcome is $+1$, our target subspace becomes a two-qubit subspace spanned by $|00\rangle$ and $(|01\rangle + |10\rangle)/\sqrt{2}$. All post-measurement subspaces are listed in Appendix C.

Secondly, we design two-qubit subspace verification strategies based on the outcome of the first measurement. With the analysis in Section IV A, we define the following two-qubit verification operators based on the different measurement outcomes, while more details can be found in Appendix C:

- $\mathcal{V}^+_{Z_i}$ is an unverifiable subspace and we define

$$M^+_{Z_i} = \mathbb{1} - |11\rangle\langle 11|. \tag{20}$$

- $\mathcal{V}_{Z_i}^-$ is a perfectly verifiable subspace and we define

$$M_{Z_i}^- = |00\rangle\langle 00| + |11\rangle\langle 11|. \tag{21}$$

- $\mathcal{V}_{X_i}^+$ is a perfectly verifiable subspace and we define

$$M_{X_i}^+ = |x_+x_+\rangle\langle x_+x_+| + |\bar{x}_+\bar{x}_+\rangle\langle \bar{x}_+\bar{x}_+|, \tag{22}$$

where $|x_+\rangle = \cos\alpha|0\rangle + \sin\alpha|1\rangle$, $|\bar{x}_+\rangle = \sin\alpha|0\rangle - \cos\alpha|1\rangle$, and $\alpha = \arctan\frac{-1+\sqrt{5}}{2}$.

- $\mathcal{V}_{X_i}^-$ is a verifiable subspace and we define

$$M_{X_i}^- = \mathbb{1} - \frac{1}{2}(|x_-x'_-\rangle\langle x_-x'_-| + |x'_-x_-\rangle\langle x'_-x_-|), \tag{23}$$

where $|x_-\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{3}}|1\rangle)$ and $|x'_-\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{-i\frac{\pi}{3}}|1\rangle)$.

- $\mathcal{V}_{Y_i}^+$ is a verifiable subspace and we define

$$M_{Y_i}^+ = \mathbb{1} - \frac{1}{2}(|y_+y'_+\rangle\langle y_+y'_+| + |y'_+y_+\rangle\langle y'_+y_+|), \tag{24}$$

where $|y_+\rangle = \cos\beta|0\rangle + e^{-i\gamma}\sin\beta|1\rangle$, $|y'_+\rangle = \sin\beta|0\rangle + e^{i(\gamma+\frac{\pi}{2})}\cos\beta|1\rangle$, and $\beta = \arctan\sqrt{\frac{\sqrt{17}+1}{2\sqrt{2}} + \frac{\sqrt{17}}{4} + \frac{1}{4}}$ and $\gamma = \arctan\tan^2\beta$.

- $\mathcal{V}_{Y_i}^-$ is a verifiable subspace and we define

$$M_{Y_i}^- = \mathbb{1} - \frac{1}{2}(|y_-y'_-\rangle\langle y_-y'_-| + |y'_-y_-\rangle\langle y'_-y_-|), \tag{25}$$

where $|y_-\rangle = \cos\beta|0\rangle + e^{i\gamma}\sin\beta|1\rangle$, $|y'_-\rangle = \sin\beta|0\rangle + e^{-i(\gamma+\frac{\pi}{2})}\cos\beta|1\rangle$, and β and γ are defined before.

Overall, the corresponding one-way adaptive measurement $\{M, \mathbb{1} - M\}$ induced by P_i has the form

$$M_{P,i} = P_i^+ \otimes M_{P_i}^+ + P_i^- \otimes M_{P_i}^-. \tag{26}$$

C. Pauli+2 strategy

Now, we are ready to describe the verification strategy, building on the one-way adaptive measurements constructed on the last section.

The verification strategy works as follows. First, we uniformly and randomly choose a qubit i and choose a measurement $P \in \{X, Y, Z\}$ according to some probability distribution $\mu(P)$, which is to be optimized. The reason why we choose the qubit uniformly is that both $|\mathrm{GHZ}\rangle$ and $|\mathrm{W}\rangle$ are symmetric with respect to the qubit indices. Given this choice, we construct an one-way adaptive measurement $M_{P,i}$ according to Eq. (26). Then we perform this measurement on the target quantum state and obtain a decision. We name the proposed strategy the "Pauli+2" strategy and illustrate it in Figure 2. Here, "Pauli" means that we first perform a randomly chosen Pauli measurement and "2" means that we perform a two-qubit subspace verification conditioned on the measurement outcome.

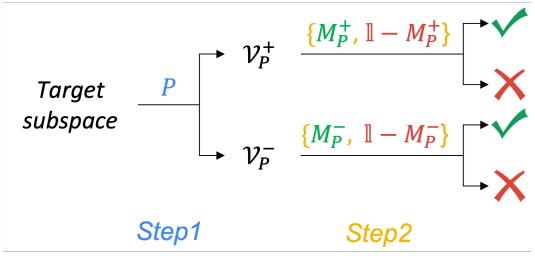


FIG. 2: The "Pauli+2" verification strategy. In the first step, we randomly perform Pauli measurement $P \in \{X, Y, Z\}$ on a randomly chosen qubit. Then, we can obtain a post-measurement subspace $\mathcal{V}_P^o$ with measurement result $o \in \{+, -\}$. Subsequently, we perform the measurement $\{M_P^o, \mathbb{1} - M_P^o\}$ defined in Section IV B. If the outcome is M_P^o, we accept; otherwise, we reject.

The verification operator of this "Pauli+2" strategy reads

$$\Omega_\mu := \sum_{P\in\{X,Y,Z\}, i\in[3]} \frac{\mu(P)}{3} M_{P,i}, \tag{27}$$

where $\mu(P)$ is a probability distribution satisfying $\sum \mu(P) = 1$ and $M_{P,i}$ is defined in Eq. (26). Obviously, $\mu(P)$ affects the performance of the Pauli+2 strategy. To find the optimal verification strategy, we solve the following optimization problem:

$$\mu^\star = \arg\max_\mu \nu(\Omega_\mu). \tag{28}$$

Numerical calculations suggest that $\nu(\Omega_{\mu^\star}) \approx 0.358$ when $\mu^\star(X) \approx 0.299$, $\mu^\star(Y) \approx 0.209$, and $\mu^\star(Y) \approx 0.4920$. Correspondingly, the required number of copies N must be

$$N \geq 2.79 \times \frac{1}{\epsilon}\ln\frac{1}{\delta} \tag{29}$$

in order to achieve a confidence level $1 - \delta$. We leave the analytic solution to Eq. (28) as an open problem.

V. STABILIZER SUBSPACE VERIFICATION

In this section, we describe two efficient protocols for verifying the stabilizer subspaces. As mentioned before, for a n-qubit system, a subspace $\mathcal{V}$ can be determined by a set of k stabilizer generators $\mathcal{G}_k$. We can construct a set of stabilizer operators $\mathcal{S}_k = \{P_{\boldsymbol{y}} : \boldsymbol{y} \in \mathbb{Z}_2^k\}$ as follows:

$$P_{\boldsymbol{y}} := \prod_{i=1}^k S_i^{y_i}. \tag{30}$$

By construction, it holds that $|\mathcal{S}_k| = 2^k$.

Protocol I works by uniformly and randomly choosing a stabilizer operator $P_{\boldsymbol{y}}$ from $\mathcal{S}_k$ and measure the target state with $P_{\boldsymbol{y}}$. If the measurement outcome is $+1$, indicating that the state lies in the positive eigenspace of $P_{\boldsymbol{y}}$, we accept; otherwise we reject. Mathematically, the verification operator of **Protocol I** reads

$$\Omega_{\mathrm{I}} := \frac{1}{2^k-1} \sum_{P\in\mathcal{S}_k\setminus\{\mathbb{1}\}} P^+, \tag{31}$$

where $P^+ := (P+\mathbb{1})/2$ is the projector onto the positive eigenspace of stabilizer operator P. We prove in Appendix D that Ω_{I} is indeed a valid verification strategy of $\mathcal{V}$. What's more, the verification efficiency, i.e. the spectral gap, of Ω_{I} satisfies

$$\nu(\Omega_{\mathrm{I}}) = \frac{2^{k-1}}{2^k-1}. \tag{32}$$

Recalling Eq. (15), to achieve a confidence level $1-\delta$, it suffices to take

$$N(\Omega_{\mathrm{I}}) = \frac{(2^k-1)}{2^{k-1}} \frac{1}{\epsilon} \ln\frac{1}{\delta} \approx 2\frac{1}{\epsilon}\ln\frac{1}{\delta} \tag{33}$$

number of state copies, which is independent with the subspace size k. Notably, this strategy necessitates at most twice as many copies as the verification strategy without local constraints. The obtained result is consistent with the result in [9] which considers the special case $k=n$. The disadvantage of **Protocol I** is self-evident: the experimenters must be able to implement a total number 2^k-1 Pauli measurement settings, which increases exponentially in the subspace size k and is challenging. This disadvantage motivates the second protocol which requires far less number of measurement settings.

Protocol II works by randomly choosing a stabilizer generator S from $\mathcal{G}_k$, each with probability $1/k$. Then, we perform the corresponding measurement and only accept the state with outcome $+1$. Mathematically, the verification operator of **Protocol II** reads

$$\Omega_{\mathrm{II}} := \frac{1}{k} \sum_{S\in\mathcal{G}_k} S^+, \tag{34}$$

where $S^+ := (S+\mathbb{1})/2$ is the projector onto the positive eigenspace of stabilizer generator S. In Appendix D, we prove that Ω_{II} is a valid verification strategy of $\mathcal{V}$. Subsequently, we examine the verification efficiency of Ω_{II}, which is

$$\nu(\Omega_{\mathrm{II}}) = \frac{1}{k}. \tag{35}$$

Therefore, it suffices to take

$$N(\Omega_{\mathrm{II}}) = k\frac{1}{\epsilon}\ln\frac{1}{\delta} \tag{36}$$

number of state copies to achieve a confidence level $1-\delta$. The obtained result is also consistent with the result in [9] which consider the special case $k=n$. **Protocol II** requires much less measurement settings than **Protocol I** but it consumes $k/2$ times more state copies. This indicates a fundamental trade-off between the total number of required state copies and the number of measurement settings, which deserves further investigation.

In the following, we present the *first* verification strategies for the genuinely entangled stabilizer subspaces induced by the prominent five-qubit code and the toric code.

Five-qubit code. Consider the GESS induced by the *five-qubit code* [25]. This subspace is generated by the following 4 generators,

$$\begin{aligned} S_1 &= X_1Z_2Z_3X_4, \quad S_2 = X_2Z_3Z_4X_5, \\ S_3 &= X_1X_3Z_4Z_5, \quad S_4 = Z_1X_2X_4Z_5. \end{aligned} \tag{37}$$

To verify such a subspace, **Protocol I** requires $2^4-1=15$ measurement settings, which are determined by 15 stabilizer operators (excluding $\mathbb{1}$) defined in Eq. (30). To achieve a confidence level $1-\delta$, we need $15/(8\epsilon)\ln 1/\delta$ state copies. On the other hand, **Protocol II** only 4 measurement settings, determined by 4 generators defined in Eq. (37). However, **Protocol II** requires more state copies, specifically $4/\epsilon\ln 1/\delta$,to achieve the same confidence level $1-\delta$.

Toric code. Consider the GESS induced by the *toric code* [25]. A toric code can be presented by a $L\times L$ lattice, where each edge represents a qubit. The corresponding stabilizer generators can be divided into two groups: (i) those associated with each lattice vertex v, with X acting on every qubit associated with an edge attached to the given vertex, and (ii) those associated with each plaquette p of the lattice, with Z acting on each qubit represented by an edge surrounding the plaquette. Mathematically, they can be written as

$$S_v = \prod_{i\in v} X_i, \quad S_p = \prod_{i\in p} Z_i. \tag{38}$$

There are $2L^2-2$ stabilizer generators in total. Therefore, **Protocol I** requires $2^{2L^2-2}-1$ measurement settings and $\approx 2/\epsilon\ln 1/\delta$ state copies to achieve a confidence level $1-\delta$. **Protocol II** only requires $2L^2-2$ measurement settings but needs $(2L^2-2)/\epsilon\ln 1/\delta$ state copies to achieve the same confidence level $1-\delta$.

Though we only provide two examples of GESSs, it should be noted that for any arbitrary GESSs, we can construct the corresponding verification strategies in a similar manner.

VI. CONCLUSIONS

This work devotes to the efficient verification of GES. We established a general verification framework and provided efficient verification strategies for two special types of GES. Firstly, we proposed a "Pauli+2" strategy based on one-way adaptive measurements to verify the GES spanned the 3-qubit GHZ state and W state. This strategy requires $\approx 2.79/\epsilon\ln 1/\delta$ copies of states to achieve confidence $1-\delta$. Then, we investigated GES determined by k stabilizer generators and constructed two non-adaptive strategies using only a few Pauli

measurements. **Protocol I** requires $2^k - 1$ measurement settings, constructed from the full stabilizer group, and consumes $\approx 2/\epsilon \ln 1/\delta$ copies of quantum states. Notably, this complexity is independent with the size of the system. **Protocol II** requires only k measurement settings, constructed from the stabilizer generators solely, and consumes $k/\epsilon \ln 1/\delta$ copies of quantum states. Notably, we presented the first verification protocols for the genuinely entangled stabilizer subspaces induced by the prominent five-qubit code and the toric code.

ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Program of China (Grant Nos. 2019YFA0308700 and 2022YFF0712800), the Jiangsu Key R&D Program Project (Grant No. BE2023011-2), the Fundamental Research Funds for the Central Universities (Grant No. 2242022k60001), and the National Natural Science Foundation of China (Grant Nos. 61960206005 and 61871111).

[1] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. Quantum certification and benchmarking. *Nature Reviews Physics*, 2(7):382–390, July 2020.

[2] Martin Kliesch and Ingo Roth. Theory of quantum system certification. *PRX Quantum*, 2(1):010201, January 2021.

[3] Steven T. Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Physical Review Letters*, 106(23):230501, June 2011.

[4] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, October 2020.

[5] Andreas Elben, Benoît Vermersch, Rick van Bijnen, Christian Kokail, Tiff Brydges, Christine Maier, Manoj Joshi, Rainer Blatt, Christian F. Roos, and Peter Zoller. Cross-platform verification of intermediate scale quantum devices. *Physical Review Letters*, 124(1):010504, January 2020.

[6] D. Zhu, Z. P. Cian, C. Noel, A. Risinger, D. Biswas, L. Egan, Y. Zhu, A. M. Green, C. Huerta Alderete, N. H. Nguyen, Q. Wang, A. Maksymov, Y. Nam, M. Cetina, N. M. Linke, M. Hafezi, and C. Monroe. Cross-platform comparison of arbitrary quantum states. *Nature Communications*, 13(1):6620, November 2022.

[7] Congcong Zheng, Xutao Yu, and Kun Wang. Cross-platform comparison of arbitrary quantum processes. *npj Quantum Information*, 10(1):1–9, January 2024.

[8] You Zhou, Pei Zeng, and Zhenhuan Liu. Single-copies estimation of entanglement negativity. *Physical Review Letters*, 125(20):200502, November 2020.

[9] Sam Pallister, Noah Linden, and Ashley Montanaro. Optimal verification of entangled states with local measurements. *Physical Review Letters*, 120(17):170502, April 2018.

[10] Huangjun Zhu and Masahito Hayashi. Optimal verification and fidelity estimation of maximally entangled states. *Physical Review A*, 99(5):052346, May 2019.

[11] Fernando G. S. L. Brandão and Reinaldo O. Vianna. Separable multipartite mixed states: Operational asymptotically necessary and sufficient conditions. *Physical Review Letters*, 93(22):220503, November 2004.

[12] Andreas Elben, Richard Kueng, Hsin-Yuan (Robert) Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, and Benoît Vermersch. Mixed-state entanglement from local randomized measurements. *Physical Review Letters*, 125(20):200501, November 2020.

[13] Xiao-Dong Yu, Jiangwei Shang, and Otfried Gühne. Statistical methods for quantum state verification and fidelity estimation. *Advanced Quantum Technologies*, 5(5):2100126, 2022.

[14] Kun Wang and Masahito Hayashi. Optimal verification of two-qubit pure states. *Physical Review A*, 100(3):032315, September 2019.

[15] Zihao Li, Yun-Guang Han, and Huangjun Zhu. Optimal verification of greenberger-horne-zeilinger states. *Physical Review Applied*, 13(5):054002, May 2020.

[16] Ninnat Dangniam, Yun-Guang Han, and Huangjun Zhu. Optimal verification of stabilizer states. *Physical Review Research*, 2(4):043323, December 2020.

[17] Siyuan Chen, Wei Xie, and Kun Wang. Memory Effects in Quantum State Verification. *arXiv:2312.11066*, December 2023.

[18] Zihao Li, Yun-Guang Han, Hao-Feng Sun, Jiangwei Shang, and Huangjun Zhu. Verification of phased dicke states. *Physical Review A*, 103(2):022601, February 2021.

[19] Daniel Alsina and Mohsen Razavi. Absolutely maximally entangled states, quantum-maximum-distance-separable codes, and quantum repeaters. *Physical Review A*, 103(2):022402, February 2021.

[20] Gilad Gour and Nolan R. Wallach. Entanglement of subspaces and error-correcting codes. *Physical Review A*, 76(4):042309, October 2007.

[21] Felix Huber and Markus Grassl. Quantum codes of maximal distance and highly entangled subspaces. *Quantum*, 4:284, June 2020.

[22] Zahra Raissi, Christian Gogolin, Arnau Riera, and Antonio Acín. Optimal quantum error correcting codes from absolutely maximally entangled states. *Journal of Physics A: Mathematical and Theoretical*, 51(7):075301, January 2018.

[23] Akshata H Shenoy and R Srikanth. Maximally nonlocal subspaces. *Journal of Physics A: Mathematical and Theoretical*, 52(9):095302, February 2019.

[24] Maciej Demianowicz. Universal construction of genuinely entangled subspaces of any size. *Quantum*, 6:854, November 2022.

[25] Flavio Baccari, Remigiusz Augusiak, Ivan Šupić, and Antonio Acín. Device-independent certification of genuinely entangled subspaces. *Physical Review Letters*, 125(26):260507, Decem-

ber 2020.
[26] Owidiusz Makuta and Remigiusz Augusiak. Self-testing maximally-dimensional genuinely entangled subspaces within the stabilizer formalism. *New Journal of Physics*, 23(4):043042, April 2021.
[27] Maciej Demianowicz and Remigiusz Augusiak. From unextendible product bases to genuinely entangled subspaces. *Physical Review A*, 98(1):012313, July 2018.
[28] Maciej Demianowicz, Grzegorz Rajchel-Mieldzioć, and Remigiusz Augusiak. Simple sufficient condition for subspace to be completely or genuinely entangled. *New Journal of Physics*, 23(10):103016, October 2021.
[29] Maciej Demianowicz and Remigiusz Augusiak. Entanglement of genuinely entangled subspaces and states: Exact, approximate, and numerical results. *Physical Review A*, 100(6):062318, December 2019.
[30] Huangjun Zhu, Yunting Li, and Tianyi Chen. Efficient verification of ground states of frustration-free hamiltonians. *Quantum*, 8:1221, January 2024.
[31] Tianyi Chen, Yunting Li, and Huangjun Zhu. Efficient verification of affleck-kennedy-lieb-tasaki states. *Physical Review A*, 107(2):022616, February 2023.
[32] Eric Chitambar, Runyao Duan, and Min-Hsiu Hsieh. When do local operations and classical communication suffice for two-qubit state discrimination? *IEEE Transactions on Information Theory*, 60(3):1549–1561, March 2014.

Appendix A: Proof of the subspace verification

1. Proof of Lemma 3

Proof of Lemma 3. The ***necessity*** is obvious. If $\sigma \in \mathrm{span}\{|\psi_j\rangle\}$, then we have

$$\sigma = \sum_{jl} \sigma_{jl} |\psi_j\rangle\langle\psi_l| \quad \Rightarrow \quad \sum_j \langle\psi_j|\sigma|\psi_j\rangle = 1. \tag{A1}$$

Now we turn to show the ***sufficiency***. For an arbitrary matrix σ, it can be written as

$$\sigma = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|, \quad \sum_i \lambda_i = 1, \quad \lambda_i \geq 0. \tag{A2}$$

For each eigenstate $|\phi_i\rangle$, we have

$$|\phi_i\rangle = \sin\theta_i |\Psi_i\rangle + \cos\theta_i |\Psi_i^\perp\rangle, \tag{A3}$$

where $\sum_j |\langle\psi_j|\Psi_i\rangle|^2 = 1$, and $|\Psi_i^\perp\rangle$ is orthogonal to $|\Psi_i\rangle$. Then, σ can also be written as

$$\sigma = \sum_i \lambda_i \left(\sin^2\theta_i |\Psi_i\rangle\langle\Psi_i| + \sin\theta_i\cos\theta_i |\Psi_i\rangle\langle\Psi_i^\perp| + \sin\theta_i\cos\theta_i |\Psi_i^\perp\rangle\langle\Psi_i| + \cos^2\theta_i |\Psi_i^\perp\rangle\langle\Psi_i^\perp|\right). \tag{A4}$$

With the trace constraint, we have

$$\sum_i \lambda_i \sin^2\theta_i = 1 \quad \Rightarrow \quad \sin\theta_i = 1, \ \forall\, \theta_i, \quad \Rightarrow \quad \sigma = \sum_i \lambda_i |\Psi_i\rangle\langle\Psi_i|, \tag{A5}$$

which hints that $\sigma \in \mathrm{span}\{|\psi_i\rangle\}$. □

2. Proof of Lemma 4

Proof of Lemma 4. With perfect completeness condition, there exist a set of orthogonal bases $\{|\psi_l^\perp\rangle\}_l$ in the complementary subspace of the target subspace, such that Ω can be written as

$$\Omega = \Pi + \sum \omega_l |\psi_l^\perp\rangle\langle\psi_l^\perp|, \tag{A6}$$

otherwise $\forall\ \sigma \in \mathrm{span}\{|\psi_j\rangle\}, \mathrm{Tr}[\Omega\sigma] = 1$ does not hold. We define the projected effective verification operator as

$$\widehat{\Omega} := (\mathbb{1} - \Pi)\Omega(\mathbb{1} - \Pi) = \sum \omega_l |\psi_l^\perp\rangle\langle\psi_l^\perp|. \tag{A7}$$

Therefore, we have

$$\mathrm{Tr}[\Omega\Pi] = \mathrm{Tr}[\Pi^2] + \mathrm{Tr}[\widehat{\Omega}\Pi] = \mathrm{Tr}[\Pi] = \mathrm{rank}(\Pi). \tag{A8}$$

□

3. Proof of the Theorem 5

Proof of the Theorem 5. For a fixed set $\{|\psi_l^\perp\rangle\}$, an arbitrary quantum state σ with $\mathrm{Tr}[\Pi\sigma] = r$ can always be written as

$$\sigma = r\Psi + (1-r)\Psi^\perp + \sum_{jl}\left(c_{jl}|\psi_j\rangle\langle\psi_l^\perp| + c_{jl}^*|\psi_l^\perp\rangle\langle\psi_j|\right), \tag{A9}$$

where Ψ and $\Psi^\perp$ are the states in the $\mathrm{span}\{|\psi_j\rangle\}$ and $\mathrm{span}\{|\psi_l^\perp\rangle\}$, respectively. Then, such a state will pass the test with probability

$$\begin{aligned}\Pr\{\text{"pass"}|\sigma\} &= \mathrm{Tr}[\Omega\sigma] &\text{(A10)}\\ &= r\,\mathrm{Tr}[\Omega\Psi] + (1-r)\,\mathrm{Tr}[\widehat{\Omega}\Psi^\perp] &\text{(A11)}\\ &\leq r + (1-r)\lambda_{\max}(\widehat{\Omega}). &\text{(A12)}\end{aligned}$$

The above inequality becomes an equality if

$$\Psi^\perp = |\psi_{\max}^\perp\rangle\langle\psi_{\max}^\perp|, \tag{A13}$$

where $|\psi_{\max}^\perp\rangle$ is the eigenstate of $\widehat{\Omega}$ corresponding to the largest eigenvalue $\lambda_{\max}(\widehat{\Omega})$. Thus, for a fixed Ω,

$$\max_{\sigma:\mathrm{Tr}[\Pi\sigma]=r}\Pr\{\text{"pass"}|\sigma\} = r + (1-r)\lambda_{\max}(\widehat{\Omega}), \tag{A14}$$

which is achieved by any density matrix of the form

$$\sigma = r\Psi + (1-r)|\psi_{\max}^\perp\rangle\langle\psi_{\max}^\perp| + \sum_{jl}\left(c_{jl}|\psi_j\rangle\langle\psi_l^\perp| + c_{jl}^*|\psi_l^\perp\rangle\langle\psi_j|\right). \tag{A15}$$

Note that the pure state $\sigma = |\phi\rangle\langle\phi|$ for

$$|\phi\rangle = \sqrt{r}|\psi'\rangle + \sqrt{1-r}|\psi_{\max}^\perp\rangle, \tag{A16}$$

where $|\psi'\rangle$ is the linear combination of vectors $|\psi_j\rangle$, is of this form. Therefore, we can only consider pure states in the following analysis.

Now, for a fixed $\bar{\epsilon} \geq \epsilon > 0$, we define a state $\sigma = |\phi_{\bar{\epsilon}}\rangle\langle\phi_{\bar{\epsilon}}|$ with $|\phi_{\bar{\epsilon}}\rangle = \sqrt{1-\bar{\epsilon}}|\phi\rangle + \sqrt{\bar{\epsilon}}|\phi^\perp\rangle$, where $|\phi\rangle$ is the linear combination of vectors $\{|\psi_j\rangle\}_j$ and $\langle\phi|\phi^\perp\rangle = 0$. Then, we define that the worst-case passing probability as

$$\begin{aligned}p(\Omega) &:= \max_{\sigma:\mathrm{Tr}[\Omega\sigma]\leq 1-\epsilon}\Pr\{\text{"pass"}|\sigma\} &\text{(A17)}\\ &= \max_{\sigma:\mathrm{Tr}[\Omega\sigma]\leq 1-\epsilon}\mathrm{Tr}[\Omega\sigma] &\text{(A18)}\\ &= \max_{\bar{\epsilon}\geq\epsilon,|\phi^\perp\rangle} 1 - \bar{\epsilon} + \bar{\epsilon}\langle\phi^\perp|\widehat{\Omega}|\phi^\perp\rangle &\text{(A19)}\\ &= 1 - (1-\lambda_{\max}(\widehat{\Omega}))\epsilon. &\text{(A20)}\end{aligned}$$

□

Appendix B: Proof of two-qubit subspace verification

Before the detailed analysis, we introduce some necessary preliminaries. A general 2-qubit pure state $|\psi\rangle$ can be uniquely represented by the 2×2 matrix ψ given by

$$|\psi\rangle = \mathbb{1}\otimes\psi|\Phi\rangle, \tag{B1}$$

where $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. One measure of the entanglement possessed by $|\psi\rangle$ is its *concurrence*, which is defined by $C(\psi) = |\det(\psi)| \leq 1$. With this representation, we have the following properties.

1. If $C(\psi) = 0$, $|\psi\rangle$ is a product state.

2. With the condition $\langle\alpha|\beta\rangle = 0$, we have

$$\langle\Phi|(\mathbb{1}\otimes\alpha)(\mathbb{1}\otimes\beta)|\Phi\rangle = 0 \quad \Rightarrow \quad \mathrm{Tr}[\alpha^\dagger\beta] = 0. \tag{B2}$$

3. If the subspace spanned by $|\alpha\rangle$ and $|\beta\rangle$ is a *local subspace*, we have [32, Lemma 1].

$$C(\alpha) = C(\beta). \tag{B3}$$

Then, we begin our analysis of two-qubit subspace verification.

1. When a two-qubit subspace is verifiable with local constraints?

Now consider a subspace $\mathcal{V}$ spanned by two orthogonal states, $\{|\psi_0\rangle, |\psi_1\rangle\}$ and its complementary subspace is denoted as $\mathcal{V}^\perp$. We can assume that $|\psi_1\rangle$ is a *product state* without loss of generality, as the maximal dimension of two-qubit CES is 1. We have the following lemma for relationship between the number of the product states in the $\mathcal{V}$ and $\mathcal{V}^\perp$.

Lemma 7. *The number of the different product states in the $\mathcal{V}$ is equal to its in the $\mathcal{V}^\perp$.*

Proof. Firstly, if there are two product states in $\mathcal{V}$, labeled as

$$|a_1\rangle\otimes|a_0\rangle, \quad |b_1\rangle\otimes|b_0\rangle, \tag{B4}$$

where $|a_i\rangle, |b_i\rangle$ are single-qubit states, then, there are also two product states in $\mathcal{V}^\perp$,

$$|\bar{a}_1\rangle\otimes|\bar{b}_0\rangle, \quad |\bar{b}_1\rangle\otimes|\bar{a}_0\rangle, \tag{B5}$$

where $|\bar{a}_i\rangle\langle\bar{a}_i| + |a_i\rangle\langle a_i| = \mathbb{1}$ (like wise for $|\bar{b}_i\rangle$), $i = 0, 1$.

Then, assume that $|\psi_1\rangle$ is the only one product state in $\mathcal{V}$. If there are two different product states in $\mathcal{V}^\perp$, then, with the previous analysis, there are two different product states in $\mathcal{V}$, which conflicts with the assumption. Therefore, there are also only one product state in $\mathcal{V}^\perp$. □

With the following lemma, we can easily compute the number of product states in $\mathcal{V}$.

Lemma 8. *For a two-qubit subspace $\mathcal{V}$ spanned by two states $|\alpha\rangle$ and $|\beta\rangle$ (not necessary orthogonal), where $|\beta\rangle$ is an entangled state, if $\alpha\beta^{-1}$ has two different eigenvalues, then there are two different product states in this subspace.*

Proof. The problem of finding all product states in $\mathcal{V}$ can be expressed as

$$\det(\alpha + \lambda\beta) = 0 \tag{B6}$$

$$\det(\alpha\beta^{-1} + \lambda\mathbb{1})\det(\beta) = 0 \tag{B7}$$

$$\det(\alpha\beta^{-1} + \lambda\mathbb{1}) = 0. \tag{B8}$$

If there are two different solution of λ, i.e., $\alpha\beta^{-1}$ has two different eigenvalues, then we have two different projector states in $\mathcal{V}$. □

Then, we try to show that whether $\mathcal{V}$ is verifiable depend on the number of product states in it. If there are two different product states in $\mathcal{V}^\perp$, then we can span $\mathcal{V}^\perp$ with local states. It hints that we can verify this subspace with two test projectors:

$$M_i = \mathbb{1} - |\tau_i\rangle\langle\tau_i|, \quad i = 0, 1, \tag{B9}$$

where $|\tau_i\rangle$ are the product states in $\mathcal{V}^\perp$. Therefore, $\mathcal{V}$ is *verifiable* and the corresponding verification operator is

$$\Omega = \frac{1}{2}\sum_i M_i = \mathbb{1} - \frac{1}{2}(|\tau_0\rangle\langle\tau_0| + |\tau_1\rangle\langle\tau_1|). \tag{B10}$$

Specially, if these two states are orthogonal, i.e., $\langle\phi_0|\phi_1\rangle = 0$, then the verification operator becomes

$$\Omega = \mathbb{1} - (|\tau_0\rangle\langle\tau_0| + |\tau_1\rangle\langle\tau_1|). \tag{B11}$$

We call this kind of subspace a *perfectly verifiable* subspace. On the other hand, if there are only one product states in $\mathcal{V}^\perp$, then we can not span $\mathcal{V}^\perp$ with local states. We call this kind of subspace an *unverifiable* subspace. It hints that we can only reject this product in the test, i.e., the corresponding verification operator is

$$\Omega = \mathbb{1} - |\tau\rangle\langle\tau|, \tag{B12}$$

where $|\tau\rangle$ is the only product state in $\mathcal{V}^\perp$.

2. Spectral gap analysis

Here we analysis the complexity of the strategy proposed in the previous subsection.

- Firstly, it is obvious that for *unverifiable* subspace, the spectral gap of Ω defined in the Eq. (B12) is 1. It means that $|\psi_2\rangle$ always can fool this strategy, where $C(\psi_2) = C(\psi_0)$ and $\langle\psi_i|\psi_2\rangle = 0$ for $i = 0, 1$.
- Secondly, it is also easy to find that for *perfectly verifiable* subspace, the spectral gap of Ω defined in Eq. (B11) is 0, i.e., no state can fool this strategy.
- Lastly, we have the following lemma for the spectral gap of Ω defined in the Eq. (B10).

Lemma 9. *Suppose there are two different product states $|a_1a_0\rangle$ and $|b_1b_0\rangle$ (not orthogonal) in $\mathcal{V}$, the spectral gap of Ω defined in the Eq.* (B10) *is*

$$\nu(\Omega) = \frac{1}{2}(1 + |\langle a_1a_0|b_1b_0\rangle|^2). \tag{B13}$$

Proof. The product states in the $\mathcal{V}^\perp$ are $|\bar{a}_1\bar{b}_0\rangle$ and $|\bar{b}_1\bar{a}_0\rangle$. The verification operator is

$$\Omega = \mathbb{1} - \frac{1}{2}\left(|\bar{a}_1\bar{b}_0\rangle\langle\bar{a}_1\bar{b}_0| + |\bar{b}_1\bar{a}_0\rangle\langle\bar{b}_1\bar{a}_0|\right), \tag{B14}$$

where $|\bar{a}_i\rangle\langle\bar{a}_i| + |a_i\rangle\langle a_i| = \mathbb{1}$ (like wise for $|\bar{b}_i\rangle$), $i = 0, 1$. Note that each state in $\mathcal{V}^\perp$ can be written as the linear combination of $|\bar{a}_1\bar{b}_0\rangle$ and $|\bar{b}_1\bar{a}_0\rangle$. So we can define that $|\phi\rangle = x|\bar{a}_1\bar{b}_0\rangle + y|\bar{b}_1\bar{a}_0\rangle \in \mathcal{V}^\perp$ without normalization, and have

$$\begin{aligned}
&\langle\phi|\Omega|\phi\rangle \\
=&\langle\phi|\phi\rangle - \frac{1}{2}\left|\langle\bar{a}_1\bar{b}_0|\phi\rangle\right|^2 - \frac{1}{2}\left|\langle\bar{b}_1\bar{a}_0|\phi\rangle\right|^2 && \text{(B15)}\\
=&\langle\phi|\phi\rangle - \frac{1}{2}\left[x^2 + 2\Re\left(x^*y\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle\right) + y^2|\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2\right] - \frac{1}{2}\left[x^2|\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2 + 2\Re\left(xy^*\langle\bar{b}_1\bar{a}_0|\bar{a}_1\bar{b}_0\rangle\right) + y^2\right] && \text{(B16)}\\
=&\langle\phi|\phi\rangle - \frac{1}{2}(x^2 + y^2)\left(1 + |\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2\right) - 2\Re\left(x^*y\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle\right) && \text{(B17)}\\
=&\frac{1}{2}(x^2 + y^2)\left(1 - |\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2\right). && \text{(B18)}
\end{aligned}$$

With normalization, we have

$$\begin{aligned}
\frac{\langle\phi|\Omega|\phi\rangle}{\langle\phi|\phi\rangle} &= \frac{1}{2}\frac{(x^2 + y^2)\left(1 - |\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2\right)}{x^2 + y^2 + 2\Re\left(x^*y\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle\right)} && \text{(B19)}\\
&= \frac{1}{2}\frac{1 - |\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2}{1 + 2\Re\left(\frac{x^*y}{x^2+y^2}\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle\right)} && \text{(B20)}\\
&= \frac{1}{2}\frac{1 - |\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2}{1 + 2\Re\left(\frac{1}{\frac{x}{y} + \frac{y^*}{x^*}}\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle\right)} && \text{(B21)}\\
&\leq \frac{1}{2}(1 - |\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2), && \text{(B22)}
\end{aligned}$$

when $x = 0$ or $y = 0$, the equality is achieved. Therefore, the spectral gap of strategy Ω is

$$\begin{aligned}
\nu(\Omega) &= \frac{1}{2}(1 + |\langle\bar{a}_1\bar{b}_0|\bar{b}_1\bar{a}_0\rangle|^2) && \text{(B23)}\\
&= \frac{1}{2}(1 + |\langle a_1a_0|b_1b_0\rangle|^2). && \text{(B24)}
\end{aligned}$$

□

Appendix C: Proof of the special case

In this section, we show the detail analysis of our special case: the subspace spanned by the following two states,

$$|\mathrm{GHZ}\rangle = (|000\rangle + |111\rangle)/\sqrt{2}, \tag{C1a}$$

$$|\mathrm{W}\rangle = (|001\rangle + |010\rangle + |100\rangle)/\sqrt{3}. \tag{C1b}$$

Firstly, we compute all 2-qubit post-measurement subspaces. Due to the symmetry of GHZ state and W state, we obtain the same post-measurement subspace with same measurements and outcomes, no matter which qubit is performed the first measurement. Therefore, we we omit the subscript i without loss of generality in the following. And all post-measurement states of different measurements and outcomes are illustrated in the Table I.

first measurement		post-measurement states	
Pauli	outcome	$\|\mathrm{GHZ}\rangle$	$\|\mathrm{W}\rangle$
Z	$+$	$\|00\rangle$	$\frac{1}{\sqrt{2}}(\|01\rangle + \|10\rangle)$
	$-$	$\|11\rangle$	$\|00\rangle$
X	$+$	$\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$	$\frac{1}{\sqrt{3}}(\|01\rangle + \|10\rangle + \|00\rangle)$
	$-$	$\frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$	$\frac{1}{\sqrt{3}}(\|01\rangle + \|10\rangle - \|00\rangle)$
Y	$+$	$\frac{1}{\sqrt{2}}(\|00\rangle - i\|11\rangle)$	$\frac{1}{\sqrt{3}}(\|01\rangle + \|10\rangle - i\|00\rangle)$
	$-$	$\frac{1}{\sqrt{2}}(\|00\rangle + i\|11\rangle)$	$\frac{1}{\sqrt{3}}(\|01\rangle + \|10\rangle + i\|00\rangle)$

TABLE I: Post-measurement states for the subspace spanned by $\{|\mathrm{GHZ}\rangle, |\mathrm{W}\rangle\}$ defined in Eq. (C1).

Subsequently, we need to construct the verification strategies of post-measurement subspaces, which have been introduced in Appendix B. In the following, we show the concrete analysis case by case and we label the post-measurement subspace with measurement $P \in \{X, Y, Z\}$ and outcome $o \in \{+, -\}$ as $\mathcal{V}_P^o$.

For the subspace $\mathcal{V}_Z^+$, there is only one product state $|11\rangle$ in its complementary subspace. Thus, it is an unverifiable subspace and the corresponding verification operator is

$$M_Z^+ = \mathbb{1} - |11\rangle\langle 11|. \tag{C2}$$

For the subspace $\mathcal{V}_Z^-$, obviously, it is a perfectly verifiable subspace with

$$M_Z^- = |00\rangle\langle 00| + |11\rangle\langle 11|. \tag{C3}$$

For the subspace $\mathcal{V}_X^+$, we can find two product states, $|x_+x_+\rangle$ and $|\bar{x}_+\bar{x}_+\rangle$, in its complementary subspace, where

$$|x_+\rangle = \cos\alpha|0\rangle + \sin\alpha|1\rangle, \quad |\bar{x}_+\rangle = \sin\alpha|0\rangle - \cos\alpha|1\rangle, \quad \alpha = \arctan\frac{-1+\sqrt{5}}{2}. \tag{C4}$$

Additionally, we have $\langle x_+x_+|\bar{x}_+\bar{x}_+\rangle = 0$. Thus, it is also a perfectly verifiable subspace with verification operator

$$M_X^+ = |x_+x_+\rangle\langle x_+x_+| + |\bar{x}_+\bar{x}_+\rangle\langle \bar{x}_+\bar{x}_+|. \tag{C5}$$

For the subspace $\mathcal{V}_Z^-$, we can find two product states, $|x_-x'_-\rangle$ and $|x'_-x_-\rangle$, in its complementary subspace, where

$$|x_-\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{3}}|1\rangle), \quad |x'_-\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{-i\frac{\pi}{3}}|1\rangle). \tag{C6}$$

As $\langle x_-x'_-|x'_-x_-\rangle \neq 0$, V_Z^- is a verifiable subspace with

$$M_Z^- = \mathbb{1} - \frac{1}{2}\left(|x_-x'_-\rangle\langle x_-x'_-| + |x'_-x_-\rangle\langle x'_-x_-|\right) \tag{C7}$$

For the subspace $\mathcal{V}_Y^+$, we can find two product states, $|y_+y'_+\rangle$ and $|y'_+y_+\rangle$, in its complementary subspace, where

$$|y_+\rangle = \cos\beta|0\rangle + e^{-i\gamma}\sin\beta|1\rangle, \quad |y'_+\rangle = \sin\beta|0\rangle + e^{i(\gamma+\frac{\pi}{2})}\cos\beta|1\rangle, \tag{C8}$$

$$\beta = \arctan\sqrt{\frac{\sqrt{\sqrt{17}+1}}{2\sqrt{2}} + \frac{\sqrt{17}}{4} + \frac{1}{4}}, \quad \gamma = \arctan\tan^2\beta. \tag{C9}$$

As $\langle y_+y'_+|y'_+y_+\rangle \neq 0$, V_Y^+ is a verifiable subspace with

$$M_Y^+ = \mathbb{1} - \frac{1}{2}(|y_+y'_+\rangle\langle y_+y'_+| + |y'_+y_+\rangle\langle y'_+y_+|). \tag{C10}$$

For the subspace $\mathcal{V}_Y^-$, we can find two product states, $|y_-y'_-\rangle$ and $|y'_-y_-\rangle$, in its complementary subspace, where

$$|y_-\rangle = \cos\beta|0\rangle + e^{i\gamma}\sin\beta|1\rangle, \quad |y'_-\rangle = \sin\beta|0\rangle + e^{-i(\gamma+\frac{\pi}{2})}\cos\beta|1\rangle. \tag{C11}$$

As $\langle y_-y'_-|y'_-y_-\rangle \neq 0$, V_Y^- is a verifiable subspace with

$$M_Y^- = \mathbb{1} - \frac{1}{2}(|y_-y'_-\rangle\langle y_-y'_-| + |y'_-y_-\rangle\langle y'_-y_-|). \tag{C12}$$

Appendix D: Proof of the stabilizer subspace verification

In this section, we prove the two verification strategies of stabilizer subspace. We prove the strategy with stabilizer operators first, then provide the proof of strategy with stabilizer generators.

1. Proof of Protocol I

For the subspace $\mathcal{V}$ determined by $\mathcal{G}_k$, we define a set of orthogonal bases in $\mathcal{V}$ as

$$\{|\psi_1\rangle, \cdots, |\psi_{2^{n-k}}\rangle\}. \tag{D1}$$

The set of stabilizer operators is defined as $\mathcal{S}_k$, and the projector onto $\mathcal{V}$ can be defined in the following two ways,

$$\Pi_V = \frac{1}{2^k}\sum_{P\in\mathcal{S}_k} P = \sum_{j=1}^{2^{n-k}} |\psi_j\rangle\langle\psi_j|. \tag{D2}$$

We know that a feasible verification strategy Ω must be in the following form:

$$\Omega = \sum_{j=1}^{2^{n-k}} |\psi_j\rangle\langle\psi_j| + \sum_{l=1}^{2^n-2^{n-k}} \omega_l|\psi_l^\perp\rangle\langle\psi_l^\perp| \tag{D3}$$

where $\{|\psi_1^\perp\rangle, \cdots, |\psi_{2^n-2^{n-k}}^\perp\rangle\}$ is a set of orthogonal bases in complementary subspace of $\mathcal{V}$. Additionally, the spectral gap of Ω is

$$\nu(\Omega) = 1 - \max_l \ \omega_l. \tag{D4}$$

Now, we begin our proof. For each $P \in \mathcal{S}_k \setminus \{\mathbb{1}\}$, we have

$$P = P^+ - P^-, \quad P^+ + P^- = \mathbb{1}, \tag{D5}$$

where P^+ (P^-) is the projector onto the positive (negative) eigenspace of P. With the above decomposition, we have

$$\sum_{j=1}^{2^{n-k}} |\psi_j\rangle\langle\psi_j| = \frac{1}{2^k}\sum_{P\in\mathcal{S}_k\setminus\{\mathbb{1}\}} \left(P^+ - P^-\right) + \frac{1}{2^k}\mathbb{1} \tag{D6}$$

$$= \frac{1}{2^k}\sum_{P\in\mathcal{S}_k\setminus\{\mathbb{1}\}} \left(2P^+ - \mathbb{1}\right) + \frac{1}{2^k}\mathbb{1} \tag{D7}$$

$$= \frac{1}{2^{k-1}}\sum_{P\in\mathcal{S}_k\setminus\{\mathbb{1}\}} P^+ - \left(1 - \frac{1}{2^{k-1}}\right)\mathbb{1}. \tag{D8}$$

Then, we have

$$\sum_{P\in\mathcal{S}_k\setminus\{\mathbb{1}\}} P^+ = 2^{k-1}\sum_{j=1}^{2^{n-k}} |\psi_j\rangle\langle\psi_j| + \left(2^{k-1}-1\right)\mathbb{1} \tag{D9}$$

$$= 2^{k-1}\sum_{j=1}^{2^{n-k}} |\psi_j\rangle\langle\psi_j| + \left(2^{k-1}-1\right)\left(\sum_j |\psi_j\rangle\langle\psi_j| + \sum_l |\psi_l^\perp\rangle\langle\psi_l^\perp|\right) \tag{D10}$$

$$= (2^k-1)\sum_{j=1}^{2^{n-k}} |\psi_j\rangle\langle\psi_j| + \left(2^{k-1}-1\right)\sum_l |\psi_l^\perp\rangle\langle\psi_l^\perp|. \tag{D11}$$

Finally, we derive the desired equation

$$\frac{1}{2^k-1}\sum_{P\in\mathcal{S}_k\setminus\{\mathbb{1}\}} P^+ = \sum_{j=1}^{2^{n-k}} |\psi_j\rangle\langle\psi_j| + \frac{2^{k-1}-1}{2^k-1}\sum_l |\psi_l^\perp\rangle\langle\psi_l^\perp|, \tag{D12}$$

which hints the Ω_{I} defined in Eq. (31) is feasible and

$$\nu(\Omega_{\mathrm{I}}) = 1 - \frac{2^{k-1}-1}{2^k-1} = \frac{2^{k-1}}{2^k-1}. \tag{D13}$$

2. Proof of Protocol II

With the definition of Ω_{II} in Eq.(34), we have

$$\mathrm{Tr}[\Omega_{\mathrm{II}}\Pi_V] = \frac{1}{2^k\cdot k}\sum_{S\in\mathcal{G}_k, P\in\mathcal{S}_k}\mathrm{Tr}[PS^+] \tag{D14}$$

$$= \frac{1}{2^k\cdot k}\left(\frac{1}{2}\sum_{S\in\mathcal{G}_k, P\in\mathcal{S}_k}\mathrm{Tr}[P] + \mathrm{Tr}[PS]\right) \tag{D15}$$

$$= \frac{1}{2^{k+1}\cdot k}\left[k\cdot 2^n + \sum_{S\in\mathcal{G}_k} 2^n\right] \tag{D16}$$

$$= \frac{2^{n+1}\cdot k}{2^{k+1}\cdot k} = 2^{n-k} = \mathrm{rank}(\Pi_V). \tag{D17}$$

Thus, Ω_{II} satisfies perfect completeness condition defined in Lemma 4. Subsequently, we analyze the spectral gap of Ω_{II}. We define a set of complete stabilizer generators

$$\mathcal{G}_n = \{\underbrace{S_1,\cdots,S_k}_{\mathcal{G}_k}, S_{k+1},\cdots,S_n\}. \tag{D18}$$

Then, we can construct a set of orthogonal bases $|C_w\rangle$ with n-bit strings $\{w\}$, where

$$|C_{\boldsymbol{w}}\rangle\langle C_{\boldsymbol{w}}| = \prod_{j=1}^{n}\frac{\mathbb{1}+(-1)^{w_j}S_j}{2}, \quad \boldsymbol{w}\in\mathbb{Z}_2^n. \tag{D19}$$

Obviously, $|C_{\boldsymbol{w}}\rangle$ is also a stabilizer state for all $\boldsymbol{w}$ [16]. And there is a subset $W\subseteq\mathbb{Z}_2^n$, for all $\boldsymbol{w}\in W$, $|C_{\boldsymbol{w}}\rangle\in\mathcal{V}$. In other word, for a fixed $\boldsymbol{w}\in W$, the first k bits of it are all zeros. Then, we can define arbitrary state in $\mathcal{V}^\perp$ as

$$|\Psi^\perp\rangle = \sum_{\boldsymbol{w}\in W^\perp}\alpha_w|C_w\rangle, \quad \sum_{\boldsymbol{w}}|\alpha_{\boldsymbol{w}}|^2 = 1, \tag{D20}$$

and we have

$$\langle \Psi^\perp | \Omega_{\mathrm{II}} | \Psi^\perp \rangle = \frac{1}{k} \sum_{i=1}^{k} \langle \Psi^\perp | S_i^+ | \Psi^\perp \rangle \tag{D21}$$

$$= \frac{1}{k} \sum_{i=1}^{k} \sum_{\boldsymbol{w}, \boldsymbol{w}' \in W^\perp} \alpha_{\boldsymbol{w}}^* \alpha_{\boldsymbol{w}'} \langle C_{\boldsymbol{w}} | S_i^+ | C_{\boldsymbol{w}'} \rangle \tag{D22}$$

$$= \frac{1}{k} \sum_{i=1}^{k} \sum_{\boldsymbol{w}, \boldsymbol{w}' \in W^\perp} \alpha_{\boldsymbol{w}}^* \alpha_{\boldsymbol{w}'} \delta_{\boldsymbol{w}\boldsymbol{w}'} \epsilon_{i,\boldsymbol{w}'} \tag{D23}$$

$$= \frac{1}{k} \sum_{i=1}^{k} \sum_{\boldsymbol{w} \in W^\perp} |\alpha_{\boldsymbol{w}}|^2 \epsilon_{i,\boldsymbol{w}}, \tag{D24}$$

where $W^\perp = \mathbb{Z}_2^n \setminus W$, $S_i^+ |C_{\boldsymbol{w}}\rangle = \epsilon_{i,\boldsymbol{w}} |C_{\boldsymbol{w}}\rangle$, and

$$\epsilon_{i,\boldsymbol{w}} = \begin{cases} 1 & i\text{-th bit of } \boldsymbol{w} \text{ is } 0 \\ 0 & \text{else} \end{cases} . \tag{D25}$$

Therefore, we have

$$\langle \Psi^\perp | \Omega_{\mathrm{II}} | \Psi^\perp \rangle = \frac{1}{k} \sum_{\boldsymbol{w} \in W^\perp} |\alpha_{\boldsymbol{w}}|^2 \left(\sum_{i=1}^{k} \epsilon_{i,\boldsymbol{w}} \right) \tag{D26}$$

$$\leq \frac{k-1}{k} \sum_{\boldsymbol{w} \in W^\perp} |\alpha_{\boldsymbol{w}}|^2 = \frac{k-1}{k}, \tag{D27}$$

with the fact that for the first k bits of $\boldsymbol{w} \in W^\perp$, there are at most $k-1$ bits equal to 0. Additionally, it should be note that the above equality is achievable. Thus, we have

$$\nu(\Omega_{\mathrm{II}}) = 1 - \frac{k-1}{k} = \frac{1}{k}. \tag{D28}$$

Overlapping Tomography of Quantum Processes

Yi Hu[1 3] Congcong Zheng[1 2 3] Xiaojun Wang[1 3 4] Xutao Yu[1 2 4] Ping Xu[5] Kun Wang[5 *]

[1] *School of Information Science and Engineering, Southeast University, Nanjing 210096, China*
[2] *State Key Lab of Millimeter Waves, Southeast University, Nanjing 211189, China*
[3] *Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, China*
[4] *Purple Mountain Laboratories, Nanjing 211111, China*
[5] *Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China*

1 Overview of the results

We focus on local quantum processes of a multipartite quantum process and consider two questions:

1. How are the outputs of certain subsystems affected when inputs of other subsystems are altered? and
2. What resources are needed to simulate the process's impact on each subsystem?

A visual depiction of the these problems is given in Figure 1. We completely answer these questions and make the following contributions.

Main result 1: We propose a unified theory of *reduced quantum processes* with mathematically rigorous definitions and equivalent characterizations in different representations. According to this theory, to obtain information about local processes, all other subsystems interacting with the subsystem of interest must be initialized in a maximally mixed state.

Main result 2: We introduce a framework called Quantum Process Overlapping Tomography (QPOT), comprising three different methods, to efficiently characterize all k-reduced quantum processes of an n-qubit quantum process. *This framework generalizes Cotler and Wilczek's results [1] from quantum states (static resources) to quantum processes (dynamics resources).* Specifically, **Method 1** requires the fewest measurement settings, $3 + (9^k - 3) \cdot \log(2n)$. While it requires n ancilla qubits and the preparation of maximally entangled states, the other two methods do not. **Method 2** requires approximately $k \cdot n^{k-1} \cdot 12^k$ measurement settings. For general quantum processes, although the generalization of Method 2 still works, it remains resource-intensive. **Method 3** requires only $2^l \cdot 12^k \cdot \log(n)$ measurement settings, where $l = \sum_i l_i$, assuming the target quantum process has i local interactions, each with l_i qubits ($\sum_i l_i \leq n$). All of these methods represent a significant improvement over direct quantum process tomography, which requires approximately $\binom{n}{k} e^{\mathcal{O}(k)} \sim n^k \cdot e^{\mathcal{O}(k)}$ settings if k is small compared to n. Our methods are experimentally confirmed in IBM hardware.

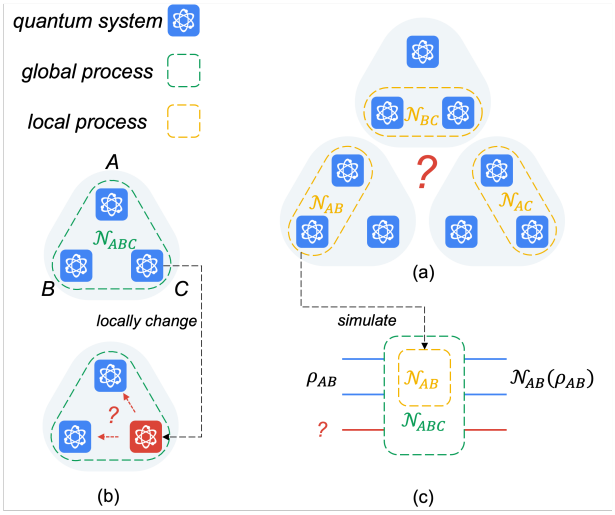


Figure 1: (a) Given a quantum process $\mathcal{N}_{ABC}$ acting on this system, how to determine its local processes $\mathcal{N}_{AB}, \mathcal{N}_{BC}$ and $\mathcal{N}_{AC}$? (b) For a global input composed of local inputs, how would a local output be influenced by other local inputs? (c) What resources are required to simulate local processes?

Our findings provide a powerful toolbox to analyze the reduced quantum processes and characterize them efficiently. We believe the results are beneficial to the broader audience of AQIS, *especially to those who are working in characterizing, verifying, and validating quantum devices of large scale.*

A full technical version can be found in the attached technical PDF.

2 Reduced quantum processes

Given a global process, determining its local processes can be a challenging task. Let $[n] := \{1, \cdots, n\}$ and $\boldsymbol{s} \in [n]$ be a set of integers. Let $A_{[n]} \equiv A_1 \cdots A_n$ be a quantum register of n qubits and $A_{\boldsymbol{s}} = \otimes_{s \in \boldsymbol{s}} A_s$ be a

*nju.wangkun@gmail.com

subset qubits of the quantum register $A_{[n]}$. We can define a reduced quantum state $\rho_{A_{\boldsymbol{s}}}$ of the global state $\rho_{A_{[n]}}$ as

$$\rho_{A_{\boldsymbol{s}}} = \mathrm{Tr}_{A_{\overline{\boldsymbol{s}}}}\left[\rho_{A_{[n]}}\right], \tag{1}$$

where $\overline{\boldsymbol{s}} := [n]\backslash\boldsymbol{s}$. It inspires us to explore the local processes of each subsystem by tracing out the other subsystems from the global process. Thus, all local quantum processes $\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}\left(\rho_{A_{\boldsymbol{s}}}\right)$ of a global process $\mathcal{N}_{A_{[n]}\to B_{[n]}}$ can be directly obtained by tracing out the other subsystems of a localizable quantum process:

$$\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}\left(\rho_{A_{\boldsymbol{s}}}\right) = \mathrm{Tr}_{B_{\overline{\boldsymbol{s}}}}\left[\mathcal{N}_{A_{[n]}\to B_{[n]}}\left(\rho_{A_{[n]}}\right)\right]. \tag{2}$$

The formula in Eq. (2) remains applicable for extracting information about local processes from an input-output flow perspective. However, the input state of Eq. (2) is a global state. Naturally, we have a problem that how to define a local process with local states. In the technical version, we define the k-reduced quantum process as follows.

Definition 1 (k-reduced quantum process) *Let $\mathcal{N}_{A_{[n]}\to B_{[n]}}(\cdot) := U(\cdot)U^\dagger$ be an n-qubit quantum unitary channel given by the unitary $U_{A_{[n]}}$. Let $\boldsymbol{s} \subseteq [n]$ be a subset of qubit indices of size k. The k-reduced quantum process $\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}$ of $\mathcal{N}$, acting on the qubit indices $\boldsymbol{s}$, is defined as follows:*

$$\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}(\rho_{A_{\boldsymbol{s}}}) := \mathrm{Tr}_{B_{\overline{\boldsymbol{s}}}}\left[U\left(\rho_{A_{\boldsymbol{s}}}\otimes\frac{\mathbb{1}_{A_{\overline{\boldsymbol{s}}}}}{2^{n-k}}\right)U^\dagger\right], \tag{3}$$

where $\rho_{A_{\boldsymbol{s}}}$ is an arbitrary quantum state in $A_{\boldsymbol{s}}$ and $\overline{\boldsymbol{s}} := [n]\backslash\boldsymbol{s}$.

There are different but equivalent representations of quantum processes, e.g., Choi representation and Pauli transfer matrix representation. In the technical version, we characterize the reduced quantum processes in these different representations and demonstrate their equivalence.

3 Quantum process overlapping tomography: Theory

There are $\binom{n}{k}$ k-reduced quantum processes for an n-qubit quantum process. To efficiently obtain all k-reduced quantum processes, we introduce three methods in the following subsections, accompanied by a detailed protocol and complexity analysis.

3.1 Method 1

Here, we introduce our initial protocol, leveraging the overlapping tomography technique mentioned. Recall that a quantum process can be expressed by a Choi state, and the Choi state of a k-qubit quantum process is a $2k$-qubit quantum state. Method 1 is a direct application of state overlapping tomography based on [1]. Following the calculations in [1], Method 1 requires $3+(9^k-3)\cdot\log_2(2n)$ measurement settings.

The primary drawback of this method is the construction of the Choi state, which needs n ancilla qubits to prepare a $2n$-qubit maximally entangled state, necessitating a quantum device of double the size. This leads to increased resource consumption and is impractical for large system. The other two method is designed without the ancilla qubits.

3.2 Method 2

Here, we consider a localizable quantum process first, where any reduced quantum process is a local quantum process. This implies that we can characterize all reduced quantum processes independently. The goal of Method 2 is to maximize the parallelism. Initially, we partition the n qubits into n/k groups, each containing k qubits. For each group, standard quantum process tomography is performed, necessitating 12^k measurement settings. All n/k groups employ the same measurement settings simultaneously. Therefore, we can obtain all k-reduced quantum processes with approximately $\binom{n}{k}\cdot 12^k/(n/k) \sim k\cdot n^{k-1}\cdot 12^k$ measurement settings.

However, with Definition 1, we know that if there are interactions between subsystems, the reduced quantum processes cannot be directly obtained. Suppose there are m qubits that interact with this k-reduced quantum process. We have to simulate a maximally mixed state of m qubits, 2^m orthonormal bases needs to be created. Therefore, the complexity for characterizing all k-reduced quantum processes is $\sum_{i=1}^{n^k}\cdot 2^{m_i}\cdot 12^k$, where m_i is the number of qubits that interact with the i-th k-reduced quantum process. Although the generalization of Method 2 still works for general quantum processes, it remains resource-intensive.

3.3 Method 3

Now, we tackle the challenge of characterizing all k-reduced quantum processes when a global process comprises many interactions among local processes. Suppose an n-qubit quantum process has i local interactions, each contains l_i qubits. To ensure that any arbitrary k-reduced quantum process contains a mixture of maximally mixed states, we prepare all local subsystems involved in interactions in maximally mixed states and then combine them. The procedure is as follows:

Step 1: We generate overlapping bases using the $\log(n)$ perfect hash functions. Therefore, there are a total of $12^k\cdot\log(n)$ bases, which cover 12 measurement settings of any single-qubit process tomography. This procedure is sufficient to characterize all k-reduced quantum processes of localizable quantum processes, as they are all composed of single-qubit processes.

Step 2: In this step, we iterate over each qubit involved in interactions with other qubits and append a complementary basis based on its current basis to all overlapping bases generated in Step 1. For example, during the iteration for qubit 0, we will add $\{|1\rangle, X\}$ if the current preparation and measurement basis of qubit 0 is $\{|0\rangle, X\}$. Each iteration doubles the length of the overlapping basis, so after all iterations, the size of the overlapping basis will be multiplied by $2^{\sum_i l_i}$. In contrast to Method 2, which involves summing over

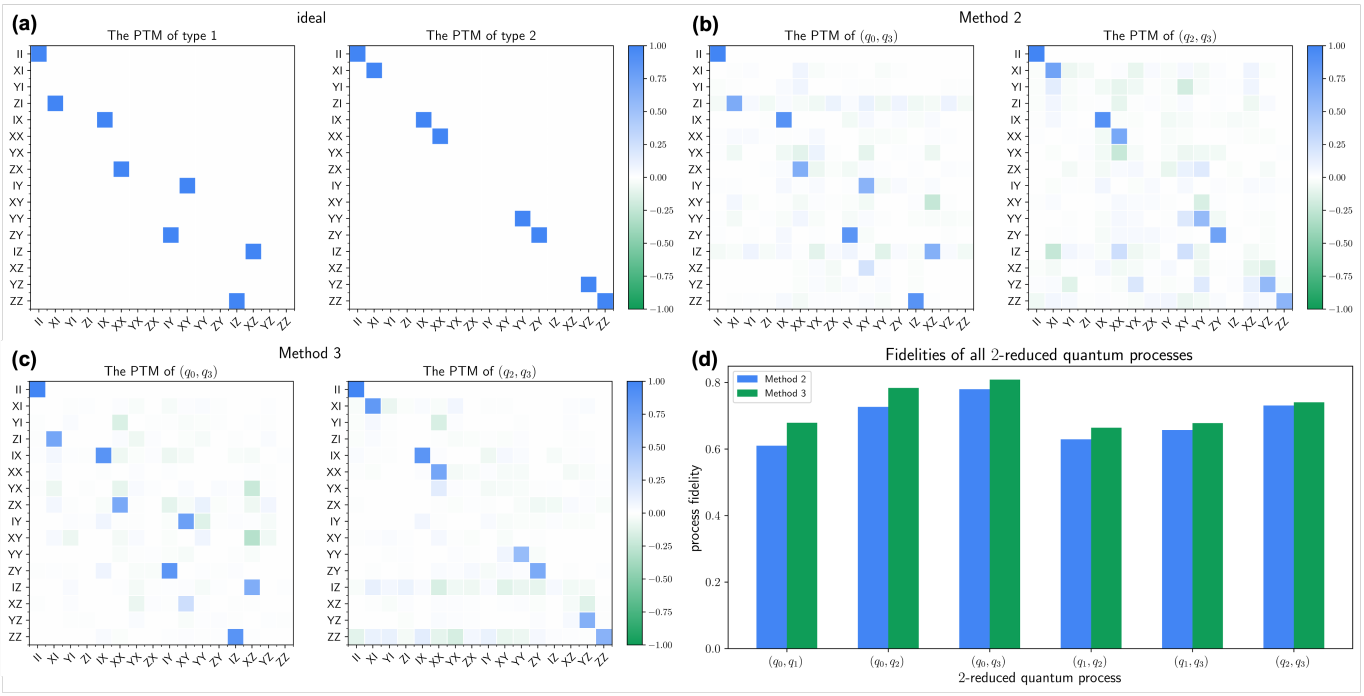


Figure 2: Experiment results of 4-qubit GHZ state preparation process on the *IBM-brisbane* device. (q_i, q_j) denotes the 2-reduced quantum process acting on i-th and j-th qubit. **(a)** Two types of ideal 2-reduced quantum processes. Type 1 contains (q_0, q_1), (q_0, q_2) and (q_0, q_3). On the other hand, type 2 contains (q_1, q_2), (q_1, q_3) and (q_2, q_3). **(b)** The 2-reduced quantum processes (q_0, q_3) and (q_2, q_3) obtained by Method 2. **(c)** The 2-reduced quantum processes (q_0, q_3) and (q_2, q_3) obtained by Method 3. **(d)** The reduced process fidelity of Method 2 and Method 3.

all k-reduced quantum processes and includes numerous repeated sums, in Method 3, $\sum_i l_i$ represents the sum over all local interactions (noting that $\sum_i l_i \leq n$). The total number of measurement settings required is $2^{\sum_i l_i} \cdot 12^k \cdot \log(n)$.

4 Quantum process overlapping tomography: Experiment

In this section, we provide an illustrative example of the process for preparing a 4-qubit GHZ state. We examine 2-reduced quantum processes derived from this quantum process as an illustration, utilizing the notation (q_i, q_j) to denote the 2-reduced process acting on the i-th and j-th qubits. There exist two types of 2-reduced quantum processes, denoted as type 1 and type 2, respectively. Their PTMs are depicted in Figure 2 (a). Type 1 quantum process comprises 2-reduced quantum processes: $(q_0, q_1), (q_0, q_2), (q_0, q_3)$, where the qubits are directly interacted with CNOT gates. Type 2 quantum process encompasses $(q_1, q_2), (q_1, q_3), (q_2, q_3)$, where the qubits are indirectly interacted.

Then, we implement Method 2 and Method 3 on the *IBM-brisbane* device and certify this device. We present two PTMs for each method, as depicted in Figure 2 (b) and (d), respectively. To quantitatively assess the similarity, we calculate the process fidelity for each 2-reduced quantum process obtained by different methods. For different 2-reduced processes, Method 3 exhibits higher fidelities, possibly due to differences in implementation times. Additionally, we observe that the reduced processes with the highest fidelity for each type are the same: (q_0, q_3) and (q_2, q_3). This suggests that q_3 may be more stable than the other qubits.

Acknowledgements

Y.H. and C.-C.Z. contributed equally to this work. Part of this work was done when K.W. was at the Institute for Quantum Computing, Baidu Research. This work was supported by the National Key Research and Development Program of China (Grant Nos. 2019YFA0308700 and 2022YFF0712800), the Jiangsu Key R&D Program Project (Grant No. BE2023011-2), the Fundamental Research Funds for the Central Universities (Grant No. 2242022k60001), and the National Natural Science Foundation of China (Grant Nos. 61960206005 and 61871111). We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

References

[1] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Physical Review Letters*, 124(10):100401, 2020.

Quantum Process Overlapping Tomography

Yi Hu[1,3], Congcong Zheng[1,2,3], Xiaojun Wang[1,3,4], Xutao Yu[1,2,3,4], Ping Xu[5], and Kun Wang[5]

[1]School of Information Science and Engineering, Southeast University, Nanjing 210096, China
[2]State Key Lab of Millimeter Waves, Southeast University, Nanjing 211189, China
[3]Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, China
[4]Purple Mountain Laboratories, Nanjing 211111, China
[5]Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China

Quantum process tomography is the gold standard for fully characterizing quantum processes, yet it is resource-intensive. In this work, we shift our focus from global to local quantum processes, which are beneficial for practical scenarios such as distributed quantum computing and quantum networks. We term these local processes as *reduced quantum processes* and present a comprehensive theory to describe them. To efficiently characterize all k-reduced quantum processes of an n-qubit quantum process, we introduce a framework called Quantum Process Overlapping Tomography (QPOT), which comprises three methods. Method 1 requires the fewest measurement settings, $3+(9^k-3)\cdot\log(2n)$. While it requires n ancilla qubits and the preparation of maximally entangled states, the other two methods do not. We consider the localizable process first and propose Method 2, which requires approximately $k\cdot n^{k-1}\cdot 12^k$ measurement settings. For general quantum processes, although the generalization of Method 2 still works, it remains resource-intensive. Therefore, we propose Method 3, which is more efficient than Method 2. Suppose the target quantum process has i local interactions, each with l_i qubits ($\sum_i l_i \leq n$). Method 3 requires only $2^l\cdot 12^k\cdot\log(n)$ measurement settings, where $l=\sum_i l_i$. All of these methods represent a significant improvement over direct quantum process tomography, which requires approximately $\binom{n}{k}e^{\mathcal{O}(k)}\sim n^k\cdot e^{\mathcal{O}(k)}$ settings if k is small compared to n. The efficacy of our methods is confirmed through experiments conducted on IBM hardware, aligning well with our theoretical predictions.

Contents

Xutao Yu: yuxutao@seu.edu.cn
Kun Wang: nju.wangkun@gmail.com

1 Introduction

Over the last decade, significant progress has been made in the development of quantum devices. These advancements include systems with entangled noisy qubits at an intermediate scale across various physical platforms, such as photons, trapped ions, and superconductors [1–4], with further advancements on the horizon. In the coming years, it is expected that quantum devices will scale up to hundreds or thousands of qubits, unlocking a broad spectrum of quantum computing applications with profound scientific and technological implications. To enhance the performance of current quantum devices, it is essential to employ methods capable of characterizing complex noisy processes. This capability is crucial for advancing error mitigation techniques in near-term applications [5–8]. Quantum Process Tomography (QPT) serves as a standard method for the diagnostic and comprehensive characterization of quantum processes, and it has been extensively studied and applied in various experimental settings [9–15]. While QPT demonstrates considerable efficacy, it is associated with substantial resource consumption. The requirement for a full quantum process tomography involves an informationally complete set of measurement settings, leading to a number of measurements that grows exponentially with the system size. Specifically, it necessitates $e^{\mathcal{O}(n)}$ measurement settings to characterize a process in an n-qubit quantum system [16]. As a result, experimentally implementing quantum process tomography remains impractical even for systems with dozens of qubits [17].

However, in many practical scenarios, the focus may not necessarily be on exploring the entirety of the quantum process but rather on specific parts of interest. For example, in applications like distributed quantum computing [18, 19] and quantum networks [20], each party possesses a local quantum processor connected to others via quantum links. To enhance the quality of global quantum operations, it is necessary to diagnose and mitigate errors on these local processors. Moreover, spatially correlated errors, such as depolarizing errors in nearest-neighbor qubits, are inherent in near-term quantum hardware [21–27]. Understanding and analyzing such error processes can help reduce logical error rates.

Local quantum processes are commonly referred to as tensor-product processes. This means that a process is considered local with respect to, for example, a bipartite system, if the entire system process can be expressed in tensor-product form $\mathcal{N} = \mathcal{N}_1 \otimes \mathcal{N}_2$, where $\mathcal{N}_1$ and $\mathcal{N}_2$ act on density operators in the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. In other words, the total operation can be decomposed into operations that act only locally on each subsystem. To characterize such local processes, we can independently perform quantum process tomography at separate locations to understand their characteristics. Taking this one step further, what if there are correlations between subsystems? Can we investigate these local processes through individual process tomography alone, or do we need additional operations? How can we rigorously describe the local processes of such a system?

To address these issues, we extend the concept of local quantum processes to more general cases. We first introduce a unified theory of *reduced quantum processes* to describe all general local processes. Without loss of generality, we explore the interaction between subsystems, ranging from no interaction to full interaction, and develop a comprehensive theory of reduced quantum processes. According to this theory, to obtain information about local processes, all other subsystems interacting with the subsystem of interest must be initialized in a maximally mixed state. Consider a k-qubit local quantum process (with no correlation) within an n-qubit system, which represents a specific case of reduced quantum processes. Characterizing each local quantum process requires $e^{\mathcal{O}(k)}$ measurement settings. Hence, naively, we would require approximately $\binom{n}{k}e^{\mathcal{O}(k)} \sim n^k \cdot e^{\mathcal{O}(k)}$ settings if k is small compared to n. For systems with interactions, the complexity grows exponentially with the scale of interactions. For instance, in a 4-qubit system with two local interactions—between qubits 0 and 1, and qubits 2 and 3—to characterize a 2-reduced quantum process involving qubits 1 and 2, qubits 0 and 3 must be initialized in a maximally mixed state. To simulate the behavior of maximally mixed state, one needs to prepare qubit 0 and 3 in all $2^2 = 4$ two-qubit computational bases to construct the local process of qubits 1 and 2. In this case, the number of measurement settings amounts to $4 \cdot e^{\mathcal{O}(2)}$. Consequently, the complexity of characterizing all 2-reduced quantum processes scales approximately as $\sim 4 \cdot n^2 \cdot e^{\mathcal{O}(2)}$. Generally, if the average number of qubits that need to be initialized in a maximally mixed state for characterizing a k-reduced quantum process is l, the overall complexity becomes $\sim 2^l \cdot n^k \cdot e^{\mathcal{O}(k)}$. As n grows, directly conducting quantum process tomography for all these reduced quantum processes becomes impractical.

In this work, we introduce an generalize framework, Quantum Process Overlapping Tomography (QPOT), designed for simultaneously characterizing all reduced quantum processes of a system. It leverages an overlapped basis generation technique, drawing inspiration from and generalizing the results of quantum overlapping tomography of quantum states originally developed by Cotler and Wilczek [28], which has been further expanded and refined in subsequent works [29–33]. We propose 3 methods to characterize all k-reduced quantum processes. Method 1 is the most efficient one and only requires $3 + (9^k - 3) \cdot \log(2n)$ measurement settings. However, it requires an additional n ancilla qubits and the preparation of $2n$-qubit maximally entangled states, which are impractical as n grows. The other two methods are designed without these requirements. Similarly, we consider local quantum processes first and Method 2 requires approximately $k \cdot n^{k-1} \cdot 12^k$ measurement settings. For general quantum processes, the generalization of Method 2 requires approximately $\sum_{i=1}^{n^k} 2^{m_i} \cdot 12^k$ measurement settings, where m_i is the number of qubits that interact with the i-th reduced quantum process. Although Method 2 still works, it remains resource-intensive. Therefore, we propose Method 3, which is more efficient than Method 2. Suppose the target quantum process has i local interactions, each with l_i qubits ($\sum_i l_i \leq n$). Method 3 requires only $2^l \cdot 12^k \cdot \log(n)$ measurement settings, where $l = \sum_i l_i$.

The rest of the paper is organized as follows. Section 2 presents preliminary concepts required by this paper. Section 3 provides a rigorous mathematical description of reduced quantum processes. Section 4 introduces the quantum process overlapping tomography framework to characterize the reduced quantum processes. Section 4 experimentally validates the quantum process overlapping tomography framework in IBM quantum devices.

2 Preliminaries

We review some definitions and notations used in this paper to elucidate our work.

2.1 Quantum processes

Recall that a general quantum process acting on a finite d-dimensional Hilbert space is a completely positive trace-preserving (CPTP) linear map [34, 35]. This description captures the evolution of a quantum system under a physical process, where the input ρ_A and output ρ_B are density operators. We denote such a process as $\mathcal{N}_{A\to B}$. The Trace-Preserving (TP) condition ensures that, after undergoing certain evolution, the state remains identifiable with unit probability, a reasonable assumption in physical reality. The Completely Positive (CP) requirement arises in situations where system A is considered as part of a joint system AR by tracing out any reference system R. It should be noted that after some process on system A, the positivity of not only the density operator ρ_A, but also ρ_{AR} acting on the joint system, should be preserved. There are various equivalent representations of CPTP maps. In this work, we predominantly employ the following four representations to describe an n-qubit quantum process $\mathcal{N}$: Stinespring representation, Kraus representation, Choi representation, and Pauli transfer matrix representation [36].

The most intuitive understanding of a linear map satisfying CPTP conditions is that it allows the system of interest to undergo an evolution alongside an external environment, where the entire system experiences a unitary evolution. This concept is known as the *Stinespring representation*. One can describe the process as follows:

$$\mathcal{N}_{A\to B}(\cdot) = \mathrm{Tr}_R\left[U\left((\cdot)\otimes\rho_R\right)U^\dagger\right], \tag{1}$$

where U is an unitary operator acting on the whole quantum system AR which comprises the principal system of interest and the environment (reference system), Tr_R denotes the partial trace over the environment, and ρ_R is the initial state of the environment. The Stinespring representation inherently satisfies the CPTP conditions: The entire system is characterized by a density operator and the partial trace ensures the preservation of its positivity.

The Stinespring representation can also be formulated in an operator-sum form, commonly known as the *Kraus representation*, expressed as:

$$\mathcal{N}_{A\to B}(\cdot) = \sum_i K_i(\cdot)K_i^\dagger, \tag{2}$$

where $\{K_i\}_i$ are linear operators acting on the Hilbert space $\mathcal{H}_A$ and the completely positive condition requires $\sum_i K_i^\dagger K_i = \mathbb{1}$, where $\mathbb{1}$ is the identity matrix. A correspondence between the Kraus and Stinespring representations can be established by identifying the operator K_i as the linear operator $\langle e_i|U|\rho_R\rangle$ on $\mathcal{H}_A$, where $\{|e_i\rangle\}_i$ forms an orthonormal basis for the environment's Hilbert space. When the CPTP map is represented by a unitary operator U, the process simplifies to a unitary transformation, given by $\mathcal{N}_{A\to B}(\cdot) = U(\cdot)U^\dagger$.

Another representation linking quantum processes to quantum states is derived from the Choi-Jamiołkowski (CJ) isomorphism, which enables the representation of CPTP maps as density operators. applying it to half of a maximally entangled $2n$-qubit state, defined as:

$$J_{A'B} = (\mathrm{id}_{A'}\otimes\mathcal{N}_{A\to B})(\Gamma_{A'A}), \tag{3}$$

where $\mathrm{id}_{A'}$ is the identity map acting on the ancillary system A', $|\Gamma\rangle_{A'A} := \sum_{i=0}^{d-1}|i\rangle_{A'}\otimes|i\rangle_A/\sqrt{d}$ is the $2n$-qubit maximally entangled state on the joint system $A'A$, $d = 2^n$, and $\{|i\rangle_A\}_i$ is an orthonormal basis of $\mathcal{H}_A$. The CP condition renders the Choi state a positive semidefinite operator, and the TP condition requires that $\mathrm{Tr}_B\, J_{A'B} = \mathbb{1}_{A'}/d$.

The last representation worth noting is the *Pauli Transfer Matrix (PTM)* representation, originating from the technique of experimentally determining a quantum process, defined as:

$$(R_\mathcal{N})_{ij} := \frac{1}{d}\mathrm{Tr}\left[P_i\mathcal{N}_{A\to B}(P_j)\right], \tag{4}$$

where we denote $\{P_i\}_i$ as the set of Pauli operators acting on an n-qubit quantum system, with $P_i \in \{I, X, Y, Z\}^{\otimes n}$ and $i = 1, ..., d^2$. By definition, $R_{\mathcal{N}}$ is a $d^2] \times d^2$ matrix. A direct transformation between the Choi operator and the PTM can be found as follows:

$$(R_{\mathcal{N}})_{ij} = \mathrm{Tr}\left[J_{AB}(P_j^T \otimes P_i)\right], \tag{5}$$

$$J_{AB} = \frac{1}{d^2} \sum_{ij} (R_{\mathcal{N}})_{ij} \left(P_j^T \otimes P_i\right), \tag{6}$$

where T denote the matrix transposition with respect to the orthonormal basis defining $|\Gamma\rangle_{A'A}$.

2.2 Quantum process tomography

In this section, we provide a brief overview of the standard quantum process tomography (QPT), which primarily involves preparing informationally complete inputs and subsequently measuring the output states using quantum state tomography [36]. The objective of QPT is to reconstruct the entire PTM of a target quantum process.

To measure the PTM of an n-qubit quantum process $\mathcal{N}$, one first needs to prepare d^2 linearly-independent states that can span the operator space of all density matrices. Subsequently, for each state, quantum state tomography is performed using d^2 measurement bases, resulting in a total of d^4 measurement settings. We denote the preparation states as $\{\rho_i\}_i$ and the measurement operators as $\{E_j\}_j$. Experimentally, one can select $\{\rho_i\}_i = \{|0\rangle, |1\rangle, |+\rangle, |r\rangle\}^{\otimes n}$, where $|+\rangle$ and $|r\rangle$ correspond to eigenstates of the Pauli X and Y operators with eigenvalue $+1$, respectively. The measurement operators $\{E_j\}_i$ can be chosen as $\{X, Y, Z\}^{\otimes n}$. Therefore, $4^n \times 3^n = 12^n$ measurement settings can be used to reconstruct the PTM of a quantum process. Generally, one performs a sequence of measurements and obtains outcome probabilities according to the Born's rule:

$$p_{ij} := \mathrm{Tr}\left[E_j \mathcal{N}\left(\rho_i\right)\right]. \tag{7}$$

We can rewrite Eq. (7) in superoperator formalism by defining a vector $|\rho\rangle\!\rangle$ whose elements are $\langle\!\langle i|\rho\rangle\!\rangle := \frac{1}{d}\mathrm{Tr}\left[P_i\rho\right]$ and $\langle\!\langle E|j\rangle\!\rangle = \mathrm{Tr}\left[EP_j\right]$. In this way, Pauli operator P_i can be represented as $|i\rangle\!\rangle$ and we can rewrite p_{ij} as

$$p_{ij} = \langle\!\langle E_j|R_{\mathcal{N}}|\rho_i\rangle\!\rangle = \sum_{kl} \langle\!\langle E_j|k\rangle\!\rangle\langle\!\langle k|R_{\mathcal{N}}|l\rangle\!\rangle\langle\!\langle l|\rho_i\rangle\!\rangle, \tag{8}$$

with the fact that $\sum_k |k\rangle\!\rangle\langle\!\langle k| = \sum_l |l\rangle\!\rangle\langle\!\langle l| = \mathbb{1}$. Because the vectors $\langle\!\langle E_j|$ and $|\rho_i\rangle\!\rangle$ are chosen by the experimenter, the matrices $\langle\!\langle E_j|k\rangle\!\rangle$ and $\langle\!\langle l|\rho_i\rangle\!\rangle$ are known in advance. Vectorizing the $R_{\mathcal{N}}$ matrix as $\boldsymbol{r_{\mathcal{N}}}$ and $\{|E_j\rangle\!\rangle\langle\!\langle\rho_i|\}_{ij}$ as $\boldsymbol{S}$, Eq. (8) can be compactly expressed as

$$\boldsymbol{p} = \boldsymbol{S^T r_{\mathcal{N}}}. \tag{9}$$

The PTM can be obtained by linear inversion estimation as

$$\boldsymbol{r_{\mathcal{N}}} = \left(\boldsymbol{S^T}\right)^{-1} \boldsymbol{p}. \tag{10}$$

In cases where $\boldsymbol{S}$ is not full-rank, least-squares estimation can be used to obtain the PTM by

$$\boldsymbol{r_{\mathcal{N}}} = \left(\boldsymbol{S S^T}\right)^{-1} \boldsymbol{S p}. \tag{11}$$

2.3 Overlapping tomography of quantum states

Quantum overlapping tomography (QOT) is a technique designed to efficiently obtain all k-body reduced density matrices (k-RDMs) of an n-partite quantum state, originally introduced by Cotler

and Wilczek [28]. Since then, extended work building upon this approach has emerged, finding applications in fields like quantum chemistry and many-body physics [29–33]. Here, we offer a concise overview of this method.

QOT can characterizes all k-RDMs of an n-qubit state with approximately $e^{\mathcal{O}(n)}\log(n)$ single-qubit measurement settings. The key insight is that measuring a particular subsystem can provide significant information about all other subsystems that overlap with it. This insight suggests the potential for efficient information extraction through parallel measurements, achieved by designing informationally complete measurement bases and suitable data post-processing. The pivotal mathematical tool for generating these measurement bases is the (n,k) families of perfect hash functions [32, 33, 37–50], extensively studied in theoretical computer science. In the context of overlapping tomography, these functions partition n qubits into k groups, with qubits in the same group undergoing the same measurement. Subsequently, measurement settings are assigned in a structured manner to ensure coverage of the complete tomographic basis for any k-RDM.

The main contribution of this work is that, we generalize Cotler and Wilczek's results [28] from quantum states to quantum processes and leverage the technique of QOT to conduct quantum process tomography effectively for all reduced quantum processes.

3 Reduced quantum processes

Given a global quantum process, determining its local processes on individual qubits can be a challenging task. In the context of distributed quantum computing, a global quantum operation is typically executed through the combination of local operations performed by spatially separated quantum processors. In such cases, the local processes might seem known, as they originate from the operation of local processors. However, it is important to remember that these local processors may be entangled due to some interactions. Such interactions imply that changes in one qubit can influence the outputs of others, complicating the characterization of local processes. We address this challenge from the perspective of a multipartite system. When considering a global quantum operation on such a system, understanding its local processes involves two key questions:

1. How are the outputs of certain subsystems affected when inputs of other subsystems are altered? and
2. What resources are needed to simulate the process's impact on each subsystem?

A visual depiction of the local processes problem is given in Figure 1.

We begin with a simple case: *localizable quantum processes*, which correspond to those that can be expressed in a tensor-product form. Consider a multipartite system $A = A_1 \cdots A_n$ with an input state $\rho_{A_1\cdots A_n} = \rho_{A_1} \otimes \cdots \otimes \rho_{A_n}$, where $\rho_{A_1}, \cdots, \rho_{A_n}$ represent the inputs for each subsystem. We denote the quantum process acting on this multipartite system as $\mathcal{N}_{A_1\cdots A_n \to B_1\cdots B_n}(\rho_{A_1\cdots A_n})$. This work only considers qubit systems for simplicity, where $A_1 \cong \cdots \cong A_n \cong B_1 \cong \cdots \cong B_n \cong \mathbb{C}^2$. However, we note that the theory can easily be generalized to qudit systems. A localizable quantum process can be written as:

$$\mathcal{N}_{A_1\cdots A_n \to B_1\cdots B_n}(\rho_{A_1\cdots A_n}) = \mathcal{N}_{A_1\to B_1}(\rho_{A_1}) \otimes \cdots \otimes \mathcal{N}_{A_n\to B_n}(\rho_{A_n}), \tag{12}$$

where $\mathcal{N}_{A_1\to B_1}(\rho_{A_1}), \cdots, \mathcal{N}_{A_n\to B_n}(\rho_{A_n})$ are local quantum processes acting on subsystems. It is clear that changes in one subsystem do not influence the others, and local resources suffice to construct each local quantum process. So, why focus on a multipartite system? The reason is that we can always express the input state of the entire system as the tensor product of the input states of each subsystem, which can be equivalently represented as $\rho_{A_i} = \mathrm{Tr}_{A_1\cdots A_{i-1}A_{i+1}\cdots A_n}[\rho_{A_1\cdots A_n}]$.

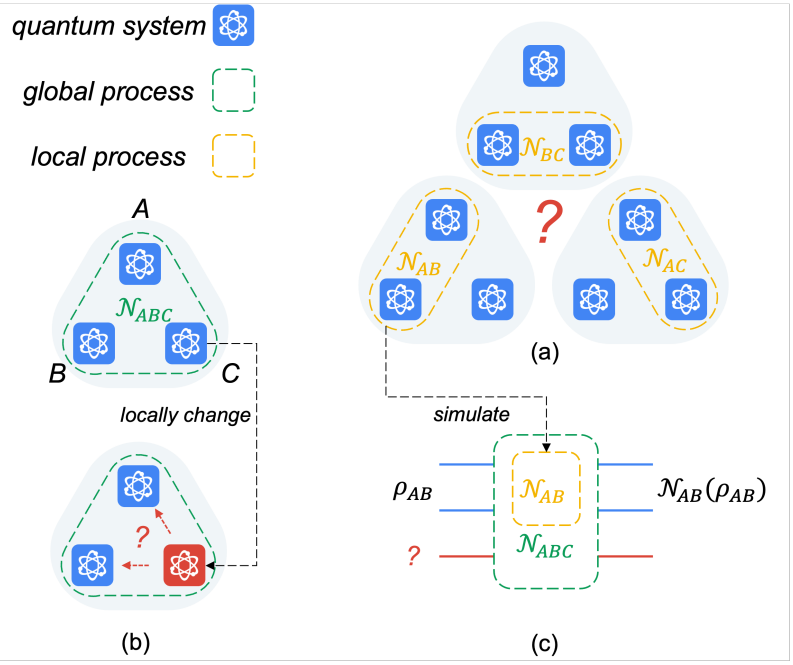


Figure 1: (a) Given a global process $\mathcal{N}_{ABC}$ acting on a multipartite system ABC, how can one determine its local processes $\mathcal{N}_{AB}$,$\mathcal{N}_{BC}$ and $\mathcal{N}_{AC}$? (b) For a global input composed of local inputs, how would a local output (e.g., $\mathrm{Tr}_C\left[\mathcal{N}_{ABC}\left(\rho_{ABC}\right)\right]$) be influenced by other local inputs? (c) To simulate local processes (e.g., $\mathcal{N}_{AB}\left(\rho_{AB}\right)$), what resources are required?

This also applies to the output state. Moreover, the input and output states of a party can be detected locally without assistance from other parties. Viewing this from an input-output flow perspective, it inspires us to explore the local processes of each subsystem by tracing out the other subsystems from the global process. Thus, all local quantum processes $\mathcal{N}_{A_i\to B_i}\left(\rho_{A_i}\right)$ can be directly obtained by tracing out the other subsystems of a localizable quantum process:

$$\mathcal{N}_{A_i\to B_i}\left(\rho_{A_i}\right) := \mathrm{Tr}_{B_1\cdots B_{i-1}B_{i+1}\cdots B_n}\left[\mathcal{N}_{A_1\cdots A_n\to B_1\cdots B_n}\left(\rho_{A_1\cdots A_n}\right)\right]. \tag{13}$$

We can extend this approach to more general quantum processes, not limited to those expressible in the form of Eq. (12). The formula in Eq. (13) remains applicable for extracting information about local processes from an input-output flow perspective. We will demonstrate the equivalence between different representations of $\mathcal{N}_{A_1\cdots A_n\to B_1\cdots B_n}\left(\rho_{A_1\cdots A_n}\right)$ in the subsequent sections. Since the form in Eq. (13) resembles that of a reduced quantum state when we replace the entire process with a quantum state, we term the local processes as *reduced quantum process*. In the following, we shall provide more formal definitions of a reduced quantum process.

3.1 Definition of reduced quantum process

Let $[n] := \{1, \cdots, n\}$ and $\boldsymbol{s} \in [n]$ be a set of integers. Let $A_{[n]} \equiv A_1 \cdots A_n$ be a quantum register of n qubits and $A_{\boldsymbol{s}} = \otimes_{s\in\boldsymbol{s}} A_s$ be a subset qubits of the quantum register $A_{[n]}$. The concept of reduced quantum process is mathematically defined as follows.

Definition 1 (k-reduced quantum process)**.** *Let $\mathcal{N}_{A_{[n]}\to B_{[n]}}(\cdot) := U(\cdot)U^\dagger$ be an n-qubit quantum unitary channel given by the unitary $U_{A_{[n]}}$. Let $\boldsymbol{s} \subseteq [n]$ be a subset of qubit indices of size k. The k-reduced quantum process $\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}$ of $\mathcal{N}$, acting on the qubit indices $\boldsymbol{s}$, is defined as follows:*

$$\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}(\rho_{A_{\boldsymbol{s}}}) := \mathrm{Tr}_{B_{\bar{\boldsymbol{s}}}}\left[U\left(\rho_{A_{\boldsymbol{s}}}\otimes\frac{\mathbb{1}_{A_{\bar{\boldsymbol{s}}}}}{2^{n-k}}\right)U^\dagger\right], \tag{14}$$

where $\rho_{A_{\boldsymbol{s}}}$ is an arbitrary quantum state in $A_{\boldsymbol{s}}$ and $\overline{\boldsymbol{s}} := [n] \backslash \boldsymbol{s}$.

From the above definition, it is obvious that for a global process $\mathcal{N}_{A_{[n]} \to B_{[n]}}$, the information of a reduced process $\mathcal{N}_{A_{\boldsymbol{s}} \to B_{\boldsymbol{s}}}$ can be extracted from the global process by initializing all other subsystems in a maximally mixed state.

Since quantum process has different yet equivalent representations, reduced quantum process should also has different representations. In the following subsections, we give different representations of the reduced quantum process and prove their equivalence.

3.2 Choi representation of reduced quantum processes

We present the first mathematical description of a reduced quantum process using the Choi representation, as the Choi operator of a quantum process is unique. Consequently, there is only one form of such a reduced quantum process. We define a k-partite reduced quantum process as a k-reduced quantum process. In the remainder of this paper, we will adopt this notation.

Definition 2 (k-reduced Choi state)**.** *Let $\mathcal{N}_{A_{[n]} \to B_{[n]}}$ be an n-qubit quantum channel whose Choi state $J_{A'_{[n]} B_{[n]}}$ is defined in Eq. (3). Let $\boldsymbol{s} \subseteq [n]$ be a subset of qubit indices of size k. The k-reduced Choi state of J, induced by the qubit indices $\boldsymbol{s}$, is defined as follows:*

$$J_{A'_{\boldsymbol{s}} B_{\boldsymbol{s}}} := \mathrm{Tr}_{A'_{\overline{\boldsymbol{s}}} B_{\overline{\boldsymbol{s}}}} \left[J_{A'_{[n]} B_{[n]}} \right]. \tag{15}$$

We ascertain that the reduced Choi state $J_{A'_{\boldsymbol{s}} B_{\boldsymbol{s}}}$ in Eq. (15) represents an valid quantum process as it satisfies the CPTP conditions. The proof can be found in Appendix A.

Proposition 3. *The reduced Choi state $J_{A'_{\boldsymbol{s}} B_{\boldsymbol{s}}}$ defined in Eq. (15), satisfies $J_{A'_{\boldsymbol{s}} B_{\boldsymbol{s}}} \geq 0$ and $\mathrm{Tr}_{B_{\boldsymbol{s}}} J_{A'_{\boldsymbol{s}} B_{\boldsymbol{s}}} = \mathbb{1}_{A'_{\boldsymbol{s}}} / 2^k$, where $k = |\boldsymbol{s}|$.*

Interestingly, we show that the Choi state of the reduced quantum process given in Definition 1 is exactly the k-reduced Choi state given in Definition 2. The proof given in Appendix B.

Proposition 4 (Choi representation of reduced quantum process)**.** *Let $\mathcal{N}_{A_{[n]} \to B_{[n]}}(\cdot) := U(\cdot)U^\dagger$ be an n-qubit quantum unitary channel given by the unitary $U_{A_{[n]}}$. Let $\boldsymbol{s} \subseteq [n]$ be a subset of qubit indices of size k and let $\mathcal{N}_{A_{\boldsymbol{s}} \to B_{\boldsymbol{s}}}$ be the k-reduced quantum process of $\mathcal{N}$ as defined in Eq. (14). It holds that the Choi state of $\mathcal{N}_{A_{\boldsymbol{s}} \to B_{\boldsymbol{s}}}$ is exactly given by $J_{A'_{\boldsymbol{s}} B_{\boldsymbol{s}}}$ defined in Eq. (15).*

3.3 PTM representation of the reduced quantum process

As a quantum process can be represented in PTM form, the k-reduced quantum processes can also be defined in this form.

Definition 5 (k-reduced PTM)**.** *Let $\mathcal{N}_{A_{[n]} \to B_{[n]}}$ be an n-qubit quantum process whose PTM $R_{\mathcal{N}_{A_{[n]} \to B_{[n]}}}$ is defined in Eq. (4). Let $\boldsymbol{s} \subseteq [n]$ be a subset of qubit indices of size k. The k-reduced PTM of $R_{\mathcal{N}_{A_{[n]} \to B_{[n]}}}$, induced by the qubit indices $\boldsymbol{s}$, is defined as follows:*

$$\left(R_{\mathcal{N}_{A_{\boldsymbol{s}} \to B_{\boldsymbol{s}}}}\right)_{ij} = \left(R_{\mathcal{N}_{A_{[n]} \to B_{[n]}}}\right)_{\left(\sum_{b \in [k]} i_b \cdot 4^{n - \boldsymbol{s}_b}\right)\left(\sum_{b \in [k]} j_b \cdot 4^{n - \boldsymbol{s}_b}\right)} \tag{16}$$

Here, we represent the indices i, j of the PTM matrix as quaternary k-bit strings $i = i_1 \cdots i_k, j = j_1 \cdots j_k$, and $\boldsymbol{s}_b$ as the b-th element of $\boldsymbol{s}$.

Definition 5 reveals that the k-reduced PTM can be obtained from the PTM of the entire quantum process. Additionally, it should be noted that the k-reduced quantum processes defined in Definition 5 are consistent with Definition 1 and Definition 2. We prove that they can transform into each other, as shown in the following proposition. The proof can be found in Appendix C.

Proposition 6. *Let $\mathcal{N}_{A_{[n]}\to B_{[n]}}$ be an n-qubit quantum process and $\boldsymbol{s}\subseteq[n]$ be a subset of qubit indices of size k. The transformation between three definition of k-reduced quantum process $\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}$ reads*

$$\left(R_{\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}}\right)_{ij}=\frac{1}{2^k}\operatorname{Tr}\left[P_i\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}(P_j)\right]=\operatorname{Tr}\left[J_{A'_{\boldsymbol{s}}B_{\boldsymbol{s}}}P_j^T\otimes P_i\right], \tag{17}$$

where $\mathcal{N}_{A_{\boldsymbol{s}}\to B_{\boldsymbol{s}}}(\rho_{A_{\boldsymbol{s}}})$ is defined in Eq.(14), $J_{A'_{\boldsymbol{s}}B_{\boldsymbol{s}}}$ is the k-reduced Choi state defined in Eq. (15), $k=|\boldsymbol{s}|$.

To validate our theory and efficiently obtain all k-reduced PTMs for an n-qubit system, we introduce several methods in the following section, accompanied by a detailed protocol and complexity analysis.

4 Quantum process overlapping tomography: Theory

This section devotes to characterizing *all* k-reduced quantum processes of a given n-qubit quantum unitary process. To achieve the target, we introduce a general framework called Quantum Process Overlapping Tomography, comprising three tomographic methods each having its own feature.

4.1 Method 1

We have shown in Section 3.2 that the reduced quantum process admit an elegant Choi representation that is directly related to the Choi state of the global quantum process. We leverage this relationship to transform reduced quantum process tomography into reduced quantum state tomography. By employing overlapping tomography techniques [28], we significantly reduce the total number of required measurement settings.

Method 1 works as follows. A k-reduced quantum process can be expressed by a k-reduced Choi state as shown in Eq. (15). Characterizing all k-reduced quantum processes is equivalent to obtaining all k-reduced Choi states. It's worth noting that the Choi state of an n-qubit quantum process is a $2n$-qubit quantum state, making a k-reduced Choi state a $2k$-qubit reduced quantum state. Therefore, we need to measure a total number of $\binom{2n}{2k}$ reduced density matrices. Explicitly, we first construct a (n',k') family of perfect hash functions, which will be utilized to group the qubits. There is extensive literature focusing on explicitly generating such (n',k') families of perfect hash functions [32, 33, 37–50]. Here, we take $n'=2n, k'=2(k=1)$ as an example and illustrate with $(2n,2)$ family of perfect hash functions.

Demonstration for $k'=2$. The $(2n,2)$ family of perfect hash functions comprises $q=\log_2(2n)$ functions, denoted as $f_1,...,f_q$. Each function maps $[n]\to\{0,1\}$ and is defined as

$$f_i(j)=i\text{th digit in the binary expansion of }(j-1). \tag{18}$$

Here, the number of the qubit $(j-1)$ is implicitly represented by a q-bit string. In general, every qubit is assigned a label of 0 or 1 by each perfect hash function. After this assignment, we use the labels of qubits to generate measurement bases, following the procedure described in Section IV of [28]. An illustration of this method is shown in Figure 2.

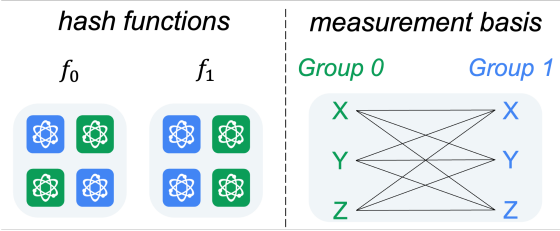


Figure 2: Illustration for Method 1. As an example with $k' = 2$, after the assignment by the perfect hash function, we color the qubit labeled with 0 in green and those with 1 in blue. All green(blue) qubits are measured in the same basis. We then iterate the green-blue basis over all $3^2 = 9$ measurement settings: $\{XX, XY, XZ, YX, YY, YZ, ZX, ZY, ZZ\}$.

Following [28], we know that tomographying $\binom{2n}{2k}$ reduced density matrices requires measurement settings of the size

$$3 + \left(3^{2k} - 3\right) \cdot \log_2 (2n) = 3 + (9^k - 3) \cdot \log_2 (2n) . \tag{19}$$

In contrast, independently measuring all these density matrices would necessitate approximately measurement settings of the size

$$\binom{2n}{2k} \cdot 3^{2k} \sim \left(36 \cdot n^2\right)^k . \tag{20}$$

Thus, overlapping tomography method offers significant advantages.

The primary drawback of **Method 1** is the need to prepare a $2n$-qubit maximally entangled state, leading to increased resource consumption. Furthermore, operating a larger system with higher entanglement may introduce more errors, potentially compromising the performance of overlapping tomography. In the following, we introduce two more methods that are ancilla free.

4.2 Method 2

Here we introduce **Method 2** that directly conducts quantum process tomography for k-reduced quantum processes by exploring parallelism, without referring to quantum state tomography as in **Method 1**. For localizable quantum processes, this method can efficiently harness the benefits of parallelism.

Method 2 works as follows. First, we partition the n qubits into n/k groups, each containing k qubits. Then, standard quantum process tomography is performed for each group, necessitating 12^k measurement settings. Note that all n/k groups employ the same measurement settings simultaneously. Explicitly, the 12 measurement settings are:

$$\underbrace{\{|0\rangle, |1\rangle, |+\rangle, |r\rangle\}}_{\textbf{preparation basis}} \otimes \underbrace{\{X, Y, Z\}}_{\textbf{measurement basis}}. \tag{21}$$

The grouping procedure is visualized in Figure 3. It is clear that such parallelism can reduce the total number of measurements by a factor of n/k.

However, if there are interactions between subsystems, the local processes cannot be directly obtained unless all other interacted subsystems are initialized in a maximally mixed state. Therefore, this parallelism is applicable only to localizable quantum processes, with the required measurement settings approximately being $\binom{n}{k} \cdot 12^k/(n/k) \sim k \cdot n^{k-1} \cdot 12^k$.

To characterize a k-reduced quantum process of a general quantum process, suppose there are m qubits that interact with this k-reduced quantum process. To simulate a maximally mixed

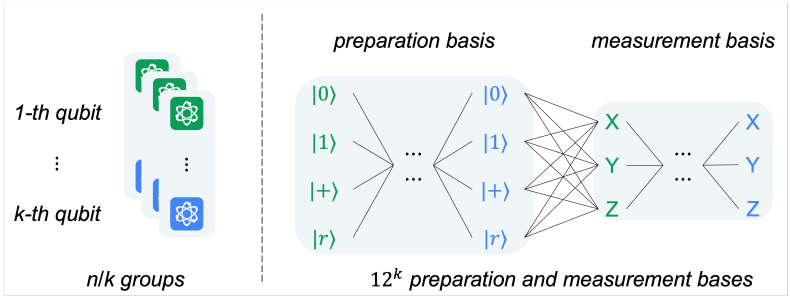


Figure 3: Visualization of Method 2: n qubits are divided into n/k groups, with each group consisting of k qubits that can represent a k-reduced quantum process. All k-reduced quantum processes undergo the same quantum process tomography using the standard 12^k measurement settings.

state of m qubits, a combination of 2^m orthonormal bases needs to be created. The complexity of characterizing such a k-reduced quantum process is $2^m \cdot 12^k$, which remains consistent for other k-reduced quantum processes. Therefore, the complexity for characterizing all k-reduced quantum processes is

$$\sum_{i=1}^{\binom{n}{k}} \cdot 2^{m_i} \cdot 12^k \sim \sum_{i=1}^{n^k} \cdot 2^{m_i} \cdot 12^k, \tag{22}$$

where m_i is the number of qubits that interact with the i-th k-reduced quantum process, and there are in total $\binom{n}{k}$ k-reduced quantum processes, so i ranges from 1 to $\binom{n}{k}$.

4.3 Method 3

Method 2 behaves bad when the number of qubits interacting with the target reduced quantum process becomes large. To resolve this problem, we propose **Method 3** that leverages interactions between subsystems to achieve efficient characterization.

Suppose an n-qubit quantum process has i local interactions, each with dimensions l_i satisfying $\sum_i l_i \leq n$. For example, in a 7-qubits process, the whole unitary can be written as $U = U_0 \otimes U_3 \otimes U_{14} \otimes U_{256}$, where U_{14} and U_{256} (the subscripts indicate qubits they act on) are two unitaries corresponding to local interactions with dimensions of 2 and 3, respectively. As discussed earlier, we must construct a valid statistical mixture of maximally mixed states for these subsystems. To ensure that any arbitrary k-reduced quantum process contains a mixture of maximally mixed states, we prepare all local subsystems involved in interactions in maximally mixed states and then combine them. **Method 3** works as follows.

Step 1: Generate overlapping bases using the perfect hash functions, following the procedure outlined in Section IV of [28]. However, we substitute the $\{X, Y, Z\}$ basis with 12 QPT measurement settings. As a result, there are a total of $12^k \cdot \log(n)$ bases, which cover 12 measurement settings of any single-qubit process tomography. This procedure is sufficient to characterize all k-reduced quantum processes of localizable quantum processes, as they are all composed of single-qubit processes. This procedure is depicted in Figure 4.

Step 2: Iterate over each qubit involved in interactions with other qubits and append a complementary basis based on its current basis to all overlapping bases generated in **Step 1**. For example, during the iteration for qubit 0, we will add $\{|1\rangle, X\}$ if the current preparation and measurement basis of qubit 0 is $\{|0\rangle, X\}$. Each iteration doubles the length of the overlapping basis, so after all iterations, the size of the overlapping basis will be multiplied by $2^{\sum_i l_i}$. In contrast to Method 2, which involves summing over all k-reduced quantum processes and includes numerous repeated

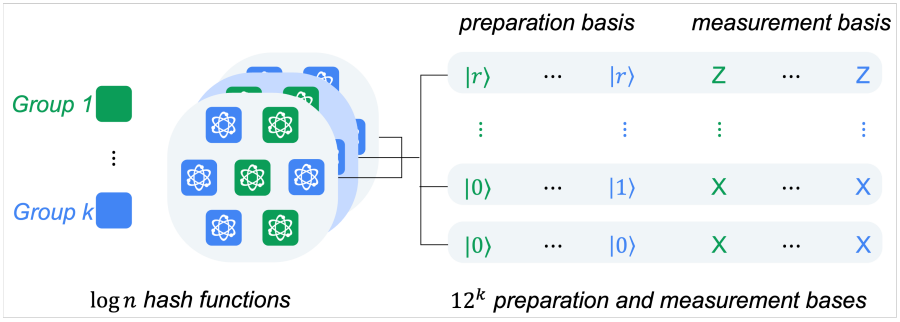


Figure 4: Illustration for Method 3 **Step 1**. We use perfect hash functions to group n qubits into k groups, where every qubit in one group uses the same single-qubit measurement setting. For each perfect hash function, we iterate the measurement setting of k groups over all 12^k measurement settings.

sums, in this method, $\sum_i l_i$ represents the sum over all local interactions (noting that $\sum_i l_i \leq n$). It's crucial to emphasize that this factor is determined by the local interactions of a quantum process, independent of the size of the entire process. Therefore, for a global process in a large system with specific local interactions, this is a favorable scale. The total number of measurement settings required is $2^{\sum_i l_i} \cdot 12^k \cdot \log(n)$.

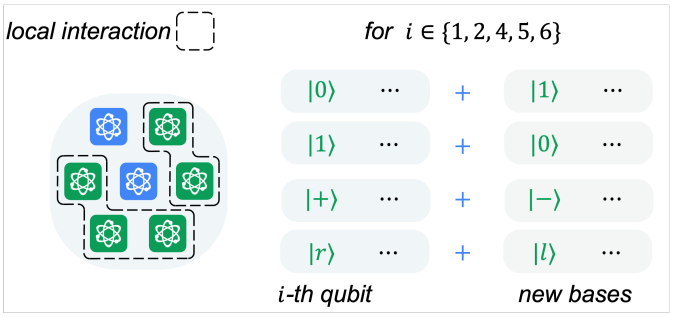


Figure 5: Illustration for Method 3 **Step 2**. Taking the previous 7-qubit example where $U = U_0 \otimes U_3 \otimes U_{14} \otimes U_{256}$, we identify two local interactions: U_{14} with dimension 2 and U_{256} with dimension 3. In this case, we should iterate over qubits $\{1, 2, 4, 5, 6\}$, and throughout all overlapping bases, we append a complement basis with respect to qubit in interation.

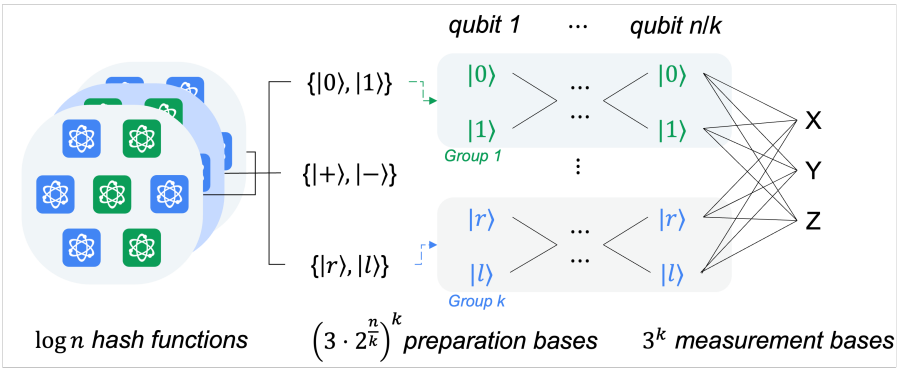


Figure 6: Illustration for Method 3 with full interactions. For qubits in one group, we prepare 3 maximally mixed states by mixing the qubits in the $\{|0\rangle, |1\rangle\}, \{|+\rangle, |-\rangle\}, \{|r\rangle, |l\rangle\}$ bases. For each maximally mixed state, we measure all qubits separately in the X, Y, Z bases. Therefore, there are in total $\left(2^{n/k} \cdot 3 \cdot 3\right)^k = 2^n \cdot 9^k$ measurement bases for each perfect hash function.

Additionally, for quantum processes with full interactions, where every qubit interacts with each other either directly or indirectly, a more efficient protocol can be designed. This protocol involves mixing the qubits within the same group, as determined by perfect hash functions, into a maximally mixed state. An illustration of this approach is depicted in Figure 6. The total number of measurement bases will be $2^n \cdot 9^k \cdot \log(n)$. In contrast, the method described in Step 2 yields $2^n \cdot 12^k \cdot \log(n)$.

4.4 Comparison of methods

	Sample Complexity	Ancilla Required	Overlapping Measurement
Method 1	$3 + (9^k - 3) \cdot \log(2n)$	✓	✓
Method 2	$\sum_{i=1}^{n^k} \cdot 2^{m_i} \cdot 12^k$	✗	✗
Method 3	$2^{\sum_i l_i} \cdot 12^k \cdot \log(n)$	✗	✓

Table 1: Comparison between three quantum process overlapping tomography methods. We primarily concern the sample complexity, whether it requires ancilla qubits, and whether it uses overlapping measurement. Here, n is the number of qubits in the global system, k is the number of qubits of the reduced system, m_i is the number of qubits that interact with the i-th k-reduced quantum process, and l_i is the dimension of the i-th local interaction.

We compare the three methods mentioned above and summarize the key results in Table 1. In terms of sample complexity, **Method 1** offers the best result, which is $3 + (9^k - 3) \cdot \log(2n)$, followed by **Method 3**, and then **Method 2**. The reason is that both **Method 1** and **Method 3** construct a set of overlapping measurements via perfect hash functions, which maximize the parallelism. Additionally, the summation of i in **Method 2** includes many repeated counts, where **Method 3** demonstrates significant advantages. On the other hand, both **Method 2** and **Method 3** require no ancillary qubits, whereas **Method 1** requires an ancilla system of n qubits to construct a maximally entangled state. Both ancillary qubits and maximally entangled states are challenging or even impossible for large quantum system. Therefore, we will only consider **Method 2** and **Method 3** in the following experiment.

5 Quantum process overlapping tomography: Experiment

This section devotes to *experimentally* validating the quantum process overlapping tomography framework proposed in Section 4. Specifically, we apply the methods to reconstruct *all* 2-reduced processes of the quantum process preparing a 4-qubit GHZ state. The corresponding quantum preparation process is visualized in Figure 7. The total number of 2-reduced processes is $\binom{4}{2} = 6$.

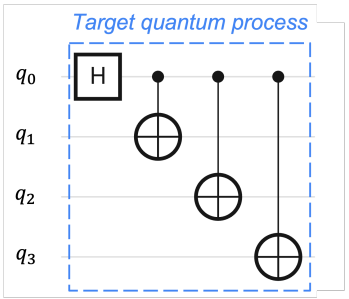


Figure 7: The quantum process for preparing a 4-qubit GHZ state.

We use (q_i, q_j) to denote the 2-reduced process acting on the i-th and j-th qubits. Intuitively, owing to the symmetry inherent in the quantum circuit, one might speculate that there exist two distinct types of the 2-reduced quantum processes: one type comprising the reduced processes $(q_0, q_1), (q_0, q_2)$ and (q_0, q_3), and the other type including $(q_1, q_2), (q_1, q_3)$ and (q_2, q_3). Our objective is to ascertain the validity of this conjecture. Initially, we compute the ideal PTM of the entire process using the IBM Qiskit package [51]. Subsequently, employing Eq. (16), we calculate all PTMs of the 2-reduced processes. The results corroborate our hypothesis: there exist two types of 2-reduced quantum processes, denoted as type 1 and type 2, respectively. Their PTMs are depicted in Figure 8(a). Type 1 quantum process comprises 2-reduced quantum processes: $(q_0, q_1), (q_0, q_2), (q_0, q_3)$, where the qubits are directly interacted with CNOT gates. Type 2 quantum process encompasses $(q_1, q_2), (q_1, q_3), (q_2, q_3)$, where the qubits are indirectly interacted.

Then, we implement **Method 2** and **Method 3** on the IBM-Brisbane quantum device and certify their performances. We present two PTMs for each method, as depicted in Figure 8(b) and (d), respectively. From the results, the PTMs obtained by **Method 2** and **Method 3** closely approximate the ideal PTM. To quantitatively assess the similarity, we calculate the process fidelity for each 2-reduced quantum process obtained by different methods. The process fidelity is defined as follows:

$$F(\mathcal{N}, \mathcal{E}) := F_s(J_\mathcal{N}, J_\mathcal{E}), \tag{23}$$

where $\mathcal{N}$, $\mathcal{E}$ are two quantum processes, $J_\mathcal{N}, J_\mathcal{E}$ are the Choi states of $\mathcal{N}, \mathcal{E}$, and F_s is the state fidelity defined as $F_s(\rho, \sigma) := (\mathrm{Tr}[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}])^2$. The results of these fidelities can be found in Figure 8(d). For different 2-reduced processes, Method 3 exhibits higher fidelities, possibly due to differences in implementation times. Additionally, we observe that the reduced processes with the highest fidelity for each type are the same: (q_0, q_3) and (q_2, q_3). This suggests that q_3 may be more stable than the other qubits.

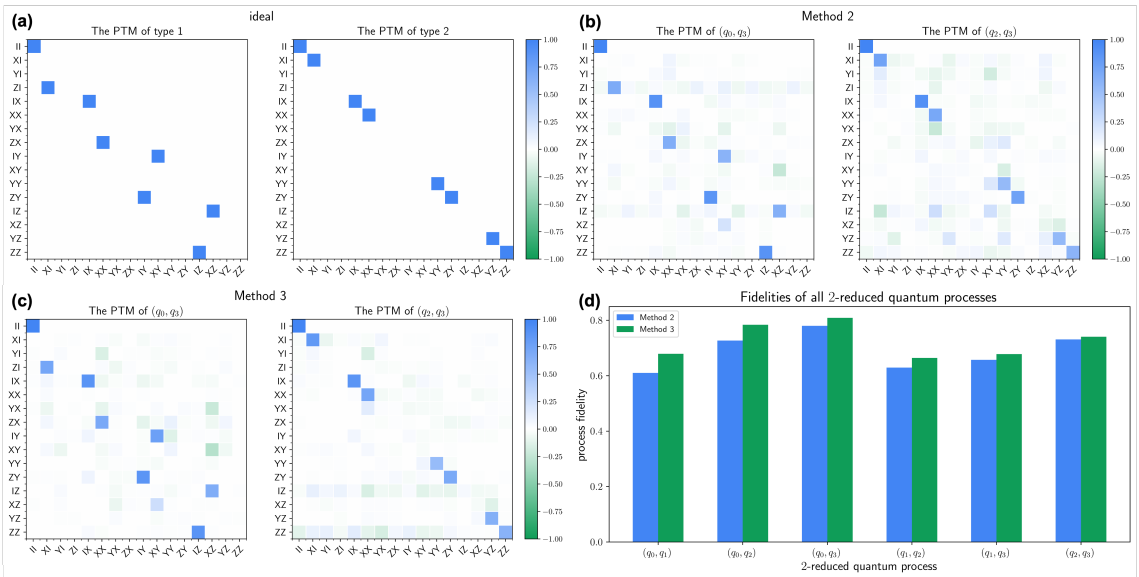


Figure 8: Quantum process overlapping tomography of 4-qubit GHZ state preparation process on the *IBM-brisbane* quantum device. (q_i, q_j) denotes the 2-reduced quantum process acting on i-th and j-th qubit. **(a)** Two types of ideal 2-reduced quantum processes. Type 1 contains (q_0, q_1), (q_0, q_2) and (q_0, q_3). On the other hand, type 2 contains (q_1, q_2), (q_1, q_3) and (q_2, q_3). **(b)** The 2-reduced quantum processes (q_0, q_3) and (q_2, q_3) obtained by **Method 2**. **(c)** The 2-reduced quantum processes (q_0, q_3) and (q_2, q_3) obtained by **Method 3**. **(d)** The reduced process fidelities of **Method 2** and **Method 3**.

6 Conclusions

We systematically investigated the concept of *reduced quantum processes* of a global quantum process, generalizing the concept of reduced quantum states to the quantum process domain. First, we provided a rigorous mathematical description of reduced quantum processes and derived equivalent characterizations in different quantum process presentations. Then, we introduced a general framework called quantum process overlapping tomography to fully characterize the reduced quantum processes. This framework makes use of perfect hash functions and comprises three tomographic methods each having its own feature. At last, we experimentally validated our quantum process overlapping tomography framework on IBM quantum devices. The obtained experimental tomographic results align perfectly with the theory predictions.

Acknowledgements

Y.H. and C.-C.Z. contributed equally to this work. Part of this work was done when K.W. was at the Institute for Quantum Computing, Baidu Research. This work was supported by the National Key Research and Development Program of China (Grant Nos. 2019YFA0308700 and 2022YFF0712800), the Jiangsu Key R&D Program Project (Grant No. BE2023011-2), the Fundamental Research Funds for the Central Universities (Grant No. 2242022k60001), and the National Natural Science Foundation of China (Grant Nos. 61960206005 and 61871111). We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

References

[1] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.

[2] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.

[3] Google AI Quantum, Collaborators*†, Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B Buckley, et al. Hartree-fock on a superconducting qubit quantum computer. *Science*, 369(6507):1084–1089, 2020.

[4] Yulin Wu, Wan-Su Bao, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, et al. Strong quantum computational advantage using a superconducting quantum processor. *Physical Review Letters*, 127(18):180501, 2021.

[5] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.

[6] Christian Kokail, Christine Maier, Rick van Bijnen, Tiff Brydges, Manoj K Joshi, Petar Jurcevic, Christine A Muschik, Pietro Silvi, Rainer Blatt, Christian F Roos, et al. Self-verifying variational quantum simulation of lattice models. *Nature*, 569(7756):355–360, 2019.

[7] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.

[8] Matthew P Harrigan, Kevin J Sung, Matthew Neeley, Kevin J Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C Bardin, Rami Barends, Sergio Boixo, et al. Quantum ap-

proximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, 2021.

[9] M Riebe, K Kim, P Schindler, Thomas Monz, PO Schmidt, TK Körber, W Hänsel, Hartmut Häffner, Christian F Roos, and Rainer Blatt. Process tomography of ion trap quantum gates. *Physical Review Letters*, 97(22):220407, 2006.

[10] Yaakov S Weinstein, Timothy F Havel, Joseph Emerson, Nicolas Boulant, Marcos Saraceno, Seth Lloyd, and David G Cory. Quantum process tomography of the quantum fourier transform. *The Journal of Chemical Physics*, 121(13):6117–6133, 2004.

[11] Jeremy L O'Brien, Geoff J Pryde, Alexei Gilchrist, Daniel FV James, Nathan K Langford, Timothy C Ralph, and Andrew G White. Quantum process tomography of a controlled-not gate. *Physical Review Letters*, 93(8):080502, 2004.

[12] Radoslaw C Bialczak, Markus Ansmann, Max Hofheinz, Erik Lucero, Matthew Neeley, Aaron D O'Connell, Daniel Sank, Haohua Wang, James Wenner, Matthias Steffen, et al. Quantum process tomography of a universal entangling gate implemented with josephson phase qubits. *Nature Physics*, 6(6):409–413, 2010.

[13] M Howard, J Twamley, C Wittmann, T Gaebel, F Jelezko, and J Wrachtrup. Quantum process tomography and Linblad estimation of a solid-state qubit. *New Journal of Physics*, 8(3):33, 2006.

[14] Trystan Surawy-Stepney, Jonas Kahn, Richard Kueng, and Madalin Guta. Projected least-squares quantum process tomography. *Quantum*, 6:844, 2022.

[15] Seth T Merkel, Jay M Gambetta, John A Smolin, Stefano Poletto, Antonio D Córcoles, Blake R Johnson, Colm A Ryan, and Matthias Steffen. Self-consistent quantum process tomography. *Physical Review A*, 87(6):062119, 2013.

[16] Masoud Mohseni, Ali T Rezakhani, and Daniel A Lidar. Quantum-process tomography: Resource analysis of different strategies. *Physical Review A*, 77(3):032322, 2008.

[17] Giacomo Torlai, Christopher J Wood, Atithi Acharya, Giuseppe Carleo, Juan Carrasquilla, and Leandro Aolita. Quantum process tomography with unsupervised learning and tensor networks. *Nature Communications*, 14(1):2858, 2023.

[18] J Ignacio Cirac, AK Ekert, Susana F Huelga, and Chiara Macchiavello. Distributed quantum computation over noisy channels. *Physical Review A*, 59(6):4249, 1999.

[19] David P DiVincenzo. The physical implementation of quantum computation. *Fortschritte der Physik: Progress of Physics*, 48(9-11):771–783, 2000.

[20] H Jeff Kimble. The quantum internet. *Nature*, 453(7198):1023–1030, 2008.

[21] Robin Harper, Steven T Flammia, and Joel J Wallman. Efficient learning of quantum noise. *Nature Physics*, 16(12):1184–1188, 2020.

[22] Steven T Flammia and Joel J Wallman. Efficient estimation of pauli channels. *ACM Transactions on Quantum Computing*, 1(1):1–32, 2020.

[23] Héctor Bombin, Ruben S Andrist, Masayuki Ohzeki, Helmut G Katzgraber, and Miguel A Martin-Delgado. Strong resilience of topological codes to depolarization. *Physical Review X*, 2(2):021004, 2012.

[24] Naomi H Nickerson and Benjamin J Brown. Analysing correlated noise on the surface code using adaptive decoding algorithms. *Quantum*, 3:131, 2019.

[25] Andrew S Darmawan and David Poulin. Tensor-network simulations of the surface code under realistic noise. *Physical Review Letters*, 119(4):040502, 2017.

[26] Nishad Maskara, Aleksander Kubica, and Tomas Jochym-O'Connor. Advantages of versatile neural-network decoding for topological codes. *Physical Review A*, 99(5):052351, 2019.

[27] David K Tuckett, Stephen D Bartlett, and Steven T Flammia. Ultrahigh error threshold for surface codes with biased noise. *Physical Review Letters*, 120(5):050505, 2018.

[28] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Physical Review Letters*, 124(10):100401, 2020.

[29] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E O'Brien. Nearly optimal measurement scheduling for partial tomography of quantum states. *Physical Review X*, 10(3):031064, 2020.

[30] Filip B Maciejewski, Flavio Baccari, Zoltán Zimborás, and Michał Oszmaniec. Modeling and mitigation of cross-talk effects in readout noise with applications to the quantum approximate optimization algorithm. *Quantum*, 5:464, 2021.

[31] Bruna GM Araújo, Márcio M Taddei, Daniel Cavalcanti, and Antonio Acín. Local quantum overlapping tomography. *Physical Review A*, 106(6):062441, 2022.

[32] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E O'Brien. Nearly optimal measurement scheduling for partial tomography of quantum states. *Physical Review X*, 10(3):031064, 2020.

[33] Filip B Maciejewski, Flavio Baccari, Zoltán Zimborás, and Michał Oszmaniec. Modeling and mitigation of cross-talk effects in readout noise with applications to the quantum approximate optimization algorithm. *Quantum*, 5:464, 2021.

[34] Mark M Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.

[35] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[36] Daniel Greenbaum. Introduction to quantum gate set tomography. *arXiv preprint arXiv:1509.02921*, 2015.

[37] Kurt Mehlhorn. *Data structures and algorithms 1: Sorting and searching*, volume 1. Springer Science & Business Media, 2013.

[38] Michael L Fredman, János Komlós, and Endre Szemerédi. Storing a sparse table with 0 (1) worst case access time. *Journal of the ACM*, 31(3):538–544, 1984.

[39] Michael L Fredman and János Komlós. On the size of separating systems and families of perfect hash functions. *SIAM Journal on Algebraic Discrete Methods*, 5(1):61–68, 1984.

[40] Noga Alon. Explicit construction of exponential sized families of k-independent sets. *Discrete Mathematics*, 58(2):191–193, 1986.

[41] János Korner and Katalin Marton. New bounds for perfect hashing via information theory. *European Journal of Combinatorics*, 9(6):523–530, 1988.

[42] Jeanette P Schmidt and Alan Siegel. The spatial complexity of oblivious k-probe hash functions. *SIAM Journal on Computing*, 19(5):775–786, 1990.

[43] Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *Journal of the ACM*, 42(4):844–856, 1995.

[44] Moni Naor, Leonard J Schulman, and Aravind Srinivasan. Splitters and near-optimal derandomization. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 182–191. IEEE, 1995.

[45] M Atici, SS Magliveras, DR Stinson, and W-D Wei. Some recursive constructions for perfect hash families. *Journal of Combinatorial Designs*, 4(5):353–363, 1996.

[46] Zbigniew J Czech, George Havas, and Bohdan S Majewski. Perfect hashing. *Theoretical Computer Science*, 182(1-2):1–143, 1997.

[47] Simon R Blackburn and Peter R Wild. Optimal linear perfect hash families. *Journal of Combinatorial Theory, Series A*, 83(2):233–250, 1998.

[48] Simon R Blackburn. Perfect hash families: probabilistic methods and explicit constructions. *Journal of Combinatorial Theory, Series A*, 92(1):54–60, 2000.

[49] Douglas R Stinson, Ruizhong Wei, and L Zhu. New constructions for perfect hash families and related structures using combinatorial designs and codes. *Journal of Combinatorial Designs*, 8(3):189–200, 2000.

[50] Noga Alon and Shai Gutner. Balanced families of perfect hash functions and their applications. *ACM Transactions on Algorithms*, 6(3):1–12, 2010.

[51] IBM Quantum Platform https://quantum-computing.ibm.com/.

A Proof of Proposition 3

Proof of Proposition 3. Since the n-qubit Choi state $J_{A'_1\cdots A'_n B_1\cdots B_n}$ is a positive semidefinite operator, and partial trace preserve positivity, it follows that $J_{A'_1\cdots A'_k B_1\cdots B_k}$ is also a positive semidefinite operator. We use the single-qubit reduced Choi state as an example. For generality, we consider the n-qubit process as an arbitrary quantum process in the form of the Kraus representation, as shown in Eq. (2). We explicitly expand the single-qubit reduced Choi state in the following form:

$$
\begin{aligned}
J_{A'_1B_1} &= \mathrm{Tr}_{A'_2\cdots A'_n B_2\cdots B_n}\, J_{A'_1\cdots A'_n B_1\cdots B_n} \\
&= \mathrm{Tr}_{A'_2\cdots A'_n B_2\cdots B_n}\left[\frac{1}{2^n}\sum_{i,j}|i\rangle\langle j|_{A'}\otimes\sum_l K_l|i\rangle\langle j|_A K_l^\dagger\right] \\
&= \frac{1}{2^n}\sum_{i,j}\mathrm{Tr}_{A'_2\cdots A'_n}\left[|i\rangle\langle j|_{A'}\right]\otimes\mathrm{Tr}_{B_2\cdots B_n}\left[\sum_l K_l|i\rangle\langle j|_A K_l^\dagger\right] \\
&= \frac{1}{2^n}\sum_{i,j}\mathrm{Tr}_{A'_2\cdots A'_n}\left[(|i_1\rangle\otimes|i_{n-1}\rangle)(\langle j_1|\otimes\langle j_{n-1}|)\right]\otimes\mathrm{Tr}_{B_2\cdots B_n}\left[\sum_l K_l(|i_1\rangle\otimes|i_{n-1}\rangle)(\langle j_1|\otimes\langle j_{n-1}|)K_l^\dagger\right] \\
&= \frac{1}{2^n}\sum_{i,j}|i_1\rangle\langle j_1|\langle i_{n-1}|j_{n-1}\rangle\otimes\mathrm{Tr}_{B_2\cdots B_n}\left[\sum_l K_l(|i_1\rangle\otimes|i_{n-1}\rangle)(\langle j_1|\otimes\langle j_{n-1}|)K_l^\dagger\right] \\
&= \frac{1}{2^n}\sum_{i,j}|i_1\rangle\langle j_1|\langle i_{n-1}|i_{n-1}\rangle\otimes\mathrm{Tr}_{B_2\cdots B_n}\left[\sum_l K_l(|i_1\rangle\otimes|i_{n-1}\rangle)(\langle j_1|\otimes\langle i_{n-1}|)K_l^\dagger\right] \\
&= \frac{1}{2^n}\sum_{i_1,j_1}\sum_{i_{n-1}}|i_1\rangle\langle j_1|\langle i_{n-1}|i_{n-1}\rangle\otimes\mathrm{Tr}_{B_2\cdots B_n}\left[\sum_l K_l(|i_1\rangle\langle j_1|\otimes|i_{n-1}\rangle\langle i_{n-1}|)K_l^\dagger\right] \\
&= \frac{1}{2^n}\sum_{i_1,j_1}\sum_{i_{n-1}}|i_1\rangle\langle j_1|\langle i_{n-1}|i_{n-1}\rangle\otimes\mathrm{Tr}_{B_2\cdots B_n}\left[\sum_l K_l(|i_1\rangle\langle j_1|\otimes\mathbb{1}A_2\cdots A_n)K_l^\dagger\right] \\
&= \sum_{i_1,j_1}|i_1\rangle\langle j_1|_{A'}\otimes\mathrm{Tr}_{B_2\cdots B_n}\left[\sum_l K_l(\frac{1}{2}|i_1\rangle\langle j_1|_A\otimes\frac{1}{2^{n-1}}\mathbb{1}A_2\cdots A_n)K_l^\dagger\right], \qquad (24)
\end{aligned}
$$

Next, we prove that $\mathrm{Tr}_{B_1} J_{A'_1B_1} = \frac{1}{2}\mathbb{1}_{A'_1}$:

$$
\begin{aligned}
&\mathrm{Tr}_{B_1}\, J_{A_1'B_1} \\
&= \sum_{i_1,j_1} |i_1\rangle\langle j_1|_{A'} \otimes \mathrm{Tr}_{B_1}\, \mathrm{Tr}_{B_2\cdots B_n} \left[\sum_l K_l(\frac{1}{2}|i_1\rangle\langle j_1|_A \otimes \frac{1}{2^{n-1}} \mathbb{1} A_2 \cdots A_n) K_l^\dagger \right] \\
&= \sum_{i_1,j_1} |i_1\rangle\langle j_1|_{A'} \otimes \mathrm{Tr} \left[\sum_l K_l(\frac{1}{2}|i_1\rangle\langle j_1|_A \otimes \frac{1}{2^{n-1}} \mathbb{1} A_2 \cdots A_n) K_l^\dagger \right] \\
&= \sum_{i_1,j_1} |i_1\rangle\langle j_1|_{A'} \otimes \mathrm{Tr} \left[\sum_l K_l(\frac{1}{2}|i_1\rangle\langle j_1|_A \otimes \frac{1}{2^{n-1}} \sum_{i_{n-1}} |i_{n-1}\rangle\langle i_{n-1}|_A) K_l^\dagger \right] \\
&= \sum_{i_1,j_1} |i_1\rangle\langle j_1|_{A'} \otimes \sum_l \sum_{i_{n-1}} \frac{1}{2^n} (\langle j_1| \otimes \langle i_{n-1}|) K_l^\dagger K_l (|i_1\rangle \otimes |i_{n-1}\rangle) \\
&= \sum_{i_1,j_1} |i_1\rangle\langle j_1|_{A'} \sum_{i_{n-1}} \frac{1}{2^n} \langle j_1|i_1\rangle \\
&= \frac{1}{2} \sum_{i_1} |i_1\rangle\langle i_1|_{A'} \\
&= \frac{1}{2} \mathbb{1}_{A_1'}
\end{aligned} \tag{25}
$$

Thus, the reduced state $J_{A_1'B_1}$ satisfies $J_{A_1'B_1} \geq 0$ and $\mathrm{Tr}_{B_1}\, J_{A_1'B_1} = \frac{1}{2}\mathbb{1}_{A_1'}$. The proof for the k-reduced Choi state $J_{A_1'\cdots A_k'B_1\cdots B_k}$ follows a similar logic. □

B Proof of Proposition 4

Proof of Proposition 4. We will use the single-qubit case as an example here, as the proof follows the same procedure when considering multiple qubits. The Choi state of the reduced quantum process $\mathcal{N}_{A_1\to B_1}$, as defined by Eq.(3), is

$$
\begin{aligned}
&\left(\mathrm{id}_{A_1'} \otimes \mathcal{N}_{A_1\to B_1}\right) \left(\Gamma_{A_1A_1'}\right) \\
&= \left(\mathrm{id}_{A_1'} \otimes \mathcal{N}_{A_1\to B_1}\right) \frac{1}{2} \sum_{i_1j_1} |i_1i_1\rangle\langle j_1j_1| \\
&= \sum_{i_1,j_1} |i_1\rangle\langle j_1|_{A'} \otimes \mathrm{Tr}_{B_2\cdots B_n} \left[\sum_l K_l(\frac{1}{2}|i_1\rangle\langle j_1|_A \otimes \frac{1}{2^{n-1}} \mathbb{1}_{A_2\cdots A_n}) K_l^\dagger \right]
\end{aligned} \tag{26}
$$

which is the same as Eq. (24). □

C Proof of Proposition 6

Proving Proposition 6 requires the following Lemma.

Lemma 7. *Let P_{AB} and Q_A be two linear operators. It holds that*

$$
\mathrm{Tr}_B[P_{AB}]Q_A = \mathrm{Tr}_B[P_{AB}(Q_A \otimes \mathbb{1}_B)]. \tag{27}
$$

Proof. Let $\{|i\rangle_A\}$ and $|k\rangle_B\}$ be orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, and let $d_A \equiv \dim(\mathcal{H}_A), d_B \equiv \dim(\mathcal{H}_B)$. Thus, $P_{AB} \in \mathcal{L}(\mathcal{H}_{AB})$ and $Q_A \in \mathcal{L}(\mathcal{H}_A)$. We can expand Q_A in terms

of the orthonormal basis $\{|i\rangle_A\}$:

$$Q_A = \sum_{ij=0}^{d_A-1} \alpha_{ij} |i\rangle\langle j|_A \tag{28}$$

and P_{AB} in terms of orthonormal basis $\{|i\rangle_A\} \otimes \{|k\rangle_B\}$:

$$P_{AB} = \sum_{ij=0}^{d_A-1} \sum_{kl=0}^{d_B-1} \beta_{ijkl} \left(|i\rangle_A \otimes |k\rangle_B\right) \left(\langle j|_A \otimes \langle l|_B\right). \tag{29}$$

Now we can expand

$$\begin{aligned}
&\mathrm{Tr}_B[P_{AB}]Q_A \\
&= \sum_{k=0}^{d_B-1} \langle k|_B \sum_{ij=0}^{d_A-1} \sum_{kl=0}^{d_B-1} \beta_{ijkl} \left(|i\rangle_A \otimes |k\rangle_B\right) \left(\langle j|_A \otimes \langle l|_B\right) \sum_{ij=0}^{d_A-1} \alpha_{ij} |i\rangle\langle j|_A |k\rangle_B \\
&= \sum_{ij=0}^{d_A-1} \sum_{kl=0}^{d_B-1} \alpha_{ij}\beta_{ijkl} \langle k|k\rangle_B \left(|i\rangle_A \otimes \mathbb{1}_B\right) \left(\langle j|_A \otimes \langle l|_B\right) |i\rangle\langle j|_A |k\rangle_B \\
&= \sum_{ij=0}^{d_A-1} \sum_{kl=0}^{d_B-1} \alpha_{ij}\beta_{ijkl} \left(|i\rangle_A \otimes \mathbb{1}_B\right) \left(\langle j|_A \otimes \mathbb{1}_B\right) |i\rangle\langle j|_A \langle l|k\rangle_B \\
&= \sum_{i=0}^{d_A-1} \sum_{k=0}^{d_B-1} \alpha_{ii}\beta_{iikk} |i\rangle\langle i|_A
\end{aligned} \tag{30}$$

and

$$\begin{aligned}
&\mathrm{Tr}_B[P_{AB}(Q_A \otimes \mathbb{1}_B)] \\
&= \sum_{k=0}^{d_B-1} \langle k|_B \sum_{ij=0}^{d_A-1} \sum_{kl=0}^{d_B-1} \beta_{ijkl} \left(|i\rangle_A \otimes |k\rangle_B\right) \left(\langle j|_A \otimes \langle l|_B\right) \sum_{ij=0}^{d_A-1} \alpha_{ij} \left(|i\rangle\langle j|_A \otimes \mathbb{1}_B\right) |k\rangle_B \\
&= \sum_{ij=0}^{d_A-1} \sum_{kl=0}^{d_B-1} \alpha_{ij}\beta_{ijkl} \langle k|k\rangle_B \left(|i\rangle_A \otimes \mathbb{1}_B\right) \left(\langle j|_A \otimes \langle l|_B\right) \left(|i\rangle\langle j|_A \otimes |k\rangle\langle k|_B\right) |k\rangle_B \\
&= \sum_{ij=0}^{d_A-1} \sum_{kl=0}^{d_B-1} \alpha_{ij}\beta_{ijkl} \left(|i\rangle\langle j|_A \otimes \langle l|_B\right) \left(|i\rangle\langle j|_A \otimes |k\rangle_B\right) \langle k|k\rangle_B \\
&= \sum_{i=0}^{d_A-1} \sum_{k=0}^{d_B-1} \alpha_{ii}\beta_{iikk} |i\rangle\langle i|_A.
\end{aligned} \tag{31}$$

From Eq. (30) and (31) we can prove Eq. (27). □

Now we are ready to prove Proposition 6.

Proof of Proposition 6. We take an example of single-qubit case. With the reduced Choi state $J_{A'_1B_1}$ defined in Eq. (15), the element of PTM representation can be computed by

$$\begin{aligned}
\left(R_{\mathcal{N}(A_1\to B_1)}\right)_{i_1 j_1} &= \mathrm{Tr}\left[J_{A'_1B_1}(P_{j_1}^T \otimes P_{i_1})\right] \\
&= \mathrm{Tr}\left[\mathrm{Tr}_{A'_2\cdots A'_nB_2\cdots B_n}\left[J_{A'_1\cdots A'_nB_1\cdots B_n}\right](P_{j_1}^T \otimes P_{i_1})\right] \\
&\qquad \text{Using Lemma 7} \\
&= \mathrm{Tr}\left[\mathrm{Tr}_{A'_2\cdots A'_nB_2\cdots B_n}\left[J_{A'_1\cdots A'_nB_1\cdots B_n}(P_{j_1\cdot 4^{n-1}}^T \otimes P_{i_1\cdot 4^{n-1}})\right]\right] \\
&= \mathrm{Tr}\left[J_{A'_1\cdots A'_nB_1\cdots B_n} P_{j_1\cdot 4^{n-1}}^T \otimes P_{i_1\cdot 4^{n-1}}\right] \\
&= \left(R_{\mathcal{N}(A_1\cdots A_n\to B_1\cdots B_n)}\right)_{(i_1\cdot 4^{n-1})(j_1\cdot 4^{n-1})},
\end{aligned} \tag{32}$$

where i_1, j_1 range from 0 to 3. Thus, the Choi representation and the PTM representation can be transformed into each other. Then, we consider Definition 1. Through Definition 1, we can also deduce the PTM representation of the reduced quantum process $\mathcal{N}_{A_1 \to B_1}$ as:

$$
\begin{aligned}
\left(R_{\mathcal{N}(A_1 \to B_1)}\right)_{i_1 j_1} &= \frac{1}{d_1} \operatorname{Tr}\left[P_{i_1} \mathcal{N}(P_{j_1})\right] \\
&= \frac{1}{d_1} \operatorname{Tr}\left[P_{i_1} \operatorname{Tr}_{B_2 \cdots B_n}\left[U(P_{j_1} \otimes \frac{1}{d_{n-1}} \mathbb{1}_{A_2 \cdots A_n}) U^\dagger\right]\right] \\
&\qquad \text{Using Lemma 7} \\
&= \frac{1}{d_1} \operatorname{Tr}\left[\operatorname{Tr}_{B_2 \cdots B_n}\left[\frac{1}{d_{n-1}} P_{i_1 \cdot 4^{n-1}} U P_{j_1 \cdot 4^{n-1}} U^\dagger\right]\right] \\
&= \frac{1}{d} \operatorname{Tr}\left[P_{i_1 \cdot 4^{n-1}} U P_{j_1 \cdot 4^{n-1}} U^\dagger\right] \\
&= \frac{1}{d} \operatorname{Tr}\left[P_{i_1 \cdot 4^{n-1}} \mathcal{N}(P_{j_1 \cdot 4^{n-1}})\right] \\
&= \left(R_{\mathcal{N}(A_1 \cdots A_n \to B_1 \cdots B_n)}\right)_{(i_1 \cdot 4^{n-1})(j_1 \cdot 4^{n-1})}
\end{aligned}
$$

which is consistent with Eq. (16).

□

A Reconfigurable Chip-Scale Quantum Key Distribution Receiver Based on Silicon Nitride

Denis Fatkhiev[1] * Hui Liu[1] Alexander Grebenchukov[1] Menno van den Hout[1]
Aaron Albores-Mejia[1 2] Chigo Okonkwo[1 2] Idelfonso Tafur Monroy[1]

[1] *Electro-Optical Communication Group, Eindhoven University of Technology, The Netherlands*
[2] *CUbIQ Technologies, Flux Building, De Groene Loper 19, Eindhoven, The Netherlands*

Abstract. A silicon nitride photonic integrated QKD receiver employing tunable couplers is demonstrated, enabling on-chip reconfigurability. The proposed approach achieves low quantum bit error rates below 1.5 % across different receiver configurations, providing a migration strategy towards multi-protocol quantum-secured communication with optimized on-chip tuning.

Keywords: optical communication, photonic integration, quantum key distribution

1 Introduction

The vast exchange of information has unquestionably become an indispensable aspect of modern society and economy. Sensitive information such as financial, health, or governmental data is a natural target for malicious actors, making its security crucial. Modern cryptography methods, including widely used public-key cryptography, rely on the computational complexity of decryption problems (factoring large numbers), whilst *harvest now, decrypt later* attacks are currently ongoing, potentially exposing data in the future due to the emergence of quantum computing [1], [2]. However, quantum key distribution (QKD) methods [3]–[5], predicated on quantum physics phenomena, may not only protect real-time communication traffic but also data that require long-term secrecy. Here, we focus on discrete-variable QKD, where the information is mapped to discrete quantum states, such as polarization, phase, etc., of a single photon. Since a QKD system is implemented on the physical layer over which a quantum channel operates, specific components, such as single-photon detectors (SPDs), single-photon sources, and low-noise analog circuits, are required. Furthermore, different protocols from the QKD "protocol family" may address distinct tasks in networks [6], hence requiring a hybrid approach and rendering the advantage of having universal receiver hardware compatible with multiple protocols. Therefore, practical deployment of QKD systems depends on scalable, cost-effective hardware foundations such as photonic integrated circuits (PICs) [7], proven for their power efficiency, miniaturization, and compatibility with state-of-the-art semiconductor technology.

In this work, we aim to tackle the challenges mentioned above by exploiting photonic integration to build a versatile receiver that can potentially be applied to various QKD protocols such as BB84 [8], differential phase shift (DPS) [9], and coherent one-way (COW) [10]. We demonstrate a low-loss silicon nitride PIC featuring tunable couplers (TCs) and an asymmetric Mach–Zehnder interferometer (AMZI). With the TC's high extinction ratio (ER) of >35 dB, we gain the ability to reconfigure the power ratio between measurement paths in a wide range. This allows for the receiver's basis selection probability adjustment and dynamic switching among multiple QKD protocols. We show low quantum bit error rates (QBERs) of 0.5 %–1.5 % using commercially available InGaAs SPDs in different receiver configurations, which are sufficient for generating high secret key rates (SKRs).

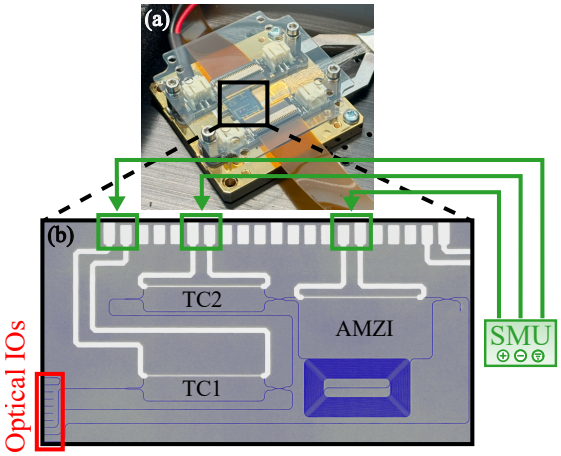


Figure 1: **(a)** Optoelectronic assembly and **(b)** microscope image of the silicon nitride PIC.

2 Integrated Receiver

The receiver assembly and the silicon nitride PIC are shown in Fig. 1a and Fig. 1b, respectively. The circuitry includes two TCs and an AMZI. The optical IOs (inputs and outputs) are implemented using a spot-size converter (SSC) array on the PIC edge. Tunability of a TC is achieved by a symmetric Mach-Zehnder interferometer containing a TOPS in one of the arms, so the relative phase of the light in the two arms can be modified, enabling control over constructive and destructive interference, determining the amount of light directed to each output of the TC. The first TC allows us to adjust the splitting ratio between two measurement paths, related to the choice of the

*d.fatkhiev@tue.nl

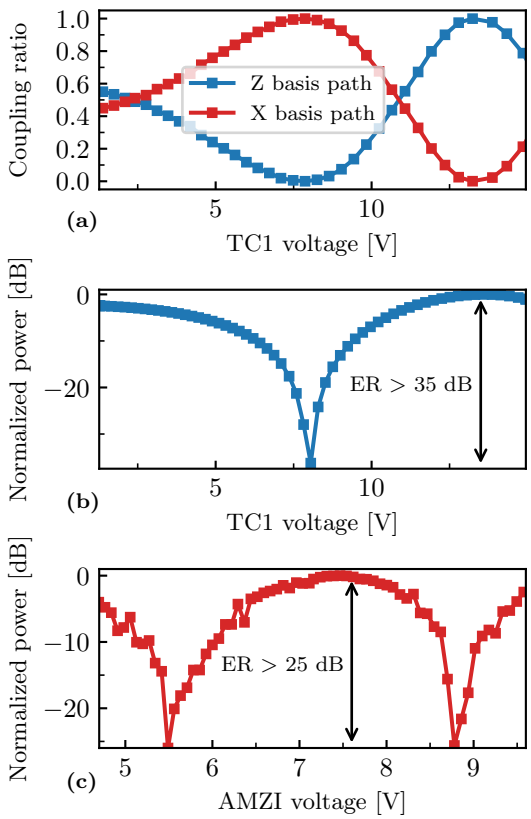


Figure 2: Integrated receiver characterization. **(a)** TC1 coupling ratio coverage; **(b)** TC1 power output varied over its TOPS voltage; **(c)** AMZI power output varied over its TOPS voltage.

measurement basis by the so-called Bob, a traditional (in quantum cryptography) misnomer for the party detecting and decoding the quantum state during the protocol (the party that generates the quantum state is called Alice). The second TC enables power balancing between arms of the AMZI to achieve higher ER by compensating for the imbalance introduced due to the presence of the delay line. The TOPS in the AMZI arm is used to fine-tune the relative phase for state demodulation efficiency.

For the fabrication, the TriPleX silicon nitride platform [11] was selected due to its low-propagation-loss

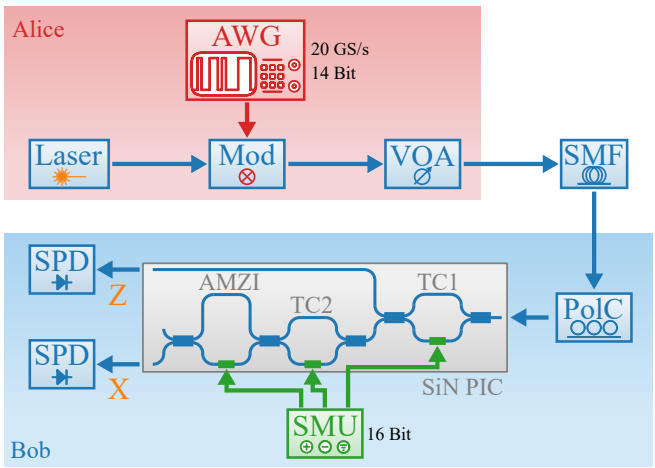


Figure 3: Experimental setup for performance evaluation of the QKD receiver.

waveguides (<0.2 dB/cm), which are important to mitigate extra losses from the ≈900 ps delay line and couplers. Overall losses through the TC1 path (Z basis) and the AMZI path (X basis) are ≈5.3 dB and ≈9.4 dB, respectively.

The PIC is wire-bonded and co-packaged with a polarization-maintaining fiber array. The insertion loss (IL) per SSC is around 1.1 dB. As shown in Fig. 1, a multichannel source measure unit (SMU) is used to independently actuate the phase shifters. The device characterization data is presented in Fig. 2. The first TC covers the whole coupling ratio range, which can be seen from Fig. 2a. The extinction ratios of the TCs and AMZI can be seen from Fig. 2b and Fig. 2c, which are >35 dB and >25 dB, respectively.

3 Quantum Key Distribution Setup

To showcase the performance of the designed PIC, we implemented the three-state time-bin BB84 protocol [12], and the experimental setup is illustrated in Fig. 3. The states are encoded in two bases: Z and X. Z basis includes two states Z_0 and Z_1, a weak coherent pulse in the first (Fig. 4a) or the second (Fig. 4b) time bin, respectively. The X basis only includes one state X_0: a superposition of the first and second time-bin with zero relative phase (Fig. 4c).

The transmitter, or Alice, consists of a low-linewidth tunable laser (<100 kHz), an arbitrary-waveform generator (AWG), and a Mach-Zehnder modulator (Mod) for the generation of optical pulses at a 2.2 GHz repetition rate with an operating wavelength of 1550 nm. The time bins have a temporal separation of ≈900 ps to match the AMZI's delay line. The optical pulse intensity

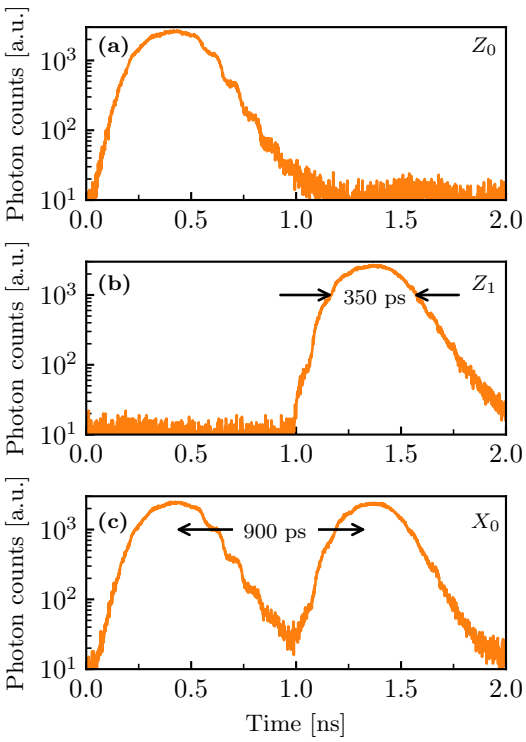


Figure 4: Histogram measurements of the states on the transmitter output: **(a)** Z_0, **(b)** Z_1, and **(c)** X_0.

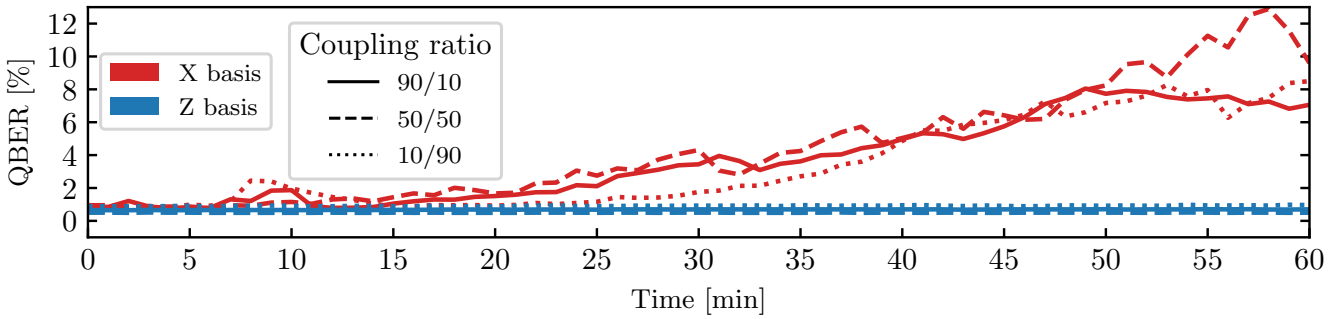


Figure 5: Evolution of QBER over time for different receiver configurations.

is significantly attenuated to a single-photon level using a variable optical attenuator (VOA). Optical pulses are then sent to the receiver through a short piece of single-mode fiber (SMF).

The receiver, or Bob, includes a polarization controller (PolC), silicon nitride PIC, and two free-running InGaAs SPDs. A PolC is employed to align the polarization of incoming pulses with the polarization axis of the PIC waveguide. Once a quantum state reaches Bob, a passive measurement basis is selected by choosing the measurement output to read (either Z or X, as shown on Fig. 3). When the Z basis is chosen, the states are directly transmitted to a SPD, which measures the arrival time of the photons. In the case of X basis pick, the coherence between two consecutive pulses is recovered by an AMZI and measured with an SPD. After this, an estimation of the QBER based on the statistics collected using SPDs is carried out.

4 Results

To examine the versatility of the receiver PIC, the approach in this work involved performing a set of QBER measurements with various coupling ratio values set on TC1, while TC2 was set to balance AMZI arms, hence providing the highest ER possible. Note that the TOPS phase shift in the AMZI arm is a free parameter we tweak before measurements to fine-tune the relative phase for precise demodulation of quantum states in the X basis.

The motivation for testing the system over different coupling ratio configurations arises from reported works on security analysis of BB84 and COW protocols [13]–[15]. Even though in published works [16], [17], coupling ratios on the receiver side were usually fixed as the used couplers were not tunable, it was shown that optimization variables related to coupling ratios alter over different quantum channel losses. For instance, in a one-decoy state BB84 protocol [14], to achieve optimal performance, the coupling ratio varies from 5 % to 35 %, and the experiment implementing 4-intensity decoy-state BB84 [18] utilizes different coupling ratio values ranging from 15 % to 65 %. Hence, coupler tunability is beneficial for QKD performance optimization.

We measured QBERs over time for three different coupling ratios of TC1: 90/10, 50/50, and 10/90, where the first value corresponds to the Z basis path and the second value to the X basis path. The results are illustrated in Fig. 5. The average QBER in the X basis (QBER-X) for each configuration remains below 1.5 % throughout the continuous 10-minute test period, which is sufficiently low for generating a high SKR. However, due to the absence of temperature control for the receiver assembly, the QBER-X (Fig. 5) slowly drifts as the delay line is sensitive to the environment, causing a relative variation in the phase. This issue can be addressed by adding feedback or thermal control [19]. Additionally, as depicted in Fig. 5, the QBER in the Z basis (QBER-Z) remains stable with an average value below 1 % for each configuration during continuous testing lasting 1 hour.

5 Conclusions and Discussion

A reconfigurable silicon nitride PIC-based QKD receiver was demonstrated. By employing flexible structures, including tunable couplers with a high extinction ratio of >35 dB and a low-loss >25 dB extinction ratio asymmetric Mach-Zehnder interferometer, a higher degree of control and tunability is achieved in the receiver. This enables adjustment of the receiver's basis selection probability, resulting in the possibility to operate in different QKD protocols and efficient quantum state decoding. We note that similar receiver architectures were used in some previous works [20], [21]. However, in our work, we employed another protocol and basis definition and explicitly analyzed the receiver in altering configurations. We reported low QBER values of 0.5 %–1.5 % in various receiver settings, which paves the way to multi-protocol QKD systems with high secret key rates. By leveraging dedicated electronic drivers [22] and integrated single-photon detectors [23], further efforts can be directed toward achieving the ultimate system-in-a-package solution.

Acknowledgements

This work was supported by the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the Quantum Delta NL and the PhotonDelta National Growth Funds Programme.

References

[1] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Review*, vol. 41, no. 2, pp. 303–332, 1999.

[2] F. Arute, K. Arya, R. Babbush, *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[3] H.-K. Lo, M. Curty, and K. Tamaki, "Secure quantum key distribution," *Nature Photonics*, vol. 8, no. 8, pp. 595–604, 2014.

[4] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, "Secure quantum key distribution with realistic devices," *Reviews of modern physics*, vol. 92, no. 2, p. 025 002, 2020.

[5] M. Mehic, L. Michalek, E. Dervisevic, *et al.*, "Quantum cryptography in 5g networks: A comprehensive overview," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 1, pp. 302–346, 2023.

[6] M. Peev, C. Pacher, R. Alléaume, *et al.*, "The secoqc quantum key distribution network in vienna," *New Journal of Physics*, vol. 11, no. 7, p. 075 001, 2009.

[7] A. Orieux and E. Diamanti, "Recent advances on integrated quantum communications," *Journal of Optics*, vol. 18, no. 8, p. 083 002, 2016.

[8] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *International Conference on Computers, Systems, and Signal Processing*, vol. 1, pp. 175–179, 1984.

[9] K. Inoue, E. Waks, and Y. Yamamoto, "Differential phase shift quantum key distribution," *Physical Review Letters*, vol. 89, p. 037 902, 3 2002.

[10] D. Stucki, N. Brunner, N. Gisin, V. Scarani, and H. Zbinden, "Fast and simple one-way quantum key distribution," *Applied Physics Letters*, vol. 87, no. 19, p. 194 108, 2005.

[11] F. Morichetti, A. Melloni, M. Martinelli, *et al.*, "Box-shaped dielectric waveguides: A new concept in integrated optics?" *Journal of Lightwave technology*, vol. 25, no. 9, pp. 2579–2589, 2007.

[12] A. Boaron, G. Boso, D. Rusca, *et al.*, "Secure quantum key distribution over 421 km of optical fiber," *Physical review letters*, vol. 121, no. 19, p. 190 502, 2018.

[13] C. C. W. Lim, M. Curty, N. Walenta, F. Xu, and H. Zbinden, "Concise security bounds for practical decoy-state quantum key distribution," *Physical Review A*, vol. 89, no. 2, p. 022 307, 2014.

[14] D. Rusca, A. Boaron, F. Grünenfelder, A. Martin, and H. Zbinden, "Finite-key analysis for the 1-decoy state qkd protocol," *Applied Physics Letters*, vol. 112, no. 17, p. 171 104, 2018.

[15] J. González-Payo, R. Trényi, W. Wang, and M. Curty, "Upper security bounds for coherent-one-way quantum key distribution," *Physical Review Letters*, vol. 125, no. 26, p. 260 510, 2020.

[16] Y. Lo, R. Woodward, T. Roger, *et al.*, "Self-tuning transmitter for quantum key distribution using machine intelligence," *Physical Review Applied*, vol. 18, no. 3, p. 034 087, 2022.

[17] N. K. Pathak, S. Chaudhary, Sangeeta, and B. Kanseri, "Phase encoded quantum key distribution up to 380 km in standard telecom grade fiber enabled by baseline error optimization," *Scientific Reports*, vol. 13, no. 1, p. 15 868, 2023.

[18] H. Liu, Z.-W. Yu, M. Zou, *et al.*, "Experimental 4-intensity decoy-state quantum key distribution with asymmetric basis-detector efficiency," *Physical Review A*, vol. 100, no. 4, p. 042 313, 2019.

[19] P. Sibson, C. Erven, M. Godfrey, *et al.*, "Chip-based quantum key distribution," *Nature communications*, vol. 8, no. 1, p. 13 984, 2017, *Supplementary information*.

[20] P. Sibson, J. E. Kennard, S. Stanisic, C. Erven, J. L. O'Brien, and M. G. Thompson, "Integrated silicon photonics for high-speed quantum key distribution," *Optica*, vol. 4, no. 2, pp. 172–177, 2017.

[21] P. Sibson, C. Erven, M. Godfrey, *et al.*, "Chip-based quantum key distribution," *Nature communications*, vol. 8, no. 1, p. 13 984, 2017.

[22] C.-X. Zhu, Z.-Y. Chen, Y. Li, *et al.*, "Experimental quantum key distribution with integrated silicon photonics and electronics," *Physical Review Applied*, vol. 17, no. 6, p. 064 034, 2022.

[23] F. Beutel, H. Gehring, M. A. Wolff, C. Schuck, and W. Pernice, "Detector-integrated on-chip qkd receiver for ghz clock rates," *npj Quantum Information*, vol. 7, no. 1, p. 40, 2021.

Efficient Transpilation of Quantum Circuits to Quantum Intermediate Representation

1st Sengthai Heng
Department of AI Convergence
Pukyong National University
Nam-gu, Busan 48513, South Korea
sengthai@pukyong.ac.kr

2nd Nagyeong Choi
Department of AI Convergence
Pukyong National University
Nam-gu, Busan 48513, South Korea
choi2019@pukyong.ac.kr

3rd Kimchhor Chiv
Department of AI Convergence
Pukyong National University
Nam-gu, Busan 48513, South Korea
kimchhor@pukyong.ac.kr

4th Youngsun Han
Department of AI Convergence
Pukyong National University
Nam-gu, Busan 48513, South Korea
youngsun@pknu.ac.kr

***Abstract*—Quantum computing holds vast potential but faces limitations, such as inefficient programming and transpilation time of quantum algorithms. In response to this demand, we present a transpiler framework for transforming Qiskit's quantum circuit objects to LLVM Quantum Intermediate Representation (QIR). Our transpiler leverages LLVMLite to efficiently transpile quantum circuits to QIR. Key features include support for single-qubit and two-qubit gate operations, parameter handling, and measurement operations. We conduct a comparative analysis of the transpilation time between our QCC transpiler and QCOR transpiler, a hybrid quantum-classical programming tool, showcasing a significant reduction in transpilation time. Through LLVMLite integration and custom instruction definitions, we achieve efficient and optimized transpilation. Our findings demonstrate a 99.3% reduction on average in transpilation time compared to an existing method. This work contributes to advancing the field of quantum programming by providing an efficient tool for transpiling quantum circuits to LLVM-based representations.**

***Index Terms*—Quantum Programming, Qiskit Circuits, LLVM QIR Transpilation**

I. Introduction

Quantum computing has the potential to revolutionize various fields, from cryptography [1] to optimization [2]. However, efficient programming and transpilation of quantum algorithms remain challenges. Existing research in quantum programming has focused on various aspects, including language design, optimization techniques, and compilation strategies. One notable approach is using intermediate representations (IRs) to facilitate efficient compilation of quantum algorithms. LLVM [3], a widely-used classical compiler infrastructure, has been adapted for quantum computing by developing LLVM Quantum Intermediate Representation (QIR). Several works [4]–[6] have been made to transpile quantum circuits from high-level programming languages to MLIR and LLVM IR. These efforts aim to bridge the gap between quantum programming frameworks and efficient execution on quantum hardware or simulators. One relevant work in this domain is the development of QCOR [6], a C++ language extension and associated compiler implementation for hybrid quantum-classical programming. QCOR leverages Pybind11 [7] and LLVM to generate QIR. However, existing transpilation tools often face challenges related to performance, scalability, and compatibility with quantum hardware architectures.

This paper addresses the transpilation challenge by presenting a transpiler for converting Qiskit's quantum circuits to LLVM QIR. We present QCC transpiler, a frontend of our QCC (quantum-classical compiler) framework, which simplifies the development process for QIR code generation by Leveraging LLVMLite [8] a lightweight LLVM binding for Python. LLVMLite is a Python-LLVM interface that provides APIs to construct LLVM IR code, optimizes and generates code, and integrates with Python's data structures. We introduce key features that support single-qubit and two-qubit gate operations, parameter handling (for rotation gates), and measurement operations. Additionally, we compare the transpilation time of our transpiler with the existing method, demonstrating efficiency improvements. This work advances quantum programming by providing an efficient tool to compile circuits to LLVM-based representations. Through LLVMLite integration and custom instruction definitions, we achieve efficient and optimized transpilation. Our findings showcase a 99.3% reduction on average in transpilation time compared to an existing method. This work contributes to advancing the field of quantum programming by offering an efficient tool for transpiling quantum circuits to LLVM-based representations.

II. Transpiler Design and Implementation

Figure 1 shows The quantum circuit compiler works in two stages: frontend and backend. The front end translates quantum circuit code into an IR format. The backend takes this IR and optimizes and tailors it for a particular quantum processor. Finally, it outputs machine code that is understandable by that specific QPU. This work focuses on the frontend part, explicitly designing and implementing our transpiler to convert Qiskit's quantum circuits to LLVM QIR.

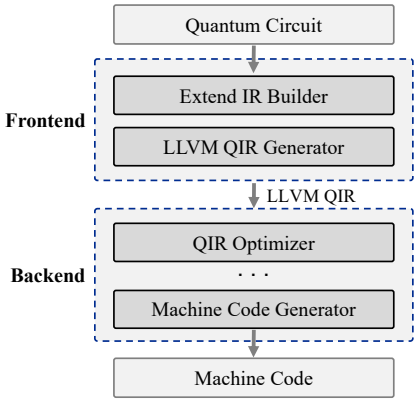


Fig. 1. Overview of Quantum Circuit Compilation Workflow

```
from qiskit import QuantumCircuit
from qvm import qir

circ = QuantumCircuit(2)
circ.h(0)
circ.cx(0, 1)

qcc_qir_str = qir.transpile(circ)
print(qcc_qir_str)
```

Code 1. Example usage of a QCC transpiler

```
define void @"circuit-160"(){
entry:
  %".2" = alloca q1
  %".3" = alloca q1
  %".4" = load q1, q1* %".2"
  %".5" = h q1 %".4"
  store q1 %".5", q1* %".2"
  %".7" = load q1, q1* %".2"
  %".8" = load q1, q1* %".3"
  %".9" = ctrl q1 %".7"
  %".10" = x q1 %".9", q1 %".8"
  store q1 %".9", q1* %".2"
  store q1 %".10", q1* %".3"
  ret void
}
```

Code 2. LLVM QIR output

A. Design Overview

Our transpiler implementation follows a modular design that is comprised of several key components to ensure extensibility and maintainability. At its core, the QIR Generator converts Qiskit quantum circuits into LLVM QIR code. We extend LLVMLite's IRBuilder with the `ExtendedIRBuilder` class, enabling it to handle quantum-specific operations and instructions. Additionally, the `QInstruction` class is inherited from LLVMLite's Instruction to introduce quantum instructions within LLVM QIR, offering instruction generation and manipulation methods. The `QubitType` class represents the quantum type within LLVM QIR. The QIR generator is then encapsulated in the `qvm` (quantum virtual machine) package. This design abstracts the complexities of LLVM QIR generation, providing a flexible foundation for customization and optimization. Code 1 shows an example of our transpiler usage. Moreover, the output from our transpiler, as depicted in Code 2, provides a glimpse into the LLVM QIR representation of a quantum circuit. In this representation, the quantum circuit comprises 2 qubits and 2 gates: the Hadamard gate `h` and the Controlled-NOT gate (`ctrl` and `x`). We implemented QCC transpiler in Python, leveraging the LLVMLite library for LLVM integration. The implementation consists of the following key steps.

B. QIR Generation

The QIR Generator module parses Qiskit quantum circuits and generates corresponding LLVM QIR code. This process involves several key steps, as illustrated in Code 3. First, a new LLVM IR module is created to contain the generated QIR code. This module is named based on the name of the input quantum circuit. Next, a new LLVM QIR function is defined within the module to represent the quantum circuit. This function serves as the entry point for the QIR code generation process. A basic block is appended to the function to contain the instructions of the quantum circuit. This basic block, named "entry", will contain the QIR instructions. An `ExtendedIRBuilder` object is created to generate QIR instructions within the basic block. This builder extends LLVMLite's IRBuilder to handle quantum-specific operations and instructions. The QIR generator also creates LLVM IR types representing qubits and classical bits within the generated QIR code. Once the necessary types and registers are defined, the generator transpiles LLVM QIR instructions corresponding to the operations in the quantum circuit. Finally, the generator completes the QIR generation process by returning the LLVM IR module containing the generated QIR code.

C. Extend IR Builder

The Extended IR Builder module extends LLVMLite's IRBuilder class to support quantum-specific gate operations and instructions. It provides convenient methods for constructing quantum instructions and managing quantum registers. The decorator function in Python is employed to extend the behavior of the `_gate_op` function to support quantum instructions, as shown in Code 4. Each decorator function takes the gate operation name as an argument and defines a corresponding method within the Extended IR Builder class. The `@_gate_op` decorator function is a wrapper for creating gate operations. The wrapped function within the decorator, named 'wrapped', defines the behavior of the generated gate operation method. It takes the qubits as input, along with an optional name for the gate operation. Inside the wrapped function, an instruction object (`QInstruction`) representing the specified gate operation is instantiated. This instruction object is then inserted into the LLVM QIR basic block associated with the Extended IR Builder.

```
...
class QIRGenerator():
 """LLVM QIR generator main class."""

  def generate(self, qc: QuantumCircuit) -> str:
   fntype = FunctionType(VoidType(), [])
   # Create a new LLVM IR module
   module = Module(name=f"file_{qc.name}")

   # Create a new LLVM IR function
   func = Function(module, fntype, name=qc.name)

   # Create a new LLVM IR basic block
   basic_block = func.append_basic_block(name="
   entry")

   # Create a Basic Block Builder
   self._builder = ExtendedIRBuilder(basic_block)

   # Create new LLVM IR qubit types
   self._qubits = self._build_qubit_registers();

   # Create new LLVM IR classical types
   self._bits = self._build_classical_registers();

   # Create new LLVM IR instructions
   self._build_instructions(qc)

   # Return the LLVM IR module
   self._builder.ret_void()

   return module
...
```

Code 3. LLVM QIR generator main class.

D. Quantum Instruction and Type

The `QInstruction` class represents quantum instructions in LLVM QIR. It encapsulates the logic for generating LLVM IR code for quantum operations and provides methods for instruction manipulation. The `QubitType` class represents the quantum type (`q1`) in LLVM QIR, extending LLVMLite's Type class.

III. Performance Evaluation

A. Experimental Setup

This section outlines our experimental setup to evaluate the transpilation performance of QCC and QCOR transpilers. The evaluation employed several tools and libraries. These included Qiskit, a comprehensive quantum computing framework for quantum circuit construction and manipulation; QCOR [6], a quantum-classical programming language compiler written in C++ and utilized Pybind11 [7] for frontend; QCC transpiler, for converting the quantum circuit to LLVM QIR.

The experimental procedure involved several steps. Initially, random quantum circuits with varying numbers of qubits and depths were generated using Qiskit's circuit generation utilities. Subsequently, the generated quantum circuits were transpiled to LLVM QIR using both QCOR and QCC transpilers, with the transpilation process timed to measure the execution time for each transpiler. Experimental parameters, including the number of qubits (ranging from 5 to 25) and the depth of circuits (ranging from 100 to 1000 supported), were

```

from llvmlite.ir.builder import IRBuilder

def _gate_op(opname, cls=QInstruction):
   """Decorator function for creating gate
   operations."""
   def wrap(fn):
       @functools.wraps(fn)
       def wrapped(self, qubits, name=''):
          if not isinstance(qubits, list):
              qubits = [qubits]
          instr = cls(self.block, QubitType(),
                  opname, qubits, name)
          self._insert(instr)
          return instr
       return wrapped
return wrap
...
class ExtendedIRBuilder(IRBuilder):
  @_gate_op('x')
  def x(self, lhs, name=''):
      """ X gate operation. """
...
```

Code 4. Decorator function for creating single-qubit gate operations.

```

from llvmlite.ir.types import Type

class _BaseQubitType(Type):
   """ Represents the base qubit type."""
   ...
   @classmethod
   def _create_instance(cls):
       cls._instance_cache = super(_BaseQubitType,
   cls).__new__(cls)

class QubitType(_BaseQubitType):
   """ The type for qubits. """
   null = '0'
   intrinsic_name = 'q1'

   def __str__(self):
       return 'q1'

   def format_constant(self, value):
       return str(value)

QubitType._create_instance()
```

Code 5. Represents the classes of qubit type.

varied to assess the scalability and efficiency of the transpilers under different circuit complexities.

B. Performance Analysis

Figure 2 compares transpilation time between the QCOR and QCC IR transpiler methods. The graph illustrates the relationship between the number of qubits (X-axis), the depth of the circuit (Y-axis), and the transpilation time in seconds (Z-axis). Based on the experimental results, the average transpilation time for generating LLVM QIR using the proposed transpiler was approximately 0.189 seconds, while the average transpilation time for the QCOR transpiler was approximately 28.303 seconds. This indicates a significant reduction in transpilation time when using our transpiler compared to QCOR,

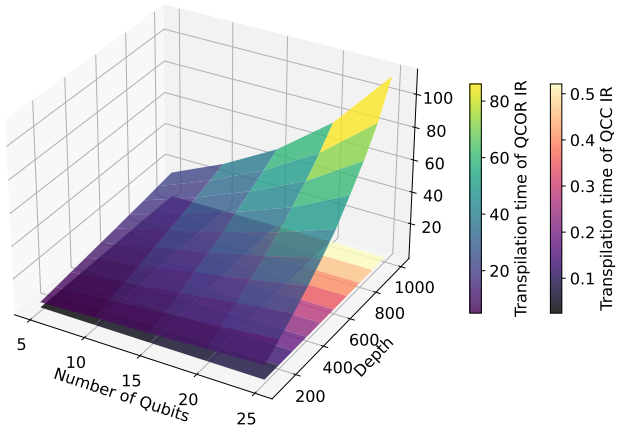


Fig. 2. Comparison of transpilation time for QCOR and QCC IR transpiler. The X-axis represents the number of qubits, the Y-axis represents the depth of the circuit, and the Z-axis represents the transpilation time in seconds.

with a percentage reduction of approximately 99.33%. This significant reduction highlights the effectiveness of our transpiler in the IR transpilation workflow for quantum circuits.

IV. Conclusion

In conclusion, we have presented a transpiler for converting Qiskit's quantum circuits to LLVM Quantum Intermediate Representation (QIR). Through experimental evaluation, we demonstrated the efficiency of our transpiler in significantly reducing transpilation time compared to the existing framework QCOR. The experimental results showcased an average execution time reduction of approximately 99.33% when using our transpiler compared to QCOR. Our transpiler performs superior transpilation time reduction across all circuit sizes tested. Notably, as the complexity of quantum circuits increases, the efficiency gains become more pronounced, showcasing the scalability of our approach. Furthermore, our framework's modular design facilitates extensibility and maintainability, allowing for future enhancements and optimizations.

Acknowledgment

This work was supported by Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00014, A Technology Development of Quantum OS for Fault-tolerant Logical Qubit Computing Environment).

References

[1] P. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*. Santa Fe, NM, USA: IEEE Comput. Soc. Press, 1994, pp. 124–134. [Online]. Available: http://ieeexplore.ieee.org/document/365700/

[2] J. Basso, E. Farhi, K. Marwaha, B. Villalonga, and L. Zhou, "The quantum approximate optimization algorithm at high depth for MaxCut on large-girth regular graphs and the sherrington-kirkpatrick model," in *17th conference on the theory of quantum computation, communication and cryptography (TQC 2022)*, ser. Leibniz international proceedings in informatics (LIPIcs), F. Le Gall and T. Morimae, Eds., vol. 232. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, pp. 7:1–7:21, iSSN: 1868-8969 tex.urn: urn:nbn:de:0030-drops-165144. [Online]. Available: https://drops.dagstuhl.de/opus/volltexte/2022/16514

[3] C. Lattner and V. Adve, "Llvm: A compilation framework for lifelong program analysis & transformation," in *Proceedings of the International Symposium on Code Generation and Optimization: Feedback-Directed and Runtime Optimization*, ser. CGO '04. USA: IEEE Computer Society, 2004, p. 75.

[4] A. McCaskey and T. Nguyen, "A mlir dialect for quantum assembly languages," in *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Los Alamitos, CA, USA: IEEE Computer Society, oct 2021, pp. 255–264. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/QCE52317.2021.00043

[5] A. JavadiAbhari, S. Patil, D. Kudrow, J. Heckey, A. Lvov, F. T. Chong, and M. Martonosi, "ScaffCC: Scalable compilation and analysis of quantum programs," *Parallel Computing*, vol. 45, pp. 2–17, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167819114001422

[6] A. Mccaskey, T. Nguyen, A. Santana, D. Claudino, T. Kharazi, and H. Finkel, "Extending c++ for heterogeneous quantum-classical computing," *ACM Transactions on Quantum Computing*, vol. 2, no. 2, jul 2021. [Online]. Available: https://doi.org/10.1145/3462670

[7] W. J. et al., "pybind/pybind11," original-date: 2015-07-05T19:46:48Z. [Online]. Available: https://github.com/pybind/pybind11

[8] "numba/llvmlite," original-date: 2014-08-07T14:51:28Z. [Online]. Available: https://github.com/numba/llvmlite

Development of a single photon source and its application at room temperature in KRISS

Kee-Suk Hong[1] * Hee-Jin Lim[1] † Wook-jae Lee[2] ‡ Jin-Kyu Yang[2] §

[1] *Korea Research Institute of Standards and Science*
[2] *Kongju National University*

Abstract. KRISS has developed and applied room temperature single photon sources utilizing diamond silicon vacancy, gallium nitride defects, and hexagonal boron nitride vacancy. The study observed varying relaxation times influencing photon stability, with GaN demonstrating high stability and hBN exhibiting high photon emission rates (over $10^6/s$).

Keywords: single photon source, quantum radiometry, quantum bio-sensing, quantum wireless communication

1 Introduction

Accurate photon measurement with high repeatability and low uncertainty is essential in few-photon metrology based on photon number [1-6]. Low photon number fluctuations and high repeatability are critical for qualifying a standard light source, but phenomena such as blinking and internal relaxations in single photon emitters can constrain these qualities [7-9], with variations observed across different materials. This study focuses on room temperature single photon emitters, including silicon vacancy in diamond (SiV), defects in gallium nitride (GaN), and vacancy in hexagonal boron nitride (hBN), which are known for their spectrally narrow and accessible platforms for single photon fluorescence.

We investigate the photon number statistics and fluctuations of these emitters, as they significantly influence the accuracy of photon flux for radiometry applications. Additionally, we compare the maximum count rates achievable with these materials using conventional confocal microscopy collection techniques. Detection count rates depend on refractive index geometries and detection techniques. While our experiments are limited by estimations of internal quantum efficiency and theoretical maximum count rates under continuous wave operation, future application-oriented studies are needed to further optimize collection efficiency, a crucial aspect of photonics.

To discern general tendencies and characteristics amidst the complexity and variety of our materials, this study is based on a substantial dataset collected from numerous emitters. Our dataset comprises two levels: the first includes basic properties for identifying single photon emitters, while the second encompasses the utilization of source characteristics in applications. Data fields in the first level include photon coincidence correlation g(2)(0) spectra and stability, which are used to authenticate single photon fluorescence. Statistical distributions of the positions of spectral peaks were collected for subsequent studies on defect states and their formations.

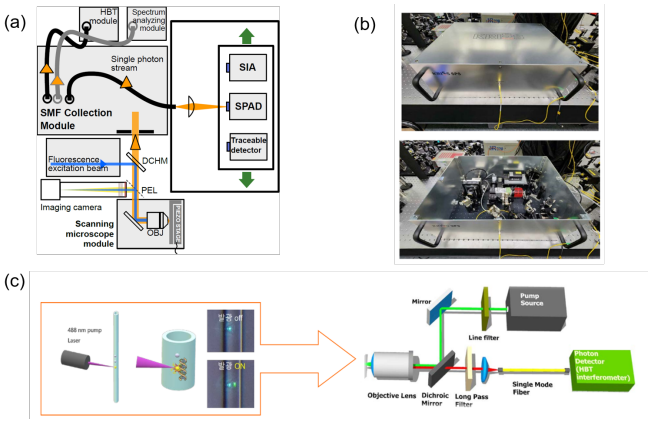


Figure 1: Presentation examples of applications in these three fields, showcasing the research findings and potential uses of each material. (a) Illustrates the experimental setup for comparing SPAD and traceable detectors, which have different units, in the context of quantum radiometry. (b) Shows the single photon source setup developed for plug-and-play quantum wireless communication. (c) Depicts the flow cytometry experimental setup used for counting DNA and RNA molecules.

2 Result and Discussion

In order to find the optimal material for realizing quantum radiometry, KRISS evaluated single photon source characteristics using various materials such as silicon vacancy nano-diamond, gallium nitride and hexagonal-boron nitride. Figure 1 presents examples of applications in these three fields, showcasing the research findings and potential uses of each single emitters. In the field of quantum radiometry, KRISS used hexagonal boron nitride (hBN) to compare the brightness range of 1 to 2 Mcps SPAD detectors(unit:CPS) with conventional detectors(unit:W), measuring detection ranges down to 1 to 1 pW levels. This allowed for a comparative analysis of the units of conventional and SPAD detectors.

For the quantum wireless communication field, GaN

*hongi2011@kriss.re.kr
†heejin.lim@kriss.re.kr
‡wookjaelee@kongju.ac.kr
§jinkyuyang@kongju.ac.kr

was utilized due to its superior long-term stability among the three materials studied. However, since the count rate was relatively low at 200 kCPS, a bulls-eye grating was employed to enhance the brightness.

Lastly, in the quantum bio-sensing field, the same confocal setup was used to develop flow cytrometry for measuring the quantities of DNA and RNA.

3 Acknowledgements

This research was supported by the R and D convergence program of NST (National Research Council of Science and Technology) of the Republic of Korea (Grant No. CAP22051-100) and Institute of Information and Communications Technology Planning and Evaluation (IITP) through the Korean Government (MSIT) (Solid-State Quantum Memory,Grant No. 2022-0-00198)

References

[1] W. Tittel and G. Weihs *Photonic entanglement for fundamental tests and quantum communication* In *Quantum Info. Comput*, vol. 1, no. 2, p. 3–56, Aug. 2001

[2] S. Chen, Y.-A. Chen, T. Strassel, Z.-S. Yuan, B. Zhao, J. Schmiedmayer, and J.-W. Pan *Deterministic and storable single-photon source based on a quantum memory.* Phys. Rev. Lett., vol. 97, p. 173004, Oct 2006.

[3] R. C. Willson *Active cavity radiometer* Appl. Opt., vol. 12, no. 4, p. 810, Apr 1973.

[4] J. E. Martin, N. P. Fox, and P. J. Key *A cryogenic radiometer for absolute radiometric measurements* Metrologia, vol. 21, no. 3, p. 147, Jan 1985.

[5] J. Y. Cheung, C. J. Chunnilall, E. R. Woolliams, N. P. Fox, J. R. Mountford, J. Wang, and P. J. Thomas *The quantum candela: a re-definition of the standard units for optical radiation* J. Mod. Opt, vol. 54, no. 2-3, p. 373, 2007.

[6] J. C. Zwinkels, E. Ikonen, N. P. Fox, G. Ulm, and M. L. Rastello *Photometry, radiometry and 'the candela': evolution in the classical and quantum world* Metrologia, vol. 47, no. 5, p. R15, Aug 2010

[7] C. J. Chunnilall, I. P. Degiovanni, S. K¨uck, I. M¨uller, and A. G. Sinclair *Metrology of single-photon sources and detectors: a review* Optical Engineering, vol. 53, no. 8, p. 1, 2014

[8] A. M. Berhane, K.-Y. Jeong, Z. Bodrog, S. Fiedler, T. Schroder, N. V. Trivino, T. Palacios, A. Gali, M. Toth, D. Englund, and I. Aharonovich *Bright room-temperature single-photon emission from defects in gallium nitride* Adv. Mater., vol. 29, no. 12, p. 1605092, 2017.

Inspecting the efficacy of quantum error correction and the virtual purification in noisy quantum metrology

Hyukgun Kwon[1 2] Youngrong Lim[3] Liang Jiang[2] Hyunseok Jeong[4]
Seung-Woo Lee[1 *] Changhun Oh[5 †]

[1] *Center for Quantum Information, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea*
[2] *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*
[3] *School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea*
[4] *Department of Physics and Astronomy, Seoul National University, Seoul 08826, Republic of Korea*
[5] *Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea*

Abstract. Quantum metrology offers a way of more accurate and precise estimation surpassing the capabilities of classical methods. However, noise often undermines the enhanced estimation performance in practical situations. Quantum error correction (QEC) - a method for correcting logical errors in quantum information processing - can be applied for quantum metrology to restore estimation accuracy from the noise. It has been shown that QEC cannot correct the noise parallel to the signal that we want to estimate, while it works well against the noise perpendicular to the signal. In this study, we apply the virtual purification method - an error mitigation approach to reduce the parallel noise and yields more accurate estimations than those obtained through QEC-applied metrology.

Keywords: Quantum Error Mitigation, Quantum Metrology, Quantum Error Correction

1 Noiseless metrology

Let us consider the canonical phase estimation where an unknown signal ϕ is embedded by the unitary operation

$$\hat{U}(\phi) = \exp\left(-i\frac{\phi}{2}\sum_{j=1}^{N}\hat{Z}^{(j)}\right) = \exp\left(-i\frac{\phi}{2}\hat{H}\right), \quad (1)$$

where $\hat{Z}^{(j)}$ is the Pauli Z operator acting on jth qubit and $\hat{H} \equiv \sum_{i=1}^{N}\hat{Z}_i$ is the signal Hamiltonian. To estimate ϕ, we consider N-qubit GHZ state as a quantum probe $|\psi_0\rangle = |0\rangle^{\otimes N} + |1\rangle^{\otimes N}$. The corresponding signal state is then

$$|\psi(\phi)\rangle = e^{-i\frac{N\phi}{2}}|0\rangle^{\otimes N} + e^{i\frac{N\phi}{2}}|1\rangle^{\otimes N}. \quad (2)$$

We emphasize that the signal state lies in the 2-dimensional Hilbert space whose basis vectors are $|0\rangle^{\otimes N}$ and $|1\rangle^{\otimes N}$. We denote the corresponding Hilbert space as $\mathcal{H}$. It also has been studied that among all possible combinations of the quantum probes and the measurements, the smallest estimation error can be achieved by measuring the signal state $|\psi(\phi)\rangle$ in the eigenbasis of $\hat{A} = \prod_{j=1}^{N}\hat{Y}^{(j)}$, where $\hat{Y}^{(j)}$ is the Pauli Y acting on ith qubit. In this study, we mainly focus on this scenario.

2 Independent Identically Distributed Dephasing error

In practical situations, the presence of noise deteriorates an estimation performance. Especially, when one cannot obtain complete information about the noise, a bias occurs during the estimation, that cannot be mitigated merely by augmenting the sample size, unlike the statistical error. Notably, as the sample size increases, the bias tends to outweigh the statistical error and becomes the predominant source of estimation error. In this study, we inspect the efficacy of the stabilizer-based QEC and the virtual purification method in terms of reducing the bias that occurs from the lack of information about the noise. We consider the independent identically distributed dephasing noise (IIDD) as a dominant noise during the estimation, which can be described as

$$\rho \;\rightarrow\; \mathcal{E}(\hat{\rho}) = \mathcal{E}^{(1)} \circ \mathcal{E}^{(2)} \circ \cdots \circ \mathcal{E}^{(N)}(\hat{\rho}), \quad (3)$$

where, $\mathcal{E}^{(i)}$ is the local Pauli noise which is defined as

$$\mathcal{E}^{(i)}(\hat{\rho}) = p_I\hat{\rho} + p_z\hat{Z}^{(i)}\hat{\rho}\hat{Z}^{(i)}. \quad (4)$$

Here $p_I + p_z = 1$. The IIDD is one of the representative noises, which is parallel to the signal, that cannot be corrected by the QEC [3]. We assume that the IIDD noise is characterized by its noise strength Δ where the probabilities p_I, p_z are functions of Δ. In addition, we consider the mild noise limit $p_I = O(1)$ and $p_z = O(\Delta)$ where Δ is assumed to be small. Under the IIDD noise, the signal state $|\psi(\phi)\rangle\langle\psi(\phi)|$ becomes the error state

$$\hat{\rho}_{\mathrm{e}}(\phi) \equiv \mathcal{E}(|\psi(\phi))\rangle\langle\psi(\phi))|) \quad (5)$$

$$= \left(\frac{1+(p_I-p_z)^N}{2}\right)|\psi(\phi)\rangle\langle\psi(\phi)| \quad (6)$$

$$+ \left(\frac{1-(p_I-p_z)^N}{2}\right)\left|\psi(\phi-\frac{\pi}{2N})\right\rangle\left\langle\psi(\phi-\frac{\pi}{2N})\right|. \quad (7)$$

Since IIDP noise commutes with the signal unitary $\hat{U}(\phi)$, our analysis includes IIDP noise occurring either before or after the signal unitary process. We emphasize that

*swleego.kist.re.kr
†changhun0218@gmail.com

the dominant eigenvector of $\hat{\rho}_{\rm e}$ is the ideal signal state $|\psi(\phi)\rangle$, and $\left|\psi(\phi-\frac{\pi}{2N})\right\rangle$ is orthogonal to $|\psi(\phi)\rangle$. When one does not apply either the QEC and the virtual purification, the corresponding bias, which we denote as $B_{\rm e}$, is [1, 2]

$$B_{\rm e} \propto {\rm Tr}\Big[\hat{\rho}_{\rm e}(\phi)\hat{A}\Big] - \langle\psi(\phi)|\hat{A}|\psi(\phi)\rangle \tag{8}$$

$$= -2Np_z\langle\psi(\phi)|\hat{A}|\psi(\phi)\rangle + O(\Delta^2) = O(\Delta). \tag{9}$$

3 Stabilizer formalism based QEC applied case

To apply the stabilizer QEC to reduce the bias, let us encode the Hilbert space $\mathcal{H}$, where the signal state lies, into the larger Hilbert space. To construct the larger Hilbert space, we assume that one can prepare the noiseless $N_{\rm A}$-qubit ancilla system and the parameter ϕ cannot be embedded in the ancilla mode, as a consequence, the signal unitary acting on the $(N+N_{\rm A})$-qubit system, is described as

$$\hat{U}_{\rm L}(\phi) = \exp\left(-i\frac{\phi}{2}(\hat{H}\otimes\hat{I}_{\rm A})\right) = \exp\left(-i\frac{\phi}{2}\hat{H}_{\rm L}\right), \tag{10}$$

where $\hat{I}_{\rm A}$ is the identity operator defined in the ancilla mode and $\hat{H}_{\rm L} \equiv \hat{H}\otimes\hat{I}_{\rm A}$. Next, let us consider a code space, which we denote as $\mathcal{C}(\mathbf{s})$, defined with the stabilizer generator set

$$\mathbf{s} = \{\hat{s}_1, \hat{s}_2, \cdots \hat{s}_{N+N_{\rm A}-1}\}, \tag{11}$$

where $\hat{s}_i$'s are the Pauli operators acting on $(N+N_{\rm A})$-qubit system. First, let us inspect the codewords of the code space $\mathcal{C}(\mathbf{s})$, which we denote as

$$|0\rangle^{\otimes N} \to |0\rangle_{\rm L} \equiv |c_0\rangle\otimes|\varphi_0\rangle_{\rm A}\,, \tag{12}$$

$$|1\rangle^{\otimes N} \to |1\rangle_{\rm L} \equiv |c_1\rangle\otimes|\varphi_1\rangle_{\rm A}\,. \tag{13}$$

The original quantum probe $|\psi_0\rangle$ is encoded in the logical quantum probe which we denote as $|\psi_{\rm L0}\rangle$

$$|\psi_0\rangle \to |\psi_{\rm L0}\rangle = |0\rangle_{\rm L} + |1\rangle_{\rm L}\,, \tag{14}$$

where the corresponding logical signal state is

$$|\psi_{\rm L}(\phi)\rangle = \hat{U}_{\rm L}(\phi)\,|\psi_{\rm L0}\rangle \tag{15}$$

$$= \left(e^{-i\frac{\phi}{2}\hat{H}}\,|c_0\rangle\right)\otimes|\varphi_0\rangle_{\rm A} + \left(e^{-i\frac{\phi}{2}\hat{H}}\,|c_1\rangle\right)\otimes|\varphi_1\rangle_{\rm A} \tag{16}$$

$$= e^{-i\frac{N\phi}{2}}\,|0\rangle_{\rm L} + e^{i\frac{N\phi}{2}}\,|1\rangle_{\rm L}\,. \tag{17}$$

As a result, the codewords should be $|0\rangle_{\rm L} = |0\rangle^{\otimes N}\otimes|\varphi_0\rangle_{\rm A}$ and $|1\rangle_{\rm L} = |1\rangle^{\otimes N}\otimes|\varphi_1\rangle_{\rm A}$. Next, let us inspect the relation between the stabilizer generators and the signal Hamiltonian. Since the logical signal state lies on the code space, the following equation should be satisfied:

$$\hat{s}_i\,|\psi_{\rm L}(\phi)\rangle = |\psi_{\rm L}(\phi)\rangle \longrightarrow \hat{U}_{\rm L}^{\dagger}(\phi)\hat{s}_i\hat{U}_{\rm L}(\phi)\,|\psi_{\rm L0}\rangle = |\psi_{\rm L0}\rangle\,, \tag{18}$$

i.e., all the stabilizer generators should commute with $\hat{H}_{\rm L}$. Equivalently, one can easily show that to commute with $\hat{H}_{\rm L}$, all the stabilizer generators should commute with $\hat{Z}^{(j)}\otimes\hat{I}_{\rm A}$ for all $j = 1, 2, \cdots, N$. Therefore, all the Pauli Z errors cannot be corrected by the QEC scheme since all the stabilizers commute with all the Pauli Z operators. As a result, QEC cannot reduce the bias occurring from the dephasing noise, which results in the same bias as the error case:

$$B_{\rm QEC} = B_{\rm e} \propto {\rm Tr}\Big[\hat{\rho}_{\rm e}(\phi)\hat{A}\Big] - \langle\psi(\phi)|\hat{A}|\psi(\phi)\rangle \tag{19}$$

$$= -2Np_z\langle\psi(\phi)|\hat{A}|\psi(\phi)\rangle + O(\Delta^2) = O(\Delta). \tag{20}$$

4 Virtual Purification

Instead of directly attaining the purified error state $\frac{\hat{\rho}_{\rm e}^n}{{\rm Tr}[\hat{\rho}_{\rm e}^n]}$, (here $\hat{\rho}_{\rm e}^n$ is n squares of the error state), the virtual purification allows one to obtain the expectation value of an observable $\hat{A}$ over the purified state $\frac{{\rm Tr}[\hat{A}\hat{\rho}_{\rm e}^n]}{{\rm Tr}[\hat{\rho}_{\rm e}^n]}$, without any prior knowledge about the noise. When the virtual distillation is applied to the quantum metrology, the bias becomes

$$B_{\rm VP} \propto {\rm Tr}\left[\frac{\hat{\rho}_{\rm e}^n}{{\rm Tr}[\hat{\rho}_{\rm e}^n]}\hat{A}\right] - \langle\psi(\phi)|\hat{A}|\psi(\phi)\rangle \tag{21}$$

$$= -2N(p_z)^n\langle\psi(\phi)|\hat{A}|\psi(\phi)\rangle + O(\Delta^{n+1}) = O(\Delta^n), \tag{22}$$

where n is the mitigation order of the virtual purification [1, 2]. By comparing Eqs. (20) and (22), one can find that the virtual purification there is a noisy estimation scheme that the virtual purification outperforms the stabilizer-based QEC, in terms of reducing the bias.

References

[1] H. Kwon *et al.*, Efficacy of virtual purification-based error mitigation on quantum metrology Phys. Rev. A **109**, 022410 (2024).

[2] K. Yamamoto *et al.*, Error-mitigated quantum metrology via virtual purification, Phys. Rev. Lett. **129**, 250503 (2022).

[3] S. Zhou *et al.*, Achieving the Heisenberg limit in quantum metrology using quantum error correction, Nat. Commun. **9**, 78 (2018).

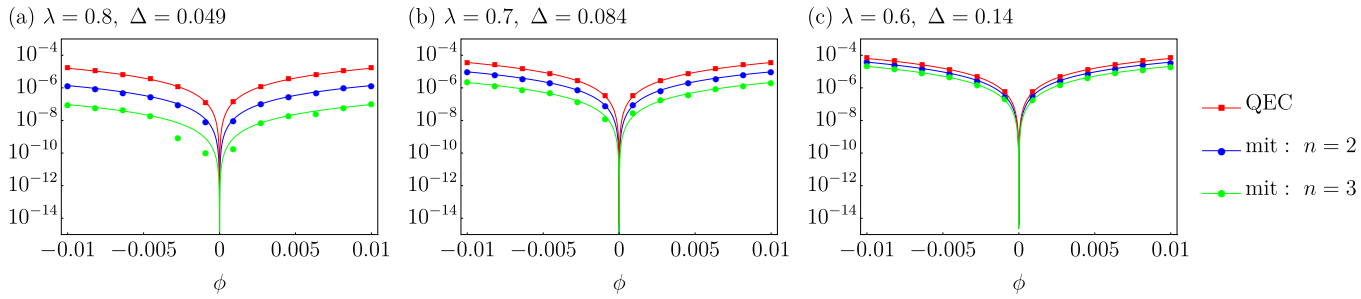


Figure 1: (a)-(c) Simulations of bias (with log scale) exploiting GHZ state ($N = 5$) as a quantum probe in the presence of dephasing noise with different noise strengths. We use $N_s = 10^9$ numbers of samples. The lines are theoretical values of the bias errors and the markers are the simulated values.

Generic Bell inequalities with many local measurements

Junghee Ryu[1] *

[1] *Division of National Supercomputing,*
Korea Institute of Science and Technology Information, Daejeon, 34141, Korea

Abstract. Violations of Bell inequalities imply that local realistic theories cannot predict the correlations produced by quantum theory. Since Bell's original discovery for the two-particle systems, there have been many studies on generalizing Bell inequalities to more complex systems, such as multi-particle systems. We here present Bell inequalities involving many measurements and show that the quantum violation can be calculated using a Greenberger-Horne-Zeilinger entangled state. We introduce a geometric method to calculate the violations of our Bell inequalities.

Keywords: Bell's theorem, Bell inequalities, Greenberger-Horne-Zeilinger state

1 Introduction

Bell derived a constraint on correlations for two remote systems that local hidden variable theories must obey, and he showed that the constraint can be violated by quantum mechanics in case of two coupled spin-1/2 particles and suitable local measurements [1]. This is known as Bell's theorem. Since Bell's original discovery, there have been many theoretical and experimental efforts to verify Bell's theorem [2, 3, 4, 5]. Nowadays, it is known that violations of Bell inequalities are essential conditions for various quantum information protocols to beat their classical counterparts, for example, quantum random number generation, quantum cryptography, reducing communication complexity, and so on [6]. Therefore, considerable efforts have been devoted to studying Bell inequalities theoretically and experimentally [7].

We here investigate Bell's theorem involving many observables by suggesting the generic Bell inequalities for $(3, M, D)$ systems. A Bell inequality is said to be generic in the sense that it is directly connected with the Greenberger-Horne-Zeilinger (GHZ) theorem. We show that the maximal quantum expectations of our generic Bell operators can be achieved by the GHZ entangled state. In order to increase the number of local measurements, we deploy the quantum Fourier transformation and the phase shift operation. To calculate the upper bound of local realistic descriptions for the Bell inequalities, we introduce a geometric approach. We shall show the violations of the generic Bell inequalities involving three ($M = 3$) and four ($M = 4$) measurement settings.

2 Results

We suggest a generic Bell operator for the tripartite D-dimensional system involving three measurement settings, which reads

$$\hat{\mathcal{B}}_3 = \frac{1}{3^3} \sum_{n=1}^{D-1} \sum_{\gamma=0}^{2} \bigotimes_{j=1}^{3} \sum_{\eta_j=0}^{2} \Omega^{\gamma\eta_j} \omega^{n\eta_j/3} \hat{X}_j^n(\eta_j/3), \quad (1)$$

where $\Omega = \exp(2\pi i/3)$ and $\omega = \exp(2\pi i/D)$. Note that nth powers operator $\hat{X}_j^n(\eta_j/3)$ is the η_jth observable of jth party and it reads

$$\hat{X}(\nu) = \omega^{-\nu} \left(\sum_{n=0}^{D-2} |n+1\rangle \langle n| + \omega^{\nu D} |0\rangle \langle D-1| \right). \quad (2)$$

The upper bound of the function $\hat{\mathcal{B}}_3$ can be achieved by the generalized GHZ state $|\psi\rangle = \frac{1}{\sqrt{D}} \sum_{n=0}^{D-1} |n, n, n\rangle$ as

$$\langle\psi| \hat{\mathcal{B}}_3 |\psi\rangle = D - 1. \quad (3)$$

It is because the GHZ state corresponds to the eigenstate of the composite observables $\bigotimes_{j=1}^{3} \hat{X}_j^n(\eta_j/3)$ in Eq. (1), see also Ref. [8].

Local realistic theories assume that the measurement outcomes are predetermined before the actual measurements and any physical influences on one side propagate at most at the speed of light. By definition, the outcome of measurement $X_j(\eta_j/3)$ is predetermined as its eigenvalue $\omega^{\alpha(j,\eta_j)}$, where $\alpha(j, \eta_j)$ is integer. As a result, the Bell function based on LHVs is given by

$$\mathcal{B}_{3\text{LHV}} = \frac{1}{3^2} \sum_{\vec{\eta}}^{2} \delta_3(\tilde{\eta}) \delta_D(\tilde{\eta}/3 + \tilde{\alpha}) D - 1, \quad (4)$$

with $\tilde{\eta} = \sum_{j=1}^{3} \eta_j$ and $\tilde{\alpha} = \sum_{j=1}^{3} \alpha(j, \eta_j)$. Here $\delta_D(\alpha) = 1$ if $\alpha \equiv 0 \mod D$ and $\delta_D(\alpha) = 0$ otherwise. We suggest a geometrical approach to obtain the classical upper bound. To describe our idea, we first apply to the $(3, 2, D)$ system. The classical Bell function is given by

$$\mathcal{B}_{2\text{LHV}} = \frac{D}{4} [\delta_D(a_1 + a_2 + a_3) + \delta_D(a_1 + b_2 + b_3 + 1) + \delta_D(b_1 + a_2 + b_3 + 1) + \delta_D(b_1 + b_2 + a_3 + 1)] - 1. \quad (5)$$

It was shown in Ref. [9] that the classical upper bound, that is, the right hand side in Eq. (5) reads $3D/4 - 1$. We shall reproduce the result by using our geometrical approach.

Consider a square with eight points depicted in Fig. 1(a). Each point indicates the integer a_j(or b_j) and each line implies the four delta functions in Eq. (5), respectively. The variables by dotted lines are assigned to the same values. Now, let us assign the integers to satisfy the functions as $\delta_D(\mathbf{a}) = 1$ in a clockwise direction.

*junghee@kisti.re.kr

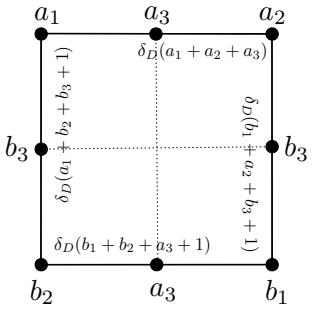


Figure 1: Geometrical method to calculate the classical upper bound of two measurements settings. The four delta functions of the generic Bell inequalities in Eq. (5) are represented by the solid lines and each point denotes the integer that is related to the predetermined value by the local realistic description.

1. The variables a_1 and a_3 are freely chosen, but $a_2 = -a_1 - a_3$ to satisfy the $\delta_D(a_1 + a_2 + a_3) = 1$.
2. The value of b_1 is given by $a_1 + a_3 - b_3 - 1$ for the $\delta_D(b_1 + a_2 + b_3 + 1) = 1$.
3. We must assign the value of $-2a_3 - a_1 + b_3$ to the b_2 to satisfy the $\delta_D(b_1 + b_2 + a_3 + 1) = 1$.
4. Finally, we get the following delta function: $\delta_D(-2a_3 + 2b_3 + 1)$.

The last delta function cannot be a unity for even D-dimensional system. The linear congruence $2(-a_3+b_3)+1 \equiv 0 \mod D$ has no solution for even D because $g = \gcd(2, \text{even } D) = 2$ and $2 \nmid -1$, where the notation $a \nmid b$ means that a does not divide b. As a result, the last delta function becomes zero, i.e., $\delta_D(-2a_3+2b_3+1) = 0$. As a result, the maximal number of the delta functions to be unity is 3, that is, we have $\mathcal{B}_{2\text{LHV}} \leq 3D/4 - 1$.

Figure 2 represents a hexagon that is employed to solve the $(3, 3, D)$ case, where the nine delta functions can be obtained. Similarly to the two observables case, each point indicates the variables in the delta functions and the variables linked by dotted line have the same values. The nine solid lines are represented to all delta functions. Following the same argument as for two observables case, we can obtain a loop consisting of the six delta functions respectively. It turns out that the linear congruence obtained from the loop has no solution for $D = 3d$, where d is integer, and therefore the only five delta functions can be unity. Following to the further calculations, we finally obtain the classical upper bound of the general Bell inequality for $(3, 3, D)$ system as $\mathcal{B}_{3\text{LHV}} \leq 7D/9 - 1$, which is contradiction to the quantum upper bound $(D - 1)$.

Note that a loop is significant in our method in the sense that we assign the values in consecutive order to satisfy each delta function, and consequently by the loop we obtain the local realistic constraint. In this poster presentation, we will describe the calculations for three and four measurements cases in detail.

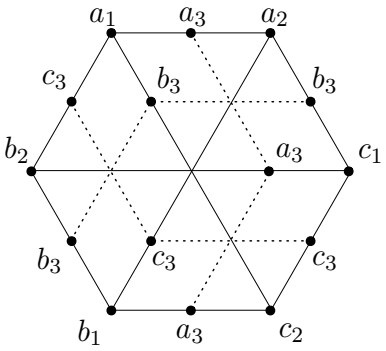


Figure 2: Three measurements case. The hexagon is employed to represent the nine delta functions.

References

[1] J. S. Bell, Physics **1**, 195 (1964).

[2] B. Hensen *et al.*, Nature **526**, 682 (2015).

[3] M. Giustina *et al.*, Phys. Rev. Lett. **115**, 250401 (2015).

[4] L. K. Shalm, *et al.*, Phys. Rev. Lett. **115**, 250402 (2015).

[5] W. Rosenfeld, *et al.*, Phys. Rev. Lett. **119**, 010402 (2017).

[6] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).

[7] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

[8] J. Ryu, C. Lee, Z. Yin, R. Rahaman, D. G. Angelakis, J. Lee, and M. Żukowski, Phys. Rev. A **89**, 024103 (2014).

[9] W. Son, J. Lee and M. S. Kim, Phys. Rev. Lett. **96**, 060406 (2006).

An Efficient Quantum Circuit Construction Method for Mutually Unbiased Bases in n-Qubit Systems

Yu Wang[1] * Dongsheng Wu[1 2] †

[1] *Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing 101408, P. R. China*
[2] *Yau Mathematical Sciences Center, Tsinghua University, Beijing, 100084, China*

Abstract. Mutually unbiased bases (MUBs) are commonly viewed as maximal incompatibility and complementarity in quantum information theory, which contributes to various applications like quantum state tomography, error correction, entanglement detection, and quantum cryptography. The quantum Fourier transformation circuit (or $H^{\otimes n}$) produces a basis mutually unbiased with the computational one but can not directly generalize to a complete set of 2^n+1 MUB circuits. Based on a set of 2^n+1 MUB formular given by the Galois-Fourier method, we design an efficient algorithm to generate each of 2^n+1 quantum MUBs circuits on n-qubit systems within $O(n^3)$ time. The 2^n nontrivial circuits (excluding $I^{\otimes n}$) consist of a maximum of $(n^2+7n)/2$ H, S, and CZ gates, structured as $-H-S-CZ-$. Alternatively, they can be implemented using $H^{\otimes n}$ and a diagonal operation. On average, the count of S gates, CZ gates, and CZ gates with distance u amounts to $3n/2$, $(n^2-n)/4$, and $(n-u)/2$, respectively. Moreover, we have observed that the entanglement segment comprises $2n-3$ fixed modules, and the 2^n circuits satisfy some intriguing "linear" relations. Precisely, the knowledge of n special MUB circuits is enough to construct all 2^n+1 MUB circuits. The strength of this new construction lies in its efficiency and simplicity, paving the way for implementing a complete set of MUBs in diverse quantum information processing tasks on high-dimensional quantum systems.

Keywords: Quantum circuits, mutually unbiased bases, quantum tomography

1 Introduction

Quantum measurement is the exclusive method for obtaining information about quantum systems, forming a crucial link for understanding microscopic quantum states through empirical observations [1]. Projective measurements onto mutually unbiased bases (MUBs) [2] are widely utilized in quantum information science. Preparing an eigenstate of one basis, its distribution is uniform across any other MUB, highlighting their maximal incompatibility and complementarity [3, 4, 5]. MUBs are useful in quantum tomography [6, 7, 8, 9], uncertainty relations [10, 11, 12, 13], quantum cryptography [14, 15, 16, 17, 18], quantum error correction [19, 20, 21], and entanglement identification [22, 23, 24, 25, 26], to name a few.

Two MUBs can always be constructed in any finite-dimensional Hilbert space [2]. However, the maximum number of MUBs is limited to $d+1$, which remains an open question in quantum information theory [27]. When d is prim power, $d+1$ MUBs can be constructed [7]. For dimension $d=6$, strong numerical evidence indicates that there are no four MUBs [28, 29, 30, 31]. Some research focuses on the structure behind complete $(d+1)$ MUBs sets and incomplete sets [32, 33, 34].

To measure the state ρ using a projective measurement onto one MUB $\{U_j|k\rangle : k=0,\cdots,d-1\}$, we can apply the unitary $U_j^\dagger$ to ρ and subsequently measure in the computational basis. We aim to efficiently implement 2^n+1 MUB circuits in n-qubit systems, starting with circuits involving two MUBs. Even two MUBs usually work in a lot of quantum information tasks, the 2^n+1 MUB circuits, together with the computational measurement, are essential as minimal and optimal resources for reconstructing all unknown n-qubit states [6, 7]. Additionally, while many d-dimensional Quantum Key Distribution (QKD) protocols, such as the BB84 protocol, use only two MUBs [35], employing $d+1$ MUBs enhances QKD robustness, particularly against correlated errors [36, 37].

Let the first MUB circuit be $I^{\otimes n}$. The second MUB circuit could be $H^{\otimes n}$ or a Fourier transformation circuit requiring $O(n^2)$ gates [38]. These circuits are integral to numerous prominent quantum algorithms, including the Deutsch-Jozsa algorithm [39, 40, 41], Shor's factorization algorithm [42], Grover's search algorithm [43], and the HHL algorithm [44], among others. However, these two MUB circuits alone cannot generate the complete set of 2^n+1 MUB circuits directly. Previous works constructed a new second MUB circuit V [45, 46, 47, 48, 49]. It's interesting that repeating V with two times, three times until 2^n times, the complete set of 2^n+1 MUB circuit can be obtained. However, the gate number for some circuits could be exponential.

In this work, we introduce a numerical method for identifying both complete and incomplete MUBs using complex Hadamard matrices and diagonal matrices. We chose the 2^n MUBs formula obtained by the Galois-Fourier approach [50] to generate the complete MUBs circuits. Each nontrivial MUB circuit is constructed with the $H^{\otimes n}$ and a diagonal operation, structured as $-H-S-CZ-$. We propose an efficient computational method to decompose each MUB circuit using $O(n^2)$ gates within $O(n^3)$ time. An interesting entanglement structure is presented. We find a linear relation where the knowledge of n special MUB circuits describes all

*ming-jing-happy@163.com
†wudongsheng14@mails.ucas.ac.cn

$2^n + 1$ MUB circuits. We calculate the average occurrence of various gate types and analyze the distribution of MUB state coefficients. Finally, we suggest several avenues for further exploration and discussion. The circuit construction method holds the potential to enhance the utilization of MUBs in the realms of quantum information and quantum computing tasks in the future. And the method could offer deeper insights into MUBs' structural properties.

2 Preliminaries and a Numerical Method Conjecture

Definition 1 (MUB). *A set of two normalized eigenbases $\{|\psi_j\rangle\}_{j=0}^{d-1}$ and $\{|\phi_k\rangle\}_{k=0}^{d-1}$ are called mutually unbiased (MU) if the following condition holds for each j, k:*

$$|\langle\psi_j|\phi_k\rangle|^2 = \frac{1}{d} \tag{1}$$

Given a set of M eigenbases labeled as $\{\{|\psi_j^k\rangle\}_{j=0}^{d-1} : k = 0, \cdots, M-1\}$, if any two bases within this set are MU, then the set is said to contain M MUBs. For prime power dimensions d, a set containing maximum $d+1$ MUBs can always be found.

Definition 2 (Complex Hadamard matrix). *Given a unitary matrix U, it is called a complex Hadamard matrix if each matrix element U_{jk} satisfies the following condition:*

$$|U_{jk}|^2 = \frac{1}{d} \tag{2}$$

Any complex Hadamard matrix can produce the second basis mutually unbiased with computation one $\{|k\rangle\}_{k=0}^{d-1}$.

Fix a basis $\{|\psi_j^0\rangle\}_{j=0}^{d-1}$ from a set of MUBs, it corresponds to a unitary operation $I = \sum_{k=0}^{d-1} |\psi_j^0\rangle\langle\psi_j^0|$. For any other MUB $\{|\psi_j^k\rangle\}_{j=0}^{d-1}$, it corresponds to unitary operation $U_k = \sum_{k=0}^{d-1} |\psi_j^k\rangle\langle\psi_j^0|$.

Corollary 1. *Finding a set of M MUBs is equivalent to finding $M-1$ unitary operations $\{U_0 = I, U_1, \cdots, U_{M-1}\}$ such that $U_j^\dagger U_k$ is a complex Hadamard matrix for each different $j, k = 0, \cdots, M-1$.*

The Corollary yields a numerical method to construct a set of MUBs [32, 33, 34], illustrated as Fig.(1). We may as well let the first row of the unitary operations be real number for the freedom choice of global phase. If we have infinite computational resources, the process in Fig.(1) will produce all the set of MUBs.

Similar to the numerical method to construct symmetric informationally complete measurement (SIC-POVM) [51], the problem is to find d^2 unit complex vectors $\{|\phi_j\rangle\}_{j=1}^{d^2}$ such that the following condition is satisfied for different j, k

$$|\langle\phi_j|\phi_k\rangle|^2 = 1/(d+1) \tag{3}$$

The existence problems of $d+1$ MUBs and SIC-POVM with d^2 elements are identified as two open questions in quantum information theory [27]. Zauner's conjecture [52, 53] simplify the computation process in Eq.(3) by finding one fiducial state $|\phi_0\rangle$. Define $X = \sum_{k=0}^{d-1} |k+1\rangle\langle k|$, $Z = \sum_{k=0}^{d-1} e^{2\pi\sqrt{-1}k/d}|k\rangle\langle k|$. If we can find $|\phi_0\rangle$ such that

$$|\langle\phi_0|X_j Z_k|\phi_0\rangle|^2 = 1/(d+1) \tag{4}$$

for $j, k = 0, \cdots, d-1$, the SIC-POVM is then constructed. Recently, a necessary condition for the existence of fiducial state is given [54].

We think the effort to find MUBs in Corollary 1 can be modified as follows.

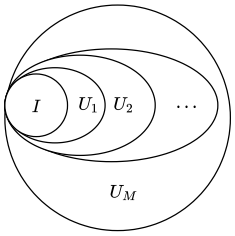


Figure 1: Method 1: First, we choose a complex Hadamard matrix U_1. Then we find another Hadamard matrix U_2 such that $U_1^\dagger U_2$ is still a complex Hadamard matrix. We continue this process until we find the final matrix U_M.

Method 1. *In order to construct the unitary operations $\{U_0 = I, U_1, \cdots, U_{M-1}\}$ in Corollary 1, we can find a complex Hadamard matrix U_1 and search $M-2$ diagonal matrices D_k, where $k = 1, \cdots, d-2$. The diagonal element is chosen from*

$$\{e^{\pi\sqrt{-1}/d}, e^{2\pi\sqrt{-1}/d}, \cdots, e^{(2d-1)\pi\sqrt{-1}/d}, 1\}$$

If the following condition is held,

$$U_1^\dagger D_j^\dagger D_k U_1 \text{ for different } j, k \tag{5}$$

The operations $\{U_0 = I, U_1, D_1U_1, \cdots, D_{M-2}U_1\}$ can generate M MUBs.

This method can avoid the knowledge of the mathematical theory of finite rings and field to construct MUBs. For n-qubit case, $d = 2^n$, we let the diagonal elements be $\{\pm 1, \pm\sqrt{-1}\}$ and let $U_1 = H^{\otimes n}$, the numerical experiment shows that we can always construct the complete set of MUBs using Method 1. Besides, the solutions are not unique. For example, the four nontrivial MUB circuits for $n = 3$ can be the following after representing the diagonal matrix with S gates and CZ gates.

However, it seems like things are heading towards two extremes. The results [48, 49] cost polynomial computations with some circuits decomposed of exponential gates. Using method 1 costs exponential computations with polynomial decomposed gates for small n.

To construct each MUB circuit within polynomial time and using a polynomial number of gates, we turn to the

formulas for 2^n nontrivial MUBs. We find the formula by Wootters and Fields [7], the Galois Rings formula [55], the Galois-Fourier formula [50], or the method involving the division of 4^n-1 Pauli observables [56, 57]. The Galois-Fourier formula [50] directly meets our requirements. And we find some structures for these MUB circuits.

3 Results

Result 1. *New decomposition method and circuit structure. The circuit for the computational basis is $I^{\otimes n}$. For $j=0,\ldots,2^n-1$, every nontrivial $U(j)$ can be determined explicitly and decomposed into the following circuit sequence: $-H-S-CZ-$.*

Analysis. According to [50, Eq.(2.70)], the j-th MUBs consists of elements

$$|e_k^j\rangle = \frac{1}{\sqrt{2^n}} \sum_{l=0}^{2^n-1} |l\rangle (-1)^{k\odot l} \cdot \alpha_l^j$$

We made a permutation to obtain states

$$|f_k^j\rangle = \frac{1}{\sqrt{2^n}} \sum_{l=0}^{2^n-1} |l\rangle (-1)^{k\cdot l^T} \cdot \alpha_l^j.$$

The parameter α_l^j are recalculated

$$\alpha_l^j = \prod_{r=0}^{n-1} \left(\sqrt{-1}\right)^{a_r(j) l_r} \cdot \prod_{0\leqslant s<t\leqslant n-1} (-1)^{b_{s,t}(j) l_s l_t} \tag{6}$$

Then

$$U(j) = U_{CZ}(j) \cdot U_S(j) \cdot H^{\otimes n}, \tag{7}$$

where

$$H^{\otimes n} = \underbrace{H \otimes \cdots \otimes H}_{n \text{ times}}, \quad U_S(j) = S^{a_0(j)} \otimes \cdots \otimes S^{a_{n-1}(j)}$$

and

$$U_{CZ}(j) = \prod_{0\leq s<t\leq n-1} CZ(s,t)^{b_{s,t}(j)}, \tag{8}$$

with $CZ(s,t)$ being the n-qubit CZ operation with q_s as the control qubit and q_t as the target qubit.

Result 2. *Entanglement structure. If gate $CZ(s,t)$ appears at circuit $U(j)$, then gate $CZ(s',t')$ should also appear with $s+t=s'+t'$. Thus the entanglement parts are divided into $2n-3$ modules.*

Result 3. *Time efficiency. Given an input qubit number n and a random $j\in\{0,1,\ldots,2^n-1\}$, the time complexity to generate the circuit $U(j)$ is $O(n^3)$.*

Result 4. *Linear property. There are "linear" relations between the 2^n MUBs circuits $U(j)$. Specifically, given the knowledge of n circuits $U(2^0), U(2^1), \cdots, U(2^{n-1})$, any circuit $U(j)$ can be deduced, $0\leq j\leq 2^n-1$.*

Result 5. *Gates efficiency and average count. The maximal number of gates in an arbitrary circuit $U(j)$ is $(n^2+7n)/2$. The average number of S gates is $3n/2$, the average number of CZ gates is $(n^2-n)/4$, and the number of CZ gates of distance u (represented as $CZ(s,t)$ with $t-s=u$) is $(n-u)/2$.*

Result 6. *Balanced parameter distributions.*

For these 4^n nontrivial MUB states, each component belongs to $\{\pm1,\pm\sqrt{-1}\}/\sqrt{2^n}$ and these coefficients are evenly distributed.

When we look at all the coefficients at all rows of 2^n nontrivial MUBs, the elements in the first row are always $1/\sqrt{2^n}$.

Given any U_j we look at the coefficients in the column. They are evenly distributed from $\{\pm1\}/\sqrt{2^n}$ or from $\{\pm\sqrt{-1}\}/\sqrt{2^n}$.

References

[1] Vladimir B Braginsky and Farid Ya Khalili. *Quantum measurement*. Cambridge University Press, 1995.

[2] Julian Schwinger. Unitary operator bases. *Proceedings of the National Academy of Sciences*, 46(4):570–579, 1960.

[3] Niels Bohr et al. *The quantum postulate and the recent development of atomic theory*, volume 3. Printed in Great Britain by R. & R. Clarke, Limited, 1928.

[4] Lorenzo Maccone, Dagmar Bruß, and Chiara Macchiavello. Complementarity and correlations. *Physical review letters*, 114(13):130401, 2015.

[5] Sébastien Designolle, Paul Skrzypczyk, Florian Fröwis, and Nicolas Brunner. Quantifying measurement incompatibility of mutually unbiased bases. *Physical review letters*, 122(5):050402, 2019.

[6] ID Ivonovic. Geometrical description of quantal state determination. *Journal of Physics A: Mathematical and General*, 14(12):3241, 1981.

[7] William K Wootters and Brian D Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191(2):363–381, 1989.

[8] RBA Adamson and Aephraim M Steinberg. Improving quantum state estimation with mutually unbiased bases. *Physical review letters*, 105(3):030406, 2010.

[9] Gustavo Lima, Leonardo Neves, R Guzmán, Esteban S Gómez, WAT Nogueira, Aldo Delgado, A Vargas, and Carlos Saavedra. Experimental quantum tomography of photonic qudits via mutually unbiased basis. *Optics Express*, 19(4):3542–3552, 2011.

[10] Hans Maassen and Jos BM Uffink. Generalized entropic uncertainty relations. *Physical review letters*, 60(12):1103, 1988.

[11] Manuel A Ballester and Stephanie Wehner. Entropic uncertainty relations and locking: Tight bounds for mutually unbiased bases. *Physical Review A*, 75(2):022319, 2007.

[12] Serge Massar and Philippe Spindel. Uncertainty relation for the discrete fourier transform. *Physical review letters*, 100(19):190401, 2008.

[13] Shengjun Wu, Sixia Yu, Klaus Mølmer, et al. Entropic uncertainty relation for mutually unbiased bases. *Physical Review A*, 79(2):022104, 2009.

[14] Nicolas J Cerf, Mohamed Bourennane, Anders Karlsson, and Nicolas Gisin. Security of quantum key distribution using d-level systems. *Physical review letters*, 88(12):127902, 2002.

[15] I-Ching Yu, Feng-Li Lin, and Ching-Yu Huang. Quantum secret sharing with multilevel mutually (un) biased bases. *Physical Review A*, 78(1):012344, 2008.

[16] Andrea Casaccino, Ernesto F Galvao, and Simone Severini. Extrema of discrete wigner functions and applications. *Physical Review A*, 78(2):022310, 2008.

[17] Mhlambululi Mafu, Angela Dudley, Sandeep Goyal, Daniel Giovannini, Melanie McLaren, Miles J Padgett, Thomas Konrad, Francesco Petruccione, Norbert Lütkenhaus, and Andrew Forbes. Higher-dimensional orbital-angular-momentum-based quantum key distribution with mutually unbiased bases. *Physical Review A*, 88(3):032305, 2013.

[18] Máté Farkas, Jędrzej Kaniewski, and Ashwin Nayak. Mutually unbiased measurements, hadamard matrices, and superdense coding. *IEEE Transactions on Information Theory*, 2023.

[19] A Robert Calderbank, Eric M Rains, Peter W Shor, and Neil JA Sloane. Quantum error correction and orthogonal geometry. *Physical Review Letters*, 78(3):405, 1997.

[20] A Robert Calderbank, Eric M Rains, Peter M Shor, and Neil JA Sloane. Quantum error correction via codes over gf (4). *IEEE Transactions on Information Theory*, 44(4):1369–1387, 1998.

[21] Daniel Gottesman. Fault-tolerant quantum computation with higher-dimensional systems. In *NASA International Conference on Quantum Computing and Quantum Communications*, pages 302–313. Springer, 1998.

[22] Christoph Spengler, Marcus Huber, Stephen Brierley, Theodor Adaktylos, and Beatrix C Hiesmayr. Entanglement detection via mutually unbiased bases. *Physical Review A*, 86(2):022311, 2012.

[23] D Giovannini, J Romero, Jonathan Leach, A Dudley, A Forbes, and Miles J Padgett. Characterization of high-dimensional entangled systems via mutually unbiased measurements. *Physical review letters*, 110(14):143601, 2013.

[24] Paul Erker, Mario Krenn, and Marcus Huber. Quantifying high dimensional entanglement with two mutually unbiased bases. *Quantum*, 1:22, 2017.

[25] Jkedrzej Kaniewski, Ivan Šupić, Jordi Tura, Flavio Baccari, Alexia Salavrakos, and Remigiusz Augusiak. Maximal nonlocality from maximal entanglement and mutually unbiased bases, and self-testing of two-qutrit quantum systems. *Quantum*, 3:198, 2019.

[26] Armin Tavakoli, Máté Farkas, Denis Rosset, Jean-Daniel Bancal, and Jedrzej Kaniewski. Mutually unbiased bases and symmetric informationally complete measurements in bell experiments. *Science advances*, 7(7):eabc3847, 2021.

[27] Paweł Horodecki, Łukasz Rudnicki, and Karol Życzkowski. Five open problems in quantum information theory. *PRX Quantum*, 3(1):010101, 2022.

[28] Paul Butterley and William Hall. Numerical evidence for the maximum number of mutually unbiased bases in dimension six. *Physics Letters A*, 369(1-2):5–8, 2007.

[29] Ingemar Bengtsson, Wojciech Bruzda, Åsa Ericsson, Jan-Åke Larsson, Wojciech Tadej, and Karol Życzkowski. Mutually unbiased bases and hadamard matrices of order six. *Journal of mathematical physics*, 48(5), 2007.

[30] Stephen Brierley and Stefan Weigert. Constructing mutually unbiased bases in dimension six. *Physical Review A*, 79(5):052316, 2009.

[31] Philippe Raynal, Xin Lü, and Berthold-Georg Englert. Mutually unbiased bases in six dimensions: The four most distant bases. *Physical Review A*, 83(6):062303, 2011.

[32] Prabha Mandayam, Somshubhro Bandyopadhyay, Markus Grassl, and William K Wootters. Unextendible mutually unbiased bases from pauli classes. *arXiv preprint arXiv:1302.3709*, 2013.

[33] Dardo Goyeneche. Mutually unbiased triplets from non-affine families of complex hadamard matrices in dimension 6. *Journal of Physics A: Mathematical and Theoretical*, 46(10):105301, 2013.

[34] Dardo Goyeneche and Santiago Gomez. Mutually unbiased bases with free parameters. *Physical Review A*, 92(6):062325, 2015.

[35] Charles H Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical computer science*, 560:7–11, 2014.

[36] Fumin Wang, Pei Zeng, Jiapeng Zhao, Boris Braverman, Yiyu Zhou, Mohammad Mirhosseini, Xiaoli Wang, Hong Gao, Fuli Li, Robert W Boyd, et al. High-dimensional quantum key distribution based on mutually partially unbiased bases. *Physical Review A*, 101(3):032340, 2020.

[37] Takuya Ikuta, Seiseki Akibue, Yuya Yonezu, Toshimori Honjo, Hiroki Takesue, and Kyo Inoue. Scalable implementation of (d+ 1) mutually unbiased bases for d-dimensional quantum key distribution. *Physical Review Research*, 4(4):L042007, 2022.

[38] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[39] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992.

[40] David Collins, KW Kim, and WC Holton. Deutsch-jozsa algorithm as a test of quantum computation. *Physical Review A*, 58(3):R1633, 1998.

[41] Daowen Qiu and Shenggen Zheng. Revisiting deutsch-jozsa algorithm. *Information and Computation*, 275:104605, 2020.

[42] Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.

[43] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.

[44] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical review letters*, 103(15):150502, 2009.

[45] Hoi Fung Chau. Unconditionally secure key distribution in higher dimensions by depolarization. *IEEE Transactions on Information Theory*, 51(4):1451–1468, 2005.

[46] Rod Gow. Generation of mutually unbiased bases as powers of a unitary matrix in 2-power dimensions. *arXiv preprint math/0703333*, 2007.

[47] Oliver Kern, Kedar S Ranade, and Ulrich Seyfarth. Complete sets of cyclic mutually unbiased bases in even prime-power dimensions. *Journal of Physics A: Mathematical and Theoretical*, 43(27):275305, 2010.

[48] Ulrich Seyfarth and Kedar S Ranade. Construction of mutually unbiased bases with cyclic symmetry for qubit systems. *Physical Review A*, 84(4):042327, 2011.

[49] U Seyfarth, LL Sanchez-Soto, and G Leuchs. Practical implementation of mutually unbiased bases using quantum circuits. *Physical Review A*, 91(3):032102, 2015.

[50] Thomas Durt, Berthold-Georg Englert, Ingemar Bengtsson, and Karol Życzkowski. On mutually unbiased bases. *International journal of quantum information*, 8(04):535–640, 2010.

[51] Joseph M Renes, Robin Blume-Kohout, Andrew J Scott, and Carlton M Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004.

[52] Gerhard Zauner. Grundzüge einer nichtkommutativen designtheorie. *Ph. D. dissertation, PhD thesis*, 1999.

[53] Andrew J Scott. Sics: Extending the list of solutions. *arXiv preprint arXiv:1703.03993*, 2017.

[54] Meng Cao, Tenghui Deng, and Yu Wang. Dynamical quantum state tomography with time-dependent channels. *Journal of Physics A: Mathematical and Theoretical*, 57(21):215301, 2024.

[55] Andreas Klappenecker and Martin Rötteler. Constructions of mutually unbiased bases. In *Finite Fields and Applications: 7th International Conference, Fq7, Toulouse, France, May 5-9, 2003. Revised Papers*, pages 137–144. Springer, 2004.

[56] Bandyopadhyay, Boykin, Roychowdhury, and Vatan. A new proof for the existence of mutually unbiased bases. *Algorithmica*, 34:512–528, 2002.

[57] Jay Lawrence, Časlav Brukner, and Anton Zeilinger. Mutually unbiased binary observable sets on n qubits. *Physical Review A*, 65(3):032320, 2002.