Universal Cost Bound of Quantum Error Mitigation Based on Quantum Estimation Theory

Kento Tsubouchi, Takahiro Sagawa, and Nobuyuki Yoshioka

Phys. Rev. Lett. **131**, 210601 — Published 22 November 2023

DOI: 10.1103/PhysRevLett.131.210601

Universal cost bound of quantum error mitigation based on quantum estimation theory

Kento Tsubouchi,[1,*] Takahiro Sagawa,[1,2] and Nobuyuki Yoshioka[1,3,4,†]

[1] *Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

[2] *Quantum-Phase Electronics Center (QPEC), The University of Tokyo, 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*

We present a unified approach to analyzing the cost of various quantum error mitigation methods on the basis of quantum estimation theory. By analyzing the quantum Fisher information matrix of a virtual quantum circuit that effectively represents the operations of quantum error mitigation methods, we derive for a generic layered quantum circuit under a wide class of Markovian noise that, unbiased estimation of an observable encounters an exponential growth with the circuit depth in the lower bound on the measurement cost. Under the global depolarizing noise, we in particular find that the bound can be asymptotically saturated by merely rescaling the measurement results. Moreover, we prove for random circuits with local noise that the cost grows exponentially also with the qubit count. Our numerical simulations support the observation that, even if the circuit has only linear connectivity, such as the brick-wall structure, each noise channel converges to the global depolarizing channel with its strength growing exponentially with the qubit count. This not only implies the exponential growth of cost both with the depth and qubit count, but also validates the rescaling technique for sufficiently deep quantum circuits. Our results contribute to the understanding of the physical limitations of quantum error mitigation and offer a new criterion for evaluating the performance of quantum error mitigation techniques.

Introduction.— One of the central problems in quantum technology is to establish control and understanding of unwanted noise, since an accumulation of errors may eventually spoil the practical advantage of quantum devices. In the case of quantum computing, an elegant framework of quantum error correction has been developed as a fundamental countermeasure [1–8], while it remains years to decades ahead when we can reliably implement provably advantageous quantum algorithms. A realistic and powerful alternative for near-future devices is to employ the art of quantum error mitigation (QEM); instead of consuming an excessive number of qubits to *correct* the bias caused by noise via interleaved measurement and feedback, we aim to *mitigate* their effect via appropriate post-processing in trade of an increased number of measurements.

A wide variety of QEM methods have been proposed: zero-noise extrapolation [9–11], probabilistic error cancellation [10, 12, 13], virtual distillation [14–17], (generalized) quantum subspace expansion [18–21], symmetry verification/expansion [22–24], and learning-based error mitigation [25, 26], to name a few (Refer to Ref. [27, 28] for review). The growing number of demonstrations by both numerical and experimental means shows that the QEM has become vital [11, 29–31]. Meanwhile, there are so far only a few guiding principles to choose from existing QEM methods [32, 33], due to the limited theoretical understanding of their fundamental aspects. It is an urgent task to understand what is the limit of QEM, in particular, the required resource to recover the desired quantum circuit output.

We find that quantum estimation theory provides a powerful tool to address this problem. Quantum estimation theory claims that, given an unbiased estimator of a physical observable, its estimation uncertainty can be characterized by the quantum Fisher information [34–36]. For example, the sampling cost for constructing an unbiased estimator for *noiseless* quantum states from measurements in *noisy* quantum states can be bounded using the quantum Fisher information [37]. While this strongly implies that the quantum estimation theory yields a tool to analyze the trade-off cost to recover the desired quantum operation, it has remained totally unknown how to investigate realistic computation models such as quantum circuits, in which the holistic effect of the error cannot be expressed by a single noise channel in general. Moreover, since QEM methods are mostly not purely classical post-processing but also require additional quantum operations, the existing framework is not straightforwardly applicable.

In this Letter, we aim to fill these gaps by extending the applicability of quantum estimation theory. By analyzing the quantum Fisher information matrix of an enlarged virtual quantum circuit which translates the operations of QEM methods, we show that the lower bound of the sampling cost for unbiased QEM grows exponentially with the circuit depth L for a generic layered quantum circuit under a wide class of noise (Theorem 1). Furthermore, for random layered circuits under local noise, we show that the cost grows exponentially also with the qubit count n (Theorem 2). We have also numerically verified that noise channels in the large

* tsubouchi@noneq.t.u-tokyo.ac.jp

† nyoshioka@ap.t.u-tokyo.ac.jp

depth regime may be effectively described by the global depolarizing channel whose strength grows exponentially with n, for which we provide an optimal technique to suppress the effect of noise. These results surpass some prior work suggesting some exponential growth (not necessarily the sampling cost of QEM) under the local depolarizing noise [32, 33, 38] from both theoretical and practical points of view: our result not only provides the first mathematical proof for a necessary condition for unbiased QEM under a wide range of noise, but also provides practical guidelines toward cost-optimal QEM.

Problem setup.— Analysis of sample complexity via the quantum estimation theory assumes operations to be expressed as quantum channels. Therefore, it is beneficial to embed QEM operations into a quantum circuit. Below, we first define a noiseless and noisy layered quantum circuit, and then present the concept of a virtual quantum circuit that encodes QEM operations.

Let $\hat{\rho} = \mathcal{U}_L \circ \cdots \circ \mathcal{U}_1(\hat{\rho}_0)$ $(\mathcal{U}_l(\cdot) = \hat{U}_l \cdot \hat{U}_l^\dagger)$ be an unknown n-qubit target state generated from L layers of noiseless unitary gates $\{\mathcal{U}_l\}_{l=1}^L$ operating on an initial state $\hat{\rho}_0$. The target state $\hat{\rho}$ can be parameterized by the generalized Bloch vector [39] $\boldsymbol{\theta} \in \mathbb{R}^{4^n-1}$ as

$$\hat{\rho} = \frac{1}{2^n}\hat{I} + 2^{(-1-n)/2}\boldsymbol{\theta}\cdot\hat{\boldsymbol{P}}, \tag{1}$$

where $\hat{I} \equiv \hat{\sigma}_0^{\otimes n}$ and $\hat{\boldsymbol{P}} = \{\hat{P}_i\}_{i=1}^{2^{2n}-1}$ is an array of non-trivial tensor product of Pauli operators $\hat{P}_i \in \{\hat{\sigma}_0, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}^{\otimes n} \setminus \{\hat{\sigma}_0\}^{\otimes n}$.

An n-qubit noisy layered circuit is defined to have the following structure: (i) noiseless preparation of initial state $\hat{\rho}_0$ [40], (ii) L layers of noisy unitary operations $\{\mathcal{E}_l \circ \mathcal{U}_l\}_{l=1}^L$ with $\mathcal{E}_l$ assumed to be a Markovian error, and (iii) noiseless POVM measurement $\mathcal{M}_0$ aimed to estimate the expectation value of a traceless observable $\hat{X} = \boldsymbol{x}\cdot\hat{\boldsymbol{P}}$ with $\boldsymbol{x} \in \mathbb{R}^{4^n-1}$. Each noise channel $\mathcal{E}_l$ maps a generalized Bloch vector as

$$\mathcal{E}_l : \boldsymbol{\theta} \mapsto A_l\boldsymbol{\theta} + \boldsymbol{c}_l, \tag{2}$$

where $(A_l)_{ij} = 2^{-n}\mathrm{tr}[\hat{P}_i\mathcal{E}_l(\hat{P}_j)]$ is the unital part of the Pauli transfer matrix of $\mathcal{E}_l$ and $(\boldsymbol{c}_l)_i = 2^{(1-3n)/2}\mathrm{tr}[\hat{P}_i\mathcal{E}_l(\hat{I})]$ quantifies the non-unital action of the noise [37]. We also define noise strength $\Gamma(\mathcal{E}_l) \equiv \|A_l\|^{-1}$ with $\|A_l\| = \max_{\boldsymbol{e}\in\mathbb{R}^{2^{2n}-1}} \frac{\|A_l\boldsymbol{e}\|}{\|\boldsymbol{e}\|}$ and $\|\boldsymbol{e}\| = \sum_i |e_i|^2$, which represents the minimal degree of shrinkage of the generalized Bloch sphere caused by $\mathcal{E}_l$, and $\gamma = \min_l \{\Gamma(\mathcal{E}_l)\}_l$ as the minimal strength among the different noise.

The objective of QEM methods is to remove the effect of the noise channels $\{\mathcal{E}_l\}_{l=1}^L$ so that we have an unbiased estimator of traceless observable $\hat{X}$, or namely $\langle\hat{X}\rangle \equiv \mathrm{tr}[\hat{\rho}\hat{X}] = 2^{(n-1)/2}\boldsymbol{\theta}\cdot\boldsymbol{x}$. Since the essence of QEM is to run noisy quantum circuits with implementable modifications into the gates, errors, and classical post-processing, we can construct a *virtual* quantum circuit which encompasses the functionality of QEM methods.

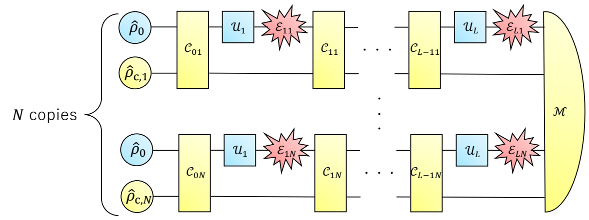

FIG. 1. A virtual quantum circuit structure that gives an equivalent representation of most existing QEM methods. The blue (red) coloring denotes that the operation is ideal (noisy), while the operations with yellow coloring explicitly involve QEM operations.

As is shown in Fig. 1, the virtual circuit involves N copies of noisy layered circuits with three-fold modifications from the original one: (boosted) noise $\mathcal{E}_{lm}$ in the l-th layer of m-th copy such that $\Gamma(\mathcal{E}_{lm}) \geq \Gamma(\mathcal{E}_l) \geq \gamma$, classical register $\hat{\rho}_{c,m}$ coupled with the system qubits via the additional operation $\mathcal{C}_{lm}$, and finally, the POVM measurement $\mathcal{M}$ performed on the entire copies to output the estimator of $\langle\hat{X}\rangle$. Classical register $\hat{\rho}_{c,m}$ is initialized with probabilistic mixtures of computational bases as $\hat{\rho}_{c,m} = \sum_i p_{mi}\,|i\rangle\langle i|$, and additional operation $\mathcal{C}_{lm}$ performs unitary operation $\mathcal{C}_{lmi}$ according to the state of the classical register as $\mathcal{C}_{lm} = \sum_i \mathcal{C}_{lmi} \otimes |i\rangle\langle i|$. Note that the virtual circuit structure excludes the quantum error correction. This is because we only allow $\mathcal{C}_{lm}$ to be unitary operation according to the state of the classical registers. We further describe in SM how various QEM methods can be mapped into this virtual circuit structure [41].

The cost of QEM can be defined as the number of copies N of the noisy circuits, or the sample complexity, which roughly can be interpreted as the number of measurements on the actual setup. Our goal is to derive the lower bound on the cost N required to perform unbiased estimation of $\langle\hat{X}\rangle$, by analyzing the evolution of the quantum Fisher information matrix of quantum states generated by the virtual circuit. Note that the lower bound described below also holds even when we think of measurement error and measurement error mitigation [42–44]. This is because noisy measurement followed by the process of measurement error mitigation can be seen as a single POVM measurement.

Main Results.— In order to achieve our goal, we reexpress the m-th copy of the quantum state in the virtual circuit as

$$\mathcal{E}'_m(\hat{\rho}(\boldsymbol{\theta}) \otimes \hat{\rho}_{c,m}), \tag{3}$$

where $\mathcal{E}'_m$ is an *effective* noise channel defined by compiling all the gates as $\mathcal{E}'_m = \mathcal{E}_{Lm} \circ \mathcal{U}_L \circ \mathcal{C}_{L-1m} \circ \cdots \circ \mathcal{E}_{1m} \circ \mathcal{U}_1 \circ \mathcal{C}_{0m} \circ \mathcal{U}_1^{-1} \circ \cdots \circ \mathcal{U}_L^{-1}$. This compilation allows us to calculate the quantum Fisher information matrix of the state right before the measurement. To be concrete, we analyze the SLD Fisher information matrix J [45] of the quantum state $\bigotimes_{m=1}^N \mathcal{E}'_m(\hat{\rho}(\boldsymbol{\theta}) \otimes \hat{\rho}_{c,m})$.

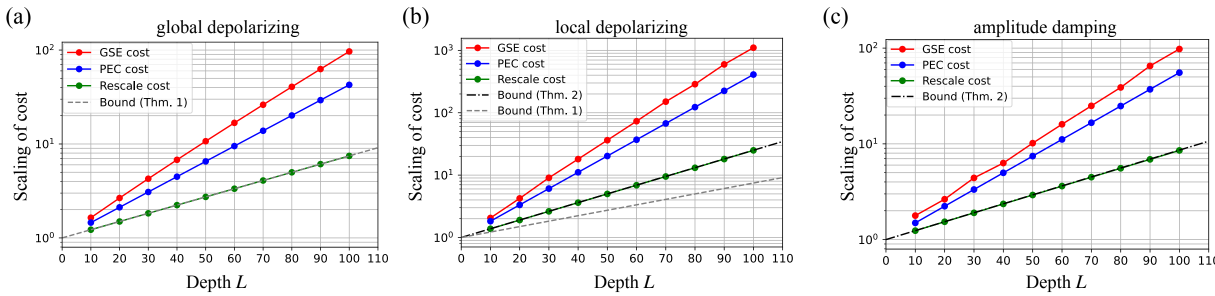

FIG. 2. Scaling of the cost to perform QEM methods for random Clifford circuit of $n = 2$ qubits under (a) global depolarizing noise, (b) local depolarizing noise, and (c) local amplitude damping noise with error rate $p = 0.01$. The red, blue, and green lines denote the sampling overhead of generalized subspace expansion [21, 46] using power subspace, the probabilistic error cancellation as derived in Ref. [47], and the rescaling technique as explained in the main text. The rescaling factor is $(1-p)^{-L}$ and $(1-p)^{-3nL4^{n-1}/(4^n-1)}$ for global and local depolarizing noise, and $(1-p)^{-2nL4^{n-1}/(4^n-1)}$ for amplitude damping noise, respectively. Bound (Thm. 1) and Bound (Thm. 2) represent the lower bound of the cost obtained from Theorem 1 and Theorem 2, respectively. The explicit scaling of Bound (Thm.2) is given in Eq. (8). Note that GSE and the rescaling methods do not completely eliminate the errors for (b) and (c), while we confirm a significant reduction of bias.

We find that J can be bounded as $J \lesssim \sum_m \Gamma(\mathcal{E}'_m)^{-2} I \lesssim N\gamma^{-2L}$, which implicates the exponential decay of J with the circuit depth L. By combining this fact with the quantum Cramér-Rao inequality, which relates J with the standard deviation ε of an unbiased estimator [48], we immediately obtain the following theorem for the cost N of the unbiased QEM (See SM [41] for the proof):

Theorem 1. *Suppose that the noise $\mathcal{E}_{lm}$ satisfies the following conditions for all l and m:*

(I) For all $\hat{\rho} \neq \hat{\sigma}$, $\mathcal{E}_{lm}(\hat{\rho}) \neq \mathcal{E}_{lm}(\hat{\sigma})$.

(II) For all $\hat{\rho}$, $\mathcal{E}_{lm}(\hat{\rho})$ is full rank, that is, $\mathcal{E}_{lm}(\hat{\rho})$ is a positive definite matrix whose eigenvalues are all greater than zero.

Then, the cost N required for any unbiased estimator of $\langle \hat{X} \rangle$ with standard deviation ε constructed from QEM that can be translated into the virtual quantum circuit in Fig. 1 satisfies

$$N \geq \frac{\|\boldsymbol{x}\|^2}{\varepsilon^2} \beta \gamma^{2L}, \tag{4}$$

where β is the largest $0 < \beta < 1$ such that $\mathcal{E}_{lm}(\hat{\rho}) - \frac{\beta}{2^n}\hat{I} \geq 0$ for all $\hat{\rho}$, l, and m. Suppose further that the noise $\mathcal{E}_{lm}$ is unital, that is, $\mathcal{E}_{lm}(\frac{\hat{I}}{2^n}) = \frac{\hat{I}}{2^n}$. Then, the cost N satisfies

$$N \geq \frac{\|\boldsymbol{x}\|^2}{\varepsilon^2}\left(1-(1-\beta)^L\right)\gamma^{2L} \sim \frac{\|\boldsymbol{x}\|^2}{\varepsilon^2}\gamma^{2L}. \tag{5}$$

Theorem 1 shows that, if $\gamma > 1$, the cost N of the unbiased QEM grows exponentially with the circuit depth L no matter how we choose $\mathcal{E}_{lm}$ (with their strength bounded from below), $\mathcal{C}_{lm}$, and $\mathcal{M}$. We can indeed show $\gamma > 1$ for unital noise under the condition (II) (See SM [41] for details).

Let us make a few remarks on the condition of Theorem 1. Condition (I) is a necessary condition for successful QEM, meaning that the information of the quantum state is not completely destroyed by noise. Condition (II) means that the variance of all observable of any state after the noise is applied is non-zero. In other words, for any observable and quantum state, the cost of obtaining an unbiased estimator from the measurement of the noisy state is greater than zero. We also remark that β is a constant that represents how far away the generalized Bloch sphere is from the original surface due to the noise.

It is noteworthy that the lower bound stated in Theorem 1 is for a generic layered quantum circuit. Since it also involves circuits that only weakly entangle qubits, the lower bound (4) does not depend on the qubit count n. However, if the quantum circuit scrambles the quantum state strong enough, we expect that every noise affects the measurement outcome; we must pay overhead to eliminate every local noise and thus encounter dependence on n. In fact, under local noise we can tighten the bound as in the following informal theorem (See SM for details [41]):

Theorem 2. *Let $U_1, U_2, ..., U_{L-1}, U_L$ be n-qubit unitary gate drawn from a set of random unitary that form unitary 2-design [49] and $\mathcal{E}_l$ be a local noise. Then, there is exponential growth with both qubit count n and depth L in the average over the number of copies N required to perform unbiased estimation of $\langle \hat{X} \rangle$ over $\{U_1, ..., U_L\}$.*

Applications.— Here, we compare the obtained bounds and the practical performance of QEM methods under realistic noise channels to determine the efficiency of existing methods. For the sake of illustrativeness, we consider three typical noise channels: the global and local depolarizing noise as representative of unital noise, and amplitude damping noise as representative of nonunital noise.

First, we consider the case where all unitary gates are followed by the global depolarizing noise $\mathcal{E}_{lm} : \boldsymbol{\theta} \mapsto (1 - p_{lm})\boldsymbol{\theta}$, where the error rate is lower bounded as $p_{lm} \geq$

p. Since the global depolarizing noise channel is unital and satisfies the assumptions of Theorem 1 with minimal noise strength $\gamma = \frac{1}{1-p}$ and $\beta = p$, the cost N required for the unbiased estimator of the expectation value $\langle \hat{X} \rangle$ constructed from QEM shall satisfy

$$N \geq \frac{\|\boldsymbol{x}\|^2}{\varepsilon^2}\left(1-(1-p)^L\right)\left(\frac{1}{1-p}\right)^{2L} \tag{6}$$

$$\sim \frac{\|\boldsymbol{x}\|^2}{\varepsilon^2}\left(\frac{1}{1-p}\right)^{2L}. \tag{7}$$

We can show that Eq. (7) can be saturated in the limit of large L. By setting $p_{lm} = p$ and ignoring the classical registers and the additional operations, the effective noise channel $\mathcal{E}'_j$ can be seen as the global depolarizing noise channel with error rate $1-(1-p)^L$. Since the measurement on the observable $\hat{X}$ yields $\langle \hat{X} \rangle^{\text{noisy}} = 2^{(n-1)/2}(1-p)^L \boldsymbol{\theta}\cdot\boldsymbol{x}$, we achieve unbiased estimation by rescaling the measurement result as $(1-p)^{-L}\langle \hat{X} \rangle^{\text{noisy}}$. Since the estimation variance on $\langle \hat{X} \rangle^{\text{noisy}}$ is $\|\boldsymbol{x}\|^2$ in the limit of large L, the sampling cost to estimate $\langle \hat{X} \rangle$ approaches $\frac{\|\boldsymbol{x}\|^2}{\varepsilon^2}\left(\frac{1}{1-p}\right)^{2L}$, which satisfies the lower bound of Theorem 1. We compare these results in Fig. 2 (a) with other error mitigation methods that also allow unbiased estimation.

Next, we consider the case of local noise $\mathcal{E}_{lm} = (\mathcal{E}^{(0)}_{lm})^{\otimes n}$ with $\mathcal{E}^{(0)}_{lm} : \boldsymbol{\theta} \mapsto (1-p_{lm})\boldsymbol{\theta}$ for local depolarizing and $(\theta_x, \theta_y, \theta_z) \mapsto (\sqrt{1-p_{lm}}\theta_x, \sqrt{1-p_{lm}}\theta_y, (1-p_{lm})\theta_z + p_{lm})$ for amplitude damping noise, where $\boldsymbol{\theta} = (\theta_x, \theta_y, \theta_z)$ denotes the Bloch vector and the error rate is lower bounded as $p_{lm} \geq p$. From Theorem 1, we can show that the cost N required by any unbiased estimator of the expectation value $\langle \hat{X} \rangle$ constructed from QEM satisfies Eq. (7) in the case of local depolarizing noise. For a random circuit whose unitary gate is drawn from unitary 2-design such as n-qubit Clifford group [49], we can even tighten this bound in the average case as

$$\mathbb{E}[N] \geq \begin{cases} O\left(\left(1+\frac{3}{2}\frac{4^n}{4^n-1}p\right)^{nL}\right) & \text{(local dep.)} \\ O\left(\left(1+\frac{4^n}{4^n-1}p\right)^{nL}\right) & \text{(amp. damping)} \end{cases} \tag{8}$$

from Theorem 2. We compare these results in Fig. 2 (b)(c) with some QEM methods.

While the scaling of Eq. (8) is derived under the assumption of unitary 2-design, our numerical simulation suggests that the bound shall hold for even wider class of quantum circuits. Concretely, as is presented in Fig. 3, the effect of each noise becomes indiscriminable from that of the global depolarizing noise whose error rate grows exponentially with n in the large-L regime, even when any of $\{\mathcal{U}_l\}$ does not constitute unitary 2-design. These results are in agreement with the phenomenological argument provided in Ref. [50] that, noise in deep layered circuits shall be modeled by global depolarizing noise with its strength fluctuating as $O(1/\sqrt{L})$. These facts not only give us another evidence for scaling as in Eq. (8) but also imply that, although we cannot remove bias completely, we may optimally suppress the effect of noise by just rescaling the measurement results as in the case of global depolarizing noise. We also applied our results for local dephasing noise, and showed that such a picture also holds as well (See SM for details [41]).

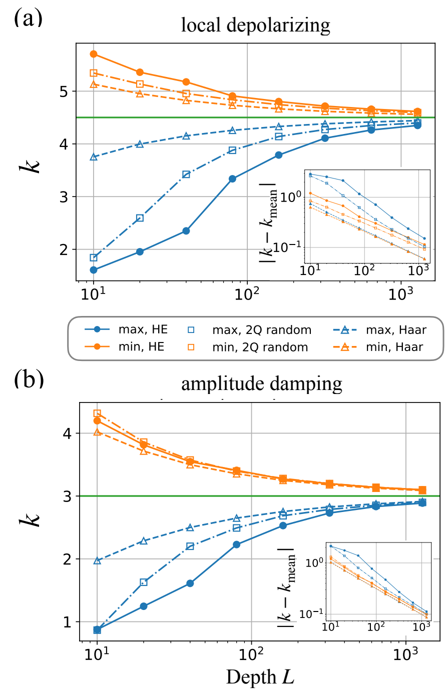

FIG. 3. Convergence of (a) local depolarizing and (b) amplitude damping into global depolarizing noise under random circuits of $n = 6$ qubits with error rate $p = 0.0001$. We denote by $(1-p)^{kL}$ the singular values of the unital part of the Pauli transfer matrix for the effective noise channel $\mathcal{E}'_m$ at each depth L, where k for the maximal and minimal ones are plotted in this figure. As is highlighted in the inset, we find that all k's approach the geometric mean k_{mean} of the singular values for each noise channel with its fluctuation scaling as $O(1/\sqrt{L})$, implying the convergence to the global depolarizing noise. For instance, $k_{\text{mean}} = 3n4^{n-1}/(4^n-1)$ for local depolarizing and $k_{\text{mean}} = 2n4^n/(4^n-1)$ for amplitude damping. Here, we consider three class of random circuits: hardware-efficient ansatz with random parameters, 2-qubit random unitary between random pairs, and Haar random unitary (See SM for details [41]).

Conclusion.— In this Letter, we have presented a theoretical analysis of quantum error mitigation (QEM) to reveal two unavoidable cost bound for unbiased QEM: exponential growth with depth L for generic layered quantum circuits, and furthermore exponential growth with qubit count n for random/chaotic quantum circuits. The lower bound is shown to be saturated under global depolarizing noise by just rescaling the measurement result, while numerical results suggest that other noise may also be mitigated as well when the circuit is sufficiently deep, since the noise including both unital ones and nonunital ones may converge to the global depolarizing noise.

We envision a rich variety of future directions. Here we mention the most important two in order. The first is to develop even more knowledge of cost-optimal QEM, especially in the early fault-tolerant regime. Even for the fault-tolerant quantum computer, a slight amount of logical errors may remain in the circuit (especially in the early regime). The implemented quantum circuits will be much deeper than those of NISQ, and thus the convergence of logical errors to global depolarizing noise is expected to be stronger. Thus, we believe that we can use our results to develop ways to utilize long-term quantum computation in the most efficient way.

The second is to incorporate the influence of bias in the estimators. QEM methods in reality are not designed to completely remove the effect of the noise, and a slight bias is allowed to remain in the estimation results. In such situations, we can expect a trade-off relationship between the cost, bias, and uncertainty of the estimator. Extending the results on single parameter estimation [51] is left as an interesting future work.

Acknowledgements.— The authors wish to thank Ryuji Takagi, Hiroyasu Tajima, and Mile Gu for insightful discussions and for sharing a preliminary version of their manuscript. The authors are also grateful to fruitful discussions with Sergey Bravyi, Suguru Endo, Keisuke Fujii, Liang Jiang, Yosuke Mitsuhashi, Changhun Oh, Zlatko Minev, Kunal Sharma, Yasunari Suzuki, and Kristan Temme. This work was supported by JST ERATO-FS Grant Number JPMJER2204, JST Grant Number JPMJPF2221, JST CREST Grant Number JPMJCR23I4, and JST ERATO Grant Number JPMJER2302, Japan. K.T. is supported by Worldleading Innovative Graduate Study Program for Materials Research, Industry, and Technology (MERITWINGS) of the University of Tokyo. T.S. is supported by JSPS KAKENHI Grant Number JP19H05796, Japan. N.Y. wishes to thank JST PRESTO No. JPMJPR2119. T. S. and N.Y. acknowledge the support from IBM Quantum.

Note added.— During the completion of our manuscript, we became aware of an independent work by Takagi et al. [52], which also showed the exponential growth of the cost N with circuit depth based on analysis of discriminability between quantum states. Also, Quek et al. [53] has theoretically analyzed the exponential scaling of sample complexity regarding both qubit counts and circuit depth via statistical learning theory. We note that, for non-unital noise, our average bound is quadratically tighter than the bound obtained by Refs. [53]. See SM [41] for more details.

[1] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. 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 rate, SIAM Journal on Computing **38**, 1207 (2008).

[4] D. A. Lidar and T. A. Brun, *Quantum error correction* (Cambridge university press, 2013).

[5] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information* (American Association of Physics Teachers, 2002).

[6] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Extending the lifetime of a quantum bit with error correction in superconducting circuits, Nature **536**, 441 (2016).

[7] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, G. J. Norris, C. K. Andersen, M. Müller, A. Blais, C. Eichler, and A. Wallraff, Realizing repeated quantum error correction in a distance-three surface code, Nature **605**, 669 (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, C. Guo, S. Guo, L. Han, N. Li, S. Li, Y. Li, F. Liang, J. Lin, H. Qian, H. Rong, H. Su, L. Sun, S. Wang, Y. Wu, Y. Xu, C. Ying, J. Yu, C. Zha, K. Zhang, Y.-H. Huo, C.-Y. Lu, C.-Z. Peng, X. Zhu, and J.-W. Pan, Realization of an error-correcting surface code with superconducting qubits, Phys. Rev. Lett. **129**, 030501 (2022).

[9] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[10] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[11] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature **567**, 491 (2019).

[12] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X **8**, 031027 (2018).

[13] E. v. d. Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse pauli-lindblad models on noisy quantum processors, arXiv preprint arXiv:2201.09866 (2022).

[14] 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).

[15] B. Koczor, Exponential error suppression for near-term quantum devices, Phys. Rev. X **11**, 031057 (2021).

[16] M. Huo and Y. Li, Dual-state purification for practical quantum error mitigation, Phys. Rev. A **105**, 022427 (2022).

[17] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles,

Qubit-efficient exponential suppression of errors, arXiv preprint arXiv:2102.06056 (2021).
[18] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, Physical Review A **95** (2017).
[19] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, Decoding quantum errors with subspace expansions, Nature Communications **11**, 636 (2020).
[20] N. Yoshioka, T. Sato, Y. O. Nakagawa, Y.-y. Ohnishi, and W. Mizukami, Variational quantum simulation for periodic materials, Phys. Rev. Research **4**, 013052 (2022).
[21] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, Generalized quantum subspace expansion, Phys. Rev. Lett. **129**, 020502 (2022).
[22] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, Low-cost error mitigation by symmetry verification, Phys. Rev. A **98**, 062339 (2018).
[23] S. McArdle, X. Yuan, and S. Benjamin, Error-mitigated digital quantum simulation, Phys. Rev. Lett. **122**, 180501 (2019).
[24] Z. Cai, Quantum Error Mitigation using Symmetry Expansion, Quantum **5**, 548 (2021).
[25] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, Quantum **5**, 592 (2021).
[26] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, Learning-based quantum error mitigation, PRX Quantum **2**, 040330 (2021).
[27] 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).
[28] 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, arXiv preprint arXiv:2210.00921 (2022).
[29] R. Sagastizabal, X. Bonet-Monroig, M. Singh, M. A. Rol, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O'Brien, and L. DiCarlo, Experimental error mitigation via symmetry verification in a variational quantum eigensolver, Phys. Rev. A **100**, 010302 (2019).
[30] J. Sun, X. Yuan, T. Tsunoda, V. Vedral, S. C. Benjamin, and S. Endo, Mitigating realistic noise in practical noisy intermediate-scale quantum devices, Phys. Rev. Applied **15**, 034026 (2021).
[31] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim, Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system, Nature Communications **11**, 587 (2020).
[32] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, npj Quantum Information **8**, 1 (2022).
[33] S. Wang, P. Czarnik, A. Arrasmith, M. Cerezo, L. Cincio, and P. J. Coles, Can error mitigation improve trainability of noisy variational quantum algorithms?, arXiv preprint arXiv:2109.01051 (2021).
[34] C. W. Helstrom, Quantum detection and estimation theory, Journal of Statistical Physics **1**, 231 (1969).
[35] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, Vol. 1 (Springer Science & Business Media, 2011).
[36] M. Hayashi, *Quantum information* (Springer, 2006).
[37] Y. Watanabe, T. Sagawa, and M. Ueda, Optimal measurement on noisy quantum systems, Phys. Rev. Lett. **104**, 020401 (2010).
[38] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation, arXiv preprint quant-ph/9611028 (1996).
[39] G. Kimura, The bloch vector for n-level systems, Physics Letters A **314**, 339 (2003).
[40] While it is straightforward to include SPAM errors as well, here we assume the preparation to be done perfectly for the sake of simplicity.
[41] See Supplementary Materials for more details (URL to be added), which includes Refs. [54–62].
[42] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).
[43] J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočnik, A. Wallraff, *et al.*, Rapid high-fidelity multiplexed readout of superconducting qubits, Physical Review Applied **10**, 034040 (2018).
[44] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, arXiv preprint arXiv:2006.14044 (2020).
[45] C. W. Helstrom, Minimum mean-squared error of estimates in quantum statistics, Physics letters A **25**, 101 (1967).
[46] B. Yang, N. Yoshioka, H. Harada, S. Hakkaku, Y. Tokunaga, H. Hakoshima, K. Yamamoto, and S. Endo, Dualgse: Resource-efficient generalized quantum subspace expansion, arXiv preprint arXiv:2309.14171 (2023).
[47] R. Takagi, Optimal resource cost for error mitigation, Phys. Rev. Research **3**, 033178 (2021).
[48] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Physical Review Letters **72**, 3439 (1994).
[49] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Physical Review A **80**, 012304 (2009).
[50] D. Qin, Y. Chen, and Y. Li, Error statistics and scalability of quantum error mitigation formulas, arXiv preprint arXiv:2112.06255 (2021).
[51] J. Liu and H. Yuan, Valid lower bound for all estimators in quantum parameter estimation, New Journal of Physics **18**, 093009 (2016).
[52] R. Takagi, H. Tajima, and M. Gu, Universal sample lower bounds for quantum error mitigation, arXiv preprint arXiv:2208.09178 (2022).
[53] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, arXiv preprint arXiv:2210.11505 (2022).
[54] T. Sagawa, *Entropy, Divergence, and Majorization in Classical and Quantum Thermodynamics*, Vol. 16 (Springer Nature, 2022).
[55] M. Ozawa, Entanglement measures and the hilbert–schmidt distance, Physics Letters A **268**, 158 (2000).
[56] D. Weingarten, Asymptotic behavior of group integrals in the limit of infinite rank, Journal of Mathematical

Physics **19**, 999 (1978).

[57] B. Collins, Moments and cumulants of polynomial random variables on unitarygroups, the itzykson-zuber integral, and free probability, International Mathematics Research Notices **2003**, 953 (2003).

[58] B. Collins and P. Śniady, Integration with respect to the haar measure on unitary, orthogonal and symplectic group, Communications in Mathematical Physics **264**, 773 (2006).

[59] Z. Webb, The clifford group forms a unitary 3-design, arXiv preprint arXiv:1510.02769 (2015).

[60] H. Zhu, Multiqubit clifford groups are unitary 3-designs, Physical Review A **96**, 062336 (2017).

[61] J. Jiang, K. Wang, and X. Wang, Physical implementability of linear maps and its application in error mitigation, Quantum **5**, 600 (2021).

[62] M. J. Kastoryano and K. Temme, Quantum logarithmic sobolev inequalities and rapid mixing, Journal of Mathematical Physics **54**, 052202 (2013).