Universal Sampling Lower Bounds for Quantum Error Mitigation

Ryuji Takagi,[1,2,*] Hiroyasu Tajima,[3,4,†] and Mile Gu[2,5,6,‡]
[1]*Department of Basic Science, The University of Tokyo, Tokyo 153-8902, Japan*
[2]*Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore*
[3]*Department of Communication Engineering and Informatics, University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585, Japan*
[4]*JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan*
[5]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore*
[6]*MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit UMI 3654, Singapore*

(Received 15 November 2022; accepted 10 October 2023; published 22 November 2023)

Numerous quantum error-mitigation protocols have been proposed, motivated by the critical need to suppress noise effects on intermediate-scale quantum devices. Yet, their general potential and limitations remain elusive. In particular, to understand the ultimate feasibility of quantum error mitigation, it is crucial to characterize the fundamental sampling cost—how many times an arbitrary mitigation protocol must run a noisy quantum device. Here, we establish universal lower bounds on the sampling cost for quantum error mitigation to achieve the desired accuracy with high probability. Our bounds apply to general mitigation protocols, including the ones involving nonlinear postprocessing and those yet to be discovered. The results imply that the sampling cost required for a wide class of protocols to mitigate errors must grow exponentially with the circuit depth for various noise models, revealing the fundamental obstacles in the scalability of useful noisy near-term quantum devices.

DOI: 10.1103/PhysRevLett.131.210602

Introduction.—As recent technological developments have started to realize controllable small-scale quantum devices, a central problem in quantum information science has been to pin down what can and cannot be accomplished with noisy intermediate-scale quantum (NISQ) devices [1]. One of the most relevant issues in understanding the ultimate capability of quantum hardware is to characterize how well noise effects could be circumvented. This is especially so for NISQ devices, as today's quantum devices generally cannot accommodate full quantum error correction that requires scalable quantum architecture. As an alternative to quantum error correction, quantum error mitigation has recently attracted much attention as a potential tool to help NISQ devices realize useful applications [2,3]. It is thus of primary interest from practical and foundational viewpoints to understand the ultimate feasibility of quantum error mitigation.

Quantum error mitigation protocols generally involve running available noisy quantum devices many times. The collected data is then postprocessed to infer classical information of interest. While this avoids the engineering challenge in error correction, it comes at the price of *sampling cost*—computational overhead in having to sample a noisy device many times. This sampling cost represents the crucial quantity determining the feasibility of quantum error mitigation. If the required sampling cost becomes too large, then such quantum error mitigation protocol becomes infeasible under a realistic time constraint. Various prominent quantum error mitigation methods face this problem, where sampling cost grows exponentially with circuit size [4–8]. The crucial question then is whether there is hope to come up with a new error mitigation strategy that avoids this hurdle or if this is a universal feature shared by all quantum error mitigation protocols. To answer this question, we need a characterization of the sampling cost that is universally required for the general error-mitigation protocols, which has hitherto been unknown.

Here, we provide a solution to this problem. We derive lower bounds for the number of samples fundamentally required for general quantum error mitigations to realize the target performance. We then show that the required samples for a wide class of mitigation protocols to error-mitigate layered circuits under various noise models—including the depolarizing and stochastic Pauli noise—must grow exponentially with the circuit depth to achieve the target performance. This turns the conjecture that quantum error mitigation would generally suffer from the exponential sampling overhead into formal relations, extending the previous results on the exponential resource overhead required for noisy circuits without postprocessing [9–11]. We accomplish these by employing an information-theoretic approach, which establishes the novel connection between the state distinguishability and operationally motivated error-mitigation performance measures. Our results place the fundamental limitations imposed on the capability of general error-mitigation strategies that include existing protocols [5–8,12–31] and the ones yet to be discovered, being analogous to the performance converse bounds established in several other disciplines—such as thermodynamics [32–34], quantum communication [35,36], and quantum resource theories [37,38]—that contributed to characterizing the ultimate operational capability allowed in each physical setting.

Our work complements and extends several recent advancements in the field. Reference [39] introduced a general framework of quantum error mitigation and established lower bounds for the maximum estimator spread, i.e., the range of the outcomes of the estimator, imposed on all error mitigation in the class, which provides a *sufficient* number of samples to ensure the target accuracy. Those bounds were then employed to show that the maximum spread grows exponentially with the circuit depth to mitigate local depolarizing noise. Reference [40] showed a related result where for the class of error-mitigation strategies that only involve linear postprocessing, in which the target expectation value can be represented by a linear combination of the actually observed quantities, either the maximum estimator spread or the sample number needs to grow exponentially with the circuit depth to mitigate local depolarizing noise. The severe obstacle induced by noise in showing a quantum advantage for variational quantum algorithms has also recently been studied [41–43]. Our results lift the observations made in these works to rigorous bounds for the *necessary* sampling cost required for general error mitigation, including the ones involving nonlinear postprocessing that constitute a large class of protocols [7,12–17,19,21,22,24,25,27,29–31,44].

Framework.—Suppose we wish to obtain the expectation value of an observable $A \in \mathbb{O}$ for an ideal state $\rho \in \mathbb{S}$ where $\mathbb{O}$ and $\mathbb{S}$ are some sets of observables and states. We assume that the ideal quantum state ρ is produced by a unitary quantum circuit $\mathcal{U}$ applied to the initial state $\rho_{\rm ini} \in \mathbb{S}_{\rm in}$ as $\rho = \mathcal{U}(\rho_{\rm ini})$ where $\mathbb{S}_{\rm in}$ is the set of possible input states. The noise in the circuit, however, prevents us from preparing the state ρ exactly. We consider quantum error mitigation protocols that aim to estimate the true expectation value under the presence of noise in the following manner [39] (see also Fig. 1).

In the mitigation procedure, one can first modify the circuit, e.g., use a different choice of unitary gates with potential circuit simplification, apply nonadaptive operations (enabling, e.g., dynamical decoupling [45,46] and Pauli twirling [6]), and supply ancillary qubits—the allowed modifications are determined by the capability of the available device. Together with the noise present in the modified circuit, this turns the original unitary $\mathcal{U}$ into some quantum channel $\mathcal{F}$, which produces a *distorted state* ρ'. The distorted state can be represented in terms of the ideal state ρ by $\rho' = \mathcal{E}(\rho)$, where we call $\mathcal{E} \coloneqq \mathcal{F} \circ \mathcal{U}^\dagger$ an *effective noise channel*.

The second step consists of collecting N samples $\{\mathcal{E}_n(\rho)\}_{n=1}^N$ of distorted states represented by a set of effective noise channels $\mathbb{E} \coloneqq \{\mathcal{E}_n\}_{n=1}^N$ and applying a *trailing quantum process* $\mathcal{P}_A$ over them. The effective noise channels in $\mathbb{E}$ can be different from each other in general, as noisy hardware could have different noise profiles each time, or could purposely change the noise strength [5,47]. The trailing process $\mathcal{P}_A$ then outputs an estimate represented by a random variable $\hat{E}_A(\rho)$ for the true expectation value $\mathrm{Tr}(A\rho)$. The main focus of our study is the *sampling number* N, the total number N of distorted states used in the error mitigation process.

We quantify the performance of an error-mitigation protocol by how well the protocol can estimate the expectation values for a given set $\mathbb{O}$ of observables and a set $\mathbb{S}$ of ideal states, which we call *target observables* and *target states* respectively. We keep the choices of these sets general, and they can be flexibly chosen depending on one's interest. For instance, if one is interested in error mitigation protocols designed to estimate the Pauli observables (e.g., virtual distillation [15,16,21]), $\mathbb{O}$ can be chosen as the set of Pauli operators. As the trailing process includes a measurement depending on the observable, an error-mitigation strategy with target observables $\mathbb{O}$ is equipped with a family of trailing processes $\{\mathcal{P}_A\}_{A\in\mathbb{O}}$. Similarly, our results hold for an arbitrary choice of $\mathbb{S}$, where one can, for instance, choose this as the set of all quantum states, which better describes the protocols such as probabilistic error cancellation [5,47–52], or as the set of states in a certain subspace, which captures the essence of subspace expansion [12,14,17].

This framework includes many error-mitigation protocols proposed so far [5–8,12–31]. It is worth noting that our framework includes protocols that involve nonlinear postprocessing of the measurement outcomes. Error-mitigation protocols typically work by (i) making some set of (usually Pauli) measurements for observables $\{O_i\}_i$, (ii) estimating their expectation values $\{\langle O_i\rangle\}_i$ for distorted states, and (iii) applying a classical postprocessing function f over them. The protocols with linear postprocessing functions, i.e., the ones with the form $f(\langle O_i\rangle_i) = \sum_i c_i\langle O_i\rangle$, are known to admit simpler analysis [39,40], but numerous protocols—including virtual distillation [15,16,21], symmetry verification [13], and subspace expansion [12,14,17]—come with nonlinear postprocessing functions. In our framework, the sampling number N is the *total* number of samples used, where we consider the output represented by $\hat{E}_A(\rho)$ as our final guess and thus do not generally assume repeating some procedure many times and take a statistical average. This enables us to have any postprocessing absorbed in the trailing process $\mathcal{P}_A$, making our results valid for the protocols with nonlinear postprocessing functions.

We also remark that our framework includes protocols with much more operational power than existing protocols, as we allow the trailing process to apply any coherent interaction over all distorted states. Our results thus provide fundamental limits on the sampling overhead applicable to an arbitrary protocol in this extended class of error-mitigation protocols.

Sampling lower bounds.—We now consider the required samples to ensure the target performance. The performance of quantum error mitigation can be defined in multiple ways. Here, we consider two possible performance quantifiers that are operationally relevant.

Our first performance measure is the combination of the accuracy of the estimate and the success probability. This closely aligns with the operational motivation, where one would like an error mitigation strategy to be able to provide a good estimate for each observable in $\mathbb{O}$ and an ideal state in $\mathbb{S}$ at a high probability. This can be formalized as a condition

$$\mathrm{Prob}(|\mathrm{Tr}(A\rho)-\hat{E}_A(\rho)|\leq\delta)\geq 1-\varepsilon,\ \ \forall\,\rho\in\mathbb{S},\ \ \forall\,A\in\mathbb{O} \tag{1}$$

where δ is the target accuracy and $1-\varepsilon$ is the success probability (see also Fig. 1).

The problem then is to identify lower bounds on the number N of distorted states needed to achieve this condition as a function of δ and ε. We address this by

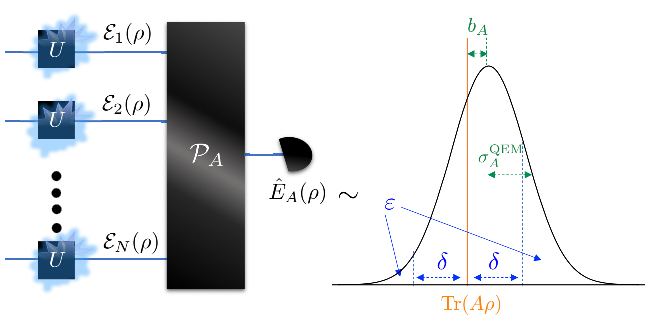

FIG. 1. Framework of quantum error mitigation. For an ideal state $\rho\in\mathbb{S}$ and an observable $A\in\mathbb{O}$ of interest, we first prepare N copies of distorted states $\{\mathcal{E}_n(\rho)\}_{n=1}^N$, where $\mathbb{E}=\{\mathcal{E}_n\}_{n=1}^N$ is the set of effective noise channels. A trailing quantum process $\mathcal{P}_A$ is then applied to N distorted states, producing the final estimation of $\mathrm{Tr}(A\rho)$ represented by a random variable $\hat{E}_A(\rho)$. We quantify the error-mitigation performance in two ways by studying the property of the distribution of $\hat{E}_A(\rho)$; the first is the combination of the accuracy δ and the success probability $1-\varepsilon$, and the second is the combination of the bias $b_A(\rho)\coloneqq\langle\hat{E}_A(\rho)\rangle-\mathrm{Tr}(A\rho)$ and the standard deviation σ_A^{QEM} of $\hat{E}_A(\rho)$.

observing that the trailing process of quantum error mitigation is represented as an application of a quantum channel and thus can never increase the state distinguishability. To formulate our result, let us define the observable-dependent distinguishability with respect to a set $\mathbb{O}$ of observables as

$$D_{\mathbb{O}}(\rho,\sigma)\coloneqq\max_{A\in\mathbb{O}}|\mathrm{Tr}[A(\rho-\sigma)]|. \tag{2}$$

This quantity can be understood as the resolution in distinguishing two quantum states by using the measurements of the observables in $\mathbb{O}$. We note that when $\mathbb{O}=\{A|0\leq A\leq\mathbb{I}\}$ [53], the quantity in (2) becomes the trace distance $D_{\mathrm{tr}}(\rho,\sigma)=\frac{1}{2}\|\rho-\sigma\|_1$ [54].

We then obtain the following sampling lower bounds applicable to an arbitrary given set $\mathbb{E}$ of effective noise channels. (Proof in Appendix A in the Supplemental Material [55].)

Theorem 1.—Suppose that an error-mitigation strategy achieves (1) with some $\delta\geq 0$ and $0\leq\varepsilon\leq 1/2$ with N distorted states characterized by the effective noise channels $\mathbb{E}=\{\mathcal{E}_n\}_{n=1}^N$. Then, the sample number N is lower bounded as

$$\begin{aligned} N &\geq \max_{\substack{\rho,\sigma\in\mathbb{S}\\ D_{\mathbb{O}}(\rho,\sigma)\geq 2\delta}}\ \min_{\mathcal{E}\in\mathbb{E}}\frac{\log\left[\frac{1}{4\varepsilon(1-\varepsilon)}\right]}{\log\left[1/F(\mathcal{E}(\rho),\mathcal{E}(\sigma))\right]},\\ N &\geq \max_{\substack{\rho,\sigma\in\mathbb{S}\\ D_{\mathbb{O}}(\rho,\sigma)\geq 2\delta}}\ \min_{\mathcal{E}\in\mathbb{E}}\frac{2(1-2\varepsilon)^2}{\ln 2\cdot S(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))}, \end{aligned} \tag{3}$$

where $F(\rho,\sigma)\coloneqq\|\sqrt{\rho}\sqrt{\sigma}\|_1^2$ is the (square) fidelity and $S(\rho\|\sigma)\coloneqq\mathrm{Tr}(\rho\log\rho)-\mathrm{Tr}(\rho\log\sigma)$ is the relative entropy.

This result tells that if the noise effect brings states close to each other, it incurs an unavoidable sampling cost to error mitigation. The minimization over $\mathbb{E}$ chooses the effective noise channel that least reduces the infidelity and the relative entropy, respectively. On the other hand, the maximum over the ideal states represents the fact that to mitigate two states ρ and σ that are separated further than 2δ in terms of observables in $\mathbb{O}$, the sample number N that achieves the accuracy δ and the success probability $1-\varepsilon$ must satisfy the lower bounds with respect to ρ and σ. The maximization over such ρ and σ then provides the tightest lower bound. This also reflects the observation that error mitigation accommodating a larger set $\mathbb{O}$ of target observables would require a larger number of samples.

We remark that although the set $\mathbb{E}$—which depends on how one modifies the noisy circuit—ultimately depends on a specific error-mitigation strategy in mind, fixing $\mathbb{E}$ to a certain form already provides useful insights as we see later in the context of noisy layered circuits. We also stress that the above bounds hold for an arbitrary choice of $\mathbb{E}$, providing the general relation between the error mitigation performance and the information-theoretic quantity.

The bounds in Theorem 1 depend on the accuracy δ implicitly through the constraints on ρ and σ in the maximization. For instance, if one sets $\delta = 0$, one can find that both bounds diverge, as the choice of $\sigma = \rho$ would be allowed in the maximization. In Appendix B of [53], we report an alternative bound that has an explicit dependence on the accuracy δ.

Let us now consider our second performance measure based on the standard deviation and the bias of the estimate. Let $\sigma_A^{\mathrm{QEM}}(\rho)$ be the standard deviation of $\hat{E}_A(\rho)$ for an observable $A \in \mathbb{O}$, which represents the uncertainty of the final estimate of an error mitigation protocol. Since a good error mitigation protocol should come with a small fluctuation in its outcome, the standard deviation of the underlying distribution for the estimate can serve as a performance quantifier. However, the standard deviation itself is not sufficient to characterize the error mitigation performance, as one can easily come up with a useless strategy that always outputs a fixed outcome, which has zero standard deviation. This issue can be addressed by considering the deviation of the expected value of the estimate from the true expectation value called bias, defined as $b_A(\rho) \coloneqq \langle \hat{E}_A(\rho) \rangle - \mathrm{Tr}(A\rho)$ for a state $\rho \in \mathbb{S}$ and an observable $A \in \mathbb{O}$ (see also Fig. 1).

To assess the performance of error-mitigation protocols, we consider the worst-case error among possible ideal states and measurements. This motivates us to consider the maximum standard deviation $\sigma_{\max}^{\mathrm{QEM}} \coloneqq \max_{A \in \mathbb{O}} \max_{\rho \in \mathbb{S}} \sigma_A^{\mathrm{QEM}}(\rho)$ and the maximum bias $b_{\max} \coloneqq \max_{A \in \mathbb{O}} \max_{\rho \in \mathbb{S}} b_A(\rho)$. Then, we obtain the following sampling lower bound in terms of these performance quantifiers. (Proof in Appendix C of [53].)

Theorem 2.—The sampling cost for an error-mitigation strategy with the maximum standard deviation $\sigma_{\max}^{\mathrm{QEM}}$ and the maximum bias $b_{\max}$ is lower bounded as

$$N \geq \max_{\substack{\rho,\sigma \in \mathbb{S} \\ D_{\mathbb{O}}(\rho,\sigma) - 2b_{\max} \geq 0}} \min_{\mathcal{E} \in \mathbb{E}} \frac{\log \left[1 - \frac{1}{\left(1 + \frac{2\sigma_{\max}^{\mathrm{QEM}}}{D_{\mathbb{O}}(\rho,\sigma) - 2b_{\max}} \right)^2} \right]^{-1}}{\log \left[1/F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \right]}. \quad (4)$$

This result represents the trade-off between the standard deviation, bias, and the required sampling cost. To realize the small standard deviation and bias, error mitigation needs to use many samples; in fact, the lower bound diverges at the limit of $\sigma_{\max}^{\mathrm{QEM}} \to 0$ whenever there exist states $\rho, \sigma \in \mathbb{S}$ such that $D_{\mathbb{O}}(\rho, \sigma) \geq 2b_{\max}$. On the other hand, a larger bias results in a smaller sampling lower bound, indicating a potential to reduce the sampling cost by giving up some bias.

The bounds in Theorems 1,2 are universally applicable to arbitrary error mitigation protocols in our framework. Therefore, our bounds are not expected to give good estimates for a given specific error-mitigation protocol in general, just as there is a huge gap between the Carnot efficiency and the efficiency of most of the practical heat engines. Nevertheless, it is still insightful to investigate how our bounds are compared to existing mitigation protocols. In Appendix D [53], we compare the bound in Theorem 1 to the sampling cost for several error-mitigation methods, showing that our bound can provide nontrivial lower bounds with the gap being the factor of 3 to 6. Although this does not guarantee that our bound behaves similarly for other scenarios in general, this ensures that there is a setting in which the bound in Theorem 1 can provide a nearly tight estimate. We further show in Appendix E [53] that the scaling of the lower bound in Theorem 2 with noise strength can be achieved by the probabilistic error cancellation method in a certain scenario. This shows that probabilistic error cancellation serves as an optimal protocol in this specific sense, complementing the recent observation on the optimality of probabilistic error cancellation established for the maximum estimator spread measure [39].

Noisy layered circuits.—The above results clarify the close relation between the sampling cost and state distinguishability. As an application of our general bounds, we study the inevitable sample overhead to mitigate noise in the circuits consisting of multiple layers of unitaries. Although we here focus on the local depolarizing noise, our results can be extended to a number of other noise models as we discuss later.

Suppose that an M-qubit quantum circuit consists of layers of unitaries, each of which is followed by a local depolarizing noise, i.e., a depolarizing noise of the form $\mathcal{D}_p = (1-p)id + p\mathbb{I}/2$, where p is a noise strength, applies to each qubit. We aim to estimate ideal expectation values for the target states $\mathbb{S}$ and observables $\mathbb{O}$ by using N such noisy layered circuits. Although the noise strength can vary for different locations, we suppose that L layers are followed by the local depolarizing noise with noise strength of at least γ. We call these layers $U_1, U_2, \ldots, U_L$ and let $\gamma_{n,l,m}$ denote the noise strength of the local depolarizing noise on the mth qubit after the lth unitary layer U_l in the nth noisy circuit, where $m \leq M$, $l \leq L$, $n \leq N$. This gives the expression of the local depolarizing noise after lth layer in the nth noisy circuit as $\bigotimes_{m=1}^{M} \mathcal{D}_{\gamma_{n,l,m}}$, where $\gamma_{n,l,m} \geq \gamma \; \forall \, n, l, m$.

Here, we focus on the error-mitigation protocols that apply an arbitrary trailing process over N distorted states and any unital operations (i.e., operations that preserve the maximally mixed state) before and after U_l (Fig. 2). This structure ensures that error correction does not come into play here, as the size of input and output spaces of the intermediate unital channels is restricted to M qubits, as well as that unital channels do not serve as good decoders for error correction.

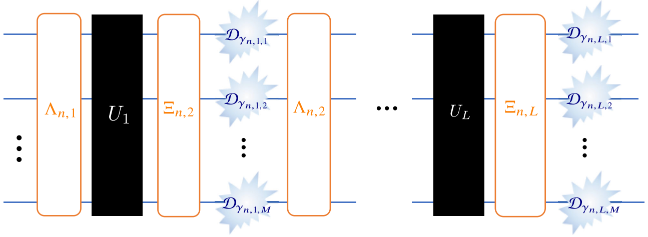

FIG. 2. Each distorted state (nth copy depicted in the figure) is produced by a circuit with L layers $U_1, \ldots, U_L$ followed by a local depolarizing noise with noise strength at least γ, i.e., $\gamma_{n,l,m} \geq \gamma, \ \forall\ n, l, m$. Each layer U_l can be sandwiched by additional unital operations $\Lambda_{n,l}$ and $\Xi_{n,l}$. Other layers and depolarizing channels with noise strength smaller than γ are absorbed in these operations as they are also unital.

We show that the necessary number of samples required to achieve the target performance grows exponentially with the number of layers in both performance quantifiers introduced above.

Theorem 3.—Suppose that an error-mitigation strategy described above is applied to an M-qubit circuit to mitigate local depolarizing channels with strength at least γ that follow L layers of unitaries, and achieves (1) with some $\delta \geq 0$ and $0 \leq \varepsilon \leq 1/2$. Then, if there exist at least two states $\rho, \sigma \in \mathbb{S}$ such that $D_{\mathbb{O}}(\rho, \sigma) \geq 2\delta$, the required sample number N is lower bounded as

$$N \geq \frac{(1-2\varepsilon)^2}{2\ln(2) M (1-\gamma)^{2L}}. \quad (5)$$

The proof can be found in Appendix F [53]. This result particularly shows that the required number of samples must grow exponentially with the circuit depth L. We remark that the bound always holds under the mild condition, i.e., $D_{\mathbb{O}}(\rho, \sigma) \geq 2\delta$ for some $\rho, \sigma \in \mathbb{S}$. This reflects that, to achieve the desired accuracy δ satisfying this condition, error mitigation really needs to extract the expectation values about the observables in $\mathbb{O}$ and the states in $\mathbb{S}$, prohibiting it from merely making a random guess.

In Appendix G [53], we obtain a similar exponential growth of the required sample overhead for a fixed target bias and standard deviation. We also obtain in Appendix H [53] alternative bounds that are tighter in the range of small ε.

With a suitable modification of allowed unitaries and intermediate operations, we extend these results to a wide class of noise models, including stochastic Pauli, global depolarizing, and thermal noise. The case of thermal noise particularly provides an intriguing physical interpretation: the sampling cost N required to mitigate thermal noise after time t is characterized by the loss of free energy $N = \Omega(1/[F(\rho_t) - F_{\rm eq}])$, where ρ_t is the state at time t and $F_{\rm eq}$ is the equilibrium free energy. This in turn shows that the necessary sampling cost grows as $N = \Omega(e^{\alpha_{\rm ent} t})$, where $\alpha_{\rm ent}$ is a constant characterized by the minimum entropy production rate. We provide details on these extensions in Appendix I of the Supplemental Material [53].

We remark that Theorem 1 (and related results discussed in the Supplemental Material [53]) extends the previous results showing the exponential resource overhead required for noisy circuits without postprocessing [9–11]. In Appendix J, we provide further clarifications about the differences between the settings considered in the previous works and ours.

Conclusions.—We established sampling lower bounds imposed on the general quantum error-mitigation protocols. Our results formalize the idea that the reduction in the state distinguishability caused by noise and error-mitigation processes leads to the unavoidable computational overhead in quantum error mitigation. We then showed that error-mitigation protocols with certain intermediate operations and an arbitrary trailing process require the number of samples that grows exponentially with the circuit depth to mitigate various types of noise. We presented these bounds with respect to multiple performance quantifiers—accuracy and success probability, as well as the standard deviation and bias—each of which has its own operational relevance.

Our bounds provide fundamental limitations that universally apply to general mitigation protocols, clarifying the underlying principle that regulates error-mitigation performance. As a trade-off, they may not give tight estimates for a given specific error-mitigation strategy, analogously to many other converse bounds established in other fields that typically give loose bounds for most specific protocols. A thorough study to identify in what setting our bounds can give good estimates will make an interesting future research direction.

Note added.— Recently, we became aware of an independent work by Tsubouchi *et al.* [81] that obtained a result related to our Theorem S.2 in Appendix G [53], in which they showed an alternative exponential sample lower bound applicable to error-mitigation protocols that achieve zero bias using quantum estimation theory.

We thank Suguru Endo, Shintaro Minagawa, Masato Koashi, and Daniel Stilck França for helpful discussions, and Kento Tsubouchi, Takahiro Sagawa, and Nobuyuki Yoshioka for sharing a preliminary version of their manuscript. We acknowledge the support of the Singapore Ministry of Education Tier 1 Grants No. RG146/20 and No. RG77/22 (S), the No. NRF2021-QEP2- 02-P06 from the Singapore Research Foundation and the Singapore Ministry of Education Tier 2 Grant No. T2EP50221-0014 and the No. FQXi R-710-000-146-720 Grant "Are quantum agents more energetically efficient at making predictions?" from the Foundational Questions Institute and Fetzer Franklin Fund (a donor-advised fund of Silicon Valley Community Foundation). R. T. was also supported by the Lee Kuan Yew Postdoctoral Fellowship at Nanyang Technological University Singapore. H. T. is supported by JSPS Grants-in-Aid for Scientific Research No. JP19K14610 and No. JP22H05250, JST PRESTO No. JPMJPR2014, and JST MOONSHOT No. JPMJMS2061.

---

*ryuji.takagi@phys.c.u-tokyo.ac.jp
†hiroyasu.tajima@uec.ac.jp
‡mgu@quantumcomplexity.org

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).
[2] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).
[3] 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).
[4] X. Yuan, Z. Zhang, N. Lütkenhaus, and X. Ma, Simulating single photons with realistic photon sources, Phys. Rev. A **94**, 062305 (2016).
[5] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).
[6] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).
[7] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X **8**, 031027 (2018).
[8] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, Phys. Rev. A **103**, 042605 (2021).
[9] W. G. Unruh, Maintaining coherence in quantum computers, Phys. Rev. A **51**, 992 (1995).
[10] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation, arXiv:quant-ph/9611028.
[11] G. M. Palma, K.-a. Suominen, and A. Ekert, Quantum computers and dissipation, Proc. R. Soc. A **452**, 567 (1996).
[12] 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, Phys. Rev. A **95**, 042308 (2017).
[13] 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).
[14] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, Decoding quantum errors with subspace expansions, Nat. Commun. **11**, 636 (2020).
[15] B. Koczor, Exponential error suppression for near-term quantum devices, Phys. Rev. X **11**, 031057 (2021).
[16] 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).
[17] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, Generalized quantum subspace expansion, Phys. Rev. Lett. **129**, 020502 (2022).
[18] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, Quantum **5**, 592 (2021).
[19] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, Learning-based quantum error mitigation, PRX Quantum **2**, 040330 (2021).
[20] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unified approach to data-driven quantum error mitigation, Phys. Rev. Res. **3**, 033098 (2021).
[21] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, Qubit-efficient exponential suppression of errors, arXiv:2102.06056.
[22] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, Error mitigation via verified phase estimation, PRX Quantum **2**, 020317 (2021).
[23] M. Huo and Y. Li, Dual-state purification for practical quantum error mitigation, Phys. Rev. A **105**, 022427 (2022).
[24] D. Bultrini, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unifying and benchmarking state-of-the-art quantum error mitigation techniques, Quantum **7**, 1034 (2023).
[25] Z. Cai, Quantum error mitigation using symmetry expansion, Quantum **5**, 548 (2021).
[26] K. Wang, Y.-A. Chen, and X. Wang, Mitigating quantum errors via truncated Neumann series, Sci. China Inf. Sci. **66**, 180508 (2023).
[27] Z. Cai, Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications, npj Quantum Inf. **7**, 80 (2021).
[28] A. Mari, N. Shammah, and W. J. Zeng, Extending quantum probabilistic error cancellation by noise scaling, Phys. Rev. A **104**, 052607 (2021).
[29] R. Mezher, J. Mills, and E. Kashefi, Mitigating errors by quantum verification and postselection, Phys. Rev. A **105**, 052608 (2022).
[30] E. Fontana, I. Rungger, R. Duncan, and C. Cîrstoiu, Spectral analysis for noise diagnostics and filter-based digital error mitigation, arXiv:2206.08811.
[31] Y. Xiong, S. X. Ng, and L. Hanzo, Quantum error mitigation relying on permutation filtering, IEEE Trans. Commun. **70**, 1927 (2022).
[32] S. Carnot, Reflections on the motive power of fire, and on machines fitted to develop that power, Paris: Bachelier **108**, 1824 (1824).
[33] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. **5**, 183 (1961).
[34] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, Proc. Natl. Acad. Sci. U.S.A. **112**, 3275 (2015).
[35] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted classical capacity of noisy quantum channels, Phys. Rev. Lett. **83**, 3081 (1999).
[36] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Fundamental limits of repeaterless quantum communications, Nat. Commun. **8**, 15043 (2017).
[37] B. Regula and R. Takagi, Fundamental limitations on distillation of quantum channel resources, Nat. Commun. **12**, 4411 (2021).

[38] K. Fang and Z.-W. Liu, No-go theorems for quantum resource purification: New approach and channel theory, PRX Quantum **3**, 010337 (2022).

[39] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, npj Quantum Inf. **8**, 114 (2022).

[40] 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:2109.01051.

[41] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[42] D. Stilck França and R. García-Patrón, Limitations of optimization algorithms on noisy quantum devices, Nat. Phys. **17**, 1221 (2021).

[43] 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).

[44] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, Scalable mitigation of measurement errors on quantum computers, PRX Quantum **2**, 040326 (2021).

[45] L. Viola and S. Lloyd, Dynamical suppression of decoherence in two-state quantum systems, Phys. Rev. A **58**, 2733 (1998).

[46] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, Phys. Rev. Lett. **82**, 2417 (1999).

[47] F. Buscemi, M. Dall'Arno, M. Ozawa, and V. Vedral, Direct observation of any two-point quantum correlation function, arXiv:1312.4240.

[48] R. Takagi, Optimal resource cost for error mitigation, Phys. Rev. Res. **3**, 033178 (2021).

[49] B. Regula, R. Takagi, and M. Gu, Operational applications of the diamond norm and related measures in quantifying the non-physicality of quantum maps, Quantum **5**, 522 (2021).

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

[51] 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. Appl. **15**, 034026 (2021).

[52] C. Piveteau, D. Sutter, and S. Woerner, Quasiprobability decompositions with reduced sampling overhead, npj Quantum Inf. **8**, 12 (2022).

[53] Here, we consider the standard matrix inequality with respect to the positive semidefinite cone, i.e., $M_1 \leq M_2$ for two matrices M_1 and M_2 means that the matrix $M_2 - M_1$ is positive semidefinite.

[54] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).

[55] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.210602 for detailed proofs and discussions of our main results, which includes Refs. [56–80].

[56] C. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[57] F. Hiai, M. Ohya, and M. Tsukada, Sufficiency, KMS condition and relative entropy in von Neumann algebras, Pac. J. Math. **96**, 99 (1981).

[58] D. Reeb, M. J. Kastoryano, and M. M. Wolf, Hilbert's projective metric in quantum information theory, J. Math. Phys. (N.Y.) **52**, 082201 (2011).

[59] M. J. Kastoryano and K. Temme, Quantum logarithmic Sobolev inequalities and rapid mixing, J. Math. Phys. (N.Y.) **54**, 052202 (2013).

[60] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd, The quantum Wasserstein distance of order 1, IEEE Trans. Inf. Theory **67**, 6627 (2021).

[61] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).

[62] H. Tajima, N. Shiraishi, and K. Saito, Uncertainty relations in implementation of unitary operations, Phys. Rev. Lett. **121**, 110403 (2018).

[63] R. Katsube, M. Hotta, and K. Yamaguchi, A fundamental upper bound for signal to noise ratio of quantum detectors, J. Phys. Soc. Jpn. **89**, 054005 (2020).

[64] M. Hayashi, *Quantum Information Theory: Mathematical Foundation* (Springer, New York, 2016).

[65] K. M. R. Audenaert and J. Eisert, Continuity bounds on the quantum relative entropy, J. Math. Phys. (N.Y.) **46**, 102104 (2005).

[66] R. Carbone and A. Martinelli, Logarithmic Sobolev inequalities in non-commutative algebras, Inf. Dim. Anal. Quant. Prob. Rel. Topics **18**, 1550011 (2015).

[67] M. J. Kastoryano and F. G. S. L. Brandão, Quantum Gibbs samplers: The commuting case, Commun. Math. Phys. **344**, 915 (2016).

[68] A. Müller-Hermes, D. Stilck França, and M. M. Wolf, Relative entropy convergence for depolarizing channels, J. Math. Phys. (N.Y.) **57**, 022202 (2016).

[69] A. Müller-Hermes, D. Stilck França, and M. M. Wolf, Entropy production of doubly stochastic quantum channels, J. Math. Phys. (N.Y.) **57**, 022203 (2016).

[70] 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, J. Math. Phys. (N.Y.) **62**, 061901 (2021).

[71] S. Beigi, N. Datta, and C. Rouzé, Quantum reverse hypercontractivity: Its tensorization and application to strong converses, Commun. Math. Phys. **376**, 753 (2020).

[72] Á. Capel, C. Rouzé, and D. Stilck França, The modified logarithmic Sobolev inequality for quantum spin systems: Classical and commuting nearest neighbour interactions, arXiv:2009.11817.

[73] C. Hirche, C. Rouzé, and D. Stilck França, On contraction coefficients, partial orders and approximation of capacities for quantum channels, Quantum **6**, 862 (2022).

[74] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, J. Math. Phys. (N.Y.) **54**, 122203 (2013).

[75] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of n-level systems, J. Math. Phys. (N.Y.) **17**, 821 (1976).

[76] K. Temme, F. Pastawski, and M. J. Kastoryano, Hypercontractivity of quasi-free quantum semigroups, J. Phys. A **47**, 405303 (2014).

[77] K. Temme and M. J. Kastoryano, How fast do stabilizer Hamiltonians thermalize? arXiv:1505.07811.

[78] Á. Capel, A. Lucia, and D. Pérez-García, Quantum conditional relative entropy and quasi-factorization of the relative entropy, J. Phys. A **51**, 484001 (2018).

[79] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete, The chi2-divergence and mixing times of quantum Markov processes, J. Math. Phys. (N.Y.) **51**, 122201 (2010).

[80] QISKIT: Device backend noise model simulations.

[81] K. Tsubouchi, T. Sagawa, and N. Yoshioka, preceding Letter, Universal cost bound of quantum error mitigation based on quantum estimation theory, Phys. Rev. Lett. **131**, 210601 (2023).