ARTICLE   OPEN

Check for updates

Fundamental limits of quantum error mitigation

Ryuji Takagi [1 ✉], Suguru Endo [2 ✉], Shintaro Minagawa [3 ✉] and Mile Gu[1,4 ✉]

The inevitable accumulation of errors in near-future quantum devices represents a key obstacle in delivering practical quantum advantages, motivating the development of various quantum error-mitigation methods. Here, we derive fundamental bounds concerning how error-mitigation algorithms can reduce the computation error as a function of their sampling overhead. Our bounds place universal performance limits on a general error-mitigation protocol class. We use them to show (1) that the sampling overhead that ensures a certain computational accuracy for mitigating local depolarizing noise in layered circuits scales exponentially with the circuit depth for general error-mitigation protocols and (2) the optimality of probabilistic error cancellation among a wide class of strategies in mitigating the local dephasing noise on an arbitrary number of qubits. Our results provide a means to identify when a given quantum error-mitigation strategy is optimal and when there is potential room for improvement.

INTRODUCTION

Recent advances in quantum technologies have resulted in the availability of noisy intermediate-scale quantum (NISQ) devices, promising advantages of quantum information processing by controlling tens to hundreds of qubits[1,2]. However, inevitable noise remains a critical roadblock for their practical use; every gate has a chance of error, and their continuing accumulation will eventually destroy any potential quantum advantage. While quantum error correction enables in-principle means to suppress such error indefinitely, they involve measuring error syndromes and making adaptive corrections. In contrast, NISQ devices often cannot adaptively execute quantum operations.

This technological hurdle has motivated the study of quantum error mitigation, resulting in a diverse collection of alternative techniques (e.g., zero-error noise extrapolation[3–8], probabilistic error cancellation[3,9–13], and virtual distillation[14–19]). All share in common that they avoid adaptive operations. Instead, error-mitigation algorithms suppress errors by sampling available noisy devices many times and classically post-processing these measurement outcomes. Such techniques generally have drastically reduced technological requirements, providing potential near-term solutions for suppressing errors in other NISQ algorithms (e.g., variational algorithms for estimating the ground state energy in quantum chemistry[20–23]).

The performance of these algorithms is typically analyzed on a case-by-case basis. While this is crucial for understanding the value of a particular methodology in a specific practical context, it leaves open a fundamental question: What is the ultimate potential of quantum error mitigation? The motivation to answer this question parallels the development of heat engines. There, Carnot's theorem allows us to understand the ultimate efficiency of all possible heat engines[24], allowing us to know what is physically forbidden and enabling a universal means to understand what specific engines have the greatest room for potential improvement.

Here, we initiate a research program toward characterizing the ultimate limits of quantum error mitigation. We propose a framework to formally define error mitigation as any strategy that requires no adaptive quantum operations (see Fig. 1). We introduce maximum estimator spread as a universal benchmark for error-mitigation performance—a quantity that tells us how many extra runs of a NISQ device guarantee that outputs are within some desired accuracy threshold. We then derive fundamental lower bounds for this spread—that no current or yet-undiscovered error-mitigation strategy can violate. Our bounds are represented in terms of the reduction in the distinguishability of quantum states due to the noise effect, providing an operational understanding of the cost for error mitigation.

We then illustrate two immediate consequences of our general bounds. The first is in the context of mitigating local depolarizing noise in variational quantum circuits[20,25]. We show that the maximum estimator spread grows exponentially with circuit depth for the general error-mitigation protocol, confirming a suspicion that the well-known exponential growing estimation error observed in several existing error-mitigation techniques[3,26] is a consequence of the fundamental obstacle shared by the general error-mitigation strategies. Our second study shows that probabilistic error cancellation—a prominent method of error mitigation—minimizes the maximum estimator spread when mitigating local dephasing noise acting on an arbitrary number of qubits. These results showcase how our bounds can help rule out what error-mitigation performance targets are unphysical, and identify what methods are already near-optimal.

RESULTS

Framework

Our framework begins by introducing a formal definition of error mitigation. Consider an ideal computation described by (1) application of some circuit U to some input ψ_{in} (2) measurement of the output state ψ in some arbitrary observable A (see Fig. 1A). In realistic situations, however, there is noise, such that we have only access to NISQ devices capable of preparing some certain distorted states $\mathcal{E}(\psi)$. The aim is then to retrieve desired output data specified by $\langle A \rangle = \text{Tr}(A\psi)$. Here, we assume $-\mathbb{I}/2 \leq A \leq \mathbb{I}/2$ without loss of generality. This is because any observable O can be
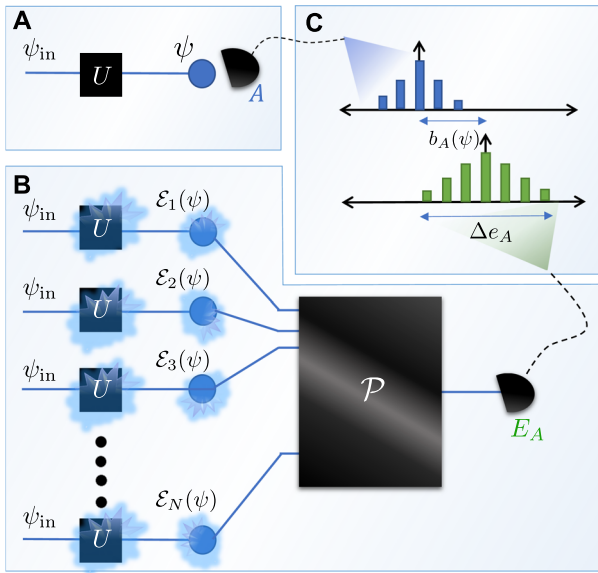
Fig. 1 Quantum error mitigation. A A major goal of many near-term algorithms is to estimate the expectation value of some observable A, when acting on the output ψ of some idealized computation U applied to some input ψ_{in}. **B** However, noise prevents the exact synthesis of ψ. Quantum error-mitigation protocols assist to estimate the true expectation value $\langle A \rangle = \text{Tr}(A\psi)$ without using the adaptive quantum operations necessary in general error correction. This is done by (1) using available NISQ devices to synthesize N distorted quantum states $\{\mathcal{E}_n(\psi)\}_{n=1}^N$ and (2) acting some physical process $\mathcal{P}$ on these distorted states to produce a random variable E_A that approximates A. This procedure can then be repeated over M rounds to draw M samples of E_A, whose mean is used to estimate $\langle A \rangle$. **C** We can characterize the efficacy of such protocol by (1) its spread Δe_A, the difference between maximum and minimum possible values of E_A and (2) the bias $b_A(\psi) = \langle E_A \rangle - \langle A \rangle$. Here we derive ultimate lower bounds on Δe_A for each given bias that no such error-mitigation protocol can exceed, as well as tighter bounds when $\mathcal{P}$ is restricted only to coherent interactions over Q noisy devices at a time. This then tells us how many times $\mathcal{P}$ must be executed to estimate $\langle A \rangle$ within some desired accuracy and failure probability.

shifted and rescaled to some A satisfying this condition, from which full information of O can be recovered. For instance, if we are interested in a non-identity Pauli operator P, which has eigenvalues ± 1, we instead consider an observable $A = P/2$. Note also that while ψ is pure in many practically relevant instances, our analysis applies equally when ψ is mixed.

We consider NISQ devices with no capacity to execute adaptive quantum operations. That is, they cannot enact different quantum operations conditioned on a measurement outcome. We then refer to an algorithm aimed to estimate $\langle A \rangle$ under such constrained devices as an error-mitigation strategy. Each error-mitigation strategy involves sampling NISQ devices configured in N settings for some integer N. Denote the states generated by these configurations by $\mathcal{E}_1(\psi), \dots, \mathcal{E}_N(\psi)$, with effective noise channels $\{\mathcal{E}_i\}_{i=1}^N$, where these effective noise channels can be different from each other in general. The effective noise channel is a non-adaptive operation that connects an ideal state to a distorted state and may be different from the actual noise channel that happens in the NISQ device. Nevertheless, one can always find such an effective noise channel given the descriptions of the actual noise channels and the idealized circuit U. The strategy then further describes some physical process $\mathcal{P}$—which is independent of either the input ψ_{in} or the ideal output ψ—that takes these distorted states as input and outputs some classical estimate random variable E_A of $\text{Tr}(A\psi)$ (see Fig. 1B). The aim is to generate
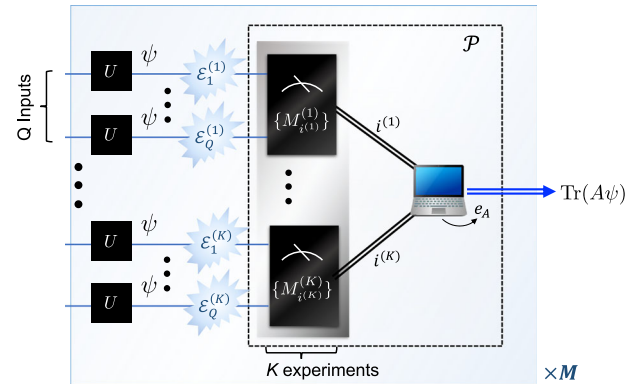

Fig. 2 Schematic of a (Q, K)-error mitigation protocol. A (Q, K)-error mitigation protocol is motivated when practical considerations limit the maximum number of distorted states that our mitigation process $\mathcal{P}$ can coherently interact to Q. A general approach then divides these into $K = \lceil N/Q \rceil$ groups of size Q. To estimate $\langle A \rangle$ of some ideal output state ψ, each round of (Q, K)-mitigation involves first using available NISQ devices to generate Q copies of each distorted states $\mathcal{E}_q^{(k)}(\psi)$, for each of $k = 1, \dots K$. These distorted states are then grouped together as inputs into K experiments, where each group consists of a single copy of each $\mathcal{E}_q^{(k)}(\psi)$. The kth experiment then involves applying some general (possibly entangling) POVM $\{M_{i^{(k)}}^{(k)}\}$ on the kth grouping, resulting in measurement outcome $i^{(k)}$. Classical computing is then deployed to produce an estimate $e_A(i^{(1)}, \dots, i^{(K)})$ whose average after M rounds of the above process is used to estimate $\text{Tr}(A\psi)$. Note that there can be additional quantum operations before the POVM measurements $\{M_{i^{(k)}}^{(k)}\}$, but these can be absorbed into the description of the POVMs without loss of generality.

E_A such that its expected value $\langle E_A \rangle$ is close to $\text{Tr}(A\psi)$. Each round of the protocol involves generating a sample of E_A. M rounds of this procedure then enable us to generate M samples of E_A, whose mean is used to estimate $\text{Tr}(A\psi)$.

Each error-mitigation strategy can then be entirely described by its choice of $\mathcal{P}$ and $\{\mathcal{E}_i\}_{i=1}^N$. Our most fundamental bound pertain to all possible choices. However, we can often make these bounds tighter in situations where further practical limitations constrain how many distorted states $\mathcal{P}$ can coherently interact. Error mitigation protocols under such constraints typically select $N = KQ$ to a multiple of Q, such that the N distorted states are divided into K clusters, each containing Q distorted states. We label these as $\{\mathcal{E}_q^{(k)}(\psi)\}_{q=1, k=1}^{Q, K}$ for convenience. $\mathcal{P}$ is then constrained to represent (1) local measurement procedures $M^{(k)}$ that can coherently interact distorted states within the kth cluster (i.e., $\{\mathcal{E}_q^{(k)}(\psi)\}_{q=1}^Q$) to produce some classical interim outputs $i^{(k)}$ and (2) classical post-processing function e_A that transform the interim outputs $\{i^{(k)}\}_{k=1}^K$ into a sample of E_A.

We name such a protocol as (Q, K)-error mitigation, and refer to the generation of each $i^{(k)}$ as an experiment. Each round of a (Q, K)-error mitigation protocol thus contains K experiments on systems of up to Q distorted states. We also summarize the above procedure in Fig. 2 and give a formal mathematical definition in Methods. Figure 3 and accompanying captions discuss how several prominent error-mitigation methods fit into this framework.

Several comments on our error-mitigation framework are in order. We first note that, for a given set of noisy circuits that result in effective noise channels $\{\mathcal{E}_i\}_{i=1}^N$, our framework assumes to apply an additional process $\mathcal{P}$ after the noisy circuits and does not include processes within the initial noisy circuits. Our framework thus excludes error correction, which employs adaptive processes
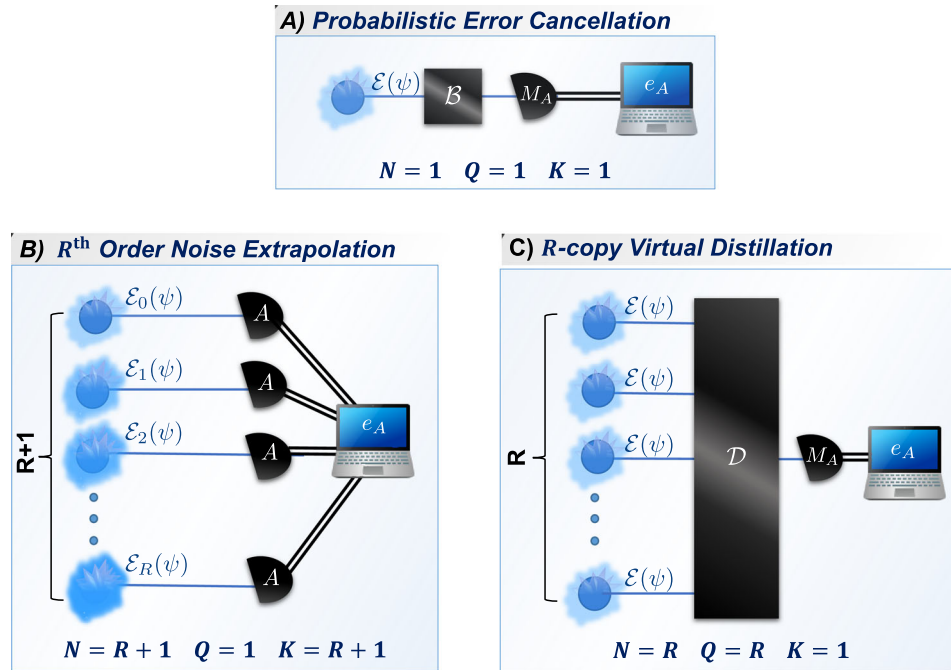
Fig. 3 Error-mitigation protocols. Our framework encompasses all commonly used error-mitigation protocols, a sample of which we outline here. **A** Probabilistic error cancellation[3] assumes we can only act a single coherent state each round, where it seeks to undo a given noise map $\mathcal{E}$ by applying a suitable stochastic operation $\mathcal{B}$. Thus it corresponds to the case of $Q = K = 1$. **B** Rth order noise extrapolation assumes[3,4] the capacity to synthesize $R + 1$ NISQ devices whose outputs represent distortions of ψ at various noise strengths. It then uses individual measurements of an observable A on these distorted states to estimate the observable expectation value on the zero-noise limit. Thus it is an example where $Q = 1$ and $K = R + 1$. **C** Meanwhile, R-copy virtual distillation[14,15] involves running an available NISQ device R times to synthesize R copies of a distorted state $\mathcal{E}(\psi)$. Coherent interaction $\mathcal{D}$ over these copies followed by a suitable measurement M_A then enables improved estimation of $\langle A \rangle$. Thus it is an example where $K = 1$ and $Q = R$. In the main text and Methods, we provide a detailed account of each protocol and how it fits within our framework.

integrated into noisy circuits. This allows our framework to differentiate error mitigation from error correction and makes it useful to investigate the limitations imposed particularly on the former.

One might think that this would overly restrict the scope of error mitigation, which could also use some processes in noisy circuits. This can be avoided by considering that such processes are already integrated into the description of effective noise channels $\{\mathcal{E}_i\}_{i=1}^N$. In other words, the effective noise channel can be considered as a map that connects an ideal state to a distorted state affected by not only a noise channel but non-adaptive processes accessible to a given near-term device; the error mitigation process $\mathcal{P}$ is then an additional process that follows them. This is manifested in the Rth order noise extrapolation in Fig. 3B, in which R different noise levels realized on a near-term device are represented by the set $\{\mathcal{E}_i\}_{i=1}^R$ of effective noise channels.

More broadly, taking appropriate effective noise channels allows our framework to include error-mitigation protocols that employ modified circuits. Namely, if $\{\mathcal{N}_i\}_{i=1}^N$ are the noisy circuits that an error-mitigation protocol employs and $\mathcal{U}$ is the ideal circuit, then such an error-mitigation strategy is encompassed in our framework with $\mathcal{E}_i = \mathcal{N}_i \circ \mathcal{U}^\dagger$. This, for instance, includes the conventional strategy of probabilistic error cancellation applied to a noisy circuit, in which a probabilistic operation is applied after every noisy gate.

We also remark that our framework leaves the freedom of how to choose the round number M and the sample number $N = KQ$ per round for a given shot budget; if the total shot budget is T, one is free to choose any N and M such that $T = NM$. As we describe shortly, our results in Theorem 1 and Corollary 2 are concerned with the number of rounds M, and they apply to any

choice of shot allocation. However, our results become most informative by choosing as large M (equivalently, as small N) as possible. The strategies in Fig. 3 admit small N's that do not scale with the total shot budget, representing examples for which our results give fruitful insights into their round number M. On the other hand, some strategies that employ highly nonlinear computation on the measurement outcomes (e.g., exponential noise extrapolation[11], subspace expansion[27]) require a large N, in which case our results on the round number M can have a large gap from the actual sampling cost.

Our framework also allows one to assume some pre-knowledge prior to the error-mitigation process. For instance, this includes the information about the underlying noise or some pre-computation that error-mitigation process can use in its strategy. The results in Theorem 1 and Corollary 2 then give information about the round number M given such pre-knowledge. Since the process of obtaining the pre-knowledge itself may be considered as a part of error-mitigation process, there are many possible divisions between the pre-computation and the error-mitigation process. Our results apply to any choice of pre-knowledge, and this can be flexibly chosen depending on one's interest. For instance, R-copy virtual distillation can be considered as a $(R, 1)$-error mitigation (that is, $N = R$) as in Fig. 3C under the pre-knowledge of an eigenvalue of the noisy state, which is one of the settings discussed in ref. [14] (see also Methods). This pre-knowledge allows for a small choice of N, making the estimation of the round number M by our method insightful. Another example includes the Clifford Data Regression[28], which can employ a linear regression based on a pattern learned from a training set. By considering the first learning step as the pre-computation, our results provide a meaningful bound for the sampling cost in the

latter stage in which the output from the circuit of interest is compared to the model estimated from the training set.

Up to the flexibility described above, our framework encompasses a broad class of error-mitigation strategies proposed so far[3,4,11,14,15,27–32].

Quantifying performance

The performance of an error-mitigation protocol is determined by how well the random variable E_A governing each estimate aligns with $\mathrm{Tr}(A\psi)$. We can characterize this by (1) its bias, representing how close $\langle E_A \rangle$ is to the ideal expectation value $\mathrm{Tr}(A\psi)$ and (2) its spread, representing the amount of intrinsic randomness within E_A.

A protocol's bias quantifies the absolute minimum error with which it can estimate $\mathrm{Tr}(A\psi)$, given no restrictions on how many rounds it can run (i.e., samples of E_A it can draw). Mathematically, this is represented by the difference $b_A(\psi) = \langle E_A \rangle - \mathrm{Tr}(A\psi)$. Since the error-mitigation strategy should work for an arbitrary state ψ and observable A, we can introduce the maximum bias:

$$b_{\max} := \max_{-\mathbb{I}/2 \leq A \leq \mathbb{I}/2} \max_\psi |\langle E_A \rangle - \mathrm{Tr}(A\psi)| \tag{1}$$

to bound the bias of an error-mitigation protocol in estimating expectation values over all output states and observables of interest. Hereafter, we will also assume $b_{\max} \leq 1/2$, as this condition must be satisfied for any meaningful error-mitigation protocol. This is because a maximum bias of 1/2 can always be achieved by the trivial "error-mitigation" protocol that outputs $e_A = 0$ regardless of ψ or A.

Of course, having $b_{\max} = 0$ still does not guarantee an effective error-mitigation protocol. Each sample of E_A will also deviate from $\mathrm{Tr}(A\psi)$ due to intrinsic random error. The greater this randomness, the more samples we need from E_A to ensure that the mean of our samples is a reliable estimate of its true expectation value $\langle E_A \rangle$. The relation is formalized by Hoeffding's inequality[33]. Namely, suppose $\{x_i\}_{i=1}^M$ are M samples of a random variable X with $x_i \in [a, b]$, the number M of samples that ensures an estimation error $|\langle X \rangle - \sum_i x_i/M| < \delta$ with probability $1 - \varepsilon$ is given by $\frac{|a-b|^2}{2\delta^2} \log(2/\varepsilon) \propto |a - b|^2$. In our context, the latter quantity corresponds to the maximum spread in the outcomes of estimator function e_A defined by:

$$\Delta e_{\max} := \max_{-\mathbb{I}/2 \leq A \leq \mathbb{I}/2} \Delta e_A, \tag{2}$$

where Δe_A is the difference between the maximum and minimum possible values that E_A can take, i.e., $\Delta e_A := e_{A,\max} - e_{A,\min}$ where $e_{A,\max} := \max_{i^{(1)}\ldots i^{(K)}} e_A(i^{(1)}\ldots i^{(K)})$ and $e_{A,\min} := \min_{i^{(1)}\ldots i^{(K)}} e_A(i^{(1)}\ldots i^{(K)})$.

$\Delta e_{\max}$ thus directly relates to the sampling cost of an error-mitigation protocol. Given an error-mitigation protocol whose estimates have maximum spread $\Delta e_{\max}$, it uses sample E_A of order $\mathcal{O}(\Delta e_{\max}^2 \log(1/\varepsilon)/\delta^2)$ times to ensure that its estimate of $\langle E_A \rangle$ has accuracy δ and failure rate ε. Therefore, we may think of $\Delta e_{\max}$ as a measure of computational cost or feasibility. Its exponential scaling with respect to the circuit depth, for example, would imply eventual intractability in mitigating associated errors in a class of non-shallow circuits.

We note that if the variance of E_A happens to be small, the actual sampling cost required to achieve the accuracy δ and failure rate ε can be smaller than the estimate based on the maximum spread. In this sense, $\Delta e_{\max}$ quantifies the round number M that one would practically use in the worst-case scenario. However, knowing the variance of E_A beforehand is a formidable task in general, and the worst-case estimate gives a useful benchmark to assess the feasibility of a given error-mitigation strategy in such situations.

Fundamental limits

Our main contribution is to establish a universal lower bound on $\Delta e_{\max}$. Our bound then determines the number of times an error-mitigation method samples E_A (and thus the number of times we invoke a NISQ device) to estimate A within some tolerable error.

To state the bound formally, we utilize measures of state distinguishability. Consider the scenario where Alice prepares a quantum state in either ρ and σ and challenges Bob to guess which is prepared. The trace distance $D_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$ (where $\|\cdot\|_1$ is the trace norm) then represents the quantity such that Bob's optimal probability of guessing correctly is $\frac{1}{2}(1 + D_{\mathrm{tr}}(\rho, \sigma))$. When ρ and σ describe states on K-partite systems $S_1 \otimes \cdots \otimes S_K$, we can also consider the setting in which Bob is constrained to local measurements, resulting in the optimal guessing probability $\frac{1}{2}(1 + D_{\mathrm{LM}}(\rho, \sigma))$ where D_{LM} is the local distinguishability measure[34] (see also Methods). In our setting, we identify each local subsystem S_k with a system corresponding to the kth experiment in Fig. 2. We are then in a position to state our main result:

Theorem 1 *Consider an arbitrary (Q, K)-mitigation protocol with maximum bias $b_{\max}$. Then, its maximum spread $\Delta e_{\max}$ is lower bounded by:*

$$\Delta e_{\max} \geq \max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi) - 2b_{\max}}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)} \tag{3}$$

where $\tilde{\psi}_Q^{(K)} := \otimes_{k=1}^K \otimes_{q=1}^Q \left[\mathcal{E}_q^{(k)}(\psi)\right]$ and $\tilde{\phi}_Q^{(K)} := \otimes_{k=1}^K \otimes_{q=1}^Q \left[\mathcal{E}_q^{(k)}(\phi)\right]$ are distorted states corresponding to the QK copies of some ideal outputs ψ and ϕ, and $\mathcal{E}_q^{(k)}$ is the effective noise channel for the qth input in the kth experiment.

Combining this with Hoeffding's inequality leads to the following bound on the sampling cost.

Corollary 2 *Consider an arbitrary (Q, K)-mitigation protocol with maximum bias $b_{\max}$. Then, an estimation error of $b_{\max} + \delta$ is realized with probability $1 - \varepsilon$ when the number of samples M satisfies:*

$$M \geq \frac{\Delta e_{\max}^2 \log(2/\varepsilon)}{2\delta^2}$$
$$\geq \left[\max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi) - 2b_{\max}}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)}\right]^2 \frac{\log(2/\varepsilon)}{2\delta^2} \tag{4}$$

where $\tilde{\psi}_Q^{(K)} := \otimes_{k=1}^K \otimes_{q=1}^Q \left[\mathcal{E}_q^{(k)}(\psi)\right]$ and $\tilde{\phi}_Q^{(K)} := \otimes_{k=1}^K \otimes_{q=1}^Q \left[\mathcal{E}_q^{(k)}(\phi)\right]$.

Theorem 1 and Corollary 2 offer two qualitative insights. The first is the potential trade-off between sampling cost and systematic error—we may reduce the sampling cost by increasing tolerance for bias. The second is a direct relation between sampling cost and distinguishability—the more a noise source degrades distinguishability between states, the more costly the error is to mitigate.

The intuition behind this relation rests on the observation that the error-mitigation process is a quantum channel. Thus, any error-mitigation procedure must obey data-processing inequalities for distinguishability. On the other hand, error mitigation aims to improve our ability to estimate expectation values of various observables, which would enhance our ability to distinguish between noisy states. The combination of these observations then implies that distinguishability places a fundamental constraint on required sampling costs to mitigate error. For details of the associated proof, see Methods.

Observe that our bound involves the local distinguishability $D_{\mathrm{LM}}(\rho, \sigma)$ rather than the standard trace distance $D_{\mathrm{tr}}(\rho, \sigma)$. This is due to the constraints we placed of $\mathcal{P}$ that limits it to coherently interacting the outputs of a finite number of NISQ devices—reflecting the hybrid nature of quantum error mitigation utilizing quantum and classical resources in tandem. Notably, these quantities coincide for the most powerful NISQ devices (the ones allowing coherent interactions between all N noisy initial states).

This case then corresponds to the most fundamental bound:

$$\Delta e_{\max} \geq \max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi) - 2b_{\max}}{D_{\mathrm{tr}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)}, \quad (5)$$

which represents the ultimate performance limits of all (Q, K) error-mitigation protocols that coherently operate on $N = QK$ distorted states each round.

We also remark that our framework can give tighter bounds when available error-mitigation methods involve specific states and observables (see Eq. (36)).

Alternative bounds

While the bounds derived above in terms of distinguishability have a clear operational meaning, its evaluation in realistic settings can face two significant hurdles. (1) It involves evaluating the distinguishability between two quantum states whose dimensions scale exponentially with KQ, making its evaluation costly for protocols that require many NISQ samples per round. (2) It requires that we have tomographic knowledge of the effective noise channels $\mathcal{E}_q^{(k)}$.

One potential means around this is to identify bounds on the distinguishability measures that alleviate such hurdles. For example, since $D_{\mathrm{tr}}(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)}$ for any pair of states ρ and σ where $F(\rho, \sigma) := \left(\mathrm{Tr}\sqrt{\sigma^{1/2}\rho\sigma^{1/2}}\right)^2$ is the (squared) fidelity[35], this, together with Eq. (5), implies:

$$\Delta e_{\max} \geq \max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi) - 2b_{\max}}{\sqrt{1 - \prod_{q=1}^Q \prod_{k=1}^K F\left(\mathcal{E}_q^{(k)}(\psi), \mathcal{E}_q^{(k)}(\phi)\right)}}. \quad (6)$$

This form only involves the computation of the trace distance and fidelity of single-copy states, both of which can be computed by semidefinite programming[36].

Meanwhile, the need for tomographic knowledge of $\mathcal{E}_q^{(k)}$ can be mitigated by using subfidelity[37]:

$$E(\rho, \sigma) := \mathrm{Tr}(\rho\sigma) + \sqrt{2\left[\{\mathrm{Tr}(\rho\sigma)\}^2 - \mathrm{Tr}(\rho\sigma\rho\sigma)\right]}. \quad (7)$$

The subfidelity bounds $F(\rho, \sigma)$ from below, and thus also lower bounds the maximum spread:

$$\Delta e_{\max} \geq \max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi) - 2b_{\max}}{\sqrt{1 - \prod_{q=1}^Q \prod_{k=1}^K E\left(\mathcal{E}_q^{(k)}(\psi), \mathcal{E}_q^{(k)}(\phi)\right)}}. \quad (8)$$

subfidelity between two unknown states can be measured by a quantum computer using a circuit of constant depth[38,39] (see also

Methods). This obviates the need for tomographical data, while its low depth means that the noise in this process is typically much smaller than the noise in our circuits of interest. We remark that, instead of using the subfidelity, one could use an alternative quantity that lower bounds the fidelity that can be estimated by NISQ devices, e.g., truncated fidelity[40]. Such techniques could enable benchmarking protocols that allow us to rule out a candidate NISQ device should our bounds suggest their error profiles are too adverse to support any viable means of error mitigation.

In addition, the maximum in the right-hand sides of (6) and (8) do not need to be evaluated exactly; any choice of states ψ and ϕ provides a valid lower bound for the maximum spread. While these alternative bounds may not be as tight, they still serve as universal lower bounds that can put non-trivial constraints on the error-mitigation performance (see Remark 2 in Supplementary Note 1 and Supplementary Note 3).

Error-mitigating layered circuits

Quantitatively, the above bounds enable us to determine the ultimate performance limits of error mitigation given a particular set of imperfect quantum devices specified by error channels $\{\mathcal{E}_q^{(k)}\}$. We now illustrate how this enables the identification of sampling overheads when performing error mitigation on a common class of NISQ algorithms—layered circuits used extensively in variational quantum eigensolvers[41]. Variational algorithms typically assume a quantum circuit consisting of multiple layers of unitary gates $\{U_l\}_{l=1}^L$ acting on an n-qubit system. Indeed, as designed with NISQ applications in mind, they are key candidates for benchmarking of error-mitigation protocols[7,42,43].

In particular, consider a local depolarizing noise[25,44], in which the depolarizing channel $\mathcal{D}_\epsilon(\rho) := (1 - \epsilon)\rho + \epsilon\mathbb{I}/2$ acts on each qubit. A general approach to mitigate this error is to employ a (Q, K)-mitigation protocol for some Q and K, in which the kth experiment involves depolarizing noise with noise strength ϵ_k (Fig. 4).

Taking $U = U_L \cdots U_2 U_1$ in Fig. 2 and applying Theorem 1 to this setting, we obtain the following bound (see Supplementary Note 1 for the proof).

Theorem 3 *For an arbitrary (Q, K)-error mitigation with maximum bias $b_{\max}$ applied to n-qubit circuits with L-layer unitaries under local depolarizing noise, the maximum spread is lower bounded as:*

$$\Delta e_{\max} \geq \frac{1 - 2b_{\max}}{\sqrt{2\ln 2}\sqrt{nQ}\,K}\left(\frac{1}{1 - \epsilon_{\min}}\right)^L, \quad (9)$$

where $\epsilon_{\min} := \min_k \epsilon_k$ is the minimum noise strength among K experiments.
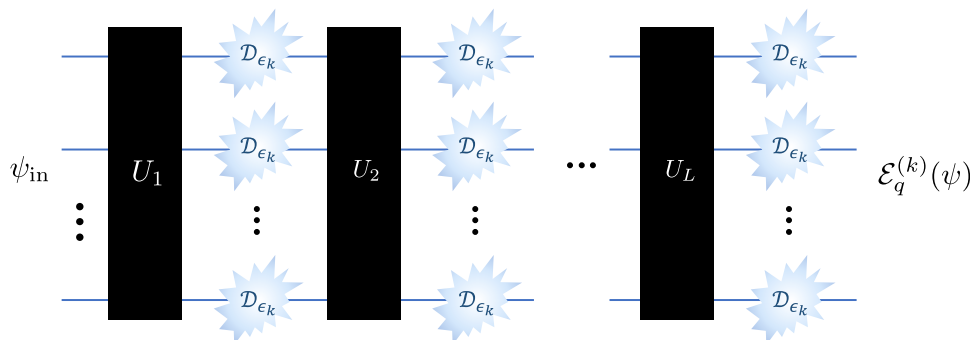

Fig. 4 Noise mitigation in layered circuits. Layered circuits are used extensively in variational algorithms for NISQ devices. They involve repeated layers of gates, each consisting of some unitary U_l. A standard noise model for such circuits involves the action of local depolarizing noise $\mathcal{D}_\epsilon$ on each qubit during each layer of the circuit. The kth experiment in a general (Q, K)-protocol involves running this circuit Q times to produce a distorted state $\otimes_{q=1}^Q \mathcal{E}_q^{(k)}(\psi)$ with some noise strength ϵ_k—which possibly varies over different experiments. The protocol then measures each $\otimes_{q=1}^Q \mathcal{E}_q^{(k)}(\psi)$ for $k = 1, \ldots, K$ and outputs an estimate E_A through classical post-processing of the measurements results.

Theorem 3 suggests that error-mitigation strategies encompassed in our framework will use exponentially many samples with respect to the circuit depth L. This validates our intuition that information should quickly get degraded due to the sequential noise effects, incurring exponential overhead to remove the accumulated noise effect.

We also remark that, although we here focus on the exponential growth of the maximum spread with respect to the circuit depth L for the sake of generality, one can expect that the maximum spread grows exponentially with the total gate number $nQKL$ rather than just the layer number L in many practical cases.

Protocol benchmarking

Theorems 1 and 3 place strategy-independent bounds on the maximum spread for each Q and K and available noise channels $\mathcal{E}_q^{(k)}$, enabling us to identify the ultimate potential of error mitigation under various noise settings and operational constraints. Comparing this limit with that achieved by specific known methods of error mitigation then provides a valuable benchmark, helping us assess their optimality and quantify the potential room for improvement. We illustrate this here by considering probabilistic error cancellation[3], while we discuss how our framework can be applied to other prominent error-mitigation protocols in Methods.

Probabilistic error cancellation is an error-mitigation protocol that produces an estimate of $\mathrm{Tr}(A\psi)$ using a distorted state $\mathcal{E}(\psi)$ each round (see Fig. 3A). It then fulfills the criteria of being a $(1, 1)$-protocol, i.e., $Q = K = 1$. Here, we assume that the description of the noise channels is given as pre-knowledge, in which case the estimator becomes unbiased, i.e., $b_{max} = 0$. Probabilistic error cancellation operates by identifying a complete basis of processes $\{\mathcal{B}_j\}_j$ such that $\mathcal{E}^{-1} = \sum_j c_j \mathcal{B}_j$ for some set of real (but possibly negative) numbers $\{c_j\}_j$. Setting $\gamma := \sum_j |c_j|$, the protocol then (1) applies $\mathcal{B}_j$ to the noisy state $\mathcal{E}(\psi)$ with probability $p_j = |c_j|/\gamma$, (2) measures A to get outcome a_j, and (3) multiplies each outcome by $\gamma \, \mathrm{sgn}(c_j)$ and takes the average.

In the context of our framework, we can introduce a quantum operation $\mathcal{B}$ that represents first initializing a classical register to a state j with probability p_j and applying $\mathcal{B}_j$ to $\mathcal{E}(\psi)$ conditioned on j. Meanwhile, M_A represents an A-measurement of the resulting quantum system combined with a measurement of the register, resulting in the outcome pair (a_j, j). Taking $e_A^{\mathrm{PEC}}((a_j, j)) = \gamma \, \mathrm{sgn}(c_j) a_j$, we see that the maximum spread of this estimator is given by:

$$\Delta e_{max}^{\mathrm{PEC}} = \gamma, \tag{10}$$

a well-studied quantity that is already associated with the sampling overhead of probabilistic error cancellation[3].

The optimal sampling cost γ_{opt} is then achieved by minimizing such γ over all feasible $\{\mathcal{B}_j\}_j$[45]. Once computed for a specific noise channel $\mathcal{E}$, we can compare it to the lower bounds in Theorem 1 to determine if there is possible room for improvement.

Let us now consider local dephasing noise on an n-qubit system, where the dephasing noise $\mathcal{Z}_\epsilon(\rho) := (1 - \epsilon)\rho + \epsilon Z\rho Z$ acts on each qubit. We find that the optimal cost is obtained as:

$$\gamma_{opt} = \Delta e_{max}^{\mathrm{PEC}} = \frac{1}{(1 - 2\epsilon)^n}. \tag{11}$$

This can be compared to the bound for Δe_{max} from Theorem 1 that applies to every mitigation protocol with $Q = K = 1$. Note that, since $K = 1$, $D_{\mathrm{LM}} = D_{\mathrm{tr}}$. We then get:

$$\max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi)}{D_{\mathrm{tr}}(\mathcal{Z}_\epsilon(\psi), \mathcal{Z}_\epsilon(\phi))} \geq \frac{1}{(1 - 2\epsilon)^n}. \tag{12}$$

Detailed computation to obtain (11) and (12) can be found in Supplementary Note 2. Remarkably, the two quantities—the maximum spread for the probabilistic error cancellation and the lower bound for arbitrary unbiased mitigation strategies with
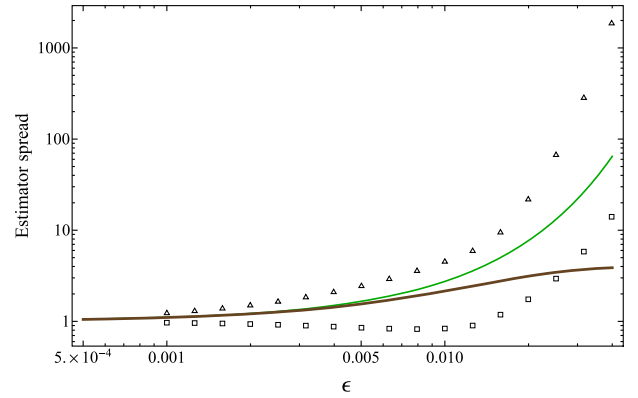

Fig. 5 The estimator spreads to mitigate local dephasing noise on a 50-qubit system. Solid green curve: Δe_{max} for probabilistic error cancellation and the lower bound for unbiased $(1, 1)$-mitigation protocols, which coincide as explained in the main text. Brown curve: Δe_A with $A = \frac{1}{2} \otimes_{i=1}^n X_i$ for 2-copy virtual distillation with GHZ state inputs and the lower bound for $(2, 1)$-mitigation protocols with the same bias, which coincide as explained in Supplementary Note 2. Triangles and rectangles: Δe_A with $A = \frac{1}{2} \otimes_{i=1}^n X_i$ for 11th order noise extrapolation with GHZ state inputs (triangles) and a lower bound for $(1, 12)$-mitigation protocols with the same bias (rectangles).

$Q = K = 1$—exactly coincide. This shows that probabilistic error cancellation achieves the ultimate performance limit of unbiased $(1, 1)$-protocols for correcting local dephasing noise for an arbitrary qubit number n.

We can also consider the d-dimensional depolarizing noise $\mathcal{D}_\epsilon^d(\rho) = (1 - \epsilon)\rho + \epsilon \mathbb{I}/d$. The bound from Theorem 1 for this noise is obtained as:

$$\max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi)}{D_{\mathrm{tr}}(\mathcal{D}_\epsilon^d(\psi), \mathcal{D}_\epsilon^d(\phi))} = \frac{1}{1 - \epsilon}, \tag{13}$$

which is slightly lower than $\Delta e_{max}^{\mathrm{PEC}} = \frac{1 + (1 - 2/d^2)\epsilon}{1 - \epsilon}$[45–47], with difference being $O(\epsilon)$. This suggests that probabilistic error cancellation is nearly optimal for this noise model, while still leaving the possibility for a better protocol to exist.

We can also apply similar techniques to study the performance of other prominent error-mitigation protocols. Here, we plot the estimator spread for probabilistic error cancellation, virtual distillation, and noise extrapolation, and their corresponding lower bounds for local dephasing noise (Fig. 5) and global depolarizing noise (Fig. 6). We note that, for virtual distillation and extrapolation, we evaluated (36) that allows us to bound Δe_A in (2) with a specific observable A of interest. We provide details for the evaluation of these values in Supplementary Note 2. We can observe that both protocols perform near-optimal limits at the low-error regime. At the high-error regime, their performance can diverge significantly from our lower bounds depending on underlying noise models and mitigation strategies. We emphasize that such divergences are expected because of the high generality of our lower bounds. Narrowing the gaps between the fundamental lower bounds and achievable maximum spread, e.g., finding more examples such as probabilistic error cancellation for local dephasing noise, will be a natural direction for future work.

DISCUSSION

Our work aimed to identify the ultimate performance limits of quantum error mitigation—a large class of techniques designed to estimate the outputs of ideal quantum circuits by post-processing measurement data from imperfect counterparts. This involved identifying a universal performance measure—applicable to any such error-mitigation protocols—that captures how many extra
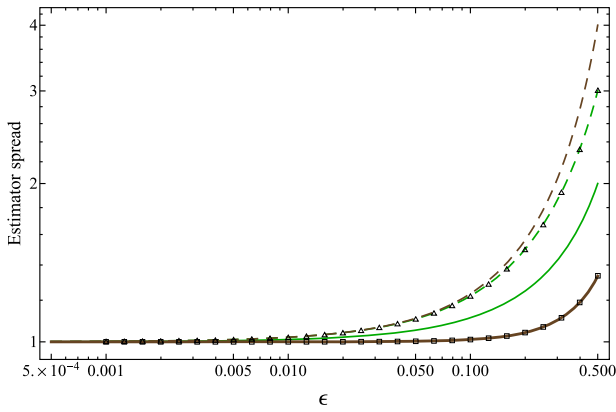
Fig. 6 The estimator spreads to mitigate global depolarizing noise on a 50-qubit system. Green curves: $\Delta e_{\max}$ for probabilistic error cancellation (dashed) and the lower bound for unbiased (1, 1)-mitigation protocols (solid). Brown curves: Δe_A with $A = \frac{1}{2} \otimes_{i=1}^{n} X_i$ for 2-copy virtual distillation with GHZ state inputs (dashed) and the lower bound for (2, 1)-mitigation protocols with the same bias (solid). Triangles and rectangles: Δe_A with $A = \frac{1}{2} \otimes_{i=1}^{n} X_i$ for 1st order noise extrapolation (triangles) and a lower bound for (1, 2)-mitigation protocols with the same bias (rectangles).

executions of available NISQ devices the protocol uses to ensure that its estimates are sufficiently close with some required probability of success. We then derived ultimate performance limits that pertain to all such error mitigation methods. The significance of our bounds parallels that of various fundamental converse bounds in quantum information (e.g., quantum communication[48–50] and thermodynamics[51–53]), representing the ultimate performance limits that quantum error-mitigation protocols can never surpass. Our bounds particularly demonstrate that probabilistic error cancellation is optimal in the maximum spread to mitigate local dephasing noise among all unbiased error-mitigation protocols that involve no coherent interactions between multiple copies of distorted states, and imply that the exponential growth in the maximum spread on mitigating noise in layered circuits is an unavoidable feature shared by the general error-mitigation protocols.

We note that our performance bounds have focused on the scaling of M, representing how many rounds an error-mitigation protocol should be run to get a reliable estimate of some observable $\langle A \rangle$. Although this analysis is sufficient for many present methods of error mitigation, it is possible to also improve estimates of $\langle A \rangle$ by scaling the number of distorted outputs we process in a single round (e.g., extrapolation[11] and subspace expansion[27]). While our framework in Fig. 1 encompasses such methodologies—and as such all bounds on estimation error apply—full understanding of the performance of such protocols would involve further investigation on how estimation error scales with respect to N or K. This then presents a natural direction for future research.

Our results also offer potential insights into several related fields. Non-Markovian dynamics have shown promise in decreasing sampling costs in error mitigation[54]. Since non-Markovianity is known to be deeply related to the trace distance[55], our newly established relations between trace distance and quantum error mitigation hint at promising relations between the two fields. The second direction is to relate our general framework of quantum error mitigation to the established theory of quantum error correction. Quantum error correction concerns algorithms that prevent degrading the trace distance between suitably encoded logical states, while our results indicate that less reduction in trace distance can enable smaller error mitigation costs. Thus, our work provides a toolkit for identifying fundamental bounds in the transition from error mitigation to error correction as we proceed

from NISQ devices toward scalable quantum computing. This then complements presently active research in error suppression that combines the two techniques[56–59]. Beyond error suppression, quantum protocols in many diverse settings also share the structure of classical post-processing of quantum measurements—from quantum metrology and illumination to hypothesis testing and stochastic analysis[60–64]. Our framework—suitably extended—could thus identify new performance bounds in each of these settings.

METHODS

Formal definition of (Q, K)-error mitigation

Here, we give a formal definition of (Q, K)-error mitigation as a quantum operation. Since POVM measurements in different experiments are independent of each other, the whole measurement process can be represented as a tensor product of each POVM. Then, the classical post-processing following the measurement is a classical-classical channel such that the expected value of the output will serve as an estimate of the desired expectation value. We can then formalize an error-mitigation process as a concatenation of these two maps.

Definition 4 ((Q, K)-error mitigation). For an arbitrary observable A satisfying $-\mathbb{I}/2 \leq A \leq \mathbb{I}/2$, a (Q, K)-mitigation protocol—involving Q inputs and K experiments—is a concatenation of quantum-classical channel Λ_A and classical-classical channel $\hat{e}_A$ as $\hat{e}_A \circ \Lambda_A$. Here, Λ_A has a form:

$$\Lambda_A(\cdot) = \sum_{\mathbf{i}} \mathrm{Tr}(\cdot\, M_{i^{(1)}}^{(1)} \otimes \cdots \otimes M_{i^{(K)}}^{(K)}) |\mathbf{i}\rangle \langle \mathbf{i}| \tag{14}$$

where $\{M_{i^{(k)}}^{(k)}\}$ is the POVM for the kth experiment acting on Q copies of n-qubit noisy states, and $\mathbf{i} := i^{(1)} \ldots i^{(K)}$ denotes a collection of measurement outcomes with $|\mathbf{i}\rangle = |i^{(1)} \ldots i^{(K)}\rangle$ being a classical state acting on K subsystems. The channel $\hat{e}_A$ implements a K-input classical function e_A such that:

$$\sum_{\mathbf{i}} p_{\mathbf{i}} e_A(\mathbf{i}) = \mathrm{Tr}(A\psi) + b_A(\psi) \tag{15}$$

for some function $b_A(\psi)$ called bias, and:

$$p_{\mathbf{i}} := \prod_{k=1}^{K} \mathrm{Tr}[\mathcal{E}_1^{(k)}(\psi) \otimes \cdots \otimes \mathcal{E}_Q^{(k)}(\psi) M_{i^{(k)}}^{(k)}] \tag{16}$$

is the probability of getting outcomes $\mathbf{i} = i^{(1)} \ldots i^{(K)}$ for the input noisy states $\{\mathcal{E}_q^{(k)}(\psi)\}_{q=1,k=1}^{Q,K}$.

Proof of Theorem 1—The intuition behind Theorem 1 lies in the intimate relation between the effect of error mitigation and distinguishability of quantum states. Recall that the goal of quantum error mitigation is to estimate the expectation value of an arbitrary observable A for an arbitrary ideal state ψ only using the noisy state $\mathcal{E}(\psi)$. Although $\mathrm{Tr}(A\mathcal{E}(\psi))$ can deviate from $\mathrm{Tr}(A\psi)$, error mitigation correctly allows us to estimate $\mathrm{Tr}(A\psi)$, which appears to have eliminated noise effects. Since each error-mitigation strategy should also work for another state ϕ, it should be able to remove the noise and estimate $\mathrm{Tr}(A\phi)$ out of $\mathrm{Tr}(A\mathcal{E}(\phi))$. Does this "removal" of noise imply that error mitigation can help distinguish $\mathcal{E}(\psi)$ and $\mathcal{E}(\phi)$?

The subtlety of this question can be seen by looking at how quantum error mitigation works. The estimation of $\mathrm{Tr}(A\mathcal{E}(\psi))$ without error mitigation is carried out by making a measurement with respect to the eigenbasis of $A = \sum_a a |a\rangle \langle a|$, which produces a probability distribution $p(a|\mathcal{E}(\psi), A)$ over possible outcomes $\{a\}$. Because of the noise, the expectation value of this distribution is shifted from $\mathrm{Tr}(A\psi)$. Similarly, the same measurement for a state $\mathcal{E}(\phi)$ produces a probability distribution $p(a|\mathcal{E}(\phi), A)$, whose expectation value may also be shifted from $\mathrm{Tr}(A\phi)$. An error-mitigation protocol applies additional operations, measurements and classical post-processing to produce other probability distributions $p_{\mathrm{EM}}(a|\mathcal{E}(\psi), A)$ and $p_{\mathrm{EM}}(a|\mathcal{E}(\phi), A)$ whose expectation values get closer to the original ones. As a result, although the expectation values of the two error-mitigated distributions get separated from each other, they also get broader, which may increase the overlap between the two distributions, possibly making it even harder to distinguish two distributions (see Fig. 7).

One can see that this intuition that error mitigation does not increase the distinguishability is indeed right by looking at the whole error-mitigation process as a quantum channel. Then, the data-processing inequality implies that the distinguishability between any two states should not be increased by the application of quantum channels. This motivates us to rather use this observation as a basis to put a lower bound for the necessary overhead.
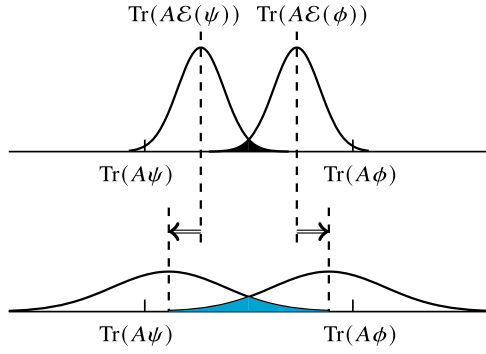
Fig. 7 Error mitigation and distinguishability. The top schematic illustrates the probability distribution of an observable A for two noisy states $\mathcal{E}(\psi)$ and $\mathcal{E}(\phi)$. The expectation values are shifted from the true values due to the noise effects. As in the bottom schematic, error mitigation converts them to other distributions whose expectation values are closer to the true values than the initial noisy distributions are. However, the converted distributions get broader, and the overlap between two distributions increases in general.

Let us recall that the trace distance admits the following form:

$$D_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$$
$$= \max_{0 \leq M \leq \mathbb{I}} \mathrm{Tr}[M(\rho - \sigma)], \tag{17}$$

and similarly the local distinguishablity measure can be written as[34]:

$$D_{\mathrm{LM}}(\rho, \sigma) = \max_{\{M_i\} \in \mathrm{LM}} \frac{1}{2}\|\mathcal{M}(\rho) - \mathcal{M}(\sigma)\|_1$$
$$= \max_{\{M, \mathbb{I} - M\} \in \mathrm{LM}_2} \mathrm{Tr}[M(\rho - \sigma)] \tag{18}$$

where LM is the set of POVMs that take the form $M_{j^{(1)}}^{(1)} \otimes \cdots \otimes M_{j^{(K)}}^{(K)}$, where $M_{j^{(k)}}^{(k)}$ represents some POVM local to system S_k, and LM_2 is the set of two-outcome measurements realized by local measurements together with classical post-processing. The second forms for the above measures particularly tell that they quantify how well two states can be distinguished by accessible quantum measurements. By definition, it is clear that:

$$D_{\mathrm{tr}}(\rho, \sigma) \geq D_{\mathrm{LM}}(\rho, \sigma) \tag{19}$$

for all states ρ and σ, and the inequality often becomes strict[65,66].

The local distinguishability measure satisfies the data-processing inequality under all local measurement channels. Namely, for all states ρ and σ defined on a composite system $\otimes_{k=1}^{K} S_k$, and for an arbitrary quantum-classical channel $\Lambda(\cdot) = \sum_i \mathrm{Tr}\left(\cdot M_{j^{(1)}}^{(1)} \otimes \cdots \otimes M_{j^{(K)}}^{(K)}\right)|i\rangle\langle i|$,

$$D_{\mathrm{LM}}(\Lambda(\rho), \Lambda(\sigma)) = \max_{\mathcal{M} \in \mathrm{LM}} \frac{1}{2}\|\mathcal{M} \circ \Lambda(\rho) - \mathcal{M} \circ \Lambda(\sigma)\|_1$$
$$\leq \max_{\mathcal{M} \in \mathrm{LM}} \frac{1}{2}\|\mathcal{M}(\rho) - \mathcal{M}(\sigma)\|_1 \tag{20}$$
$$= D_{\mathrm{LM}}(\rho, \sigma)$$

where in the inequality we used that the set of local measurement channels is closed under concatenation.

Let us define:

$$\tilde{\psi}_Q^{(K)} := \otimes_{k=1}^{K} \otimes_{q=1}^{Q} \left[\mathcal{E}_q^{(k)}(\psi)\right],$$
$$\tilde{\phi}_Q^{(K)} := \otimes_{k=1}^{K} \otimes_{q=1}^{Q} \left[\mathcal{E}_q^{(k)}(\phi)\right]. \tag{21}$$

Since the channel Λ_A in Definition 4 is a local measurement channel, we employ (20) to get:

$$D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right) \geq D_{\mathrm{LM}}\left(\Lambda_A\left(\tilde{\psi}_Q^{(K)}\right), \Lambda_A\left(\tilde{\phi}_Q^{(K)}\right)\right)$$
$$= D_{\mathrm{LM}}(\hat{p}, \hat{q}) \tag{22}$$

where:

$$\hat{p} = \sum_i p_i |i\rangle\langle i|, \quad \hat{q} = \sum_i q_i |i\rangle\langle i| \tag{23}$$

and p_i and q_i are classical distributions defined in (16) for ψ and ϕ respectively, which satisfy:

$$\sum_i p_i e_A(i) = \mathrm{Tr}(A\psi) + b_A(\psi),$$
$$\sum_i q_i e_A(i) = \mathrm{Tr}(A\phi) + b_A(\phi). \tag{24}$$

When $\hat{p}$ and $\hat{q}$ are tensor products of classical states, i.e., $\hat{p} = \hat{p}^{(1)} \otimes \cdots \otimes \hat{p}^{(K)}$ and $\hat{q} = \hat{q}^{(1)} \otimes \cdots \otimes \hat{q}^{(K)}$, it holds that:

$$D_{\mathrm{LM}}(\hat{p}, \hat{q}) = D_{\mathrm{tr}}(\hat{p}, \hat{q}). \tag{25}$$

This can be seen as follows. Let $M^\star$ be the optimal POVM element achieving the trace distance in (17). Then, we get:

$$D_{\mathrm{tr}}(\hat{p}, \hat{q}) = Tr[M^\star(\hat{p} - \hat{q})]$$
$$= Tr[\Delta(M^\star)(\hat{p} - \hat{q})] \tag{26}$$

where:

$$\Delta(\cdot) := \sum_i |i\rangle\langle i| \cdot |i\rangle\langle i| \tag{27}$$

is a classical dephasing channel. The effective POVM element $\Delta(M^\star)$ has the form:

$$\Delta(M^\star) = \sum_i \langle i|M^\star|i\rangle|i\rangle\langle i|. \tag{28}$$

Since each $|i\rangle\langle i|$ is a local POVM element and $0 \leq \langle i|M^\star|i\rangle \leq 1$ because $0 \leq M^\star \leq \mathbb{I}$, the two-outcome measurement $\{\Delta(M^\star), \mathbb{I} - \Delta(M^\star)\}$ can be realized by a local measurement and classical post-processing, and thus belongs to LM_2. This, together with (18), implies $D_{\mathrm{tr}}(\hat{p}, \hat{q}) \leq D_{\mathrm{LM}}(\hat{p}, \hat{q})$, and further combining (19) gives (25).

Combining (22) and (25) gives:

$$D_{\mathrm{tr}}(\hat{p}, \hat{q}) \leq D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right) \tag{29}$$

We now connect (29) to the expression (24) of the expectation value and bias. Let us first suppose $\mathrm{Tr}(A\psi) + b_A(\psi) \geq \mathrm{Tr}(A\phi) + b_A(\phi)$. Let $\mathcal{I}^\star := \{i \mid p_i - q_i \geq 0\}$ and let $\overline{\mathcal{I}}^\star$ be the complement set. Let us also define $A' = A + \mathbb{I}/2$, which satisfies $0 \leq A' \leq \mathbb{I}$ due to $-\mathbb{I}/2 \leq A \leq \mathbb{I}/2$. Then, we get:

$$\mathrm{Tr}[A'(\psi - \phi)] + b_A(\psi) - b_A(\phi)$$
$$= \mathrm{Tr}[(A + \mathbb{I}/2)(\psi - \phi)] + b_A(\psi) - b_A(\phi)$$
$$= \mathrm{Tr}[A(\psi - \phi)] + b_A(\psi) - b_A(\phi)$$
$$= \sum_i (p_i - q_i)e_A(i) \tag{30}$$
$$\leq \sum_{i \in \mathcal{I}^\star}(p_i - q_i)e_{A,\max} + \sum_{i \in \overline{\mathcal{I}}^\star}(p_i - q_i)e_{A,\min}$$
$$= D_{\mathrm{tr}}(\hat{p}, \hat{q})(e_{A,\max} - e_{A,\min})$$

where in the third line we used (24), in the fourth line we used the maximum and minimum estimator values:

$$e_{A,\max} := \max_i e_A(i), \quad e_{A,\min} := \min_i e_A(i), \tag{31}$$

and in the last line we used that:

$$\sum_{i \in \mathcal{I}^\star}(p_i - q_i) = -\sum_{i \in \overline{\mathcal{I}}^\star}(p_i - q_i) \tag{32}$$

and that the trace distance reduces to the total variation distance:

$$D_{\mathrm{tr}}(\hat{p}, \hat{q}) = \sum_{i:p_i - q_i \geq 0}(p_i - q_i) \tag{33}$$

for all classical states $\hat{p} = \sum_i p_i |i\rangle\langle i|$ and $\hat{q} = \sum_i q_i |i\rangle\langle i|$. Combining (29) and (30), we get:

$$e_{A,\max} - e_{A,\min} \geq \frac{\mathrm{Tr}[A'(\psi - \phi)] + b_A(\psi) - b_A(\phi)}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)}. \tag{34}$$

On the other hand, if $\mathrm{Tr}(A\psi) + b_A(\psi) \leq \mathrm{Tr}(A\phi) + b_A(\phi)$, we flip the role of ψ and ϕ to get:

$$e_{A,\max} - e_{A,\min} \geq -\frac{\mathrm{Tr}[A'(\psi - \phi)] + b_A(\psi) - b_A(\phi)}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)}. \tag{35}$$

Defining $\Delta e_A := e_{A,\max} - e_{A,\min}$, these two can be summarized as:

$$\Delta e_A \geq \frac{|\mathrm{Tr}[A'(\psi - \phi)] + b_A(\psi) - b_A(\phi)|}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)}. \tag{36}$$

Optimizing over A, ϕ, and ψ on both sides, we reach:

$$
\begin{aligned}
\Delta e_{\max} &\geq \max_{\psi, \phi} \max_{-\mathbb{I}/2 \leq A \leq \mathbb{I}/2} \frac{|\mathrm{Tr}[A'(\psi - \phi)] + b_A(\psi) - b_A(\phi)|}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)} \\
&= \max_{\psi, \phi} \max_{-\mathbb{I}/2 \leq A \leq \mathbb{I}/2} \frac{\mathrm{Tr}[A'(\psi - \phi)] + b_A(\psi) - b_A(\phi)}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)} \\
&\geq \max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi) + b_{A^\star}(\psi) - b_{A^\star}(\phi)}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)} \\
&\geq \max_{\psi, \phi} \frac{D_{\mathrm{tr}}(\psi, \phi) - 2b_{\max}}{D_{\mathrm{LM}}\left(\tilde{\psi}_Q^{(K)}, \tilde{\phi}_Q^{(K)}\right)}
\end{aligned} \tag{37}
$$

where in the second line we used that we can always take the numerator positive by appropriately flipping ψ and ϕ, in the third line we fixed $A'^\star = A^\star + \mathbb{I}/2$ to the one that achieves the trace distance $\mathrm{Tr}[A'^\star(\psi - \phi)] = D_{\mathrm{tr}}(\psi, \phi)$ as in (17), and in the fourth line we used the definition of $b_{\max}$. □

Measuring subfidelity

To estimate the subfidelity (7) for n-qubit states ρ and σ, it suffices to measure the two quantities, $\mathrm{Tr}(\rho\sigma)$ and $\mathrm{Tr}(\rho\sigma\rho\sigma)$, which can be measured by a quantum computer[38,39]. For readers' convenience, here we summarize several methods that can measure the subfidelity and see that the measurement can be done by a constant-depth quantum circuit.

Let us begin by $\mathrm{Tr}(\rho\sigma)$. Note that $\mathrm{Tr}(\rho\sigma) = \mathrm{Tr}(S\rho \otimes \sigma)$ where S is the n-qubit SWAP operator defined by $S|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$ with $|\psi\rangle$ and $|\phi\rangle$ being arbitrary n-qubit pure states. This can be famously measured by the SWAP test[38] that uses one ancillary qubit and n-qubit SWAP gate controlled on the ancillary qubit. Since the n-qubit SWAP gate can be realized by swapping individual qubits, the SWAP test runs with n uses of qubit SWAP gates controlled on the ancillary qubit, taking the circuit depth n.

One can significantly reduce the circuit depth by employing the destructive SWAP test[67]. Note that $\mathrm{Tr}(\rho\sigma) = \mathrm{Tr}(S_2^{\otimes n}\rho \otimes \sigma)$ where $S_2 := \sum_{i,j=0}^1 |ij\rangle\langle ji|$ is the qubit SWAP operator. This is obtained by measuring $\rho \otimes \sigma$ with respect to the eigenbasis of $S_2^{\otimes n}$, which is just a tensor product of the eigenbasis of S_2. Therefore, such a measurement can be accomplished by individually measuring a pair of qubits from ρ and σ with respect to the eigenbasis of S_2, for which one can use, e.g., Bell measurement. These measurements can run in parallel and thus only needs a constant depth circuit with respect to n (in fact, depth 2) that involves n two-qubit gates.

We remark that, at this point, we have already obtained a valid lower bound of $\Delta e_{\max}$ because the second term in (7) is positive, only improving the lower bound. Nevertheless, evaluating the second term, which involves $\mathrm{Tr}(\rho\sigma\rho\sigma)$, can significantly improve the bound particularly when ρ and σ are highly noisy and their purity is small.

$\mathrm{Tr}(\rho\sigma\rho\sigma)$ can be measured by a similar strategy to the one for $\mathrm{Tr}(\rho\sigma)$ with two copies of ρ and σ. Instead of the SWAP operator S, consider the CYCLE operator C defined as $C\left(\otimes_{i=1}^4 |\psi_i\rangle\right) = \otimes_{i=1}^4 |\psi_{i+1}\rangle$ where $|\psi_i\rangle$, $i = 1, 2, 3, 4$ is an arbitrary n-qubit pure state with $|\psi_5\rangle := |\psi_1\rangle$. Then, it is straightforward to check that $\mathrm{Tr}(\rho\sigma\rho\sigma) = \mathrm{Tr}(C\rho \otimes \sigma \otimes \rho \otimes \sigma)$. This can be measured by a generalization of the SWAP test where CYCLE gate C is controlled on the single ancillary qubit. Similarly to the case of SWAP, the CYCLE gate C can be decomposed into $C = C_2^{\otimes n}$ where kth C_2 gate (for any $k = 1, \ldots, n$) acts on the four-qubit state that consists of the kth qubit of ρ, σ, ρ, and σ. Since C_2 can be realized by three SWAP gates, one can measure $\mathrm{Tr}(\rho\sigma\rho\sigma)$ with $3n$ uses of qubit-SWAP gates controlled on the ancillary qubit, taking the circuit depth $3n$.

Similarly to the case of $\mathrm{Tr}(\rho\sigma)$, we can realize a significant reduction in the circuit depth by making the measurement destructive. All we have to do is to measure individual four-qubit states that each C_2 gate acts on with respect to the eigenbasis of C_2. Since the measurement of each C_2 can be run in parallel and each measurement circuit has a depth independent of n, this results in a constant-depth circuit that measures $C = C_2^{\otimes n}$.

We note the apparent similarity between the construction above and the circuit used in virtual distillation[14,15]. In particular, the strategy of destructive measurement was extensively discussed in ref. [15]. It is interesting to see that a construction that is highly relevant to a specific error-mitigation protocol provides a bound applicable to a general class of error-mitigation protocols.

Applications to other error-mitigation protocols

Here, we discuss how our framework can be applied to other two prominent error-mitigation protocols, noise extrapolation and virtual distillation.

Extrapolation methods[3,4] are used in scenarios where there is no clear analytical noise model. These strategies consider a family of noise channels $\{\mathcal{N}_\xi\}_\xi$, where ξ corresponds to the noise strength. The assumption here is that the description of $\mathcal{N}_\xi$ is unknown, but we have the ability to "boost" ξ such that $\xi \geq \bar{\xi}$ where $\bar{\xi}$ is the noise strength present in some given noisy circuit. The idea is that by studying how the expectation value of an observable depends on ξ, we can extrapolate what its value would be if $\xi = 0$. In particular, the Rth order Richardson extrapolation method work as follows. Let us take constants $\{\gamma_r\}_{r=0}^R$ and $\{c_r\}_{r=0}^R$ with $1 = c_0 < c_1 < \cdots < c_R \leq 1/\bar{\xi}$ such that:

$$\sum_{r=0}^R \gamma_r = 1, \quad \sum_{r=0}^R \gamma_r c_r^t = 0 \quad t = 1, \ldots, R. \tag{38}$$

Using these constants, one can show that:

$$\sum_{r=0}^R \gamma_r \mathrm{Tr}[A\mathcal{N}_{c_r\bar{\xi}}(\psi)] = \mathrm{Tr}(A\psi) + b_A(\psi) \tag{39}$$

where $b_A(\psi) = \mathcal{O}(\bar{\xi}^{R+1})$. This allows us to estimate the true expectation value using noisy states under multiple noise levels, as long as $\bar{\xi}$ is sufficiently small.

Richardson extrapolation is an instance of $(1, R+1)$-error mitigation. In particular, we have:

$$\mathcal{E}^{(k)} = \mathcal{N}_{c_{k-1}\bar{\xi}} \quad k = 1, \ldots, R+1 \tag{40}$$

in Definition 4. For an observable $A = \sum_a a\Pi_a$ where Π_a is the projector corresponding to measuring outcome a, the POVMs $\{M_{a^{(k)}}^{(k)}\}_{k=1}^{R+1}$ and classical estimator function e_A take the forms:

$$M_{a^{(k)}}^{(k)} = \Pi_{a^{(k)}} \quad k = 1, \ldots, R+1, \tag{41}$$

$$e_A(a^{(1)}, \ldots, a^{(R+1)}) = \sum_{k=1}^{R+1} \gamma_{k-1} a^{(k)}, \tag{42}$$

where $\{\gamma_k\}_{k=0}^R$ are the constants determined by (38). One can easily check that plugging the above expressions in the form of Definition 4 leads to (39).

Because of the constraint $-\mathbb{I}/2 \leq A \leq \mathbb{I}/2$, every eigenvalue a satisfies $-1/2 \leq a \leq 1/2$. This implies that:

$$
\begin{aligned}
e_{A,\max} &\leq \frac{1}{2}\sum_{r:\gamma_r \geq 0} \gamma_r - \frac{1}{2}\sum_{r:\gamma_r < 0}\gamma_r \\
&= \frac{1}{2}\sum_{r=0}^R |\gamma_r|
\end{aligned} \tag{43}
$$

and:

$$
\begin{aligned}
e_{A,\min} &\geq -\frac{1}{2}\sum_{r:\gamma_r \geq 0} \gamma_r + \frac{1}{2}\sum_{r:\gamma_r < 0}\gamma_r \\
&= -\frac{1}{2}\sum_{r=0}^R |\gamma_r|,
\end{aligned} \tag{44}
$$

leading to $\Delta e_{\max} \leq \sum_{r=0}^R |\gamma_r|$. On the other hand, any observable A having $\pm 1/2$ eigenvalues saturates this inequality. Therefore, we get the exact expression of the maximum spread for the extrapolation method as:

$$\Delta e_{\max}^{\mathrm{EX}} = \sum_{r=0}^R |\gamma_r|. \tag{45}$$

Next, we discuss virtual distillation[14,15], which is an example of $(Q, 1)$-error mitigation. Let ψ be an ideal pure output state from a quantum circuit. We consider a scenario where the noise in the circuit acts as an effective noise channel $\mathcal{E}$ that brings the ideal state to a noisy state of the form:

$$\mathcal{E}(\psi) = \lambda\psi + \sum_{k=2}^d \lambda_k \psi_k \tag{46}$$

for a certain $\{\lambda_k\}_{k=1}^d$, where d is the dimension of the system and $\{\psi_k\}_{k=1}^d$ constructs an orthonormal basis with $\psi_1 := \psi$. We also assume that λ is given as pre-knowledge. This form reflects the intuition that, as long as the noise is sufficiently small, the dominant eigenvector should be close to the

ideal state ψ. For a more detailed analysis of the form of this spectrum, we refer readers to ref. [68].

The Q-copy virtual distillation algorithm aims to estimate $\text{Tr}(W\psi)$ for a unitary observable W satisfying $W^2 = \mathbb{I}$ (e.g., Pauli operators) by using Q copies of $\mathcal{E}(\psi)$. The mitigation circuit consists of a controlled permutation and unitary W, followed by a measurement on the control qubit with the Hadamard basis. The probability of getting outcome 0 (projecting onto $|+\rangle\langle+|$) is:

$$\begin{aligned} p_0 &= \tfrac{1}{2}\left(1 + \text{Tr}\left[W\mathcal{E}(\psi)^Q\right]\right) \\ &= \tfrac{1}{2}\left[1 + \lambda^Q \text{Tr}(W\psi) + \sum_{k=2}^{d} \lambda_k^Q \text{Tr}(W\psi_k)\right]. \end{aligned} \tag{47}$$

This implies that:

$$(2p_0 - 1)\lambda^{-Q} = \text{Tr}(W\psi) + \sum_{k=2}^{d} \left(\frac{\lambda_k}{\lambda}\right)^Q \text{Tr}(W\psi_k), \tag{48}$$

providing a way of estimating $\text{Tr}(W\psi)$ with the bias $|\sum_{k=2}^{d} (\lambda_k/\lambda)^Q \text{Tr}(W\psi_k)| \leq \sum_{k=2}^{d} (\lambda_k/\lambda)^Q$.

We can see that this protocol fits into our framework with $K = 1$ and $\mathcal{E}_q = \mathcal{E}$ for $q = 1, \dots, Q$ as follows. For an arbitrary observable A, we can always find a decomposition with respect to the Pauli operators $\{P_i\}$ as:

$$A = \sum_i c_i P_i \tag{49}$$

for some set of real numbers $\{c_i\}$. We now apply the virtual distillation circuit for P_i at probability $|c_i|/\sum_j |c_j|$ and—similarly to the case of probabilistic error cancellation—employ an estimator function defined as:

$$e_A(i0) := \gamma \text{sgn}(c_i)\lambda^{-Q} \tag{50}$$
$$e_A(i1) := -\gamma \text{sgn}(c_i)\lambda^{-Q}$$

with $\gamma := \sum_i |c_i|$, where we treat i as a part of the measurement outcome. Then, we get:

$$\sum_i [p_{i0}\, e_A(i0) + p_{i1}\, e_A(i1)] = \text{Tr}(A\psi) + b_A(\psi) \tag{51}$$

where p_{i0} is the probability (47) with $W = P_i$ multiplied by $|c_i|/\sum_j |c_j|$, $p_{i1} = 1 - p_{i0}$, and $b_A(\psi) := \sum_{k=2}^{d} (\lambda_k/\lambda)^Q \text{Tr}(A\psi_k)$. Optimizing over observables $-\mathbb{I}/2 \leq A \leq \mathbb{I}/2$, we have:

$$\Delta e_{\max}^{VD} = \max\left\{2\lambda^{-Q}\sum_i |c_i| \,\middle|\, -\mathbb{I}/2 \leq \sum_i c_i P_i \leq \mathbb{I}/2\right\} \tag{52}$$

and:

$$b_{\max}^{VD} = \sum_{k=2}^{d} \frac{1}{2}\left(\frac{\lambda_k}{\lambda}\right)^Q. \tag{53}$$

Note added to proof

During the completion of our manuscript, we became aware of an independent work by Wang et al.[69], which showed a result related to our Theorem 3 on the exponential scaling of the maximum estimator spread.

CODE AVAILABILITY

Source codes used to generate the plots are available from the corresponding author upon request.

REFERENCES

1. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
2. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505 (2019).
3. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
4. Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
5. Giurgica-Tiron, T., Hindy, Y., LaRose, R., Mari, A. & Zeng, W. J. Digital zero noise extrapolation for quantum error mitigation. *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* 306 (2020).
6. He, A., Nachman, B., de Jong, W. A. & Bauer, C. W. Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Phys. Rev. A* **102**, 012426 (2020).
7. Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491 (2019).
8. Dumitrescu, E. F. et al. Cloud quantum computing of an atomic nucleus. *Phys. Rev. Lett.* **120**, 210501 (2018).
9. Buscemi, F., Dall'Arno, M., Ozawa, M. & Vedral, V. Direct observation of any two-point quantum correlation function. Preprint at https://arxiv.org/abs/1312.4240 (2013).
10. Buscemi, F., Dall'Arno, M., Ozawa, M. & Vedral, V. Universal optimal quantum correlator. *Int. J. Quantum Inf.* **12**, 1560002 (2014).
11. Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
12. Song, C. et al. Quantum computation with universal error mitigation on a superconducting quantum processor. *Sci. Adv.* **5**, eaaw5686 (2019).
13. Zhang, S. et al. Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system. *Nat. Commun.* **11**, 587 (2020).
14. Koczor, B. Exponential error suppression for near-term quantum devices. *Phys. Rev. X* **11**, 031057 (2021).
15. Huggins, W. J. et al. Virtual distillation for quantum error mitigation. *Phys. Rev. X* **11**, 041036 (2021).
16. Czarnik, P., Arrasmith, A., Cincio, L. & Coles, P. J. Qubit-efficient exponential suppression of errors. https://doi.org/10.48550/arXiv.2102.06056 (2021).
17. Cai, Z. Resource-efficient purification-based quantum error mitigation. Preprint at https://arxiv.org/abs/2107.07279 (2021).
18. Huo, M. & Li, Y. Dual-state purification for practical quantum error mitigation. *Phys. Rev. A* **105**, 022427 (2022).
19. Xiong, Y., Ng, S. X. & Hanzo, L. Quantum error mitigation relying on permutation filtering. *IEEE Trans. Commun.* **70**, 1927 (2022).
20. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242 (2017).
21. McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. C. & Yuan, X. Quantum computational chemistry. *Rev. Mod. Phys.* **92**, 015003 (2020).
22. Cao, Y. et al. Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**, 10856 (2019).
23. McArdle, S., Yuan, X. & Benjamin, S. Error-mitigated digital quantum simulation. *Phys. Rev. Lett.* **122**, 180501 (2019).
24. Carnot, S. Reflections on the motive power of fire, and on machines fitted to develop that power. *Paris: Bachelier* **108**, 1824 (1824).
25. Wang, S. et al. Noise-induced barren plateaus in variational quantum algorithms. *Nat. Commun.* **12**, 6961 (2021).
26. Yuan, X., Zhang, Z., Lütkenhaus, N. & Ma, X. Simulating single photons with realistic photon sources. *Phys. Rev. A* **94**, 062305 (2016).
27. McClean, J. R., Kimchi-Schwartz, M. E., Carter, J. & de Jong, W. A. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Phys. Rev. A* **95**, 042308 (2017).
28. Czarnik, P., Arrasmith, A., Coles, P. J. & Cincio, L. Error mitigation with Clifford quantum-circuit data. *Quantum* **5**, 592 (2021).
29. Bonet-Monroig, X., Sagastizabal, R., Singh, M. & O'Brien, T. E. Low-cost error mitigation by symmetry verification. *Phys. Rev. A* **98**, 062339 (2018).
30. Bravyi, S., Sheldon, S., Kandala, A., Mckay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A* **103**, 042605 (2021).
31. Yoshioka, N. et al. Generalized quantum subspace expansion. *Phys. Rev. Lett.* **129**, 020502 (2022).
32. McClean, J. R., Jiang, Z., Rubin, N. C., Babbush, R. & Neven, H. Decoding quantum errors with subspace expansions. *Nat. Commun.* **11**, 636 (2020).
33. Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13 (1963).
34. Matthews, W., Wehner, S. & Winter, A. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Commun. Math. Phys.* **291**, 813 (2009).
35. Fuchs, C. & van de Graaf, J. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inf. Theory* **45**, 1216–1227 (1999).
36. Watrous, J. *The Theory of Quantum Information* (Cambridge University Press, 2018).
37. Miszczak, J. A., Puchała, Z., Horodecki, P., Uhlmann, A. & Życzkowski, K. Sub– and super–fidelity as bounds for quantum fidelity. *Quantum Inf. Comput.* **9**, 0103 (2009).
38. Ekert, A. K. et al. Direct estimations of linear and nonlinear functionals of a quantum state. *Phys. Rev. Lett.* **88**, 217901 (2002).
39. Bacon, D., Chuang, I. L. & Harrow, A. W. Efficient quantum circuits for schur and clebsch-gordan transforms. *Phys. Rev. Lett.* **97**, 170502 (2006).

40. Cerezo, M., Poremba, A., Cincio, L. & Coles, P. J. Variational quantum fidelity estimation. *Quantum* **4**, 248 (2020).

41. Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).

42. Kim, Y. et al. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. Preprint at https://arxiv.org/abs/2108.09197 (2021).

43. Sagastizabal, R. et al. Experimental error mitigation via symmetry verification in a variational quantum eigensolver. *Phys. Rev. A* **100**, 010302 (2019).

44. Müller-Hermes, A., Stilck França, D. & Wolf, M. M. Relative entropy convergence for depolarizing channels. *J. Math. Phys.* **57**, 022202 (2016).

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

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

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

48. Bennett, C. H., Shor, P. W., Smolin, J. A. & Thapliyal, A. V. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.* **83**, 3081 (1999).

49. Pirandola, S., Laurenza, R., Ottaviani, C. & Banchi, L. Fundamental limits of repeaterless quantum communications. *Nat. Commun.* **8**, 15043 (2017).

50. Berta, M., Brandão, F. G. S. L., Christandl, M. & Wehner, S. Entanglement cost of quantum channels. *IEEE Trans. Inf. Theory* **59**, 6779 (2013).

51. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183 (1961).

52. Brandão, F., Horodecki, M., Ng, N., Oppenheim, J. & Wehner, S. The second laws of quantum thermodynamics. *Proc. Natl. Acad. Sci. USA* **112**, 3275 (2015).

53. Gour, G., Jennings, D., Buscemi, F., Duan, R. & Marvian, I. Quantum majorization and a complete set of entropic conditions for quantum thermodynamics. *Nat. Commun.* **9**, 5352 (2018).

54. Hakoshima, H., Matsuzaki, Y. & Endo, S. Relationship between costs for quantum error mitigation and non-markovian measures. *Phys. Rev. A* **103**, 012611 (2021).

55. Breuer, H.-P., Laine, E.-M., Piilo, J. & Vacchini, B. Colloquium: non-Markovian dynamics in open quantum systems. *Rev. Mod. Phys.* **88**, 021002 (2016).

56. Suzuki, Y., Endo, S., Fujii, K. & Tokunaga, Y. 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).

57. Lostaglio, M. & Ciani, A. Error mitigation and quantum-assisted simulation in the error corrected regime. *Phys. Rev. Lett.* **127**, 200506 (2021).

58. Piveteau, C., Sutter, D., Bravyi, S., Gambetta, J. M. & Temme, K. Error mitigation for universal gates on encoded qubits. *Phys. Rev. Lett.* **127**, 200505 (2021).

59. Xiong, Y., Chandra, D., Ng, S. X. & Hanzo, L. Sampling overhead analysis of quantum error mitigation: uncoded vs. coded systems. *IEEE Access* **8**, 228967 (2020).

60. Lloyd, S. Enhanced sensitivity of photodetection via quantum illumination. *Science* **321**, 1463 (2008).

61. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).

62. Audenaert, K. M., Nussbaum, M., Szkoła, A. & Verstraete, F. Asymptotic error rates in quantum hypothesis testing. *Commun. Math. Phys.* **279**, 251 (2008).

63. Binder, F. C., Thompson, J. & Gu, M. Practical unitary simulator for non-markovian complex processes. *Phys. Rev. Lett.* **120**, 240502 (2018).

64. Blank, C., Park, D. K. & Petruccione, F. Quantum-enhanced analysis of discrete stochastic processes. *npj Quantum Inf.* **7**, 1–9 (2021).

65. Lami, L., Palazuelos, C. & Winter, A. Ultimate data hiding in quantum mechanics and beyond. *Commun. Math. Phys.* **361**, 661 (2018).

66. Corrêa, W. H. G., Lami, L. & Palazuelos, C. Maximal gap between local and global distinguishability of bipartite quantum states. Preprint at https://arxiv.org/abs/2110.04387 (2021).

67. Garcia-Escartin, J. C. & Chamorro-Posada, P. Swap test and hong-ou-mandel effect are equivalent. *Phys. Rev. A* **87**, 052330 (2013).

68. Koczor, B. The dominant eigenvector of a noisy quantum state. *New J. Phys.* **23**, 123047 (2021).

69. Wang, S. et al. Can error mitigation improve trainability of noisy variational quantum algorithms? Preprint at https://arxiv.org/abs/2109.01051 (2021).

AUTHOR CONTRIBUTIONS

R.T., S.E., and S.M. came up with a preliminary idea on connecting the state distinguishability to error mitigation. R.T. conceived the project, obtained the main results, and wrote the manuscript draft. S.E. proposed a way of directly estimating a lower bound of the fidelity bound on a quantum computer, which eventually resulted in the subfidelity bound. S.M. contributed to the analysis to compare the subfidelity bound to the trace-distance bound. R.T. and M.G. wrote the manuscript. All authors contributed to the interpretation and discussion of the results.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41534-022-00618-z.

Correspondence and requests for materials should be addressed to Ryuji Takagi, Suguru Endo, Shintaro Minagawa or Mile Gu.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.