ARTICLE   OPEN

Check for updates

Error statistics and scalability of quantum error mitigation formulas

Dayue Qin [ID][1], Yanzhu Chen [ID][2] and Ying Li [ID][1]✉

Quantum computing promises advantages over classical computing in many problems. Nevertheless, noise in quantum devices prevents most quantum algorithms from achieving the quantum advantage. Quantum error mitigation provides a variety of protocols to handle such noise using minimal qubit resources. While some of those protocols have been implemented in experiments for a few qubits, it remains unclear whether error mitigation will be effective in quantum circuits with tens to hundreds of qubits. In this paper, we apply statistics principles to quantum error mitigation and analyse the scaling behaviour of its intrinsic error. We find that the error increases linearly $O(\epsilon N)$ with the gate number N before mitigation and sublinearly $O(\epsilon' N^\gamma)$ after mitigation, where $\gamma \approx 0.5$, ϵ is the error rate of a quantum gate, and ϵ' is a protocol-dependent factor. The $\sqrt{N}$ scaling is a consequence of the law of large numbers, and it indicates that error mitigation can suppress the error by a larger factor in larger circuits. We propose the importance Clifford sampling as a key technique for error mitigation in large circuits to obtain this result.

INTRODUCTION

With the recent progress that quantum computers can have more than half a hundred qubits[1,2], it is widely accepted that we are in the era of noisy intermediate-scale quantum (NISQ) technologies[3]. A prominent feature of NISQ technologies is the potential for surpassing all classical computers in certain tasks, yet they cannot realise full quantum error correction and achieve fault tolerance due to noise and the limited number of physical qubits. Under the assumption of realistic noise models, the qubit overhead is thousands of physical qubits per logical qubit to reduce the chance of a logical error to the negligible level[4,5]. This requirement of quantum error correction is considerably beyond today's technologies.

Nevertheless, we can still perform computation tasks with NISQ devices. Protocols proposed recently allow us to bypass quantum error correction, which are termed quantum error mitigation[6–17]. Unlike error correction preserving the logical quantum state, error mitigation aims at recovering the error-free measurement outcome without physically preparing the error-free state. It can extract the correct computation result from a noisy device as long as the physical quantum state is not excessively damaged by the error accumulation[18]. For example, if the state becomes the maximally mixed state due to noise, there is nothing we can do to extract any useful information about the noise-free state. Recently, quantum algorithms using shallow circuits have been developed to minimise error accumulation. Quantum simulation algorithms based on variational, Lanczos and Monte Carlo methods are promising examples of such algorithms[19–23]. Although shallow-circuit algorithms and error mitigation protocols have been successful in proof-of-principle experiments[12,24–30], it remains unexplored how they will perform as we venture into the regime of useful applications, where the computation involves more than half a hundred qubits and the device noise permits error mitigation but not yet error correction.

In this work, we address how the computation error after mitigation scales with the circuit size. In many quantum algorithms, we use quantum circuits to evaluate the expected values of observables. For example, the Hamiltonian is evaluated in the variational quantum eigensolver[20]. Because of noise, an actual quantum computer produces a biased expected value, and the bias usually increases with the circuit size due to the error accumulation. Among the error mitigation protocols, probabilistic error cancellation can completely remove the bias under ideal conditions[7,8]. Under realistic conditions, however, all protocols leave a residual bias in the computation result. This residual bias depends on the protocol and circuit depth.

To draw a conclusion regardless of the protocol, we utilise a general formalism of error mitigation. In this formalism, we recover the observable in the error-free circuit using an error mitigation formula, which is a function of observables directly measured with noisy circuits. Many such formulas are inspired by our knowledge of quantum physics, such as error extrapolation[6,7,31,32], probabilistic error cancellation[7,8] and virtual distillation[13,14,33–35]. Throughout this work, when a concrete error mitigation formula is needed for analysis, we take the three aforementioned protocols as examples. An alternative way to construct the formula is optimising a parameterised function with data of selected training circuits[36,37]. We find that the optimisation can suppress the scaling of the residual bias with respect to the circuit size.

For optimisation-based error mitigation protocols, we propose the importance Clifford sampling (ICS) as an efficient and scalable method to generate training circuits. Other than being practically useful in its own right, ICS lends us a tool to analyse the residual bias in the computation result. With its help, we show that the global depolarising model with circuit-dependent fluctuation is an effective phenomenological-error model, which describes the impact of realistic error models. Using this phenomenological model, we analyse the scaling behaviour of the residual bias. We find that the bias in the computation result after an optimised error mitigation process increases in proportion to $\sqrt{N}$, where N is the gate number. In contrast, the bias is usually proportional to N

[1]Graduate School of China Academy of Engineering Physics, Beijing 100193, China. [2]Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA.
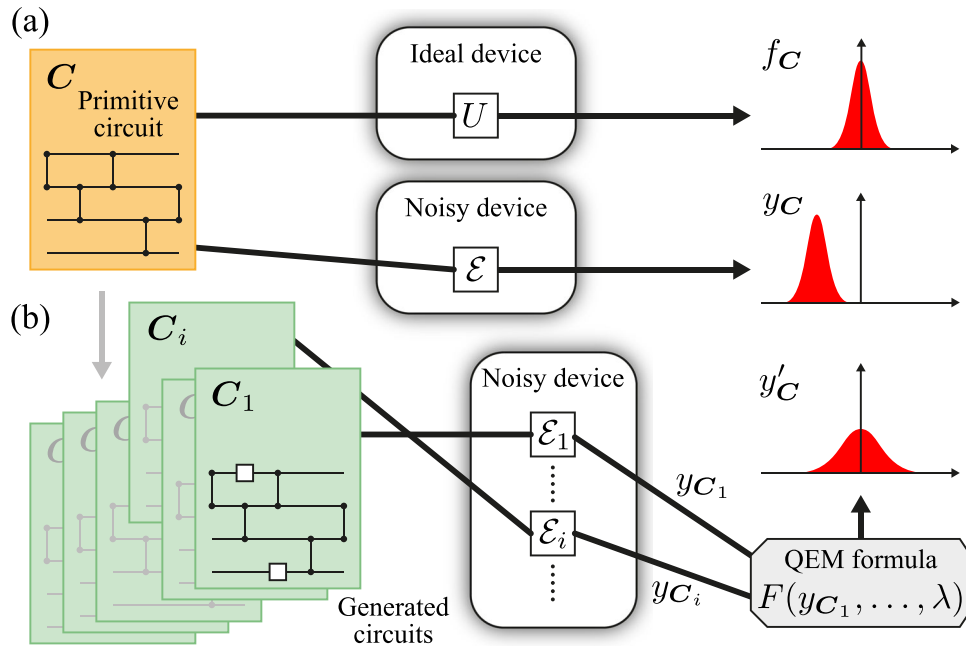✉email: yli@gscaep.ac.cn

Fig. 1 Schematic illustration of quantum error mitigation formulas. a Ideal and noisy quantum computing for the expected value of an observable. The distribution of the expected value is biased because of noise. **b** Error-mitigated quantum computing. The bias is corrected in quantum error mitigation (QEM).

without error mitigation. Because error mitigation can suppress the error by a factor increasing with the circuit size, it is a feasible technique for large circuits.

The Results section is organised as follows. After introducing the general formalism of error mitigation, we discuss the error scaling in the mitigation protocols using the global depolarising model, which will be validated subsequently as the effective phenomenological-error model. Then we propose the ICS protocol, followed by a description of the important training circuits, the algorithms to generate them and an analysis of the sampling cost. We introduce the phenomenological-error model and show that the fluctuation of the effective depolarising rate follows the $\sqrt{N}$ scaling, which is numerically verified. Finally, we show the same scaling relation between the bias and the gate number in error extrapolation, probabilistic error cancellation and virtual distillation.

RESULTS

Error mitigation formula

First, we introduce the notations. In quantum computing, a quantum circuit consists of quantum gates. Let U_j be the unitary operator of the jth gate. The circuit with N gates realises the transformation $U = U_N \cdots U_2 U_1$. Given the initial state of n qubits $|0\rangle^{\otimes n}$ and observable Q, the expected value in the error-free circuit is $f_{\mathbf{C}} = \mathrm{Tr}[Q[U](|0\rangle\langle 0|^{\otimes n})]$, where $[U](\bullet) = U \bullet U^\dagger$. Here we use $\mathbf{C} = (U_1, \ldots, U_N, Q)$ to denote the circuit with the observable specified. If the circuit is noisy, the transformation is inexact, and we use the completely positive map $\mathcal{E}$ to denote the erroneous transformation. The expected value becomes $y_{\mathbf{C}} = \mathrm{Tr}[Q\mathcal{E}(|0\rangle\langle 0|^{\otimes n})]$. Then, $y_{\mathbf{C}} - f_{\mathbf{C}}$ is the bias without error mitigation. Note that the error in the actual computing also depends on the statistical error due to finite measurement shots.

The general form of error mitigation formulas reads

$$y'_{\mathbf{C}} = F(y_{\mathbf{C}_1}, y_{\mathbf{C}_2}, \ldots, \lambda_1, \lambda_2, \ldots) \tag{1}$$

where $y'_{\mathbf{C}}$ is the result of the circuit $\mathbf{C}$ after error mitigation, $\mathbf{C}_1, \mathbf{C}_2, \ldots$ are circuits generated from the primitive circuit $\mathbf{C}$, and λ's denote parameters determined via error mitigation protocols. See

Fig. 1. In quantum computing, we evaluate $y_{\mathbf{C}_i}$ using the noisy quantum computer and calculate the error-mitigated value $y'_{\mathbf{C}}$ according to the formula. The bias after error mitigation is $y'_{\mathbf{C}} - f_{\mathbf{C}}$. Next, we show how some specific error mitigation protocols fit into the general form.

Many error mitigation protocols have been proposed. See Ref. [17] for a review. In this work, we take three protocols as examples: error extrapolation, probabilistic error cancellation and virtual distillation. These protocols are applicable to any quantum algorithm evaluating expected values and can largely reduce the error. We give a minimal description here and leave a more detailed overview to Supplementary Note 1.

In error extrapolation using a polynomial fitting function[7,31], the error mitigation formula is

$$y'_{\mathbf{C}} = \sum_i q_i y_{\mathbf{C}_i}, \tag{2}$$

where $\mathbf{C}_i$ is the primitive circuit with noise increased by a factor of r_i, and coefficients q_i are determined by noise amplification factors (i.e. r_i). For example, for the linear extrapolation with $r_1 = 1$ and $r_2 = 2$, the formula is

$$y'_{\mathbf{C}} = 2y_{\mathbf{C}_1} - y_{\mathbf{C}_2}. \tag{3}$$

In probabilistic error cancellation, the completely positive map of the error-free circuit is expressed as a linear combination of erroneous maps, i.e.

$$[U] = \sum_i q_i \mathcal{E}_i, \tag{4}$$

where q_i are quasi-probabilities, and $\mathcal{E}_i$ is the map of a noisy circuit $\mathbf{C}_i$. Here $\mathbf{C}_i$ is generated by, for example, replacing or adding some gates in the primitive circuit $\mathbf{C}$. We can work out the quasi-probability decomposition with gate set tomography data[8] or in a learning manner[36]. Given the decomposition, the error mitigation formula is the same as Eq. (2), but coefficients and circuits are different from error extrapolation.

In virtual distillation, k copies of the erroneous state ρ are used to evaluate the observable in a distilled state without physically

preparing it. Given the primitive circuit **C** that prepares the state ρ, the circuit $\mathbf{C}_1$ is to evaluate $y_{\mathbf{C}_1} = \mathrm{Tr}(Q\rho^k)$, and the circuit $\mathbf{C}_2$ is to evaluate $y_{\mathbf{C}_2} = \mathrm{Tr}(\rho^k)$. Then the error mitigation formula reads

$$y'_{\mathbf{C}} = \frac{y_{\mathbf{C}_1}}{y_{\mathbf{C}_2}}. \tag{5}$$

It is similar in related protocols, e.g. verified phase estimation[34] and dual-state purification[35].

Bias in the global depolarising model

Before considering realistic error models, we take the global depolarising model as an example to discuss the bias in error mitigation formulas. In this section, we show that, if the error mitigation protocols are perfectly implemented, probabilistic error cancellation and learning-based error mitigation can reduce the bias to zero, while linear extrapolation and virtual distillation with two copies can reduce the bias from $O(N\epsilon)$ to $O(N^2\epsilon^2)$, where N is the gate number and ϵ is the depolarising rate per gate. In the section of "Phenomenological-error model" we will show that the global depolarising model successfully captures the influence of realistic noise and can be used as a phenomenological model.

In the global depolarising model, the j-th gate with error is described by the map $\mathcal{G}_j = (1-\epsilon)[U_j] + \epsilon\mathcal{D}$ acting on the whole input state, where ϵ is the gate depolarising rate, $\mathcal{D}(\bullet) = \mathrm{Tr}(\bullet)\rho_m$ is the depolarising map, and $\rho_m = \mathbb{1}/2^n$ is the maximally mixed state. Without loss of generality, we assume that the observable is a traceless operator, and we have $y_{\mathbf{C}} = (1-\epsilon)^N f_{\mathbf{C}} = f_{\mathbf{C}} + O(\epsilon N)$. The bias increases linearly with the gate number when N is significantly smaller than ϵ^{-1}. In the limit of large N, the bias approaches a finite value if the observable is bounded.

We take linear extrapolation as an example of error extrapolation. We can construct two noisy circuits using original gates and double-noise gates, respectively. Let $\mathcal{G}'_j = (1-2\epsilon)[U_j] + 2\epsilon\mathcal{D}$ be the gate with the doubled depolarising rate, two circuits labelled by $i = 1, 2$ produce expected values $y_{\mathbf{C}_i} = \mathrm{Tr}[Q\mathcal{E}_i(|0\rangle\langle 0|^{\otimes n})]$, where $\mathcal{E}_1 = \mathcal{G}_N \cdots \mathcal{G}_2\mathcal{G}_1$ and $\mathcal{E}_2 = \mathcal{G}'_N \cdots \mathcal{G}'_2\mathcal{G}'_1$. Then, Eq. (3) leads to the error-mitigated expected value

$$\begin{aligned} y'_{\mathbf{C}} &= 2(1-\epsilon)^N f_{\mathbf{C}} - (1-2\epsilon)^N f_{\mathbf{C}} \\ &= f_{\mathbf{C}} + O(\epsilon^2 N^2). \end{aligned} \tag{6}$$

We can find that the bias in the linear extrapolation formula increases quadratically with the gate number because the linear extrapolation eliminates the first-order contribution of errors.

In probabilistic error cancellation, we take the quasi-probability decomposition of each gate as

$$[U_j] = \frac{1}{1-\epsilon}\mathcal{G}_j - \frac{\epsilon}{1-\epsilon}\mathcal{D}. \tag{7}$$

This decomposition means that we can correct the error by stochastically replacing the original gate $\mathcal{G}_j$ with the depolarising map $\mathcal{D}$ according to a quasi-probability distribution. The decomposition formula of the entire circuit reads

$$\begin{aligned} [U] &= \prod_{j=1}^{N}\left[\frac{1}{1-\epsilon}\mathcal{G}_j - \frac{\epsilon}{1-\epsilon}\mathcal{D}\right] \\ &= \frac{1}{(1-\epsilon)^N}\mathcal{E}_1 - \frac{\epsilon}{(1-\epsilon)^N}\mathcal{E}_2 + \cdots, \end{aligned} \tag{8}$$

where $\mathcal{E}_1 = \mathcal{G}_N \cdots \mathcal{G}_2\mathcal{G}_1$ corresponding to the primitive circuit, $\mathcal{E}_2 = \mathcal{G}_N \cdots \mathcal{G}_2\mathcal{D}$ in which the first gate is replaced, and so on. Then the error mitigation formula is

$$y'_{\mathbf{C}} = \frac{1}{(1-\epsilon)^N}y_{\mathbf{C}_1} - \frac{\epsilon}{(1-\epsilon)^N}y_{\mathbf{C}_2} + \cdots = f_{\mathbf{C}}. \tag{9}$$

Here, we have used that $y_{\mathbf{C}_i} = 0$ if any gate is replaced with $\mathcal{D}$. Therefore, the residual bias is zero.

Lastly, we consider virtual distillation. The final state of N gates with the depolarising error is

$$\rho = (1-\epsilon_t)U|0\rangle\langle 0|^{\otimes n}U^\dagger + \epsilon_t\rho_m, \tag{10}$$

where $\epsilon_t = 1 - (1-\epsilon)^N$. Take the second-order virtual distillation (i.e. $k = 2$) as an example, the error-mitigated expected value is

$$\begin{aligned} y'_{\mathbf{C}} &= \frac{(1-\epsilon_t)^2 + 2^{1-n}(1-\epsilon_t)\epsilon_t}{(1-\epsilon_t)^2 + 2^{1-n}(1-\epsilon_t)\epsilon_t + 2^{-n}\epsilon_t^2}f_{\mathbf{C}} \\ &= f_{\mathbf{C}} + O(\epsilon^2 N^2). \end{aligned} \tag{11}$$

Therefore, the bias in the second-order virtual distillation increases quadratically with the gate number, which is the natural consequence of the second-order distillation formalism.

So far we have been considering ideal conditions. Under realistic conditions, imperfections in the implementation cause an additional contribution to the bias. For example, zero-bias probabilistic error cancellation requires exact knowledge about the depolarising rate. If the depolarising rate is thought to be ϵ' instead of its actual value ϵ and we work out the error mitigation formula with ϵ', we have $y'_{\mathbf{C}} = (1-\epsilon)^N/(1-\epsilon')^N f_{\mathbf{C}}$. Then, the bias of the error mitigation formula is $O((\epsilon'-\epsilon)N)$, which is finite and increases linearly with the gate number. It is similar for error extrapolation, in which the bias scales linearly if the noise is not increased exactly as designed.

Next, we analyse the bias in learning-based error mitigation. The optimisation of an ansatz function is a flexible approach for working out a proper error mitigation formula. Various ansatz functions have been proposed[36–38]. In this work, we consider a general framework of this approach and focus on the scaling of the bias with respect to the gate number.

One way to compose an ansatz function is by modifying a specific-form formula. Taking the linear error extrapolation as an example, we parameterise the formula as

$$y'_{\mathbf{C}} = \lambda y_{\mathbf{C}_1} + (1-\lambda)y_{\mathbf{C}_2}. \tag{12}$$

We determine λ by minimising the bias for a set of circuits, which are called training circuits. To evaluate the bias, the error-free expected value must be known. This condition limits the choice of training circuits. We can use only one training circuit **T** and the corresponding data $(y_{\mathbf{T}_1}, y_{\mathbf{T}_2}, f_{\mathbf{T}})$ to determine λ for the ansatz considered here. The bias of the training circuit is minimised at

$$\lambda^* = \frac{f_{\mathbf{T}} - y_{\mathbf{T}_2}}{y_{\mathbf{T}_1} - y_{\mathbf{T}_2}}. \tag{13}$$

For the global depolarising model, the optimal parameter is $\lambda^* = [1 - (1-2\epsilon)^N]/[(1-\epsilon)^N - (1-2\epsilon)^N]$. If we take $\lambda = \lambda^*$ in the error mitigation formula, the bias is zero for all circuits with the same gate number N. Therefore, the linear error extrapolation becomes bias-free after the optimisation.

It is similar for other error mitigation protocols. For probabilistic error cancellation, we can take the depolarising rate ϵ in Eq. (9) as the variational parameter, assuming the actual depolarising rate is unknown. We can find the optimal value of ϵ with data of a training circuit, and the optimal value must be the actual depolarising rate. Then, the error mitigation formula taking the optimal parameter is bias-free for all circuits. For virtual distillation, we can choose the ansatz $y'_{\mathbf{C}} = \lambda\frac{y_{\mathbf{C}_1}}{y_{\mathbf{C}_2}}$. According to Eq. (11), the bias is zero when λ cancels the factor before $f_{\mathbf{C}}$.

We have seen that the learning-based approach can reduce the bias in error mitigation. According to the global depolarising model, the bias is zero in all examples. We get this perfect result because the global depolarising model is free of fluctuation, i.e. errors of all gates have the same impact on the expected value. The impact is a factor of $1-\epsilon$. Without the fluctuation, there are many simple error mitigation formulas that can simultaneously and completely correct the bias for all circuits.

In error models with fluctuation, the optimised error mitigation formula has a finite bias, and the bias increases with the gate number. Usually, errors are localised in many actual quantum computing systems, e.g. superconducting qubits and trapped ions. The error associated with a gate only affects qubits at the location of the gate (rather than the entire quantum register as in the global depolarising model). The contribution of an error to the bias depends on its location and the circuit. For example, if the observable is the Pauli operator X of qubit-1, errors localised on qubit-2 do not affect the observable; A phase-flip error before the measurement changes the sign of X but preserves the sign if we modify the circuit by inserting a Hadamard gate before the measurement. The fluctuation of error contributions causes a finite bias, i.e. the error mitigation formula cannot simultaneously compensate for all errors for all circuits. Assuming we can successfully compensate for the average contribution of errors, the residual bias is due to the fluctuation across different circuits. We find that in a large class of error mitigation formulas, the fluctuation-caused bias is proportional to $\sqrt{N}$. Later, we will show that the global depolarising model with fluctuation is an effective phenomenological model to characterise the impact of errors in realistic error models, see Fig. 2.

Importance Clifford sampling

In this section, we address the question of how to efficiently sample large training circuits by proposing sampling algorithms whose resource costs scale linearly with the circuit size. These training circuits are Clifford circuits sharing the same circuit frame as the original noisy circuit, for which the ideal measurements take non-zero expected values.

A classical computer can efficiently simulate Clifford circuits, in which all gates are Clifford gates. Because the error-free expected value f_C of a Clifford circuit is computable[39,40], we can take them as training circuits. However, not every Clifford circuit is suitable. We take Eq. (13) as an example. If the training circuit T has a zero expected value, i.e. $f_T = 0$, erroneous expected values are all zero, i.e. $y_{T_1} = y_{T_2} = 0$. In this case, we cannot use the equation to determine the optimal parameter. Therefore, to find the optimal parameter, we need a training circuit T whose expected value is non-zero.
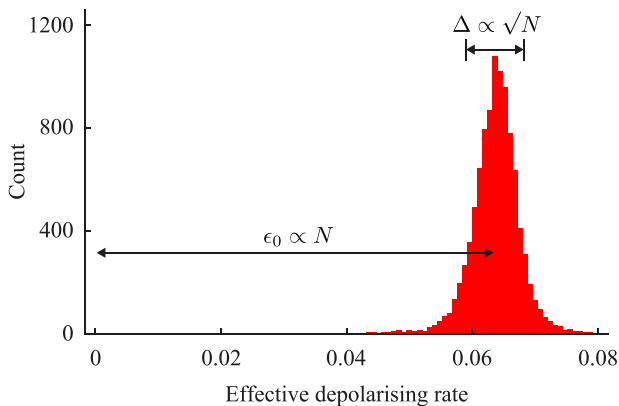

Fig. 2 Distribution of the effective depolarising rate in the phenomenological-error model. In the model, the impact of errors in a noisy circuit is characterised by the global depolarising model with the circuit-dependent depolarising rate ϵ_C. The histogram is generated using six-qubit periodic-cycling circuits with 72 two-qubit gates under the gate depolarising noise. The error rate per gate is 0.001. Single-qubit gates are randomly sampled from the set of single-qubit unitaries with the weight f_C^2. The average depolarising rate is proportional to the gate number N, and the standard deviation is proportional to $\sqrt{N}$.

It is general that some training circuits are more important than others in the learning-based approach. To optimise the error mitigation formula, we need a measure of its overall performance in various circuits. We take the mean squared error (MSE) as an example, which reads

$$L_{\mathbb{R}} = \langle (y_C - f_C)^2 \rangle_{\mathbb{R}}, \tag{14}$$

where $\langle g(C) \rangle_{\mathbb{R}} \equiv \frac{1}{|\mathbb{R}|} \sum_{C \in \mathbb{R}} g(C)$ is the average of the real-valued circuit function $g(C)$ over the circuit set $\mathbb{R}$. Importance sampling is a crucial technique in statistics, in which the probability of a sample is proportional to the magnitude of its value, i.e. $(y_C - f_C)^2$ in MSE. According to importance sampling, we prefer training circuits with a larger bias over those with a smaller bias. The larger bias circuits, i.e. error-sensitive circuits, can provide more information about noise in the circuit.

The question of sampling training circuits has two parts. The first part is how to efficiently generate an error-sensitive circuit. The second part is how to draw samples according to a distribution. We address the first part in the "Circuit generation" section and the second part in the "Circuit frame" and "Sampling algorithms" sections.

Circuit generation

There are different approaches of generating an error-sensitive circuit. For example, we can randomly select a circuit and calculate the expected value, and we take it as a training circuit only if the expected value is non-zero. This approach works only when the circuit size is small because circuits with a non-zero expected value are rare in large Clifford circuits. An approach usually used in randomised benchmarking is reversing the transformation by adding an additional unitary at the end of the circuit[41]. We will not take this approach because the additional unitary may significantly increase the total gate number in multi-qubit circuits. We want to generate training circuits with a specific gate number, such that the error mitigation formula is optimised for circuits with the same gate number.

In the following, we focus on the case that the observable Q is a Pauli operator. In the standard model of quantum computing, qubits at the end of the circuit are measured in the computation basis, i.e. the Pauli operator Z is measured. One can adjust the measurement basis by inserting gates before the measurement. For example, by inserting single-qubit Clifford gates before the measurement, we can measure any Pauli operator. For a general observable, a way to evaluate its expected value is by expressing it as a linear combination of Pauli operators and computing the expected value of each term.

The expected value of a Pauli operator in a Clifford circuit takes three values 0 and ± 1. We can reexpress the error-free expected value as $f_C = \mathrm{Tr}(Q_U |0\rangle\langle 0|^{\otimes n})$, where $Q_U = U^\dagger Q U$ is the effective observable. When U is Clifford, Q_U is a Pauli operator. Let $P_i = I, X, Y, Z$ be the single-qubit Pauli operator on qubit-i, $Q_U = \pm P_1 \otimes P_2 \otimes \cdots \otimes P_n$. Then $f_C = \pm \prod_{i=1}^{n} \langle 0|P_i|0\rangle$. If any single-qubit Pauli operator P_i is X or Y, the expected value is zero. If all P_i are I or Z, $f_C = \pm 1$, and the sign is the same as Q_U. For a randomly generated Clifford circuit, it is likely that some single-qubit Pauli operators contained in Q_U are X or Y, i.e. $f_C = 0$.

We can deterministically generate an error-sensitive circuit as follows. The setup is shown in Fig. 3. The overall unitary transformation of the circuit is $U = U'U_0$, where $U_0 = R_1 \otimes R_2 \otimes \cdots \otimes R_n$ is one layer of single-qubit gates, and R_i is the gate on qubit-i. First, given the gate number, we generate a random Clifford circuit, which realises the unitary U'. If $U_0 = \mathbb{1}$, the effective observable is $Q_{U'} = \pm P'_1 \otimes P'_2 \otimes \cdots \otimes P'_n$. Given Q and U', we can efficiently work out this expression of $Q_{U'}$ on a classical computer. Second, we determine single-qubit gates in U_0: we take a Clifford R_i satisfying $R_i^\dagger P'_i R_i = \pm Z, I$. For the final circuit $U = U'U_0$, single-qubit Pauli operators in its effective observable Q_U are either I or Z. Then, the expected value is $f_C = \pm 1$.
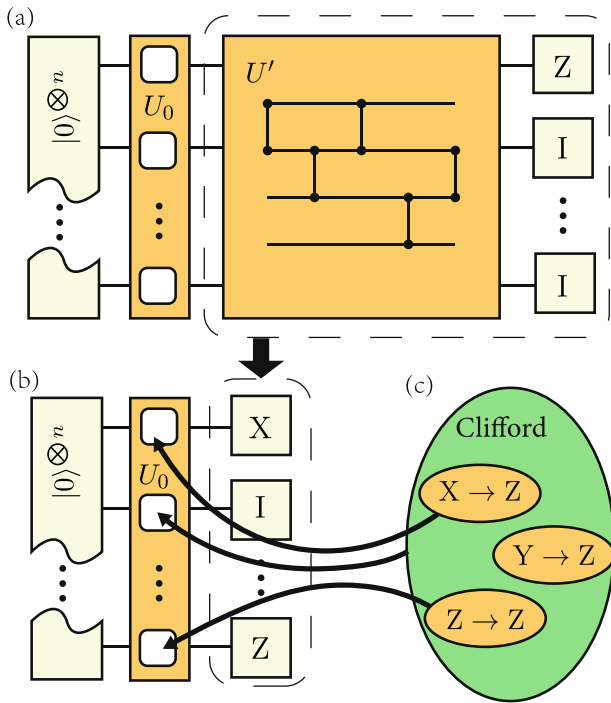
Fig. 3 **Error-sensitive circuit generation.** We compose an error-sensitive circuit with two sections U_0 and U', as shown in (**a**). U' is a Clifford operator. The observable is a Pauli operator, e.g. $Q = Z \otimes I \otimes \cdots \otimes I$. U' and the observable is equivalent to an effective observable $Q_{U'} = X \otimes I \otimes \cdots \otimes Z$, as shown in (**b**). Gates in U_0 are taken from the group of single-qubit Clifford gates. We choose the gates such that all non-identity Pauli operators in $Q_{U'}$ are mapped to $\pm Z$, as shown in (**c**).

Circuit frame

In the learning-based error mitigation, we aim at an optimised error mitigation formula that works for a set of circuits, including training circuits and circuits useful in some computation tasks. Choosing the target circuit set is important. When the circuit set is larger, it is harder to find a formula suitable for every circuit. Therefore, we want to be focusing on a circuit set relevant to some tasks to minimise bias. A way to construct a task-relevant circuit set is by taking circuits with the same pattern of multi-qubit Clifford gates, see Fig. 4. This pattern is called the circuit frame. In many quantum computing systems, such as superconducting qubits and trapped ions, the error rates of single-qubit gates are much lower than multi-qubit gates. Errors occurring in a circuit are mainly determined by multi-qubit gates. Therefore, all the circuits with the same frame have approximately the same errors, and we are able to correct them using the same error mitigation formula.

In the fixed-frame circuit set, single-qubit gates are variables. As shown in Fig. 4, the frame includes the qubit initialisation, multi-qubit Clifford gates and measurement. Fixing these operations, we change single-qubit gates to generate the circuit set. We call each variable single-qubit gate a slot. In Ref. [36], a setup with slots after each multi-qubit gate is proposed. Here we reduce the slot number to minimise the circuit set. We only take locations of single-qubit non-Clifford gates in the task circuit as slots and add two layers of slots after the initialisation and before the measurement, respectively. The reason is that a sequence of Clifford gates not interrupted by any non-Clifford gate can be treated as one multi-qubit Clifford gate.

The minimised slots have sufficient degrees of freedom for implementing Pauli twirling and probabilistic error cancellation for general error models. A Pauli error is an unwanted Pauli transformation stochastically occurring in the circuit. In Pauli
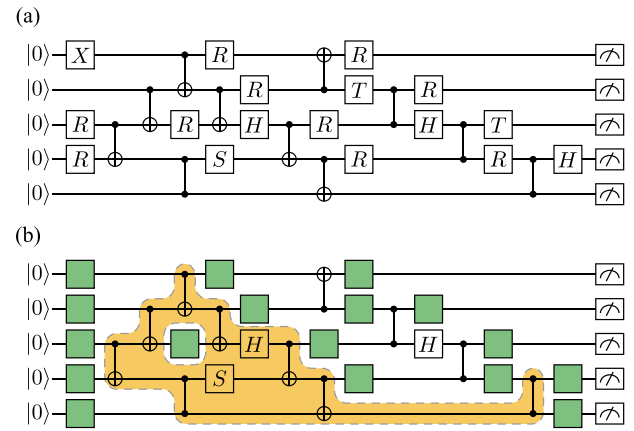

Fig. 4 **Quantum circuit and circuit frame. a** The circuit for a specific task. Single-qubit gates X, H and S are Clifford, and gates T and R are non-Clifford. **b** The task-dependent circuit frame. Green boxes are slots for variable single-qubit gates. Clifford gates in the yellow region with dashed borders form a composite Clifford gate.

twirling, we convert general errors into Pauli errors by randomly applying Pauli gates before and after each Clifford gate. We can correct a Pauli error by applying a Pauli gate to undo the error. Relevant discussions can be found in ref. [36].

With the frame determined, a circuit depends on the choice of single-qubit gates. Let $\boldsymbol{C} = (U_1, \ldots, U_N, Q)$ be a circuit (with two layers of single-qubit gates after the initialisation and before the measurement, respectively). The corresponding frame is $F = (\ldots, U_i, \bullet_k, \ldots, U_j, \bullet_q, \ldots, Q)$, where U_i is a gate on the frame, and $\bullet_k$ denotes a slot on qubit-k. In other words, F is the same as $\boldsymbol{C}$ except that gates in slots are replaced with $\bullet_k$. Formally, if $S = \{i_1, i_2, \ldots\}$ are labels of slots and $K = \{k_{i_1}, k_{i_2}, \ldots\}$ are corresponding qubits, the frame is $F = (F_1, \ldots, F_N, Q)$, where $F_i = U_i$ if $i \notin S$, and $F_i = \bullet_{k_i}$ if $i \in S$. Then, we can reexpress the circuit as $\boldsymbol{C} = [F, R_1, R_2, \ldots]$, where R_l is the single-qubit gate in the l-th slot, i.e. $U_{i_l} = I^{\otimes (k_{i_l} - 1)} \otimes R_l \otimes I^{\otimes (n - k_{i_l})}$.

To generate training circuits of the fixed frame, we can randomly draw the gate on each slot from the 24 single-qubit Clifford gates. Because the frame is formed of Clifford gates, the entire circuit constructed in this way is Clifford. It is likely that such a random circuit has a zero expected value. We can work out a circuit with a non-zero expected value by adjusting the first-layer gates, i.e. gates after the initialisation, as described in in the previous section. We give details of this procedure in Algorithm 1.

Algorithm 1. Generation of error-sensitive circuits.

1: **function** ESCIRCUIT($F, \overline{R}$)
2:     Compose the candidate circuit $\boldsymbol{C}' = [F, I, \ldots, I, R_{n+1}, \ldots, R_{N_R}]$.
3:     Calculate $Q_{U'} = U'^\dagger Q U'$.
4:     Calculate $(P'_1, P'_2, \ldots, P'_n)$ according to $Q_{U'} = \pm P'_1 \otimes P'_2 \otimes \cdots \otimes P'_n$.
5:     **for** $i = 1$ to n **do**
6:         **repeat**
7:             Choose a random R_i from C_1.
8:         **until** $R_i^\dagger P'_i R_i = \pm Z, I$
9:     Compose the error-sensitive circuit $\boldsymbol{C} = [F, R_1, \ldots, R_n, R_{n+1}, \ldots, R_{N_R}]$.
10:     Output $\boldsymbol{C}$.

Sampling algorithms

We give two algorithms for sampling error-sensitive Clifford circuits in Algorithms 2 and 3. For clarity, we use the following notations in the algorithms. F is the circuit frame, Q is the

observable, n is the qubit number, N_R is the slot number, and N_T is the sample number. C_1 is the single-qubit Clifford group with 24 elements. $U = U_N \cdots U_2 U_1$ is the unitary transformation of the circuit $\mathbf{C} = (U_1, \ldots, U_N, Q) = [F, R_1, R_2, \ldots]$. We use $\overline{R} = (R_{n+1}, R_{n+2}, \ldots, R_{N_R})$ to denote an ordered set of single-qubit Clifford gates, and $R_1, R_2, \ldots, R_n$ are gates in the first-layer slots. $w(\mathbf{C})$ is the weight of the Clifford circuit $\mathbf{C}$: $Q_U = U^\dagger Q U = \pm P_1 \otimes P_2 \otimes \cdots \otimes P_n$ is a tensor product of Pauli operators, then $w(\mathbf{C})$ is the number of non-identity Pauli operators in the product, i.e.

$$w(\mathbf{C}) \equiv n - \sum_{i=1}^{n} \delta_{I,P_i}, \tag{15}$$

where $\delta_{I,P_i} = 1$ if $P_i = I$, and $\delta_{I,P_i} = 0$ otherwise. In Algorithm 3, we employ the Metropolis-Hasting algorithm to realise a uniform distribution of error-sensitive circuits, which requires a conditional distribution $g(\overline{R}'|\overline{R})$ for suggesting a candidate sample. For example, we can take the conditional distribution as follows: we update gates in some randomly selected slots with newly generated random gates and keep gates in other slots unchanged.

Algorithm 2. Non-uniform importance Clifford sampling.
1: Input F.
2: **for** $t = 1$ to N_T **do**
3:     **for** $i = n+1$ to N_R **do**
4:         Choose a random R_i from C_1.
5:     Call ESCIRCUIT $F, \overline{R}$ to generate $\mathbf{C}$.
6:     Output $\mathbf{C}_t = \mathbf{C}$.

Algorithm 3. Uniform importance Clifford sampling.
1: Input F, a conditional distribution $g(\overline{R}'|\overline{R})$ and an initial slot-gate pattern $\overline{R}^{(0)}$.
2: Set $t = 0$.
3: Call ESCIRCUIT $F, \overline{R}^{(0)}$ to generate $\mathbf{C}$.
4: Take $\mathbf{C}_0 = \mathbf{C}$.
5: **for** $t = 1$ to N_T **do**
6:     Generate a random candidate of slot-gate pattern $\overline{R}^{(t)}$ according to $g(\overline{R}^{(t)}|\overline{R}^{(t-1)})$.
7:     Call ESCIRCUIT $F, \overline{R}^{(t)}$ to generate $\mathbf{C}$.
8:     Calculate the acceptance probability

$$A = \min\left(1, \frac{3^{-w(\mathbf{C})}}{3^{-w(\mathbf{C}_{t-1})}} \frac{g(\overline{R}^{(t-1)}|\overline{R}^{(t)})}{g(\overline{R}^{(t)}|\overline{R}^{(t-1)})}\right).$$

9:     Generate a uniform random number $u \in [0, 1]$.
10:    Accept and set $\mathbf{C}_t = \mathbf{C}$ if $u \leq A$.
11:    Reject and set $\mathbf{C}_t = \mathbf{C}_{t-1}$ if $u > A$.
12:    Output $\mathbf{C}_t$.

There is a relation between Clifford sampling and unitary sampling which allows us to estimate the bias distribution in general unitary circuits using Clifford circuits. We use $\mathbb{C}$ to denote the set of Clifford circuits and $\mathbb{U}$ to denote the set of all unitary circuits with the same frame. For a frame with N_R slots, the total number of Clifford circuits is $|\mathbb{C}| = 24^{N_R}$, i.e. each slot takes one of 24 single-qubit Clifford gates. In $\mathbb{U}$, each slot can take any single-qubit unitary. When errors are independent of the choice of single-qubit gates, MSEs are the same for the two circuit sets, i.e. $L_{\mathbb{U}} = L_{\mathbb{C}}$[42]. Because the set $\mathbb{C}$ is large, we need to use the Monte Carlo method to evaluate $L_{\mathbb{C}}$.

There is a similar relation between ICS and unitary sampling. Error-sensitive circuits are a subset of all Clifford circuits, denoted by $\mathbb{C}^{ES}$. According to Algorithm 1, given slot gates $\overline{R} = (R_{n+1}, R_{n+2}, \ldots, R_{N_R})$, the number of error-sensitive circuits is $8^{w(\mathbf{C})} 24^{n-w(\mathbf{C})}$. If $P'_i = I$, $R_i^\dagger P'_i R_i = I$ for all 24 single-qubit Clifford gates, which contributes a factor of 24; if $P'_i \neq I$, $R_i^\dagger P'_i R_i = \pm Z$ for 8 single-qubit Clifford gates, which contributes a factor of 8. The number of different $\overline{R}$'s is 24^{N_R-n}, then the total number of error-

sensitive circuits is

$$|\mathbb{C}^{ES}| = \sum_{j=1}^{24^{N_R-n}} 8^{w(\mathbf{C}_j)} 24^{n-w(\mathbf{C}_j)}, \tag{16}$$

where $\mathbf{C}_j$ are circuits with different $\overline{R}$'s. In a Clifford circuit, a Pauli error either preserves the Pauli observable or flips its sign. As a result, non-sensitive Clifford circuits do not respond to Pauli errors, i.e $y_{\mathbf{C}} = f_{\mathbf{C}}$ if $f_{\mathbf{C}} = 0$. Therefore,

$$L_{\mathbb{U}} = L_{\mathbb{C}} = \eta L_{\mathbb{C}^{ES}}, \tag{17}$$

for Pauli error models, where $\eta \equiv |\mathbb{C}^{ES}|/|\mathbb{C}|$ is the proportion of error-sensitive circuits in all Clifford circuits.

The distribution of error-sensitive circuits from Algorithm 2 is non-uniform. Because we uniformly choose slot gates in $\overline{R}$, the probability of an error-sensitive circuit $\mathbf{C}$ is

$$\begin{aligned} P_{nu}(\mathbf{C}) &= 24^{-(N_R-n)} 8^{-w(\mathbf{C})} 24^{-[n-w(\mathbf{C})]} \\ &= 24^{-N_R} 3^{w(\mathbf{C})}. \end{aligned} \tag{18}$$

Therefore, the probability of $\mathbf{C}$ is proportional to $3^{w(\mathbf{C})}$. If we use Algorithm 2 to sample circuits, we can evaluate $L_{\mathbb{C}^{ES}}$ according to

$$L_{\mathbb{C}^{ES}} = \eta^{-1} \mathrm{E}[3^{-w(\mathbf{C})}(y_{\mathbf{C}} - f_{\mathbf{C}})^2]_{nu}, \tag{19}$$

where the expected value is taken over the distribution $P_{nu}(\mathbf{C})$.

We can generate a uniform distribution of error-sensitive circuits as shown in Algorithm 3. In the uniform distribution, the probability of an error-sensitive circuit is $P_u(\mathbf{C}) = |\mathbb{C}^{ES}|^{-1}$. Then, we can evaluate $L_{\mathbb{C}^{ES}}$ with $L_{\mathbb{C}^{ES}} = \mathrm{E}[(y_{\mathbf{C}} - f_{\mathbf{C}})^2]_u$, where the expected value is taken over the distribution $P_u(\mathbf{C})$. By changing the formula of the acceptance probability, we can use the same algorithm to generate other distributions of error-sensitive circuits.

We now summarise the algorithms and analyse their classical-computing costs. Algorithm 1 is used to generate an error-sensitive circuit. Provided with an observable Q and a frame with n qubits and N two-qubit gates, Algorithm 1 includes operations that conjugate Q (line 3) via $O(N)$ Clifford gates and a conditioned random selection for the single-qubit gates in the first layer (line 5 to 8). The time cost of the conjugating operations is $O(nN)$ according to the efficient simulation algorithm for Clifford gates[39], and the time cost of selecting gates in the first layer is $O(n)$. Thus, the cost of Algorithm 1 is $O(nN)$. Algorithm 2 and Algorithm 3 are used to sample error-sensitive circuits according to the non-uniform distribution $P_{nu}(\mathbf{C})$ and uniform distribution $P_u(\mathbf{C})$, respectively. To generate N_T circuits, the costs for both algorithms are $O(N_T nN)$, because the elementary building block of both algorithms is nothing but the circuit generation given in Algorithm 1, which is repeated for N_T times. The numerical result in Supplementary Note 3 demonstrates that the number of error-sensitive circuits N_T required to perform learning-based error mitigation does not increase (as far as we have observed) with either the number of gates or the number of qubits. Overall, the cost scales linearly with the number of qubits and the number of gates. Noting that the sampling algorithms assume that two-qubit gates are Clifford and errors are independent of single-qubit gates. We give discussion in Supplementary Note 4 about the implementation of the algorithms when the assumptions are not satisfied.

Phenomenological-error model

In this section, we introduce the phenomenological-error model which quantifies the bias caused by realistic errors in a circuit. Then, we show that the phenomenological-error model can be effectively represented by a global depolarising model with fluctuation, and the fluctuation is $O(1/\sqrt{N})$ times smaller than the depolarising rate. This result suggests that, if we are able to use error mitigation to cancel the impact of the effective global

depolarising error, we can reduce the bias caused by realistic errors by a factor of $O(1/\sqrt{N})$.

Before introducing our phenomenological-error model, we give a brief overview of realistic error models. Consider a quantum gate with the unitary operator U_i, the error-free output state of the gate is $[U_i]\rho_i$, where ρ_i is the input state. When the gate is imperfect, we can always express the output state with error as $\mathcal{N}_i[U_i]\rho_i$ (assuming the noisy circuit is a Markov process), where the completely positive map $\mathcal{N}_i$ describes the effect of noise associated with the gate. In the global depolarising model, $\mathcal{N}_i = (1-\epsilon)[\mathbb{1}] + \epsilon\mathcal{D}$. In realistic error models, $\mathcal{N}_i$ is usually caused by local processes, such as dephasing, dissipation and imperfections in the coherent evolution. If the gate acts on qubit-1 and qubit-2, the noise mainly affects these two qubits. Taking a Pauli error model as an example, the noise map reads

$$\mathcal{N}_i = \left(1 - \frac{16\epsilon}{15}\right)[I^{\otimes n}] + \frac{16\epsilon}{15}\mathcal{D}_{1,2}, \tag{20}$$

where

$$\mathcal{D}_{1,2} \equiv \frac{1}{16}\sum_{P_1,P_2=I,X,Y,Z}[P_1 \otimes P_2 \otimes I^{\otimes(n-2)}]. \tag{21}$$

We call this particular Pauli error model the gate depolarising model, in which probabilities of Pauli errors are the same. We can rewrite this summation-form error model into the product form

$$\mathcal{N}_i = \prod_{P_1,P_2=I,X,Y,Z}\left[(1-p)[I^{\otimes n}] + p[P_1 \otimes P_2 \otimes I^{\otimes(n-2)}]\right], \tag{22}$$

where $p \simeq \epsilon/15$. In the product form, the noise map is a product of 15 independent maps, and we call each of them a Pauli error channel.

The global depolarising model with fluctuation can characterise the impact of realistic errors in large circuits. Given a circuit $\mathbf{C}$, the error-free final state is $\rho_0 = U|0\rangle\langle 0|^{\otimes n}U^\dagger$. In our error model, the erroneous final state is $\rho = (1-\epsilon_{\mathbf{C}})\rho_0 + \epsilon_{\mathbf{C}}\rho_m$, where $\epsilon_{\mathbf{C}}$ is the circuit-dependent depolarising rate. According to this model, we have $y_{\mathbf{C}} = (1-\epsilon_{\mathbf{C}})f_{\mathbf{C}}$. If we allow $\epsilon_{\mathbf{C}}$ to be any value (rather than limited in the interval $[0,1]$), this error model is a general phenomenological-error model. Given any $f_{\mathbf{C}}$ and $y_{\mathbf{C}}$, the corresponding depolarising rate is $\epsilon_{\mathbf{C}} = 1 - y_{\mathbf{C}}/f_{\mathbf{C}}$. Note that the bias is $\epsilon_{\mathbf{C}}f_{\mathbf{C}}$, which is always finite even when $f_{\mathbf{C}} = 0$ and $\epsilon_{\mathbf{C}}$ is infinite.

We write the circuit-dependent depolarising rate as two terms, the average and fluctuation, i.e. $\epsilon_{\mathbf{C}} = \epsilon_0 + \delta\epsilon_{\mathbf{C}}$, where

$$\epsilon_0 \equiv \frac{\langle \epsilon_{\mathbf{C}} f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}}{\langle f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}} \tag{23}$$

is the average depolarising rate with the weight $f_{\mathbf{C}}^2$, and $\delta\epsilon_{\mathbf{C}}$ is the circuit-dependent fluctuation. We characterise the fluctuation with the weighted standard deviation

$$\Delta \equiv \sqrt{\frac{\langle \delta\epsilon_{\mathbf{C}}^2 f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}}{\langle f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}}}. \tag{24}$$

The key result is that Δ increases with the gate number as $O(N^\gamma)$, and $\gamma \approx 0.5$, see Fig. 2.

In the rest part of this section, we show theoretically that the standard deviation Δ is proportional to $\sqrt{N}$ using a Pauli error model. In the next two sections, we introduce an error mitigation protocol inspired by the phenomenological-error model, then we verify the scaling behaviour in numerical simulations of the gate depolarising model, composite error models involving Pauli, amplitude damping and coherent errors, and a model with single-qubit-gate-dependent errors. The $\sqrt{N}$ scaling is observed in all the error models.

We focus on Pauli errors to analyse the fluctuation in the phenomenological-error model. For general errors, we can use Pauli twirling to convert them into Pauli errors. If error mitigation is concatenated with error correction, logical errors after correction are mainly Pauli errors[43]. Suppose errors are independent of single-qubit gates, we have the following relations,

$$\langle f_{\mathbf{C}}^2 \rangle_{\mathbb{U}} = \langle f_{\mathbf{C}}^2 \rangle_{\mathbb{C}} = \eta\langle f_{\mathbf{C}}^2 \rangle_{\mathbb{C}^{ES}}, \tag{25}$$

$$\langle f_{\mathbf{C}}y_{\mathbf{C}} \rangle_{\mathbb{U}} = \langle f_{\mathbf{C}}y_{\mathbf{C}} \rangle_{\mathbb{C}} = \eta\langle f_{\mathbf{C}}y_{\mathbf{C}} \rangle_{\mathbb{C}^{ES}}, \tag{26}$$

$$\langle y_{\mathbf{C}}^2 \rangle_{\mathbb{U}} = \langle y_{\mathbf{C}}^2 \rangle_{\mathbb{C}} = \eta\langle y_{\mathbf{C}}^2 \rangle_{\mathbb{C}^{ES}}, \tag{27}$$

where $\mathbb{U}$, $\mathbb{C}$ and $\mathbb{C}^{ES}$ are circuit sets with the same frame. In the above equations, the first equal sign follows because the Clifford group is a unitary-2 design[42,44], and therefore $\langle \bullet \rangle_{\mathbb{U}} = \langle \bullet \rangle_{\mathbb{C}}$ holds if $\bullet$ is a polynomial of degree two in the gate unitaries. The second equal sign is a consequence of $f_{\mathbf{C}} = 0$ when $\mathbf{C} \notin \mathbb{C}^{ES}$ and $\eta = |\mathbb{C}^{ES}|/|\mathbb{C}|$. Using $f_{\mathbf{C}} = \pm 1$ for error-sensitive circuits, we can obtain

$$\eta = \langle f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}, \tag{28}$$

$$\epsilon_0 = \langle \epsilon_{\mathbf{C}} \rangle_{\mathbb{C}^{ES}}, \tag{29}$$

$$\Delta = \sqrt{\langle \delta\epsilon_{\mathbf{C}}^2 \rangle_{\mathbb{C}^{ES}}}. \tag{30}$$

These relations allow us to study ϵ_0 and Δ with error-sensitive circuits.

For simplicity, we consider an error model where two-qubit gates are the dominant sources of errors in actual quantum computing devices. We assume that the initialisation, single-qubit gates and measurement are perfect. In a two-qubit gate, we assume that the probability of Pauli errors are the same, i.e. the gate depolarising model. We use N' to denote the number of two-qubit gates.

The effect of local Pauli errors is equivalent to that of global depolarising errors in error-sensitive circuits. The unitary transformation of a circuit with N gates is $U = U_N \cdots U_1$. If a Pauli error σ occurs after the ith gate, the transformation becomes $U' = U_N \cdots U_{i+1}\sigma U_i \cdots U_1 = \sigma'_{\mathbf{C}}U$, where $\sigma'_{\mathbf{C}} = U_N \cdots U_{i+1}\sigma U_{i+1}^\dagger \cdots U_N^\dagger$ is the Pauli error propagated to the end of the circuit. Because gates are Clifford, $\sigma'_{\mathbf{C}}$ is also a Pauli operator, i.e. any Pauli error in the circuit is equivalent to a Pauli error at the end of the circuit. If the probability of the Pauli error is p, i.e. the error channel is $(1-p)[\mathbb{1}] + p[\sigma]$, the final state of the circuit is transformed from ρ_0 to $(1-p)\rho_0 + p[\sigma'_{\mathbf{C}}]\rho_0$. Then there are two cases: If $\sigma'_{\mathbf{C}}$ and the Pauli observable Q are commutative, the expected value is preserved under the Pauli error; otherwise, the expected value is changed from $f_{\mathbf{C}}$ to $(1-2p)f_{\mathbf{C}}$, i.e. the equivalent depolarising rate is $2p$.

The overall depolarising rate depends on the number of Pauli error channels. Each two-qubit gate contributes 15 Pauli error channels according to the product form of the Pauli error model. For a circuit with N' two-qubit gates, there are $M = 15N'$ error channels. Let $(1-p)[\mathbb{1}] + p[\sigma_k]$ be the k-th error channel, $(1-p)[\mathbb{1}] + p[\sigma'_{k,\mathbf{C}}]$ is the corresponding error channel at the end of the circuit. We use the binary number $t_k(\mathbf{C})$ to denote whether the k-th error channel affect the observable, i.e. $t_k(\mathbf{C}) = 0$ if $\sigma'_{k,\mathbf{C}}$ and Q are commutative, and $t_k(\mathbf{C}) = 1$ otherwise. Then, the expected value is changed to $\prod_{k=1}^M (1-2p)^{t_k(\mathbf{C})}f_{\mathbf{C}}$. The equivalent depolarising rate is

$$\epsilon_{\mathbf{C}} = 1 - \prod_{k=1}^M (1-2p)^{t_k(\mathbf{C})} = \sum_{k=1}^M 2t_k(\mathbf{C})p + O(p^2). \tag{31}$$

The average depolarising rate is proportional to the gate number, and the standard deviation is proportional to the square root of the gate number. We can understand this phenomenon as follows. If we choose the circuit randomly from the circuit set, each error channel is switched on and off randomly, i.e. each t_k takes a random value. Under the assumption that t_k are independent and

identically distributed random variables, the distribution of $\epsilon_{\boldsymbol{C}}$ is binomial. Let P be the probability of $t_k = 1$ and neglect $O(p^2)$ terms, the average depolarising rate is $\epsilon_0 \simeq 2pMP$, and the standard deviation is $\Delta \simeq \sqrt{2pMP(1-P)}$. Note that M is proportional to the gate number.

In large circuits, the global depolarising model with the depolarising rate ϵ_0 is an approximate phenomenological-error model. When we sample circuits composed of noisy gates, the circuit plays the role of a sampler, i.e. the impact of each gate error is a random variable dependent on the circuit configuration. In a certain regime, the total impact is the summation of individual gate errors. When the gate number is larger, the number of random variables in the summation is larger. According to the law of large numbers, the relative standard deviation of the summation decreases with the number of random variables, i.e.

$$\frac{\Delta}{\epsilon_0} \propto \frac{1}{\sqrt{M}}, \tag{32}$$

where $M \propto N' \sim N$. Therefore, $\epsilon_{\boldsymbol{C}}$ is in the vicinity of ϵ_0 with a high probability in large circuits.

The analysis above has shown that local gate errors can be represented by a fluctuating global depolarising error, and the ratio of the fluctuation Δ to the depolarising rate ϵ_0 is in proportion to $1/\sqrt{N}$. This result will be verified by the numerical simulations in the next two sections. We will show that, if the effective global depolarising error is removed by error mitigation, the remaining error (caused by the fluctuation) scales with the gate number as $1/\sqrt{N}$. In addition, we numerically illustrate the error propagation model used in the above analysis. We show that the overall effect of propagated gate errors will become close to the global depolarising error and the relative difference between them decreases as $1/\sqrt{N}$. We leave the numerical result of error propagation to Supplementary Note 2.

The analysis in this section assumes a small total error rate pM. Under this assumption, we can neglect contributions from the second order in Eq. (31). In the section of "Numerical results of the scaling behaviour", we randomly take total error rates from about 0.003 to 0.3, and we observe the $\sqrt{N}$ scaling behaviour. We remark that a modest total error rate is a general requirement of quantum error mitigation[45,46]. Unlike quantum error correction, which actively detects and corrects errors in the circuit, most quantum error mitigation protocols correct the result by post-processing the noisy experimental data. When the total error rate is high, i.e. the fidelity approaches zero, the raw data lose the information about the correct quantum state, from which post-processing cannot recover the information. For example, in probabilistic error cancellation, the sampling overhead is exponential in the number of gates given a constant error rate per gate[7,8].

Error mitigation according to the phenomenological-error model

According to the phenomenological-error model, the effective depolarising rate in large circuits is ϵ_0 with a small fluctuation. We can mitigate errors by compensating the effect of ϵ_0. We use the root mean square error (RMSE) as the measure of the overall accuracy of an error mitigation formula in a circuit set. Before error mitigation, RMSE of unitary circuits with the same frame is $\sqrt{\langle (y_{\boldsymbol{C}} - f_{\boldsymbol{C}})^2 \rangle_{\mathbb{U}}} = \sqrt{\eta(\epsilon_0^2 + \Delta^2)} \simeq \sqrt{\eta}\epsilon_0$, which increases linearly with the gate number. Using the error mitigation formula $y'_{\boldsymbol{C}} = (1 - \epsilon_0)^{-1} y_{\boldsymbol{C}}$, we can reduce RMSE to $\sqrt{\langle (y'_{\boldsymbol{C}} - f_{\boldsymbol{C}})^2 \rangle_{\mathbb{U}}} = (1 - \epsilon_0)^{-1} \sqrt{\eta}\Delta \simeq \sqrt{\eta}\Delta$, which increases sublinearly with the gate number. Because $\epsilon_0 = 1 - \langle y_{\boldsymbol{C}} f_{\boldsymbol{C}} \rangle_{\mathbb{C}^{ES}}$, we can measure ϵ_0 (and Δ) by uniformly sampling error-sensitive circuits. Actually, because the fluctuation is small, we can even take $\hat{\epsilon}_0 = 1 - y_{\boldsymbol{C}} f_{\boldsymbol{C}}$ for one randomly generated error-sensitive circuit $\boldsymbol{C} \in \mathbb{C}^{ES}$, and it is likely that the error mitigation formula still
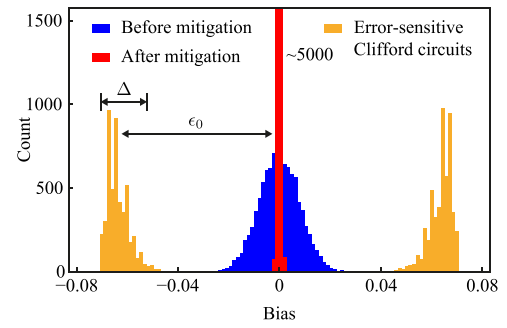

Fig. 5 Distributions of the bias for six-qubit periodic-cycling circuits with 72 two-qubit gates under the gate depolarising noise. The error rate per gate is 0.001. Before error mitigation, the bias distribution of unitary circuits (the blue histogram) has a shape similar to the Gaussian distribution, and the bias distribution of error-sensitive circuits (the orange histogram) is concentrated at two values. When we mitigate errors according to the average depolarising rate ϵ_0, we move the two peaks to the centre, and the residual bias is determined by the width of the two peaks. Because of the equivalence between the importance Clifford sampling and unitary sampling, the bias of unitary circuits is significantly reduced after error mitigation (the red histogram).

works. This phenomenological-error-model inspired (PEMI) error mitigation protocol is illustrated in Fig. 5.

Similar protocols that mitigate errors according to the global depolarising model have been proposed in Refs. [37,47,48]. In these protocols, the effective depolarising rate is measured in different ways. Before considering general error mitigation formulas, we take the PEMI protocol as an example to verify the phenomenological-error model, because the bias of this protocol is directly related to the fluctuation.

In the PEMI protocol, we can further reduce RMSE by optimising the error mitigation formula. If we take

$$y'_{\boldsymbol{C}} = \frac{1 - \epsilon_0}{(1 - \epsilon_0)^2 + \Delta^2} y_{\boldsymbol{C}}, \tag{33}$$

RMSE after mitigation is reduced to

$$\sqrt{\langle (y'_{\boldsymbol{C}} - f_{\boldsymbol{C}})^2 \rangle_{\mathbb{U}}} = \frac{\sqrt{\eta}\Delta}{\sqrt{(1 - \epsilon_0)^2 + \Delta^2}}. \tag{34}$$

Numerical results of the scaling behaviour

In this section, we numerically test the PEMI error mitigation formula and verify the scaling behaviour of ϵ_0 and Δ. Results of other error mitigation formulas will be given in the next section.

To demonstrate the scaling behaviour, we generate three families of circuits. In periodic-cycling circuits, two-qubit gates are arranged according to a fixed pattern, and we increase the circuit depth by repeating the pattern. Therefore, periodic-cycling circuits are deterministic. In linear-network circuits, two-qubit gates only act on the nearest neighbouring qubits on a one-dimensional qubit array, and we randomly place two-qubit gates in the circuit. In all-to-all-network circuits, two-qubit gates are also arranged randomly but they can act on any pair of qubits.

We use three types of error models in our numerical calculations: the gate depolarising model with a randomly selected error rate, randomly generated composite error models and a model with single-qubit-gate-dependent errors. The gate depolarising model is used to derive the phenomenological-error model, but the conclusion holds for other error models. The composite error model involves gate depolarising, dephasing, amplitude damping and coherent errors, which are the typical
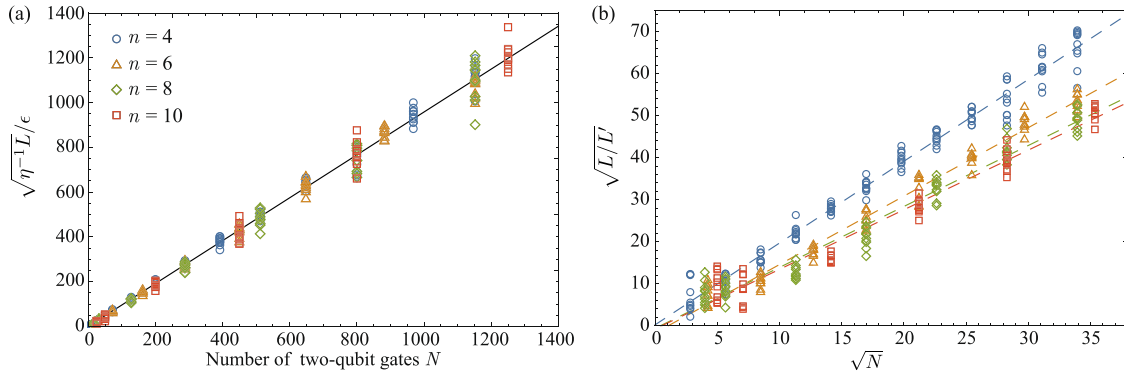
Fig. 6 Root mean square errors of linear-network circuits with the gate depolarising model. a Root mean square error $\sqrt{L}$ before error mitigation. **b** Root mean square error $\sqrt{L'}$ after error mitigation. In the numerical simulation, we randomly generate a circuit frame with n qubits and N two-qubit gates, and we randomly take the error rate per gate ϵ. We generate 1000 Clifford circuits according to Algorithm 2 to estimate the phenomenological-error rate and then generate 1000 random unitary circuits to compute L and L'.

error sources in actual devices. We generate different composite error models by randomly choosing the weight of each component and observe the same scaling behaviour as the gate depolarising model. The equivalence between Clifford sampling and unitary sampling is also used in deriving the phenomenological-error model, which is under the condition that errors are single-qubit-gate independent. In the numerical result, we find that the conclusion on the scaling behaviour holds even if errors are single-qubit-gate dependent. See the Methods section for details of numerical calculations.

By compensating the average depolarising rate, we can reduce RMSE from $\sqrt{L} = \sqrt{\langle (y_{\mathbf{C}} - f_{\mathbf{C}})^2 \rangle_{\mathbb{U}}} \simeq \sqrt{\eta} \epsilon_0$ to $\sqrt{L'} = \sqrt{\langle (y'_{\mathbf{C}} - f_{\mathbf{C}})^2 \rangle_{\mathbb{U}}} \simeq \sqrt{\eta} \Delta$. According to the discussion in the section of "Phenomenological-error model", $\epsilon_0 \propto N$ and $\Delta \propto \sqrt{N}$. Therefore, RMSE is reduced in error mitigation by a factor of $\Delta / \epsilon_0 \propto 1/\sqrt{N}$. We verify these scaling behaviours by applying the error mitigation formula in Eq. (33) to randomly generated circuits with up to ten qubits and more than a thousand two-qubit gates. To implement the formula, ϵ_0 and Δ are measured by sampling error-sensitive circuits. RMSEs before and after error mitigation $\sqrt{L}$ and $\sqrt{L'}$ are calculated and plotted in Figs. 6 and 7. For the model with single-qubit-gate-dependent errors, we directly calculate and plot ϵ_0 and Δ in Fig. 8. We can find that numerical results are consistent with scaling behaviours predicted by the phenomenological-error model. In addition, we perform experiments on IBM quantum computers[49] and observe good agreement between the numerical and experimental results. We include the experimental results in Supplementary Note 6.

In Fig. 7, the error suppression ratio $\sqrt{L/L'}$ for all-to-all-network circuits meets $\sqrt{L/L'} = a\sqrt{N}$ and a is a positive number independent of the qubit number. However, in Fig. 6, we find that a for linear-network circuits decreases with the qubit number. The difference between all-to-all-network and linear-network circuits is that two-qubit gates in linear-network circuits are short-range, thus it requires more gates for the error on one qubit to propagate across the circuit network.

The error suppression ratio $\sqrt{L/L'}$ are obtained via averaging random unitary circuits, which usually have near-zero expected values. However, in common quantum applications such as variational quantum eigensolver, the expected value is far from zero, which is atypical for random unitary circuits. Thus, we come to ask the question of whether the average suppression ratio of random unitary circuits is also the error suppression ratio of these atypical circuits. To answer this question, we numerically investigate the dependence of the error suppression ratio on the error-free expectation. The numerical result is illustrated in Supplementary Note 5, and the answer is which demonstrates

that the average error suppression ratio can be applied to these atypical circuits.

We note that the $\sqrt{N}$ scaling of error-mitigated result relies on a modest total error rate. This condition is essential for quantum error mitigation methods to work properly[45,46] and is considered as a general requirement of NISQ computation[3]. For each data point in Figs. 6 and 7, we randomly choose the error rate per gate ϵ such that the total error rate $N\epsilon$ is in the interval about 0.003 to 0.3.

Error scaling in optimised error mitigation formulas

In this section, we utilise the phenomenological-error model to show that one can suppress the scaling of the residual bias in a learning-based manner. For imperfect error extrapolation and probabilistic error cancellation, the error scaling after the optimisation is $\propto \sqrt{N}$. The imperfections are due to the imperfect control of noise in error extrapolation and inaccurate knowledge of the error model in probabilistic error cancellation. For virtual distillation, the result is similar.

First, we analyse the error scaling of error extrapolation. An error mitigation formula usually involves multiple circuits. For each of them, we can effectively characterise the impact of noise using our phenomenological-error model. Taking the linear error extrapolation as an example, the two circuits $\mathbf{C}_1$ and $\mathbf{C}_2$ are the same as the primitive circuit $\mathbf{C}$, but the noise level is doubled in $\mathbf{C}_2$. In the phenomenological-error model of the circuit $\mathbf{C}_i$, the average depolarising rate is ϵ_i, the rate fluctuation is $\delta\epsilon_{\mathbf{C},i}$ and the standard deviation is Δ_i. Because $\mathbf{C}_1$ and $\mathbf{C}_2$ are the same circuit, their fluctuations are correlated: Suppose effective depolarising rates are approximately proportional to the noise level, we have $\epsilon_2 \simeq 2\epsilon_1$ and $\delta\epsilon_{\mathbf{C},2} = 2\delta\epsilon_{\mathbf{C},1}$. Therefore, the fluctuation-caused bias depends on the covariance matrix $K_{i,j} \equiv \eta^{-1} \langle \delta\epsilon_{\mathbf{C},i} \delta\epsilon_{\mathbf{C},j} f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}$.

For the linear extrapolation formula in Eq. (12), RMSE after mitigation depends on average depolarising rates ϵ_i and the covariance matrix K, i.e.

$$\sqrt{\langle (y'_{\mathbf{C}} - f_{\mathbf{C}})^2 \rangle_{\mathbb{U}}} = \sqrt{\eta[(E^\dagger \Lambda - 1)^2 + \Lambda^\dagger K \Lambda]}, \qquad (35)$$

where $E = (1 - \epsilon_1, 1 - \epsilon_2)^{\mathsf{T}}$ and $\Lambda = (\lambda, 1 - \lambda)^{\mathsf{T}}$. Taking $\lambda = \epsilon_2/(\epsilon_2 - \epsilon_1)$, we can remove the contribution of average depolarising rates, and RMSE becomes $\sqrt{\langle (y'_{\mathbf{C}} - f_{\mathbf{C}})^2 \rangle_{\mathbb{U}}} = \sqrt{\eta \Lambda^\dagger K \Lambda} \leq \sqrt{\eta(\Delta_1^2 + \Delta_2^2)} \|\Lambda\| \propto \sqrt{N}$. Here, we have used that K is positive semi-definite, Δ_1^2 and Δ_2^2 are diagonal elements of K, and $\|\Lambda\| \simeq \sqrt{5}$ does not change significantly with the gate number. Note that this upper bound holds even if the noise is not increased as designed, and we can further reduce RMSE by optimising the parameter λ. In Fig. 9, we plot RMSE before and after error mitigation. In the optimised error mitigation formula, we take $\lambda = \epsilon_2/(\epsilon_2 - \epsilon_1)$. The numerical result is
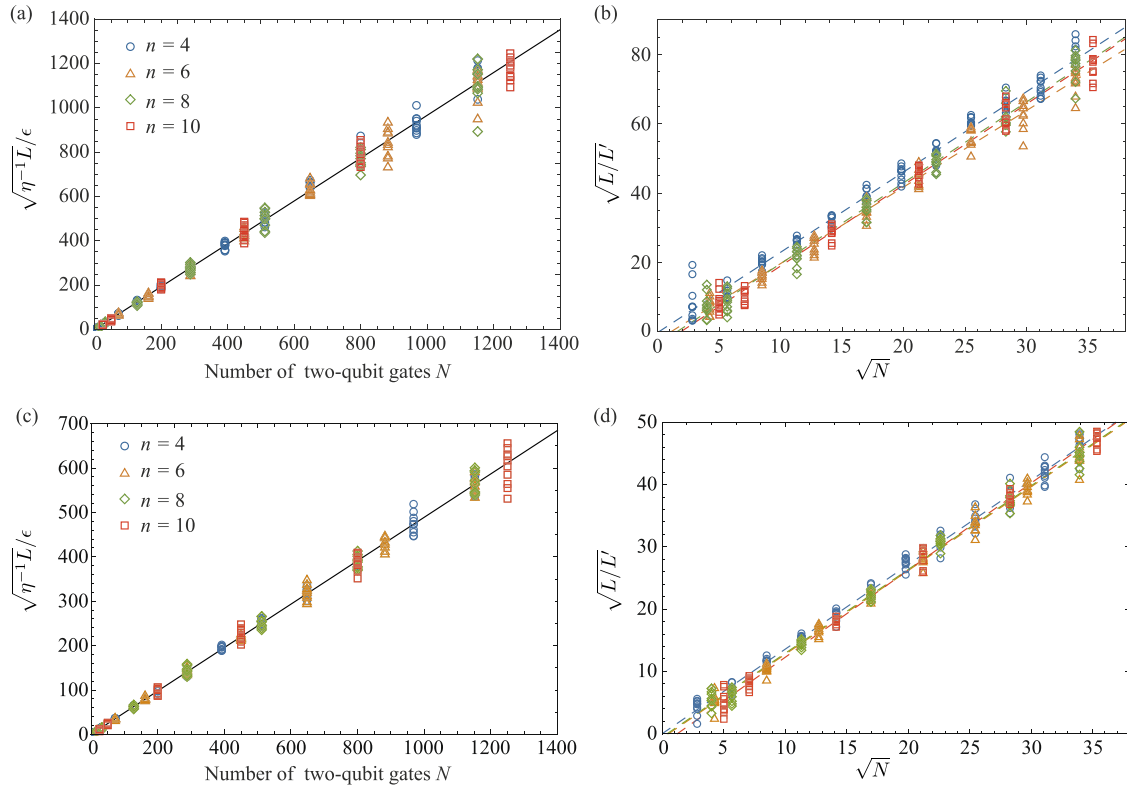
Fig. 7 Root mean square errors of all-to-all-network circuits. a, c Root mean square error $\sqrt{L}$ before error mitigation. **b, d** Root mean square error $\sqrt{L'}$ after error mitigation. The results in (**a**) and (**b**) are obtained with the gate depolarising model, and the results in (**c**) and (**d**) are obtained with the composite model. In the numerical simulation, we randomly generate a circuit frame with n qubits and N two-qubit gates, and we randomly take the error rate per gate ϵ. We generate 1000 Clifford circuits according to Algorithm 2 to estimate the phenomenological-error rate and then generate 1000 random unitary circuits to compute L and L'.
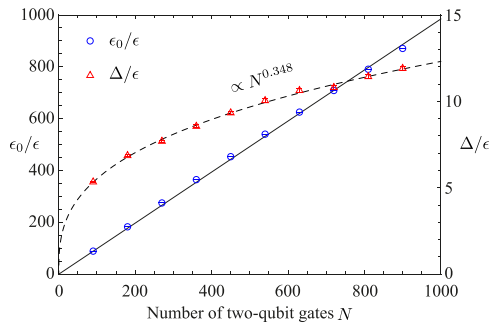
Fig. 8 Average depolarising rate ϵ_0 and standard deviation Δ in six-qubit periodic-cycling circuits. The axis on the left corresponds to ϵ_0 and the axis on the right corresponds to Δ. The error rate per two-qubit gate is $\epsilon = 2 \times 10^{-4}$, and the error rate of a single-qubit gate R is $0.1\pi^{-1}\epsilon \arccos\frac{|\mathrm{Tr}(R)|}{2}$. The error bar represents one standard deviation.

consistent with the scaling behaviour predicted by the phenomenological-error model.

Theorem 1. Consider the general extrapolation formula in Eq. (2), let ϵ_i, $\delta\epsilon_{\mathbf{C},i}$ and Δ_i be the average depolarising rate, rate fluctuation and standard deviation of the circuit $\mathbf{C}_i$, respectively, then

$$\min_{\{q_i\}} \sqrt{\langle (y'_{\mathbf{C}} - f_{\mathbf{C}})^2 \rangle_{\mathbb{U}}} \leq \frac{\sqrt{\eta E^\dagger K E}}{\|E\|^2} \leq \frac{\sqrt{\eta \sum_i \Delta_i^2}}{\|E\|}.$$

where $E = (1 - \epsilon_1, 1 - \epsilon_2, \dots)^{\mathsf{T}}$, $K_{i,j} = \eta^{-1}\langle \delta\epsilon_{\mathbf{C},i}\delta\epsilon_{\mathbf{C},j}f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}$ and $\eta = \langle f_{\mathbf{C}}^2 \rangle_{\mathbb{U}}$.

The proof is straightforward. Let $\Lambda = (q_1, q_2, \dots)^{\mathsf{T}}$, the expression of RMSE is the same as Eq. (35). We can prove the theorem by taking $\Lambda = E/\|E\|^2$.

Second, we investigate the error scaling of probabilistic error cancellation. In probabilistic error cancellation, we reconstruct the transformation of the ideal circuit as a linear combination of transformations of noisy circuits. A practical way is decomposing each ideal gate in the circuit as a linear combination of noisy gates. In general, we can work out the decomposition as follows. If U_i is the unitary operator of the ideal gate, the completely positive map of the noisy gate is $\mathcal{N}_i[U]$. We can cancel the noise by applying an inverse noise $\widetilde{\mathcal{N}}_i^{-1} = \sum_k q_{i,k}\mathcal{E}_{i,k}$ after the noisy gate, and the overall effective gate is $\widetilde{\mathcal{N}}_i^{-1}\mathcal{N}_i[U]$. Here, $\mathcal{E}_{i,k}$ are some noisy gates, i.e. we insert the gate $\mathcal{E}_{i,k}$ after the gate $\mathcal{N}_i[U]$ with the quasi-probability $q_{i,k}$. If $\widetilde{\mathcal{N}}_i^{-1} = \mathcal{N}_i^{-1}$, the error in the gate is completely removed; otherwise, effective noise in the gate is $\widetilde{\mathcal{N}}_i^{-1}\mathcal{N}_i$.

We consider a Pauli error model with gate depolarising errors and dephasing errors as an example. For a two-qubit gate on qubit-1 and qubit-2, the noise map is

$$\begin{aligned} \mathcal{N}_i &= \left(1 - \frac{16\epsilon_d}{15} - \epsilon_z\right)[I^{\otimes n}] \\ &\quad + \frac{16\epsilon_d}{15}\mathcal{D}_{1,2} + \frac{\epsilon_z}{2}([Z_1] + [Z_2]), \end{aligned} \tag{36}$$

where $Z_i = [I^{\otimes(i-1)} \otimes Z \otimes I^{\otimes(n-i)}]$. Suppose our knowledge about the noise map is inaccurate and we correct the error according to
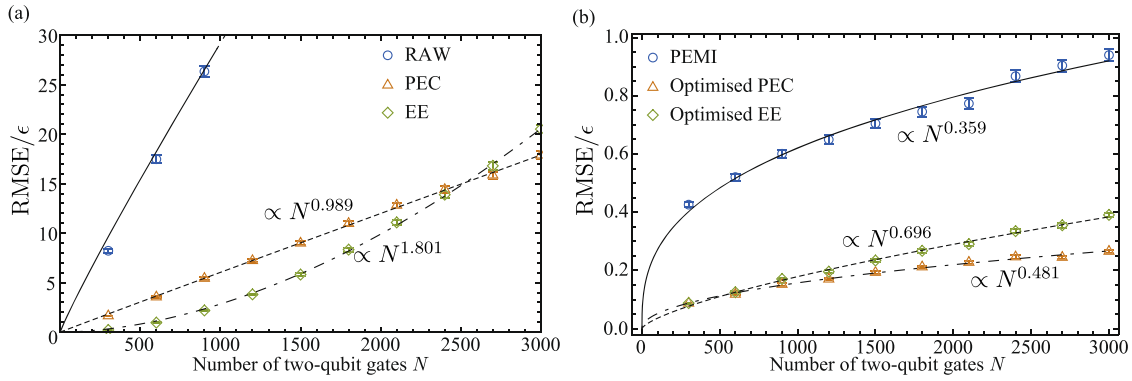
Fig. 9 Root mean square errors (RMSE) in error mitigation protocols. The result is obtained using ten-qubit periodic-cycling circuits under the error model in Eq. (36) with $\epsilon_d = 8 \times 10^{-5}$ and $\epsilon_z = 2 \times 10^{-5}$, and we use 1000 Clifford circuits generated via Algorithm 2 for the training and 1000 unitary circuits to compute the RMSE. The error bar represents one standard deviation. In the raw result without error mitigation, RMSE increases linearly with the gate number. In error extrapolation (EE), noise is increased imperfectly: $\epsilon_d = 1.8 \times 10^{-4}$ and $\epsilon_z = 2 \times 10^{-5}$ in the error model with a doubled error rate, i.e. only the gate depolarising component is increased. In probabilistic error cancellation (PEC), we take the inverse map in Eq. (37) according to an inaccurate error model with only gate depolarising errors, i.e. we take $\lambda = -16\epsilon_d/(15 - 16\epsilon_d)$ before the optimisation and the optimal value after the optimisation.

the gate depolarising model, we have

$$\widetilde{\mathcal{N}}_i^{-1} = (1 - \lambda)[I^{\otimes n}] + \lambda \mathcal{D}_{1,2}. \qquad (37)$$

When $\lambda = -16\epsilon_d/(15 - 16\epsilon_d)$ and $\epsilon_z = 0$, we can correct all errors in the gate; otherwise, the effective gate has a finite error rate.

We can suppress the error scaling in imperfect probabilistic error cancellation by optimisation. For an error mitigation formula worked out according to an inaccurate error model, we can treat it as having a virtual quantum computer, in which the error model is given by $\widetilde{\mathcal{N}}_i \mathcal{N}_i$. Then, we can describe the error in this virtual machine using the phenomenological-error model and reduce the bias using the PEMI protocol. We can use the formula $y''_{\mathbf{C}} = (1 - \epsilon'_0)y'_{\mathbf{C}}$, where ϵ'_0 and $y'_{\mathbf{C}}$ are respectively the average depolarising rate and expected value in the virtual machine. Then the residual bias of $y''_{\mathbf{C}}$ is determined by the standard deviation Δ' of the virtual machine. Actually, it is not necessary to modify the formula to suppress the error scaling. For example, we can take λ in Eq. (37) as a variational parameter and optimise it in ICS. The numerical result in Fig. 9 shows that RMSE of probabilistic error cancellation with the optimised λ scales as $\propto \sqrt{N}$.

Third, we investigate the error scaling of virtual distillation. The virtual distillation formula is nonlinear unlike error extrapolation and cancellation. For a general error mitigation formula, suppose the truncation on the Taylor expansion is valid, we have

$$y'_{\mathbf{C}} \simeq F(a_1 f_{\mathbf{C}_1}, a_2 f_{\mathbf{C}_2}, \cdots) + \sum_i \frac{\partial F}{\partial y_{\mathbf{C},i}} \delta\epsilon_{\mathbf{C},i}, \qquad (38)$$

where $a_i = 1 - \epsilon_i$. In Eq. (38), we have considered the general error mitigation formula in Eq. (1) and $y_{\mathbf{C},i} = (1 - \epsilon_i + \delta\epsilon_{\mathbf{C},i})f_{\mathbf{C}_i}$. If we can remove the zeroth-order term (contribution of average depolarising rates) by taking proper variational parameters in the formula, the bias is determined by fluctuations. For virtual distillation, $F(a_1 f_{\mathbf{C}_1}, a_2 f_{\mathbf{C}_2}) = a_1 f_{\mathbf{C}_1}/(a_2 f_{\mathbf{C}_2})$, therefore, we can compensate average depolarising rates by a factor. In the numerical simulation, we determine the factor by taking the original virtual distillation formula $y'_{\mathbf{C}} = y_{\mathbf{C}_1}/y_{\mathbf{C}_2}$ as a virtual machine and concatenating it with the PEMI protocol according to the formula $y''_{\mathbf{C}} = (1 - \epsilon'_0)y'_{\mathbf{C}}$, where ϵ'_0 is the average depolarising rate of $y'_{\mathbf{C}}$. We find that RMSE of the optimised formula scales as N^a and $a < 1/2$ as shown in Fig. 10.

The remaining error after virtual distillation changes from the coherent mismatch[14] to decoherence error when the gate number increases. With the error-mitigation formula $y'_{\mathbf{C}} = \mathrm{Tr}(Q\rho^2)/\mathrm{Tr}(\rho^2)$, the decoherence error is reduced from $N\epsilon$ (gate number times error rate per gate) to $(N\epsilon)^2$, while the coherent mismatch is not
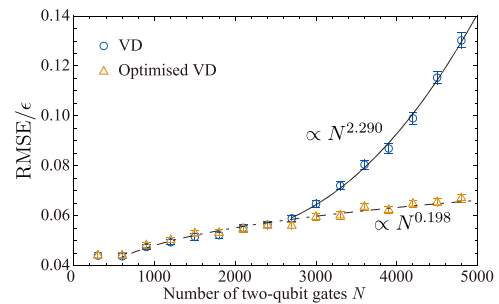

Fig. 10 Root mean square errors (RMSE) in virtual distillation (VD) protocols. The error bar represents one standard deviation. Other details such as the circuit configuration and error model are the same as Fig. 9.

suppressed, about which we give a short introduction in Supplementary Note 1.3. Because the remaining decoherence error increases quadratically with the gate number, the coherent mismatch is the dominant component in the remaining error when the gate number is small, and the decoherence error is the dominant component when the gate number is large. This change in the type of error could explain the bifurcation in Fig. 10, and the result suggests that the optimisation protocol can further reduce the remaining decoherence error but not the coherent mismatch.

In the numerical simulations, we have taken into account imperfect implementations in probabilistic error cancellation and error extrapolation. Assuming the implementation is perfect, probabilistic error cancellation can reduce RMSE to zero, and error extrapolation can reduce RMSE to a much lower level. Note that perfect implementation requires the exact knowledge of the error model or exact control of the error model. In virtual distillation, we have only taken into account errors in those gates that prepare the state ρ and neglected errors in those gates that implement virtual distillation, e.g. the controlled-swaps in Ref. [14].

DISCUSSION

In this work, we show that the residual bias in the computation result after error mitigation scales with the gate number N as $O(\epsilon' N^\gamma)$ if the error mitigation formula is optimised. Here, $\gamma \approx 0.5$, and ϵ' is a parameter depending on the error rate of quantum gates and the error mitigation formula. In contrast, the bias in the computation result before error mitigation scales linearly with N. The two scaling relations lead to a somewhat surprising result: We

can suppress the computation error by a larger factor in larger circuits.

In the analysis, we introduce a phenomenological-error model characterising errors as the global depolarisation with fluctuation, which captures the impact of realistic noise on the computation result. For the optimisation of an error mitigation formula, we propose ICS as an efficient method of generating training circuits, where only those Clifford circuits sensitive to Pauli errors are selected. The optimised formula removes the average contribution of noise and leaves the fluctuation proportional to $\sqrt{N}$. We verify this result with the numerical simulation of various circuits, error models and error mitigation formulas, from which we observe that the scaling behaviour is universal.

Despite the encouraging scaling of bias in error mitigation, we point out that the circuit size is still limited by the quality of quantum devices. On a quantum device with a finite error rate per gate, the bias increases with the circuit size. Although the bias scaling after error mitigation is advantageous in comparison with the linear error accumulation before mitigation, at certain circuit sizes the computation result becomes sufficiently random that error mitigation cannot faithfully recover the information. Therefore, the efficacy of error mitigation is conditional on the quality of the quantum device. In general, the minimum requirement for error mitigation to take effect is a non-zero fidelity between the error-free and erroneous circuits, and the performance is better with higher fidelity. Beyond this, the impact of the unmitigated error rate on the accuracy of the mitigated result depends on the mitigation method. In probabilistic error cancellation, for example, the variance in calculating the expectation value of the result increases with the error rate. Another example is that, after the virtual distillation using two copies, the bias in the expectation value scales quadratically with the error rate. Once the device can implement the circuit with sufficiently high fidelity (which is not necessarily close to one but we take a fidelity of 0.9 as an example), error mitigation can improve the computation result to a much higher accuracy (equivalent to quantum computing with fidelity of 0.99 if the error is reduced by a factor of ten).

In scalable quantum computers, we can adopt quantum error correction to increase the fidelity of logical qubits. Protocols concatenating error correction with error mitigation have been proposed recently[50–52]. Fault-tolerant devices will enable the implementation of much deeper circuits than NISQ hardware. Our result of the scaling behaviours suggests that error mitigation can perform even better in the fault-tolerant regime than in the NISQ regime.

METHODS

Circuits
We use three families of circuits: periodic-cycling circuits, linear-network circuits and all-to-all-network circuits.

Periodic-cycling circuits. The qubit array has n qubits, and n is even. All qubits are initialised in the state $|0\rangle$. After initialisation, a layer of single-qubit gates is placed, see Supplementary Figure 1(a). The circuit pattern is periodic, and each period has two layers of two-qubit gates. In the first layer, a controlled-Z gate is applied on qubit-$(2i-1)$ and qubit-$(2i)$, where $i = 1, 2, …, n/2$. In the second layer, a controlled-Z gate is applied on qubit-$(2i-1)$ and qubit-$(2i-2)$, and qubit-0 and qubit-n are the same qubits. After each two-qubit gate, a single-qubit gate is applied to each of the two qubits. The observable O is Z of the first qubit. All single-qubit gates are taken as slots in the corresponding circuit frame.

Linear-network circuits. Except for the pattern of two-qubit gates and observable, the setup is the same as periodic-cycling circuits. All two-qubit gates are controlled-Z gates. For each of them, we randomly generate an integer $i \in [1, n]$ and apply the two-qubit gate on qubit-$(i-1)$ and qubit-i, see Supplementary

Figure 1(b). The observable is $O = P_1 \otimes P_2 \otimes \cdots \otimes P_n$, where $P = I, Z$ is taken randomly.

All-to-all-network circuits. It is similar to linear-network circuits. For each of the two-qubit gates, we randomly generate two different integers $i, j \in [1, n]$ and apply the two-qubit gate on qubit-i and qubit-j, see Supplementary Figure 1(c).

Error models
Several error models are used in the numerical simulations.

Gate depolarising model. The model is given in Eq. (20), and only two-qubit gates have errors. This model is used to generated data shown in Figs. 2, 5 and 6. In Figs. 2 and 5, we take $\epsilon = 0.001$. In Figs. 6, for each data point, we randomly generate a circuit (and the corresponding circuit frame) and an error rate. For a circuit with N two-qubit gates, we generate a random real number $\eta \in [-2.5, -0.5]$, and we take $\epsilon = 10^\eta/N$ as the error rate per gate. Notice that 10^η is the total error rate.

Composite error model. Only two-qubit gates have errors. For a two-qubit gate U, the gate with errors is

$$\mathcal{A}_2\mathcal{A}_1[R_{2,Z}R_{2,Y}R_{2,X}][R_{1,Z}R_{1,Y}R_{1,X}]\mathcal{Z}_2\mathcal{Z}_1\mathcal{N}[U],$$

where $\mathcal{N}$ is the gate depolarising error in Eq. (20) with the error rate ϵ_d, $\mathcal{Z}_i = (1 - \epsilon_{i,z})[I] + \epsilon_{i,z}[Z]$ is the dephasing error on qubit-i, $R_{i,P} = e^{-i\frac{\theta_{i,P}}{2}P}$ is a single-qubit rotation on qubit-i, and

$$\mathcal{A}_i = \left[\frac{1+Z}{2} + \sqrt{1 - \epsilon_{i,a}}\frac{1-Z}{2}\right] \\ + \left[\sqrt{\epsilon_{i,a}}\frac{X+iY}{2}\right] \quad (39)$$

is the amplitude damping on qubit-i. This model is used to generate data shown in Fig. 7 (c) and (d). For each data point, we randomly generate the error model parameters as follows. For a circuit with N two-qubit gates, we generate a random real number $\eta \in [-2.5, -0.5]$, and we take $\epsilon = 10^\eta/N$ as the error rate per gate. Then, we take $\epsilon_d = (1 + 0.2\kappa_d)\epsilon/9$, $\epsilon_{i,z} = (1 + 0.2\kappa_{i,z})\epsilon/9$, $\theta_{i,P} = \kappa_{i,P}\epsilon/9$ and $\epsilon_{i,a} = (1 + 0.2\kappa_{i,a})\epsilon/6$. Each κ is taken randomly in the interval $[-1, 1]$.

Gate-dependent error model. In this model, both single-qubit and two-qubit gates have errors. The error model is the gate depolarising model. For two-qubit gates, the noise map is given by Eq. (20). For a single-qubit gate R, the gate with error is $\mathcal{S}[R]$, where

$$\mathcal{S} = \left(1 - \frac{4\epsilon_s}{3}\right)[I] + \frac{\epsilon_s}{3}\sum_{P=I,X,Y,Z}[P], \quad (40)$$

and $\epsilon_s = 0.1\pi^{-1}\epsilon\arccos\frac{|\text{Tr}(R)|}{2}$. This model is used to generate data shown in Fig. 8, and we estimate ϵ_0 and Δ using 10000 unitary circuits in $\mathbb{U}$.

Gate depolarising and dephasing model. The model is given in Eq. (36), and only two-qubit gates have errors. This model is used to generate data shown in Figs. 9 and 10. In the numerical simulation, we approximate the error model with $\mathcal{Z}_2\mathcal{Z}_1\mathcal{N}$ for simplicity in coding, which only causes a small difference and will not change the conclusion.

The above error models take into consideration kinds of physical noise processes and are able to simulate noises in realistic quantum devices. The depolarising error $\mathcal{N}$ and dephasing error $\mathcal{Z}$ simulates the relaxation process and the dephasing process[53,54], which are the main contributions to noise in realistic quantum devices. Amplitude damping $\mathcal{A}$ refers to the infidelity caused by energy dissipation. Random rotations R refer to coherent errors caused by imperfect controls. This composite model takes into consideration all the above realistic imperfections and it was demonstrated in Ref. [42] that the composite model can produce error distributions resembling that in experiments on a superconducting quantum processor. The single-qubit-gate-dependent error model $\mathcal{S}$ is the single-qubit depolarising error with an error rate depending on the gate parameters. This error

model takes into consideration the realistic situation that gate errors increase with the gate time. Additionally, we will make a direct comparison between the experimental results and simulation results in Supplementary Note 6 and show that experimental results are consistent with simulation results.

Error mitigation protocols

We verified the scaling behaviour by simulating various error mitigation protocols. The formula in Eq. (33) is used to generate data shown in Figs. 5, 6 and 7. The PEMI protocol in Fig. 9 is $y'_C = (1 - \epsilon_0)^{-1} y_C$. In optimised error extrapolation, we take $\lambda = \epsilon_2/(\epsilon_2 - \epsilon_1)$. In optimised probabilistic error cancellation, we take $\lambda = -16\epsilon_d/(15 - 16\epsilon_d) - 2\epsilon_z$. We have searched for the optimal λ using ICS data and found that the numerical optimal value is close to it. In optimised virtual distillation in Fig. 10, the formula is $y''_C = (1 - \epsilon'_0)^{-1} y'_C$. To implement optimised error mitigation formulas, we estimate ϵ_0, Δ, ϵ_1, ϵ_2 or ϵ'_0 using 1000 error-sensitive circuits, according to Algorithm 2. Then, we generate 1000 unitary circuits with the same frame to estimate RMSE.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The codes that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
2. Gong, M. et al. Quantum walks on a programmable two-dimensional 62-qubit superconducting processor. *Science* **372**, 948–952 (2021).
3. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
4. Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
5. O'Gorman, J. & Campbell, E. T. Quantum computation with realistic magic-state factories. *Phys. Rev. A* **95**, 032338 (2017).
6. Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
7. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
8. Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
9. Bonet-Monroig, X., Sagastizabal, R., Singh, M. & O'Brien, T. E. Low-cost error mitigation by symmetry verification. *Phys. Rev. A* **98**, 062339 (2018).
10. McArdle, S., Yuan, X. & Benjamin, S. Error-mitigated digital quantum simulation. *Phys. Rev. Lett.* **122**, 180501 (2019).
11. McClean, J. R., 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).
12. Colless, J. I. et al. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Phys. Rev. X* **8**, 011021 (2018).
13. Huggins, W. J. et al. Virtual distillation for quantum error mitigation. *Phys. Rev. X* **11**, 041036 (2021).
14. Koczor, B. Exponential error suppression for near-term quantum devices. *Phys. Rev. X* **11**, 031057 (2021).
15. Kwon, H. & Bae, J. A hybrid quantum-classical approach to mitigating measurement errors in quantum algorithms. *IEEE Trans. Comput.* **70**, 1401–1411 (2021).
16. Smart, S. E. & Mazziotti, D. A. Efficient two-electron ansatz for benchmarking quantum chemistry on a quantum computer. *Phys. Rev. Res.* **2**, 023048 (2020).

17. Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
18. Takagi, R., Endo, S., Minagawa, S. & Gu, M. Fundamental limits of quantum error mitigation. *npj Quantum Inf* **8**, 114 (2022).
19. Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).
20. McClean, J. R. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **18**, 023023 (2016).
21. Motta, M. et al. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nat. Phys.* **16**, 205–210 (2020).
22. Yang, Y., Lu, B.-N. & Li, Y. Accelerated quantum monte carlo with mitigated error on noisy quantum computer. *PRX Quantum* **2**, 040361 (2021).
23. Huggins, W. J. et al. Unbiasing fermionic quantum Monte Carlo with a quantum computer. *Nature* **603**, 416–420 (2022).
24. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
25. Arute, F. et al. Hartree-fock on a superconducting qubit quantum computer. *Science* **369**, 1084–1089 (2020).
26. Dumitrescu, E. F. et al. Cloud quantum computing of an atomic nucleus. *Phys. Rev. Lett.* **120**, 210501 (2018).
27. Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).
28. Song, C. et al. Quantum computation with universal error mitigation on a superconducting quantum processor. *Sci. Adv.* **5**, eaaw5686 (2019).
29. Zhang, S. et al. Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system. *Nat. Commun.* **11**, 587 (2020).
30. Kim, Y. et al. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. *Nat. Phys.* **2023**, 1–8 (2023).
31. Giurgica-Tiron, T., Hindy, Y., LaRose, R., Mari, A. & Zeng, W. J. Digital zero noise extrapolation for quantum error mitigation. In 2020 IEEE International Conference on Quantum Computing and Engineering (QCE), vol. 2020, 306–316 (IEEE, 2020).
32. 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).
33. Arrasmith, A., Czarnik, P., Cincio, L. & Coles, P. Qubit-efficient exponential suppression of errors. In APS March Meeting Abstracts, vol. 2022, S40.001 (2022).
34. O'Brien, T. E. et al. Error mitigation via verified phase estimation. *PRX Quantum* **2**, 020317 (2021).
35. Huo, M. & Li, Y. Dual-state purification for practical quantum error mitigation. *Phys. Rev. A* **105**, 022427 (2022).
36. Strikis, A., Qin, D., Chen, Y., Benjamin, S. C. & Li, Y. Learning-based quantum error mitigation. *PRX Quantum* **2**, 040330 (2021).
37. Czarnik, P., Arrasmith, A., Coles, P. J. & Cincio, L. Error mitigation with Clifford quantum-circuit data. *Quantum* **5**, 592 (2021).
38. Gordon, M. Unifying and benchmarking state-of-the-art quantum error mitigation techniques. In APS March Meeting Abstracts, vol. 2022, S40.012 (2022).
39. Aaronson, S. & Gottesman, D. Improved simulation of stabilizer circuits. *Phys. Rev. A* **70**, 052328 (2004).
40. Anders, S. & Briegel, H. J. Fast simulation of stabilizer circuits using a graph state representation. *Phys. Rev. A* **73**, 022334 (2006).
41. Magesan, E., Gambetta, J. M. & Emerson, J. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.* **106**, 180504 (2011).
42. Wang, Z. et al. Scalable evaluation of quantum-circuit error loss using clifford sampling. *Phys. Rev. Lett.* **126**, 080501 (2021).
43. Bravyi, S., Englbrecht, M., König, R. & Peard, N. Correcting coherent errors with surface codes. *npj Quantum Inf* **4**, 55 (2018).
44. Dankert, C., Cleve, R., Emerson, J. & Livine, E. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A* **80**, 012304 (2009).
45. Cai, Z. et al. Quantum Error Mitigation. Preprint at http://arxiv.org/abs/2210.00921 (2022).
46. Qin, D., Xu, X. & Li, Y. An overview of quantum error mitigation formulas. *Chinese Phys. B* **31**, 090306 (2022).
47. Vovrosh, J. et al. Simple mitigation of global depolarizing errors in quantum simulations. *Phys. Rev. E* **104**, 035309 (2021).
48. Urbanek, M. et al. Mitigating depolarizing noise on quantum computers with noise-estimation circuits. *Phys. Rev. Lett.* **127**, 270502 (2021).
49. IBM Quantum. https://quantum-computing.ibm.com/.
50. 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).
51. 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).
52. Lostaglio, M. & Ciani, A. Error mitigation and quantum-assisted simulation in the error corrected regime. *Phys. Rev. Lett.* **127**, 200506 (2021).

53. Ioffe, L. & Mézard, M. Asymmetric quantum error-correcting codes. *Phys. Rev. A* **75**, 032345 (2007).
54. Wang, P. et al. Single ion qubit with estimated coherence time exceeding one hour. *Nat Commun* **12**, 233 (2021).
55. Jones, T. & Benjamin, S. C. QuESTlink—Mathematica embiggened by a hardware-optimised quantum emulator. *Quantum Sci. Technol.* **5**, 034012 (2020).

AUTHOR CONTRIBUTIONS

D.Y.Q., Y.C. and Y.L. together conceived the ideas. Y.L. developed the theory. D.Y.Q. and Y.L. performed the numerical simulation. D.Y.Q. and Y.C. implement the experiment. D.Y.Q., Y.C. and Y.L. prepared the manuscript.

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/s41534-023-00707-7.

Correspondence and requests for materials should be addressed to Ying Li.

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.