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 $O(cN)$ with the gate number $N$ before mitigation and sublinearly $O(c'N^\gamma)$ after mitigation, where $\gamma \approx 0.5$, $c$ is the error rate of a quantum gate, and $c'$ 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 technique for error mitigation in large circuits to obtain this result.

Error statistics and scalability of quantum error mitigation formulas

Dayue Qin$^1$, Yanzhu Chen$^2$ and Ying Li$^1$

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 $O(cN)$ with the gate number $N$ before mitigation and sublinearly $O(c'N^\gamma)$ after mitigation, where $\gamma \approx 0.5$, $c$ is the error rate of a quantum gate, and $c'$ 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 technique for error mitigation in large circuits to obtain this result.
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 detailed overview to Supplementary Note 1.

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\(^7\),\(^3\) the error mitigation formula is

$$y'_C = \sum q_i y_{C_i},$$  

(2)

where \(C_i\) 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

$$y'_C = 2y_{C_1} - y_{C_2}.$$  

(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.

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

(4)

where \(q_i\) are quasi-probabilities, and \(\mathcal{E}_i\) is the map of a noisy circuit \(C_i\). Here \(C_i\) is generated by, for example, replacing or adding some gates in the primitive circuit \(C\). We can work out the quasi-probability decomposition with gate set tomography data\(^8\) or in a learning manner\(^3\). 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 \(\rho\) are used to evaluate the observable in a distilled state without physically
preparing it. Given the primitive circuit $\mathbf{C}$ that prepares the state $\rho$, the circuit $\mathbf{C}_i$ is to evaluate $\gamma_{\mathbf{C}_i} = \text{Tr}(\mathbf{Q}_{\mathbf{C}_i})$, and the circuit $\mathbf{C}_o$ is to evaluate $\gamma_{\mathbf{C}_o} = \text{Tr}(\rho^o)$. Then the error mitigation formula reads

$$\gamma_{\mathbf{C}_o} = \frac{\gamma_{\mathbf{C}_o}}{\gamma_{\mathbf{C}_i}}.$$  

(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 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 $\text{O}(N\epsilon(C))$ to $\text{O}(N\epsilon^2(C))$, where $N$ is the gate number and $\epsilon$ 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 $\mathcal{G}_j = (1-\epsilon)\mathcal{U}_j + \epsilon\mathcal{D}$ acting on the whole input state, where $\epsilon$ is the gate depolarising rate, $\mathcal{D}(\bullet) = \text{Tr}(\bullet)\rho_m$ is the depolarising map, and $\rho_m = 1/2^{D}$ is the maximally mixed state. Without loss of generality, we assume that the observable is a traceless operator, and we have $\gamma_j = (1-\epsilon)^N\gamma_{\mathbf{C}_o} = f_c + \text{O}(\epsilon N)$. The bias increases linearly with the gate number when $N$ is significantly smaller than $1/\epsilon$. 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 $\mathcal{G}'_j = (1 - 2\epsilon)\mathcal{U}_j + 2\epsilon\mathcal{D}$ be the gate with the doubled depolarising rate, two circuits labelled by $i = 1, 2$ produce expected values $\gamma_{\mathbf{C}_i} = \text{Tr}(\mathbf{Q}_{\mathbf{C}_i}(0|0)^{\otimes n})$, where $\mathcal{E}_1 = \mathcal{G}_N \cdots \mathcal{G}_i \mathcal{G}_i$ and $\mathcal{E}_2 = \mathcal{G}_N \cdots \mathcal{G}_i \mathcal{G}_i \mathcal{D}$. Then, Eq. (3) leads to the error-mitigated expected value

$$\gamma'_{\mathbf{C}_i} = 2(1-\epsilon)^N f_c - (1-2\epsilon)^N f_c = f_c + \text{O}(\epsilon^2 N^2).$$  

(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

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

(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

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

(8)

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

$$\gamma_{\mathbf{C}_o} = \frac{1}{1-\epsilon} \gamma_{\mathbf{C}_i} - \frac{\epsilon}{1-\epsilon} \gamma_{\mathbf{C}_o} + \cdots = f_c.$$  

(9)

Here, we have used that $\gamma_{\mathbf{C}_o} = 0$ 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

$$\rho = (1-\epsilon_i)\mathcal{U}_j(0|0)^{\otimes n}U_j + \epsilon_i \rho_m.$$  

(10)

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

$$\gamma_{\mathbf{C}_o} = \frac{(1-\epsilon)^N(1-\epsilon^2)^N}{(1-\epsilon^2)^N} f_c = f_c + \text{O}(\epsilon^2 N^2).$$  

(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 $\epsilon'$ instead of its actual value $\epsilon$ and we work out the error mitigation formula with $\epsilon'$, we have $\gamma_{\mathbf{C}_o}' = (1-\epsilon')^N/(1-\epsilon')^N f_c$. Then, the bias of the error mitigation formula is $\text{O}(\epsilon - \epsilon')$, 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

$$\gamma_{\mathbf{C}_o} = \lambda \gamma_{\mathbf{C}_i} + (1-\lambda)\gamma_{\mathbf{C}_o}.$$  

(12)

We determine $\lambda$ 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 $\mathcal{T}$ and the corresponding data $(\gamma_{\mathcal{T}_i}, \gamma_{\mathcal{T}_o}, f_{\mathcal{T}})$ to determine $\lambda$ for the ansatz considered here. The bias of the training circuit is minimised at

$$\lambda^* = \frac{f_{\mathcal{T}} - \gamma_{\mathcal{T}_o}}{\gamma_{\mathcal{T}_i} - \gamma_{\mathcal{T}_o}}.$$  

(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 $\epsilon$ in Eq. (9) as the variational parameter, assuming the actual depolarising rate is unknown. We can find the optimal value of $\epsilon$ 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 $\gamma_{\mathbf{C}_o}' = \lambda \gamma_{\mathbf{C}_o}'$ by $\lambda$. According to Eq. (11), the bias is zero when $\lambda$ cancels the factor $\epsilon/c$. Before $f_c$.

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 $f_e$ of a Clifford circuit is computable\cite{34,35}, we can simulate 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. $\gamma_T = 0$. 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.

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

$$L_R = \langle (\gamma e - f_c)^2 \rangle_R,$$

where $(\gamma(C))_R \equiv \sum_{\gamma(C,R)} \gamma(C)$ is the average of the real-valued function $\gamma(C)$ over the circuit set $R$. 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. $\langle (\gamma e - f_c)^2 \rangle$ 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\cite{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 $\pm 1$. We can reexpress the error-free expected value as $f_e = \text{Tr}(Q_0 | 0 \rangle \langle 0 | )$, where $Q_0 = U^0 QU$ is the effective observable. When $U$ is Clifford, $Q_0$ is a Pauli operator. Let $P_i = I, X, Y, Z$ be the single-qubit Pauli operator on qubit-$i$, $Q_0 = \pm P_x \otimes P_x \otimes \cdots \otimes P_x$. Then $f_e = \pm \prod_{i=1}^n \langle 0 | P_i | 0 \rangle$. If any single-qubit Pauli operator $P_i$ is $X$ or $Y$, the expected value is zero. If all $P_i$ are $I$ or $Z$, $f_e = \pm 1$, and the sign is the same as $Q_0$. For a randomly generated Clifford circuit, it is likely that some single-qubit Pauli operators contained in $Q_0$ are $X$ or $Y$, i.e. $f_e = 0$.

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 $U = U_0 U_0$, where $U_0 = R_t \otimes R_t \otimes \cdots \otimes R_t$ is one layer of single-qubit gates, and $R_t$ is the gate on qubit-$i$. First, given the gate number, we generate a random Clifford circuit, which realises the unitary $U'$. If $U_0 = 1$, the effective observable is $Q_0 = \pm P_x \otimes P_x \otimes \cdots \otimes P_x$. Given $Q$ and $U'$, we can efficiently work out this expression of $Q_0$ on a classical computer. Second, we determine single-qubit gates in $U_0$; we take a Clifford $R_t$ satisfying $R_t | R_t \otimes R_t | = \pm 2 | Z |$. For the final circuit $U = U' U_0$, single-qubit Pauli operators in its effective observable $Q_0$ are either $I$ or $Z$. Then, the expected value is $f_e = \pm 1$.

![Fig. 2 Distribution of the effective depolarising rate in the phenomenological-error model.](image-url)

In the model, the impact of errors in a noisy circuit is characterised by the global depolarising model with the circuit-dependent depolarising rate $\gamma$. 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 $f_e^2$. The average depolarising rate is proportional to the gate number $N$, and the standard deviation is proportional to $\sqrt{N}$. This is the phenomenon of fluctuation in realistic error models, see Fig. 2.
treated as one multi-qubit Clifford gate. Clifford gates not interrupted by any non-Clifford gate can be measurement, respectively. The reason is that a sequence of number to minimise the circuit set. We only take locations of each multi-qubit gate is proposed. Here we reduce the slot transformation stochastically occurring in the circuit. In Pauli 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 much lower than multi-qubit gates. Errors occurring in a circuit are shown in Fig. 4, the frame includes the qubit initialisation, multi-qubit gates, and the observable is equivalent to an effective Pauli observable. As shown in (d), gates in slots after the initialisation and before the measurement, respectively. The corresponding frame is \( F = \{ \ldots, U_i, \ldots, U_q, Q \} \), where \( U_i \) is a gate on the frame, and \( \chi_i \) denotes a slot on qubit-\( i \). In other words, \( F \) is the same as \( C \) except that gates in slots are replaced with \( \chi_i \). Formally, if \( S = \{i_1, i_2, \ldots\} \) are labels of slots and \( K = \{k_1, k_2, \ldots\} \) are corresponding qubits, the frame is \( F = \{ F_1, \ldots, F_N, Q \} \), where \( F_i = U_i \) for \( i \notin S \), and \( F_i = \chi_i \) if \( i \in S \). Then, we can reexpress the circuit as \( C = [F, R_1, R_2, \ldots] \), where \( R_i \) is the single-qubit gate in the \( i \)-th slot, i.e. \( U_i = \Pi_{\chi_i} \). 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 the previous section. We give details of this procedure in Algorithm 1.

**Algorithm 1. Generation of error-sensitive circuits.**

1. **function** 
2. 3. **end**

**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
single-qubit Clifford gates, which contributes a factor of 8. The number of non-identity Pauli operators in the product, i.e. the sample number.

When errors are independent of the choice of observable, \( n \) is the qubit number, \( N_f \) is the slot number, and \( N_T \) is the sample number. \( C_i \) is the single-qubit Clifford group with 24 elements. \( U = U_f \cdots U_1 \) is the unitary transformation of the circuit \( C = (U_f, \ldots, U_1, Q) = \{ F, R_1, R_2, \ldots \} \). We use \( \tilde{R} = (R_{n+1}, R_{n+2}, \ldots, R_N) \) 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. \( w(C) \) is the weight of the Clifford circuit \( C \). \( QU = U QU = \pm P_I \otimes P_2 \otimes \cdots \otimes P_n \) is a tensor product of Pauli operators, then \( w(C) \) is the number of non-identity Pauli operators in the product, i.e.

\[
w(C) = n - \sum_{i=1}^{n} \delta_{iR},
\]

where \( \delta_{iR} = 1 \) if \( P_i = I \), and \( \delta_{iR} = 0 \) otherwise. In Algorithm 3, we employ the Metropolis-Hasting algorithm to realise a uniform distribution of error-sensitive circuits, which requires a conditional distribution \( g(R|\tilde{R}) \) 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 \( N_f \) do
3. for \( i = n + 1 \) to \( N_T \) do
4. Choose a random \( R \) from \( C_i \).
5. Call ESCIRCUIT \( F, R \) to generate \( C \).
6. Output \( C_\tilde{R} = C \).

**Algorithm 3. Uniform importance Clifford sampling.**
1. Input \( F \), a conditional distribution \( g(R|\tilde{R}) \) and an initial slot-gate pattern \( \tilde{R}^{(0)} \).
2. Set \( t = 0 \).
3. Call ESCIRCUIT \( F, \tilde{R}^{(0)} \) to generate \( C \).
4. Take \( C_0 = C \).
5. for \( t = 1 \) to \( N_f \) do
6. Generate a random candidate of slot-gate pattern \( \tilde{R}^{(t)} \) according to \( g(R|\tilde{R}^{(t-1)}) \).
7. Call ESCIRCUIT \( F, \tilde{R}^{(t)} \) to generate \( C \).
8. Calculate the acceptance probability
\[
A = \min \left(1, \frac{3^{-w(C)}}{3^{-w(C)}} \frac{g(R|\tilde{R}^{(t-1)})}{g(R|\tilde{R}^{(t)})} \right).
\]
9. Generate a uniform random number \( u \in [0, 1] \).
10. Accept and set \( C_\tilde{R} = C \) if \( u < A \).
11. Reject and set \( C_\tilde{R} = C_{t-1} \) if \( u > A \).
12. Output \( C_\tilde{R} \).

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 \( C \) to denote the set of Clifford circuits and \( U \) to denote the set of all unitary circuits with the same frame. For a frame with \( N_f \) slots, the total number of Clifford circuits is \( |C| = 24^{N_f} \). The cost of Algorithm 1 is \( O(N_f^2) \). Algorithm 2 and Algorithm 3 are used to sample error-sensitive circuits according to the non-uniform distribution \( P_{nu}(C) \) and uniform distribution \( P_u(C) \), respectively.

There is a similar relation between ICS and unitary sampling. Error-sensitive circuits are a subset of all Clifford circuits, denoted by \( C^{ES} \). According to Algorithm 1, given slot gates \( \tilde{R} = (R_{n+1}, R_{n+2}, \ldots, R_N) \), the number of error-sensitive circuits is 
\[
8^{\tilde{w}(C)} 24^{n-\tilde{w}(C)},
\]
where \( \tilde{w}(C) \) is the weight of the Clifford circuit \( C \). If \( P_i = I \), \( R_i P_i R_i = I \) for all 24 single-qubit Clifford gates, which contributes a factor of 24; If \( P_i \neq I \), \( R_i P_i R_i = \pm Z \) for 8 single-qubit Clifford gates, which contributes a factor of 8. The number of different \( \tilde{R} \)s is \( 24^{N_f-n} \), then the total number of error-sensitive circuits is
\[
|C^{ES}| = \sum_{j=1}^{24^{N_f-n}} 8^{\tilde{w}(C)} 24^{n-\tilde{w}(C)}.
\]

where \( C_i \) are circuits with different \( \tilde{R} \)'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. \( \gamma = f_c \) if \( f_c = 0 \). Therefore,
\[
L_U = L_C = nL_C,\tag{17}
\]
for Pauli error models, where \( \eta \equiv |C^{ES}| / |C| \) is the proportion of error-sensitive circuits in all Clifford circuits.

The distribution of error-sensitive circuits from Algorithm 2 is non-uniform. We uniformly choose slot gates in \( \tilde{R} \), the probability of an error-sensitive circuit \( C \) is
\[
P_{nu}(C) = 24^{-((N_f-n)\gamma)} 24^{\gamma-\gamma} 3^{w(C)} \tag{18}
\]
Therefore, the probability of \( C \) is proportional to \( 3^{w(C)} \). If we use Algorithm 2 to sample circuits, we can evaluate \( L_C \) according to
\[
L_C = n^{-1} E(3^{w(C)} (y_C - f_c)^2) \text{nu}, \tag{19}
\]
where the expected value is taken over the distribution \( P_{nu}(C) \).

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 \( P_u(C) = |C^{ES}| / |C| \). Then, we can evaluate \( L_C \) with \( L_C = E((y_C - f_c)^2) \text{nu} \), where the expected value is taken over the distribution \( P_u(C) \). 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\(^{19}\), 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 \( P_{nu}(C) \) and uniform distribution \( P_u(C) \), respectively.

To generate \( N_f \) circuits, the costs for both algorithms are \( O(N_f nN) \), because the elementary building block of both algorithms is nothing but the circuit generation given in Algorithm 1, which is repeated for \( N_f \) times. The numerical result in Supplementary Note 3 demonstrates that the number of error-sensitive circuits \( N_f \) 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
Pauli twirling to convert them into Pauli errors. If error mitigation is concatenated with error correction, logical errors after correction are mainly Pauli errors. Suppose errors are independent of single-qubit gates, we have the following relations,

\[ (f^2_x)_U = (f^2_x)_C = \eta (f^2_x)_{C_E}, \]
\[ (f^2_y)_C = (f^2_y)_C = \eta (f^2_y)_{C_E}, \]
\[ (f^2_z)_U = (f^2_z)_C = \eta (f^2_z)_{C_E}, \]

where \( U, C \) and \( C_E \) 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, and therefore \( (f^2_x)_U = (f^2_x)_C \) holds if \( f \) is a polynomial of degree two in the gate unitaries. The second equal sign is a consequence of \( f_C = 0 \) when \( C \neq C_E \) and \( \eta = |\langle C | C_E \rangle | \). Using \( f_C = \pm 1 \) for error-sensitive circuits, we can obtain

\[ \eta = (f^2_x)_U, \]
\[ e_0 = (f^2_y)_{C_E}, \]
\[ \Delta = \sqrt{\Delta_2 (f^2_z)_{C_E}}. \]

These relations allow us to study \( e_0 \) and \( \Delta \) 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 \( \sigma_k \) occurs after the \( i \)-th gate, the transformation becomes \( U' = U_N \cdots U_{i+1} u_i U_{i} \cdots U_1 = \sigma_k U \), where \( \sigma_k = U_N \cdots U_{i+1} U_{i+1}^* \cdots U_1^* \) is the Pauli error propagated to the end of the circuit. Because gates are Clifford, \( \sigma_k \) 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 \( (1 - p) [1] + p [\sigma_k], \) the final state of the circuit is transformed from \( \rho_0 \) to \( (1 - p) \rho_0 + p [\sigma_k^*] \rho_0. \) Then there are two cases: If \( \sigma_k^* \) and the Pauli observable \( Q \) commute, the expected value is preserved under the Pauli error; otherwise, the expected value is changed from \( f_C \) to \( (1 - 2p) f_C, \) 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 \( (1 - p) [1] + p [\sigma_k] \) be the \( k \)-th error channel, \( (1 - p) [1] + p [\sigma_k^t] \) is the corresponding error channel at the end of the circuit. We use the binary number \( t(C) \) to denote whether the \( k \)-th error channel affects the observable, i.e. \( t_k(C) = 0 \) if \( \sigma_k^t \) and \( Q \) are commutative, and \( t_k(C) = 1 \) otherwise. Then, the expected value is changed to \( \prod_{k=1}^M (1 - 2p)^{t_k(C)} f_C. \) The equivalent depolarising rate is

\[ e_C = 1 - \prod_{k=1}^M (1 - 2p)^{t_k(C)} = \sum_{p=0}^M 2t_k(C) p + O(p^2). \]

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 \( \epsilon_C \) is binomial. Let \( P \) be the probability of \( t_0 = 1 \) and neglect \( \mathcal{O}(P^2) \) terms, the average depolarising rate is \( \epsilon_0 \approx 2PM \), and the standard deviation is \( \Delta \approx \sqrt{2PM(1 - P)} \). Note that \( M \) is proportional to the gate number.

In large circuits, the global depolarising model with the depolarising rate \( \epsilon_0 \) 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.

\[
\frac{\Delta}{\epsilon_0} \approx \frac{1}{\sqrt{M}} \quad (32)
\]

where \( M \gg N' \sim N \). Therefore, \( \epsilon_C \) is in the vicinity of \( \epsilon_0 \) 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 \( \Delta \) to the depolarising rate \( \epsilon_0 \) 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. 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\(^{74} \).

### Error mitigation according to the phenomenological-error model

According to the phenomenological-error model, the effective depolarising rate in large circuits is \( \epsilon_0 \) with a small fluctuation. We can mitigate errors by compensating the effect of \( \epsilon_0 \). 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 \( \sqrt{\langle (y_C - f^2) \rangle} = \sqrt{\langle \epsilon_0^2 \rangle} + \sqrt{\langle \Delta^2 \epsilon_0 \rangle} \), which increases linearly with the gate number. Using the error mitigation formula \( y_C' = (1 - \epsilon_0) y_C + \epsilon_0 \), we can reduce RMSE to \( \sqrt{\langle (y_C' - f^2) \rangle} = \sqrt{\langle 1 - \epsilon_0 \rangle^2 y_C^2 + \frac{\epsilon_0^2}{(1 - \epsilon_0)^2} \Delta^2 \epsilon_0} \), which increases sublinearly with the gate number. Because \( \epsilon_0 = 1 - y_C f^2 \), we can measure \( \epsilon_0 \) (and \( \Delta \)) by uniformly sampling error-sensitive circuits. Actually, because the fluctuation is small, we can even take \( \epsilon_0 = 1 - y_C f^2 \) for one randomly generated error-sensitive circuit \( C \in C^P \), and it is likely that the error mitigation formula still works. This phenomenological-error-model inspired (PEMI) error mitigation protocol is illustrated in Fig. 5.

![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 \( \epsilon_0 \), 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).](image-url)

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

\[
\epsilon_0' = \frac{1 - \epsilon_0}{(1 - \epsilon_0)^2 + \Delta^2} \epsilon_0.
\]

RMSE after mitigation is reduced to

\[
\sqrt{\langle (y_C' - f^2) \rangle} = \frac{\sqrt{\Delta}}{\sqrt{1 - \epsilon_0^2 + \Delta^2}}.
\]

#### Numerical results of the scaling behaviour

In this section, we numerically test the PEMI error mitigation formula and verify the scaling behaviour of \( \epsilon_0 \) and \( \Delta \). 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 errors.
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 $\sqrt{E-\langle(y_e-f_e)^2\rangle_{\mathcal{R}}}$ to $\sqrt{E-\langle(y_e-f_e)^2\rangle_{\mathcal{R}}}$, where $E$ is in the interval about $0.003$ to $0.3$. Taking the phenomenological-error model as an example, the RMSE for all-to-all networks is $\propto N$ and $\Delta \approx \sqrt{N}$. Therefore, RMSE is reduced in error mitigation by a factor of $\Delta/\epsilon_0 \approx 1/\sqrt{N}$. We verify these scaling behaviours by applying the error mitigation formula in Eq. (35) to randomly generated circuits with up to ten qubits and more than a thousand two-qubit gates. To implement the formula, $\epsilon_0$ and $\Delta$ are measured by sampling error-sensitive circuits. RMSEs before and after error mitigation $\sqrt{E}$ and $\sqrt{E'}$ are calculated and plotted in Figs. 6 and 7. For the model with single-qubit-gate-dependent errors, we directly calculate and plot $\epsilon_0$ and $\Delta$ 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 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 $\sqrt{E/E'}$ for all-to-all-network circuits meets $\sqrt{E/E'} = a\sqrt{N}$ 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 $\sqrt{E/E'}$ 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 and is considered as a general requirement of NISQ computation. For each data point in Figs. 6 and 7, we randomly choose the error rate per gate $\epsilon$ 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 $\propto 1/\sqrt{N}$. 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 extrapolation formula as an example, the two circuits $C_1$ and $C_2$ are the same as the primitive circuit $C$, but the noise level is doubled in $C_2$. In the phenomenological-error model of the circuit $C$, the average depolarising rate is $\epsilon_r$, the rate fluctuation is $\Delta \epsilon_r$, and the standard deviation is $\Delta$. Because $C_1$ and $C_2$ are the same circuit, their fluctuations are correlated: Suppose effective depolarising rates are approximately proportional to the noise level, we have $\epsilon_2 \approx 2\epsilon_1$ and $\Delta \epsilon_2 \approx 2\Delta \epsilon_1$. Therefore, the fluctuation-caused bias depends on the covariance matrix $K_{ij} \equiv n^{-1}\left(\Delta \epsilon_1 \Delta \epsilon_2 \langle f_{C_1} f_{C_2} \rangle_{\mathcal{U}} / \mathcal{U}\right)$.

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

$$
\sqrt{\langle (y_e-f_e)^2 \rangle_{\mathcal{U}}} = \sqrt{\eta \langle (E' + \Lambda - 1)^2 \rangle + \Lambda' K' A},
$$

where $E = (1 - \epsilon_1, 1 - \epsilon_2)^T$ and $\Lambda = (\lambda, 1 - \lambda)^T$. Taking $\lambda = \epsilon_2 / (\epsilon_2 - \epsilon_1)$, we can remove the contribution of average depolarising rates, and RMSE becomes $\sqrt{\langle (y_e-f_e)^2 \rangle_{\mathcal{R}}} = \sqrt{\eta \langle (\Delta \epsilon_1 + \Delta \epsilon_2)^2 \rangle / \mathcal{U}}$. Here, we have used that $K$ is positive semi-definite, $\Delta^2_1$ and $\Delta^2_2$ are diagonal elements of $K$, and $\|\Lambda\| \approx \sqrt{5}$ 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 $\lambda$. 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
Theorem 1. Consider the general extrapolation formula in Eq. (2), let $e_\Delta \delta_{e,C_i}$ and $\Delta$ be the average depolarising rate, rate fluctuation and standard deviation of the circuit $C_i$, respectively, then

$$\min_{(\delta_e)} \sqrt{(\langle e - f_e \rangle)^2} \leq \frac{\sqrt{\eta E^2 KE}}{\|E\|} \leq \frac{\sqrt{\eta \sum \Delta^2}}{\|E\|}.$$ 

where $E = (1 - \epsilon_1, 1 - \epsilon_2, \ldots)^T$, $K_i = \eta^{-1} \langle \delta_e C, \delta e C_i \rangle_U^2$ and $\eta = (f_e^2)_U$.

The proof is straightforward. Let $\Lambda = (q_1, q_2, \ldots)^T$, 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 $\mathcal{N}_i[U_i]$. We can cancel the noise by applying an inverse noise $\mathcal{N}_i^{-1} \sum q_i E_i$ after the noisy gate, and the overall effective gate is $\mathcal{N}_i^{-1} \mathcal{N}_i[U_i]$. Here, $E_i$ are some noisy gates, i.e. we insert the gate $E_i$ after the gate $\mathcal{N}_i[U_i]$ with the quasi-probability $q_i$.

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

$$\mathcal{N}_i = (1 - \delta_{15} - \epsilon_2 |1\rangle \langle 1|) + \frac{16\delta_{12}}{15} (|+\rangle \langle +| + |Z_1| + |Z_2|),$$

where $Z_i = |i^{n-1} 0 \rangle \otimes Z \otimes |0^{n-i-2}\rangle$. Suppose our knowledge about the noise map is inaccurate and we correct the error according to
the residual bias of the gate depolarising rate and expected value in the virtual machine. Then the bias is determined by taking proper variational parameters in the formula, removing the zeroth-order term (contribution of average depolarising rates) by a factor. In the numerical simulation, we determine the factor by taking the original virtual distillation formula to suppress the error scaling. For example, we can take λ = 16εd/(15 − 16εd) before the optimisation and the optimal value after the optimisation.

When λ = −16εd/(15 − 16εd) and εd = 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 $\mathcal{N} \cdot \mathcal{N}$. Then, we can describe the error in this virtual machine using the phenomenological-error model and reduce the bias using the PEC protocol. We can use the formula

$$\gamma'_{D} = (1 - \epsilon_{0}' \delta_{0}) \gamma_{D},$$

where $\epsilon_{0}'$ and $\gamma_{0}'$ are respectively the average depolarising rates and expected value in the virtual machine. Then the residual bias of $\gamma_{D}'$ is determined by the standard deviation $\Delta'$ of the virtual machine. Actually, it is not necessary to modify the formula to suppress the error scaling. For example, we can take $\lambda$ 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 $\lambda$ scales as $\sim \sqrt{N}$.

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

$$\gamma'_{D} \approx F(a_{1} f_{c}, a_{2} f_{c}, \cdots) + \sum_{i} \frac{\partial F}{\partial f_{i}} \delta_{0} f_{c},$$

(38)

where $a_{i} = 1 - \epsilon_{i}$. In Eq. (38), we have considered the general error mitigation formula in Eq. (1) and $\gamma_{c} = (1 - \epsilon_{i} + \delta_{0} f_{c}) f_{c}$. 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, $F(a_{1} f_{c}, a_{2} f_{c}) = a_{1} f_{c} / (a_{2} f_{c})$, 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 $\gamma'_{D} = \gamma_{D} f_{c}$ as a virtual machine and concatenating it with the PEC protocol according to the formula $\gamma'_{D} = (1 - \epsilon_{0}') \gamma_{D}$, where $\epsilon_{0}'$ is the average depolarising rate of $\gamma_{D}'$. We find that RMSE of the optimised formula scales as $N^{\alpha}$ and $\alpha < 1/2$ as shown in Fig. 10.

The remaining error after virtual distillation changes from the coherent mismatch to decoherence error when the gate number increases. With the error-mitigation formula $\gamma'_{D} = \text{Tr}(Q \rho^{2}) / \text{Tr}(\rho^{2})$, 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 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 $|\rho\rangle$ 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, $\gamma = 0.5$, and $\epsilon'$ 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-(2\( i \) − 1) and qubit-(2\( i \)), where \( i = 1, 2, \ldots, n/2 \). In the second layer, a controlled-Z gate is applied on qubit-(2\( i \) − 1) and qubit-(2\( i \) − 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 \( \epsilon = 0.99 \). 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 \( 10^\eta \) 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

\[
A_\eta U = A_1 U_A R \left[ R_Z \right] U_A R \left[ R_X \right] Z Z N' U,
\]

where \( N' \) is the gate depolarising error in Eq. (20) with the error rate \( \epsilon_{dp, Z} = \left( 1 - \epsilon_{Z, Z} \right) [I] + \epsilon_{Z, Z} [Z] \) is the dephasing error on qubit-\( i \), \( R_{Z} = e^{\frac{\eta Z_\eta}{\epsilon}} \) is a single-qubit rotation on qubit-\( i \), and

\[
A_i = \left[ 1 - \frac{\epsilon}{2} + \sqrt{1 - \frac{\epsilon}{4}} \right] + \left[ \sqrt{\epsilon} + \frac{\epsilon}{2} \right]
\]

(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 \( \epsilon_{Z, Z} = \left( 1 + 0.2 \kappa_{\eta} \right) / 9 \), \( \epsilon_{Z, Z} = \left( 1 + 0.2 \kappa_{\eta} \right) / 9 \), \( \theta_{\eta} = \kappa_{\eta} / 9 \), and \( \epsilon_{Z, Z} = \left( 1 + 0.2 \kappa_{\eta} \right) / 6 \). Each \( \kappa \) 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 \( S = |R| \), where

\[
S = \left( 1 - \frac{4 \epsilon_{s}}{3} \right) |I| + \frac{\epsilon_{s}}{3} \sum_{p=|X, Y, Z|} |P|,
\]

(40)

and \( \epsilon_{s} = 0.1 \eta^{-1} \epsilon \arccos \left( \frac{\eta(\eta)}{2} \right) \). This model is used to generate data shown in Fig. 8, and we estimate \( \epsilon_{s} \) and \( \Delta \) using 10000 unitary circuits in \( U \).

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 \( Z Z N' \) 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 noise in realistic quantum devices. The depolarising error \( N' \) and dephasing error \( Z \) simulates the relaxation process and the dephasing process\(^{53,54}\), which are the main contributions to noise in realistic quantum devices. Amplitude damping \( A \) 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 \( S \) is the single-qubit depolarising error with an error rate depending on the gate parameters. This error
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.

Received: 10 February 2022; Accepted: 31 March 2023;
Published online: 11 April 2023

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., Mari, 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., Sabag, Z., Singh, M. & O’Brien, T. E. Low-cost error mitigation by symmetry verification. Phys. Rev. A 98, 062339 (2018).
10. McArthurd, 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-robust 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. & Mazzotti, D. A. Efficient two-electron ansatz for benchmarking quantum chemistry on a quantum computer. Phys. Rev. Res. 2, 020048 (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, eaaw5886 (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., Bachor, 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, S4001 (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, S40012 (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., Englebrecht, 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., Fuji, K. & Tokunaga, Y. Quantum error mitigation as a universal error reduction technique: Applications from the nisq to the fault-tolerant quantum computing era. 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. & Canti, A. Error mitigation and quantum-assisted simulation in the error corrected regime. Phys. Rev. Lett. 127, 200506 (2021).
ACKNOWLEDGEMENTS

The authors thank Hang Ren for the discussions. We acknowledge the use of simulation toolkit QuESTlink\(^{55}\) for this work. We acknowledge the use of IBM Quantum services for this work. D.Y.Q. and Y.L. are supported by the National Natural Science Foundation of China (Grant Nos. 11875050 and 12088101) and NSAF (Grant No. U1930403). Y.C. acknowledges support from US Department of Energy (Award No. DE-SC0019318).

Note.—When preparing the manuscript, we notice a recent preprint arXiv:2111.14907 that reports the global depolarising model as an effective model of noisy quantum circuits. This work studies the distribution of measurement outcomes in circuits with single-qubit noise channels. In comparison, our work studies expected-value computing using circuits with two-qubit noise channels as the dominant error source. We focus on properties of circuits with the same circuit frame, and we use the effective model in error mitigation. Our final result is on the bias scaling of error mitigation formulas.

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023