

Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors

Received: 8 September 2022

Accepted: 30 March 2023

Published online: 08 May 2023

 Check for updates

Ewout van den Berg , Zlatko K. Minev, Abhinav Kandala  & Kristan Temme  

Noise in quantum computers can result in biased estimates of physical observables. Accurate bias-free estimates can be obtained using probabilistic error cancellation, an error-mitigation technique that effectively inverts well-characterized noise channels. Learning correlated noise channels in large quantum circuits, however, has been a major challenge and has severely hampered experimental realizations. Our work presents a practical protocol for learning and inverting a sparse noise model that is able to capture correlated noise and scales to large quantum devices. These advances allow us to demonstrate probabilistic error cancellation on a superconducting quantum processor, thereby providing a way to measure noise-free observables at larger circuit volumes.

As a result of continuous improvement in quantum hardware and control systems, quantum processors are now able to provide more qubits with longer coherence times and better gate fidelities^{1–3}. Despite these improvements, the levels of noise in current quantum processors still limit the depth of quantum circuits and reduce the accuracy of measured observables. Nevertheless, there is a growing number of quantum applications that run on noisy quantum processors and still provide competitive results^{4–8}. Fault tolerance using quantum error correction or similar techniques would solve many noise-related issues, but, until this is achieved, quantum error mitigation^{9–12} may very well be the best way forward. Unlike error correction, which ensures that quantum circuits can be executed faithfully, error mitigation only aims to produce accurate expectation values $\langle A \rangle$ of observables A .

One of the earliest and most general protocols for error mitigation is probabilistic error cancellation (PEC)⁹. To implement the error-mitigated action $\mathcal{U}(\rho) = U\rho U^\dagger$ of an ideal gate U applied to a quantum state ρ on a device where only noisy operations $\mathcal{U} \circ \Lambda$ are available, the protocol first requires an accurate noise model Λ . The action of the ideal gate would then be obtained by applying the mathematical inverse Λ^{-1} before the noisy gate. Although Λ^{-1} is not a physical operation, it can be expressed as a linear combination of gates and state-preparation operations^{9,12}. The PEC protocol implements this linear combination on average by promoting it to a quasi-probability

distribution. Sampling the distribution generates N physical circuit instances and results in an expectation value $\langle \hat{A}_N \rangle$ that is unbiased and completely removes the effect of Λ . However, this comes at the expense of an increased sampling overhead that we denote by γ , which captures the noise strength and the resulting increase of the standard deviation.

Despite the method's theoretical appeal^{12–21}, practical challenges have limited its demonstration to the one- and two-qubit level^{22,23}. The main difficulty has been the accurate representation of the noise in a full device, which is particularly complicated by cross-talk errors that occur during the parallel application of gates. This has led to protocols where a quasi-probability distribution for mitigation is determined by minimizing the deviation of a set of measured and exact expectation values²⁴. Fully scalable implementations of PEC require a noise model Λ that accurately captures correlated errors across all qubits, has a compact representation that can be learned efficiently, and has an inverse representation that enables tractable sampling from the associated quasi-probability distribution.

We address these challenges in the context of quantum circuits that consist of l layers of noisy two-qubit gates interleaved with layers of single-qubit gates. Each layer $i = 1, \dots, l$ consists of a noisy operator $\tilde{\mathcal{U}}_i$ and is error mitigated by Λ_i^{-1} , as shown in Fig. 1a. The noise channel Λ_i is specific to the gates in layer i and is assumed to be a Pauli channel.

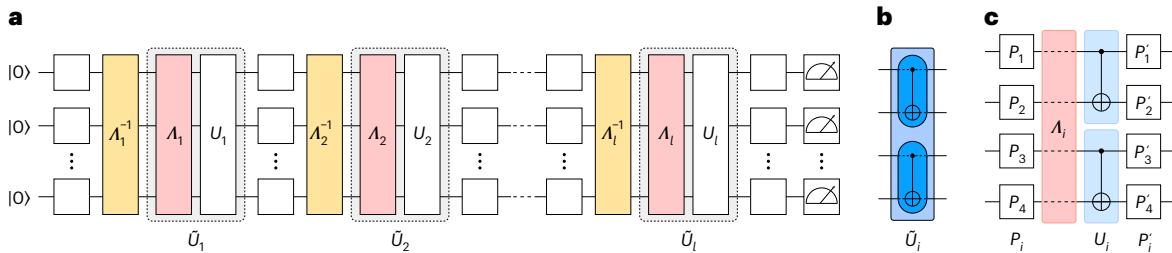


Fig. 1 | Context of the noise model. **a**, Ideal error mitigation of a circuit consisting of layers of noisy two-qubit gates interleaved with layers of single-qubit gates. **b**, Example of a layer consisting of two noisy CX gates.

c, Expansion of the same layer in terms of the ideal gates U_i and noise channel Λ_i , flanked with Pauli-twirl gates P_i and $P'_i = U_i P_i U_i^\dagger$, where P_i is sampled uniformly at random.

If needed, this can be ensured using Pauli twirling^{25–29}, as illustrated in Figs. 1b and 1c for an example with four qubits and two CX gates.

We present an efficient mitigation scheme that models the noise across each layer of two-qubit gates as a sparse Pauli–Lindblad error model. In our experiments, the model includes only weight-one and weight-two Pauli terms whose support coincides with the quantum processor’s connectivity. The parameters of the resulting model scale linearly with the number of qubits, which ensures that the model is efficiently represented and easy to learn. The inverse noise model is obtained simply by negating the model coefficients and gives rise to a quasi-probability distribution on Pauli matrices. We provide an efficient algorithm for sampling this distribution in linear time with the number of model coefficients. The mitigation Pauli terms can be combined with those used for twirling as well as with the single-qubit operations in the interleaved layers. The error mitigation scheme therefore maintains the original circuit structure and changes only the classical distribution of the single-qubit gates.

Pauli–Lindblad noise model

We model a given n -qubit Pauli noise channel Λ that arises from a sparse set of local interactions by a Lindblad generator³⁰ $\mathcal{L}(\rho) = \sum_{k \in \mathcal{K}} \lambda_k (P_k \rho P_k^\dagger - \rho)$, where \mathcal{K} is chosen as a poly(n) sized subset of all 4^n Pauli matrices, and $\lambda_k \geq 0$ denotes the model coefficient corresponding to Pauli matrix P_k . The Lindbladian is only chosen as a parameterization of Pauli channels and does not model the continuous system evolution³¹. The resulting model $\Lambda(\rho) = \exp[\mathcal{L}](\rho)$ then factors as the composition (Supplementary Information Section III)

$$\Lambda(\rho) = \bigcirc_{k \in \mathcal{K}} (w_k \cdot + (1 - w_k) P_k \cdot P_k^\dagger) \rho, \quad (1)$$

where $w_k = 2^{-1}(1 + e^{-2\lambda_k})$. Here \cdot stands as the place holder $T(\cdot)\rho = T(\rho)$ and $\bigcirc_{k \in [1, m]} T_k = T_1 \circ \dots \circ T_m$ denotes the composition of maps T_k . The model terms \mathcal{K} are chosen to reflect the noise interactions in the quantum processor, and their number, which determines the model complexity and expressivity, typically scales polynomially in n and therefore allows us to represent noise models for the full device by a small set of non-negative coefficients λ_k .

The fidelity of a Pauli matrix P_b with respect to Λ is given by $f_b = \frac{1}{2^n} \text{Tr}(P_b^\dagger \Lambda(P_b))$. Defining the symplectic inner product $\langle b, k \rangle_{\text{sp}}$ to be 0 if the Pauli matrices P_b and P_k commute and 1 otherwise, we can concisely express the relationship between model coefficients λ and the set $f = \{f_b\}_{b \in \mathcal{B}}$ of fidelities for an arbitrary set of Pauli matrices \mathcal{B} as $\log(f) = -2M(\mathcal{B}, \mathcal{K})\lambda$, where the logarithm is applied elementwise and the entries of binary matrix $M(\mathcal{B}, \mathcal{K})$ are given by $M_{b,k} = \langle b, k \rangle_{\text{sp}}$. For a given λ , this allows us to evaluate the fidelity of any set of Pauli matrices \mathcal{B} . More importantly, though, the relationship allows us to fit physical model parameters, $\lambda \geq 0$, given the fidelity estimates \hat{f} for a set of benchmark Pauli matrices \mathcal{B} by solving a non-negative least-squares

problem in $\log(\hat{f})$ (see Supplementary Information Section III.3 for more details).

Various methods of learning the fidelities of Pauli channels are known^{32–35} and have been implemented experimentally³⁶. The central idea in these methods is that the same noise process is repeated up to d times and the corresponding Pauli expectation values are measured at every depth. The fidelities for the noise channel can then be extracted from the decay rates in the resulting curves in a way that is robust to state-preparation and measurement (SPAM) errors. In Supplementary Information Section IV.2 we provide theoretical guarantees for the sample complexity for learning the error model. Under mild conditions on the minimal fidelity of the noise channel and the level of SPAM errors, we provide the following result for all the fidelities predicted by the model. Assume that the channel can be represented with the model Pauli matrices from set \mathcal{K} , and that the channel fidelities for Pauli matrices in \mathcal{B} are learned by benchmarking up to depth d with at least $2e^{-2} \log(2|\mathcal{B}|/\delta)$ circuit instances for each of the relevant measurement bases. Then it holds with probability at least $1 - \delta$ that the estimates \hat{f}_j of all fidelities f_j are bounded by

$$C_\epsilon^{-\tau} \leq f_j \hat{f}_j^{-1} \leq C_\epsilon^\tau, \quad (2)$$

with $C_\epsilon = \left(\frac{1+4\epsilon}{1-4\epsilon}\right)$ and $\tau = \sqrt{|\mathcal{K}| \cdot |\mathcal{B}|}/(\sigma_{\min}(M(\mathcal{B}, \mathcal{K}))d)$, where $\sigma_{\min}(M)$ is the smallest singular value of M .

Experimental model fitting

To illustrate the learning protocol, we first benchmark the four-qubit layer with two CX gates shown in Fig. 1b on a 27-transmon-qubit, fixed-connectivity processor with a heavy-hex topology, with qubits as indicated at the top of Fig. 2a. For all our experiments, we apply dynamical decoupling sequences during idle times of qubits in the layer. These idle times arise when one or more gates in the layer are substantially faster than the slowest one or when a qubit in the layer does not contain a gate (see also Supplementary Information Section VII.3). Repeated application of a noise channel in the context of self-adjoint two-qubit Clifford gates, such as CX and CZ gates, generally results in pairwise products of fidelities. Although inserting appropriate single-qubit gates between applications can increase the number of individual Pauli fidelity estimates, pairwise fidelities will always remain, leading to indeterminacy of model coefficients; for instance, we can express the pairwise fidelity $f_a f_b$ as $(\alpha f_a)(f_b/\alpha)$ for any α . We address this indeterminacy either through direct estimation of missing fidelities by measuring a single layer, at the cost of an additive error in the estimate and sensitivity to state preparation and readout errors, or through symmetry relations that hold under the reasonable assumption on the noise (Supplementary Information Section V).

With this in mind, we benchmark the four-qubit layer for increasing depths up to d in nine different bases to obtain all necessary data. Each

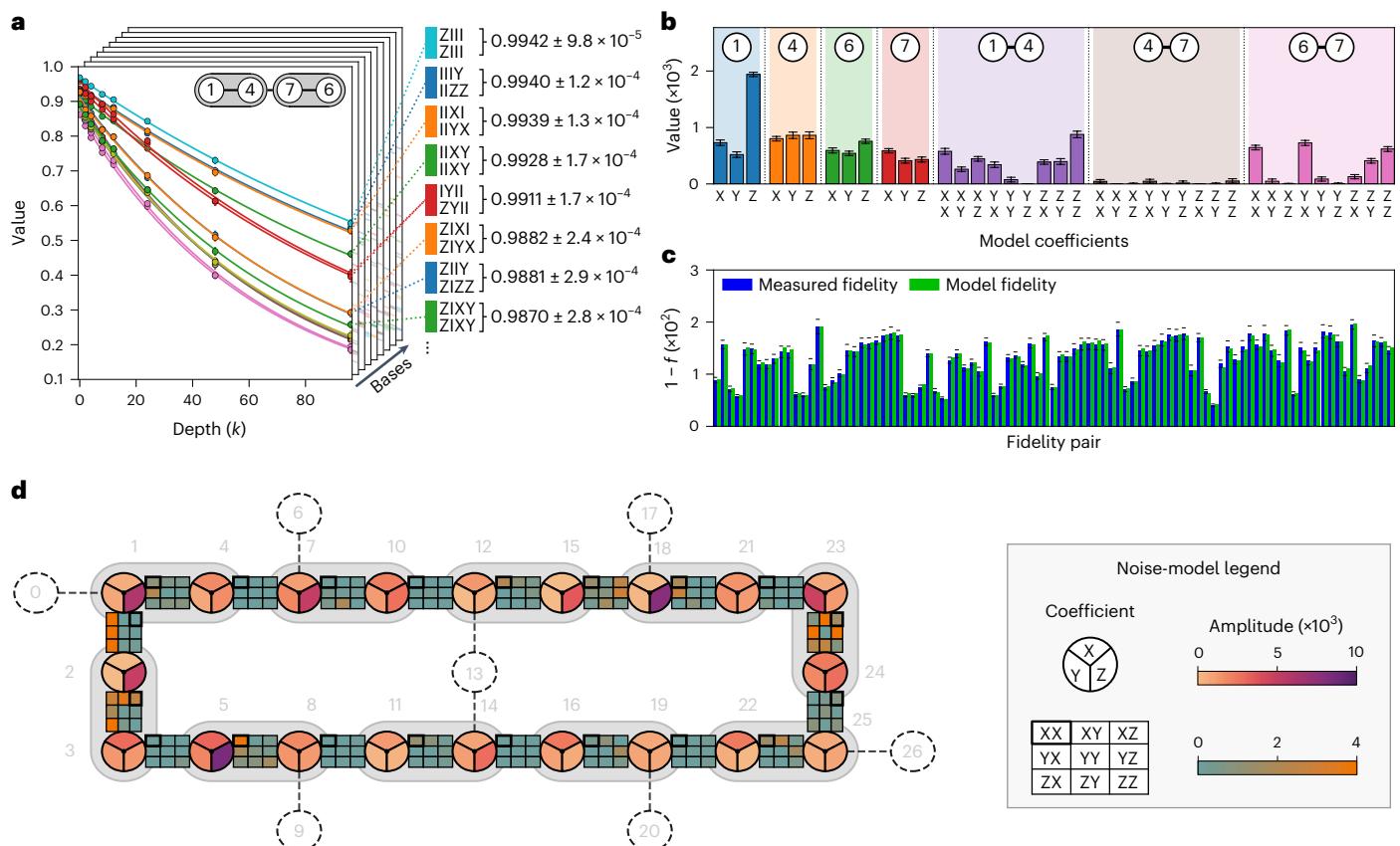


Fig. 2 | Learning the noise channel. **a**, The first step in learning our noise model, in this case for the four-qubit layer depicted in the top inset with two concurrent CX gates, is to measure a set of observables with increasing numbers of circuit repetitions k (even) up to some maximum depth d . This requires measurements in nine different bases, illustrated by the stacked planes. Associated with each observable P_0 is a fidelity of the form $\alpha_0(f_1 f_2)^{k/2}$, where α_0 is a constant that captures the state-preparation and measurement error, and f_1 and f_2 are the fidelities of the noise channel for two Pauli terms. We estimate the values of the different fidelity pairs in a consistent manner by fitting exponentially decaying curves through the data points of all observables that include the same pair, which may arise in multiple bases and possibly different observables, whose curves may have a different offset values α_0 as a result of differences in Pauli support and variability of SPAM errors across the device. The legend on the right-hand side illustrates the fidelity estimates for several pairs along with error

bars at one standard deviation around the estimate (largely covered by the markers). All standard deviations reported in this figure are based on a 100-fold bootstrap, obtained by resampling the circuit instances and shots. **b**, Model coefficients obtained using a non-negative least-squares fit of the log fidelities plus error bars at one standard deviation. **c**, Plot of one minus the fidelity for each of the measured fidelity pairs and horizontal error bars at three standard deviations, along with the corresponding fidelities from the learned noise model. **d**, Visualization of the sparse noise model of a 20-qubit layer with 10 concurrent CX gates (shaded pairs) overlaid on the topology of the IBM_HANOI quantum processor. Circles denote qubits (labelled by numbers); coloured wedges in the circle visualize the single-body X, Y and Z Lindblad coefficients (see legend (top)). Two-body coefficients, for example XX, for adjacent qubit pairs are visually represented by a 3×3 matrix (see legend (bottom)). The first Pauli character corresponds to the qubit adjacent to the highlighted tile.

data point in Fig. 2a represents an estimated observable in a given basis, averaged over 100 random circuit instances with 256 shots each. We then fit exponentially decaying curves through the data points corresponding to each unique fidelity pair $f_a f_b$, and augment the fidelities obtained this way with fidelity estimates resulting from the symmetry condition. From this, we obtain the model coefficients λ , shown in Fig. 2b, using an adapted non-negative least-squares fitting procedure that uses the modified relation $\log(f_1 f_2) = -2(M_1 + M_2)\lambda$ to reflect the use of pairwise fidelities (Supplementary Information Section V). As seen in Fig. 2c, the fidelities of the resulting model closely match the measured fidelities. This provides confidence that the selected model captures the noise accurately.

To illustrate scalability of the method we used the same protocol to learn the noise model for a 20-qubit layer involving ten concurrent CX gates. Figure 2d depicts the layer and the resulting model coefficients. The illustration visualizes the sparse-model coefficients as a map over the quantum processor. We emphasize that learning the 20-qubit noise model takes the same number of circuit instances as that of the 4-qubit model.

Probabilistic error cancellation

Once the noise model has been learned, it can be used to mitigate the noise using the PEC method⁹. The protocol implements the channel inverse A_i^{-1} through quasi-probabilistic sampling for each of the l layers. The inverse of the map A is given by $A^{-1}(\rho) = \exp[-\mathcal{L}](\rho)$, which is a non-physical map:

$$A^{-1}(\rho) = \gamma \bigcup_{k \in \mathcal{K}} (w_k \cdot -(1 - w_k)P_k \cdot P_k^\dagger) \rho \quad (3)$$

with sampling overhead $\gamma = \exp(\sum_{k \in \mathcal{K}} 2\lambda_k)$. This amounts exactly to inverting each individual factor in equation (1) due to commutativity of the factors. The product structure allows for a direct way of sampling the map. For each $k \in \mathcal{K}$, we sample the identity with probability w_k or apply the Pauli matrix P_k otherwise. We record the number of times m we have applied a non-identity Pauli matrix, compute a final Pauli matrix as the product of all sampled terms. Repeating this for each noise channel $i = 1, \dots, l$ with respective m_i and γ_i values, we construct a circuit instance in which each noisy layer is preceded with the corresponding

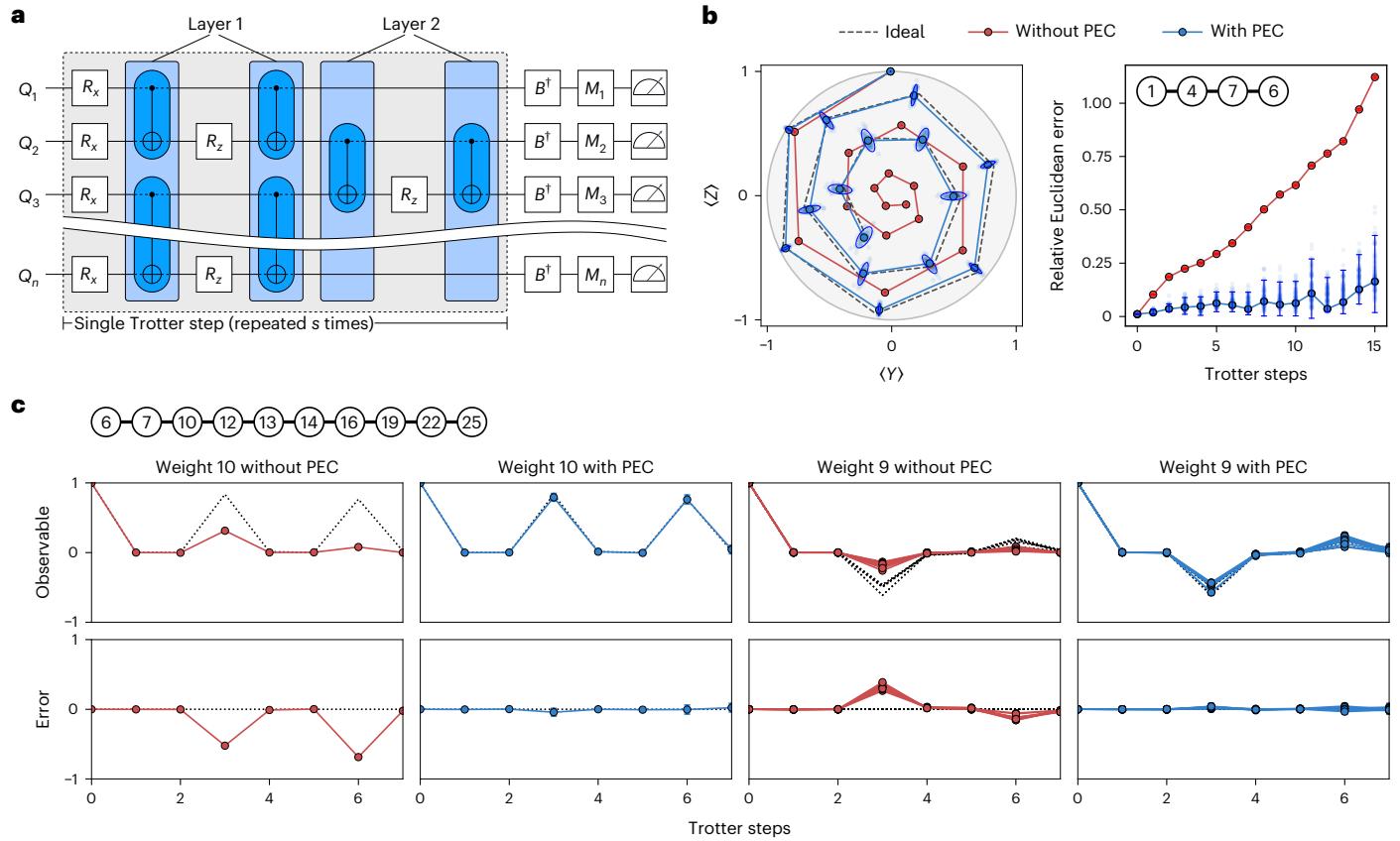


Fig. 3 | Error-mitigated time evolution of Ising spin chains. **a**, Trotter circuit for the Ising Hamiltonian in equation (4) over a one-dimensional n -qubit lattice. The shaded box represents a single Trotter step and is repeated s times, with associated $R_x(2\delta t)$ and $R_z(-2J\delta t)$ rotations. Each step comprises two instances of two unique CX layers. The B^\dagger gates select the measurement basis and the M gates facilitate our model-free readout-error mitigation³⁷. **b**, Time evolution of the Ising model for an $n = 4$ spin lattice with and without probabilistic error correction (PEC) for 15 Trotter steps; $h = 1$, $J = 0.15$, and $\delta t = 1/4$. Left, Trotterized time evolution of the global magnetization $\mathbf{M} := \sum_n (\langle X_n \rangle, \langle Y_n \rangle, \langle Z_n \rangle) / N$ shown in the Y - Z Bloch plane. The experimentally measured evolution (dots and solid lines) is

compared with the ideal noise-free one (dashed lines). Right, The error between ideal and measured magnetization vectors, in terms of the relative Euclidean distance $\|\mathbf{M} - \mathbf{M}_{\text{ideal}}\|_2 / \|\mathbf{M}_{\text{ideal}}\|_2$. The bootstrap-estimated error distribution for each data point is shown as clouds (light dots). Error bars (and ellipsoids) are given at two standard deviations around the data points. All standard deviations in this figure are based on a 100-fold bootstrap of the circuit instances and shots. **c**, Time evolution of the Ising model on a one-dimensional ten-qubit lattice sites (top) with $h = 1$, $J = 0.5236$ and $\delta t = 1/4$. All weight-10 (left) and weight-9 Pauli-Z observables (right) are plotted along with the ground truth (dashed). The error bars at two standard deviations are largely covered by the markers.

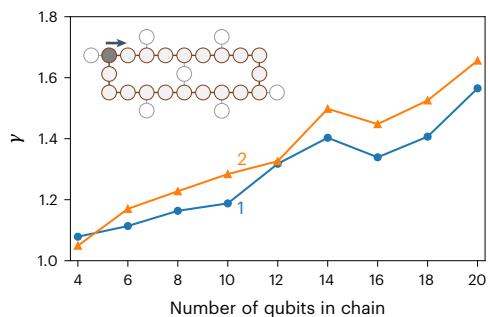


Fig. 4 | Mitigation sampling overhead. Sampling overhead γ for the two Ising layers as a function of the number of qubits in the Ising lattice chain. The chain is depicted as an inset at the top. The first qubit in the chain is at the top-left (dark shading) and the chain proceeds clockwise (arrow). Layers 1 and 2 have complementary CX gates on alternating pairs of qubits in the chain (see Fig. 3).

sampled Pauli matrix. The measurement outcome of the circuit is then multiplied by $\prod_{i=1}^l (-1)^{m_i} \gamma_i$. On average, this implements the inverse maps and produces an unbiased expectation value with sampling overhead $\gamma(l) = \prod_{i=1}^l \gamma_i$. In Supplementary Information Section VI.2, we

derive an error bound on the final expectation value that considers the errors in all steps of the procedure. The bound states that, given a quantum circuit with l layers whose learning layer satisfies equation (2), we can estimate the ideal expectation value $\langle A \rangle$ of an observable A with $\|A\| \leq 1$ by the average mitigated estimate $\langle \hat{A}_N \rangle$ using N error-mitigated circuit instances, such that

$$|\langle A \rangle - \langle \hat{A}_N \rangle| \leq (C_e^{\ell} - 1) + \gamma(l) \sqrt{2 \log(2/\delta')/N}$$

is satisfied with probability at least $1 - \delta'$. For modest noise, C_e can be expected to be close to one, which leads to a scaling that is only weakly exponential in l and τ . The sampling overhead $\gamma(l)$ dictates the resources needed to obtain a reliable estimator⁹.

Quantum simulation of the Ising model

As a practical application for noise mitigation with our proposed noise model, we consider time evolution of the one-dimensional transverse-field Ising model due to the Hamiltonian

$$H = -J \sum_j Z_j Z_{j+1} + h \sum_j X_j = -J H_{ZZ} + h H_X, \quad (4)$$

where J denotes the exchange coupling between neighbouring spins and h represents the transverse magnetic field. Unitary time evolution e^{-iHt} can be approximated by a first-order Trotter decomposition $(e^{iH_Z t/s} e^{-iH_X t/s})^s$ with s segments. We perform the time evolution on a linear chain of qubits, where we implement the unitary $\exp(iJ(Z_j Z_{j+1})\delta t)$ with $\delta t = t/s$ as a quantum circuit consisting of an $R_Z(-2J\delta t)$ rotation on qubit $j+1$ between two CX gates with control and target qubits j and $j+1$. Similarly, $\exp(-ihH_X\delta t)$ decomposes into a product of single-qubit rotations $R_X(2h\delta t)$ on each qubit j (for more details, see ref. 6). This results in circuits of the form shown in Fig. 3a. The circuit contains two unique layers of CX gates, one starting at even and one at odd locations in the qubit chain. Once the noise models for the two layers are learned, we generate random circuit instances. We apply readout-error mitigation on all observables (see ref. 37 for more on readout mitigation). To counter time-dependent fluctuations in the noise, we relearn the noise model after fixed intervals (see also Supplementary Information Section VII). The final observables are obtained after averaging.

As a first experiment, we consider the Ising-model dynamics for a spin chain with four sites with $h=1$ and $J=0.15$. Learning of the first layer was detailed in Fig. 2a–c and resulted in a sampling overhead factor $\gamma_1=1.031$. The same procedure, applied to the second layer, gave $\gamma_2=1.038$. We use the learned noise for error mitigation of the Ising-model dynamics in Fig. 3b. The number of mitigated circuit instances for each $s=0, 1, \dots, 15$ in Fig. 3b is given by $\min(200, 40\gamma^{2s})$, where $\gamma=\gamma_1\gamma_2=1.071$. Since we apply Pauli twirling in the unmitigated experiment, we used the same number of circuits instances for simplicity. All circuit instances were measured 1,024 times.

For each of the s Trotter steps, we compute the global magnetization component $\langle Z \rangle_s$ as the overall average of all weight-one Pauli-Z observables, and likewise for $\langle X \rangle_s$ and $\langle Y \rangle_s$. The resulting Y and Z magnetization components are plotted in Fig. 3b (left) along with the results obtained without PEC and exact simulation. We compare the relative Euclidean distance for the estimated and exact global magnetization in Fig. 3b (right).

Our second experiment considers the simulation of a one-dimensional lattice on ten qubits with $h=1$ and $J=0.5236$ for up to seven Trotter steps. High-weight observables are highly noise sensitive and serve as a demanding test of the method. In Fig. 3c, we compare the results for weight-9 and -10 Pauli-Z observables obtained with and without PEC. Mitigated observables exhibit vanishing residuals.

Discussion and conclusions

The remarkable accuracy of the error-mitigated observables in Fig. 3 provides strong evidence for the validity of our sparse noise model and learning protocol. It is nonetheless important to discuss potential limitations of our method, such as the sampling overhead. In particular, the variance in the estimator scales with the square of the sampling overhead factor $\gamma(l)$. This factor depends on the number of qubits (Fig. 4) as well as the circuit depth in terms of the number of layers. We can define a qubit- and depth-normalized version of the scaling factor, $\bar{\gamma}$, which allows us to conveniently express the sampling overhead for l layers on n qubits as $\bar{\gamma}^{nl}$. This normalized parameter itself can also be used as a metric to represent quantum processor performance; improvements in the hardware quality are reflected in lower $\bar{\gamma}$ values, which, in turn, translate into potentially dramatic decreases in the sampling overhead (see also Supplementary Information Section VI.3). Our work serves as a powerful example of how classical run-time overheads can be traded for tremendously improved quantum computation on noisy processors. However, this also highlights the importance of improving total circuit execution time³⁸, which will reduce the practical PEC overhead.

In conclusion, our results demonstrate for the first time, to our knowledge, a practical path to extend probabilistic error cancellation to remove the noise-induced bias from high-weight observables across the full circuit (see Fig. 3c). This is made possible by our sparse learning

protocol, which provides a versatile noise representation with rigorous theoretical bounds and near-constant learning with number of qubits, and an effective noise-inversion scheme. The accuracy of the model-reconstructed noise-fidelity pairs, as shown in Fig. 2c, and our error-mitigated observables validate the view that the Lindbladian learning is accurate, efficient and scalable. We expect our learning protocol to be a powerful characterization and benchmarking tool, and, more-broadly, to enable the study and mitigation of noise in quantum processors at a new scale.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41567-023-02042-2>.

References

1. Zhang, E. J. et al. High-performance superconducting quantum processors via laser annealing of transmon qubits. *Sci. Adv.* **8**, eabi6690 (2022).
2. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
3. Wu, Y. et al. Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.* **127**, 180501 (2021).
4. Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).
5. Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
6. Kim, Y. et al. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. *Nat. Phys.* <https://doi.org/10.1038/s41567-022-01914-3> (2023).
7. Havlíček, Vojtěch et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**, 209–212 (2019).
8. Schuld, M. & Killoran, N. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.* **122**, 040504 (2019).
9. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
10. Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* **7**, 021050 (2017).
11. Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).
12. Endo, S., Benjamin, S. C. & Li, Y. Practical quantum error mitigation for near-future applications. *Phys. Rev. X* **8**, 031027 (2018).
13. Guo, Y. & Yang, S. Quantum error mitigation via matrix product operators. *PRX Quantum* **3**, 040313 (2022).
14. Piveteau, C., Sutter, D. & Woerner, S. Quasiprobability decompositions with reduced sampling overhead. *Npj Quantum Inf.* **8**, 12 (2022).
15. Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. *J. Phys. Soc. Japan* **90**, 032001 (2021).
16. 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).
17. Takagi, R. Optimal resource cost for error mitigation. *Phys. Rev. Res.* **3**, 033178 (2021).
18. Takagi, R., Endo, S., Minagawa, S. & Gu, M. Fundamental limits of quantum error mitigation. *Npj Quantum Inf.* **8**, 114 (2022).
19. Lostaglio, M. & Ciani, A. Error mitigation and quantum-assisted simulation in the error corrected regime. *Phys. Rev. Lett.* **127**, 200506 (2021).

20. Xiong, Y., Chandra, D., Ng, S. X. & Hanzo, L. Sampling overhead analysis of quantum error mitigation: uncoded vs. coded systems. *IEEE Access* **8**, 228967–228991 (2020).
21. 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).
22. Song, C. et al. Quantum computation with universal error mitigation on a superconducting quantum processor. *Sci. Adv.* **5**, eaaw5686 (2019).
23. Zhang, S. et al. Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system. *Nat. Commun.* **11**, 587 (2020).
24. Strikis, A., Qin, D., Chen, Y., Benjamin, S. C. & Li, Y. Learning-based quantum error mitigation. *PRX Quantum* **2**, 040330 (2021).
25. Bennett, C. H. et al. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).
26. Knill, E. Fault-tolerant postselected quantum computation: threshold analysis. Preprint at <https://arxiv.org/abs/quant-ph/0404104> (2004).
27. Kern, O., Alber, G. & Shepelyansky, D. L. Quantum error correction of coherent errors by randomization. *Eur. Phys. J. D* **32**, 153–156 (2005).
28. Geller, M. R. & Zhou, Z. Efficient error models for fault-tolerant architectures and the Pauli twirling approximation. *Phys. Rev. A* **88**, 012314 (2013).
29. Wallman, J. J. & Emerson, J. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**, 052325 (2016).
30. Breuer, H.-P. & Petruccione, F. *The Theory of Open Quantum Systems* (Oxford Univ. Press, 2002); <https://doi.org/10.1093/acprof:oso/9780199213900.001.0001>
31. Sun, J. et al. Mitigating realistic noise in practical noisy intermediate-scale quantum devices. *Phys. Rev. Appl.* **15**, 034026 (2021).
32. Flammia, S. T. & Wallman, J. J. Efficient estimation of Pauli channels. *ACM Trans. Quantum Comput.* **1**, 3 (2020).
33. Erhard, A. et al. Characterizing large-scale quantum computers via cycle benchmarking. *Nat. Commun.* **10**, 5347 (2019).
34. Kimmel, S., da Silva, M. P., Ryan, C. A., Johnson, B. R. & Ohki, T. Robust extraction of tomographic information via randomized benchmarking. *Phys. Rev. X* **4**, 011050 (2014).
35. Helsen, J., Xue, X., Vandersypen, L. M. K. & Wehner, S. A new class of efficient randomized benchmarking protocols. *Npj Quantum Inf.* **5**, 71 (2019).
36. Harper, R., Flammia, S. T. & Wallman, J. J. Efficient learning of quantum noise. *Nat. Phys.* **16**, 1184–1188 (2020).
37. van den Berg, E., Minev, Z. K. & Temme, K. Model-free readout-error mitigation for quantum expectation values. *Phys. Rev. A* **105**, 032620 (2022).
38. Wack, A. et al. Quality, speed, and scale: three key attributes to measure the performance of near-term quantum computers. Preprint at <https://arxiv.org/abs/2110.14108> (2021).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

Data availability

Data are available from the authors upon reasonable request. Supplementary Fig. 7 in the supplementary information shows data representative of the performance of the cloud-based device IBM_HANOI, which is accessible online through the open-source Qiskit backend.

Acknowledgements

We thank S. Bravyi, D. T. McClure and J. M. Gambetta for helpful discussions. Research in characterization and noise learning was sponsored in part by the Army Research Office and was accomplished under Grant Number W911NF-21-1-0002. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Author contributions

K.T. and E.B. developed the theory. E.B., Z.M. and A.K. ran the experiments. All authors designed the experiments, analysed the data and wrote the paper.

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/s41567-023-02042-2>.

Correspondence and requests for materials should be addressed to Kristan Temme.

Peer review information *Nature Physics* thanks Ming Gong and Suguru Endo for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.