Quantum computation with universal error mitigation on superconducting quantum processor

Chao Song\textsuperscript{1}, Jing Cui\textsuperscript{1}, H. Wang\textsuperscript{1}, J. Hao\textsuperscript{2}, H. Feng\textsuperscript{2}, and Ying Li\textsuperscript{3}\textsuperscript{1} Interdisciplinary Center for Quantum Information and Zhejiang Province Key Laboratory of Quantum Technology and Device, Department of Physics, Zhejiang University, Hangzhou 310027, China, \textsuperscript{2} Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, \textsuperscript{3} Graduate School of China Academy of Engineering Physics, Beijing 100193, China

(Dated: December 23, 2020)

Medium-scale quantum devices that integrate about hundreds of physical qubits are likely to be developed in the near future. However, such devices will lack the resources for realizing quantum fault tolerance. Therefore, the main challenge of exploring the advantage of quantum computation is to minimize the impact of device and control imperfections without encoding. Quantum error mitigation is a solution satisfying the requirement. Here, we demonstrate an error mitigation protocol based on gate set tomography and quasiprobability decomposition. One- and two-qubit circuits are tested on a superconducting device, and computation errors are successfully suppressed. Because this protocol is universal for digital quantum computers and algorithms computing expected values, our results suggest that error mitigation can be an essential component of near-future quantum computation.

INTRODUCTION

Quantum computers are quantum-mechanical devices capable of solving problems that are believed to be intractable for classical computers. The most essential issue in practicing quantum computation is to deal with imperfections of the device and control that cause computation errors. Quantum error correction can suppress the chance of errors to an arbitrarily low level, which however is beyond the scope of near-future technologies\textsuperscript{[1, 2]}. For shallow algorithms executed on near-future quantum devices\textsuperscript{[3–9]}, quantum error mitigation (QEM) methods\textsuperscript{[10, 11]} are recently proposed to attain a computation result with error minimized, which are practical because encoding is not required. The QEM protocol based on a combination of gate set tomography and quasiprobability decomposition is one of those methods\textsuperscript{[12]}. We use Pauli transfer matrix (PTM) representation to notate quantum states, quantum gates and measurements as commonly adopted in quantum tomography. We define $\sigma_i$ as the $i^{th}$ operator from the $n$-qubit Pauli basis $\mathcal{P} = \{I, X, Y, Z\}^\otimes n$, where $I, X, Y, Z$ denote the identity and three Pauli matrices, respectively. In PTM representation, a state $\rho$ is expressed as a column vector $|\rho\rangle$ with elements $|\rho\rangle_i = \text{Tr}(\sigma_i \rho)$; An observable $Q$ is expressed as a row vector $\langle Q|$ with elements $\langle Q|_i = \text{Tr}(\sigma_i Q)/2^n$; The superoperator $U$ of a gate is expressed as a matrix $U_{ij} = \text{Tr}(\sigma_i U \sigma_j)/2^n$. Thus the expected value of the observable $Q$ in the state $\rho$ going through a sequence of gates $U_1, ..., U_N$ reads as follows:

$$\text{Tr}(QU_N...U_1\rho) = \langle(Q|U_N...U_1|\rho)\rangle.$$

In GST, quantum gates are reconstructed in a set of experiments. In each experiment, one of the gates is applied on an initial state, and then a measurement is performed to read the value of an observable [see Fig. 2(a)]. We select $\{|0\rangle, |1\rangle, |0+i1\rangle\}^\otimes n = \{\rho_j\}$ as initial states and Pauli operators in $\mathcal{P}$ as observables. Using the PTM representation, we can express the set of initial states as the state preparation matrix $A_{ij}^{\text{exp}} = \langle\sigma_i|\rho_j^{\text{exp}}\rangle$ and similarly express the set of observables as the readout matrix $B_{ij}^{\text{exp}} = \langle\sigma_i^{\text{exp}}|\sigma_j\rangle$. Here and below the operations with no superscripts are ideal and those with the superscript “exp” are physical and with errors as experimentally accessible ones with errors. GST can be seen as a self-consistent extension of the quantum process tomography, which takes into account all the errors occurred in the experimental operations including state preparations, quantum gates and measurements.
and the matrix characterizing the superoperator $\mathcal{U}$, i.e., $\mathcal{U}^{\exp} = B^{\exp} A^{\exp} U^{\exp}$, can be obtained by applying the relevant operations in sequence in the experiment.

Experimental operations of $A^{\exp}$, $B^{\exp}$ and $U^{\exp}$ can be reconstructed by analyzing the data of $g^{\exp}$ and $\tilde{U}^{\exp}$. However, we cannot exactly reconstruct experimental operations due to the insufficient information encoded in $g^{\exp}$ and $U^{\exp}$ in the presence of both state preparation and measurement errors. In our device, fidelities of the state preparation and single-qubit gates are much higher than other operations. Therefore, we assume that initial states are error-free and take $\tilde{A} = \langle \langle \sigma_i | \rho_j \rangle \rangle$ as a decent guess of $A^{\exp}$, where the caret symbol is introduced to differentiate an estimate from the physical operation itself. Then according to the linear inversion method [14–16], we have the estimates for $B^{\exp}$ and $U^{\exp}$ as $\tilde{B} = g^{\exp} \tilde{A}^{-1}$ and $\tilde{U} = B^{-1} U^{\exp} \tilde{A}^{-1}$, respectively. The difference between the physical gate $U^{\exp}$ and its estimate $\tilde{U}$ depends on state preparation errors (i.e. the difference between $A^{\exp}$ and $\tilde{A}$). We note that this difference is not important because of not only the high fidelity of state preparation in our device but also the self-consistency of GST. Even if state preparation errors are significant, assuming the error-free state preparation does not affect the accuracy of quantum computation using QEM [12]. With these results obtained from the GST experiment, we can decompose any ideal measurement and gates into experimentally achieved operations.

Given the decomposition formulas [12], we randomly generate circuits modified from the original circuit of the computation task and implement these random circuits to obtain the computation result with error mitigated, because errors in random circuits cancel with each other when taking the average. In this paper, we only decompose and replace the measurement and two-qubit gate in the original circuit while the state preparation and single-qubit gates are unchanged. We do this because the error of the previous two is larger than the latter by an order of magnitude. To be explicit, by heralding the ground state for qubit preparation and measuring the heating rate [17], we estimate the ground state preparation error to be below 0.25%; error rates of single-qubit gates are calibrated to be below 0.25% in randomized benchmarking [18–20]. In comparison, the readout error is about 3.5% for the ground state and 5.7% for the excited state; the two-qubit gates applied in this paper have errors around 7%. We remark that gate fidelities can be further boosted on our device, however, which is unnecessary for the purpose of QEM demonstration.

**MITIGATING READOUT ERROR IN ONE-QUBIT COMPUTATION**

We first test the effect of QEM with a one-qubit computation, whose circuit is shown in Fig. 3(a). In this circuit, we initialize the qubit by heralding the state $|0\rangle$, with a state fidelity above 0.997. Then the gate $X_{\pi/2} = e^{-i\pi/2X}$, whose gate fidelity is calibrated to be 0.998 by randomized benchmarking, is applied to rotate the qubit state around the x-axis by $\pi/2$, after which the operator $Z$ is measured. The measurement is denoted by $M_Z$ in the circuit, which yields an outcome of $-1$ or $1$. We repetitively implement the circuit for 3,000 times and take the average as the expected value of $Z$. We then repeat the same procedure to obtain 100 expected values. The histogram analysis of the resulting data is shown in the upper panel of Fig. 3(f). The average over 100 expected values gives $\langle Z \rangle^{\exp} = 0.027 \pm 0.016$.

Since the state preparation and single-qubit gate are both quite precise, the relatively large deviation of $\langle Z \rangle^{\exp}$ from zero, i.e. the ideal result, is mainly due to the readout error, for which we intend to mitigate by decomposing the readout operation. To work out the decomposition formula, we obtain the Gram matrix $g^{\exp}$ in the experiment, and the result is shown in Fig. 2(b). Assuming the error-free state preparation, we can take a reasonable estimate of $A^{\exp}$ as

$$\tilde{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 \end{pmatrix},$$

where the Pauli operator basis is sorted in the order $I, X, Y, Z$. With $\tilde{B} = g^{\exp} \tilde{A}^{-1}$, we can decompose the observable $Z$ in the form $Z = \sum_{i} q_i \hat{\sigma}_i$, where $\langle \langle \hat{\sigma}_i \rangle \rangle$ are rows of $\tilde{B}$, and the quasi-probabilities $q_i$ are elements of the vector $q = \langle \langle Z | \tilde{B}^{-1} \rangle \rangle$. Quasi-probabilities are real but can be greater than unity and be negative. In the quantum computation with QEM, we implement random circuits for 3,000 times. In each of them, the measurement of $Z$ in the original circuit is replaced by the measurement of $\sigma_i^{\exp}$ [see Fig. 3(b)] with the probability $|q_i|/ \sum_i |q_i|$. Here, $\sigma_i^{\exp}$ is the physical measurement whose estimate is $\hat{\sigma}_i$. The measurement outcome is $+1$ or $-1$ (the outcome is always $+1$ in the measurement of $\sigma_0^{\exp}$, i.e., the identity operator $I$). When taking the average of measurement outcomes, each outcome is multiplied by a weight factor of $w_i = sgn(q_i) \cdot \sum_i |q_i|$. We take this weighted average as the expected value of $Z$. Then we repeat the above procedure to obtain 100 expected values, and the average over all these expected values yields $\langle Z \rangle^{QEM} = 0.003 \pm 0.022$ as shown in the lower panel of Fig. 3(f). The accuracy of the computation is successfully improved compared with the computation without QEM.
FIG. 1. (a) Flowchart of the universal quantum error mitigation, which has two stages: gate set tomography (GST) and the random circuit computation. (b) Layout of four qubits actively used in the experiment. The information is encoded in Q1 and Q2, and the other two qubits QA1 and QA2 are ancillary qubits. (c) Controlled-$\phi$-phase gate $C_{\phi}$ realized using the dressed state gate $U_{\phi}$ and single-qubit gates. The single-qubit gate $P_{\phi} = e^{-i\frac{\phi}{2}P}$, where $P = X, Y, Z$. (d) Reset gate $R_{\psi}$ realized using an ancillary qubit QA1 or QA2, which reinitialize the qubit Q1 or Q2 in the state $|\psi\rangle = G|0\rangle$.

MITIGATING READOUT AND ENTANGLING GATE ERRORS IN TWO-QUBIT COMPUTATION

Now we turn to a two-qubit computation, taking the deterministic quantum computation with pure states (DQC1) \cite{21} as an example. The circuit is shown in Fig. 2(c). A main error source in this circuit is the controlled-$\phi$-phase gate $C_{\phi} = \frac{1+iZ}{2} \otimes I + \frac{1-iZ}{2} \otimes Z_{\phi}$, where $Z_{\phi} = e^{-i\frac{\phi}{2}Z}$. This gate is realized with a two-qubit dressed state gate $U_{\phi}$ \cite{22} plus ten single-qubit gates as shown in Fig. 1(c). The two-qubit dressed state gate essentially achieves a controlled-$\phi$-phase gate in the $X$ basis, and single-qubit gates are used to transform the $X$ basis into the $Z$ basis. Fidelities of different $C_{\phi}$ gates are $0.958 \pm 0.010$, $0.935 \pm 0.011$, $0.920 \pm 0.011$, and $0.915 \pm 0.011$ for $\phi = \pi/4$, $\pi/2$, $3\pi/4$, and $\pi$, respectively, characterized using GST. We remark that gate fidelities can be further boosted on our device, however, which is unnecessary for the purpose of QEM demonstration.

To mitigate the error in $C_{\phi}$, we need first to work out the decomposition formula. Given the ideal superoperator $U$ representing $C_{\phi}$, the decomposition formula reads $U = \sum_{j} q_{j}^{U} \hat{U}_{j}$, where $q_{j}^{U}$ is a set of quasi-probabilities as introduced previously and $\hat{U}_{j}$ is the estimate of a set of experimentally achieved operations $U_{j}^{\text{exp}}$ as defined below and in Table. \[1\]

In our experiment, 257 operations are used for decomposing an ideal $C_{\phi}$ gate. The first 256 operations are generated from the tensor product of 16 single-qubit operations, which include measurement and reset gates, as listed in Table. \[1\]. The 257th operation is the gate $C_{\phi}$ modified by the Pauli twirling as we will explain soon. We reconstruct the experimental operations of $C_{\phi}$ and single-qubit measurement-reset gates in GST, while we simply assume single-qubit gates are error-free, because single-qubit gates can be experimentally implemented with high fidelity. Unlike the state preparation, assuming error-free single-qubit gates can potentially cause inaccuracy in the quantum computation with QEM. PTM of the controlled-$\pi$-phase gate $C_{\pi}$ obtained using GST is illustrated in Fig. 2(c) as an example. These 257 operations are linearly independent, which ensures that the decomposition solutions can always be found by solving a system of linear equations. In all solutions we choose the one with the minimum in $\sum_{k} |q_{k}^{U}|$ to minimize the variance of the computation result.

The measurement-reset operation is occasionally in random circuits [see Table. \[1\]]. In order to minimize the time of reset, we realize the reset gate using ancilla qubits QA1 and QA2 on the same chip [see Fig. 1(b)]. Each re-
set operation uses an ancilla qubit initially prepared in the ground state $|0\rangle$, then a swap gate is applied to reinitialize the target qubit when the reset is requested [13], following which a single-qubit gate $G$ with the fidelity above 0.997 rotates the qubit to the state $|\psi\rangle$ as shown in Fig. 4(d). The whole measurement-reset operation typically has a fidelity of around 0.916.

Pauli twirling converts the error in a gate into stochastic Pauli errors [23–25], which can reduce the variance of the computation result [12]. The circuit of the gate $C_\phi$ with Pauli twirling is shown in Fig. 4(d). We sandwich $C_\phi$ in between four Pauli gates (two for each qubit) which are randomly chosen but conserves the gate $C_\phi$ up to a global phase difference. If all gates are error-free, the two-qubit gate realized in this way is still $C_\phi$, i.e. $[C_\phi] = \sum_{a,b} p_{a,b} |\sigma_a \otimes \sigma_b C_\phi \sigma_a \otimes \sigma_b\rangle$. Here we use the bracket notation to denote a superoperator $[|\rho\rangle] = C_\phi \rho C_\phi^\dagger$, i.e. $\{\sigma\}$ are single-qubit Pauli gates chosen to satisfy $\sigma_a \otimes \sigma_b C_\phi \sigma_a \otimes \sigma_b = \eta C_\phi$, $\eta$ can be any phase factor, and $p_{a,b}$ is the probability. If gates have errors, the two-qubit gate will be effectively changed by the Pauli twirling. Using the twirled gate as the 257th operation in the decomposition, we need the estimate of the twirled gate, which is $\hat{U} = \sum_{a,b} p_{a,b} |\sigma_a \otimes \sigma_b \rangle \langle \sigma_a \otimes \sigma_b|$, where $\hat{U}$ is the estimate of the experimental operation of $C_\phi$ obtained in the experiment. For $\phi = \pi$, the distribution of Pauli gates is uniform, i.e. $p_{a,b} = 1/16$. For other $C_\phi$ gates where $\phi \neq \pi$, Pauli gates are chosen from a subset: We take $p_{a,b} = 1/4$ if $\sigma_a \otimes \sigma_b \in \{I, Z\} \otimes \{I, Z\}$, and $p_{a,b} = 0$ otherwise.

In the two-qubit computation with QEM, to mitigate both readout and two-qubit gate errors, the gate $C_\phi$ is randomly replaced by the gate $\hat{U}^{\text{stoch}}$ with the probability $|\langle q_j | \hat{U} \rangle|^2$, and the measurement of $X$ is replaced by the measurement of $\hat{X}^{\text{stoch}}$ with the probability $|\langle q_{j'} | \hat{X} \rangle|^2$, where $q'' = \langle X \rangle B^{-1}$. Similar to the one-qubit case, the measurement outcome of each random circuit is multiplied by a weight factor of $w_{j,i} = \text{sgn}(q_j q_{j'}') \cdot \sum_k |q_k'| \cdot \sum_l |q_l|$, and we take the weighted average as the computation result as shown in Fig. 4(e).

In the experiment, we adjust the phase $\phi$ of $C_\phi$ and measure the expected value of $X$. When implementing the computation with QEM, we randomly sample a circuit according to the decompositions of both $C_\phi$ and the observable $X$. Representative decomposed sampling circuits are shown in Fig. 4(e). The experiment result is shown in Fig. 4(g), which demonstrates a significant improvement on the computation accuracy.

The most significant improvement is obtained at $\phi = \pi/2$, in which case the difference between the computation result and the ideal value is reduced from 0.1690 to 0.0102 by using QEM. To estimate the fidelity required to achieve the same computation accuracy, we consider a quantum system with depolarizing error channels [24] and assume that the state preparation and single-qubit gates are ideal. The depolarizing error channel either preserves or completely destroys the information with cer-
taining probabilities [27], which does not characterize our device. We choose the depolarizing model because it takes all possible errors into account with equal probability. In the depolarizing model, the two-qubit gate and measurement with the fidelity \( \sim 99.3\% \) are required to achieve the computation accuracy 0.0102, which is comparable to the highest fidelity reported in the superconducting qubit system [20, 28].

**DISCUSSIONS**

For multi-qubit devices, GST of the entire device is not practical, because the experiment time increases exponentially with the qubit number. Similarly, the number of operations for decomposing a multi-qubit gate also increases exponentially with the qubit number. Single-qubit and two-qubit gates are sufficient for the universal quantum computation. Therefore, if errors are uncorrelated, we only need to implement GST up to two qubits and decompose two-qubit gates, as demonstrated in our experiment. Errors are uncorrelated if the evolution of two qubits under a two-qubit gate is independent from the evolution of other qubits. As a result of the evolution, the quantum operation on the entire device can be factorized into the product of an operation on the two qubits and operations on other qubits. It is similar for single-qubit gates. In our device, the primary dephasing noise is dominated by fluctuators in the form of magnetic moments, whose influence is local in each individual physical qubit, and, therefore, the dephasing-induced errors are uncorrelated between qubits. In our experiment, we have neglected error correlations in GST, so that single-qubit operations are characterized in single-qubit tomography even in the two-qubit experiment. Neglecting error correlations sacrifices accuracy of QEM. A significant effect of correlations on the computation result is not observed in our experiment.

We have experimentally demonstrated that the universal QEM protocol can significantly reduce the error in quantum computation on a noisy quantum device. The protocol in our experiment does not require sub-threshold error rate or tremendous additional physical qubit resource. Compared with the algorithm-specified protocol [7] and the extrapolation of gate time [9], the combination of gate set tomography and quasi-probability decomposition is not restricted to the algorithm or error model. A few techniques in QEM are explored: estimate of the state preparation matrix according to ideal states, Pauli twirling for randomizing the error, approximate GST and decomposition neglecting error correlations. An important factor limiting the circuit depth in our demonstration is the variance of computation result, which depends on the error rate of quantum gates. Improvement in gate fidelity can extend the circuit depth, and relatively deep circuits can be implemented on intermediate-scale devices with a feasible fidelity [12]. By demonstrating the power of error mitigation techniques on the superconducting quantum device, our results highlight the potential of using such techniques in computation tasks on near-future quantum devices.

**Acknowledgments**

We thank Xiaobo Zhu and Dongning Zheng for fabricating the device. This work was supported by the National Basic Research Program of China (Grant No. 2017YFA0304300) and the National Natural Science Foundations of China (Grants No. 11434008 and No. 11725419). YL is supported by NSAF (Grant No. U1730449).

---

* yli@gscmap.ac.cn

[1] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, *Surface codes: Towards practical large-scale quantum computation*, Phys. Rev. A 86, 032324 (2012).

[2] J. O’Gorman and E. T. Campbell, *Quantum computation with realistic magic state factories*, Phys. Rev. A 95, 032338 (2017).

[3] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, *A variational eigenvalue solver on a photonic quantum processor*, Nat. Commun. 5, 4213 (2014).

[4] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Covenev, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, *Scalable quantum simulation of molecular energies*, Phys. Rev. X 6, 031007 (2016).

[5] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, *Quantum implementation of unitary coupled cluster for simulating molecular electronic structure*, Phys. Rev. A 95, 020501 (2017).

[6] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, *Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets*, Nature 549, 242 (2017).

[7] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, *Computation of molecular spectra on a quantum processor with an error-resilient algorithm*, Phys. Rev. X 8, 011021 (2018).

[8] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jursic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, *Quantum chemistry calculations on a trapped-ion quantum simulator*, Phys. Rev. X 8, 031022 (2018).

[9] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, *Extending the computational reach of a noisy superconducting quantum processor*, arXiv:1805.01492

[10] Y. Li and S. C. Benjamin, *Efficient variational quantum simulator incorporating active error minimization*, Phys. Rev. X 7, 021050 (2017).

[11] K. Temme, S. Bravyi, and J. M. Gambetta, *Error mti-
FIG. 3. (a) Circuit of the one-qubit computation. In the quantum error mitigation (QEM), the measurement of the observable $Z$ is replaced by random gates. (b) Random circuit of the one-qubit computation, in which the measurement in the original circuit is replaced by the measurement of the observable $\sigma_i$. (c) Circuit of the two-qubit DQCp computation. In QEM, the two-qubit gate and the measurement are replaced by random operations. (d) Circuit of the Pauli twirling. (e) Representative random circuits of the two-qubit computation. (f) Results of the one-qubit computation. The probability distribution of the computation result is plotted. Without error, the ideal result is $\langle Z \rangle = 0$. (g) Results of the two-qubit computation. Each data point is obtained using $10,000$ times to compute one average value of $X$ and repeat the computation to obtain $100$ average values. The error bar indicates the standard deviation of these average values.

[12] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X 8, 031027 (2018).

[13] C. Song, K. Xu, W. Liu, C.-P. Yang, S.-B. Zheng, H. Deng, Q. Xie, K. Huang, Q.-J. Guo, L. Zhang, P. Zhang, D. Xu, D. Zheng, X. Zhu, H. Wang, Y.-A. Chen, C.-Y. Lu, S. Han, and J.-W. Pan, 10-qubit entanglement and parallel logic operations with a superconducting circuit, Phys. Rev. Lett. 119, 180509 (2017).

[14] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, Phys. Rev. A 87, 062119 (2013).

[15] D. Greenbaum, Introduction to quantum gate set tomography, arXiv:1509.02921.

[16] R. Blume-Kohout, J. King Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault toler-
ance threshold with gate set tomography, Nat. Commun. 8, 14485 (2017).

[17] Z. Chen, J. Kelly, C. Quintana, R. Barends, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Lucero, E. Jeffrey, A. Megrant, J. Mutus, M. Neeley, C. Neill, P. J. J. O’Malley, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Korotkov, and J. M. Martinis, Measuring and suppressing quantum state leakage in a superconducting qubit, Phys. Rev. Lett. 116, 020501 (2016).

[18] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B Quantum Semiclassical Opt. 7, S347 (2005).

[19] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Phys. Rev. A 77, 012307 (2008).

[20] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O’Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature 508, 500 (2014).

[21] E. Knill, Power of one bit of quantum information, Phys. Rev. Lett. 81, 5672 (1998).

[22] Q. Guo, S.-B. Zheng, J. Wang, C. Song, P. Zhang, K. Li, W. Liu, H. Deng, K. Huang, D. Zheng, X. Zhu, H. Wang, C.-Y. Lu, and J.-W. Pan, Dephasing-insensitive quantum information storage and processing with superconducting qubits, Phys. Rev. Lett. 121, 130501 (2018).

[23] E. Knill, Fault-tolerant postselected quantum computation: threshold analysis, arXiv:quant-ph/0404104.

[24] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A 94, 052325 (2016).

[25] J. O’Gorman, N. H. Nickerson, P. Ross, J. J. L. Morton, and S. C. Benjamin, A silicon-based surface code quantum computer, npj Quantum Information 2, 15019 (2016).

[26] E. Knill, Quantum computing with realistically noisy devices, Nature 434, 39 (2005).

[27] Supplementary Information

[28] E. Jeffrey, D. Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O’Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, Fast accurate state measurement with superconducting qubits, Phys. Rev. Lett. 112, 190504 (2014).

In the experiment we herald the state $|0\rangle$ to initialize the qubit. We need to wait 1.6 $\mu$s after the measurement pulse for the readout resonator to settle down before any further operation can be applied, during which the qubit may be heated again, causing the ground state preparation error. In Fig. 52, we show the measurement of the heating rate of Q1, which indicates a ground state preparation error below 0.0025.

**Readout Error for Q1 and Q2.**

In order to measure the readout error, we repeatedly prepare the qubit to the state $|0\rangle$ or $|1\rangle$ and measure the probability of incorrect output. The results are shown in Table 51.
FIG. S2. Heating rate measurement. The sequence for measuring heating rate is shown in the inset. The first measurement heralds the ground state. After the delay $\tau$, we measure the probability of the qubit being in the excited state. The black line is a fit to the data using the formula $P_1 = \Gamma_\uparrow \left[ 1 - e^{-(\Gamma_\uparrow + \Gamma_\downarrow)\tau} \right]$, where $P_1$ is the $|1\rangle$ state population after subtracting readout error and $\Gamma_\uparrow$, $\Gamma_\downarrow$ are heating and decay rates of the qubit. We take $\Gamma_\downarrow = 1/T_1$ and optimize $\Gamma_\uparrow$ to fit the data. In our experiment, we wait 1.6 $\mu$s after heralding for the readout resonator to settle down. Taking $T_1 = 24.7$ $\mu$s and $\tau = 1.6$ $\mu$s in the equation, we obtain the ground state preparation error of 0.0023.

TABLE S1. Error rates of readout measured by repeatedly preparing the state $|0\rangle$ or $|1\rangle$ and measuring the probability of incorrect output.

|       | Q1     | Q2     |
|-------|--------|--------|
| error rate of ground state readout | 0.0343 | 0.0360 |
| error rate of excited state readout | 0.0526 | 0.0607 |

ONE-QUBIT QEM EXPERIMENT

Gram matrices $g$ for Q1 and Q2 are shown in Fig. S3(a). The decomposition coefficients (i.e. quasiprobabilities) of the measurement $M_Z$ is shown in Fig. S3(b). In the QEM experiment, we take a sample number of 3,000 to obtain an average value $\langle Z \rangle$ and repeat the experiment for 100 times to obtain the distribution of $\langle Z \rangle$. The result is shown in Fig. S3(c).

MEASUREMENT AND RESET GATES

In the two-qubit QEM experiment, we used two ancillary qubits QA1 and QA2 to reset Q1 and Q2. The Circuit for the measurement-reset gate of $M_{\uparrow\downarrow}$ and $R_{\langle 0+1 \rangle}$ on Q1 is shown in Fig. S4(a). An i-swap gate is used in the reset gate, which is shown in Fig. S4(b). The GST result for this measurement-reset gate is shown in Fig. S4(c). Gate fidelities for all measurement-reset gates used in this paper are listed in Table. S2.

DECOMPOSING $C_\phi$ GATE

When decomposing two-qubit gates, we suppose that the first 10 gates listed in Table. I in the main text are ideal. The last 6 gates, i.e. the measurement-reset gates are characterized by using GST, and the GST results are used in the decomposition. Coefficients (i.e. quasiprobabilities)
The fidelity was acquired by firstly transforming the Pauli transfer matrix to the $\chi$ matrix and then calculating $F = \frac{\text{Tr}(\chi^\text{exp}\chi^\text{ideal})}{\text{Tr}(\chi^\text{exp})\text{Tr}(\chi^\text{ideal})}$.

| gate           | Q1            | Q2            |
|----------------|---------------|---------------|
| $M_{\frac{1}{2}X}, R_{[0+1]}$ | $0.937 \pm 0.034$ | $0.889 \pm 0.043$ |
| $M_{\frac{1}{2}X}, R_{[0-1]}$ | $0.906 \pm 0.034$ | $0.871 \pm 0.031$ |
| $M_{\frac{1}{2}Y}, R_{[0+1]}$ | $0.947 \pm 0.023$ | $0.893 \pm 0.022$ |
| $M_{\frac{1}{2}Y}, R_{[0-1]}$ | $0.941 \pm 0.022$ | $0.909 \pm 0.019$ |
| $M_{\frac{1}{2}Z}, R_{[0]}$  | $0.943 \pm 0.015$ | $0.903 \pm 0.013$ |
| $M_{\frac{1}{2}Z}, R_{[1]}$  | $0.938 \pm 0.023$ | $0.915 \pm 0.019$ |

Before Pauli Twirling

After Pauli Twirling

FIG. S5. Quasiprobability distribution in the decomposition of the $C_x$ gate without (up) and with (down) the Pauli twirling.

DEPOLARIZING ERROR CHANNELS

For a two-qubit gate, if the fidelity is $F_2$ and the error channel is depolarizing, the gate succeeds (i.e. the gate does not cause any error) with the probability $1 - \epsilon_2 = (16F_2 - 1)/15$ and fails (i.e. the two-qubit state becomes the maximally mixed state) with the probability $\epsilon_2 = 16(1 - F_2)/15$. For a measurement, if the fidelity is $F_M$ and the error channel is depolarizing, the measurement succeeds (i.e. the outcome is true) with the probability $1 - \epsilon_M = 2F_M - 1$, and the measurement fails (i.e. the outcome is completely random) with the probability $\epsilon_M = 2(1 - F_M)$. Suppose that $\langle X \rangle^{\text{ideal}}$ is the computation result when all operations are error-free, the computation result becomes $\langle X \rangle^{\text{ideal}}(1 - \epsilon_2)(1 - \epsilon_M)$ when error channels are switched on, and the difference from the ideal value is $\Delta \langle X \rangle \simeq \langle X \rangle^{\text{ideal}}(\epsilon_2 + \epsilon_M) \simeq \langle X \rangle^{\text{ideal}}(3 - F_2 - 2F_M)$. When $\phi = \pi/2$, $\langle X \rangle^{\text{ideal}} = 1/2$, therefore $\Delta \langle X \rangle \simeq 0.01$ if $F_2 = F_M = 0.993$. 

We find that by applying the Pauli twirling, the cost $\sum_j |q_j|$ reduced from 8.71 to 1.77. In the two-qubit QEM experiment, we generate random circuits according to the coefficients with the Pauli twirling applied.

Before Pauli Twirling

After Pauli Twirling

FIG. S5. Quasiprobability distribution in the decomposition of the $C_x$ gate without (up) and with (down) the Pauli twirling.

TABLE S2. Gate fidelities for all measurement-reset gates used in this paper. Gates were characterized by using GST. The fidelity was acquired by firstly transforming the Pauli transfer matrix to the $\chi$ matrix and then calculating $F = \frac{\text{Tr}(\chi^\text{exp}\chi^\text{ideal})}{\text{Tr}(\chi^\text{exp})\text{Tr}(\chi^\text{ideal})}$. 

Before Pauli Twirling

After Pauli Twirling

FIG. S5. Quasiprobability distribution in the decomposition of the $C_x$ gate without (up) and with (down) the Pauli twirling.