

# Scalable evaluation of quantum-circuit error loss using Clifford sampling

Zhen Wang<sup>1,\*</sup>, Yan Zhu Chen<sup>2,3,\*</sup>, Zixuan Song<sup>1</sup>, Dayue Qin<sup>4</sup>, Hekang Li<sup>1</sup>, Qiujiang Guo<sup>1</sup>, H. Wang<sup>1</sup>, Chao Song<sup>1,†</sup>, and Ying Li<sup>4,‡</sup>

<sup>1</sup>*Interdisciplinary Center for Quantum Information,  
State Key Laboratory of Modern Optical Instrumentation,  
and Zhejiang Province Key Laboratory of Quantum Technology and Device,  
Department of Physics, Zhejiang University, Hangzhou 310027, China*

<sup>2</sup>*C. N. Yang Institute for Theoretical Physics, State University of New York at Stony Brook,  
Stony Brook, NY 11794-3840, USA*

<sup>3</sup>*Department of Physics and Astronomy, State University of New York at Stony Brook,  
Stony Brook, NY 11794-3800, USA*

<sup>4</sup>*Graduate School of China Academy of Engineering Physics, Beijing 100193, China*

A major challenge in developing quantum computing technologies is to accomplish high precision tasks by utilizing multiplex optimization approaches, on both the physical system and algorithm levels. Loss functions assessing the overall performance of quantum circuits can provide the foundation for many optimization techniques. In this paper, we use the quadratic error loss and the final-state fidelity loss to characterize quantum circuits. We find that the distribution of computation error is approximately Gaussian, which in turn justifies the quadratic error loss. It is shown that these loss functions can be efficiently evaluated in a scalable way by sampling from Clifford-dominated circuits. We demonstrate the results by numerically simulating ten-qubit noisy quantum circuits with various error models as well as executing four-qubit circuits with up to ten layers of two-qubit gates on a superconducting quantum processor. Our results pave the way towards the optimization-based quantum device and algorithm design in the intermediate-scale quantum regime.

## I. INTRODUCTION

In quantum computation, errors caused by decoherence and imperfect controls form the main obstacle to meaningful applications, such as solving integer factorization and quantum chemistry problems [1–4]. Evaluating the error severity in quantum computation is essential for improving the design of device [5–7], optimizing control parameters [8], and minimizing errors with mitigation protocols [9–11]. Various schemes of quantum system characterization have been developed. Randomized benchmarking [12–21] and quantum process tomography (QPT) [22–27] can measure the average gate fidelity and full information of a noisy quantum channel, respectively. These two methods are efficient in systems with a few qubits. Cross-entropy benchmarking is used to verify a multi-qubit system but cannot be directly applied to quantum-supremacy circuits that are unsimulatable on classical computers [28, 29]. We can infer the performance of a large system by dividing it into tractable subsystems and characterizing each subsystem individually [11, 29–33]. However, this approach only works when the crosstalk is insignificant. The temporal correlation of noise is another factor that usually limits the effectiveness of characterization techniques [27, 34–40].

Many quantum algorithms utilize multi-qubit and deep quantum circuits. Even for variational quantum computation, which is promising for near-term applications, we need to implement hundreds of gates on tens of qubits [41–45]. In this paper, we propose an intuitive method that can efficiently characterize large quantum circuits, in the presence of both spatial and temporal error correlations. The resource cost of our method scales

polynomially with the circuit size.

We take the quadratic loss function of computation error [46] as the measure of error severity, which is

$$L_{\mathbb{R}}(\mathbf{F}) \equiv \frac{1}{|\mathbb{R}|} \sum_{\mathbf{R} \in \mathbb{R}} \text{Error}(\mathbf{F}, \mathbf{R})^2. \quad (1)$$

Here  $\text{Error}(\mathbf{F}, \mathbf{R}) \equiv \text{com}(\mathbf{F}, \mathbf{R}) - \text{com}^{\text{ef}}(\mathbf{F}, \mathbf{R})$  is the computation error,  $\text{com}(\mathbf{F}, \mathbf{R})$  and  $\text{com}^{\text{ef}}(\mathbf{F}, \mathbf{R})$  are respectively results (means of an observable) in the actual noisy computation and error-free computation, and  $(\mathbf{F}, \mathbf{R})$  specifies a quantum circuit. This loss function characterizes errors in a set of circuits with the same frame operations  $\mathbf{F}$  as shown in Fig. 1(a): Frame operations include the qubit initialization, measurement and multi-qubit entangling gates (e.g. controlled-NOT and controlled-phase gates), which are usually error-prone compared with single-qubit gates. We focus on the case that entangling gates are all Clifford. Single-qubit gates denoted by  $\mathbf{R}$  are different in the set of circuits.  $\mathbb{R}$  is the set of single-qubit gate configurations. When  $\mathbb{R} = \mathbb{U}$ , single-qubit gates can be any unitary transformations, and the summation should be taken as integration with respect to Haar measure; when  $\mathbb{R} = \mathbb{C}$ , single-qubit gates are all Clifford. Taking single-qubit gates as variables is a natural way to construct ansatz circuits in variational quantum algorithms [3, 47, 48]. Loss functions in this form can be used to determine parameters in the learning-based quantum error mitigation [46, 49, 50].

In this paper, we demonstrate that the quadratic error loss  $L_{\mathbb{U}}(\mathbf{F})$  (i.e.  $\mathbb{R} = \mathbb{U}$ ) is a good objective function and can be efficiently evaluated when the circuit is large. By sampling random circuits, we study the statistics of  $\text{Error}(\mathbf{F}, \mathbf{R})$  in experiments on a quantum device with

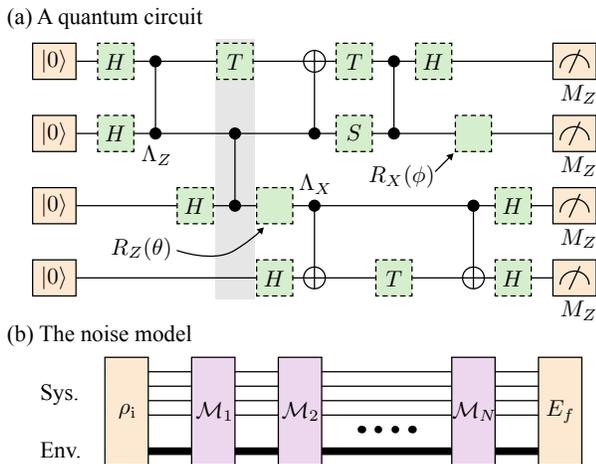


FIG. 1. (a) A quantum circuit. The circuit  $(\mathbf{F}, \mathbf{R})$  consists of frame operations  $\mathbf{F}$  and single-qubit unitary gates  $\mathbf{R}$ . Single-qubit gates are the green dashed squares, and all other operations are frame operations. In the random circuit sampling for evaluating  $L_{\mathbb{R}}(\mathbf{F})$ , frame operations are always the same, but we alter single-qubit gates to different configurations. (b) The noise model. The system (Sys.) formed by qubits and the environment (Env.) are initialized in the state  $\rho_i$ . Following the initialization, there is a sequence of completely-positive maps applied. The map  $\mathcal{M}_\tau$  describes the evolution applied at the time  $\tau$  to realize a layer (column) of quantum gates, which actually acts on the system and environment because of imperfections. Finally, we implement the measurement on each qubit. The operator of observable is  $E_f$ , which acts on both the system and environment because of imperfections.

four superconducting qubits and numerical simulations with up to ten qubits using various error models. We find that the error distribution is approximately Gaussian with zero mean when general unitary circuits are uniformly sampled from  $\mathbf{R} \in \mathbb{U}$  according to Haar measure, i.e.  $L_{\mathbb{U}}(\mathbf{F})$  is the only value that we need for characterizing the error statistics [51]. Computing the error-free result  $\text{com}^{\text{ef}}(\mathbf{F}, \mathbf{R})$  is impractical for large general unitary circuits but efficient for Clifford circuits, according to the Gottesman-Knill theorem [1, 52]. We prove  $L_{\mathbb{U}}(\mathbf{F}) = L_{\mathbb{C}}(\mathbf{F})$ , i.e. we can obtain  $L_{\mathbb{U}}(\mathbf{F})$  by only sampling Clifford circuits, under the assumption that errors in single-qubit gates are gate-independent. Note that we do not need any assumptions on frame-operation errors. The equivalence between error losses of unitary sampling and Clifford sampling is verified in both experiments and numerical simulations.

## II. FIDELITY LOSS AND HYBRID SAMPLING

In addition to the quadratic error loss, the final-state fidelity loss  $E_{\mathbb{U}}(\mathbf{F})$  can also be efficiently evaluated using Clifford sampling. The fidelity loss reads (see Ap-

pendix C)

$$E_{\mathbb{R}}(\mathbf{F}) \equiv \frac{1}{|\mathbb{R}|} \sum_{\mathbf{R} \in \mathbb{R}} [1 - \text{Fidelity}(\mathbf{F}, \mathbf{R})]. \quad (2)$$

The fidelity loss measures the overall quality of final states, compared with the quadratic error loss defined for specific computation tasks (observables).

In the fully-Clifford sampling, we assume single-qubit-gate errors are gate-independent. The weak gate dependence can be accounted for by hybridizing Clifford circuits with a few general unitary single-qubit gates. We remark that Clifford-dominated circuits can be efficiently simulated using classical computer [53]. The fidelity loss and hybrid sampling are studied analytically and numerically in Appendix C and D. In the following, we focus on the quadratic error loss and fully-Clifford sampling.

## III. FORMALISM AND CLIFFORD SAMPLING

In a quantum circuit, we can draw gates applied in parallel in the same layer (column), see Fig. 1(a). For example, the gray box is the fourth layer, which contains a  $T$  gate and a controlled-phase gate. When gates are error-free, the overall map of the fourth layer is  $\mathcal{M}_4^{\text{ef}} = [T \otimes \Lambda_Z \otimes I]$ , where  $[U](\rho) = U\rho U^\dagger$ , and  $I$  is the identity operator of a qubit. The gates are realized through time evolution. Because of the noise, the actual time evolution leads to a different map  $\mathcal{M}_4$ , which acts on not only qubits but also the environment. This is a general formalism of errors in the quantum computation, including both spatial and temporal correlations. The temporal correlation is caused by the environment. According to this formalism, we can express the actual computation result with error as  $\text{com}(\mathbf{F}, \mathbf{R}) = \text{Tr}[E_f \mathcal{M}_N \cdots \mathcal{M}_2 \mathcal{M}_1(\rho_i)]$  for an  $N$ -layer circuit [see Fig. 1(b)]. Here,  $\rho_i$ ,  $E_f$  and  $\mathcal{M}_\tau$  depend on  $\mathbf{F}$  and  $\mathbf{R}$  (Temporal correlations can cause the dependence). Qubits are measured in the computational basis, and the outcome is a binary vector  $\boldsymbol{\mu}$ . The corresponding measurement operator is  $E_{\boldsymbol{\mu}}$ . We consider the case that the computation result is the mean of a real function  $f(\boldsymbol{\mu})$ , then  $E_f = \sum_{\boldsymbol{\mu}} f(\boldsymbol{\mu}) E_{\boldsymbol{\mu}}$ .

We can express the error-free map  $\mathcal{M}_\tau^{\text{ef}}$  as a product of frame gates  $\mathcal{G}_\tau^{\text{ef}}$  and single-qubit gates  $\mathcal{R}_\tau^{\text{ef}}$ , i.e.  $\mathcal{M}_\tau^{\text{ef}} = \mathcal{G}_\tau^{\text{ef}} \mathcal{R}_\tau^{\text{ef}}$ . For example, we have  $\mathcal{G}_4^{\text{ef}} = [I \otimes \Lambda_Z \otimes I]$  and  $\mathcal{R}_4^{\text{ef}} = [T \otimes I^{\otimes 3}]$ . The actual map can always be expressed in the form  $\mathcal{M}_\tau = \mathcal{J}_\tau(\mathcal{R}_\tau^{\text{ef}} \otimes [\mathbb{1}_E]) \mathcal{K}_\tau$ . Here,  $\mathbb{1}_E$  is the identity operator of the environment, and  $\mathcal{J}_\tau$  and  $\mathcal{K}_\tau$  are maps on both the system and environment.  $\mathcal{M}_\tau$  in this form is a linear map for matrix entries of  $\mathcal{R}_\tau^{\text{ef}}$ . Therefore, we have the tensor form of the quantum computation  $\text{com}(\mathbf{F}, \mathbf{R}) = \text{Tr}[(\bar{\mathbf{R}} \otimes \bar{\mathbf{R}}^*) F]$ , where  $\bar{\mathbf{R}}$  is the tensor product of error-free single-qubit gates [e.g.  $\bar{\mathbf{R}} = H^{\otimes 3} \otimes T \otimes R_Z(\theta) \otimes \cdots$  in Fig. 1(a), in which gates are listed from top to bottom then left to right], and  $F$  is a tensor describing the effect of frame operations (see Appendix A). Errors are single-qubit-gate-independent (i.e.  $\mathbf{R}$ -independent) if  $\rho_i$ ,  $E_f$ ,  $\mathcal{J}_\tau$  and  $\mathcal{K}_\tau$

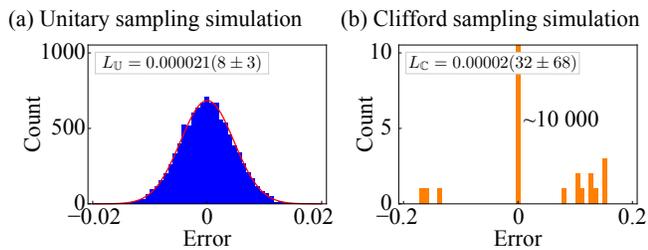


FIG. 2. Numerical results. (a) The unitary sampling of  $\text{Error}(\mathbf{F}, \mathbf{R})$ . The red curve denotes the Gaussian distribution  $\mathcal{N}(0, L_U)$ . (b) The Clifford sampling of  $\text{Error}(\mathbf{F}, \mathbf{R})$ . A ten-qubit circuit with a hundred two-qubit gates is used to compute the mean of a Pauli operator.

(for all  $\tau$ ) are constants. Then  $\text{Error}(\mathbf{F}, \mathbf{R})^2$  is a homogeneous polynomial of degree 2 in both matrix elements of single-qubit gates and their Hermitian conjugates. The Clifford group is a unitary 2-design [14, 54, 55], and therefore  $L_U(\mathbf{F}) = L_C(\mathbf{F})$ . We remark that, not only the second- but also the first- and third-order moments of the error distribution in unitary sampling can be evaluated using the Clifford sampling, because the Clifford group is also a 1-design and 3-design [56, 57].

In the Clifford sampling, we uniformly sample each single-qubit gate in the circuit from the Clifford group. We compute the error loss using the Monte Carlo summation method. There are two approaches. In the mean-value approach, we run each random circuit for multiple times on the actual quantum computer and in the simulation on a classical computer to estimate  $\text{com}(\mathbf{F}, \mathbf{R})$  and  $\text{com}^{\text{ef}}(\mathbf{F}, \mathbf{R})$ , respectively. Then, we can compute the error loss directly according to its definition. This approach is used in our experiments and numerical simulations. In the single-run approach, each random circuit only runs for once or twice. Then, the variance (due to finite sampling) of the  $L_C(\mathbf{F})$  estimator is upper bounded by  $4\|E_f\|^4/N_s$ , and  $4N_s$  circuit runs are implemented on both quantum and classical computers (see Appendix B). Our method is scalable since the variance is independent of the circuit depth and the number of qubits.

#### IV. NUMERICAL RESULTS

We implement numerical simulations of quantum circuits using the library QuESTlink [58, 59] for various error models, frame operation configurations and up to ten qubits. Error models include the depolarizing, dephasing, amplitude damping, correlated coherent, gate-dependent depolarizing, composite and experimentally-measured models. Here, we only show the results of ten qubits with the depolarizing model and a specific frame operation configuration. See Appendix E for details and results of other error models.

In Fig. 2(a), we plot the distribution of  $\text{Error}(\mathbf{F}, \mathbf{R})$  in the unitary sampling, and we can find that the distribu-

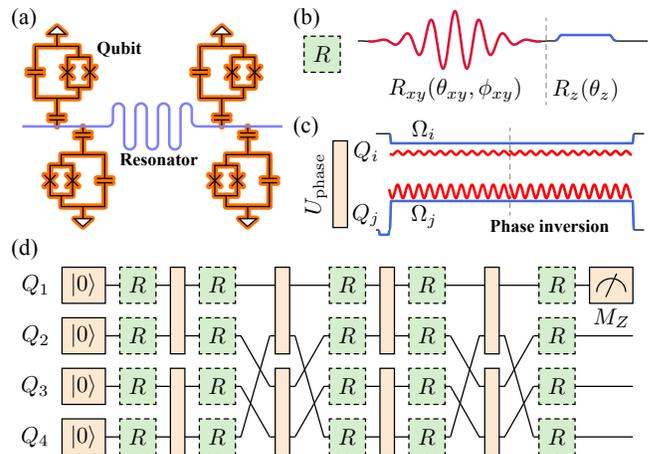


FIG. 3. Experiment setup. (a) Diagram of the superconducting qubit device. We use four Xmon qubits in the experiment. All of them are coupled to a central bus resonator. (b) A single-qubit gate is realized using a  $R_{xy}$  gate followed by a  $R_z$  gate. The  $R_{xy}$  gate has two parameters  $\theta_{xy}$  and  $\phi_{xy}$ , which are controlled by the amplitude and phase of XY pulse (red). The pulse length is 40 ns. The  $R_z$  gate has only one parameter  $\theta_z$ , which is controlled by the amplitude of the Z pulse (blue), whose length is 10 ns. (c) A two-qubit gate  $U_{\text{phase}}$  is generated by tuning two qubits into near resonance (long square Z pulse in blue) with sinusoidal microwave driving field (red) being applied on each qubit dynamically. A short Z pulse before the long Z pulse is applied on one of the qubits to align the  $x$ -axes of their Bloch spheres. We note that the driving field ( $\Omega_i$  on the qubit  $Q_i$ ) is different in  $U_{\text{phase}}$  gates on different qubits. The length of  $U_{\text{phase}}$  is around 300 ns. Phases of driving fields are inverted at the middle of the gate. (d) The quantum circuit used to demonstrate the Clifford sampling, in which we fix two-qubit gates and left single-qubit gates as variables.

tion is approximately Gaussian. This conclusion holds in all our numerical simulations and experiments, for various circuit sizes, error models and frame operation configurations. See Appendix E for a comparison between moments of  $\text{Error}(\mathbf{F}, \mathbf{R})$  and the Gaussian distribution.

The error distribution is non-Gaussian in the Clifford sampling, as shown in Fig. 2(b). In all simple error models (i.e. depolarizing, dephasing, amplitude damping, and coherent models), the distribution is discretized and concentrated at several values of the error, and most of the probability is concentrated at zero. We can understand this result as follows [46]. For Clifford circuits, if the observable to be measured  $E_f$  is a Pauli operator as in our case,  $\text{com}^{\text{ef}}(\mathbf{F}, \mathbf{R})$  takes three values 0 or  $\pm 1$ . For most of the cases,  $\text{com}^{\text{ef}}(\mathbf{F}, \mathbf{R}) = 0$ , and we always have  $\text{com}(\mathbf{F}, \mathbf{R}) = 0$  if errors are Pauli, i.e.  $\text{Error}(\mathbf{F}, \mathbf{R}) = 0$ . Therefore, for Pauli and Pauli-like errors, many Clifford circuits are error-insensitive. We can improve the efficiency of evaluating  $L_C(\mathbf{F})$  using the importance sampling by selecting error-sensitive Clifford circuits.

## V. EXPERIMENTAL RESULTS

To demonstrate the feasibility and usefulness of Clifford sampling in an actual quantum computer, we implement it on a superconducting quantum device, which is illustrated in Fig. 3(a). Four frequency-tunable Xmon qubits ( $Q_1 \sim Q_4$ ) are coupled to a central bus resonator, which mediates the effective interaction between qubits for implementing two-qubit gates. For every single-qubit gate  $R \in U(2)$ , we can decompose it into two experimentally feasible gates  $R = e^{i\alpha} R_z(\theta_z) R_{xy}(\theta_{xy}, \phi_{xy})$ , as shown in Fig. 3(b), where  $\alpha$ ,  $\theta_z$ ,  $\theta_{xy}$  and  $\phi_{xy}$  are real numbers. For two-qubit gates, we use the Clifford dressed-state gate  $U_{\text{phase}}$ , which is essentially the controlled-phase gate but in the  $X$  basis [see Fig. 3(c)] [60]. We can implement the gate  $U_{\text{phase}}$  between any pair of qubits, therefore we have six gate setups for four qubits. See Appendix F for device parameters and detailed implementation of gates.

Before applying Clifford sampling, we benchmarked fidelities of single-qubit gates and two-qubit gates. Fidelities of six  $U_{\text{phase}}$  setups measured using QPT are 95%  $\sim$  97%. In some circuits, two  $U_{\text{phase}}$  are applied in parallel. Because of crosstalk, gate fidelities are changed slightly in parallel operations. We use randomized benchmarking to measure gate fidelities of  $R_{xy}$  and  $R_z$ , which is implemented on each qubit individually as well as simultaneously on all qubits. Both approaches yield no less than 99.3% fidelities for  $X$ ,  $Y$ ,  $X/2$  and  $Z$  gates. The average error rate of single-qubit gates is at least an order of magnitude lower than two-qubit gates, thus we can safely infer that most of the noise is introduced by  $U_{\text{phase}}$ . The gate performance can be improved by optimization based on Clifford sampling. See Appendix F for benchmarking and optimization data.

We use the circuit in Fig. 3(d) as an example to implement the Clifford sampling. The observable to be measured is the probability of  $Q_1$  being in  $|0\rangle$ , i.e.  $E_f = |0\rangle\langle 0|_1 = (I_1 + Z_1)/2$ . Given a specific circuit  $(\mathbf{F}, \mathbf{R})$ , we run the circuit for 1000 times in order to estimate the probability in  $|0\rangle$ . The probability obtained in the experiment is  $P_0^{\text{exp}}$ , and its error-free value computed using the classical computer is  $P_0^{\text{ef}}$ . We note that  $P_0^{\text{exp}}$  has been corrected for readout errors (see Appendix F). Then, the computation error is  $\text{Error} = P_0^{\text{exp}} - P_0^{\text{ef}}$ .

Both unitary sampling and Clifford sampling are implemented in the experiment. For each case, 20000 random configurations of single-qubit gates  $\mathbf{R}$  are generated. In the unitary sampling, the error distribution is Gaussian as shown in Fig. 4(a), the same as in numerical simulations. However, in the Clifford sampling, the error distribution is continuous as shown in Fig. 4(b), which is obviously different from the numerical results of simple error models. In Appendix E, we give numerical results of a composite error model (a combination of coherent and amplitude damping errors) and the experimentally-measured model (from QPT). The error distribution in Clifford sampling for these two models are

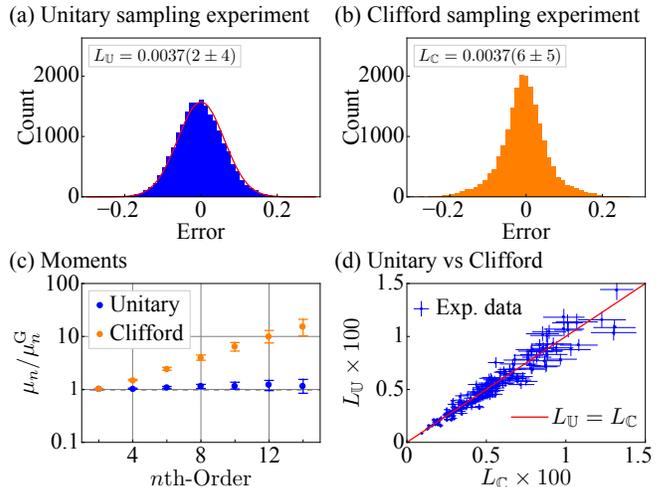


FIG. 4. Experimental results. (a) The unitary sampling of  $\text{Error}(\mathbf{F}, \mathbf{R})$ . The red curve denotes the Gaussian distribution  $\mathcal{N}(0, L_U)$ . (b) The Clifford sampling of  $\text{Error}(\mathbf{F}, \mathbf{R})$ . (c) Moments of two sampling approaches.  $\mu_n = \text{E}[\text{Error}^n]$  is the  $n$ th-order moment, and  $\mu_n^G$  is the moment of  $\mathcal{N}(0, L_U)$ . The quantum circuit used to generate data in (a), (b) and (c) is in Fig. 3(d). (d) Error losses of unitary sampling versus Clifford sampling for 50 randomly generated frame operation configurations with one to ten layers of two-qubit gates. Each layer has one or two  $U_{\text{phase}}$  gates. When only one gate is applied, we protect the other two qubits from dephasing by applying dynamically drive fields [60]. The observable is  $|0\rangle\langle 0|$  for one of four qubits. See Appendix F for details. The error bar denotes the standard deviation due to finite sampling.

continuous and in qualitative agreement with the experimental result. We plot moments up to the 14th-order in Fig. 4(c): Moments of the unitary sampling are consistent with the Gaussian distribution, and moments of the Clifford sampling converge more slowly than Gaussian. Although two distributions are different, their 2nd-order moments, i.e. the loss function values  $L_U = 0.0037(2 \pm 4)$  and  $L_C = 0.0037(6 \pm 5)$ , are the same up to the sampling noise.

In addition to the circuit in Fig. 3(d), we implemented the experiment for 50 randomly generated frame operation configurations  $\mathbf{F}$ . The error loss ( $L_U$  or  $L_C$ ) is estimated by sampling 500 single-qubit gate configurations  $\mathbf{R}$  for each frame operation configuration. The result is plotted in Fig. 4(d). Almost all data points are within  $2\sigma$  from the diagonal line, which represents  $L_U = L_C$ .  $\mathbf{R}$ -dependent errors can cause the difference between two error losses (which is observed in numerical simulations of the gate-dependent error model), however, this effect is not significant in the experimental result.

## VI. DISCUSSION

We propose to characterize quantum circuits executed on a noisy device by evaluating the quadratic error loss

and fidelity loss using the Clifford sampling method. In these two loss functions, all the temporal and spatial correlations are automatically taken into account by treating the entire circuit as a whole. We demonstrate the Clifford sampling method with both numerical simulations and experiments on a superconducting device. We prove that fully-Clifford sampling is sufficient as long as the noise is independent of single-qubit gates. Weak gate dependence can be tackled using hybrid sampling. Experimental results do not show the significant effect of gate dependence. We observe a continuous distribution of the computation error in Clifford sampling in the experiment, whereas some simple error models such as Pauli error models predict a discretized distribution. This result suggests that these models cannot correctly describe the noise in our experiment. One can verify an error model and determine its parameters by constructing loss functions to compare the actual device with the error model.

In addition to characterizing quantum circuits, our method can find application in optimizing their performance. We experimentally implemented the optimization of a Rabi frequency driving two-qubit gates in a set of four-qubit four-depth circuits. The error losses decrease by more than 10% by using Clifford sampling (see Appendix F). In the single-run approach, the sampling cost does not scale with the circuit size. Therefore, our method is promising in the multi-parameter optimization for large-scale quantum circuits. Other than optimizing parameters, our method can provide ground for choosing circuits. The circuit for a computation task may not be unique. Given a noisy quantum device, one can select a working circuit among theoretically equivalent circuits based on our loss functions. A similar idea was proposed in Ref. [50]. Compared to assessing the general performance of the device, our scheme is more application-oriented, i.e. each characterization experiment reflects the likelihood that the device performs well in solving a particular problem. Consequently, the optimization based on our characterization is tailored for specific problems and corresponding circuits.

## ACKNOWLEDGMENTS

We thank Wuxing Liu and Xu Zhang for discussions and technical support. YC thanks Tzu-Chieh Wei and Jason P. Kestner for insightful discussions. We are grateful to the authors of QuESTlink for making their package public. We acknowledge the support of the National Natural Science Foundation of China (No. 11725419 and No. 11875050), the National Key Research and Development Program of China (Grants No. 2017YFA0304300, No. 2019YFA0308100), the Zhejiang Province Key Research and Development Program (Grant No. 2020C01019) and the Basic Research Funding of Zhejiang University. YC is supported by the National Science Foundation (Grant No. PHY 1915165) and BNL LDRD #19-002. DYQ and

YL are also supported by NSAF (Grant No. U1930403).

## Appendix A: Tensor representation of quantum circuits

Let  $\mathbf{R} = (R_1, R_2, \dots, R_{N_R})$  be the list of single-qubit gates, where  $R_i \in \text{U}(2)$  is a single-qubit unitary gate. The  $i$ th single-qubit gate is applied on the  $l_i$ th qubit in the  $t_i$ th layer. The overall map of error-free single-qubit gates in the  $\tau$ th layer is  $\mathcal{R}_\tau^{\text{ef}} = [R_{\tau,1}] \otimes [R_{\tau,2}] \otimes \dots \otimes [R_{\tau,n}]$ , where  $n$  is the number of qubits, and  $R_{\tau,m}$  is the single-qubit gate on the  $m$ th qubit in the  $\tau$ th column, which is either one of  $R_i$ 's in  $\mathbf{R}$  or identity. We have  $R_{\tau,m} = \prod_{i=1}^{N_R} R_i^{\delta_{\tau,t_i} \delta_{m,l_i}}$ . The tensor product of all single-qubit gates is  $\bar{\mathbf{R}} = R_1 \otimes R_2 \otimes \dots \otimes R_{N_R}$ .

We express the quantum computation result as

$$\text{com}(\mathbf{F}, \mathbf{R}) = \text{Tr}[E_f \mathcal{M}_N \dots \mathcal{M}_2 \mathcal{M}_1(\rho_i)]. \quad (\text{A1})$$

We can always write the  $\tau$ th-column map as  $\mathcal{M}_\tau = \mathcal{J}_\tau (\mathcal{R}_\tau^{\text{ef}} \otimes [\mathbb{1}_E]) \mathcal{K}_\tau$ . Because  $\mathcal{R}_\tau^{\text{ef}}$  is invertible, we can take  $\mathcal{K}_\tau = \mathcal{I}$ , which is the identity map, and  $\mathcal{J}_\tau = \mathcal{M}_\tau ((\mathcal{R}_\tau^{\text{ef}})^{-1} \otimes [\mathbb{1}_E])$ . Then, we have

$$\begin{aligned} & \text{com}(\mathbf{F}; [R_1], [R_2], \dots, [R_{N_R}]) \\ &= \text{Tr}[E_f \mathcal{J}_N \mathcal{R}_N^{\text{ef}} \mathcal{K}_N \dots \mathcal{J}_2 \mathcal{R}_2^{\text{ef}} \mathcal{K}_2 \mathcal{J}_1 \mathcal{R}_1^{\text{ef}} \mathcal{K}_1(\rho_i)], \end{aligned} \quad (\text{A2})$$

in which we have replaced  $\mathbf{R}$  with  $([R_1], [R_2], \dots, [R_{N_R}])$ , and they are equivalent.

A single-qubit map can be decomposed as

$$[R_i] = \sum_{a,b,c,d=0,1} R_{i;a,b} R_{i;c,d}^* \mathcal{B}_{a,b}^{c,d}, \quad (\text{A3})$$

where  $R_i = \sum_{a,b=0,1} R_{i;a,b} |a\rangle\langle b|$  is a single-qubit unitary operator, and  $\mathcal{B}_{a,b}^{c,d}(\bullet) = |a\rangle\langle b| \bullet |d\rangle\langle c|$  is the natural basis of single-qubit maps. Because of the linearity of  $\text{com}(\mathbf{F}; [R_1], [R_2], \dots, [R_{N_R}])$ , we have

$$\begin{aligned} & \text{com}(\mathbf{F}; [R_1], [R_2], \dots, [R_{N_R}]) \\ &= \sum_{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}} \left( \prod_{i=1}^{N_R} R_{i;a_i, b_i} R_{i;c_i, d_i}^* \right) F_{\mathbf{a}, \mathbf{b}}^{\mathbf{c}, \mathbf{d}}, \end{aligned} \quad (\text{A4})$$

where

$$F_{\mathbf{a}, \mathbf{b}}^{\mathbf{c}, \mathbf{d}} = \text{com}(\mathbf{F}; \mathcal{B}_{a_1, b_1}^{c_1, d_1}, \mathcal{B}_{a_2, b_2}^{c_2, d_2}, \dots, \mathcal{B}_{a_{N_R}, b_{N_R}}^{c_{N_R}, d_{N_R}}). \quad (\text{A5})$$

Here,  $\mathbf{a} = (a_1, a_2, \dots, a_{N_R})$ ,  $\mathbf{b} = (b_1, b_2, \dots, b_{N_R})$ ,  $\mathbf{c} = (c_1, c_2, \dots, c_{N_R})$  and  $\mathbf{d} = (d_1, d_2, \dots, d_{N_R})$  are binary vectors. Taking

$$F = \sum_{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}} F_{\mathbf{a}, \mathbf{b}}^{\mathbf{c}, \mathbf{d}} |\mathbf{b}, \mathbf{d}\rangle\langle \mathbf{a}, \mathbf{c}|, \quad (\text{A6})$$

we have

$$\text{com}(\mathbf{F}; [R_1], [R_2], \dots, [R_{N_R}]) = \text{Tr}[(\bar{\mathbf{R}} \otimes \bar{\mathbf{R}}^*) F], \quad (\text{A7})$$

where

$$\bar{\mathbf{R}} \otimes \bar{\mathbf{R}}^* = \sum_{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}} \left( \prod_{i=1}^{N_R} R_{i;a_i, b_i} R_{i;c_i, d_i}^* \right) |\mathbf{a}, \mathbf{c}\rangle\langle \mathbf{b}, \mathbf{d}|. \quad (\text{A8})$$

## Appendix B: Monte Carlo method

Let  $f$  be the measurement outcome of the quantum circuit specified by  $(\mathbf{F}, \mathbf{R})$ , and its distribution is  $\text{Pro}(f|\mathbf{F}, \mathbf{R})$ . Then, the computing result, i.e. the mean value of  $f$ , reads

$$\text{com}(\mathbf{F}, \mathbf{R}) = \sum_f \text{Pro}(f|\mathbf{F}, \mathbf{R})f. \quad (\text{B1})$$

Similarly, the error-free computing result can be expressed as

$$\text{com}^{\text{ef}}(\mathbf{F}, \mathbf{R}) = \sum_f \text{Pro}^{\text{ef}}(f|\mathbf{F}, \mathbf{R})f. \quad (\text{B2})$$

In the Monte Carlo summation, the distribution  $\text{Pro}(f|\mathbf{F}, \mathbf{R})$  is realized using the actual quantum computer, and all other distributions, including  $\text{Pro}^{\text{ef}}(f|\mathbf{F}, \mathbf{R})$ , are realized on the classical computer.

We express the loss function as

$$\begin{aligned} L_{\text{C}}(\mathbf{F}) = \frac{1}{|\mathbb{C}|} \sum_{\mathbf{R}, f, f'} & [\text{Pro}(f|\mathbf{F}, \mathbf{R})\text{Pro}(f'|\mathbf{F}, \mathbf{R})ff' \\ & - 2\text{Pro}(f|\mathbf{F}, \mathbf{R})\text{Pro}^{\text{ef}}(f'|\mathbf{F}, \mathbf{R})ff' \\ & + \text{Pro}^{\text{ef}}(f|\mathbf{F}, \mathbf{R})\text{Pro}^{\text{ef}}(f'|\mathbf{F}, \mathbf{R})ff']. \end{aligned} \quad (\text{B3})$$

To compute the first term, we generate  $N_s$  independent and identically distributed samples  $\{(\mathbf{R}_i, f_i, f'_i)|i = 1, 2, \dots, N_s\}$  according to the distribution

$$\text{Pro}(\mathbf{R})\text{Pro}(f|\mathbf{F}, \mathbf{R})\text{Pro}(f'|\mathbf{F}, \mathbf{R}),$$

where

$$\text{Pro}(\mathbf{R}) = \frac{1}{|\mathbb{C}|}. \quad (\text{B4})$$

The estimator of the first term is

$$\hat{L}_1 = \frac{1}{N_s} \sum_{i=1}^{N_s} f_i f'_i, \quad (\text{B5})$$

where  $f_i$  and  $f'_i$  are two independent experimental outcomes obtained for the  $i$ -th sampled circuit. The variance of the estimator is

$$\text{Var}(\hat{L}_1) = \frac{1}{N_s} \text{Var}(ff'). \quad (\text{B6})$$

Let  $|f|_{\text{max}}$  be the maximum value of  $|f(\boldsymbol{\mu})|$ , we have  $ff' \leq |f|_{\text{max}}^2$ . Therefore,

$$\text{Var}(\hat{L}_1) \leq \frac{1}{N_s} |f|_{\text{max}}^4. \quad (\text{B7})$$

To compute the second term, we generate  $2N_s$  independent and identically distributed samples  $\{(\mathbf{R}_i, f_i, f'_i)|i = 1, 2, \dots, 2N_s\}$  according to the distribution

$$\text{Pro}(\mathbf{R})\text{Pro}(f|\mathbf{F}, \mathbf{R})\text{Pro}^{\text{ef}}(f'|\mathbf{F}, \mathbf{R}).$$

The estimator of the second term is

$$\hat{L}_2 = \frac{1}{2N_s} \sum_{i=1}^{2N_s} f_i f'_i, \quad (\text{B8})$$

where  $f_i$  and  $f'_i$  are one experimental outcome and one simulated outcome for the  $i$ -th sampled circuit respectively. The variance of the estimator is

$$\text{Var}(\hat{L}_2) = \frac{1}{2N_s} \text{Var}(ff') \leq \frac{1}{2N_s} |f|_{\text{max}}^4. \quad (\text{B9})$$

To compute the third term, we generate  $N_s$  independent and identically distributed samples  $\{(\mathbf{R}_i, f_i, f'_i)|i = 1, 2, \dots, N_s\}$  according to the distribution

$$\text{Pro}(\mathbf{R})\text{Pro}^{\text{ef}}(f|\mathbf{F}, \mathbf{R})\text{Pro}^{\text{ef}}(f'|\mathbf{F}, \mathbf{R}).$$

The estimator of the third term is

$$\hat{L}_3 = \frac{1}{N_s} \sum_{i=1}^{N_s} f_i f'_i, \quad (\text{B10})$$

where  $f_i$  and  $f'_i$  are two independent simulated outcomes obtained for the  $i$ -th sampled circuit. The variance of the estimator is

$$\text{Var}(\hat{L}_3) = \frac{1}{N_s} \text{Var}(ff') \leq \frac{1}{N_s} |f|_{\text{max}}^4. \quad (\text{B11})$$

The estimator of the error loss is

$$\hat{L}_{\text{C}} = \hat{L}_1 - 2\hat{L}_2 + \hat{L}_3. \quad (\text{B12})$$

The variance of the estimator is

$$\begin{aligned} \text{Var}(\hat{L}_{\text{C}}) &= \text{Var}(\hat{L}_1) + 4\text{Var}(\hat{L}_2) + \text{Var}(\hat{L}_3) \\ &\leq \frac{4}{N_s} |f|_{\text{max}}^4. \end{aligned} \quad (\text{B13})$$

## Appendix C: Fidelity loss

The fidelity loss function is

$$E_{\mathbb{R}}(\mathbf{F}) \equiv \frac{1}{|\mathbb{R}|} \sum_{\mathbf{R} \in \mathbb{R}} [1 - \text{Fidelity}(\mathbf{F}, \mathbf{R})], \quad (\text{C1})$$

where for  $\mathbb{R} = \mathbb{U}$  the summation is understood as integration with respect to Haar measure. Given the final state of the quantum circuit  $\rho(\mathbf{F}, \mathbf{R})$  and the error-free final state  $\rho^{\text{ef}}(\mathbf{F}, \mathbf{R}) = |\Psi(\mathbf{F}, \mathbf{R})\rangle\langle\Psi(\mathbf{F}, \mathbf{R})|$ , the fidelity is

$$\text{Fidelity}(\mathbf{F}, \mathbf{R}) = \text{Tr}[\rho^{\text{ef}}(\mathbf{F}, \mathbf{R})\rho(\mathbf{F}, \mathbf{R})]. \quad (\text{C2})$$

Since we have

$$\begin{aligned} \rho(\mathbf{F}, \mathbf{R}) &= \text{Tr}_{\text{E}}[\mathcal{M}_N \cdots \mathcal{M}_2 \mathcal{M}_1(\rho_i)] \\ &= \text{Tr}_{\text{E}}[\mathcal{J}_N \mathcal{R}_N^{\text{ef}} \mathcal{K}_N \cdots \mathcal{J}_2 \mathcal{R}_2^{\text{ef}} \mathcal{K}_2 \mathcal{J}_1 \mathcal{R}_1^{\text{ef}} \mathcal{K}_1(\rho_i)], \end{aligned} \quad (\text{C3})$$

$\rho(\mathbf{F}, \mathbf{R})$  is a linear map for each  $R_{i;a_i, b_i} R_{i;c_i, d_i}^*$ . It is the same for  $\rho^{\text{ef}}(\mathbf{F}, \mathbf{R})$ . When the noise is independent of  $\mathbf{R}$ ,

Fidelity( $\mathbf{F}, \mathbf{R}$ ) is  $\text{Hom}(2, 2)$ , where we adopt the notation of homogeneous polynomials from Ref. [55]. Therefore,  $E_{\mathbb{U}}(\mathbf{F}) = E_{\mathbb{C}}(\mathbf{F})$  and Clifford sampling is sufficient to produce the result for unitary sampling. Compared with the fidelity loss proposed in Ref. [46], which has a  $\text{Hom}(4, 4)$  term, the application of  $E_{\mathbb{R}}(\mathbf{F})$  in the learning-based quantum error mitigation may have problem, because the error-mitigated state  $\rho$  may not be positive semi-definite.

When the circuit is Clifford, i.e.  $\mathbf{R} \in \mathbb{C}$ , the final state  $|\Psi(\mathbf{F}, \mathbf{R})\rangle$  is a stabiliser state [52]. Suppose  $S_{\mathbf{F}, \mathbf{R}}$  is the stabiliser group of the state  $|\Psi(\mathbf{F}, \mathbf{R})\rangle$ , we have (see Ref. [46])

$$\text{Fidelity}(\mathbf{F}, \mathbf{R}) = \frac{1}{2^n} \sum_{g \in S_{\mathbf{F}, \mathbf{R}}} \text{Tr}[g\rho(\mathbf{F}, \mathbf{R})]. \quad (\text{C4})$$

By measuring the group elements  $g$ , which are Pauli operators with  $\pm$  signs, we can evaluate the fidelity and then the loss function using the Monte Carlo method. In the practical implementation the measurement error may contribute to the result, which is discussed below.

#### Measurement error in fidelity loss

To evaluate the fidelity, we need to measure the group elements  $g$ , which is realized by using an additional layer of single-qubit gates  $\mathcal{M}_{N+1} = \mathcal{J}_{N+1}(\mathcal{R}_{N+1;g}^{\text{ef}} \otimes [\mathbb{1}_{\mathbb{E}}])\mathcal{K}_{N+1}$  to change the effective measurement basis and then measuring all qubits in the  $Z$  basis. Here,  $\mathcal{R}_{N+1;g}^{\text{ef}}$  depends on the element  $g$  to be measured, and  $\mathcal{J}_{N+1}$  and  $\mathcal{K}_{N+1}$  are constants when errors are  $\mathbf{R}$ -independent. We consider the case that single-qubit gates are error-free, i.e.  $\mathcal{M}_{N+1} = \mathcal{R}_{N+1;g}^{\text{ef}} \otimes [\mathbb{1}_{\mathbb{E}}]$ .

For uncorrelated and balanced measurement errors, the measurement outcome is incorrect with a probability  $p$ , which is the same for both measurement outcome 0 and 1, and the event of measurement error is uncorrelated with other operations. Such measurement errors can be expressed as bit-flip errors occurring before the measurement with the probability  $p$ . Now, we introduce the single-qubit depolarising map

$$\mathcal{N}_1(\epsilon) = (1 - \epsilon)[I] + \frac{\epsilon}{3}([X] + [Y] + [Z]). \quad (\text{C5})$$

Because phase-flip errors do not change measurement outcomes in the  $Z$  basis, the uncorrelated and balanced measurement errors with the probability  $p$  is equivalent to applying  $\mathcal{N}_1(3p/2)$  before the measurement. Let  $p_1, p_2, \dots, p_N$  be measurement error rates of  $n$  qubits, the overall measurement-error map is  $\mathcal{N}_{\mathbb{M}} = \mathcal{N}_1(3p_1/2) \otimes \mathcal{N}_1(3p_2/2) \otimes \dots \otimes \mathcal{N}_1(3p_n/2) \otimes [\mathbb{1}_{\mathbb{E}}]$ . Then, the mean of  $g$  measured in the experiment is actually

$$\langle g \rangle_{\text{actual}} = \text{Tr}[g_Z \mathcal{N}_{\mathbb{M}} \mathcal{R}_{N+1;g}^{\text{ef}}(\rho(\mathbf{F}, \mathbf{R}))]. \quad (\text{C6})$$

Here,  $g_Z = \mathcal{R}_{N+1;g}^{\text{ef}}(g)$  is a tensor product of  $Z$  operators, which is directly measured at the end of the circuit.

The single-qubit depolarising map commutes with single-qubit unitary maps, then

$$\begin{aligned} \langle g \rangle_{\text{actual}} &= \text{Tr}[g_Z \mathcal{R}_{N+1;g}^{\text{ef}} \mathcal{N}_{\mathbb{M}}(\rho(\mathbf{F}, \mathbf{R}))] \\ &= \text{Tr}[g \mathcal{N}_{\mathbb{M}}(\rho(\mathbf{F}, \mathbf{R}))]. \end{aligned} \quad (\text{C7})$$

Therefore, the Clifford sampling measures the fidelity in the state  $\mathcal{N}_{\mathbb{M}}(\rho(\mathbf{F}, \mathbf{R}))$ , which includes the effect of measurement errors. We remark that the conditions are i) measurement errors are uncorrelated and balanced, and ii) single-qubit-gate errors are negligible.

#### Appendix D: Hybrid sampling

When the noise in the circuit depends on single-qubit gates  $\mathbf{R}$ , Clifford sampling may become insufficient for evaluating the loss functions  $L_{\mathbb{U}}$  and  $E_{\mathbb{U}}$ . We show that the hybrid sampling method provides an estimator that can tolerate weak  $\mathbf{R}$ -dependence. Let  $F^{\text{ef}}$  be the error-free frame-operation tensor, then

$$L_{\mathbb{R}} = \frac{1}{|\mathbb{R}|} \sum_{\mathbf{R} \in \mathbb{R}} \text{Tr}[(\bar{\mathbf{R}} \otimes \bar{\mathbf{R}}^*)(F - F^{\text{ef}})]^2. \quad (\text{D1})$$

For circuits with  $N_R$  single-qubit gates, we can expand  $F$  as

$$F = F^{(0)} + \sum_{i=1}^{N_R} F_i^{(1)}(R_i) + \Delta F, \quad (\text{D2})$$

where  $F^{(0)}$  is a constant,  $F_i^{(1)}(R_i)$  only depends on the  $i$ th single-qubit gate  $R_i$ , and  $\Delta F$  is small when the gate-dependence is weak. The expansion of error loss is

$$L_{\mathbb{R}} = L_{\mathbb{R}}^{(0)} + 2 \sum_{i=1}^{N_R} L_{\mathbb{R};i}^{(1)} + O(\|F_i^{(1)}\|^2, \|\Delta F\|), \quad (\text{D3})$$

where

$$L_{\mathbb{R}}^{(0)} = \frac{1}{|\mathbb{R}|} \sum_{\mathbf{R} \in \mathbb{R}} W^{(0)}, \quad (\text{D4})$$

and

$$L_{\mathbb{R};i}^{(1)} = \frac{1}{|\mathbb{R}|} \sum_{\mathbf{R} \in \mathbb{R}} W_i^{(1)}. \quad (\text{D5})$$

Here,

$$W^{(0)} = \text{Tr}[(\bar{\mathbf{R}} \otimes \bar{\mathbf{R}}^*)(F^{(0)} - F^{\text{ef}})]^2 \quad (\text{D6})$$

is  $\text{Hom}(2, 2)$  for all  $R_j$ , and

$$\begin{aligned} W_i^{(1)} &= \text{Tr}[(\bar{\mathbf{R}} \otimes \bar{\mathbf{R}}^*)(F^{(0)} - F^{\text{ef}})] \\ &\quad \times \text{Tr}[(\bar{\mathbf{R}} \otimes \bar{\mathbf{R}}^*)F_i^{(1)}(R_i)] \end{aligned} \quad (\text{D7})$$

is  $\text{Hom}(2, 2)$  for all  $R_{j \neq i}$  but not for  $R_i$ . We denote the hybrid sampling set by  $\mathbb{H}_i \equiv \{R_{j \neq i} \in \mathbb{C}(2), R_i \in \mathbb{U}(2)\}$ . Therefore, we have  $L_{\mathbb{U}}^{(0)} = L_{\mathbb{C}}^{(0)} = L_{\mathbb{H}_i}^{(0)}$ ,  $L_{\mathbb{U};i}^{(1)} = L_{\mathbb{H}_i;i}^{(1)}$  and  $L_{\mathbb{H}_j;i \neq j}^{(1)} = L_{\mathbb{C};i}^{(1)}$ .

For the hybrid sampling  $\mathbb{H}_j$ , we have

$$L_{\mathbb{H}_j} = L_{\mathbb{C}}^{(0)} + 2 \left( L_{\mathbb{U};j}^{(1)} + \sum_{i \neq j} L_{\mathbb{C};i}^{(1)} \right) + O(\|F_i^{(1)}\|^2, \|\Delta F\|). \quad (\text{D8})$$

Then, the average of hybrid sampling is

$$\begin{aligned} L_{\text{hybrid}} &\equiv \frac{1}{N_R} \sum_{i=1}^{N_R} L_{\mathbb{H}_i} \\ &= L_{\mathbb{C}}^{(0)} + 2 \frac{1}{N_R} \sum_{i=1}^{N_R} \left[ (N_R - 1) L_{\mathbb{C};i}^{(1)} + L_{\mathbb{U};i}^{(1)} \right] \\ &\quad + O(\|F_i^{(1)}\|^2, \|\Delta F\|). \end{aligned} \quad (\text{D9})$$

Note that the Clifford sampling gives

$$L_{\mathbb{C}} = L_{\mathbb{C}}^{(0)} + 2 \sum_{i=1}^{N_R} L_{\mathbb{C};i}^{(1)} + O(\|F_i^{(1)}\|^2, \|\Delta F\|). \quad (\text{D10})$$

We finally obtain the combined estimator of the error loss

$$\begin{aligned} L_{\text{combined}} &\equiv N_R L_{\text{hybrid}} - (N_R - 1) L_{\mathbb{C}} \\ &= L_{\mathbb{C}}^{(0)} + 2 \sum_{i=1}^{N_R} L_{\mathbb{U};i}^{(1)} + O(\|F_i^{(1)}\|^2, \|\Delta F\|) \\ &= L_{\mathbb{U}}^{(0)} + 2 \sum_{i=1}^{N_R} L_{\mathbb{U};i}^{(1)} + O(\|F_i^{(1)}\|^2, \|\Delta F\|) \\ &= L_{\mathbb{U}} + O(\|F_i^{(1)}\|^2, \|\Delta F\|). \end{aligned} \quad (\text{D11})$$

This strategy can be readily generalized to higher orders, where stronger and higher-order correlated gate dependence can be tolerated by including more unitary single-qubit gates in the sampled circuits. In the discussion above we focus on the quadratic error loss function, but the same logic applies to the fidelity loss function as well.

### Simulation of hybrid circuits

To compute  $L_{\mathbb{H}_j}$ , we need to simulate error-free circuits with one non-Clifford single-qubit gate on a classical computer, which is efficient as discussed in Ref. [53]. A straightforward approach is to decompose a general unitary gate as a linear combination of ten linearly-independent Clifford gates [11], i.e.  $[R] = \sum_{k=1}^{10} \alpha_k [B_k]$ , where  $B_k$  is a Clifford gate. Then, the error-free final state is a linear combination of ten stabiliser states, i.e.  $\rho^{\text{ef}} = \sum_{k=1}^{10} \alpha_k \rho_k^{\text{ef}}$ . Here,  $\rho_k^{\text{ef}}$  is the final state of the Clifford circuit in which  $R$  is replaced with  $B_k$ , and this circuit can be efficiently simulated using the classical computer. For the fidelity loss, we need to measure stabiliser operators of these ten stabiliser states in order to evaluate the fidelity in the non-stabiliser state  $\rho^{\text{ef}}$ .

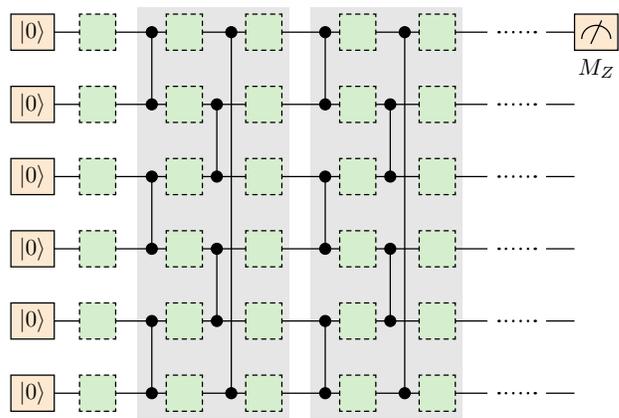


FIG. S1. Circuit used in the numerical simulation. For a circuit of  $n$  qubits, a layer of single-qubit gates contains one single-qubit gate on each qubit, and a layer of two-qubit gates has  $n/2$  controlled-phase gates ( $n$  is even). These  $n$  qubits form a ring, i.e. the  $n$ th qubit and the first qubit are coupled. Controlled-phase gates are applied on nearest-neighboring qubits. After the qubit initialization, a layer of single-qubit gates is applied. Then the circuit in the gray box is repeated for  $n$  times, i.e. the total number of controlled-phase gates is  $n^2$ . After the  $n$ th gray box, the first qubit is measured in the  $Z$  basis.

### Appendix E: Numerical simulation

Three categories of circuits are simulated using QuESTlink [58, 59]. The first category includes circuits shown in Fig. S1, and we call them standard circuits. The second category are randomly generated circuits, in which the observable is a tensor product of  $Z$  operators on randomly selected qubits. The third category is the four-qubit circuit used in the experiment, see Fig. 3(d) in the main text.

For the depolarizing, dephasing, amplitude damping, correlated coherent and gate-dependent depolarizing error models, we implement the simulation for the standard and randomly generated circuits with the qubit number four, six and ten. For each error model and qubit number, we take two different error rates. Given the error model, qubit number and error rate, we generate the standard circuit and three random circuits. Therefore, the total number of circuits is  $5 \times 3 \times 2 \times 4 = 120$ . In the paper, we only show results of the standard circuit with 10 qubits and one error rate for each error model. The complete data and codes for generating them are available at <https://github.com/yzchen-phy/clifford-sampling>.

For the composite error model and the experimentally-measured error model, we only implement the simulation for the four-qubit circuit used in the experiment. The hybrid sampling is demonstrated using the four-qubit standard circuit.

In all the simulations, we assume qubit initialization and measurement are error-free. In the gate-dependent depolarizing error model (which is used in the hybrid

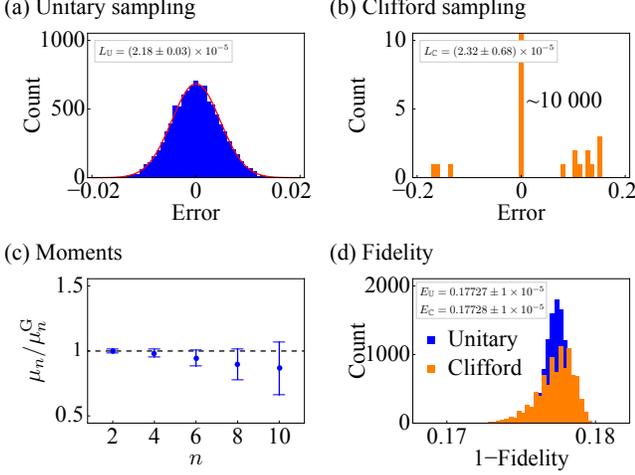


FIG. S2. Numerical results of the depolarizing error model.

sampling simulation), single-qubit gates are noisy. However, in all other error models, we assume single-qubit gates are also error-free.

### 1. Depolarizing model

We introduce the noise by adding the two-qubit depolarizing map after each two-qubit gate. The two-qubit depolarizing map reads

$$\mathcal{N}_2(\epsilon) = (1 - \epsilon)[I] + \frac{\epsilon}{15} \sum_{\sigma \in \{I, X, Y, Z\}^{\otimes 2} \setminus I^{\otimes 2}} [\sigma]. \quad (\text{E1})$$

The numerical results of the ten-qubit standard circuit with the error rate  $\epsilon = 0.002$  are shown in Fig. S2 (and Fig. 2 in the main text). Numbers of single-qubit-gate configurations are both 10000 in unitary sampling and Clifford sampling.

### 2. Dephasing model

We introduce the noise by adding the two-qubit dephasing map after each two-qubit gate. The two-qubit dephasing map reads

$$\mathcal{D}_2(\epsilon) = (1 - \epsilon)[I] + \frac{\epsilon}{3} \sum_{\sigma \in \{I, Z\}^{\otimes 2} \setminus I^{\otimes 2}} [\sigma]. \quad (\text{E2})$$

The numerical results of the ten-qubit standard circuit with the error rate  $\epsilon = 0.002$  are shown in Fig. S3. Numbers of single-qubit-gate configurations are both 10000 in unitary sampling and Clifford sampling.

### 3. Amplitude damping model

After each two-qubit gate, we introduce the noise by adding the one-qubit amplitude damping map on each

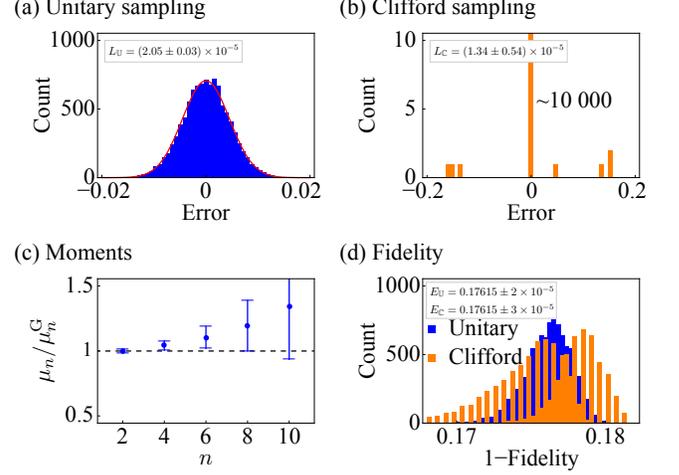


FIG. S3. Numerical results of the dephasing error model.

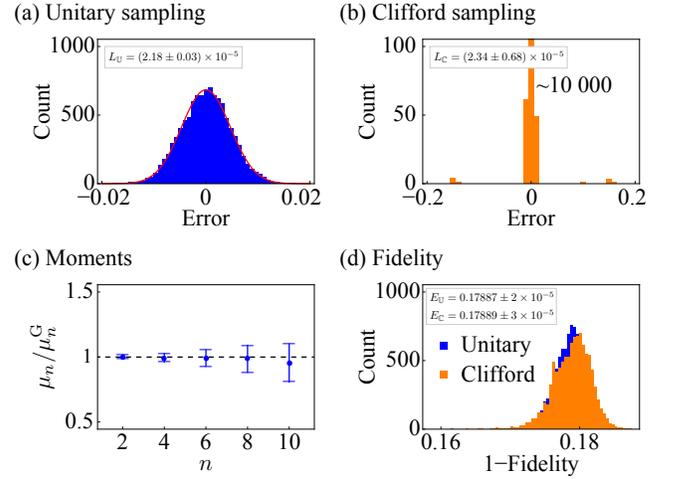


FIG. S4. Numerical results of the amplitude damping error model.

qubit. The one-qubit amplitude damping map reads

$$\mathcal{A}_1(\epsilon) = \left[ \frac{I + Z}{2} + \sqrt{1 - \epsilon} \frac{I - Z}{2} \right] + \left[ \sqrt{\epsilon} \frac{X + iY}{2} \right]. \quad (\text{E3})$$

The numerical results of the ten-qubit standard circuit with the error rate  $\epsilon = 0.002$  are shown in Fig. S4. Numbers of single-qubit-gate configurations are both 10000 in unitary sampling and Clifford sampling.

### 4. Correlated coherent model

After each two-qubit gate, we introduce the noise by adding the one-qubit coherent-error map  $[e^{\pm i\pi\epsilon Z}]$  on each qubit. In one run of the circuit, the sign is the same

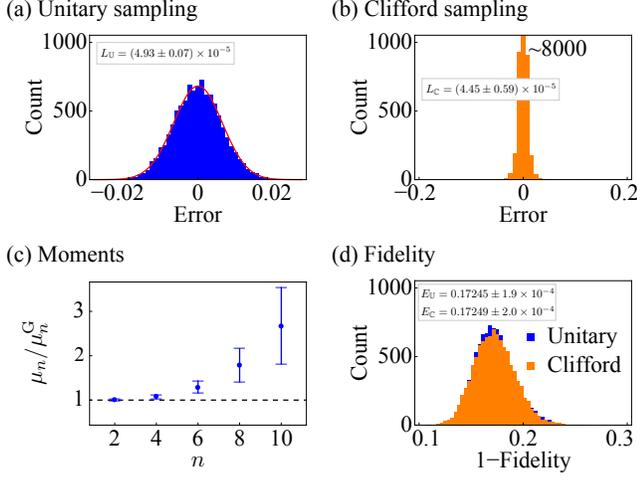


FIG. S5. Numerical results of the correlated coherent error model.

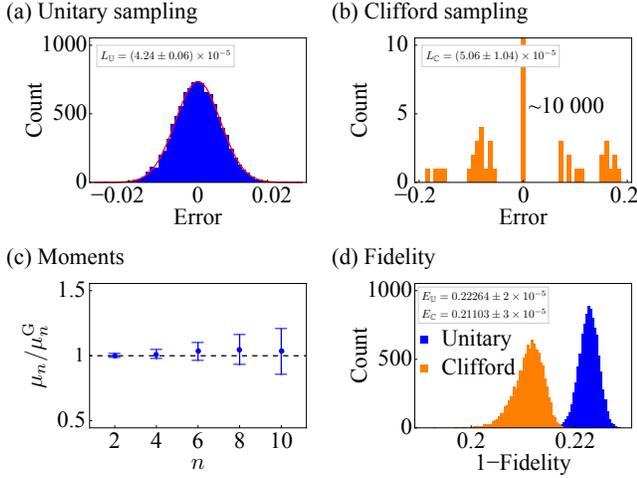


FIG. S6. Numerical results of the gate-dependent depolarizing error model.

in all coherent-error maps, which is  $+$  or  $-$  with the probability of  $1/2$ . The numerical results of the ten-qubit standard circuit with the error rate  $\epsilon = 0.01$  are shown in Fig. S5. Numbers of single-qubit-gate configurations are both 10000 in unitary sampling and Clifford sampling. We can find that the high-order moments of the error distribution slowly deviate from those of the Gaussian distribution.

## 5. Gate-dependent depolarizing model

The gate-dependent error model is based on the depolarizing error model. In addition to two-qubit depolarizing maps, we also introduce single-qubit depolarizing maps after each single-qubit gate. We decom-

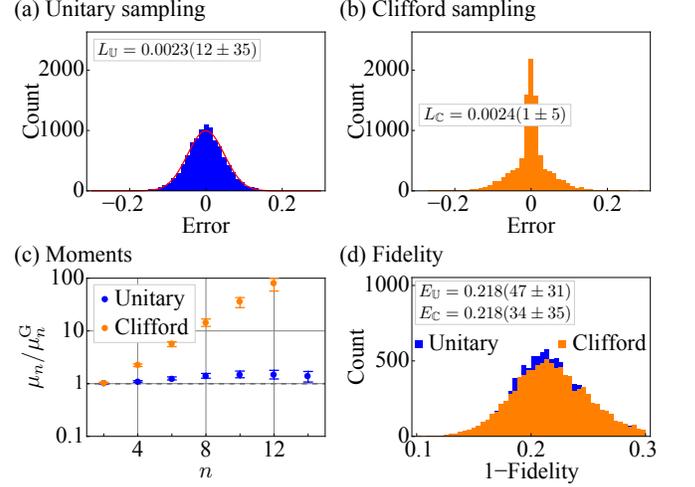


FIG. S7. Numerical results of the composite error model.

pose each single-qubit gate as  $R = e^{-i\frac{\theta_3}{2}Z} e^{-i\frac{\theta_2}{2}X} e^{-i\frac{\theta_1}{2}Z}$ . The single-qubit depolarizing map after the gate  $R$  is  $\mathcal{N}_1(\gamma(\theta_1 + \theta_2 + \theta_3)/2\pi)$ . The numerical results of the ten-qubit standard circuit with error rates  $\epsilon = 0.002$  (two-qubit error rate) and  $\gamma = \epsilon/10$  are shown in Fig. S6. Numbers of single-qubit-gate configurations are both 10000 in unitary sampling and Clifford sampling.

## 6. Composite model

The composite error model consists of coherent single-qubit rotations and the amplitude damping map. After each two-qubit gate we introduce the single-qubit map  $[e^{-i\pi\epsilon_X X}][e^{-i\pi\epsilon_Z Z}]$ , followed by a single-qubit amplitude damping  $\mathcal{A}_1(\epsilon_D)$ . We set  $\epsilon_D = 0.02$  in the entire circuit while the values of  $\epsilon_X$  and  $\epsilon_Z$  are different for each two-qubit gate and drawn from the uniform distribution between 0 and 0.04. The noise parameters for the two qubits in a two-qubit gate are the same. The numerical results of the four-qubit experimental circuit are shown in Fig. S7. Numbers of single-qubit-gate configurations are both 10000 in unitary sampling and Clifford sampling.

## 7. Experimentally-measured model

We simulate the four-qubit experimental circuit with the two-qubit gates replaced by the two-qubit maps obtained in the quantum process tomography (QPT). Other operations are assumed to be error-free. We note that the maps from QPT are not exactly trace-preserving and completely positive, as a result of sampling noise and state preparation and measurement error. Because QuESTlink validates maps, we prepared our own code to implement the numerical simulation without requiring the trace-preserving and completely positive condition. The numerical results are shown in Fig. S8. Num-

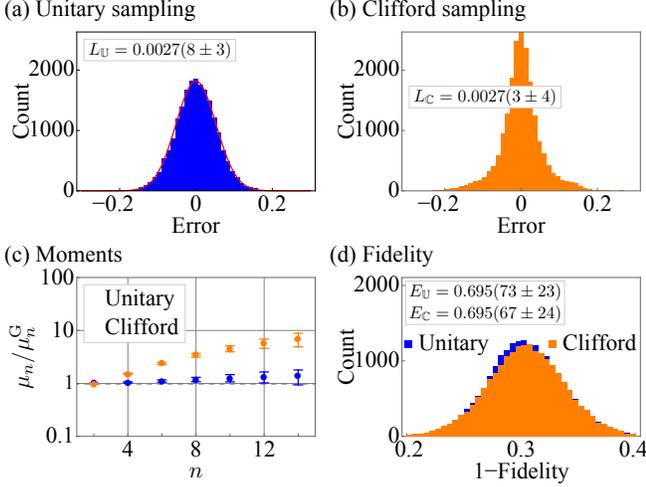


FIG. S8. Numerical results of the experimentally-measured error model.

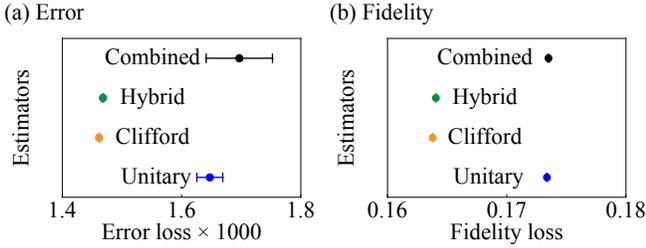


FIG. S9. Numerical results of hybrid sampling.

bers of single-qubit-gate configurations are both 20000 in unitary sampling and Clifford sampling.

## 8. Hybrid sampling

We implement the simulation of the gate-dependent depolarizing model on the four-qubit standard circuit. We take  $\epsilon = 0.01$  (two-qubit error rate) and  $\gamma = \epsilon/10$ . The number of single-qubit-gate configurations in the unitary sampling is 10000. The number of single-qubit-gate configurations in the Clifford sampling is  $10000 \times 2(N_R - 1)^2 = 4500000$ . The number of single-qubit-gate configurations in the sampling of each  $\mathbb{H}_i$  is  $10000 \times 2N_R = 320000$ . Results are shown in Fig. S9. We can find that the combined estimator is closer to the result of unitary sampling than fully-Clifford sampling.

## Appendix F: Experimental details

### 1. Device parameters

The quantum device consists of 20 frequency-tunable Xmon qubits, where four qubits labeled as  $Q_1 \sim Q_4$  are

used in this experiment to illustrate the idea of Clifford sampling. Each qubit has a Z line for tuning the qubit frequency, an XY line for exciting the qubit and a readout resonator coupled to a common readout line for the qubit-state measurement. With regard to connectivity, each qubit is capacitively coupled to the central bus resonator  $B$ , which has a fixed resonance frequency of  $\omega_B/2\pi \approx 5.248$  GHz, with the coupling strength ( $g_j$ ) listed in Tab. S1. In the experiment,  $Q_1 \sim Q_4$  are initialized in the ground state  $|0\rangle$  at their respective idle frequencies  $\omega_j$  as listed in Tab. S1, while all the other qubits are left at their respective maximum frequencies, i.e., the sweetpoints insensitive to flux noises, which are at least 1 GHz higher than the idle frequencies of  $Q_1 \sim Q_4$ . The energy relaxation time ( $T_{1,j}$ ) and the Ramsey dephasing time ( $T_{2,j}^*$ ) are listed in Tab. S1. It was observed that the dressed states of the qubits under coherent microwave fields are less sensitive to external dephasing noises [60]. Therefore the effective dephasing times of these qubits should be much longer than  $T_{2,j}^*$ , as we frequently apply the XY microwave driving fields on these qubits to implement the single-qubit rotations, the single-qubit dynamical decoupling schemes and the two-qubit  $U_{\text{phase}}$  gates within the Clifford sampling circuit.

### 2. Readout correction

The experimental scheme to directly measure the multiqubit occupation probabilities and the subsequent procedure to eliminate the readout errors were detailed in Ref. [61, 62]. To initialize the qubit in its ground state  $|0\rangle$ , we idle it for about 200  $\mu\text{s}$ , during which the residue thermal excitation is estimated to be small via a post-selection procedure. The produced  $|0\rangle$  state of the qubit has a state fidelity above 0.99 on average, following which we can apply a high-fidelity  $X$  gate ( $\pi$  rotation around  $x$ -axis of the Bloch sphere) to reliably prepare the qubit in  $|1\rangle$ .

However, due to the existence of readout errors, the directly measured occupation probability  $P_0^m$  ( $P_1^m$ ) after we reliably prepare  $Q_j$  in  $|0\rangle$  ( $|1\rangle$ ) may still be away from ideal (actual), which is noted as  $Q_j$ 's measurement fidelity in  $|0\rangle$  ( $|1\rangle$ ),  $F_{0,j}$  ( $F_{1,j}$ ). For any experimental measurement, we label the directly measured probability vector as  $\mathbf{P}^m = (P_0^m, P_1^m)^T$  and the actual one as  $\mathbf{P} = (P_0, P_1)^T$ , which map as

$$\mathbf{P}_j = \begin{pmatrix} F_{0,j} & 1 - F_{1,j} \\ 1 - F_{0,j} & F_{1,j} \end{pmatrix}^{-1} \mathbf{P}_j^m = \mathbf{F}_j^{-1} \mathbf{P}_j^m. \quad (\text{F1})$$

To eliminate the readout errors of the directly measured probability column vector  $\mathbf{P}_{Q_1 Q_2 Q_3 Q_4}^m$  for the 4-qubit joint states, we obtain the actual probability column vector by

$$\mathbf{P}_{Q_1 Q_2 Q_3 Q_4} = (\otimes_{j=1}^4 \mathbf{F}_j)^{-1} \mathbf{P}_{Q_1 Q_2 Q_3 Q_4}^m. \quad (\text{F2})$$

Table S1 lists the simultaneously obtained measurement fidelity values  $F_{0,j}$  and  $F_{1,j}$  for all four qubits.

TABLE S1. Quantum device parameters.  $\omega_j$  is the idle frequency where we initialize the state of  $Q_j$ , apply single-qubit gates and perform quantum measurement.  $g_j$  is the coupling strength between  $Q_j$  and  $B$ .  $T_{1,j}$  is the energy relaxation time and  $T_{2,j}^*$  is the Ramsey (Gaussian) depahsing time of  $Q_j$ .  $F_{0,j}$  ( $F_{1,j}$ ) is the readout fidelity for  $Q_j$  in  $|0\rangle$  ( $|1\rangle$ ), which is used to eliminate the readout errors.  $F_{R,j}^{\text{indiv}}$  is the RB gate fidelity of the  $R$  gate, where  $R \in \{X, Y, Z, X/2\}$ , characterized individually for  $Q_j$  at its idle frequency, whereas  $F_{R,j}^{\text{simul}}$  is the simultaneous RB gate fidelity obtained by running RBs characterizing the same  $R$  gate on all four qubits simultaneously [61].

Qubit	$Q_1$	$Q_2$	$Q_3$	$Q_4$
$\omega_j/2\pi$ (GHz)	4.904	4.955	4.999	5.043
$g_j/2\pi$ (MHz)	18.9	17.5	16.1	18.9
$T_{1,j}$ ( $\mu\text{s}$ )	38.7	54.4	34.3	37.4
$T_{2,j}^*$ ( $\mu\text{s}$ )	2.2	2.3	2.2	2.5
$F_{0,j}$	0.957	0.981	0.977	0.973
$F_{1,j}$	0.920	0.922	0.919	0.951
$F_{X,j}^{\text{indiv}}$	0.9983(2)	0.9967(4)	0.9993(1)	0.9991(1)
$F_{X,j}^{\text{simul}}$	0.9986(6)	0.9966(17)	0.9936(15)	0.9995(7)
$F_{Y,j}^{\text{indiv}}$	0.9987(2)	0.9978(3)	0.9993(1)	0.9994(1)
$F_{Y,j}^{\text{simul}}$	0.9984(5)	0.9964(13)	0.9937(9)	0.9993(5)
$F_{Z,j}^{\text{indiv}}$	0.9981(2)	0.9939(6)	0.9979(2)	0.9982(2)
$F_{Z,j}^{\text{simul}}$	0.9982(8)	0.9975(12)	0.9954(11)	0.9984(5)
$F_{X/2,j}^{\text{indiv}}$	0.9989(1)	0.9990(2)	0.9988(1)	0.9990(1)
$F_{X/2,j}^{\text{simul}}$	0.9992(8)	0.9993(11)	0.9987(7)	0.9996(7)

TABLE S2. Two-qubit gate parameters.  $\omega_{I,jj'}$  is the interaction frequency at which we tune  $Q_j$  and  $Q_{j'}$  on resonance to obtain  $U_{\text{phase},jj'}$ .  $\lambda_{jj'}$  is the effective coupling strength between  $Q_j$  and  $Q_{j'}$ . By applying continuous driving fields with driving strengths of  $\Omega_j$  and  $\Omega_{j'}$  on  $Q_j$  and  $Q_{j'}$ , respectively, for a fixed duration of  $T_{\text{gate},jj'}$ , we realize  $U_{\text{phase},jj'}$ . The individual (simultaneous) gate fidelity  $F$  of  $U_{\text{phase},jj'}$  is obtained by QPT on  $Q_j$ - $Q_{j'}$ , while no action ( $U_{\text{phase},kk'}$  in parallel) is applied on the other two qubits  $Q_k$ - $Q_{k'}$  at the idle frequencies ( $\omega_{I,kk'}$ ).

Qubit pair	$Q_1Q_2$	$Q_1Q_3$	$Q_1Q_4$	$Q_2Q_3$	$Q_2Q_4$	$Q_3Q_4$
$\omega_{I,jj'}/2\pi$ (GHz)	4.905	4.905	4.955	4.955	5.043	5.043
$\lambda_{jj'}/2\pi$ (MHz)	0.90	0.95	0.85	0.70	1.1	0.80
$\Omega_{j(j' )}/2\pi$ (MHz)	10.93(2.96)	3.61(12.74)	2.96(9.09)	2.96(13.39)	10.13(4.26)	7.52(12.74)
$T_{\text{gate},jj'}$ (ns)	277	274	297	356	231	326
$F(U_{\text{phase},jj'})$ (Individual)	0.967(2)	0.968(3)	0.968(3)	0.944(3)	0.951(4)	0.960(2)
$F(U_{\text{phase},jj'})$ (Simultaneous)	0.956(2)	0.961(4)	0.941(3)	0.941(3)	0.947(5)	0.962(2)

### 3. Single-qubit gate

Single-qubit gates include unitary single-qubit gates and Clifford single-qubit gates. For a unitary single-qubit gate, we randomly generate a unitary matrix  $M_U \in U(2)$ , which is distributed with Haar measure [63]. For a Clifford single-qubit gate, we randomly choose a matrix  $M_C \in C(2)$ . Here,  $C(2)$  denotes the Clifford group of one qubit. Then we convert the matrix to  $e^{i\alpha}R_{xy}(\theta_{xy}, \phi_{xy})R_z(\theta_z)$ , where  $R_z(\theta_z)$  represents a rotation by  $\theta_z$  around  $z$ -axis and  $R_{xy}(\theta_{xy}, \phi_{xy})$  represents a rotation by  $\theta_{xy}$  around the axis in the equator plane, which has an angle  $\phi_{xy}$  with respect to  $x$ -axis. Here the global phase factor  $e^{i\alpha}$  can be ignored. To characterize the gate performance, we perform individual and simultaneous randomized benchmarkings (RBs) on representative single-qubit gates such as the  $X$  ( $\theta_{xy} = \pi, \phi_{xy} = 0$ ),

$X/2$  ( $\theta_{xy} = \pi/2, \phi_{xy} = 0$ ),  $Y$  ( $\theta_{xy} = \pi, \phi_{xy} = \pi/2$ ) and  $Z$  ( $\theta_z = \pi$ ) gates, yielding gate fidelities no less than 0.993 for all four qubits (see Tab. S1).

### 4. Two-qubit $U_{\text{phase}}$ gate

The  $U_{\text{phase}}$  gate implemented on a pair of qubits was detailed in Ref. [60]. Here we discuss the case of executing parallel  $U_{\text{phase}}$  gates on four qubits. The original Hamiltonian with four qubits is

$$\begin{aligned}
H/\hbar = & \omega_B a^\dagger a + \sum_{j=1}^4 \omega_j |1_j\rangle\langle 1_j| \\
& + \sum_{j=1}^4 [g_j(\sigma_j^+ a + \sigma_j^- a^\dagger)], \quad (\text{F3})
\end{aligned}$$

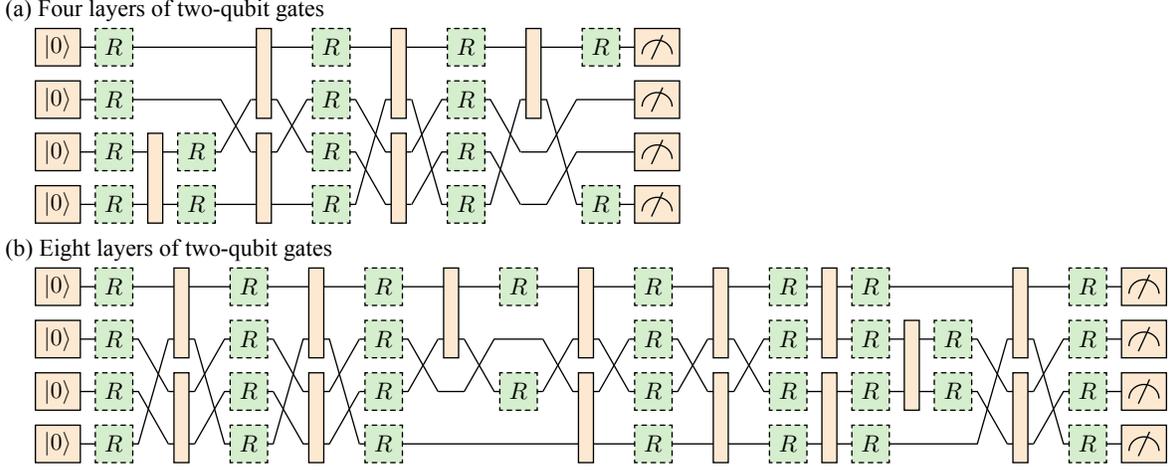


FIG. S10. Exemplary frame operation configurations. (a) Configuration with four layers of two-qubit gates. (b) Configuration with eight layers of two-qubit gates.

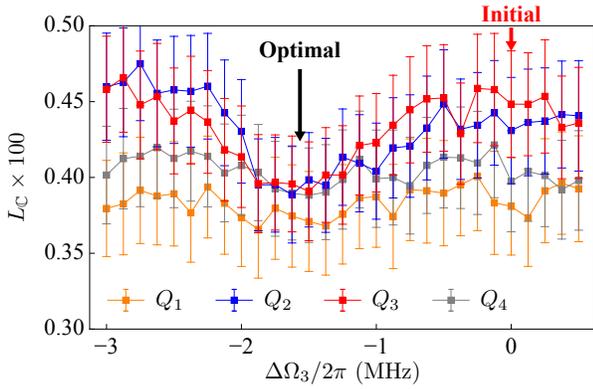


FIG. S11. Parameter optimization of the  $U_{\text{phase}}$  gate on  $Q_2$ - $Q_3$  based on Clifford sampling.

where  $\sigma_j^+$  ( $\sigma_j^-$ ) is the raising (lowering) operator of  $Q_j$  and  $a^\dagger$  ( $a$ ) is the creation (annihilation) operator of  $B$ .

Now we divide the four qubits into two pairs,  $Q_k$ - $Q_{k'}$  and  $Q_l$ - $Q_{l'}$ , and position  $Q_k$ - $Q_{k'}$  ( $Q_l$ - $Q_{l'}$ ) at the interaction frequency  $\omega_{I,kk'}$  ( $\omega_{I,ll'}$ ). Under the assumption that  $\omega_{I,kk'}$  and  $\omega_{I,ll'}$  are largely detuned and  $B$  is in its ground state, we have

$$H_1/\hbar = \sum_{j \in \{k,l\}} \lambda_{jj'} (\sigma_j^- \sigma_{j'}^+ + \sigma_j^+ \sigma_{j'}^-) + \sum_{j \in \{k,l\}} \left( \frac{g_j^2}{\Delta_{jj'}} |1_j\rangle\langle 1_j| + \frac{g_{j'}^2}{\Delta_{jj'}} |1_{j'}\rangle\langle 1_{j'}| \right), \quad (\text{F4})$$

where  $\Delta_{jj'} = \omega_{I,jj'} - \omega_B$  and  $\lambda_{jj'}$  is effective coupling strength between  $Q_j$  and  $Q_{j'}$ . Therefore the intra-pair qubits interact with each other, while the inter-pair qubits are effectively decoupled. By applying appropriate microwave driving fields on all four qubits, the Hamilto-

nian in the double-rotating frames can be written as:

$$H_{\text{eff}}/\hbar = \sum_{j \in \{k,l\}} (\lambda_{jj'} \sigma_j^+ \sigma_{j'}^- + \Omega_j e^{-i\phi_j} \sigma_j^+ + \Omega_{j'} e^{-i\phi_{j'}} \sigma_{j'}^+) + \text{H.c.}, \quad (\text{F5})$$

where  $\Omega_j$  ( $\phi_j$ ) is the Rabi frequency (phase) of the driving field on  $Q_j$ . The Hamiltonian for the two qubits ( $Q_j$ - $Q_{j'}$ ) in a pair takes exactly the form as described in Eq. (4) of Ref. [60], based on which a controlled  $\pi$ -phase gate in the dressed-state basis can be realized by choosing appropriate driving strengths and evolving the system for a fixed amount of time  $T_{\text{gate},jj'}$ , during which the phases of the driving fields are reversed at  $T_{\text{gate},jj'}/2$ . The unitary matrix in the two-qubit computational basis  $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$  describing the evolution is

$$U_{\text{phase}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & i \\ 0 & 1 & i & 0 \\ 0 & i & 1 & 0 \\ i & 0 & 0 & 1 \end{pmatrix}. \quad (\text{F6})$$

The resulting  $U_{\text{phase}}$  gates on various qubit combinations are characterized by QPT and summarized in Tab. S2.

## 5. Randomly generated frame operation configurations

As mentioned in the main text, we randomly generate 50 frame operation configurations to validate  $L_{\mathbb{U}} = L_{\mathbb{C}}$ . Here we plot two examples in Fig. S10: one with four layers and a total of six  $U_{\text{phase}}$  gates [Fig. S10(a)], and the other with eight layers and a total of fourteen  $U_{\text{phase}}$  gates [Fig. S10(b)]. For the four-layer example, there are two  $U_{\text{phase}}$  gates in each of the middle layers and one  $U_{\text{phase}}$  gate in each of the beginning and ending layers. For any layer with only one  $U_{\text{phase}}$  gate, during the  $U_{\text{phase}}$  gate

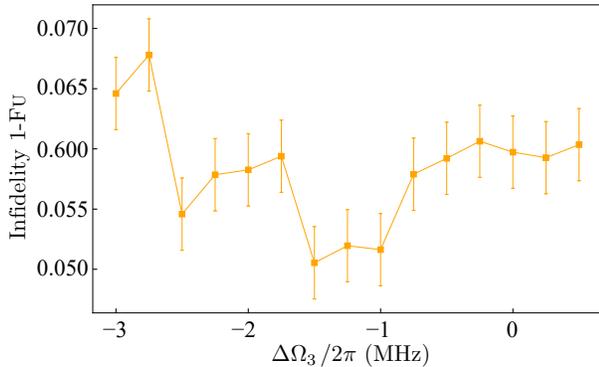


FIG. S12. Infidelity of the  $U_{\text{phase}}$  gate on  $Q_2$ - $Q_3$  characterized by QPT as a function of  $\Delta\Omega_3$ .

on two qubits for a time of about 300 ns (see Tab. S2), we apply continuous microwave fields, with the driving strengths  $\Omega/2\pi \approx 3$  MHz, on the other two qubits which are left at their idle frequencies to protect them from dephasing [60].

## 6. Clifford sampling optimization

Usually, pulse parameters of a gate are predetermined by benchmarking qubits individually. Nonetheless, the gate performance may decline, or the optimal parameters may drift when we implement multiple gates in par-

allel, because of correlations such as crosstalk. The more qubits are involved, the more significant the impact of correlations may become. Clifford sampling provides a convenient and scalable way to optimize parameters in the large-circuit quantum computation. As a demonstration, we detect the optimal parameter of the gate  $U_{\text{phase}}$  on qubits  $Q_2$  and  $Q_3$  in the experiment. An important parameter of  $U_{\text{phase}}$  is the strength of driving field, e.g. Rabi frequency  $\Omega_2$  ( $\Omega_3$ ) applied on  $Q_2$  ( $Q_3$ ). In the experiment, we find the optimal  $\Omega_3$  in the four-qubit circuit shown in Fig. 3(d) in the main text. The observable is  $E_f = |0\rangle\langle 0|_i$ , i.e. the probability in  $|0\rangle$  for one of four qubits. We change  $\Omega_3$  from its initial value  $\Omega_3^{\text{ini}} = 2\pi \times 13.39$  MHz to  $\Omega_3^{\text{ini}} + \Delta\Omega_3$ . For each value of  $\Omega_3$ , we implement 1500 random single-qubit gate configurations  $\mathbf{R}$  to estimate the Clifford error loss  $L_{\mathbb{C}}$ . The result is shown in Fig. S11. We can find that error losses of  $Q_2$  and  $Q_3$  decrease by more than 10% when  $\Delta\Omega_3$  changes from 0 to  $-2\pi \times 1.5$  MHz.

To demonstrate the effectiveness of the  $L_{\mathbb{C}}$  approach, we measure the infidelity of the  $Q_2$ - $Q_3$   $U_{\text{phase}}$  gate,  $1 - F_U$ , as a function of  $\Delta\Omega_3$ , while we apply the  $U_{\text{phase}}$  gates on  $Q_2$ - $Q_3$  and  $Q_1$ - $Q_4$  simultaneously. As shown in Fig. S12, the infidelity approaches the minimum around  $\Delta\Omega_3 = -2\pi \times 1.5$  MHz, agreeing well with the Clifford sampling optimization, proving the effectiveness of optimizing with  $L_{\mathbb{C}}$  as the loss function.

\* Z. W. and Y. C. contributed equally to this work.

† chaosong@zju.edu.cn

‡ yli@gscaep.ac.cn

- 
- [1] I. L. C. Michael A. Nielsen, *Quantum Computation and Quantum Information* (Cambridge University Pr., 2010).
- [2] P. W. Shor, “Algorithms for quantum computation: discrete logarithms and factoring,” *Proceedings 35th Annual Symposium on Foundations of Computer Science (1994)*, 10.1109/sfcs.1994.365700.
- [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] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, “Quantum computational chemistry,” *Rev. Mod. Phys.* **92**, 015003 (2020).
- [5] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, “Randomized benchmarking and process tomography for gate errors in a solid-state qubit,” *Phys. Rev. Lett.* **102**, 090502 (2009).
- [6] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, “Noise spectroscopy through dynamical decoupling with a superconducting flux qubit,” *Nat. Phys.* **7**, 565–570 (2011).
- [7] J. Braumüller, L. Ding, A. P. Vepsäläinen, Y. Sung, M. Kjaergaard, T. Menke, R. Winik, D. Kim, B. M. Niedzielski, A. Melville, J. L. Yoder, C. F. Hirjibehedin, T. P. Orlando, S. Gustavsson, and W. D. Oliver, “Characterizing and optimizing qubit coherence based on squid geometry,” *Phys. Rev. Appl.* **13**, 054079 (2020).
- [8] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O’Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, “Optimal quantum control using randomized benchmarking,” *Phys. Rev. Lett.* **112**, 240504 (2014).
- [9] Y. Li and S. C. Benjamin, “Efficient variational quantum simulator incorporating active error minimization,” *Phys. Rev. X* **7**, 021050 (2017).
- [10] K. Temme, S. Bravyi, and J. M. Gambetta, “Error mitigation for short-depth quantum circuits,” *Phys. Rev. Lett.* **119**, 180509 (2017).
- [11] S. Endo, S. C. Benjamin, and Y. Li, “Practical quantum error mitigation for near-future applications,” *Phys. Rev. X* **8**, 031027 (2018).
- [12] J. Emerson, R. Alicki, and K. Życzkowski, “Scalable noise estimation with random unitary operators,” *J. Opt. B: Quantum Semiclassical Opt.* **7**, S347–S352 (2005).
- [13] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B.

- 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).
- [14] C. Dankert, R. Cleve, J. Emerson, and E. Livine, “Exact and approximate unitary 2-designs and their application to fidelity estimation,” *Phys. Rev. A* **80**, 012304 (2009).
- [15] E. Magesan, J. M. Gambetta, and J. Emerson, “Scalable and robust randomized benchmarking of quantum processes,” *Phys. Rev. Lett.* **106**, 180504 (2011).
- [16] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, “Investigating the limits of randomized benchmarking protocols,” *Phys. Rev. A* **89**, 062321 (2014).
- [17] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, “Robust extraction of tomographic information via randomized benchmarking,” *Phys. Rev. X* **4**, 011050 (2014).
- [18] D. Lu, H. Li, D.-A. Trottier, J. Li, A. Brodutch, A. P. Krismanich, A. Ghavami, G. I. Dmitrienko, G. Long, J. Baugh, and R. Laflamme, “Experimental estimation of average fidelity of a clifford gate on a 7-qubit quantum processor,” *Phys. Rev. Lett.* **114**, 140505 (2015).
- [19] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch, “Recovering quantum gates from few average gate fidelities,” *Phys. Rev. Lett.* **121**, 170502 (2018).
- [20] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, “Direct randomized benchmarking for multiqubit devices,” *Phys. Rev. Lett.* **123**, 030503 (2019).
- [21] D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta, “Three-qubit randomized benchmarking,” *Phys. Rev. Lett.* **122**, 200502 (2019).
- [22] I. L. Chuang and M. A. Nielsen, “Prescription for experimental determination of the dynamics of a quantum black box,” *J. Mod. Opt.* **44**, 2455–2467 (1997).
- [23] J. F. Poyatos, J. I. Cirac, and P. Zoller, “Complete characterization of a quantum process: The two-bit quantum gate,” *Phys. Rev. Lett.* **78**, 390–393 (1997).
- [24] G. M. D’Ariano and P. L. Presti, “Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation,” *Phys. Rev. Lett.* **86**, 4195–4198 (2001).
- [25] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O’Brien, M. A. Nielsen, and A. G. White, “Ancilla-assisted quantum process tomography,” *Phys. Rev. Lett.* **90**, 193601 (2003).
- [26] M. Mohseni and D. A. Lidar, “Direct characterization of quantum dynamics,” *Phys. Rev. Lett.* **97**, 170501 (2006).
- [27] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography,” *Nat. Commun.* **8**, 14485 (2017).
- [28] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, “Characterizing quantum supremacy in near-term devices,” *Nat. Phys.* **14**, 595–600 (2018).
- [29] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, et al., “Quantum supremacy using a programmable superconducting processor,” *Nature* **574**, 505–510 (2019).
- [30] L. C. G. Góvia, G. J. Ribeill, D. Ristè, M. Ware, and H. Krovi, “Bootstrapping quantum process tomography via a perturbative ansatz,” *Nat. Commun.* **11**, 1084 (2020).
- [31] J. Cotler and F. Wilczek, “Quantum overlapping tomography,” *Phys. Rev. Lett.* **124**, 100401 (2020).
- [32] M. R. Geller and M. Sun, “Efficient correction of multi-qubit measurement errors,” (2020), [arXiv:2001.09980v1](https://arxiv.org/abs/2001.09980v1).
- [33] K. E. Hamilton, T. Kharazi, T. Morris, A. J. McCaskey, R. S. Bennink, and R. C. Pooser, “Scalable quantum processor noise characterization,” (2020), [arXiv:2006.01805v1](https://arxiv.org/abs/2006.01805v1).
- [34] J. J. Wallman and S. T. Flammia, “Randomized benchmarking with confidence,” *New J Phys* **16**, 103032 (2014).
- [35] M. A. Fogarty, M. Veldhorst, R. Harper, C. H. Yang, S. D. Bartlett, S. T. Flammia, and A. S. Dzurak, “Non-exponential fidelity decay in randomized benchmarking with low-frequency noise,” *Phys. Rev. A* **92**, 022326 (2015).
- [36] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, “Effect of noise correlations on randomized benchmarking,” *Phys. Rev. A* **93**, 022303 (2016).
- [37] S. Mavadia, C. L. Edmunds, C. Hempel, H. Ball, F. Roy, T. M. Stace, and M. J. Biercuk, “Experimental quantum verification in the presence of temporally correlated noise,” *npj Quantum Inf.* **4**, 7 (2018).
- [38] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, “Probing context-dependent errors in quantum processors,” *Phys. Rev. X* **9**, 021045 (2019).
- [39] A. Veitia and S. J. van Enk, “Testing the context-independence of quantum gates,” (2018), [arXiv:1810.05945v1](https://arxiv.org/abs/1810.05945v1).
- [40] M. Huo and Y. Li, “Self-consistent tomography of temporally correlated errors,” (2018), [arXiv:1811.02734v3](https://arxiv.org/abs/1811.02734v3).
- [41] D. Wecker, B. Bauer, B. K. Clark, M. B. Hastings, and M. Troyer, “Gate-count estimates for performing quantum chemistry on small quantum computers,” *Phys. Rev. A* **90**, 022305 (2014).
- [42] N. Moll, A. Fuhrer, P. Staar, and I. Tavernelli, “Optimizing qubit resources for quantum chemistry simulations in second quantization on a quantum computer,” *J. Phys. A: Math. Theor.* **49**, 295301 (2016).
- [43] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Majller, W. Riess, G. Salis, et al., “Quantum optimization using variational algorithms on near-term quantum devices,” *Quantum Sci. Technol.* **3**, 030503 (2018).
- [44] P.-L. Dallaire-Demers, J. Romero, L. Veis, S. Sim, and A. Aspuru-Guzik, “Low-depth circuit ansatz for preparing correlated fermionic states on a quantum computer,” *Quantum Sci. Technol.* **4**, 045005 (2019).
- [45] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, “Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm,” *npj Quantum Information* **6**, 10 (2020).
- [46] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, “Learning-based quantum error mitigation,” (2020), [arXiv:2005.07601v1](https://arxiv.org/abs/2005.07601v1).
- [47] 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**,

- 242–246 (2017).
- [48] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature* **567**, 209–212 (2019).
- [49] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, “Error mitigation with clifford quantum-circuit data,” (2020), [arXiv:2005.10189v1](https://arxiv.org/abs/2005.10189v1).
- [50] L. Cincio, K. Rudinger, M. Sarovar, and P. J. Coles, “Machine learning of noise-resilient quantum circuits,” (2020), [arXiv:2007.01210v1](https://arxiv.org/abs/2007.01210v1).
- [51] We observe that the mean is approximately zero in all numerical simulations and experiments, in which the observable is a Pauli operator. In general, the mean can be measured using Clifford sampling and Hybrid sampling in the same way as the loss function.
- [52] D. Gottesman, “The heisenberg representation of quantum computers,” (1998), [arXiv:quant-ph/9807006v1](https://arxiv.org/abs/quant-ph/9807006v1).
- [53] S. Bravyi and D. Gosset, “Improved classical simulation of quantum circuits dominated by clifford gates,” *Phys. Rev. Lett.* **116**, 250501 (2016).
- [54] D. Gross, K. Audenaert, and J. Eisert, “Evenly distributed unitaries: On the structure of unitary designs,” *J. Math. Phys.* **48**, 052104 (2007).
- [55] A. Roy and A. J. Scott, “Unitary designs and codes,” *Des. Codes Cryptogr.* **53**, 13–31 (2009).
- [56] Z. Webb, “The clifford group forms a unitary 3-design,” (2015), [arXiv:1510.02769v3](https://arxiv.org/abs/1510.02769v3).
- [57] H. Zhu, “Multiqubit clifford groups are unitary 3-designs,” *Phys. Rev. A* **96**, 062336 (2017).
- [58] T. Jones and S. Benjamin, “QuESTlink—mathematica embiggened by a hardware-optimised quantum emulator,” *Quantum Sci. Technol.* **5**, 034012 (2020).
- [59] T. Jones, A. Brown, I. Bush, and S. C. Benjamin, “QuEST and high performance simulation of quantum computers,” *Sci. Rep.* **9**, 10736 (2019).
- [60] 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).
- [61] C. Song, K. Xu, W. Liu, C.-p. Yang, S.-B. Zheng, H. Deng, Q. Xie, K. Huang, Q. Guo, L. Zhang, et al., “10-qubit entanglement and parallel logic operations with a superconducting circuit,” *Phys. Rev. Lett.* **119**, 180511 (2017).
- [62] C. Song, K. Xu, H. Li, Y.-R. Zhang, X. Zhang, W. Liu, Q. Guo, Z. Wang, W. Ren, J. Hao, H. Feng, H. Fan, D. Zheng, D.-W. Wang, H. Wang, and S.-Y. Zhu, “Generation of multicomponent atomic schrödinger cat states of up to 20 qubits,” *Science* **365**, 574–577 (2019), <https://science.sciencemag.org/content/365/6453/574.full.pdf>.
- [63] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nat. Methods* **17**, 261–272 (2020).