I. INTRODUCTION

How to correct errors is one of the most critical issues in practical quantum computation. In the theory of quantum fault tolerance based on quantum error correction (QEC), an arbitrarily high-fidelity quantum computation can be achieved, providing sub-threshold error rates and sufficient qubits [1]. Recently, error rates within or close to the sub-threshold regime have been demonstrated in various platforms [2–6]. These error rates are measured using either randomized benchmarking (RB) [7, 8] or quantum process tomography (QPT) [9, 10]. RB only estimates an average effect of the noise, and QPT can provide a model of error channels. Rigorously speaking, whether or not a quantum system is in the sub-threshold regime is not only determined by the error rate but also the detailed error model [11, 12], including correlations between errors [13]. Therefore, an error model describing correlated errors is important for verifying sub-threshold quantum devices. We can also optimize QEC protocols by exploring these correlations [11, 14], which is crucial for the early-stage demonstration of small-scale quantum fault tolerance. Given the limited number of qubits and error rate close to the threshold, we need to carefully choose the protocol to observe any advantage of QEC [15–23]. However, neither RB nor conventional QPT can provide an error model describing temporal correlations [6, 24–27]. In this paper, we propose a tomography method to obtain such an error model.

We may still need many years to realize a fault-tolerant quantum computer [28, 29], however noisy intermediate-scale quantum (NISQ) computers are likely to be developed in the near future [30–32]. Quantum error mitigation (QEM) is an alternative approach to high-fidelity quantum computation [33–36], which does not require encoding, therefore, is more promising than QEC on NISQ devices. In QEM, we can increase errors to learn their effect on the observable representing the computation result. Once we know how the observable changes with the level of errors, we can make an extrapolated estimate of the zero-error computation result. This extrapolation can be implemented directly on the final result or each gate using the quasi-probability decomposition formalism. The effect of errors depends on the error model. Therefore, we need to increase errors according to the model of original errors in the system, i.e. at first we need a proper error model of original errors. The error model can be obtained using gate set tomography (GST) [6, 37, 38], a self-consistent QPT protocol, with which the effect of errors on the computation result can be eliminated, under the condition that errors are not correlated [35]. However, correlations are common in quantum systems [39–41], e.g. the slow drift of laser frequency can cause time-dependent gate fidelity in ion trap systems [42–45], which limits the fidelity of quantum computation on NISQ devices. Based on our tomography method, we can mitigate errors with temporal correlations.

In this paper, we show that quantum gates with temporal correlations are fully characterized by a set of linear operators, which however may not be completely positive maps acting on reduced density matrices. These linear operators can be measured in the experiment only using computation operations themselves. Therefore, we term our method as linear operator tomography (LOT). Then we propose a QEM protocol based on the error model obtained using LOT, and we find that temporally correlated errors can be mitigated in quantum computation. A tremendous amount of experimental data may be required to obtain the exact error model. For the practical implementation, we aim at an approximate error model, and we find that efficient approximations for low-frequency time-dependent noise and context-dependent noise exist [46, 47]. Practical protocols are proposed and demonstrated numerically. Error rates estimated using RB and GST may exhibit significant difference due to temporal correlations [6]. In numerical simulations, we show that RB and LOT results coincide with each
other, and QEM can successfully suppress correlated errors. With the increased computation fidelity, we extend the complexity of algorithms that can be implemented on NISQ devices.

II. MODEL OF QUANTUM COMPUTER

To show how to describe errors with temporal correlations, we start with an example in ion trap systems. If the quantum gate is driven by laser fields, usually the gate fidelity depends on frequencies of lasers [42–45]. We use \( \lambda \) to denote laser frequencies, and we use the superoperator \( \mathcal{O}_S(\lambda) \) to denote the gate operation when laser frequencies are \( \lambda \). Because laser frequencies drift with time, \( \lambda \) is a set of random variables, and we use \( p(\lambda) \) to denote the probability distribution of \( \lambda \). Suppose that the change of laser frequencies in the time of a quantum gate is negligible, the superoperator describing one quantum gate reads \( \int d\lambda p(\lambda) \mathcal{O}_S(\lambda) \), and the superoperator describing two sequential quantum gates reads \( \int d\lambda p(\lambda) \mathcal{O}_S(\lambda) \mathcal{O}_S(\lambda) \). Here, we have assumed that the state at the beginning of the gate does not depend on \( \lambda \). Errors are correlated in the sense that the two-gate superoperator cannot be factorized as a product of two one-gate superoperators, i.e. \( \int d\lambda p(\lambda) \mathcal{O}_S(\lambda) \mathcal{O}_S(\lambda) \neq \left[ \int d\lambda p(\lambda) \mathcal{O}_S(\lambda) \right] \left[ \int d\lambda' p(\lambda') \mathcal{O}_S(\lambda') \right] \), in which case the conventional QPT cannot be applied.

To apply tomography techniques, we can factorize multi-gate superoperators by introducing the state space of laser frequencies. We use \( |\lambda\rangle_E \) to denote the state corresponding to laser frequencies \( \lambda \). We remark that the state space of \( |\lambda\rangle_E \) can be virtual, i.e. it is not necessary that \( |\lambda\rangle_E \) corresponds to a pure state in a physical Hilbert space. The state of qubits and laser frequencies can be expressed as \( \rho_{SE} = \int d\lambda \rho_S(\lambda) \otimes |\lambda\rangle_E\langle\lambda| \), where \( \rho_S(\lambda) \) is the state of qubits when laser frequencies are \( \lambda \). The one-gate superoperator can be expressed as \( \mathcal{O}_S = \int d\lambda \mathcal{O}_S(\lambda) \otimes |\lambda\rangle_E\langle\lambda| \), where \( |\lambda\rangle_E \) denotes a superoperator [42]–[45]. Then the two-gate superoperator can be expressed in the factorized form \( \mathcal{O}_{SE} \mathcal{O}_S \).

We note that multi-gate superoperators can be factorized following a similar procedure for noise with any spectrum, i.e. the change of laser frequencies in the time of a quantum gate can be nonnegligible or even significant, as we will show in Sec. V A.

Now, we introduce a general model of quantum computer. For a quantum computer with \( n \) qubits, we call the 2\(^n\)-dimensional Hilbert space of qubits the system. Degrees of freedom coupled to the system form the environment, including but not limited to all random variables determining the evolution of the system. Quantum computation is realized by a sequence of quantum gates. The gate sequence is stored in a terminal, e.g. a classical computer, and the evolution of the system-environment (SE) is controlled by the terminal as shown in Fig. 1. We use \( \chi \) to denote the state of the terminal indicating “Implement the gate \( \chi \)” and superoperator \( \mathcal{O}_{SE}(\chi) \) to denote the corresponding evolution of SE. Here \( \chi \) is a deterministic parameter rather than a random variable. By setting \( \chi \) according to the gate sequence, we realize the quantum computation. We assume that the Born-Markov approximation can be applied to the terminal and SE. For the gate sequence \( \chi_1, \chi_2, \ldots, \chi_n \), the overall evolution of SE is \( \mathcal{O}_{SE}(\chi_N) \cdots \mathcal{O}_{SE}(\chi_2) \mathcal{O}_{SE}(\chi_1) \).

This model can describe errors with temporal correlations, and it is not limited to classical random variables such as laser frequencies. The model can also describe correlations caused by the coupling to a quantum system in the environment and context-dependent errors.

In general, the evolution of the system \( \mathcal{O}_S(\lambda, \rho_{SE}) \) depends on not only \( \chi \) but also the state of SE at the beginning of the evolution \( \rho_{SE} \). If the system and environment are correlated in \( \rho_{SE} \), the evolution may not even be completely positive [48]. If the system evolution \( \mathcal{O}_S(\chi, \rho_{SE}) = \mathcal{O}_S(\chi) \) does not depend on \( \rho_{SE} \), the overall system evolution of a gate sequence is \( \mathcal{O}_S(\chi_N) \cdots \mathcal{O}_S(\chi_2) \mathcal{O}_S(\chi_1) \). In this case, the conventional QPT can be applied, we can obtain \( \mathcal{O}_S(\chi) \) up to a similarity transformation using GST [37, 38], and the computation error can be mitigated as proposed in Ref. [35]. By introducing the environment, non-Markovian processes can be reconstructed using quantum tomography [49].

From now on, we focus on states and operations of SE and neglect the subscript ‘SE’.

State, measurement, operations and Pauli transfer matrix representation

A quantum computer is characterized by a set of linear operators: the initial state \( \rho_{in} \) which is a normalized positive Hermitian operator, the measurement (i.e. measured observable) \( Q_{out} \) which is also a Hermitian operator, and a site of elementary computation operations \( \{ \mathcal{O}(\chi) \} \) which are completely-positive maps. We remark
that, $\rho\text{in}$, $Q_{\text{out}}$ and $\{O(\chi)\}$ describe the actual quantum computer (including both the system and environment) rather than an ideal quantum computer, and they are all unknown therefore need to be investigated in tomography. The quantum computation is realized by a sequence of elementary operations on the initial state. The set of operation sequences $O = \{O(x)\}$ includes all operations generated by elementary operations $\{O(\chi)\}$.

We focus on the case that the quantum computer only provides one option of the initial state $\rho\text{in}$ and one option of the observable to be measured $Q_{\text{out}}$. It is straightforward to generalize our results to the case that multiple options of initial states and observables are available.

In this paper, we use the Pauli transfer matrix representation [37, 38]: $|\rho\rangle$ is a column vector with elements $\langle\rho|\rangle_{\sigma} = \text{Tr}(\sigma\rho)$; $\langle Q\rangle$ is a row vector with elements $\langle Q|\sigma = d_H^{-1}\text{Tr}(\sigma Q)$; then an quantum operation $O$ can be expressed as a square matrix with elements $O_{\sigma,\tau} = d_H^{-1}\text{Tr}[(\sigma O(\tau)]$. Here, $\sigma$ and $\tau$ are Pauli operators or generalized Pauli operators, i.e. Hermitian operators satisfying $\text{Tr}(\sigma\tau) = d_H\delta_{\sigma,\tau}$, and $d_H$ is the dimension of the Hilbert space of SE. All these vectors and matrices are real and $d_H$-dimensional. For any state $\rho$ and observable $Q$, $\langle Q|\rho\rangle = \text{Tr}(Q\rho)$ is the mean of the observable $Q$ in the state $\rho$. For any operation $O$, $|O(\rho)\rangle = \mathcal{O}(\rho)$ is the vector corresponding to the state $O(\rho)$. Therefore, the mean of the observable $Q$ in the state $\rho$ after a sequence of quantum operations reads $\text{Tr}[Q\mathcal{O}_N\cdots\mathcal{O}_2\mathcal{O}_1(\rho)] = \langle Q|O_N\cdots\mathcal{O}_2\mathcal{O}_1|\rho\rangle$.

III. EXACT NOISE PROCESSING

Using our model of the quantum computer, we propose a tomography protocol to obtain a set of linear operators which exactly characterize the behaviour of the quantum computer even in the presence of temporal correlations. In the conventional QPT, we obtain a set of completely-positive maps to characterize quantum operations. However, in our protocol the linear operators may not be completely-positive maps.

Based on the tomography result, we also propose a QEM protocol to completely eliminate the computation error in the presence of temporal correlations. These protocols only use computation operations in the quantum computer, and additional operations are not required. At this stage the feasibility is not our concern. In following sections, we will discuss how to adapt these protocols for the purpose of practical implementation.

A. Exact linear operator tomography

Given the sufficient ability to access both the system and environment, we can treat SE as one system and apply the conventional QPT to reconstruct quantum operations on SE. Such an ability usually requires operations other than computation operations in the quantum computer, e.g. operations that manipulate the state of the environment.

In this section, we show that all necessary information for exactly characterizing computation operations can be obtained only using computation operations themselves. However we may not be able to reconstruct completely-positive maps of computation operations. We obtain a linear operator in a subspace of states for each computation operation.

Although we focus on quantum computation, our result is not limited to computation operations. For any finite set of initial states, observables and operations, their behaviour can be exactly characterized by the information that can be obtained in the tomography only involving themselves, given sufficient data from the experiment.

With only computation operations, we may not be able to access entire spaces of states and observables. Therefore, we consider three subspaces as follows. The subspace $V_{\text{in}} = \text{span}\{|\rho\rangle\mid O \in O\}$ is the span of all states that can be prepared in the quantum computer, and the subspace $V_{\text{out}} = \text{span}\{|Q_{\text{out}}\mid O \in O\}$ is the span of all observables that can be effectively measured. Note that $O$ is the set of all operations generated by elementary computation operations. We use $P_{\text{in}}$ and $P_{\text{out}}$ to denote the orthogonal projections on $V_{\text{in}}$ and $V_{\text{out}}$, respectively. The third subspace is $V = \text{span}\{|P_{\text{out}}\rho\rangle\mid O \in O\}$, and we use $P$ to denote the orthogonal projection on $V$. We remark that $V$ is not a subspace of the Hilbert space.

Our first result is that in order to fully characterize the quantum computer, we only need to reconstruct $P|\rho_{\text{in}}\rangle$, $\langle Q_{\text{out}}|O(\chi)\rangle$ and $\{PO(\chi)P\}$ in the tomography. The reason is that, for any sequence of computation operations in $O$, we have

$$\langle Q_{\text{out}}|O_N\cdots O_2O_1|\rho_{\text{in}}\rangle$$

$$= \langle Q_{\text{out}}|PO_NP\cdots PO_2PO_1P|\rho_{\text{in}}\rangle. \quad (1)$$

See Appendix for the proof. We would like to remark that the conclusion is the same for the subspace $\text{span}\{|Q_{\text{out}}|O_{\text{in}}\mid O \in O\}$.

Now we give the exact LOT protocol:

- Choose a set of states $\{|\rho_i\rangle\mid O_i \in O; i = 1, \ldots , d\}$ and a set of observables $\{Q_k\} = \langle Q_{\text{out}}|O_k\rangle | O_k \in O; k = 1, \ldots , d\}$. We always take $|\rho_1\rangle = |\rho_{\text{in}}\rangle$ and $\langle Q_1\rangle = \langle Q_{\text{out}}\rangle$. These states can be prepared in the quantum computer, and these observables can be effectively measured in the quantum computer.

- We need to assume that the dimension of the subspace $V$ is finite. We also need to assume that these states and observables satisfy the following conditions: $d$ is the dimension of the subspace $V$, $\{P|\rho_i\rangle\}$ are linearly independent, and $\langle Q_k|P\rangle$ are also linearly independent. According to the definition of the subspace $V$, states and observables
satisfying these conditions always exist and can be realized in the quantum computer using computation operations.

- Obtain matrices $g = M_{\text{out}} M_{\text{in}}$ and $\hat{O}(\chi) = M_{\text{out}} \mathcal{O}(\chi) M_{\text{in}}$ for each $\chi$ in the experiment. Here, $M_{\text{in}} = \begin{bmatrix} |\rho_1\rangle \cdots |\rho_d\rangle \end{bmatrix}$ is the matrix with $\{|\rho_i\rangle\}$ as columns, and $M_{\text{out}} = \begin{bmatrix} \langle Q_1 |^T \cdots \langle Q_d |^T \end{bmatrix}^T$ is the matrix with $\{|\langle Q_k |\rangle\}$ as rows. Each matrix element can be measured in the experiment. The element $g_{k,i} = \langle Q_k | \rho_i \rangle$ is the mean of $Q_k$ in the state $\rho_i$. The element $\mathcal{O}_{k,i}(\chi) = \langle Q_k | \hat{O}(\chi) | \rho_i \rangle$ is the mean of $Q_k$ in the state $\rho_i$ after the operation $\hat{O}(\chi)$.

Data $g$ and $\{|\hat{O}(\chi)\rangle\}$ exactly characterize the quantum computer. With states and observables satisfying conditions in the protocol, $g$ is always invertible. Then we have

$$M_{\text{out}} \mathcal{O}_N \cdots \mathcal{O}_2 \mathcal{O}_1 M_{\text{in}} = \tilde{\mathcal{O}}_N g^{-1} \cdots \tilde{\mathcal{O}}_2 g^{-1} \tilde{\mathcal{O}}_1 \quad (2)$$

for any sequence of computation operations in $\mathcal{O}$. See Appendix for the proof.

Given $g$ and $\{|\hat{O}(\chi)\rangle\}$, we can obtain an exact error model of the quantum computer.

- Choose an arbitrary $d$-dimensional invertible real matrix $\tilde{M}_{\text{in}}$, and compute $\tilde{M}_{\text{out}} = g \tilde{M}_{\text{in}}^{-1}$.
- Take $|\tilde{\rho}_{\text{in}}\rangle = \tilde{M}_{\text{in}} |\rho_{\text{in}}\rangle$ and $\langle \tilde{\rho}_{\text{out}} | = \tilde{M}_{\text{out}}^{-1} |\rho_{\text{out}}\rangle$, and compute $\hat{O}(\chi) = \tilde{M}_{\text{out}}^{-1} \tilde{O}(\chi) \tilde{M}_{\text{in}}^{-1}$ for each $\chi$.

Here, $M_{\text{in}}$ and $M_{\text{out}}$ denote the $j$th column and the $k$th row of the matrix $\tilde{M}$, respectively. The error model of the quantum computer is formed by $|\tilde{\rho}_{\text{out}}\rangle$, $\langle \tilde{\rho}_{\text{out}} |$ and $\{|\tilde{O}(\chi)\rangle\}$, which correspond to $|\rho\rangle$, $\langle \rho |$ and $\{|\mathcal{O}(\chi)\rangle\}$, respectively.

According to Eq. (2), we have

$$\langle \tilde{\rho}_{\text{out}} | \hat{O}(\chi_N) \cdots \hat{O}(\chi_1) | \tilde{\rho}_{\text{in}}\rangle = \langle \rho_{\text{out}} | \mathcal{O}(\chi_N) \cdots \mathcal{O}(\chi_1) | \rho_{\text{in}}\rangle. \quad (3)$$

The first line is the computation result according to the error model, and the second line is the computation result of the actual quantum computer, which are equal. In this sense the error model is exact. The exactness of the error model does not rely on how to choose the matrix $\tilde{M}_{\text{in}}$. If we choose a different matrix $\tilde{M}_{\text{in}}$, then we can obtain another error model $|\tilde{\rho}_{\text{out}}\rangle = S |\rho_{\text{out}}\rangle$, $\langle \tilde{\rho}_{\text{out}} | = \langle \rho_{\text{out}} | S^{-1}$ and $\hat{O}(\chi) = S \hat{O}(\chi) S^{-1}$, where $S = \tilde{M}_{\text{in}}^T \tilde{M}_{\text{in}}^{-1}$. Both error models can exactly characterize the quantum computer, because the difference between two error models is only a similarity transformation [6, 37, 38].

B. Exact quantum error mitigation

We consider the QEM approach based on the quasi-probability decomposition formalism [34, 35]. The approach works for any quantum algorithm that computes the mean of an observable. To get the error-free computation result, we implement a set of random circuits according to a specific distribution. Errors are canceled when we take the average of these random circuits. The distribution of random circuits is worked out based on the quasi-probability decomposition formula of each computation operation. For an error-free operation $\mathcal{O}^{(0)}$, we decompose it using actual computation operations $\mathcal{O}_j$ in the from $\mathcal{O}^{(0)} = \sum_j q_j \mathcal{O}_j$, where $q_j$ are quasi-probabilities. Each operation is a matrix in the Pauli transfer matrix representation. The matrix of an error-free operation can be computed using a classical computer. The matrix of an actual operation can be obtained in GST if errors are not correlated [35]. Now, we show how to implement QEM when errors have temporal correlations based on matrices obtained in LOT.

- Compute error-free vectors $|\rho_{\text{in}}^{(0)}\rangle$, $\langle \rho_{\text{out}}^{(0)} |$ and matrices $\mathcal{O}^{(0)}(\chi)$ for each $\chi$, which are $d$-dimensional:

First, we compute error-free vectors $|\rho_{\text{in}}^{(0)}\rangle_S$, $\langle \rho_{\text{out}}^{(0)} |_S$ and matrices $\mathcal{O}^{(0)}_S(\chi)$ for each $\chi$ for the system only. If the Hilbert space of the system is $d_S$-dimensional, these vectors and matrices are $d_S$-dimensional. Given expressions of error-free states, observables and gates, the computation is straightforward using the definition of the Pauli transfer matrix representation.

Then, choose a $d$-dimensional real invertible matrix $S$, and take $|\rho_{\text{in}}^{(0)}\rangle = S |\rho_{\text{in}}^{(0)}\rangle_S$, $\langle \rho_{\text{out}}^{(0)} | = \langle \rho_{\text{out}}^{(0)} | S^{-1} Q S$ and

$$\mathcal{O}^{(0)}(\chi) = S \begin{bmatrix} \mathcal{O}^{(0)}_S(\chi) & \mathcal{O}^{(0)}_S(\chi) \mathcal{O}^{(0)}_S(\chi) \end{bmatrix} S^{-1}.$$

Here, $S$ is the $(d-d_S^2)$-dimensional zero row vector, $Q$ can be any $(d-d_S^2)$-dimensional real row vector, $\mathbf{0}$ is the $(d-d_S^2) \times d_S^2$ zero matrix, and $\mathcal{O}(\chi)$ and $\mathcal{O}^{(0)}(\chi)$ can be any $d_S^2 \times (d-d_S^2)$ and $(d-d_S^2) \times (d-d_S^2)$ real matrices, respectively.

- Choose a set of states $|\rho_j^{(b)}\rangle$, a set of observables $\{|\mathcal{Q}_i^{(b)}\rangle\}$ and a set of operations $\{\mathcal{O}_j^{(b)}\}$. Compute $|\tilde{\rho}_j^{(b)}\rangle$, $\{|\tilde{\mathcal{Q}}_i^{(b)}\rangle\}$ and $\{|\tilde{\mathcal{O}}_j^{(b)}\rangle\}$ using the result of LOT.

If $|\rho_j^{(b)}\rangle = \mathcal{O} |\rho_{\text{in}}\rangle$ and $\langle \rho_j^{(b)} | = \langle \mathcal{O} | \rho_{\text{out}} \rangle$, we take $|\tilde{\rho}_j^{(b)}\rangle = \tilde{\mathcal{O}} |\rho_{\text{in}}\rangle$ and $\langle \tilde{\rho}_j^{(b)} | = \langle \mathcal{O} | \tilde{\mathcal{O}} | \rho_{\text{out}} \rangle$ accordingly. Here, $\tilde{\mathcal{O}} = \tilde{\mathcal{O}}(\chi_N) \cdots \tilde{\mathcal{O}}(\chi_1)$ if $\mathcal{O} = \mathcal{O}(\chi_N) \cdots \mathcal{O}(\chi_1)$.

These states, observables and operations must satisfy the condition that $\{|\tilde{\rho}_j^{(b)}\rangle\}$, $\{|\tilde{\mathcal{Q}}_i^{(b)}\rangle\}$ and $\{|\tilde{\mathcal{O}}_j^{(b)}\rangle\}$ span the $d$-dimensional column vector space, row vector space and matrix space, respectively.

- Find decomposition formulas $\langle \rho_{\text{in}}^{(0)} | = \sum_i q_{\rho,i} |\tilde{\rho}_i^{(b)}\rangle$, $\langle \rho_{\text{out}}^{(0)} | = \sum_k q_{Q,k} \langle \tilde{\mathcal{Q}}_k^{(b)} |$ and $\mathcal{O}^{(0)}(\chi) = \sum_j q_{\mathcal{O},j} \langle \tilde{\mathcal{O}}_j^{(b)} |$
for each \( \chi \), by solving systems of linear equations. We consider the completeness of states, observables and operations, solutions always exist. Compute \( C_{in} = \sum_i |p_{i}\rangle \langle q_{i}| \), \( C_{out} = \sum_k |q_{Q,k}\rangle \) and \( C_{\chi} = \sum_j |q_{\chi,j}\rangle \) for each \( \chi \).

- To compute \( \langle\langle Q^{(0)}_{out} | O^{(0)}(\chi_N) \cdots O^{(0)}(\chi_1) | \rho^{(0)}_{in} \rangle \rangle \) using QEM, randomly choose and implement the actual quantum circuit \( \langle\langle Q^{(b)}_{k} | O^{(b)}_{j_1} \cdots O^{(b)}_{j_l} | \rho^{(b)}_{i} \rangle \rangle \) in the quantum computer with the probability \( C^{-1} |q_{p,b}\rangle \langle q_{Q,b}| \prod_{l=1}^{N} |q_{\chi,l-j,b}\rangle |C_{\chi}\rangle \), where \( C = C_{in}C_{out} \prod_{l=1}^{N} C_{\chi} \times 1 \). For each trial, compute the effective outcome \( \mu_{eff} = \text{sgn}(q_{p,b},q_{Q,b},q_{\chi,b}) \prod_{l=1}^{N} |q_{\chi,l-j,b}\rangle |C_{\chi}\rangle \), where \( \mu \) is the outcome in the measurement of the observable \( Q_{k} \). Repeat the actual quantum circuit randomly chosen according to the probability distribution, and compute the average of \( \mu_{eff} \), which is the final result of the quantum computation.

If the actual quantum circuit \( \langle\langle Q^{(b)}_{k} | O^{(b)}_{j_1} \cdots O^{(b)}_{j_l} | \rho^{(b)}_{i} \rangle \rangle \) includes multiple measurements as we will discuss soon, \( \mu \) takes the product of outcomes of all measurements in the circuit.

Using this protocol of QEM, the distribution of the final computation result, i.e. an estimate of \( \langle\langle Q^{(0)}_{out} | O^{(0)}(\chi_N) \cdots O^{(0)}(\chi_1) | \rho^{(0)}_{in} \rangle \rangle \), is centered at the error-free value [35]. The variance of the final result distribution is \( \sim C^2/n \), where \( n \) is the number of trials for computing the average of \( \mu_{eff} \).

Complete sets of states, observables and operations always exist. We consider states \( \{|\rho_{i}\rangle\} \) and observables \( \{|Q_{k}\rangle\} \) used in LOT, which are columns of \( M_{in} \) and rows of \( M_{out} \), respectively. We can find that \( \{|\hat{\rho}_{i}\rangle\} \) and \( \{|\hat{Q}_{k}\rangle\} \) are complete, which are columns of \( \tilde{M}_{in} \) and rows of \( \tilde{M}_{out} \), respectively. Then, we can also construct a complete set of operations based on these states and observables. We consider the \( d^2 \) operations \( \{|B_{i,k} = |\rho_{j}\rangle\}_{Q_{k}} \). The operation \( |\rho\rangle\rangle Q \) means measuring the observable \( Q \) and then preparing the state \( \rho \). Because \( \{B_{i,k}\} \) are linearly independent, these operations form a complete set.

### IV. SPACE DIMENSION TRUNCATION

Usually, the environment is a high-dimensional Hilbert space. Although only the subspace \( V \) is relevant in the exact LOT, its dimension could still be too high to allow the exact LOT to be implemented. Therefore, a practical LOT protocol is approximate and requires that a low-dimensional subspace approximately characterize the quantum computer. Here, we give a sufficient condition for the existence of such a subspace.

We consider the case that \( d < \text{Tr}(P) \), i.e. states \( \{|\rho_{i}\rangle\}_{i = 1,\ldots,d} \) and observables \( \{|Q_{k}\rangle \}_{k = 1,\ldots,d} \) are not enough to implement the exact LOT. We find that the quantum computer is approximately characterized by \( \Pi_{in}M_{in}, ~ M_{out}\Pi_{in} \) and \( \Pi_{in}\mathcal{O}(\chi)\Pi_{in} \) if the subspace spanned by \( \{|\rho_{i}\rangle\} \) is approximately invariant under operations \( \{|\mathcal{O}(\chi)\rangle\} \). Here, \( \Pi_{in} \) and \( \Pi_{out} \) are orthogonal projections on subspaces span\( \{|\rho_{i}\rangle\} \) and span\( \{|Q_{k}\rangle\} \), respectively.

If \( \|\rho_{i}\| \leq N_{\rho}, ~ \|\langle\langle Q_{k}\rangle\rangle\| \leq N_{Q}, ~ \|\mathcal{O}(\chi)\| \leq N_{\mathcal{O}}, \) and \( \|\Pi_{in}\mathcal{O}(\chi)\Pi_{in} - \mathcal{O}(\chi)\Pi_{in}\| \leq \epsilon \) [50], we have

\[
\|M_{out}\mathcal{O}(\chi_N) \cdots \mathcal{O}(\chi_1)M_{in} - M_{out}\Pi_{in}\mathcal{O}(\chi_N)\Pi_{in} \cdots \Pi_{in}\mathcal{O}(\chi_1)\Pi_{in}M_{in}\|_{\max} \\
\leq N_{Q} N_{\rho} \left( (N_{\mathcal{O}} + \epsilon)^N - N_{\mathcal{O}}^N \right) \sim N_{Q} N_{\rho} N_{\mathcal{O}}^{N-1} \times N_{\rho} \epsilon \tag{4}
\]

for any sequence of elementary operations. See Appendix for the proof. Here, we always have \( N_{\mathcal{O}} = 1 \) by taking the trace norm. A small \( \epsilon \) means that the subspace is approximately invariant under elementary operations, in which case an approximate tomography is possible.

### V. APPROXIMATE MODELS OF TEMPORALLY CORRELATED ERRORS

The practical use of LOT requires that a low-dimensional approximate model exists. In this section, we consider two typical sources of temporal correlations, i.e. low-frequency noise and context-dependent noise. For both of them, low-dimensional approximate models exist.

#### A. Low-frequency noise and classical random variables

A typical source of temporally correlated errors is the stochastic variation of classical parameters as discussed in Sec. II. If the correlation time of the noise is negligible compared with the time of a quantum gate, temporal correlations in gate errors are insignificant. However, if the correlation time is comparable or even longer than the gate time, errors are correlated, i.e. a sequence of quantum gates cannot be factorized because of the low-frequency noise. We show that errors with this kind of correlations can be efficiently approximated using a low-dimensional model if moments of the parameter distribution converge rapidly.

As the same as in Sec. II, We can use

\[
\rho = \int d\lambda p(\lambda) \rho_{S}(\lambda) \otimes |\lambda\rangle\langle\lambda|_{E} \tag{5}
\]

to describe a state that depends on random variables \( \lambda \). Here, \( \lambda \) is an array with \( n \) elements that respectively denote \( n \) variables, and \( \rho_{S}(\lambda) \) is the state of the system when variables are \( \lambda \). An operation depending on random variables reads

\[
\mathcal{O}(\chi) = \mathcal{T}(\chi) \int d\lambda \mathcal{O}_{S}(\chi,\lambda) \otimes |\lambda\rangle\langle\lambda|_{E}, \tag{6}
\]

where \( \mathcal{O}_{S}(\chi,\lambda) \) is the operation on the system when variables are \( \lambda \). Compared with the expression in Sec. II,
there is an additional operation \( \mathcal{T}(\chi) \) in \( \mathcal{O}(\chi) \), which describes the stochastic evolution of variables in the time of the operation. Here \( \mathcal{T}(\chi) = [\mathbb{I}_S] \otimes \mathcal{T}_E(\chi) \), \( \mathcal{T}_E(\chi) = \int d\lambda' d\lambda \sigma_{\chi,\lambda}(\chi)|\lambda'\rangle\langle \lambda_0| \), \( \sigma_{\chi,\lambda}(\chi) \) is the transition probability density from \( \lambda \) to \( \lambda' \), and \( \mathbb{I} \) is the identity operator. Similarly, an observable depending on random variables reads

\[
Q = \int d\lambda Q_S(\lambda) \otimes |\lambda\rangle\langle \lambda'|_E, \tag{7}
\]

where \( Q_S(\lambda) \) is the observable of the system when variables are \( \lambda \).

The approximate model is given by

\[
\rho^a = \sum_{\lambda \in L} p^a(\lambda) \rho_S(\lambda) \otimes |\lambda\rangle\langle \lambda'|_E, \tag{8}
\]

\[
\mathcal{O}^a(\chi) = \mathcal{T}^a(\chi) \sum_{\lambda \in L} \mathcal{O}_S(\chi, \lambda) \otimes [|\lambda\rangle\langle \lambda'|_E], \tag{9}
\]

\[
Q^a = \sum_{\lambda \in L} Q_S(\lambda) \otimes |\lambda\rangle\langle \lambda'|_E. \tag{10}
\]

Here, \( L \) is a finite subset of random variables. If \( \lambda \) takes \( m \) values in \( L \), i.e. \( |L| = m \), the environment in the approximate model is \( m \)-dimensional. The transition operation in the approximate model is \( \mathcal{T}^a(\chi) = [\mathbb{I}_S] \otimes \mathcal{T}_E^a(\chi) \), where \( \mathcal{T}_E^a(\chi) = \sum_{\lambda,\lambda' \in L} \sum_{a} t^{a}_{\lambda',\lambda}(\chi)[|\lambda\rangle\langle \lambda'|_E] \). We remark that \( \rho_S(\lambda), \mathcal{O}_S(\chi, \lambda) \) and \( Q_S(\lambda) \) are the same as in Eqs. (5)–(7). By properly choosing the subset of random variables \( L \), the distribution \( p^a(\lambda) \) and transition matrices \( t^{a}_{\lambda',\lambda}(\chi) \), such a model can approximately characterize the quantum computer as we will show next.

We focus on the case of only one random variable, and the generalization to the case of multiple variables is straightforward. In any quantum computation platform, the effect of the noise on quantum operations should be weak, i.e. error rates are low. In this case, only low-order moments are important. Using the Taylor expansion, we have

\[
\rho_S(\lambda) = \sum_{l=0}^{\infty} \rho_S^{(l)}(\lambda) \lambda^l, \tag{11}
\]

\[
\mathcal{O}_S(\chi, \lambda) = \sum_{l=0}^{\infty} \mathcal{O}_S^{(l)}(\chi, \lambda) \lambda^l, \tag{12}
\]

\[
Q_S(\lambda) = \sum_{l=0}^{\infty} Q_S^{(l)}(\lambda^l). \tag{13}
\]

Then, any quantum computation of a mean value, i.e. the mean of an observable \( Q \) in the state \( \rho \) after a sequence of operations, can be expressed as

\[
\langle Q | \mathcal{O}(\chi_N) \cdots \mathcal{O}(\chi_1) | \rho \rangle = \sum_{l=0}^{\infty} \langle Q_S^{(l)}(\chi_N) \cdots \mathcal{O}_S^{(l)}(\chi_1) \rangle \rho_S^{(l)}, \tag{14}
\]

where \( l = (l_0, l_1, \ldots, l_N) \), \( \lambda_j \) is the value of the variable at the time of the \( j \)-th operation, and the overline denotes the average. Therefore, the behaviour of the quantum computer is determined by correlations of random variables. If these correlations can be approximately reconstructed in the approximate model, the model approximately characterizes the quantum computer.

Correlations are formally defined here. We introduce the operator \( \hat{\lambda} = \int d\lambda |\lambda\rangle\langle \lambda'|_E \). Then,

\[
\lambda_N^{l_{N+1}} \cdots \lambda_0^{l_0} = \text{Tr} \left[ \hat{\lambda}^{l_{N+1}} \mathcal{T}_E(\chi_N) \cdots \mathcal{T}_E(\chi_1) \hat{\lambda}^{l_1} \hat{\lambda}^{l_0} \rho_E \right], \tag{15}
\]

where \( \rho_E = \int d\lambda \rho(\lambda) |\lambda\rangle\langle \lambda'|_E \).

1. Second-order approximation

First, consider the case that the contribution of high-order correlations other than \( \lambda_j \) and \( \lambda_{j'} \lambda_{j''} \) is negligible, the distribution of \( \lambda \) is stationary, and the correlation time is much longer than the time from the state preparation to the measurement. In this case, only \( \lambda_j \) and \( \lambda_{j'} \lambda_{j''} \) are important. Without loss of generality, we assume that the distribution is centered at \( \lambda = 0 \), i.e. \( \lambda_j = 0 \).

Because of the long correlation time, the change of the random variable is slow, and \( \lambda_j \lambda_{j'} \lambda_{j''} \ll \sigma^2 \) is approximately a constant. Such correlations can be reconstructed in the approximate model with \( m = 2 \). We can take parameters in the approximate model as \( L = \{ \lambda = \pm \sigma \} \), \( p^a(\pm \sigma) = 1/2 \) and \( t_{\lambda,\lambda}(\chi) = \delta_{\lambda,\lambda} \).

Then, we consider the case that the correlation is not a constant but decreases with time. We assume that for each operation \( \chi \) the correlation is reduced by a factor of \( e^\Gamma(\chi) \). If the correlation decreases exponentially with time, \( \Gamma(\chi) \) is proportional to the operation time. The correlation reads \( \lambda_j \lambda_{j'} \ll \sigma^2 \exp[\sum_{i=1}^{j'-1} \Gamma(\chi_i)] \).

This two-time correlation can also be reconstructed in the approximate model with \( m = 2 \). The only difference is the transition matrix. We take the transition matrix as \( \mathcal{T}_E^a(\chi) = \mathcal{T}_E^a(\Gamma(\chi)) \), and

\[
\mathcal{T}_E^a(\Gamma) = \frac{1 + e^{-\Gamma}}{2} (-|\lambda\rangle\langle \lambda'|_E) + \frac{1 - e^{-\Gamma}}{2} (|\lambda\rangle\langle \lambda|_E). \tag{16}
\]

where \( \pm |\lambda\rangle = |\lambda = \pm \sigma \rangle \). Then we have \( \mathcal{T}_{E}^a(\Gamma') \mathcal{T}_{E}^a(\Gamma) = \mathcal{T}_{E}^a(\Gamma'' + \Gamma) \).

2. High-order approximations and multiple variables

We consider the case that the change of the random variable is negligible in the time from the state preparation to the measurement, i.e. \( \mathcal{T}_E(\chi) \simeq [\mathbb{I}_E] \). Then, correlations become \( \lambda_N^{l_{N+1}} \cdots \lambda_0^{l_0} \simeq \int d\lambda \lambda^l \), where \( l = \sum_{j=0}^{N+1} l_j \). If the contribution of correlations with \( l \leq l_{j} \) is negligible, we only need to reconstruct correlations with \( l \leq l_{j} \) in the approximate model, which is...
always possible by taking \(m = \lceil (l_t + 1)/2 \rceil\) [51]. We remark that \(m\) is the dimension of the environment in the approximate model.

It is similar for multiple random variables. If moments of the distribution converge rapidly, the Gaussian cubature approximation can be applied [52]. Then, up to \(l_t^{th}\)-order moments can be reconstructed with \(m = \binom{n_\lambda + l_t}{l_t}\), where \(n_\lambda\) is the number of random variables.

### B. Context-dependent noise

Context dependence is a kind of effects that the error in an operation depends on what operations have been performed before the operation, i.e. the environment has the memory of previous operations. Because this kind of effects is in the scope of our general model of the quantum computer in Sec. II, our results can be applied to the context-dependent noise. A list of previous operations is classical information, so the memory of previous operations can be treated as a set of classical variables whose evolution is operation-dependent. Therefore, we can use the same formalism for classical random variables to characterize the context-dependent noise. We consider two examples as follows.

In the ion trap, the temperature of ions may depend on how many gates have been performed after the last cooling operation, and the fidelity of a gate depends on the temperature. This effect can be characterized using a set of discretized variables \(\lambda = (n_1, n_2, \ldots)\). Here, \(n_i\) denotes the phonon number of the \(i^{th}\) mode. Because of the low temperature of ions, each \(n_i\) can be truncated at a small number. Suppose the evolution of the qubit state mainly depends on the distribution at the beginning of the gate, the gate can be expressed as the same as in Eq. (9), where \(\mathcal{T}(\chi)\) describes the heating process in the gate \(\chi\). The cooling operation can also be expressed in this form. Then, we can apply the approximation similar to classical random variables.

If the error in an operation only significantly depends on a few previous operations, we can use a low-dimensional environment to characterize the effect. We focus on the case that the error only depends on the last one operation, and it can be generalized to the case of depending on multiple previous operations. We characterize this effect using one discretized variable \(\lambda \in \{\chi\}\), where \(\{\chi\}\) is the list of all possible operations. The state after the operation \(\lambda\) can be expressed in the form \(\rho = \rho_\mathcal{S} \otimes |\lambda\rangle \langle \lambda|_E\). An operation can be expressed in the form \(\mathcal{O}(\chi) = \sum_\lambda \mathcal{O}_\mathcal{S}(\chi, \lambda) \otimes |\chi\rangle \langle \chi|_E\), where \(\mathcal{O}_\mathcal{S}(\chi, \lambda)\) is the operation on the system when the last operation is \(\lambda\). After the operation, the state becomes \(\mathcal{O}(\chi)|\rho\rangle = |\mathcal{O}_\mathcal{S}(\chi, \lambda)\rho_\mathcal{S}\rangle \otimes |\chi\rangle \langle \chi|_E\). We remark that it is not necessary that \(|\lambda\rangle_E\) corresponds to a pure state in a physical Hilbert space. It has been found in the experiment that a similar formalism can characterize the non-Markovian noise in the ion trap system [6].

### VI. APPROXIMATE QUANTUM TOMOGRAPHY

The exact tomography protocol is not practical because of the high-dimensional state space of the environment. In Sec. V, we show that an effective model with a low-dimensional environment state space can approximately characterize the quantum computer for typical temporally correlated noises. In this section, we discuss how to implement LOT to obtain a low-dimensional approximate model of the quantum computer. There are two approaches of self-consistent tomography, the linear inversion method (LIM) and the maximum likelihood estimation (MLE) [6, 37, 38], and we will discuss both of them.

#### A. Linear inversion method

Given sufficient data from the experiment, we can use LIM to obtain an exact model of the quantum computer as in the exact LOT. However, to obtain an approximate model, even if the approximate model exists, LIM does not always work. We suppose that \(d \times d\) matrices \(M_{in}^a\), \(M_{out}^a\) and \(\{\mathcal{O}^a(\chi)\}\) satisfy

\[
\| M_{out}^a \mathcal{O}(\chi) \cdots \mathcal{O}(\chi_1) M_{in} \| - M_{out}^a \mathcal{O}^a(\chi) \cdots \mathcal{O}^a(\chi_1) M_{in}^a \|_{max} \leq \epsilon_N \quad (17)
\]

for any sequence of elementary operations, where \(\epsilon_N\) is a small quantity depending on \(N\), and \(d < Tr(P)\). Then, these matrices form a model that approximately characterises the quantum computer. If the approximate model exists, we only need to obtain \(d \times d\) matrices \(g^a = M_{out}^a M_{in}^a\) and \(\{\tilde{\mathcal{O}}^a(\chi) = M_{out}^a \mathcal{O}^a(\chi) M_{in}^a\}\) in order to approximately characterise the quantum computer. Because \(\| g - g^a \|_{max} \leq \epsilon_0\) and \(\| \tilde{\mathcal{O}}(\chi) - \tilde{\mathcal{O}}^a(\chi) \|_{max} \leq \epsilon_1\), we can directly use \(g\) and \(\{\tilde{\mathcal{O}}(\chi)\}\) as estimates of \(g^a\) and \(\{\tilde{\mathcal{O}}^a(\chi)\}\), which can be obtained in the experiment. However, \(g^{-1}\) may be very different from \((g^a)^{-1}\), and in this case Eq. (2) may not even approximately hold. We remark that Eq. (2) always exactly holds if \(d = Tr(P)\). If Eq. (2) does not hold, LIM does not work.

LIM works for the approximate model if the following conditions are satisfied. \(\{|\rho_i^a\rangle\}\) are columns of \(M_{in}^a\), and \(\{|\langle Q_k^a|\rangle\}\) are rows of \(M_{out}^a\). Then, if \(\|\rho_i^a\| \leq N_{\rho}^a\), \(\|\langle Q_k^a|\| \leq N_{Q}^a\), \(\|\mathcal{O}^a(\chi)\| \leq N_{\Omega}\), \(\|M_{in}^a\|_{max} \leq 1\), \((1 + \epsilon_0)^{N-1}(N_0 + \epsilon_0)^N - N_\Omega\)

\[
\sim N_{Q} N_{\rho}^a \times N_{\Omega}^{N-1} [(N_0 + N_\Omega) + N_\epsilon \epsilon_0] \quad (18)
\]

for any sequence of elementary operations. See Appendix for the proof. Therefore, LIM works under conditions that \(N_\Omega \lesssim 1\), and \(\epsilon_0\) and \(\epsilon_1\) are small quantities.
We apply this result to the approximate model given by the approximately invariant subspace $\Pi_n$. We have
\[
\|\tilde{O}(\chi \nu)g^{-1}\cdots \tilde{O}(\chi_2)g^{-1}\tilde{O}(\chi_1) - M_{\text{out}} \Pi_n g^{-1} \Pi_n \cdots \Pi_n g^{-1} \Pi_n M_{\text{in}}\|_{\text{max}} \\
\leq N_Q(1 + \epsilon_1)N_p \left[ \left( N_O + \frac{\epsilon}{1 - \epsilon_1} \right)^N - (N_O + \epsilon)^N \right],
\]
where $\epsilon_1 = ||\Pi_n - \Pi_{\text{out}}||$, and we have $N_O = 1$ by taking the trace norm. Therefore, if $\|\Pi_n O(\chi) \Pi_n - O(\chi) \Pi_n\| < 1$ and $\|\Pi_n - \Pi_{\text{out}}\| < 1$ for the trace norm, LIM can be applied. Here, $\|\Pi_n - \Pi_{\text{out}}\| < 1$ means that two subspaces $\Pi_n$ and $\Pi_{\text{out}}$ are approximately the same.

In order to implement LIM, we need to find states $\{|\rho_i\rangle\}$ and observables $\{|\langle Q_k |\rangle\}$ corresponding to an approximately invariant subspace $\Pi_n$. For this purpose, we choose a set of trial states $\{|\rho_i^t\rangle\}$ and a set of trial observables $\{|\langle Q^t_k |\rangle\}$. The most interesting approximate invariant subspace is the subspace containing the initial state $|\rho_n\rangle$. If such an approximate invariant subspace exists, states in the form $O(\chi) \Pi_n \rho_n$ are all approximately within the subspace, as long as $N$ is sufficiently small. Therefore, we can choose the initial states $|\rho_n\rangle$ and states in the form $O(\chi) \Pi_n \rho_n$ as trial states. Similarly, we can choose the observable $\langle Q_{\text{out}} \rangle$ and effective observables in the form $\langle Q_{\text{out}} | O(\chi) \cdots O(\chi) | \rangle$ as trial observables.

In the ideal case, i.e. $\Pi_n$ is an exactly invariant subspace, and trial states and observables are exactly within $\Pi_n$, i.e. $\Pi_n [\rho^t_i] = [\rho^t_i]$ and $\langle Q^t_k \rangle \Pi_n = \langle Q^t_k \rangle$. Then, the rank of $g^t_i = M_{\text{out}}^t M_{\text{in}}^t$ is not greater than the dimension of the subspace $\Pi_n$. Here, $M_{\text{in}}^t$ and $M_{\text{out}}^t$ are matrices corresponding to trial states and observables, respectively. If the subspace $\Pi_n$ is approximately invariant, the matrix $g^t$ should still be close to a matrix with a rank not greater than the subspace dimension. Therefore, we can determine $M_{\text{in}}$ and $M_{\text{out}}$ by performing a truncation on the spectrum of singular values of $g^t$, i.e. we choose $M_{\text{in}}$ and $M_{\text{out}}$ corresponding to the greatest d singular values of $g^t$. Suppose the singular value decomposition is $U g^t V = \Lambda$, where $\Lambda = \text{diag}(s_1, s_2, \ldots, s_d)$ and $s_1 \geq s_2 \geq \cdots \geq s_d$, then we use states $\{|\rho_i\rangle\} = \sum_j |\rho_i^j\rangle \Pi_n$ and observables $\{|\langle Q_k |\rangle\}$ to implement LOT.

The approximate LOT protocol using LIM is given here.

- Choose a set of states $\{|\rho_i^t\rangle\} = \mathcal{O}_i |\rho_n\rangle$, $\mathcal{O}_i \in \{O_i ; i = 1, \ldots, d\}$, and a set of observables $\{|\langle Q^t_k |\rangle\} = \langle Q_{\text{out}} | O^t_k \rangle$, $O^t_k \in \{O_k ; k = 1, \ldots, d\}$. We always take $|\rho_i^t\rangle = |\rho_n\rangle$ and $\langle Q^t_k |\rangle = \langle Q_{\text{out}} |$. We always take $|\rho_i^t\rangle = |\rho_n\rangle$ and $\langle Q^t_k |\rangle = \langle Q_{\text{out}} |$.

- Obtain matrices $g^t_i = M_{\text{out}}^t M_{\text{in}}^t$ and $\tilde{O}^t_i(\chi) = M_{\text{out}}^t O(\chi) M_{\text{in}}^t$ for each $\chi$ in the experiment. Here, $M_{\text{in}}^t = \left[ |\rho_1^t\rangle \cdots |\rho_d^t\rangle \right]$ and $M_{\text{out}}^t = \left[ \langle Q_1^t | \cdots \langle Q_d^t | \right]^\top$.

- Compute the singular value decomposition $U g^t V = \Lambda$, where $\Lambda = \text{diag}(s_1, s_2, \ldots, s_d)$, and singular values are sorted in the descending order $s_1 \geq s_2 \geq \cdots \geq s_d$.

- Choose the dimension $d$. Compute $g = \text{diag}(s_1, s_2, \ldots, s_d) = \Lambda D d^d$ and $\tilde{O}^t_i(\chi) = U D U^\top$ for each $\chi$. Here, $D$ is a $d \times d^d$ matrix, and $D_{i,i'} = \delta_{i,i'}$.

- Choose an arbitrary $d$-dimensional invertible real matrix $\hat{M}_{\text{in}}$, and compute $\hat{M}_{\text{out}} = g \hat{M}_{\text{in}}^{-1}$.

- Compute $|\hat{\rho}_{\text{in}}\rangle = \sum_{i=1}^d \hat{M}_{\text{in}} |V^*_{i,1} | O_{\text{out}} | V^*_{i,1} | \Pi_{\text{out}} \Pi_n | \hat{\rho}_{\text{in}}(\chi)\rangle$, $\langle \hat{Q}_{\text{out}} | = \sum_{k=1}^d U^*_{k,1} \hat{M}_{\text{out}} | \Phi_{\text{out}} | \Pi_{\text{out}} \Pi_n | \langle \hat{Q}_k |$, and $\tilde{O}(\chi) = \hat{M}_{\text{in}}^{-1} \tilde{O}(\chi) \hat{M}_{\text{in}}^{-1}$ for each $\chi$.

### B. Maximum likelihood estimation

The alternative method for determining the error model is based on MLE. Given a model of the quantum computer with unknown parameters, MLE is to find the estimated values of the unknown parameters, such that the probability of samples observed in the experiment is maximized. Let $d$-dimensional column vector $|\tilde{\rho}_{\text{in}}(\chi)\rangle$, row vector $\langle \hat{Q}_{\text{out}} |$ and matrices $\{O(\chi, x)\}$, respectively corresponding to the initial state, measured observable and operations, be the theoretical model of the quantum computer depending on parameters $x$. Our goal is to estimate parameters $x$ based on data from the experiment. The mean of $Q_{\text{out}}$ in $\rho_{\text{in}}$ after a sequence of operations measured in the experiment is $C = \langle \hat{Q}_{\text{out}} | O(\chi, x) \cdots O(\chi, x) | \hat{\rho}_{\text{in}}(\chi)\rangle + \delta$, where $\delta$ is the deviation from the actual mean value, and the mean according to the model is $C(x) = \langle \hat{Q}_{\text{out}} | O(\chi, x) \cdots O(\chi, x) | \tilde{\rho}_{\text{in}}(\chi)\rangle$. Using the Gaussian approximation, the likelihood function to be maximized is $L(x) = \exp\left\{ - \left[ \frac{C(x) - C}{\sigma} \right]^2 \right\}$, where $\sigma$ is the standard deviation of $C$. In the practical implementation, multiple quantum circuits and corresponding mean values are needed to determine the error model. The protocol is given in here.

- Parameterize the $d$-dimensional column vector $|\tilde{\rho}_{\text{in}}(\chi)\rangle$, row vector $\langle \hat{Q}_{\text{out}} |$ and matrix $\tilde{O}(\chi, x)$ for each $\chi$ as functions of parameters $x = (x_1, x_2, \ldots)$.

- Choose $M$ circuits $\{X_1, \ldots, X_M\}$. For each circuit $X_m = (x_{m,1}, x_{m,2}, \ldots, x_{m,N_m})$, obtain $\langle \hat{Q}_{\text{out}} | O(X_m, x) \cdots O(X_m, x) | \tilde{\rho}_{\text{in}}(\chi)\rangle$ in the experiment. The result is $C_m$.

- Minimize the likelihood function $L(x) = \prod_{m=1}^M \exp\left\{ - \left[ \frac{C_m(x) - C_m}{\sigma_m} \right]^2 \right\}$, where $C_m(x) =$
and numerically implement LOT and QEM on a classical computer. In the model, gate fidelities depend on a low-frequency time-dependent variable $\lambda$, whose distribution is Gaussian. We assume that the change of the variable is negligible in the time scale of a quantum circuit, i.e. the time from the state preparation to the measurement. The initial state and the observable to be measured are error free, which are $\rho_{in} = |0\rangle\langle 0| \otimes \rho_E$ and $Q_{out} = |0\rangle\langle 0| \otimes \mathbb{I}_E$, respectively. Here, the state of the environment is $\rho_E = \int d\lambda \frac{1}{2\pi} \exp \left(\frac{\lambda^2}{2}\right) |\lambda\rangle\langle \lambda|$. Errors in single-qubit gates are depolarizing errors, and depolarizing rates depend on $\lambda$. For a unitary single-qubit gate $G$, the actual gate with error is $O_S(G,\lambda) = \mathcal{E}(\epsilon_G(\lambda))|G\rangle\langle G|$, where $\epsilon_G(\lambda)$ is the depolarizing rate, $\mathcal{E}(\epsilon) = (1 - \epsilon)|\mathbb{I}| + \epsilon D$, and $D = \frac{1}{4}(|\mathbb{I}| + |X\rangle\langle X| + |Y\rangle\langle Y| + |Z\rangle\langle Z|)$. Here, $X$, $Y$, and $Z$ are Pauli operators. Then the operation on SE for the gate $G$ is $\mathcal{O}(G) = \int d\lambda O_S(G,\lambda) \otimes |\lambda\rangle\langle \lambda|$. We consider two single-qubit gates, the Hadamard gate $H$ and the phase gate $S$, which can generate all single-qubit Clifford gates. We take $\epsilon_H(\lambda) = \epsilon_S(\lambda) = \bar{\eta}(1 - \exp(-\lambda^2))$, therefore, two gates are both optimized at $\lambda = 0$. Here, $\eta \in [0,1]$ denotes the strength of the noise. RB is the usual way of the verification of a quantum computing system [7, 8, 24]. In our simulation, we perform a sequence of $H$ and $S$ gates randomly chosen in the uniform distribution. We initialize the qubit in the state $|0\rangle$, perform the random gate sequence, and measure the probability in the state $|0\rangle$. We only take into account gate sequences that the final state is $|0\rangle$ in the case of ideal gates without error, so that the probability in the state $|0\rangle$ is expected to be 1. When errors are switched on, the probability in the state $|0\rangle$ is $F(N_g) = (1 + 1/\sqrt{1 + 2N_g^{\sigma^2}})/2$ if $N_g = 1$, where $N_g$ is the number of gates in the random gate sequence. The non-exponential decay of the probability is due to temporal correlations [25]. Without temporal correlation, the probability decreases exponentially with the gate number. If depolarizing rates are constants, i.e. $\epsilon_H(\lambda) = \epsilon_S(\lambda) = \epsilon$, we have $[1 + (1 - \epsilon)N_g]/2$.

In our simulation, we implement both LIM and MLE. We take the dimension of the state space $d = 4, 7$ to compare LOT with the conventional GST. In approximate models of classical random variables with stationary distribution (see Sec. V A), the state space is $[|d_S^2 - 1|n + 1]$-dimensional as explained in Appendix, when the system and environment Hilbert spaces are respectively $d_S$-dimensional and $m$-dimensional. Therefore, $d = 4, 7$ correspond to $m = 1, 2$ approximations, respectively. If $d = 4$, the LOT protocol is equivalent to the conventional GST protocol, because the one-dimensional environment is trivial and does not have any effect. As shown in Fig. 2, LOT with $d = 7$ can characterize the behavior of the quantum computer much more accurately than LOT with $d = 4$ (i.e. the conventional GST).

In the simulation of LOT using MLE, we parameterize the state, observable and operations as follows. The state is in the form $\tilde{\rho}_m = \sum_{\lambda \in L} p^m(\lambda)|0\rangle\langle 0| \otimes |\lambda\rangle\langle \lambda| \otimes |\mathbb{I}_E\rangle$. The observable is in the form $Q_{out} = \sum_{\lambda \in L} |0\rangle\langle 0| \otimes |\lambda\rangle\langle \lambda| \otimes |\mathbb{I}_E\rangle$. We consider two single-qubit gates, the Hadamard gate $H$ and the phase gate $S$, which can generate all single-qubit Clifford gates. We take $\epsilon_H(\lambda) = \epsilon_S(\lambda) = \bar{\eta}(1 - \exp(-\lambda^2))$, therefore, two gates are both optimized at $\lambda = 0$. Here, $\eta \in [0,1]$ denotes the strength of the noise. RB is the usual way of the verification of a quantum computing system [7, 8, 24]. In our simulation, we perform a sequence of $H$ and $S$ gates randomly chosen in the uniform distribution. We initialize the qubit in the state $|0\rangle$, perform the random gate sequence, and measure the probability in the state $|0\rangle$. We only take into account gate sequences that the final state is $|0\rangle$ in the case of ideal gates without error, so that the probability in the state $|0\rangle$ is expected to be 1. When errors are switched on, the probability in the state $|0\rangle$ is $F(N_g) = (1 + 1/\sqrt{1 + 2N_g^{\sigma^2}})/2$ if $N_g = 1$, where $N_g$ is the number of gates in the random gate sequence. The non-exponential decay of the probability is due to temporal correlations [25]. Without temporal correlation, the probability decreases exponentially with the gate number. If depolarizing rates are constants, i.e. $\epsilon_H(\lambda) = \epsilon_S(\lambda) = \epsilon$, we have $[1 + (1 - \epsilon)N_g]/2$.

In our simulation, we implement both LIM and MLE. We take the dimension of the state space $d = 4, 7$ to compare LOT with the conventional GST. In approximate models of classical random variables with stationary distribution (see Sec. V A), the state space is $[|d_S^2 - 1|n + 1]$-dimensional as explained in Appendix, when the system and environment Hilbert spaces are respectively $d_S$-dimensional and $m$-dimensional. Therefore, $d = 4, 7$ correspond to $m = 1, 2$ approximations, respectively. If $d = 4$, the LOT protocol is equivalent to the conventional GST protocol, because the one-dimensional environment is trivial and does not have any effect. As shown in Fig. 2, LOT with $d = 7$ can characterize the behavior of the quantum computer much more accurately than LOT with $d = 4$ (i.e. the conventional GST).
FIG. 3. (a) Probabilities in the ideal state $|0\rangle$ with the gate number 50. The probabilities at 9 different values of $\eta$ are obtained using the quantum computer in the numerical simulation (blue scatters). Then, by fitting the data (red curve), we can find that the probability is 1 at $\eta = 0$, which is our final computation result, i.e. an estimate of the result in the error-free quantum computation. (b) Depolarizing rates in the error model obtained in LOT using MLE with $d = 7$. Two depolarizing rates $\epsilon(G(1))$ and $\epsilon(G(2))$ ($G = H, S$) are the same for gates $S$ and $H$, which are obtained at 9 different values of $\eta$ using LOT (blue scatters). By fitting the data (red curves), we can find that depolarizing rates approach zero when $\eta = 0$. See Appendix for details of the simulation and more data.

|\lambda\rangle\langle\lambda|_E^a$. The gate $G$ with error is in the form $\hat{O}(G) = \sum_{\lambda \in E} \mathcal{E}(\epsilon_G^{a}(\lambda))|G\rangle \otimes |\lambda\rangle\langle\lambda|_E$. We take $\{p^a(\lambda)\}$ and $\{\epsilon_G^{a}(\lambda)\}$ as parameters (i.e. $x$) in MLE. The number of values that $\lambda$ can take, i.e. the dimension of the environment Hilbert space, is important, but the value of $\lambda$ is not important. For the one-dimensional environment approximation, i.e. $d = 4$, we take $L = \{1\}$; and for the two-dimensional environment approximation, i.e. $d = 7$, we take $L = \{1, 2\}$.

VIII. QUANTUM ERROR MITIGATION OF TEMPORALLY CORRELATED ERRORS

Given the result of approximate LOT, we can use the quasi-probability decomposition protocol in Sec. III B to mitigate errors. Then the accuracy of QEM is limited by the approximation in LOT. In Fig. 2, we plot the computation result with errors mitigated using the quasi-probability decomposition based on LOT with $d = 4$. We remark that the ideal value of the computation result is 1. We can find that there is still a significant error in the computation result because LOT with $d = 4$ cannot characterize temporal correlations.

Using the decomposition based on LOT with $d = 7$, the error can be reduced to a much lower level compared with $d = 4$ as shown in Fig. 2. However, we find that in this case, the cost factor $C$ increases rapidly with the number of gates. We remark that the variance of the computation result obtained in $n$ trials is $\sim C^2/n$. Therefore a larger $C$ means more trials and longer computation time. With $d = 4$, $C$ is increased by a factor of $\sim 1.01$ for each gate, so a computation using hundreds of gates could be practical. With $d = 7$, $C$ is increased by a factor of $\sim 2$ for each gate, so it is not practical.

A reason of the large cost factor is that we use unrelated random operations to compensate correlated errors. In the quasi-probability decomposition, each gate is decomposed independently, and distributions of random operations for replacing each gate are independent. However, errors mitigated by random operations are correlated. Using correlated random operations, we may be able to mitigate correlated errors. For example, for the context-dependent noise that the error in a gate depends on the last one gate (see Sec. V B), we can decompose a gate depending on what is the last gate. In this case, the distribution of random operations for replacing a gate is correlated with previous gates, and correlated errors can be efficiently mitigated with a practical cost.

Error extrapolation can efficiently mitigate correlated errors. We can boost correlated errors by either simulating errors [33] or tuning some physical parameters and then use the extrapolation to estimate the computation result without error. Because we introduce correlated errors in the error extrapolation, we expect that the cost is as low as in the case without correlation [35]. To demonstrate the error extrapolation, we suppose that the noise strength factor $\eta$ can be increased. As shown in Fig. 3(a), we can find that the computation error can be reduced to the level as low as in the quasi-probability decomposition protocol with $d = 7$. To implement the extrapolation, we take 9 different values of $\eta$, i.e. the quantum computation needs to be repeated at 9 values of $\eta$.

We can use LOT verify the extrapolation. We consider the situation that $\eta$ is a physical parameter that we can tune in the experiment, however, we do not exactly know how the error model depends on $\eta$. Then, we need to find out at which value of $\eta$ errors are minimized to implement the error extrapolation. In Fig. 3(b), we plot depolarizing rates obtain in LOT for different values of $\eta$, and we can find that depolarizing rates approach zero when $\eta = 0$ in the extrapolation. Therefore, in the error extrapolation, we should take the computation result at $\eta = 0$.

IX. CONCLUSIONS

We have proposed a tomography protocol to obtain the model of temporally correlated errors in a quantum computer. Given sufficient data from the experiment, the
model obtained in our protocol can be exact. However approximate models are favorable for practical implementation. To obtain approximate models characterizing temporal correlations, more quantities need to be measured compared with the conventional QPT and GST, but the overhead is moderate. We can use such approximate models to predict the behavior of a quantum computer much more accurately than the model obtained in GST in the presence of temporal correlations. Therefore, we can use such models to design circuits to mitigate temporally correlated errors in quantum computation, and we have proposed protocols to do so. Our protocols provide a way to quantitatively access temporal correlations in quantum computers and achieve high-fidelity quantum computation on NISQ devices.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China grant 2016YFA0301200 and the National Basic Research Program of China (under Grant No. 2014CB921403). It is also supported by Science Challenge Project (under Grant No. TZ2017003) and the National Natural Science Foundation of China (under Grants No. 11774024, No. 11534002, and No. U1530401). YL is supported by NSAF (Grant No. U1730449).

Appendix A: Exact linear operator tomography

We consider two subspaces $V_{in} = \text{span}(\{\langle Q_{out}\rangle | O \in \mathcal{O}\})$ and $V_{out} = \text{span}(\{\langle P_{out}\rangle | O \in \mathcal{O}\})$. We use $\hat{P}_{in}$ and $\hat{P}_{out}$ to denote the orthogonal projection on $V_{in}$ and $V_{out}$, respectively. Here, $V_{out} = V$ and $\hat{P}_{out} = P_{out}$. $P_{in/out}$ is the orthogonal projection on the intersection of $V_{in}$ and the orthogonal complement of $V_{out}$. $P_{in/out}$ is the orthogonal projection on the intersection of $V_{out}$ and the orthogonal complement of $V_{in}$. Then, $\hat{P}_{in} = P_{in} - P_{in/out}$ and $\hat{P}_{out} = P_{out} - P_{in/out}$.

**Lemma 1.** Let $O \in \mathcal{O}$, all the following expressions are valid.

\[
\begin{align*}
\hat{P}_{in} &= \hat{P}_{in} P_{in} = P_{in} \hat{P}_{in}, \\
\hat{P}_{out} &= \hat{P}_{out} P_{out} = P_{out} \hat{P}_{out}, \\
P_{out} P_{in} &= P_{out} \hat{P}_{in} = \hat{P}_{out} P_{in}, \\
P_{out} O P_{in} &= P_{out} \hat{P}_{in} O P_{in} = P_{out} \hat{P}_{in} O \hat{P}_{in}.
\end{align*}
\]

**Proof.** We have

\[
\begin{align*}
P_{in} P^{-1}_{in} &= P_{in} P^{-1}_{out} P_{in} = P_{in} P^{-1}_{in}, \\
P_{out} P^{-1}_{out} &= P_{out} P^{-1}_{in} P_{out} = 0, \\
P_{out} P^{-1}_{out} &= P_{out} P^{-1}_{in} P_{out} = P_{out} P^{-1}_{in}, \\
P_{in} P^{-1}_{in} &= P_{in} P^{-1}_{out} P_{in} = 0.
\end{align*}
\]

There are Eqs. (A1), Eqs. (A2) and Eqs. (A3) are valid. Using $\mathcal{O} P_{in} = P_{in} \mathcal{O} P_{in}$ and $P_{out} \mathcal{O} = P_{out} \mathcal{O} P_{out}$, we have

\[
\begin{align*}
P_{out} \mathcal{O} P_{in} &= P_{out} P_{out} \mathcal{O} P_{in} = P_{out} P_{out} \mathcal{O} P_{in} \\
&= P_{out} \hat{P}_{in} P_{out} = P_{out} \hat{P}_{in} P_{out} \\
&= P_{out} P_{out} \mathcal{O} P_{in} \hat{P}_{in} = P_{out} P_{out} \mathcal{O} \hat{P}_{in} \hat{P}_{in} \\
&= P_{out} \hat{P}_{in} \mathcal{O} P_{in} \hat{P}_{in}.
\end{align*}
\]

Similarly,

\[
\begin{align*}
P_{out} \mathcal{O} P_{in} &= P_{out} P_{out} \mathcal{O} P_{in} = \hat{P}_{out} P_{out} \mathcal{O} P_{in} \\
&= \hat{P}_{out} \mathcal{O} P_{out} \hat{P}_{in} = \hat{P}_{out} \mathcal{O} \hat{P}_{out} \hat{P}_{in} \\
&= \hat{P}_{out} \mathcal{O} \hat{P}_{out} P_{in} = \hat{P}_{out} \mathcal{O} \hat{P}_{out} P_{in}.
\end{align*}
\]

Therefore, Eq. (A4) is valid. \(\square\)

**Theorem 1.** Let $|\sigma\rangle \in V_{in}$, $|\langle H| \in V_{out}$ and $O \in \mathcal{O}$, where $i = 1, 2, \ldots, N$. Then,

\[
\langle\langle H|O_{N} \cdots O_{2} O_{1}\rangle\langle\sigma|\rangle = \langle\langle H|\hat{P}_{in} O_{N} \cdots \hat{P}_{in} O_{2} \hat{P}_{in} O_{1} \hat{P}_{in} |\sigma\rangle\rangle.
\]

**Proof.** Using Lemma 1, we have

\[
\begin{align*}
P_{out} \mathcal{O} \hat{P}_{out} P_{in} &= P_{out} \mathcal{O} \hat{P}_{out} P_{out} \mathcal{O} \hat{P}_{out} P_{in} \\
&= P_{out} \mathcal{O} \hat{P}_{out} \hat{P}_{in} P_{out} \mathcal{O} \hat{P}_{out} P_{in} \\
&= P_{out} \mathcal{O} \hat{P}_{out} \hat{P}_{in} P_{out} \mathcal{O} \hat{P}_{out} P_{in}.
\end{align*}
\]

Using $|\sigma\rangle = P_{in} |\sigma\rangle$ and $|\langle H| = (H| P_{out}$, we have

\[
\begin{align*}
\langle\langle H|O_{N} \cdots O_{2} O_{1}|\sigma\rangle\rangle = \langle\langle H|P_{out} O_{N} \cdots O_{2} O_{1} P_{in}|\sigma\rangle\rangle \\
= \langle\langle H|P_{out} O_{N} \cdots O_{2} O_{1} P_{in}|\sigma\rangle\rangle \\
= \langle\langle H|P_{out} O_{N} \cdots O_{2} O_{1} P_{in}|\sigma\rangle\rangle \\
= \langle\langle H|P_{out} O_{N} \cdots O_{2} O_{1} P_{in}|\sigma\rangle\rangle.
\end{align*}
\]
Similarly,
\[
\langle\langle H|O_N\cdots O_2 O_1|\sigma\rangle\rangle
\]
\[
= \langle\langle H|P_{\text{out}} O_N\cdots O_2 O_1|\rho_1\rangle\rangle
\]
\[
= \langle\langle H|P_{\text{out}} O_N P_{\text{in}} \cdots P_{\text{in}} O_2 P_{\text{in}} O_1 P_{\text{in}}|\sigma\rangle\rangle
\]
\[
= \langle\langle H|\hat{P}_{\text{out}} O_N P_{\text{in}} \cdots P_{\text{in}} O_2 P_{\text{in}} O_1 P_{\text{in}}|\sigma\rangle\rangle
\]
\[
= \langle\langle H|\hat{P}_{\text{out}} O_N \hat{P}_{\text{out}} \cdots \hat{P}_{\text{out}} O_2 \hat{P}_{\text{out}} O_1 \hat{P}_{\text{out}}|\sigma\rangle\rangle.
\] (A15)

\[\square\]

Theorem 2. Let \(d = \text{Tr}(P)\), and each of \(\{P|\rho_1\}\) and \(\langle\langle Q_k|P\rangle\rangle\) be a set of \(d\)-linearly-independent vectors. Then, \(g = M_{\text{out}} M_{\text{in}} = M_{\text{out}} P M_{\text{in}}\) is invertible, \(\hat{O} = M_{\text{out}} O M_{\text{in}} = M_{\text{out}} P O P M_{\text{in}}\), and
\[
C = M_{\text{out}} O N \cdots O_2 O_1 M_{\text{in}}
\]
\[
= M_{\text{out}} P O N \cdots O_2 P O_1 P M_{\text{in}}
\]
\[
= \hat{O} N g^{-1} \cdots O_2 g^{-1} \hat{O}_1.
\] (A16)

**Proof.** We remark that \(P = \hat{P}_{\text{out}}\), and the theorem is also valid for \(\hat{P}_{\text{in}}\).

According to definitions of \(M_{\text{out}}\) and \(M_{\text{in}}\), we have \(g_{k,i} = \langle\langle Q_k|\rho_i\rangle\rangle\). Because \(|\rho_i\rangle\rangle \in V_{\text{in}}\) and \(\langle\langle Q_k|\in V_{\text{out}}, \) we have \(g_{k,i} = \langle\langle Q_k|\hat{P}_{\text{out}}|\rho_i\rangle\rangle = \langle\langle Q_k|\hat{P}_{\text{out}}|\rho_i\rangle\rangle\). Here, we have used Theorem 1. Therefore, \(M_{\text{out}} M_{\text{in}} = M_{\text{out}} P M_{\text{in}}\).

Similarly, we have \(\hat{O}_{k,i} = \langle\langle Q_k|O|\rho_i\rangle\rangle = \langle\langle Q_k|\hat{P}_{\text{out}} O|\rho_i\rangle\rangle\). Therefore, \(M_{\text{out}} O M_{\text{in}} = M_{\text{out}} P O P M_{\text{in}}\).

We also have
\[
C_{k,i} = \langle\langle Q_k|O_N \cdots O_2 O_1|\rho_i\rangle\rangle
\]
\[
= \langle\langle Q_k|\hat{P}_{\text{out}} O_N \hat{P}_{\text{out}} \cdots \hat{P}_{\text{out}} O_2 \hat{P}_{\text{out}} O_1 \hat{P}_{\text{out}}|\rho_i\rangle\rangle.
\] (A17)

Therefore, the first two lines of Eq. (A16) are equal.

\(P M_{\text{in}}\) is a full rank \(d_{H}^2 \times \text{Tr}(P)\) matrix, and \(M_{\text{out}} P\) is a full rank \(\text{Tr}(P) \times d_{H}^2\) matrix. Thus, \((P M_{\text{in}})^{+} P M_{\text{in}} = I\), \(P M_{\text{in}}(P M_{\text{in}})^{+} = P\), \(M_{\text{out}} P (M_{\text{out}} P)^{+} = I\) and \((M_{\text{out}} P)^{+} M_{\text{out}} P = P\). Here, \(A^{+}\) denotes the pseudo inverse of matrix \(A\).

Using pseudo inverses, we have \(g^{-1} = (P M_{\text{in}})^{+} (M_{\text{out}} P)^{+}\), i.e. \(g = M_{\text{out}} P P M_{\text{in}}\) is invertible. Thus,
\[
\hat{O} N g^{-1} \cdots O_2 g^{-1} \hat{O}_1
\]
\[
= M_{\text{out}} P O N P M_{\text{in}}(P M_{\text{in}})^{+} (M_{\text{out}} P)^{+} \cdots
\]
\[
\times M_{\text{out}} P O_2 P M_{\text{in}}(P M_{\text{in}})^{+} (M_{\text{out}} P)^{+} M_{\text{out}} P O_1 P M_{\text{in}}
\]
\[
= M_{\text{out}} P O N P \cdots O_2 P O_1 P M_{\text{in}}.
\] (A18)

Therefore, the last two lines of Eq. (A16) are equal. \(\square\)

**Appendix B: Space dimension truncation**

We use \(\|\cdot\|\) to denote a vector norm satisfying \(||\langle\langle A|B\rangle\rangle\| \leq ||\langle\langle A\|\|\langle\langle B\rangle\rangle\||\text{ and the submultiplicative matrix norm induced by the vector norm, i.e. } ||\langle\langle O|B\rangle\rangle|| \leq ||\langle\langle O\|\|\langle\langle B\rangle\rangle\||\text{ and }||O_{1}\|\|O_{2}\|| \leq ||\langle\langle O_{1}\|\|O_{2}\||\text{.}

Two examples of such norms. First, we can take \(||\langle\langle B\rangle\rangle|| = \sqrt{\langle\langle B|B\rangle\rangle} = \text{Tr}(B^2).\) Then, \(||\langle\langle B\rangle\rangle|| = \sqrt{\sum_{i} \sigma_{i}^2}, \) where \(\sigma_{i}\) are singular values of \(B\). Second, we can take \(||\langle\langle B\rangle\rangle|| = ||B|| = \sum_{i} \sigma_{i} \geq \sqrt{\sum_{i} \sigma_{i}^2}, \) where \(||\cdot\||\text{ denotes the trace norm.}

We use \(||\cdot||_{\text{max}}\) to denote the max norm.

**Lemma 2.** Let \(N_{Q} \geq ||\langle\langle Q_{k}\rangle\rangle||\) for all \(k\), and \(N_{P} \geq ||\langle\langle \rho_{i}\rangle\rangle||\) for all \(i\). Then
\[
||M_{\text{out}} O_{N} \cdots O_{1} M_{\text{in}}||_{\text{max}} \leq N_{Q} N_{P} \prod_{j=1}^{N} ||O_{j}||. \] (B1)

**Theorem 3.** Let \(N_{Q} \geq ||\langle\langle Q_{k}\rangle\rangle||\) for all \(k\), and \(N_{P} \geq ||\langle\langle \rho_{i}\rangle\rangle||\) for all \(i\). Then, for any sequence of operations,
\[
||M_{\text{out}} O_{1} M_{\text{in}} - M_{\text{out}} O_{m+1} P_{m} M_{\text{in}}||_{\text{max}}\]
\[
\leq N_{Q} N_{P} \prod_{j=m+1}^{N} ||O_{j}|| \times \left[ \prod_{j=1}^{n} \left(||O_{j}|| + ||\delta O_{j}|| \right) - \prod_{j=1}^{n} ||O_{j}|| \right], \] (B2)

where
\[
\delta O = O_{m} O_{m+1} O_{m+2} \cdots O_{n} O_{n+1}, \]
\[
O_{m} = O_{N} \cdots O_{m+1} O_{m}, \]
\[
P_{m} = O_{m} O_{m+1} \cdots O_{n} O_{2} O_{1} O_{1} O_{m}. \] (B5)

**Proof.** Inequality (B2) holds for \(n = 0\), because
\[
||M_{\text{out}} O_{1} M_{\text{in}} - M_{\text{out}} O_{m} P_{m} M_{\text{in}}||_{\text{max}} = 0. \] (B6)

If inequality (B2) holds for \(n = m\), then it also holds for \(n = m + 1\). Now we assume that inequality (B2) holds for \(n = m\).

Because
\[
||M_{\text{out}} O_{1} M_{\text{in}}||_{\text{max}} \leq N_{Q} N_{P} \prod_{j=1}^{N} ||O_{j}||, \] (B7)

we have
\[
||M_{\text{out}} O_{m+1} O_{m+2} \cdots O_{m+1} M_{\text{in}}||_{\text{max}}\]
\[
\leq N_{Q} N_{P} \prod_{j=m+1}^{N} ||O_{j}|| \times \prod_{j=1}^{m} \left(||O_{j}|| + ||\delta O_{j}|| \right), \] (B8)

Because
\[
M_{\text{out}} O_{m+1} O_{m+2} \cdots O_{m+1} M_{\text{in}} = M_{\text{out}} O_{m+1} O_{m+2} \cdots O_{m+1} M_{\text{in}} + M_{\text{out}} O_{m+1} O_{m+2} \cdots O_{m+1} M_{\text{in}} , \] (B9)
we have
\[
\|M_{\text{out}} \overline{O}_1 M_{\text{in}} - M_{\text{out}} \overline{O}_{m+2} P_{m+1} M_{\text{in}}\|_{\text{max}} \\
\leq N_Q N_{\rho} \prod_{j=m+2}^{N} \|O_j\| \|\delta_{O_{m+1}}\| \prod_{j=1}^{m} (\|O_j\| + ||\delta_{O_j}||) \\
+ N_Q N_{\rho} \prod_{j=m+1}^{N} \|O_j\| \\
\times \left[ \prod_{j=1}^{m+1} (\|O_j\| + ||\delta_{O_j}||) - \prod_{j=1}^{m} \|O_j\| \right] \\
= N_Q N_{\rho} \prod_{j=m+2}^{N} \|O_j\| \\
\times \left[ \prod_{j=1}^{m+1} (\|O_j\| + ||\delta_{O_j}||) - \prod_{j=1}^{m+1} \|O_j\| \right], \tag{B10}
\]
i.e. inequality (B2) holds for \( n = m + 1 \).

\[\square\]

Appendix C: Linear inversion method

**Theorem 4.** \( \{l_i^n\} \) are columns of \( M_{\text{in}} \) and \( \{\{Q_k^n\}\} \) are rows of \( M_{\text{out}} \). Let \( N_Q \geq \|\{Q_k^n\}\| \) for all \( k \), and \( N_{\rho} \geq \|\{l_i^n\}\| \) for all \( i \). If \( g, M^a_{\text{in}} \) and \( M^a_{\text{out}} \) are inevitable, for any sequence of operations in \( \{O(\chi)\} \),

\[
\|\tilde{O}(\chi_1)g^{-1} \cdots \tilde{O}(\chi_2)g^{-1} \tilde{O}(\chi_1) - M^a_{\text{out}} O^a(\chi_1)M^a_{\text{in}}\|_{\text{max}} \\
\leq N^a_Q N^a_{\rho} \left[ \prod_{j=1}^{N} \|O^a(\chi_j)\| + ||\delta_{\chi_j}|| \right] \\
- \prod_{j=1}^{N} \|O^a(\chi_j)\|, \tag{C1}
\]

where
\[
\delta_g = M^a_{\text{in}}g^{-1}M^a_{\text{out}} - \mathbb{I}, \tag{C2}
\]
\[
\delta_{\chi} = (M^a_{\text{out}})^{-1} \tilde{O}(\chi)(M^a_{\text{in}})^{-1} - O^a(\chi). \tag{C3}
\]

**Proof.** We have
\[
g^{-1} = (M^a_{\text{in}})^{-1}(\mathbb{I} + \delta_g)(M^a_{\text{out}})^{-1}, \tag{C4}
\]
\[
\tilde{O}(\chi) = M^a_{\text{out}} [O^a(\chi) + \delta_{\chi}] M^a_{\text{in}}. \tag{C5}
\]

Then,
\[
\tilde{O}(\chi_1)g^{-1} \cdots \tilde{O}(\chi_2)g^{-1} \tilde{O}(\chi_1) = M^a_{\text{out}} [O^a(\chi_1) + \delta_{\chi_1}] \cdots \cdot [O^a(\chi_2) + \delta_{\chi_2}] \cdots \cdot [O^a(\chi_1) + \delta_{\chi_1}] M^a_{\text{in}}. \tag{C6}
\]

Therefore, inequality (C1) holds. \[\square\]

We now apply Theorem 4 to the approximate model given by the approximate invariant subspace \( \Pi_{\text{in}} \). Let \( \{l_i\} \) \( i = 1, 2, \ldots, d \) be an orthonormal basis of the subspace \( \Pi_{\text{in}} \), i.e. \( \{l_i\Pi_{\text{in}}\} = \delta_{l_1}, \ldots, \delta_{l_d} \) and \( \Pi_{\text{in}} = \sum_{l=1}^{d} |l\rangle \langle l| \)\( |l\rangle \langle l| \) \( l = 1, 2, \ldots, d \) be an orthonormal basis of the approximate-model space, i.e. \( \{l_i\overline{p}\} = \delta_{l_1}, \ldots, \delta_{l_d} \) and \( \overline{p} = \sum_{l=1}^{d} |l\rangle \langle l| \). Then, \( T = \sum_{l=1}^{d} |l\rangle \langle l| \) is the transformation from the actual space to the approximate-model space, and \( T^+ = \sum_{l=1}^{d} |l\rangle \langle l| \) is the inverse transformation. We have \( TT^+ = \mathbb{I} \) and \( T^+T = \Pi_{\text{in}} \). The approximate model is given by
\[
M^a_{\text{in}} = TM_{\text{in}}, \tag{C7}
\]
\[
M^a_{\text{out}} = MT^+, \tag{C8}
\]
\[
O^a(\chi) = TO(\chi)T+. \tag{C9}
\]

We take the vector norm in the approximate-model space \( \|\{A^n\}\| = \|\{(A^nT^\dagger\|^2) \) \( \|\{B^n\}\| = \|T^+\|B^n\| \). Then, \( \|\{A^n\}^2\| = \|\{(A^nT^\dagger\|^2) \) \( \|\{B^n\}^2\| = \|\|B^n\|^2\| \) is satisfied. We have
\[
\|O^a(\chi)\|^2 = \|T^+O^a(\chi)T^+|B^n\|^2\|, \tag{C10}
\]
Because
\[
\|T^+O^a(\chi)T^+|B^n\|^2\| \leq \|T^+O^aT^+|T^+|B^n\|^2\| \\
= \|T^+O^aT^+\||B^n\|^2\|, \tag{C11}
\]
we have
\[
\|O^a(\chi)\| \leq \|T^+O^aT\|. \tag{C12}
\]

Let \( O \) be an operation satisfying \( \Pi_{\text{in}}O\Pi_{\text{in}} = T^+O^aT \), we have
\[
\|T^+O^aT^+\| \leq \|\Pi_{\text{in}}O\Pi_{\text{in}}T^+\|B^n\|^2\| \\
= \|\Pi_{\text{in}}O\Pi_{\text{in}}T^+|B^n\|^2\| \\
\leq \|\Pi_{\text{in}}O\Pi_{\text{in}}T^+|\|O|\|\|\Pi_{\text{in}}O\Pi_{\text{in}}T^+|\|B^n\|^2\| \\
= \|\Pi_{\text{in}}O\Pi_{\text{in}}\||B^n\|^2\| \\
\leq \|\Pi_{\text{in}}O\Pi_{\text{in}}\||B^n\|^2\|, \tag{C13}
\]
where
\[
\delta_O = \Pi_{\text{in}}O\Pi_{\text{in}} - O\Pi_{\text{in}}. \tag{C14}
\]

Therefore, \( \|O^a(\chi)\| \leq \|\Pi_{\text{in}}O(\chi)\Pi_{\text{in}}\| + \|\delta_O(\chi)\| \). Because \( T^+O^a(\chi)T = \Pi_{\text{in}}O(\chi)\Pi_{\text{in}} \), we have \( \|O^a(\chi)\| \leq \|O(\chi)\| + \|\delta_O(\chi)\| \).

We have,
\[
\delta_g = TM_{\text{in}}g^{-1}M_{\text{out}}T^+ - \mathbb{I} = T(M_{\text{in}}g^{-1}M_{\text{out}}\Pi_{\text{in}} - \Pi_{\text{in}})T^+. \tag{C15}
\]

Because \( g \) is invertible, \( M_{\text{in}} \) and \( M_{\text{out}} \) are full rank. Thus, \( M_{\text{in}}^+M_{\text{in}} = \mathbb{I} \), \( M_{\text{out}}M_{\text{out}} = \Pi_{\text{in}} \), \( M_{\text{in}}^+M_{\text{out}} = \mathbb{I} \) and \( M_{\text{out}}^+M_{\text{out}} = \Pi_{\text{out}} \). Then,
\[
\Pi_{\text{in}}g^{-1}M_{\text{out}} = \Pi_{\text{in}}g^{-1}M_{\text{out}}M_{\text{in}}M_{\text{in}}^+ \\
= g^{-1}gM_{\text{in}}^+M_{\text{in}} = \Pi_{\text{in}}. \tag{C16}
\]

Therefore, \( \delta_g = 0 \).
Because \( g = M_{\text{out}} \Pi_{\text{in}} M_{\text{in}} \) is invertible, \( M_{\text{out}} \Pi_{\text{in}} \) is full rank. Thus, \( M_{\text{out}} \Pi_{\text{in}} (M_{\text{out}} \Pi_{\text{in}})^+ = \mathbb{1} \) and \( (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \Pi_{\text{in}} = \Pi_{\text{in}} \). Then, we have \((M_{\text{out}} T^+)^{-1} = T(M_{\text{out}} \Pi_{\text{in}})^+ \) and \((TM_{\text{in}})^{-1} = M_{\text{in}} T^+ \). Therefore,

\[
\delta_\mathbf{\chi} = (M_{\text{out}} T^+)^{-1} \tilde{O}(\mathbf{\chi}) (TM_{\text{in}})^{-1} - T \mathcal{O}(\mathbf{\chi}) T^+ = T \left[ (M_{\text{out}} \Pi_{\text{in}})^+ \tilde{O}(\mathbf{\chi}) M_{\text{in}}^+ - \Pi_{\text{in}} \mathcal{O}(\mathbf{\chi}) \Pi_{\text{in}} \right] T^+. \quad \text{(C17)}
\]

We have

\[
(M_{\text{out}} \Pi_{\text{in}})^+ \tilde{O}(\mathbf{\chi}) M_{\text{in}}^+ = (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \mathcal{O}(\mathbf{\chi}) M_{\text{in}} M_{\text{in}}^+ = (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \mathcal{O}(\mathbf{\chi}) \Pi_{\text{in}} = (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \Pi_{\text{in}} \mathcal{O}(\mathbf{\chi}) \Pi_{\text{in}} - (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \delta_{\mathcal{O}(\mathbf{\chi})} = \Pi_{\text{in}} \mathcal{O}(\mathbf{\chi}) \Pi_{\text{in}} - (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \delta_{\mathcal{O}(\mathbf{\chi})}. \quad \text{(C18)}
\]

Then,

\[
\delta_\mathbf{\chi} = -T(M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \delta_{\mathcal{O}(\mathbf{\chi})} T^+. \quad \text{(C19)}
\]

Let \( G = \Pi_{\text{in}} (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \), we have

\[
\Pi_{\text{in}} G = G, \quad \text{(C20)}
\]

\[
G \Pi_{\text{out}} = G, \quad \text{(C21)}
\]

\[
G \Pi_{\text{in}} = \Pi_{\text{in}} (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \Pi_{\text{in}} = \Pi_{\text{in}} \text{, (C22)}
\]

\[
\Pi_{\text{out}} G = \Pi_{\text{out}} (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} = M_{\text{out}}^+ M_{\text{out}} \Pi_{\text{in}} (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} = M_{\text{out}}^+ M_{\text{out}} = \Pi_{\text{out}} \text{. (C23)}
\]

Then,

\[
(\mathbb{1} - \Pi_{\text{in}} + \Pi_{\text{out}}) G = \Pi_{\text{out}}. \quad \text{(C24)}
\]

We define \( \delta_{\Pi} = \Pi_{\text{in}} - \Pi_{\text{out}} \). If \( \mathbb{1} - \delta_{\Pi} \) is invertible, we have

\[
G = (\mathbb{1} - \delta_{\Pi})^{-1} \Pi_{\text{out}}. \quad \text{(C25)}
\]

Then,

\[
T^+ \delta_\mathbf{\chi} T = -\Pi_{\text{in}} (M_{\text{out}} \Pi_{\text{in}})^+ M_{\text{out}} \delta_{\mathcal{O}(\mathbf{\chi})} \Pi_{\text{in}} = -G \delta_{\mathcal{O}(\mathbf{\chi})} = -(\mathbb{1} - \delta_{\Pi})^{-1} \Pi_{\text{out}} \delta_{\mathcal{O}(\mathbf{\chi})} = (\mathbb{1} - \delta_{\Pi})^{-1} \Pi_{\text{in}} \delta_{\mathcal{O}(\mathbf{\chi})}, \quad \text{(C26)}
\]

Therefore,

\[
||\delta_\mathbf{\chi}|| \leq ||T^+ \delta_\mathbf{\chi} T|| = ||(\mathbb{1} - \delta_{\Pi})^{-1} \Pi_{\text{in}} \delta_{\mathcal{O}(\mathbf{\chi})}|| \leq (1 - ||\delta_{\Pi}||)^{-1} ||\Pi_{\text{in}}|| \parallel \delta_{\mathcal{O}(\mathbf{\chi})} \parallel. \quad \text{(C27)}
\]

Using inequality (C1), we have

\[
||\mathcal{O} (\mathbf{\chi}) g^{-1} \cdots \mathcal{O} (\mathbf{\chi}) g^{-1} \tilde{\mathcal{O}} (\mathbf{\chi})|| - M_{\text{out}}^a \mathcal{O}^a(\mathbf{\chi}) \cdots \mathcal{O}^a(\mathbf{\chi}) M_{\text{in}}^b \leq N_Q \mathcal{O} \left( \prod_{j=1}^{N} \left( ||\mathcal{O} (\mathbf{\chi})|| + ||\delta_{\mathcal{O}(\mathbf{\chi})}|| \right) \right) + \left( 1 - ||\delta_{\Pi}|| \right)^{-1} ||\Pi_{\text{in}}|| \parallel \delta_{\mathcal{O}(\mathbf{\chi})} \parallel \leq N_Q N_P \left\{ \prod_{j=1}^{N} \left( ||\mathcal{O} (\mathbf{\chi})|| + ||\delta_{\mathcal{O}(\mathbf{\chi})}|| \right) \right\}, \quad \text{(C28)}
\]

where

\[
N_Q = \max \{ ||\langle \langle Q_s \parallel \mathbf{|}\rangle \rangle || \} = \max \{ ||\langle \langle Q_k \parallel \Pi_{\text{in}} || \rangle \rangle || \}
\]

and

\[
N_P = \max \{ ||\langle \langle \rho_s \parallel \mathbf{|}\rangle \rangle || \} = \max \{ ||\langle \langle \rho_t \parallel \Pi_{\text{in}} || \rangle \rangle || \}.
\]

**Appendix D: Vector space dimensions**

If Hilbert spaces of the system and environment are respectively \( d_S \)-dimensional and \( d_E \)-dimensional, the Hilbert space of SE is \((d_S d_E)^2\)-dimensional. Then, a column vector \( |\rho\rangle \) representing the state of SE is \((d_S^2 d_E^2)^-\)dimensional. We remark that \( d_E = m \).

For the classical random variable noise, the state is in the form \( \rho = \sum_{\lambda} \rho_{\lambda} (\mathbf{1} \otimes |\lambda\rangle \langle \lambda|) \). \( \rho_{\lambda} \) is the state of the environment (in the reduced density matrix form) only has diagonal elements. Therefore, we can use a \((d_E^2)^\perp\)-dimensional vector to represent the state, i.e. take \(|\rho\rangle = \sum_{\lambda} \rho_{\lambda} (|\lambda\rangle \langle \lambda|) s \otimes |\lambda\rangle \rangle \).\( s \) and \(|\lambda\rangle \rangle \) are column vectors representing states of the system, and \(|\lambda\rangle \rangle \) are column vectors representing states of the environment.

The state of the system can always be expressed in the form \( \rho_S = d_S^{-1} 1_S + \rho_S^\perp \), where \( \text{Tr} (\rho_S^\perp) = 0 \). Then \(|\rho_S^\perp \rangle \rangle = |\lambda\rangle \rangle + |\rho_S^\perp \rangle \rangle \) corresponds to \(|\lambda\rangle \rangle \) and \(|\rho_S^\perp \rangle \rangle \) corresponding to \(|\lambda\rangle \rangle \). The orthogonal projection on this subspace is \( P_H = |\lambda\rangle \rangle (|\lambda\rangle \rangle \langle \lambda|) \). If the distribution of \( \lambda \) is stationary, i.e. \( \{ \rho_{\lambda} \} \) are invariant under operations, we have \( P_H |\rho\rangle = P_H |\rho\rangle \) for any operation \( O \) that does not change the distribution. Therefore, if the distribution of \( \lambda \) is stationary, \( P_H |\rho\rangle \) is the only non-trivial vector in the subspace \( P_H \) that contributes to the state, and \(|\rho\rangle \) is effectively \((d_S^2 - 1) d_E + 1\)-dimensional.
Appendix E: Details of the numerical simulation

1. Actual quantum computer simulation

To simulate the behaviour of the actual quantum computer, we use the Gaussian cubature approximation to match up to the $9^{th}$-order moment, by taking $e^{-\lambda^2}$ instead of $\lambda$ as the random variable. Reducing the precision of the approximation and only matching up to the $7^{th}$-order moment, we find that the difference is negligible.

2. Linear inversion method simulation

In LIM simulation, trial states and observables are generated by gate sequences with up to 3 Hadamard and phase gates. Trial states include the state $|\rho_{in}\rangle$, two states in the form $\mathcal{O}(G_1)|\rho_{in}\rangle$, four states in the form $\mathcal{O}(G_2)\mathcal{O}(G_1)|\rho_{in}\rangle$, and eight states in the form $\mathcal{O}(G_3)\mathcal{O}(G_2)\mathcal{O}(G_1)|\rho_{in}\rangle$. Trial observables include the observable $\langle Q \rangle$, two observables in the form $\langle Q \rangle\mathcal{O}(G_1)$, four observables in the form $\langle Q \rangle\mathcal{O}(G_2)\mathcal{O}(G_1)$, and eight observables in the form $\langle Q \rangle\mathcal{O}(G_3)\mathcal{O}(G_2)\mathcal{O}(G_1)$. Here, $G_1, G_2, G_3 = H, S$.

In QEM based on the quasi-probability decomposition, we use states $\{|\rho_{i}\rangle\}$, four states in the form $\{\langle Q_{i}\rangle\}$ to decompose the initial state, and we use observables $\{\langle Q_{i}\rangle\}$ to decompose the measured observable. Here, $\{|\rho_{i}\rangle\}$ are columns of $M_{in}$, $\{\langle Q_{i}\rangle\}$ are rows of $M_{out}$. For the gate $G$, we use the corresponding operation with error $\bar{G}$ and measurement-initialisation operations $\{\mathcal{B}_{i,k}\}$ to decompose the gate. Because $\{|\rho_{i}\rangle\}$, $\{\langle Q_{i}\rangle\}$ and $\{\mathcal{B}_{i,k}\}$ are all complete, corresponding decomposition formulas always exist. Decomposition formulas are not unique. To determine decomposition formulas, we minimise weights, i.e. $w_{\rho}$, $w_{Q}$ and $w_{\chi=G}$ for each gate $G$. The invertible matrix $S$ is determined by minimising $w_{H}w_{S}$.

3. Maximum likelihood estimation simulation

Circuits for generating data $\{|C_{m}\rangle\}$ used in MLE are $\langle 0|R_{L}R_{R}|0\rangle$, $\langle 0|R_{L}HR_{R}|0\rangle$ and $\langle 0|R_{L}SR_{R}|0\rangle$, where $\langle 0|R_{L}$ and $R_{R}|0\rangle$ are gate sequences with up to 3 Hadamard and phase gates, as the same as in LIM simulation.

In QEM based on the quasi-probability decomposition, we take $\rho^{(0)} = \sum_{\lambda} P_{\lambda}|0\rangle\langle 0| \otimes |\lambda\rangle\langle \lambda|$, and $Q^{(0)} = \sum_{\lambda} |0\rangle\langle 0| \otimes |\lambda\rangle\langle \lambda|$. Therefore, we do not have to mitigate errors in the initial state and observable, because they are error-free. For a gate $G$, we take $\mathcal{O}^{(0)}(G) = \sum_{\lambda}\{G\} \otimes |\lambda\rangle\langle \lambda|_{E}$ as the gate without error.

Let $\mathcal{H} = \mathcal{O}(H)$ and $\mathcal{S} = \mathcal{O}(S)$ be operations with errors corresponding to gates $H$ and $S$, respectively. When $d = 4$, we use $B_{H} = \{\mathcal{H}, HSS, SSH, SSHSS\}$ to decompose the Hadamard gate $H$, and we use $B_{S} = \{S, SSS, SSHSS, SSHSSS\}$ to decompose the phase gate $S$. When $d = 7$, we use $B_{H}$ and $B_{H} = B_{H} \mathcal{H}$ to decompose the Hadamard gate, and we use $B_{S}$ and $B_{S}' = B_{S} \mathcal{H}$ to decompose the phase gate. Decomposition formulas always exist and are not unique. Weights $w_{H}$ and $w_{S}$ are minimised to determine decomposition formulas.

4. Error extrapolation

To fit probabilities in the ideal state $|0\rangle$, we use the fitting function $F(\eta) = a + \exp(b_{0} + b_{1}\eta + b_{2}\eta^{2} + b_{3}\eta^{3})$ to obtain the red curve in Fig. 3(a). To fit depolarizing rates, we use the fitting function $\epsilon(\eta) = a + b\eta$ to obtain red curves in Fig. 3(b).

[1] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, Cambridge, (2010).
[2] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiari, 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).
[3] X. Rong, J. Geng, F. Shi, Y. Liu, K. Xu, W. Ma, F. Kong, Z. Jiang, Y. Wu, and J. Du, Experimental fault-tolerant universal quantum gates with solid-state spins under ambient conditions, Nat. Commun. 6, 8748 (2015).
[4] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, High-fidelity quantum logic gates using trapped-ion hyperfine qubits, Phys. Rev. Lett. 117, 060504 (2016).
[5] J. P. Gaebler, T. R. Tan, Y. Lin, Y. Wan, R. Bowler, A. C. Keith, S. Glancy, K. Coakley, E. Knill, D. Leibfried, and D. J. Wineland, Superconducting quantum circuits at the surface code threshold for $^{9}$Be$^{+}$ ion qubits, Phys. Rev. Lett. 117, 060505 (2016).
[6] 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 tolerance threshold with gate set tomography, Nat. Commun. 8, 14485 (2017).
[7] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B Quantum Semiclassical Opt. 7, S347 (2005).
[8] 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).
[9] J. Poyatos, J. Cirac, P. Zoller, Complete characterization of a quantum process: the two-bit quantum gate, Phys.
FIG. 4. The difference between probabilities in the ideal state $|0\rangle$ after a sequence of randomly chosen gates. (a,b) The probability difference $F_{\text{tom}} - F_{\text{act}}$, where $F_{\text{act}}$ is the probability obtained in the actual quantum computing without quantum error mitigation (QEM), and $F_{\text{tom}}$ is the probability estimated using linear operator tomography (LOT). (c,d) The probability difference $F_{\text{QEM}} - 1$, where $F_{\text{QEM}}$ is the probability obtained in the actual quantum computing with QEM based on the quasi-probability decomposition, and the probability should be 1 if the quantum computing has no error. The probability obtained with QEM based on the error extrapolation for all gate numbers are computed, and blue curve represents the difference from the ideal value.

Rev. Lett. 78, 390 (1997).
[10] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. 44, 2455 (1997).
[11] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, Surface code quantum computing with error rates over 1%, Phys. Rev. A 83, 020302(R) (2011).
[12] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, Comparing experiments to the fault-tolerance threshold, Phys. Rev. Lett. 117, 170502 (2016).
[13] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, arXiv:quant-ph/9906129
[14] M.-X. Huo and Y. Li, Learning time-dependent noise to reduce logical errors: real time error rate estimation in quantum error correction, New J. Phys. 19, 123032 (2017).
[15] J. Chiaverini, D. Leibfried, T. Schaetz, M. D. Barrett, R. B. Blakestad, J. Britton, W. M. Itano, J. D. Jost, E. Knill, C. Langer, R. Ozeri, and D. J. Wineland, Realization of quantum error correction, Nature 432, 602 (2004).
[16] P. Schindler, J. T. Barreiro, T. Monz, V. Nebendahl, D. Nigg, M. Chwalla, M. Henrich, and Rainer Blatt, Experimental repetitive quantum error correction, Science 332, 1059 (2011).
We use $\|\cdot\|$ to denote a vector norm satisfying $|\langle \langle A | B \rangle \rangle| \leq \|\langle \langle A \| \| B \rangle \rangle\|$ and the submultiplicative matrix norm induced by the vector norm. We can take $\|\cdot\rangle\rangle\| = \|\cdot\|_1$, where $\|\cdot\|_1$ is the trace norm. Then, $\|O\| = 1$ if $O$ is trace-preserving. $\|\cdot\|_{\text{max}}$ denotes max norm.

[50] A. C. Miller, III, and T. R. Rice, *Discrete approximations of probability distributions*, Management Science 29, 352 (1983).

[51] E. A. DeVuysta and P. V. Preckelb, *Gaussian cubature: A practitioner’s guide*, Mathematical and Computer Modelling 45, 787 (2007).