

Efficient separate quantification of state preparation errors and measurement errors on quantum computers and their mitigation

Hongye Yu^{1,2} and Tzu-Chieh Wei^{1,2}

¹C. N. Yang Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, NY 11794-3840, USA

²Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, NY 11794-3800, USA

Current noisy quantum computers have multiple types of errors, which can occur in the state preparation, measurement/readout, and gate operations, as well as intrinsic decoherence and relaxation. Partly motivated by the booming of intermediate-scale quantum processors, measurement and gate errors have recently been extensively studied, and several methods of mitigating them have been proposed and formulated in software packages (e.g., IBM Qiskit). Despite this, the state preparation error and the procedure to quantify it have not yet been standardized, as state preparation and measurement errors are usually considered to be not directly separable. Inspired by a recent work of Laflamme, Lin, and Mor [Phys. Rev. A 106, 012439 (2022)], we propose a simple and resource-efficient approach to quantify separately the state preparation and readout error rates. With these two errors separately quantified, we also propose methods to mitigate them separately, especially mitigating state preparation errors with linear (with the number of qubits) complexity. As a result of the separate mitigation, we show that the fidelity of the outcome can be improved by an order of magnitude compared to the standard measurement error mitigation scheme. We also show that the quantification and mitigation schemes are resilient against gate noises and can be immediately applied to current noisy quantum computers. To demonstrate this, we present results from cloud experiments on IBM's superconducting quantum computers. The results indicate that the state preparation error rate is also an important metric for qubit metrology that can be efficiently obtained.

1 Introduction

There has been dramatic growth in multi-qubit quantum processors; some possess qubits well over one hundred. Nevertheless, they are characterized as noisy-intermediate quantum (NISQ) devices [1] as it is currently not yet feasible to implement full-scale quantum error correction on these devices to reduce logical error rates and prolong coherent computation. Hence, a short-term goal is to develop error mitigation schemes to enhance the performance of NISQ devices. Errors and noise arise in different stages of quantum computation: state preparation, gate operation, and readout/measurement. In the past few years, substantial efforts have been spent on gate error mitigation [2, 3, 4, 5, 6, 7, 8], as gate operation is the bulk of quantum computation.

To mitigate readout/measurement errors, several methods have been proposed [9, 10, 11, 12, 13] to mitigate the final readout statistics, and have improved the estimation of physical observables from quantum computers. State preparation is perhaps the most omitted among the three types of errors. At present, there does not seem to be an efficient and practical approach to even quantifying state preparation and readout errors separately. Usually, the state preparation error and measurement error are considered inseparable and are combined as the state preparation and measurement (SPAM) error. Indeed, if we only have one qubit, prepared at $|0\rangle$ but measured with outcome $|1\rangle$, we cannot tell whether the outcome error comes from the stage in the measurement or the initial state preparation. Moreover, current measurement mitigation schemes on superconducting qubits, such as in the Qiskit software [14], are still mixed with the state preparation error. Although the rate of state-preparation error is typically lower than that of the readout error, for quantum computers with state-preparation error dominating, such mixed mitigation can have large systematic errors proportional to the error rate of the state prepa-

ration.

The goal of this work is two-fold. First, we propose a simple and efficient scheme that allows separate quantification of state preparation and measurement errors. This issue was recently discussed by Laflamme, Lin, and Mor [15] employing algorithmic cooling [16, 17, 18]. Our approach extends theirs, but dramatically simplifies the procedure. Both schemes rely on coupling the original qubit to other qubits. Ours requires only one additional qubit to characterize the errors in state preparation and measurement separately. Another approach to characterize gates and SPAM errors with very few assumptions is gate set tomography [19]; however, it requires many more gate sequences and is not feasible for large-system experiments, whereas ours can be easily implemented on current noisy quantum computers even for large numbers of qubits. Secondly, with separate quantification, the measurement error, in principle, is free of the state preparation error, and the current standard mitigation scheme can be directly applied to mitigating the measurement error. Moreover, we propose a novel approach to mitigate the state-preparation error. The mitigation scheme is practical and requires an $O(N)$ overhead for first-order full-qubit mitigation. The scheme can also be done qubit-wisely, and thus the overhead can be greatly reduced if very few qubits (with relatively large state-preparation errors) need mitigation. The mitigation scheme can also be directly applied to mitigating expectation values of observables, not merely to the outcome probability distributions. Even though we focus our attention and implementation on superconducting qubits, we envision that our method can be applied to other systems as well.

We also remark that in this work, we assume that single-qubit gates, in particular, the X gate, are of high fidelity in the characterization circuits. The assumption is reasonable in practice. In current superconducting quantum computers [20], the errors of single-qubit gates are around 10^{-4} while the SPAM error of qubits can be 10^{-2} . We do not assume that the CNOT gate is noiseless, but we assume that existing error-mitigation techniques, such as zero-noise extrapolation (ZNE), can mitigate the CNOT gate error [2, 3, 4, 5, 6, 7] to sufficient accuracy if needed.

The remainder of the paper is organized as follows. In Sec. 2, we present a simple scheme using an ancilla to characterize the target qubit's state preparation and measurement errors separately. In Sec. 3, we discuss how the characterized state preparation and readout errors can be separately mitigated. In Sec. 4, we compare our proposed mitigation scheme with the standard readout mitigation (which does not characterize the state preparation error) and illustrate the difference in mitigated outcomes for simple circuits. To showcase the

practicalness of our separate characterization and mitigation schemes, we present results from experiments performed on cloud IBM quantum computers to demonstrate our theoretical proposal in Sec. 2 and 4. We make concluding remarks in Sec. 5.

2 Separate characterization of state preparation and readout errors

A single-qubit readout can be characterized by a two-element POVM $\{M_0, M_1\}$, so that a one-qubit state ρ will be measured to be '0' with probability $\text{Tr}(\rho M_0)$ and '1' with probability $\text{Tr}(\rho M_1)$. The readout errors in practice are dominantly classical flips, i.e., the POVM elements are well approximated by the following forms,

$$M_0 = \begin{pmatrix} 1 - \delta_M^0 & 0 \\ 0 & \delta_M^1 \end{pmatrix}, \quad M_1 = \begin{pmatrix} \delta_M^0 & 0 \\ 0 & 1 - \delta_M^1 \end{pmatrix}, \quad (1)$$

where the parameter δ_M^0 is the probability that the ideal '0' state will be read as '1', and δ_M^1 is the probability that the ideal '1' will be read as '0'; see, e.g., Ref. [9], which also demonstrated this on superconducting qubits of IBM's devices. We note that the POVM for the general single-qubit readout can be more general than the above form, but it can be enforced to the above form by twirling with a random Pauli Z operator (with 50% probability) before the measurement [9, 15].

In Ref. [15], a measurement-based algorithmic cooling was developed to reduce the state-preparation error, where the error ruins the preparation of a perfect $|0\rangle$, i.e.,

$$\rho_0^{[\text{err}]} \equiv (1 - \delta_{\text{SP}})|0\rangle\langle 0| + \delta_{\text{SP}}|1\rangle\langle 1|, \quad (2)$$

and the cooling was achieved by coupling to ancillary qubits via CNOTs. There, symmetric readout errors were considered, i.e., $\delta_M^0 = \delta_M^1 =: \delta_M$ and a procedure employing multiple ancillary qubits was proposed to separately characterize a target qubit's δ_{SP} and δ_M (in the limit of many ancillas), which comprises the total SPAM error $\delta_{\text{SPAM}} = \delta_{\text{SP}} + \delta_M - 2\delta_{\text{SP}}\delta_M$, giving the measurable probability of reading out '1'. The above result can be easily derived from $\delta_{\text{SPAM}} = \text{Tr}(M_1 \rho_0^{[\text{err}]})$ with symmetric error in M_i 's. We give further detailed discussions related to the algorithmic cooling in Appendix A. Naively, one might have expected that the total contribution of the SPAM error is simply the sum of two errors. However, it is important to note that the total SPAM error does contain the crossed term $-2\delta_{\text{SP}}\delta_M$ that reduces the sum of the two. Thus, it is clear that *using δ_{SPAM} to perform readout mitigation will contain the error from state preparation*, which, however, is the current standard approach.

The above discussion, therefore, motivates us to propose a simplified procedure in this section, using one ancilla to extract the state-preparation error δ_{SP} from readout errors δ_M^0 and δ_M^1 separately. First, we note that a coherent state preparation error, e.g., $\sqrt{1-\delta_{\text{SP}}}|0\rangle + \sqrt{\delta_{\text{SP}}}e^{i\phi}|1\rangle$, can be turned into the above incoherent state preparation error (2) by twirling with a Pauli Z operator randomly with 50% probability [15]. We note that the coherent error can be probed using quantum state tomography once the readout errors have been characterized and mitigated; see Sec. 2.3.2 below. We will not assume the symmetric readout errors but actively employ the twirling to ensure the incoherent form of state-preparation error. The parameters δ_M^0 and δ_M^1 are reported on the device properties of all IBM quantum computers, except that the reported values are mixed with the state preparation error. Can the true readout errors be obtained experimentally and separately from the state preparation error? Indeed, Ref. [15] describes an approach by algorithmic cooling to reduce the state preparation error, and in the limit that it becomes infinitesimally small, only the readout errors remain. However, the reduction to zero state preparation error is not practical as the required CNOT gates (which were assumed to be error-free in Ref. [15]) will introduce additional errors in practice. Our procedure below solves all these problems and can practically improve the device property characterization, separating the state preparation and readout errors.

2.1 Access δ_{SP} via noiseless CNOT gates

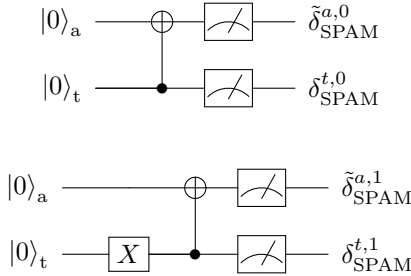


Figure 1: Circuits for measuring δ_{SP} with initial state prepared to be: (top) $|00\rangle$ and (bottom) $|01\rangle$. Here we assume the X gate is noiseless and can be considered as a part of state preparation.

Here we present an efficient way to estimate the state-preparation error δ_{SP}^t for the target qubit q_t , which only requires one additional qubit q_a , as shown in the Fig. 1. We remark that such circuits are similar to a repetition code, which has been applied to suppressing both state-preparation errors [15] and readout errors [21, 22]. In

this work, the circuits are only applied to the quantification process, not the active suppression process.

In these circuits, the SPAM error of both qubits can be easily measured. To measure the probability δ_{SPAM}^0 that a qubit is prepared with $|0\rangle$ but measured in $|1\rangle$, one only needs to measure the initialized qubit directly; to measure the probability δ_{SPAM}^1 that the qubit is prepared with $|1\rangle$ but measured in $|0\rangle$, one needs to first initialize the qubit in $|0\rangle$, apply an X gate, and then measure it. As emphasized above, the measured error δ_{SPAM} is a combination of both state-preparation error δ_{SP} and measurement errors δ_M 's, and, without assuming symmetric readout errors, we have $\delta_{\text{SPAM}}^0 = \delta_M^0(1 - \delta_{\text{SP}}) + (1 - \delta_M^1)\delta_{\text{SP}}$. Combined with similar derivations for δ_{SPAM}^1 , we have

$$\begin{aligned}\delta_{\text{SPAM}}^0 &= (1 - \delta_M^0 - \delta_M^1)\delta_{\text{SP}} + \delta_M^0, \\ \delta_{\text{SPAM}}^1 &= (1 - \delta_M^0 - \delta_M^1)\delta_{\text{SP}} + \delta_M^1,\end{aligned}\quad (3)$$

or equivalently,

$$A_{\text{SPAM}} = A_M A_{\text{SP}}, \quad (4)$$

where A 's are the corresponding error assignment matrices (see Appendix B for explicit definitions and derivations)

$$A \equiv \begin{pmatrix} 1 - \delta^0 & \delta^1 \\ \delta^0 & 1 - \delta^1 \end{pmatrix}. \quad (5)$$

After adding a noiseless CNOT gate (Fig. 1), the SPAM error of q_t does not change, whereas the ancilla qubit readout error $\tilde{\delta}_{\text{SPAM}}^{a,0/1}$ becomes

$$\begin{aligned}\tilde{\delta}_{\text{SPAM}}^{a,0} &= (1 - \delta_{\text{SPAM}}^{a,0} - \delta_{\text{SPAM}}^{a,1})\delta_{\text{SP}}^t + \delta_{\text{SPAM}}^{a,0}, \\ \tilde{\delta}_{\text{SPAM}}^{a,1} &= (1 - \delta_{\text{SPAM}}^{a,0} - \delta_{\text{SPAM}}^{a,1})\delta_{\text{SP}}^t + \delta_{\text{SPAM}}^{a,1},\end{aligned}\quad (6)$$

where the superscript 0/1 indicates the ideal output and notation $\delta^{0/1}$'s indicate their error probabilities. The left-hand sides (emphasized by $\tilde{}$) are the results with a CNOT present, and the right-handed sides are results obtained in single-qubit circuits.

These relations can be derived as follows. From the top panel of Fig. 1, $\tilde{\delta}_{\text{SPAM}}^{a,0}$ is the probability of measuring '1' on qubit a , which has two contributions due to noisy preparation of qubit t : (a) With probably $(1 - \delta_{\text{SP}}^t)$ the qubit t is prepared in $|0\rangle$ and the CNOT has no effect; thus for qubit a , we are interested in the matrix element $(A_{\text{SPAM}}^a)_{10}$ (i.e., preparing '0' and reading out '1'). (b) With probably δ_{SP}^t the qubit t is prepared in $|1\rangle$ and the CNOT has the effect of flipping qubit a ; thus for qubit a , we are interested in the matrix element $(A_{\text{SPAM}}^a)_{11}$. This gives the first equation in Eq. (6) via: $\tilde{\delta}_{\text{SPAM}}^{a,0} = (1 - \delta_{\text{SP}}^t)(A_{\text{SPAM}}^a)_{10} + \delta_{\text{SP}}^t(A_{\text{SPAM}}^a)_{11}$. Likewise, we have $\tilde{\delta}_{\text{SPAM}}^{a,1} = (1 - \delta_{\text{SP}}^t)(A_{\text{SPAM}}^a)_{01} + \delta_{\text{SP}}^t(A_{\text{SPAM}}^a)_{00}$.

Actually, these arguments can be captured in the single matrix equation:

$$\tilde{A}_{\text{SPAM}}^a = A_{\text{SPAM}}^a A_{\text{SP}}^t; \quad (7)$$

see also Appendix B for slightly formal notation that streamlines the derivation.

By using either of the two equations in (6) and the measured values of $\delta_{\text{SPAM}}^{a,0/1}$, one can solve the δ_{SP}^t . Combined with the measured SPAM errors from q_t , one can further obtain the value of measurement errors $\delta_M^{0/1}$ for qubit q_t .

We note that solving δ_M 's and δ_{SP} here involves only linear equations. One can also extend the algorithmic cooling of Laflamme, Lin, and Mor in the two-qubit setting to solve for (the symmetric) δ_M and δ_{SP} , but this will involve non-linear equations; see Appendix A. In the next section, we expand our analysis to noisy CNOT gates.

2.2 Characterization of δ_{SP} with noisy CNOT gates

As the CNOT gate noise is inevitable in current noisy quantum devices, directly applying the characterization circuits (Fig. 1) to quantum computers can lead to inaccurate results, especially when the CNOT gate noise is much higher than the state-preparation errors.

2.2.1 Obtaining δ_{SP} by mitigating gate noise

Fortunately, there are already practical schemes, such as zero-noise extrapolation (ZNE) [6] and probabilistic error cancellation (PEC) [8], that can adequately mitigate the CNOT gate noise. Since we only have one CNOT gate in the characterization circuits, the accuracy of the results after mitigation is typically below the state-preparation errors, thus one can still obtain relatively accurate δ_{SP} 's. We will discuss the effects and additional errors on our δ_{SP} -characterization scheme from the gate-noise mitigation methods with more details in 2.4.

Therefore, the δ_{SP} -characterization scheme can still work on noisy devices with an additional gate-noise mitigation procedure, although such mitigation procedure typically requires additional computational resources. Given the ZNE and PEC procedures have been incorporated in Qiskit [14], below we will present the results on real devices using these mitigation schemes. In the next section, we will introduce a new technique that can efficiently mitigate the CNOT gate noise without additional computational resources. The technique only works for specific circuits on noisy devices with symmetric error channels for Pauli X and Z types. Most noise

in current superconducting qubits is dominated by depolarizing noise [23], and thus the symmetric condition approximately holds in most superconducting quantum computers. But in any case, one can twirl the CNOT gate so that its error channel becomes a Pauli channel.

2.2.2 Efficient mitigation scheme for symmetric CNOT noise

In this section, we show that the gate-noise mitigation procedure can be greatly simplified if the noise can be described by uncorrelated single-qubit Pauli noise channels and with some symmetry. We remark that for sufficiently large correlated errors, general gate error mitigation (e.g., ZNE and PEC) may be a more practical approach.

Suppose the noise of the CNOT can be described by a noise channel Λ after the CNOT gate, which we assume is a direct product of two Pauli-like single-qubit channels

$$\Lambda(\rho) = \sum_{C \in \{I, X, Y, Z\}} p_C C(\rho) C. \quad (8)$$

By adding a random Pauli Z operator with 50% probability for each qubit before the CNOT gate in Fig. 1, we ensure that the state-preparation error produces effectively $\rho = (1 - \delta_{\text{SP}})|0\rangle\langle 0| + \delta_{\text{SP}}|1\rangle\langle 1|$, on which the Pauli Z channel has no effect but other Pauli channels act effectively like the bit-flip error, whose error assignment matrix A_{CNOT} can be combined with the measurement error

$$A_M^{\text{noisy}} \equiv A_M A_{\text{CNOT}}, \quad (9)$$

where the superscript ‘noisy’ is used to remind us of the noisy CNOT gate. For q_t in Fig. 1, the SPAM errors become

$$A_{\text{SPAM}}^{\text{noisy}} \equiv A_M^{\text{noisy}} A_{\text{SP}}. \quad (10)$$

If we assume the CNOT gate noise acts symmetrically on control and target qubits, the A_M^{noisy} matrix will not change when swapping the control qubit and target qubit of the CNOT gate. This can be justified by adding additional four noiseless Hadamard gates alongside the CNOT gate to swap the control and target qubits. Then, the error channel will change accordingly $p_x \leftrightarrow p_z$. Thus, if $p_x \approx p_z$ (e.g., when the single-qubit depolarizing noise dominates), the assumption approximately holds (see Appendix C for mitigating procedures for more general uncorrelated Pauli noise channels). In this case, one can use the SPAM error with gate noise to calculate δ_{SP} . The SPAM error directly measured from circuits in Fig. 1, using q_a as the *target* of the CNOT gate, becomes

$$\tilde{A}_{\text{SPAM}}^{a, \text{noisy}} = A_{\text{SPAM}}^{a, \text{noisy}} A_{\text{SP}}^t, \quad (11)$$

$q_t(q_a)$	$q_0(q_1)$	$q_1(q_2)$	$q_2(q_1)$	$q_3(q_5)$	$q_4(q_5)$	$q_5(q_6)$	$q_6(q_5)$
δ_{SPAM}^0	0.0108(4)	0.0134(5)	0.0086(4)	0.0087(4)	0.0120(4)	0.0162(5)	0.0086(4)
δ_{SPAM}^1	0.0514(9)	0.0394(8)	0.085(1)	0.0380(8)	0.0457(8)	0.102(1)	0.0319(7)
δ_M^0	0.0005(8)	0.0037(8)	0.0018(9)	0.0020(8)	0.0039(8)	0.010(1)	0.0021(7)
δ_M^1	0.0411(8)	0.0297(8)	0.0780(9)	0.0312(8)	0.0376(8)	0.096(1)	0.0253(7)
δ_{SP}	0.011(1)	0.0101(4)	0.0074(3)	0.0070(5)	0.0085(7)	0.0069(3)	0.0067(4)

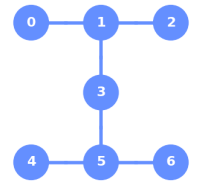


Table 1: The state preparation error and measurement error for all the qubits in `ibm_nairobi` have been characterized using methods described in the text. The qubit connection of `ibm_nairobi` is shown on the right. All of the circuits are repeated 6×10^5 times.

where $A_{\text{SPAM}}^{a,\text{noisy}}$ can be obtained by swapping CNOT and setting q_a as the *control* qubit of the CNOT gate. If one wants to characterize δ_{SP} for both qubits, one does not need additional circuits, as the circuits after swapping CNOT are exactly δ_{SP} -characterization circuits for q_a . Once obtaining $A_{\text{SPAM}}^{a,\text{noisy}}$ and $\tilde{A}_{\text{SPAM}}^{a,\text{noisy}}$, we can directly solve for δ_{SP} . As a byproduct, one can also extract the CNOT error rate δ_{CNOT} and obtain a rough estimation for the rate of the bit-flip channel.

In practice, the procedure of obtaining δ_{SP} of this method is almost the same as the noiseless case in Eq. (6). For general Pauli noise of CNOT gates, the procedure to obtain δ_{SP} is more complicated, and we leave the discussion in the appendix C. After obtaining the δ_{SP} , we use the SPAM error without the CNOT gate noise to calculate δ_M according to Eqs. (3). This can be obtained by directly preparing each qubit at $|0/1\rangle$ and then measuring it.

To validate the above arguments, we implement the circuit shown in Fig. 1 on IBM's superconducting quantum computers, and obtain the δ_{SP} via our efficient mitigation scheme, along with ZNE and PEC. The δ_{SP} result for q_{33} of `ibm_nazca` quantum computer obtained by our efficient mitigation scheme is $\delta_{\text{SP}}^{(33)} = 0.011(2)$. The results from ZNE and PEC are 0.013(2) and 0.014(6), respectively, which are consistent with our mitigation schemes. Thus, using our efficient CNOT-noise mitigation scheme merely brings a constant overhead, but can achieve similar accuracy as other gate-noise mitigation schemes. We also add more detailed comparison for our mitigation method and ZNE method in Appendix D.

2.3 Cloud experimental demonstration

2.3.1 Parallel δ_{SP} -characterization for all qubits

Since our δ_{SP} -characterization only requires two-qubit gates, the mitigation for multiple qubits can be done parallelly. For example, in a 1-D qubit chain, one can do the mitigation for even sites at the same time, then for odd sites.

We use the above method to characterize δ_{SP} for all qubits in the `ibm_nairobi`, and the results are shown in Table 1. We find that the state-preparation error,

if not separated from the readout error, is lumped in the SPAM error that is used in the current standard approach for readout mitigation.

2.3.2 One-qubit state tomography

To fully characterize the state-preparation noise channel and confirm the validness of our δ_{SP} -characterization scheme, we also perform state tomography for $|0\rangle$ state and $|1\rangle$ state after obtaining the δ_{SP} and δ_M 's. We perform the experiment on q_{12} of `ibmq_mumbai` and get $\delta_{\text{SP}} = 0.018(1)$. After the measurement mitigation with the characterized δ_M 's, the tomography results for preparing $|0\rangle$ and $|1\rangle$ are

$$\rho_0 = \begin{pmatrix} 0.983(2) & -0.003(2) - 0.016(2)i \\ -0.003(2) + 0.016(2)i & 0.017(2) \end{pmatrix}, \quad (12)$$

$$\rho_1 = \begin{pmatrix} 0.018(2) & 0.002(2) + 0.019(2)i \\ 0.002(2) - 0.019(2)i & 0.982(2) \end{pmatrix}, \quad (13)$$

where the diagonal part is consistent with the δ_{SP} obtained from our quantification scheme.

2.4 Resource estimations

In this section, we estimate the resource requirements for our δ_{SP} -characterization scheme, especially the measurement number and circuit number requirements, where we count each full-qubit measurement of one circuit as one measurement.

Firstly, Z -twirling can be implemented by four circuits: with/without the initial Z gate for the target and the ancilla qubits. The probability distribution for 50%- Z twirling can be obtained by directly averaging the results of the four circuits, which only brings a constant factor of uncertainty that can be canceled by increasing the number of measurements. Thus, the total overhead from Z -twirling is a constant number. In addition, our δ_{SP} -characterization scheme can be performed in parallel for multiple pair of qubits. Thus, the total number of measurements will not grow with the number of qubits due to the parallelization, although the complexity for classical data processing grows linearly.

In the case of noiseless CNOT, the error of our δ_{SP} -characterization is purely statistical uncertainty, which can be arbitrarily small by increasing the number of measurements N , with the corresponding uncertainty scale as $O(\frac{1}{\sqrt{N}})$.

In the noisy CNOT case, the accuracy of our scheme is upper bounded by the accuracy of gate-noise mitigation. We assume after some gate-noise mitigation scheme with a total of N_G measurements of circuits, we can obtain expectations for Z operators with systematic error Δ_G and variance σ_G^2 . As the systematic error of the δ_{SP} is lower bounded by Δ_G , our δ_{SP} -characterization scheme only works if $\Delta_G \ll \delta_{\text{SP}}$ and σ_G is sufficiently small $\sigma_G \ll \delta_{\text{SP}}$. Supposing the Δ_G can be made small compared to δ_{SP} , after performing N_{SP} repetitions of gate-noise-mitigated circuits (each repetition contains N_G measurements), the deviation $\Delta(\delta_{\text{SP}})$ from the true value of δ_{SP} is bounded, with probability $1 - p$,

$$|\Delta(\delta_{\text{SP}})| < C|\Delta_G| + O\left(\sigma_G \sqrt{\frac{\log \frac{1}{p}}{N_{\text{SP}}}}\right), \quad (14)$$

where $C > 0$ is a constant related to the true value of state-preparation errors and measurement errors. If all δ_{SPAM} 's are small, C is close to 2.

In typical gate-noise mitigation schemes (e.g., ZNE and PEC), σ_G roughly scales like $O(\frac{1}{\sqrt{N_G}})$. Thus, the total uncertainty still scales as $O(\frac{1}{\sqrt{N}})$, where $N = N_{\text{SP}}N_G$. Therefore, the uncertainty of our δ_{SP} -characterization is directly related to the uncertainty of the chosen gate-noise mitigation σ_G , and we need N_{SP} repetitions of those mitigated results (with total measurement number $N = N_{\text{SP}}N_G$) to further reduce the uncertainty from $O(\sigma_G)$ by a factor $\frac{1}{\sqrt{N_{\text{SP}}}}$.

Nevertheless, the systematic error of δ_{SP} is still lower-bounded by those of gate-noise mitigation schemes. Therefore, we require that the systematic error of gate-noise mitigation is much smaller than the δ_{SP} in order to obtain meaningful results. As we use only one CNOT in the characterization circuits, in current noisy superconducting quantum computers where CNOT gate errors can be small, the condition is roughly satisfied for typical gate-noise mitigation schemes, such as ZNE and PEC.

3 Complete mitigation for measurement error and state-preparation error

Measurement error mitigation is an essential tool to extract information from noisy quantum devices. The measurement error with only bit flips can be characterized by the error assignment matrix. In the single-qubit

case, the matrix is simply

$$A_M = \begin{pmatrix} 1 - \delta_M^0 & \delta_M^1 \\ \delta_M^0 & 1 - \delta_M^1 \end{pmatrix}. \quad (15)$$

Note that we assume that there are only bit-flip errors, or else X and Y errors have been removed by the Z randomization procedure mentioned above. Assume the ideal preparation of the initial state, the effect of the measurement error on the ideal outcome probability distribution $\mathcal{P}$ is then

$$\mathbf{P}_M = A_M \mathcal{P}, \quad (16)$$

where A_M can be the general assignment matrix characterizing n -qubit measurement error and $\mathbf{P}_M$ denotes the noisy probability distribution affected by only measurement errors. Suppose we know the exact form of A_M ; we can exactly mitigate the measurement error by

$$\mathcal{P} = A_M^{-1} \mathbf{P}_M. \quad (17)$$

However, when the state preparation error appears, the matrix A_M is usually hard to obtain exactly. In practice, the state preparation and readout errors are usually lumped together as the SPAM error δ_{SPAM} . The corresponding assignment matrix A_{SPAM} is obtained by naive calibration circuits, such as in Qiskit [14], where 2^n computational-basis states are prepared and then measured. Subsequently, the inverse of the A_{SPAM} , instead of A_M , is applied for the mitigation in practice

$$\hat{\mathbf{P}} = A_{\text{SPAM}}^{-1} \mathbf{P}_{\text{SP},M} = A_{\text{SPAM}}^{-1} A_M \mathbf{P}_{\text{SP}}. \quad (18)$$

where $\mathbf{P}_{\text{SP},M}$ denotes the noisy probability distribution affected by both state preparation and measurement errors, and $\mathbf{P}_{\text{SP}}$ denotes the noisy probability distribution affected by only state-preparation errors. Since $A_{\text{SPAM}}^{-1} A_M$ is not identity when $\delta_{\text{SP}} \neq 0$, the mitigated result $\hat{\mathbf{P}}$ differs from the δ_{SP} -only noisy distribution $\mathbf{P}_{\text{SP}}$. Suppose the assignment matrix for the state-preparation error on the initial state is A_{SP} . In single-qubit case, it is

$$A_{\text{SP}} = \begin{pmatrix} 1 - \delta_{\text{SP}} & \delta_{\text{SP}} \\ \delta_{\text{SP}} & 1 - \delta_{\text{SP}} \end{pmatrix}. \quad (19)$$

What calibration circuits measure is effectively the successive application of A_{SP} and A_M on the initial state, which gives $A_{\text{SPAM}} = A_M A_{\text{SP}}$. Thus, the result mitigated by A_{SPAM} is

$$\hat{\mathbf{P}} = A_{\text{SP}}^{-1} \mathbf{P}_{\text{SP}}. \quad (20)$$

Note that the distribution $\mathbf{P}_{\text{SP}}$ contains the state-preparation error, which takes place at the beginning of the circuit and, in general, cannot be mitigated by

applying the inverse at the end of the circuit. Thus, the mitigation by A_{SPAM}^{-1} usually over-mitigates the outcome and can give rise to some unphysical probability distributions (in addition to those caused by statistical fluctuations), such as negative probabilities. Nevertheless, the negative probability problem can be dealt with by using the nearest physical probability distribution instead [24], after applying the mitigation matrix A_{SPAM}^{-1} .

While fully characterizing those error matrices requires exponentially many circuits, if we assume all the errors are local, those error assignment matrices can be decomposed into a direct product of all single-qubit measurement errors $\{A^{(i)}\}$

$$A_M = A_M^{(1)} \otimes A_M^{(2)} \cdots A_M^{(n)}, \quad (21)$$

$$A_{\text{SP}} = A_{\text{SP}}^{(1)} \otimes A_{\text{SP}}^{(2)} \cdots A_{\text{SP}}^{(n)}, \quad (22)$$

$$A_{\text{SPAM}} = A_{\text{SPAM}}^{(1)} \otimes A_{\text{SPAM}}^{(2)} \cdots A_{\text{SPAM}}^{(n)}, \quad (23)$$

where

$$A_{\text{SPAM}}^{(i)} = \begin{pmatrix} 1 - \delta_{\text{SPAM}}^{i,0} & \delta_{\text{SPAM}}^{i,1} \\ \delta_{\text{SPAM}}^{i,0} & 1 - \delta_{\text{SPAM}}^{i,1} \end{pmatrix}. \quad (24)$$

The above conditions approximately hold and match many current superconducting qubit devices [25], where the uncorrelated errors dominate, and correlated errors among qubits are usually small. It is easy to check that $A_{\text{SPAM}}^{(i)} = A_M^{(i)} A_{\text{SP}}^{(i)}$. Thus, one can measure the δ_{SPAM} locally and then combine them to get the full A_{SPAM} matrix, which only takes $O(n)$ number of circuits. Throughout our work, we will mainly focus on uncorrelated errors so that all error assignment matrices can be decomposed into a tensor product of single-qubit ones.

Nevertheless, the state-preparation error usually does not commute with the circuit gates, which, in principle, cannot be automatically combined with the measurement errors at the end of the circuit. Thus, we need a new technique to mitigate the state-preparation error and the measurement error separately.

3.1 Measurement error mitigation

In this section, we present a similar mitigation scheme for the measurement error as Eq. 17. Here, we assume that all the errors are uncorrelated and Eqs. (21), (22), and (23) hold. We can firstly measure the SPAM error δ_{SPAM} for each qubit, then use the methods in Sec. 2.1 to obtain the state-preparation errors δ_{SP} 's and measurement errors δ_M 's. After getting the single-qubit measurement errors $\delta_M^{i,0}$ and $\delta_M^{i,1}$ for each qubit, we construct the full measurement error assignment matrix A_M according to Eq. (15) and Eq. (21). The measurement error is mitigated by applying the inverse of A_M

$$\hat{P}_{\text{SP}} = A_M^{-1} P_{\text{SP},M}. \quad (25)$$

Ideally, after such a readout error mitigation process, the remaining errors in $\hat{P}_{\text{SP}}$ purely come from state preparation and gate operations.

3.2 State preparation error mitigation

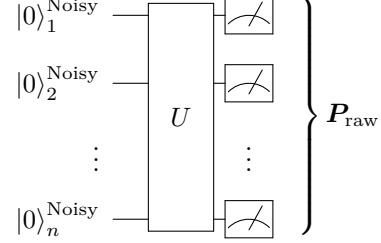


Figure 2: The target circuit to be mitigated, where each qubit is intended to be initialized at $|0\rangle$, but with some probabilities to be prepared at $|1\rangle$ due to the state-preparation errors. We label such erroneous preparation as $|0\rangle_{\text{Noisy}}$. The gates U and measurements can also contain noise. The output probability distribution is labeled as P_{raw} .

In this section, we present a new mitigation scheme to mitigate the state-preparation errors. The mitigation is applied in a fully noisy scenario: we prepare all n qubits at $|0\rangle$ with state-preparation errors, apply some noisy gates U , perform some noisy measurements at the end, and obtain a noisy probability distribution P_{raw} (see Fig. 2). The goal of our state-preparation error mitigation is to get a mitigated distribution $\hat{P}$ that is sufficiently close to the δ_{SP} -free distribution $\mathcal{P}$, where $\mathcal{P}$ can be obtained with perfect qubit initialization at all $|0\rangle$'s along with the same noisy gates U and noisy measurements as P_{raw} .

As we assume the state-preparation errors can be described by independent classical bit-flip errors, the erroneous output can be regarded as samples of classical independent random events. For each event, qubits q_j 's are perfectly initialized in $s_j \in \{0, 1\}$, undergo the same noisy gates U and noisy measurements, and output a probability distribution $\mathcal{P}_{s_1, s_2 \dots s_n}$. Such event happens with a probability $\prod_{j=1}^n q_{\text{SP}}^{(j)}(s_j, 0)$, where

$$q_{\text{SP}}^{(j)}(s, s') \equiv \begin{cases} 1 - \delta_{\text{SP}}^{(j)}, & s = s'; \\ \delta_{\text{SP}}^{(j)}, & s \neq s'; \end{cases} \quad (26)$$

and $\delta_{\text{SP}}^{(j)}$ is the state-preparation error of the j -th qubit q_j . Thus, the output probability distribution P is a combination of all possible initial assignments $\{s_j\} \in \{0, 1\}^n$ with the corresponding probability

$$P_{\text{raw}} = \sum_{\{s_j\} \in \{0, 1\}^n} \left(\prod_{j=1}^n q_{\text{SP}}^{(j)}(s_j, 0) \right) \mathcal{P}_{s_1, s_2 \dots s_n}. \quad (27)$$

We can also regard the $\mathbf{P}_{\text{raw}}$ as a function of $\delta_{\text{SP}}^{(j)}$'s, and rewrite the equation as

$$\mathbf{P}(x_1, x_2, \dots) = \sum_{\{s_j\} \in \{0,1\}^n} \left(\prod_{j=1}^n q_{x_j}(s_j, 0) \right) \mathcal{P}_{s_1, s_2 \dots s_n}, \quad (28)$$

where $q_{x_j}(s, s')$ is $1 - x_j$ if $s = s'$ and x_j otherwise, similar to Eq. (26). $\mathbf{P}_{\text{raw}}$ can then be obtained from $\mathbf{P}(x_1, x_2, \dots)$ by setting $\{x_j\} = \{\delta_{\text{SP}}^{(j)}\}$.

Note that the Eq. (28) is a linear equation with regard to $\mathcal{P}_{s_1, s_2 \dots s_n}$'s. Thus, to obtain desired δ_{SP} -free probability distribution

$$\mathcal{P} = \mathcal{P}_{0,0\dots} = \mathbf{P}(0, 0\dots), \quad (29)$$

we can, in principle, solve 2^n linear equations with 2^n choices of $\{x_1, x_2, \dots\}$ for $\mathbf{P}(x_1, x_2, \dots)$ in Eq. (28). Nevertheless, in real quantum devices, it is usually hard to tune the state-preparation error freely. Here we introduce a trick to change x_j 's without changing the actual state-preparation errors. Notice that if we add a noiseless initial X gate on qubit q_j , then the qubit is prepared in $|0\rangle$ with probability δ_{SP}^j , and in $|1\rangle$ with probability $1 - \delta_{\text{SP}}^j$, which effectively changes the state-preparation error for q_j to $1 - \delta_{\text{SP}}^j$. Thus, as long as $q_j \neq \frac{1}{2}$, we can obtain a linearly independent equation for $\mathcal{P}_{s_1, s_2 \dots s_n}$'s. By adding or not adding initial X gates for all qubits, we obtain 2^n configuration of $\{x_j\}$'s, and 2^n linear equations

$$\mathbf{P}^{s'_1, s'_2, \dots, s'_n} = \sum_{\{s_j\} \in \{0,1\}^n} \left(\prod_{j=1}^n q_{\text{SP}}^{(j)}(s_j, s'_j) \right) \mathcal{P}_{s_1, s_2 \dots s_n}, \quad (30)$$

where $s_j, s'_j \in \{0, 1\}$ indicate qubit-initialization choices and

$$\mathbf{P}^{s'_1, s'_2, \dots, s'_n} \equiv \mathbf{P} \left(s'_1(1 - \delta_{\text{SP}}^{(1)}) + (1 - s'_1)\delta_{\text{SP}}^{(1)}, \right. \\ \left. s'_2(1 - \delta_{\text{SP}}^{(2)}) + (1 - s'_2)\delta_{\text{SP}}^{(2)}, \dots \right). \quad (31)$$

The probability distribution $\mathbf{P}^{s'_1, s'_2, \dots, s'_n}$ can be experimentally obtained, by initializing the qubits in $|s'_1, s'_2, \dots, s'_n\rangle$ via adding an initial X gate to qubit q_j if $s'_j = 1$. For a small number of qubits, one can exactly solve $\mathcal{P}$'s by explicitly measuring the 2^n circuits and solving 2^n linear equations.

However, for a large number of qubits, exactly solving $\mathcal{P}$ requires exponentially many circuits and is not practical in experiments. Here, we present an approximate approach that takes a linear number of circuits to achieve $O(\delta_{\text{SP}}^2)$ accuracy. To do this, we first expand

the distribution $\mathbf{P}$ near the $\delta_{\text{SP}}^{(j)}$,

$$\mathbf{P}(x_1, x_2, \dots) \approx \mathbf{P}(\delta_{\text{SP}}^{(1)}, \delta_{\text{SP}}^{(2)}, \dots) + \sum_{i=1}^n \frac{\partial \mathbf{P}(\delta_{\text{SP}}^{(1)}, \dots)}{\partial x_i} (x_i - \delta_{\text{SP}}^{(i)}) + \dots \quad (32)$$

If we assume the $\delta_{\text{SP}}^{(i)}$'s are small and approximate the above equation to the order of $O(\delta_{\text{SP}})$, we can neglect higher order terms and get the first-order estimation $\hat{\mathbf{P}}$ of $\mathbf{P}(0, 0\dots)$

$$\hat{\mathbf{P}} = \mathbf{P}(\delta_{\text{SP}}^{(1)}, \delta_{\text{SP}}^{(2)}, \dots) - \sum_{i=1}^n \frac{\partial \mathbf{P}(\delta_{\text{SP}}^{(1)}, \dots)}{\partial x_i} \delta_{\text{SP}}^{(i)}, \quad (33)$$

where $\mathbf{P}(\delta_{\text{SP}}^{(1)}, \delta_{\text{SP}}^{(2)}, \dots) = \mathbf{P}_{\text{raw}}$ is the raw experimental result to be mitigated, and the derivatives $\frac{\partial \mathbf{P}(\delta_{\text{SP}}^{(1)}, \dots)}{\partial x_i}$ can be obtained in a similar way as solving linear equations. This can be seen by rewriting Eq. (28) into

$$\mathbf{P}(x_1, \delta_{\text{SP}}^{(2)}, \dots) = (1 - x_1)\tilde{\mathbf{P}}_{0,X} + x_1\tilde{\mathbf{P}}_{1,X}, \quad (34)$$

where we have taken q_1 as an example and

$$\tilde{\mathbf{P}}_{0,X} = \sum_{\{s_2, s_3, \dots\} \in \{0,1\}^{n-1}} \left(\prod_{j=2}^n q_{\text{SP}}^{(j)}(s_j, 0) \right) \mathcal{P}_{0, s_2 \dots s_n}, \\ \tilde{\mathbf{P}}_{1,X} = \sum_{\{s_2, s_3, \dots\} \in \{0,1\}^{n-1}} \left(\prod_{j=2}^n q_{\text{SP}}^{(j)}(s_j, 0) \right) \mathcal{P}_{1, s_2 \dots s_n}.$$

Thus, $\mathbf{P}(x_1, \delta_{\text{SP}}^{(2)}, \dots)$ is linear with x_1 . The derivative can be obtained by one additional probability distribution with $x_1 \neq \delta_{\text{SP}}^{(1)}$, and such different x_1 can be obtained by adding initial X gate on q_1 so that effectively $x_1 = 1 - \delta_{\text{SP}}^{(1)}$. We denote the probability distribution obtained by adding an initial noiseless X gate on q_i as $\mathbf{Q}_i$ (see Fig. 3).

The original distribution $\mathbf{P}_{\text{raw}}$ and the additional distribution $\mathbf{Q}_1$ (the distribution from the same circuit with additional initial X gate on the first qubit) can be viewed as the result of substituting $x_1 = \delta_{\text{SP}}^{(1)}$ and $x_1 = 1 - \delta_{\text{SP}}^{(1)}$, respectively, in Eq. (34),

$$\mathbf{P}_{\text{raw}} \equiv \mathbf{P}(\delta_{\text{SP}}^{(1)}, \dots) = (1 - \delta_{\text{SP}}^{(1)})\tilde{\mathbf{P}}_{0,X} + \delta_{\text{SP}}^{(1)}\tilde{\mathbf{P}}_{1,X}, \\ \mathbf{Q}_1 \equiv \mathbf{P}(1 - \delta_{\text{SP}}^{(1)}, \dots) = \delta_{\text{SP}}^{(1)}\tilde{\mathbf{P}}_{0,X} + (1 - \delta_{\text{SP}}^{(1)})\tilde{\mathbf{P}}_{1,X}.$$

Thus, we can obtain the derivative of x_1 by

$$\frac{\partial \mathbf{P}(\delta_{\text{SP}}^{(1)}, \dots)}{\partial x_1} = \tilde{\mathbf{P}}_{1,X} - \tilde{\mathbf{P}}_{0,X} \\ = \frac{\mathbf{Q}_1 - \mathbf{P}_{\text{raw}}}{1 - 2\delta_{\text{SP}}^{(1)}}. \quad (35)$$

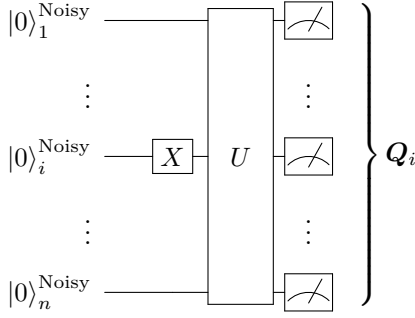


Figure 3: The first-order state-preparation error mitigation circuit on i -th qubit q_i , where an initial X gate is added to the q_i and we assume the noise brought by this additional X is negligible. The rest of the circuits, including the noisy qubit preparation, noisy gates U , and noisy measurements are the same as target circuit in Fig. 2. The output probability distribution is labeled as Q_i .

The other derivatives w.r.t. x_i 's can be obtained in a similar way. Substituting them into the Eq. (33), we arrive at the equation for the first-order state-preparation error mitigation

$$\hat{\mathbf{P}} = \mathbf{P}_{\text{raw}} + \sum_{i=1}^n \frac{\delta_{\text{SP}}^{(i)}}{1 - 2\delta_{\text{SP}}^{(i)}} (\mathbf{P}_{\text{raw}} - \mathbf{Q}_i). \quad (36)$$

Note that the state-preparation error mitigation (36) is linear among different distributions $\mathbf{P}$'s and that measurement mitigation (25) is a linear transformation within each distribution $\mathbf{P}$, so swapping the order of measurement error mitigation and state-preparation error mitigation here does not change the final results.

We remark that one can achieve r -th-order accuracy mitigation at the cost of $O(n^r)$ additional circuits with a similar procedure, where n is the qubit number. A straightforward way is to expand the Eq. (32) to higher orders. As $P(x_1, x_2, \dots)$ is a multi-linear function of $\{x_1, x_2, \dots\}$, a r -th order derivative is non-zero only if it involves exactly r number of x_j 's. To obtain the r -th order derivative with respect to $x_{m_1}, \dots, x_{m_r}$, one can follow the procedure similar to those from Eq. (34) to Eq. (35). By adding/not adding initial X gate for qubits $\{x_{m_1}, \dots, x_{m_r}\}$, one can obtain 2^r number of linear equations with respect to 2^r number of $\hat{\mathbf{P}}$'s, which can be exactly calculated by solving the 2^r linear equations. In addition, the r -th order derivative can be expressed as linear combinations of $\tilde{\mathbf{P}}$, thus can also be directly calculated after obtaining $\tilde{\mathbf{P}}$'s. For r -th order derivatives, the number of additional distributions needed is $\sum_{j=1}^r \binom{n}{j}$, and these distributions can be obtained by adding initial X gates for at most r qubits.

Since the δ_{SP} -mitigation only involves the linear combination of probability distributions, we can directly apply the mitigation scheme for observables. For an

observable O_{raw} obtained from a noisy quantum computer, we can perform the circuits with initially flipped i -th qubit to obtain O_i , and mitigate (at the first order) the state-preparation errors by

$$\hat{O} = O_{\text{raw}} + \sum_{i=1}^n \frac{\delta_{\text{SP}}^{(i)}}{1 - 2\delta_{\text{SP}}^{(i)}} (O_{\text{raw}} - O_i). \quad (37)$$

In addition, as the δ_{SP} -mitigation scheme does not interfere with each probability distribution, the scheme is compatible with most gate-noise mitigation and measurement error mitigation techniques. To do further mitigation for gate noises and measurement errors, one can firstly apply corresponding mitigation techniques to $\mathbf{P}_{\text{raw}}$ and $\mathbf{Q}_i$, and then do the δ_{SP} -mitigation according to Eq. (36). Same procedure can also be applied to the additional gate and measurement noise mitigations for observables.

3.3 Resource estimations

Compared to the results without δ_{SP} -mitigation, our δ_{SP} -mitigation scheme requires additional n circuits, where n is the qubit number. The δ_{SP} -mitigation reduce the error rate to $O(\delta_{\text{SP}}^2)$, whereas the error rate is $O(\delta_{\text{SP}})$ without mitigation. On the other hand, when doing the δ_{SP} -mitigation according to Eq. (36), the uncertainty of the final results $\hat{\mathbf{P}}$ grows with $O(\sqrt{n}\delta_{\text{SP}})$, where we assume all state-preparation errors are almost identical and close to some $\delta_{\text{SP}} < \frac{1}{3}$ and n is sufficiently large. To remedy this and restore the same uncertainty before δ_{SP} -mitigation, we require additional $O(n\delta_{\text{SP}}^2)$ multiples of measurements to reduce the statistical uncertainty. Thus, if one strictly requires the mitigated $\hat{\mathbf{P}}$ to achieve the same uncertainty as the $\mathbf{P}_{\text{raw}}$, the total additional overhead of measurements is $O(\delta_{\text{SP}}^2 n^2)$.

We remark that the overhead can be greatly reduced if not all qubits are highly noisy. In the first order mitigation, the correction terms for δ_{SP} of each qubit are separated. Thus, one can perform qubit-wise mitigation only for those qubits with large δ_{SP} 's. For example, if only q_j contains large state-preparation error and other qubits are almost perfectly prepared with negligible δ_{SP} 's, one can mitigate the state-preparation errors for q_j only by

$$\hat{\mathbf{P}} = \mathbf{P}_{\text{raw}} + \frac{\delta_{\text{SP}}^{(j)}}{1 - 2\delta_{\text{SP}}^{(j)}} (\mathbf{P}_{\text{raw}} - \mathbf{Q}_j), \quad (38)$$

with the remainder error being $O((\delta_{\text{SP}}^{(j)})^2 + \sum_{i \neq j} \delta_{\text{SP}}^{(i)})$. Therefore, the total overhead of the mitigation can be reduced to $O(\delta^2 m^2)$, where m counts the number of qubits with sufficiently large $\delta_{\text{SP}} \sim \delta$ that need mitigations.

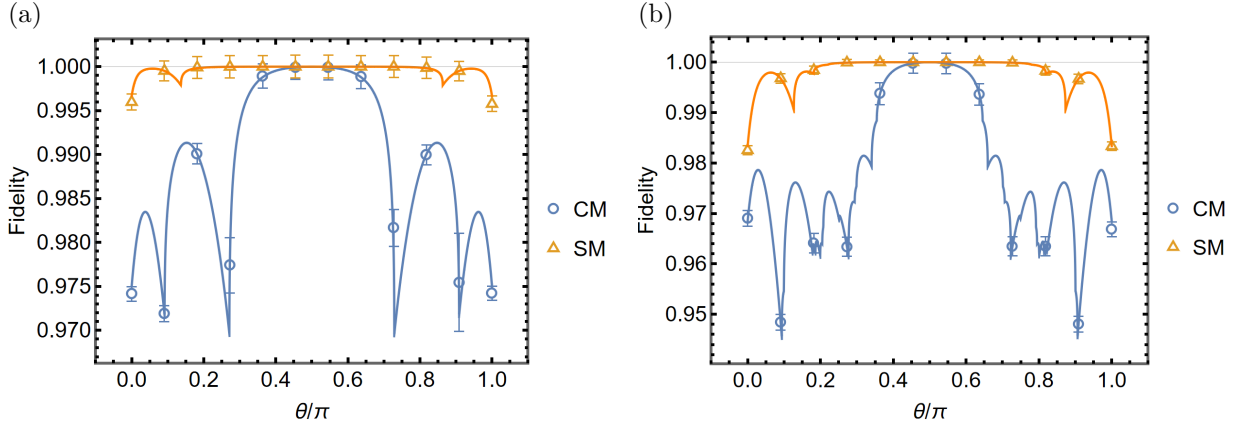


Figure 4: Simulation results for comparison between two mitigation schemes for the (a) 2-qubit circuit (Fig. 6) and (b) 4-qubit circuit (Fig. 7). We set $\delta_{\text{SP}} = 0.05$, $\delta_M^0 = 0.04$, and $\delta_M^1 = 0.06$ for all qubits. The circuits for calculating δ_{SP} are repeated 4×10^6 times, and the target circuits are repeated 6.4×10^5 times.

4 comparison between two mitigation schemes

In the current standard practice on superconducting qubits, readout mitigation is carried out first by measuring the mitigation assignment matrix A_{SPAM} obtained by measuring the distribution in the computational basis when states are prepared in all different computational states. Then, the inverse of the assignment matrix is applied to the actual measurement distribution. However, as mentioned before, the preparation of computational states incurs errors, and these are not separately characterized and are thus lumped together in the readout mitigation matrix. We call this combined mitigation (CM) and refer to ours as separate state preparation and readout mitigation (SM).

4.1 Summary of the two mitigation schemes

First, we summarize the procedure of CM as follows:

1. Execute the quantum circuit U on qubits $\{q_i\}$ and get the noisy raw probability distribution $\mathbf{P}_{\text{raw}}$;
2. Run the standard SPAM error calibration circuits (initializing each qubit at $|0\rangle/|1\rangle$ followed by measurements) on qubits $\{q_i\}$ and get the SPAM error assignment matrix A_{SPAM} ;
3. Apply the inverse of A_{SPAM} to the raw distribution and get the mitigated quasi-probability distribution $\tilde{\mathbf{P}} = A_{\text{SPAM}}^{-1} \mathbf{P}_{\text{raw}}$;
4. Find the nearest probability distribution $\hat{\mathbf{P}}$ from $\tilde{\mathbf{P}}$ in case of the existence of negative probabilities in $\tilde{\mathbf{P}}$. Then, the output $\hat{\mathbf{P}}$ is the desired mitigated probability distribution.

Next, we summarize our proposed mitigation procedure (SM) as follows:

1. Execute the quantum circuit U on qubits $\{q_i\}$ and get the noisy raw probability distribution $\mathbf{P}_{\text{SP},M}$;
2. Run the calibration circuits in sec. 2.1 on each qubit pair and separately get the state preparation error $\{\delta_{\text{SP}}^{(i)}\}$ and readout error assignment matrix A_M ;
3. Add an initial X gate on i th qubit q_i and then apply the same circuit U for each qubit to get the calibration distribution $\{\mathbf{Q}_i\}$;
4. Perform the measurement error mitigation by applying the inverse of A_M to the raw distribution $\mathbf{P}_{\text{SP},M}$ and calibration distribution $\{\mathbf{Q}_i\}$ s, and get the mitigated quasi-probability distribution $\tilde{\mathbf{P}}_{\text{SP}} = A_M^{-1} \mathbf{P}_{\text{SP},M}$ and $\{\tilde{\mathbf{Q}}_i = A_M^{-1} \mathbf{Q}_i\}$;
5. Use the Eq. (36) to get the δ_{SP} -mitigated quasi-probability distribution $\tilde{\mathbf{P}}$ from $\tilde{\mathbf{P}}_{\text{SP}}$, $\{\tilde{\mathbf{Q}}_i\}$, and $\{\delta_{\text{SP}}^{(i)}\}$: $\tilde{\mathbf{P}} = \tilde{\mathbf{P}}_{\text{SP}} + \sum_{i=1}^n \frac{\delta_{\text{SP}}^{(i)}}{1-2\delta_{\text{SP}}^{(i)}} (\tilde{\mathbf{P}}_{\text{SP}} - \tilde{\mathbf{Q}}_i)$;
6. Find the nearest probability distribution $\hat{\mathbf{P}}$ from $\tilde{\mathbf{P}}$ in case of the existence of negative probabilities in $\tilde{\mathbf{P}}$. Then, the output $\hat{\mathbf{P}}$ is the desired mitigated probability distribution.

In both CM and SM, we assume all the state-preparation errors and measurement errors are uncorrelated for different qubits so that their assignment matrices can be written as a tensor product of single-qubit ones. We remark that, the total complexity and resource requirements for CM are the same as obtaining $\tilde{\mathbf{P}}_{\text{SP}} = A_M^{-1} \mathbf{P}_{\text{SP},M}$ in SM. The only difference is that in CM, we use A_{SPAM} to mitigate the measurement errors,

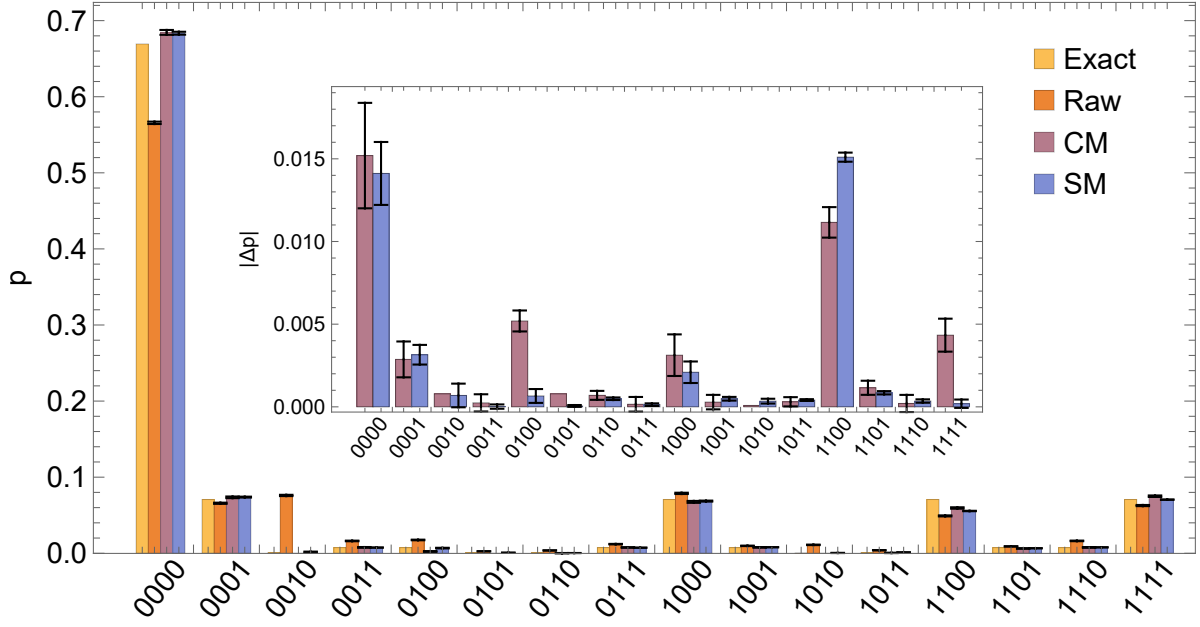


Figure 5: Experimental results, expected exact results, and mitigated results, using both combined mitigation (CM) and separate mitigation (SM), for the 4-qubit circuit shown in Fig. 7 with $\theta = \pi/5$. The outcome labels on the horizontal axis are in the order $[q_4, q_3, q_2, q_1]$. We also illustrate the absolute differences $|\Delta p|$ from exact values for CM and SM in the inset. The results are obtained from ibmq_mumbai (whose machine layout is shown above the histogram), with the initial layout of our experiment being the four qubits $[21, 18, 15, 12]$ (highlighted in green in Fig. 8). The circuits for calculating $\{\delta_{SP}\}$ are repeated 6×10^5 times, and the target circuits are repeated 10^5 times.

whereas in SM, we use A_M instead, which is closer to the true measurement errors. Since SM requires additional circuits and measurements for Q_i 's, it contains an additional overhead for measurement number. According to the analysis in Sec. 3.3, if $\delta_{SP}^2 n \ll 1$, the additional overhead is roughly $O(n)$. For $\delta_{SP}^2 n \gg 1$, the additional overhead is $O(\delta_{SP}^2 n^2)$.

We remark that if one wants to further perform mitigation on gate noise, one can do that on top of the mitigated distribution $\hat{P}$.

4.2 Results from simulator

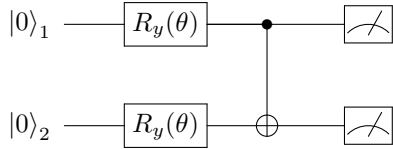


Figure 6: The 2-qubit benchmark circuit for comparing two mitigation schemes CM and SM.

Here, we compare mitigating the state preparation error and measurement error separately and together for a simple circuit shown in Fig. 6. (A four-qubit example

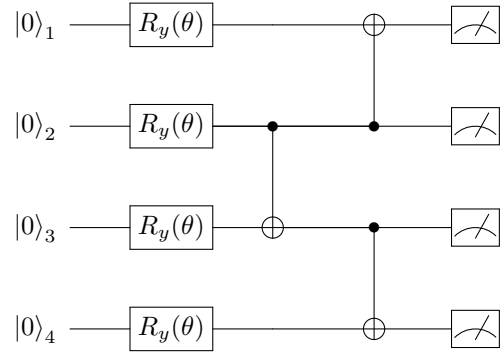


Figure 7: The 4-qubit benchmark circuit for comparing two mitigation schemes CM and SM.

circuit that will be tested later is shown in Fig. 7.) The R_y gate is defined as $R_y(\theta) = \exp(-i\theta Y/2)$ and the ideal output state is

$$|\hat{\psi}\rangle = \cos^2 \frac{\theta}{2} |00\rangle + \sin^2 \frac{\theta}{2} |01\rangle + \sin \frac{\theta}{2} \cos \frac{\theta}{2} |10\rangle + \sin \frac{\theta}{2} \cos \frac{\theta}{2} |11\rangle. \quad (39)$$

Since the measurement can only give probability distributions instead of wave functions, we define the over-

lap (or fidelity) between two probability distributions as [26]

$$F(\mathbf{P}, \mathbf{P}') = \left(\sum_{i=1}^{2^n} \sqrt{p_i p'_i} \right)^2, \quad (40)$$

where p_i is the probability in the distribution $\mathbf{P}$ for the state $|i\rangle$. The above definition is related to their Heilinger distance [27] $H(\mathbf{P}, \mathbf{P}')$ via $\sqrt{F(\mathbf{P}, \mathbf{P}')} = 1 - H^2(\mathbf{P}, \mathbf{P}')$ between the two distributions. The fidelity of an output distribution $\mathbf{P}_\theta$ is then defined as the overlap with the ideal probability distribution $\hat{\mathbf{P}}$

$$f(\theta) = F(\mathbf{P}_\theta, \hat{\mathbf{P}}), \quad (41)$$

where $\hat{\mathbf{P}} = (\cos^4 \frac{\theta}{2}, \sin^4 \frac{\theta}{2}, \frac{1}{4} \sin^2 \theta, \frac{1}{4} \sin^2 \theta)$. We use a simulator with a state preparation error $\delta_{\text{SP}} = 0.05$ and $\delta_M^0 = 0.04, \delta_M^1 = 0.06$ for both qubits.

The theoretical calculation is straightforward but gives rise to messy expressions for both mitigation schemes, except for $\theta = 0$, at which we have

$$f_{\text{CM}}(0) = 1 - \frac{\delta_{\text{SP}}^{(1)}}{2}, \quad (42)$$

$$f_{\text{SM}}(0) = 1 - \frac{4}{3} \delta_{\text{SP}}^{(1)} \delta_{\text{SP}}^{(2)}, \quad (43)$$

and, at $\theta = \pi/2$, we have

$$f_{\text{CM}}(\frac{\pi}{2}) = f_{\text{SM}}(\frac{\pi}{2}) = 1. \quad (44)$$

Therefore, we expect that, near $\theta = 0$, the two mitigation schemes achieve fidelity with the ideal distribution at different orders, with the CM infidelity roughly approaching the first order $O(0.05)$ and SM infidelity approaching the second order $O(0.05^2)$ in most region. The simulation results for both the two-qubit and four-qubit circuits are shown in Fig. 4, where SM has better fidelity and less fluctuation in a large proportion of the parameter region in both cases.

4.3 Experimental demonstration on superconducting quantum computers

In this section, we experimentally implement the δ_{SP} characterizing circuits and obtain the estimated values of δ_{SP} on superconducting quantum computers of IBM's cloud platform. Then, we perform quantum state tomography (with readout mitigation) to confirm that the δ_{SP} values we obtain match the tomography results. Finally, we compare the two mitigation schemes on a 4-qubit circuit and demonstrate that the SM obtains better probability fidelity than the CM.

To demonstrate the direct application to circuit runs on current NISQ devices, we use the circuit shown in Fig. 7 to benchmark both mitigation

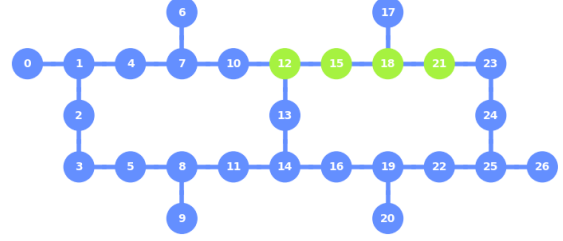


Figure 8: The qubit geometry for ibmq_mumbai. We run the benchmark circuit on qubits [21,18,15,12] (highlighted in green).

schemes. We choose $\theta = \pi/5$. We run the circuit on the ibmq_mumbai with the initial qubit layout [21,18,15,12]. Using the technique shown in Sec. 2.2, we obtained the state preparation errors $\{\delta_{\text{SP}}\} = \{0.007(3), 0.007(2), 0.007(1), 0.029(1)\}$. Then, we carry out the procedure for the CM and SM, respectively. The raw and mitigated probability results are shown in Fig. 5. The CM gives 99.6(4)% probability fidelity while the SM gives 99.9(2)% probability fidelity. The simulation without gate noise, but with state-preparation errors and measurement errors, the same as the experimental data, indicates a 99.6273% to 99.9998% improvement in fidelity, close to the experimental results.

5 Conclusion

We have proposed a resource-efficient approach to separately quantify the state preparation and readout error rates. With these two errors separately quantified, we have also described efficient methods to mitigate them, especially mitigating state preparation errors with linear complexity. In addition to theoretical analysis, we have additionally presented results from cloud experiments on IBM's superconducting quantum computers, and the results well match our theoretical predictions.

In this work, we have assumed high fidelity of X gates, which is usually the case in real quantum devices [20]. We also proposed an efficient way to obtain δ_{SP} in the presence of CNOT gate noise, which, in the experiment, gives consistent results compared to those obtained by the ZNE and PEC. Moreover, we believe our methods can be easily extended beyond superconducting quantum computers to other physical systems.

The experimental results show practical applications of our quantification and mitigation scheme for δ_{SP} on current noisy quantum computers. Although the improvement of fidelity is limited due to the relatively small δ_{SP} compared to the δ_M on IBM's superconduct-

ing quantum computers, the improvement would be much evident when the δ_M is reduced to a similar scale as δ_{SP} in the future. We advocate that in reporting the properties of quantum bits, the state preparation error rate should also be provided as an important metric (at least the portion in the incoherent mixture of $|1\rangle\langle 1|$ when preparing $|0\rangle$; but, in principle, quantum state tomography can be used for further characterization).

The work focuses on single-qubit errors, which can be extended to multi-qubit correlated errors. One will need to add twirling to the initial state preparation so that its error becomes incoherent, which will require an exponentially increasing number of circuits. One can also straightforwardly replace all the error assignment matrices with the full $2^n \times 2^n$ matrix, although this will not be practical for large n . However, one may consider dividing the system into groups of qubits that have significant correlations and then characterizing different groups independently, similar to the two-qubit case in [12]. One may also implement the truncation schemes in [25] to reduce the complexity of mitigation further. In general, methods to efficiently deal with multiqubit-correlated errors in state preparation are worth exploring in the future.

Code availability. The codes and data for the cloud experiments in this work are available upon reasonable request. The codes for demonstrating separate quantification are available at [github](#).

Acknowledgments

We thank Yusheng Zhao and Nicholas Bronn for many useful discussions. This work was partly supported by the National Science Foundation under Grant No. PHY 2310614, in particular, on the part of the state preparation error’s characterization and mitigation, and by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under Contract No. DE-SC0012704, in particular, on the part of the algorithmic cooling. This research also used resources from the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725, and the Brookhaven National Laboratory operated IBM-Q Hub. The results presented in this work do not reflect the views of IBM and its employees.

References

- [1] John Preskill. “Quantum Computing in the NISQ era and beyond”. *Quantum* **2**, 79 (2018).
- [2] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. “Error mitigation for short-depth quantum circuits”. *Phys. Rev. Lett.* **119**, 180509 (2017).
- [3] Suguru Endo, Simon C. Benjamin, and Ying Li. “Practical quantum error mitigation for near-future applications”. *Phys. Rev. X* **8**, 031027 (2018).
- [4] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski. “Cloud quantum computing of an atomic nucleus”. *Phys. Rev. Lett.* **120**, 210501 (2018).
- [5] Abhinav Kandala, Kristan Temme, Antonio D Córcoles, Antonio Mezzacapo, Jerry M Chow, and Jay M Gambetta. “Error mitigation extends the computational reach of a noisy quantum processor”. *Nature* **567**, 491–495 (2019).
- [6] Tudor Giurgica-Tiron, Yousef Hindy, Ryan LaRose, Andrea Mari, and William J Zeng. “Digital zero noise extrapolation for quantum error mitigation”. In 2020 IEEE International Conference on Quantum Computing and Engineering (QCE). Pages 306–316. IEEE (2020).
- [7] Youngseok Kim, Christopher J Wood, Theodore J Yoder, Seth T Merkel, Jay M Gambetta, Kristan Temme, and Abhinav Kandala. “Scalable error mitigation for noisy quantum circuits produces competitive expectation values”. *Nature Physics* **19**, 752–759 (2023).
- [8] Ewout Van Den Berg, Zlatko K Mineev, Abhinav Kandala, and Kristan Temme. “Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors”. *Nature physics* **19**, 1116–1121 (2023).
- [9] Yanzhu Chen, Maziar Farahzad, Shinjae Yoo, and Tzu-Chieh Wei. “Detector tomography on ibm quantum computers and mitigation of an imperfect measurement”. *Physical Review A* **100**, 052315 (2019).
- [10] Filip B Maciejewski, Zoltán Zimborás, and Michał Oszmaniec. “Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography”. *Quantum* **4**, 257 (2020).
- [11] Michael R Geller and Mingyu Sun. “Toward efficient correction of multiqubit measurement errors: Pair correlation method”. *Quantum Science and Technology* **6**, 025009 (2021).
- [12] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C McKay, and Jay M Gambetta. “Mitigating

- measurement errors in multiqubit experiments”. *Physical Review A* **103**, 042605 (2021).
- [13] Ewout Van Den Berg, Zlatko K Mineev, and Kristan Temme. “Model-free readout-error mitigation for quantum expectation values”. *Physical Review A* **105**, 032620 (2022).
 - [14] Qiskit contributors. “Qiskit: An open-source framework for quantum computing” (2023).
 - [15] Raymond Laflamme, Junan Lin, and Tal Mor. “Algorithmic cooling for resolving state preparation and measurement errors in quantum computing”. *Phys. Rev. A* **106**, 012439 (2022).
 - [16] Jonathan Baugh, Osama Moussa, Colm A Ryan, Ashwin Nayak, and Raymond Laflamme. “Experimental implementation of heat-bath algorithmic cooling using solid-state nuclear magnetic resonance”. *Nature* **438**, 470–473 (2005).
 - [17] CA Ryan, O Moussa, J Baugh, and R Laflamme. “Spin based heat engine: demonstration of multiple rounds of algorithmic cooling”. *Physical review letters* **100**, 140501 (2008).
 - [18] Leonard J Schulman, Tal Mor, and Yossi Weinstein. “Physical limits of heat-bath algorithmic cooling”. *Physical review letters* **94**, 120501 (2005).
 - [19] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Kenneth Rudinger, Jonathan Mizrahi, Kevin Fortier, and Peter Maunz. “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography”. *Nature communications* **8**, 14485 (2017).
 - [20] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout Van Den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, et al. “Evidence for the utility of quantum computing before fault tolerance”. *Nature* **618**, 500–505 (2023).
 - [21] Jakob M Günther, Francesco Tacchino, James R Wootton, Ivano Tavernelli, and Panagiotis Kl Barkoutsos. “Improving readout in quantum simulations with repetition codes”. *Quantum Science and Technology* **7**, 015009 (2021).
 - [22] Rebecca Hicks, Bryce Kobrin, Christian W Bauer, and Benjamin Nachman. “Active readout-error mitigation”. *Physical Review A* **105**, 012419 (2022).
 - [23] Miroslav Urbanek, Benjamin Nachman, Vincent R Pascuzzi, Andre He, Christian W Bauer, and Wibe A de Jong. “Mitigating depolarizing noise on quantum computers with noise-estimation circuits”. *Physical review letters* **127**, 270502 (2021).
 - [24] John A Smolin, Jay M Gambetta, and Graeme Smith. “Efficient method for computing the maximum-likelihood quantum state from measurements with additive gaussian noise”. *Physical review letters* **108**, 070502 (2012).
 - [25] Paul D Nation, Hwajung Kang, Neereja Sundaresan, and Jay M Gambetta. “Scalable mitigation of measurement errors on quantum computers”. *PRX Quantum* **2**, 040326 (2021).
 - [26] Michael A Nielsen and Isaac L Chuang. “Quantum computation and quantum information”. Cambridge university press. (2010).
 - [27] Ernst Hellinger. “Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen.”. *Journal für die reine und angewandte Mathematik* **1909**, 210–271 (1909).

A Comparison and Extension of the Laflamme-Lin-Mor algorithmic-cooling scheme for separate quantification of state preparation and readout errors

In this appendix, we briefly review the algorithmic-cooling method proposed in [15] and compare it with our δ_{SP} -characterization scheme.

A.1 Brief summary of algorithmic-cooling scheme for quantifying δ_{SP}

The measurement-based algorithmic-cooling method [15] is based on algorithmically discarding the higher-energy state $|1\rangle$ from the measurement outcome of ancillary qubits, so that it can cool the system closer to the ground state $|0\rangle$ (see Fig. 9).

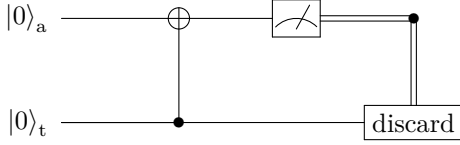


Figure 9: Circuits for algorithmic-cooling scheme with one ancilla, where the results of target qubit is discarded if the ancilla qubit is measured at $|1\rangle$.

We assume $\delta_{\text{SP}} = \delta$ for all qubits in this section for simplicity. One can verify that the effective state-preparation error is reduced to $O(\delta^2)$ after discarding all the results where the ancilla qubit is measured at $|1\rangle$. Such procedure can be further extended to k ancilla qubits, where each ancilla qubit is prepared at $|0\rangle$ (with error δ), connected with the target qubit with a CNOT gate similar to Fig. 9. After discarding all the results where any ancilla qubit is measured at $|1\rangle$, the effective state-preparation error for q_t is reduced to $O(\delta^k)$. If k is sufficiently large, the effective SPAM error of q_t is almost purely measurement errors, thus can be measured directly. Then one can solve δ_{SP} according to Eq. (3), where δ_{SPAM} can be measured directly and δ_M can be measured after discarding $|1\rangle_a$ results from large- k algorithmic-cooling circuit.

Compared to our δ_{SP} -characterization scheme, the algorithmic-cooling scheme focuses more on reducing the effects of δ_{SP} via post-selection from measuring ancilla qubits. After δ_{SP} is reduced to almost 0, the measurement error can be directly measured, and δ_{SP} can then be solved. In our characterization scheme, the state-preparation error and measurement error are both obtained by directly solving the linear equations

Eqs. (6) and (3). However, the algorithmic-cooling scheme requires multiple ancilla qubits, with multiple CNOT gates connected to the target qubit, and, in practice, contains more noise and errors from noisy CNOT gates and is hard to implement if the target qubit is not physically connected with multiple qubits. In our δ_{SP} -characterization scheme, we only require one additional ancilla qubit and one CNOT gate so that the noise from CNOT is much easier to deal with by gate-noise mitigation schemes. Thus, our scheme is more practical to be implemented on current noisy quantum computers.

A.2 Quantifying δ_{SP} via algorithmic-cooling with one ancilla

In this section, we show that it is possible to extract the state preparation error rate in the algorithmic-cooling scheme with one ancilla. We assume for simplicity that the readout error is symmetric (i.e., $\delta_M^0 = \delta_M^1$) as in Ref. [15] or else we can use the symmetrized readout protocol. We will use scripts to label the two qubits 1 and 2. First, we use 1 as the control and 2 as the target and go through the procedure of post-selecting ‘0’ for qubit 2. This reduces the state preparation error of qubit 1 to $\tilde{\delta}_{\text{SP},1} = \delta_{\text{SP},1} f_{12}$, where the factor f_{12} was obtained and defined in Ref. [15], and we quote it here,

$$f_{12} \equiv \frac{2(\delta_{\text{SP},1} + \delta_{M,1} - 2\delta_{\text{SP},1}\delta_{M,1})}{1 + (1 - 2\delta_{\text{SP},1})(1 - 2\delta_{M,1})(1 - 2\delta_{\text{SP},2})}, \quad (45)$$

where the subscripts are used to denote which qubits. The total measured SPAM error (after post-selecting qubit 2) is $\tilde{\delta}_{\text{SPAM},1} = \tilde{\delta}_{\text{SP},1} + \delta_{M,1}$.

Next, we use 2 as the control and 1 as the target and go through the procedure of post-selecting ‘0’ for qubit 1. This reduces the state preparation error of qubit 2 to $\tilde{\delta}_{\text{SP},2} = \delta_{\text{SP},2} f_{21}$, where

$$f_{21} \equiv \frac{2(\delta_{\text{SP},2} + \delta_{M,2} - 2\delta_{\text{SP},2}\delta_{M,2})}{1 + (1 - 2\delta_{\text{SP},2})(1 - 2\delta_{M,2})(1 - 2\delta_{\text{SP},1})}. \quad (46)$$

The total measured SPAM error (after post-selecting qubit 2) is $\tilde{\delta}_{\text{SPAM},2} = \tilde{\delta}_{\text{SP},2} + \delta_{M,2}$. Together with two other measurable quantities, $\delta_{\text{SPAM},i} = \delta_{\text{SP},i} + \delta_{M,i} - 2\delta_{\text{SP},i}\delta_{M,i}$ (for $i = 1, 2$), we have a total of four equations for four unknowns, with which one can solve in principle.

We have performed numerical tests and confirmed that the above equations can be solved. For example, if we measure the four SPAM errors (using the symmetrized readout protocol to average the two readout errors) and obtain the measured results:

$$\{\delta_{\text{SPAM},1}, \delta_{\text{SPAM},2}, \tilde{\delta}_{\text{SPAM},1}, \tilde{\delta}_{\text{SPAM},2}\} \\ = \{0.22923, 0.277084, 0.222717, 0.275828\}. \quad (47)$$

Then, solving the above equations leads to a single solution:

$$\{\delta_{\text{SP},1}, \delta_{\text{M},1}, \delta_{\text{SP},2}, \delta_{\text{M},2}\} = \{0.0242911, 0.215404, 0.0172848, 0.269102\} \quad (48)$$

With additional characterization of the asymmetry in the readout, we can obtain a complete and separate characterization of SPAM errors.

We note that the algorithmic cooling approach here requires solving coupled nonlinear equations, whereas the method presented in the main text requires solving only linear equations.

B Vector representations of single qubit spin-flip error and the proof for the δ_{SP} equation (6)

Consider a bit-flip error, with δ^0 probability flipping the state $|0\rangle$ to $|1\rangle$ and δ^1 for $|1\rangle$ to $|0\rangle$. We can define a vector for such error $D \equiv \{\delta^0, \delta^1\}$, with a map $f(D) = A$ to its corresponding assignment matrix

$$f(D) = f(\{\delta^0, \delta^1\}) = A = \begin{pmatrix} 1 - \delta^0 & \delta^1 \\ \delta^0 & 1 - \delta^1 \end{pmatrix}. \quad (49)$$

The combination of error D_1 on top of error D_2 is defined as the matrix product of their corresponding assignment matrices A_1 and A_2 ,

$$D_1 \circ D_2 \equiv f^{-1}(f(D_1)f(D_2)) = f^{-1}(A_1 A_2) \quad (50)$$

or written explicitly

$$\begin{aligned} D_1 \circ D_2 &= \{\delta_1^0, \delta_1^1\} \circ \{\delta_2^0, \delta_2^1\} \\ &\equiv \{\delta_1^0 + \delta_2^0(1 - \delta_1^0 - \delta_1^1), \\ &\quad \delta_1^1 + \delta_2^1(1 - \delta_1^0 - \delta_1^1)\}. \end{aligned} \quad (51)$$

Note that A is invertible iff $\delta^0 + \delta^1 \neq 1$. Thus, error vectors with the combination rule $(\{D, \delta^0 + \delta^1 \neq 1\}, \circ)$ form a group. $D_a \circ D_b$ is usually not equal to $D_b \circ D_a$, unless $\delta_a^0/\delta_a^1 = \delta_b^0/\delta_b^1$. Particularly, for those errors satisfy $\delta^0 = \delta^1$ (e.g. state-preparation errors (19)), they commute with each other. We can derive the equations (6) as follows

$$\begin{aligned} \tilde{D}_{\text{SPAM}}^a &= D_M^a \circ D_{\text{SP}}^t \circ D_{\text{SP}}^a \\ &= D_M^a \circ D_{\text{SP}}^a \circ D_{\text{SP}}^t \\ &= D_{\text{SPAM}}^a \circ D_{\text{SP}}^t, \end{aligned} \quad (52)$$

where we use the relation $D_{\text{SPAM}}^a = D_M^a \circ D_{\text{SP}}^a$. By measuring D_{SPAM}^a and $\tilde{D}_{\text{SPAM}}^a$, one can solve the $D_{\text{SP}}^t = (D_{\text{SPAM}}^a)^{-1} \circ \tilde{D}_{\text{SPAM}}^a$.

C Obtain δ_{SP} for general uncorrelated CNOT gate noise

For general CNOT gate noise, we consider the case without correlating error, where we assume the noise is effectively a channel applied after the CNOT operation. For general noise channels, one can apply Pauli-twirling to effectively turn the noise channels into Pauli channels. Then the single-qubit error channel can be written as a direct product of two Pauli-like single-qubit channels

$$\Lambda(\rho) = \sum_{a \in \{I, X, Y, Z\}} p_a P_a(\rho) P_a. \quad (53)$$

With same argument, by adding a random Pauli Z operator with 50% probability for each qubit before the CNOT gate, we ensure that the state-preparation error produces effectively $\rho = (1 - \delta_{\text{SP}})|0\rangle\langle 0| + \delta_{\text{SP}}|1\rangle\langle 1|$, on which the Pauli Z channel has no effect but other Pauli channels act effectively like the bit-flip error. For a CNOT gate with qubit q_2 controlling q_1 , the error assignment matrix of q_1 can be written as

$$A_{\text{CNOT}}^{(q_2, q_1)q_1} = \begin{pmatrix} 1 - \delta_{\text{CNOT}}^{(q_2, q_1)q_1} & \delta_{\text{CNOT}}^{(q_2, q_1)q_1} \\ \delta_{\text{CNOT}}^{(q_2, q_1)q_1} & 1 - \delta_{\text{CNOT}}^{(q_2, q_1)q_1} \end{pmatrix}. \quad (54)$$

With this additional error, the procedure of quantifying δ_{SP} becomes more complicated. The circuits we use are as follows:

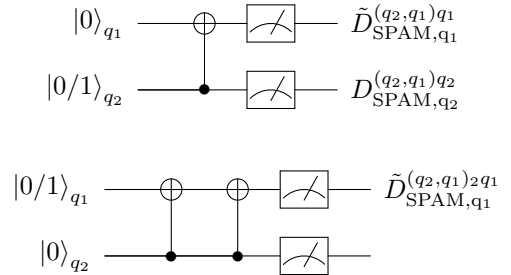


Figure 10: Circuits for measuring δ_{SP} with CNOT gate noise.

We can get the following relations from the circuits

$$\tilde{D}_{\text{SPAM}, q_1}^{(q_2, q_1)q_1} = D_{\text{SPAM}}^{q_1} \circ D_{\text{CNOT}}^{(q_2, q_1)q_1} \circ D_{\text{SP}}^{q_2} \quad (55)$$

$$D_{\text{SPAM}, q_2}^{(q_2, q_1)q_2} = D_{\text{SPAM}}^{q_2} \circ D_{\text{CNOT}}^{(q_2, q_1)q_2} \quad (56)$$

$$\tilde{D}_{\text{SPAM}, q_1}^{(q_2, q_1)2q_1} = D_{\text{SPAM}}^{q_1} \circ (D_{\text{CNOT}}^{(q_2, q_1)q_1})^2 \circ D_{\text{CNOT}}^{(q_2, q_1)q_2} \quad (57)$$

where $D_{\text{SPAM}}^{q_1}$ and $D_{\text{SPAM}}^{q_2}$ can be directly measured. From the equation (56) we can solve the $D_{\text{CNOT}}^{(q_2, q_1)q_2}$. Then by inserting it into (57) we can get $D_{\text{CNOT}}^{(q_2, q_1)q_1}$. Finally, we insert $D_{\text{CNOT}}^{(q_2, q_1)q_1}$ into (55) and get the $D_{\text{SP}}^{q_2}$.

D Detailed comparison for calculating δ_{SP} with and without CNOT gate noise extrapolation

In this appendix, we compare the estimation of δ_{SP} from noisy-CNOT results (11) and mitigated-CNOT results. To mitigate CNOT gate noise, we use the zero-noise extrapolation technique [5], where CNOT gates are repeated $m = 1, 3, 5, \dots$ times to get different noise-scale results. After that, we use linear extrapolation to extrapolate them to zero-noise ($m = 0$) point and get the mitigated data. The extrapolating results for calculating $\delta_{\text{SP}}^{(12)}$ via ancilla qubit q_{10} of ibmq_mumbai are shown in Fig. 11. From the mitigated results, we get the $\delta_{\text{SP}}^{(12)} = 0.041(4)$, which is very close to the results from unmitigated data $\delta_{\text{SP}}^{(12)} = 0.042(2)$ (shown in Sec. 2.2).

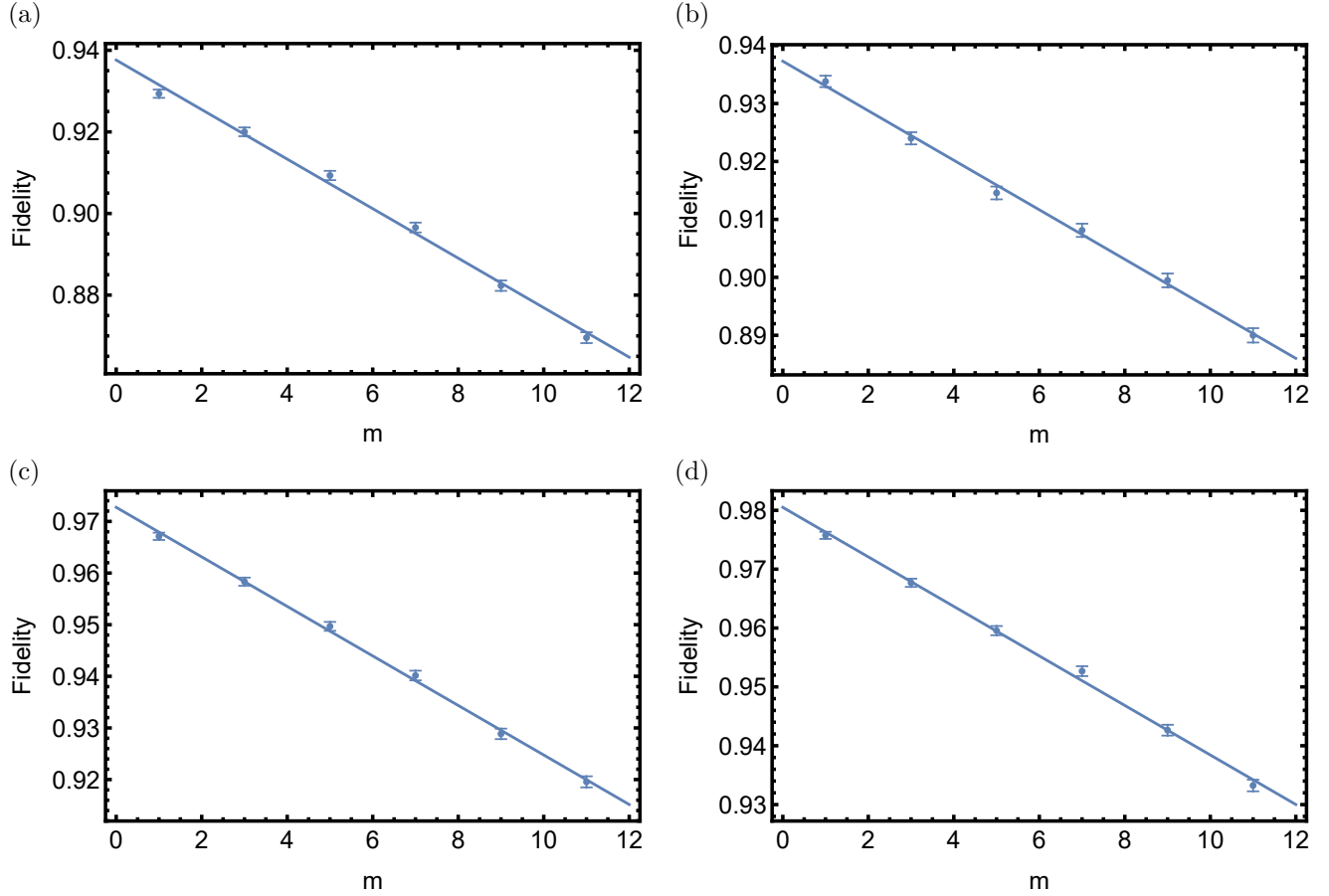


Figure 11: Zero-noise extrapolation for δ_{SP} characterization circuit with q_{12}, q_{10} on ibmq_mumbai. Each circuit is repeated 64000 times. (a) Fidelity of q_{10} with initial state $|00\rangle$ for circuit 1 with q_{12} as target qubit. The fitted result gives $1 - \tilde{\delta}_{\text{SPAM}}^{10,0,\text{miti}} = 0.938(1)$. (b) Fidelity of q_{10} with initial state $|01\rangle$ for circuit 1 with q_{12} as target qubit. The fitted result gives $1 - \tilde{\delta}_{\text{SPAM}}^{10,1,\text{miti}} = 0.937(1)$. (c) Fidelity of q_{10} with initial state $|00\rangle$ for circuit 1 with q_{10} as target qubit. The fitted result gives $1 - \delta_{\text{SPAM}}^{10,0,\text{miti}} = 0.973(1)$. (d) Fidelity of q_{10} with initial state $|01\rangle$ for circuit 1 with q_{10} as target qubit. The fitted result gives $1 - \delta_{\text{SPAM}}^{10,0,\text{miti}} = 0.980(1)$.