

## Benchmarking the quality of multiplexed qubit readout beyond assignment fidelity

Andras Di Giovanni<sup>1,\*</sup>, Adrian Skasberg Aasen<sup>1,2,3</sup>, Jürgen Lisenfeld<sup>1</sup>, Martin Gärttner<sup>1,3</sup>, Hannes Rotzinger<sup>1,4</sup>, and Alexey V. Ustinov<sup>1,4</sup>

<sup>1</sup>Physikalisches Institut, Karlsruher Institut für Technologie, Kaiserstraße 12, 76131 Karlsruhe, Germany

<sup>2</sup>Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany

<sup>3</sup>Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany

<sup>4</sup>Institut für QuantenMaterialien und Technologien, Karlsruher Institut für Technologie, Kaiserstraße 12, 76131 Karlsruhe, Germany

(Received 21 February 2025; revised 30 May 2025; accepted 15 September 2025; published 15 October 2025)

The accurate measurement of quantum two-level objects (qubits) is crucial for developing quantum computers. Over the last decade, the measure of choice for benchmarking readout routines for superconducting qubits has been assignment fidelity; however, this method only focuses on the preparation of computational basis states and therefore does not provide a complete characterization of the readout. Here, we expand the focus to the use of detector tomography to fully characterize multiqubit readout of superconducting transmon qubits. The impact of different readout parameters on the rate of information extraction is studied using quantum-state reconstruction infidelity as a proxy. The results are then compared with assignment fidelities, showing good agreement for separable two-qubit states. We therefore propose the rate of infidelity convergence as a validation tool for assignment fidelity and a more comprehensive benchmark for single- and multiqubit readout optimization. To make the best use of experimental resources, we investigate the most efficient distribution of a limited shot budget between detector tomography and state reconstruction within the context of single- and two-qubit experiments. To address the growing interest in three-qubit gates and test the scalability of the validation tool, we perform three-qubit quantum-state tomography that goes beyond conventional readout-error-mitigation methods and find a factor-of-30 reduction in quantum infidelity. Our results demonstrate that qubit readout correlations are not induced by a significantly reduced state distinguishability. Consequently, correlation coefficients can serve as a valuable tool in qubit readout optimization.

DOI: 10.1103/dpft-lxtx

### I. INTRODUCTION

State-of-the-art quantum computing experiments have started pushing the boundary to fault tolerance [1,2]. Although gate-error-mitigation approaches such as Pauli twirling or zero-noise extrapolation [3,4] remain important for most experiments, these methods have been shown to be of limited use to extract noiseless expectation values [5]. Significant research effort has gone into optimizing single- and multiqubit gate fidelities [6–9]; however, it is becoming increasingly clear that readout errors also

pose an important obstacle to general-purpose quantum computing [10]. On the analog quantum simulation side, precise readout is just as important to prevent biasing of reconstructed observables. Therefore, recent research has focused on understanding [11,12] and improving the multiplexed dispersive readout for superconducting qubits [13,14]. In this work, we aim to contribute to understanding and mitigating readout errors with the help of quantum detector tomography [15–18]. This approach assumes that state preparation in the quantum system is accurate, and it seeks to mitigate readout errors by reconstructing the experimental measurement operator instead of assuming that the readout is accurately modeled by projective computational basis measurements.

In particular, correlated readout errors have recently been observed in superconducting qubit systems. References [19,20] consider correlated bitflip errors and energy relaxation, respectively. Both discuss physical mechanisms

\*Contact author: andras.digiovanni@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](#). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

that may occur during readout, effectively manifesting as correlated readout, which would be captured by readout correlation coefficients. A further example could be fluctuations in the mutual electric ground. Numerical simulations of correlated readout errors from experiments and noise models have been reported in Ref. [21]. Furthermore, Ref. [22] discusses methods to decrease electromagnetic crosstalk enabled readout correlations in superconducting flux qubits. Reference [14] provides experimental evidence for qubit dephasing due to correlated readout errors.

We apply a comprehensive readout-error-mitigation protocol, introduced in Ref. [23], to a multiqubit quantum-state tomography experiment. The protocol mitigates both coherent and correlated readout errors, going beyond traditional techniques that address only classical readout errors [17,24,25] and enables very precise state reconstruction. Our experiments are implemented on a superconducting qubit chip with four transmons, which we manipulate and read out using frequency-division multiplexing [26]. This architectural choice simplifies wiring, but it can be prone to significant readout crosstalk, and it is hence an ideal testbed for multiqubit readout-error mitigation. By employing a traveling-wave parametric amplifier (TWPA) [27], we can perform single-shot readout of the qubits. We introduce a readout quality metric that uses the number of shots necessary to reach a specified state reconstruction infidelity, quantifying the efficiency of extracting information from the system. The proposed metric takes into account coherent errors, also referred to as quantum errors, which we compare with assignment fidelities, defined as the probabilities of correctly assigning labels 0 or 1 to the prepared computational basis states.

Given the substantial number of repetitive measurements required for detector tomography, it is important to avoid spending time on overly precise calibration of the reconstructed positive operator-valued measures (POVMs). Therefore, we also investigate different distributions of a fixed shot budget between quantum detector tomography (QDT) and quantum-state tomography (QST) to determine which results in the lowest achievable infidelity. Furthermore, we demonstrate two- and three-qubit density matrix reconstruction that takes into account coherent and correlated readout noise [23], which opens the way to reducing readout errors in complete benchmarking of two- and three-qubit gates. This is motivated by renewed interest in multiqubit entangling gates in the community [28–30].

An advantage of detector tomography is that it explicitly reconstructs the POVM corresponding to the multiqubit readout process. This enables extraction of multiqubit readout correlation coefficients of the system, regardless of their origin, which could include off-resonant drive of the resonator, mechanical vibrations [19], or ionizing radiation [31]. In multiplexed readout, a perfect measurement of a qubit should not depend on the state of its neighbors, which

makes readout correlation a suitable indicator of readout quality [22].

The readout correlation coefficients should be independent of the state distinguishability in the readout plane. By lowering the readout pulse amplitude, effectively lowering state distinguishability, we experimentally demonstrate that readout correlation coefficients satisfy this property.

## II. PRELIMINARIES

### A. Generalized measurements

To characterize a noisy detector, it is essential to adopt a framework that goes beyond projective measurement. This approach can be formulated by a set of operators possessing the following properties:

$$M_i^\dagger = M_i, \quad M_i \geq 0, \quad \text{and} \quad \sum_i M_i = \mathbb{1}, \quad (1)$$

where  $\mathbf{M} = \{M_i\}$  is the POVM with elements  $M_i$ . When taking the expectation values of the measurement operators, one can retrieve the probabilities of obtaining different outcomes  $i$  in a measurement process,

$$p_i = \text{Tr}(\rho M_i). \quad (2)$$

### B. Detector tomography

Characterization of the noisy measurement device is achieved by performing detector tomography [15]. The goal of detector tomography is to reconstruct the generalized measurement operators  $\mathbf{M} = \{M_i\}$  that satisfy the constraints in Eq. (1) through experimentally measured outcome probabilities

$$\frac{n_{is}}{N} = p_{is} = \text{Tr}(\rho_s M_i), \quad (3)$$

where  $N$  is the total number of measurements and  $n_{is}$  is the number of outcomes  $i$  observed from calibration state  $\rho_s$ . To enable full detector tomography, the set of calibration states must be informationally complete. In our experiment, we chose the symmetric and informationally complete tetrahedron states

$$\begin{aligned} |\psi_1\rangle &= |0\rangle, \\ |\psi_2\rangle &= \frac{1}{\sqrt{3}} |0\rangle + \sqrt{\frac{2}{3}} |1\rangle, \\ |\psi_3\rangle &= \frac{1}{\sqrt{3}} |0\rangle + \sqrt{\frac{2}{3}} e^{i\frac{2\pi}{3}} |1\rangle, \\ |\psi_4\rangle &= \frac{1}{\sqrt{3}} |0\rangle + \sqrt{\frac{2}{3}} e^{i\frac{4\pi}{3}} |1\rangle. \end{aligned} \quad (4)$$

With the full set of measurement probabilities in Eq. (3), one can estimate the most likely set of operators through

a maximum-likelihood estimator, as outlined, for example, in Ref. [32].

### C. Readout-error-mitigated quantum-state tomography

With a reconstruction of a noisy, experimentally measured POVM, it is possible to use the POVM to construct a readout-error-mitigated density matrix. The approach we use is described in Ref. [23] and is based on directly integrating the full set of estimated measurement operators into a density matrix estimator. The protocol is specifically based on estimating the density matrix using the adapted likelihood function,

$$\mathcal{L}_{M^{\text{estm}}}(\rho) \propto \Pi_i \text{Tr}(\rho M_i^{\text{estm}})^{n_i} = \left( \Pi_i (p_i)^{\hat{p}_i} \right)^N, \quad (5)$$

with either a maximum-likelihood estimator [33] or a Bayesian-mean estimator [34].

This approach allows for beyond-classical readout-error mitigation as it naturally accommodates coherent errors, and it does not require any channel inversion or numerical optimization while still guaranteeing physical density matrix reconstruction.

### D. Multiplexed qubit readout

Multiplexed readout [26] is a central architectural choice for superconducting quantum processors. By coupling multiple resonators to a single transmission line, one can save circuit footprint on the chip. In conventional transmon processors, transmission lines account for a significant chip surface area. As resonators now share their transmission line, their readout pulses are sent through the cryostat on the same microwave line, and they are also amplified and mixed simultaneously after reading out the qubit state. This architectural choice needs fewer amplifiers and other microwave equipment; however, it has a significant drawback: resonators on a single transmission line cannot be too close in frequency, otherwise they will be measured with crosstalk. A readout pulse synthesized for one resonator will also drive a neighboring one, leading to a bias in the derived qubit state. Recently, deep-learning methods have been explored to increase single-shot assignment fidelity in multiplexed readout operations [35].

### E. Quantifying correlated readout errors

A major challenge in scaling up quantum hardware is induced correlations and crosstalk, e.g., due to multiplexed qubit readout. A useful tool for characterizing readout correlations was introduced in Refs. [12,18], where, by using reconstructed two-qubit POVMs, asymmetric correlation coefficients  $c_{j \rightarrow i}$  could be extracted. These coefficients quantify how much the measurement results from qubit  $i$

are affected by the state of qubit  $j$ . For a more detailed discussion of the definition of the correlation coefficients, see Appendix A.

Since the correlation coefficients are asymmetric, we also define a symmetric version,  $c_{i \leftrightarrow j} = (c_{j \rightarrow i} + c_{i \rightarrow j})/2$ . The coefficients can be modified to measure only the classical correlations, represented as  $c_{i \leftrightarrow j}^C$ , which are determined exclusively by the diagonal elements in the measurement operators. The full correlation coefficients contain the effect of both the diagonal and off-diagonal elements, which we call quantum correlation coefficients, represented by  $c_{i \leftrightarrow j}^Q$ .

### F. Infidelity

To quantify the quality of a reconstructed density matrix, we use the quantum infidelity,  $I(\rho, \sigma)$ , as a figure of merit. In the case of pure target states  $\sigma$ , the infidelity with respect to the estimated state  $\rho$  is given by

$$I(\rho, \sigma) = 1 - \text{Tr}(\rho\sigma). \quad (6)$$

## III. EXPERIMENTAL REALIZATION

### A. Experimental setup

We investigate a device with four transmon qubits [36], each coupled individually to a resonator for readout. The device is measured below 15 mK in a dry dilution cryostat. The frequencies of the qubits and resonators, as well as typical coherence times can be found in Appendix B. We define System A as the two-qubit subsystem comprising q0 and q1, System B as the subsystem comprising q1 and q2, and finally System C as the subsystem comprising q2 and q3, as shown in Fig. 1. A description of the readout chain in the context of readout errors is found in Appendix C.

### B. Experimental protocol

To reconstruct the POVM that describes the measurement process, we perform detector tomography. We prepare calibration states on the qubits using single-qubit gates. The calibration states are explicitly given in Sec. II B. For the phase offset between the state-preparation and readout control pulses, we use virtual  $Z$  gates. The calibration states are measured in a set of different bases corresponding to the projectors onto the eigenbasis of the three Pauli operators  $X$ ,  $Y$ , and  $Z$ . The POVM is then reconstructed in postprocessing according to Ref. [23].

Once the detector tomography has been completed, we prepare different quantum states to which we can apply quantum readout-error mitigation. The prepared states are randomly selected from the set of product states. Each single-qubit state is generated by uniformly sampling from the surface of the corresponding Bloch sphere, referred to as Haar random states. Haar random states are created by

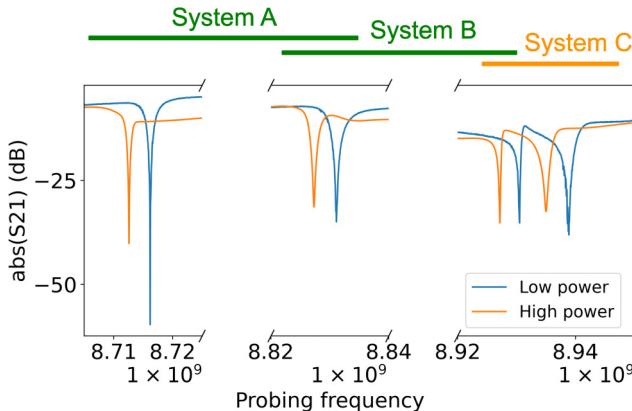


FIG. 1. Resonator spectroscopy of the four-qubit system. At sufficiently low readout powers, each qubit couples coherently to its resonator, enabling dispersive readout. Systems A, B, and C are two-qubit systems of neighboring resonator-qubit systems. An important feature is the closeness of resonators in System C, symbolized by the orange color. From left to right, the corresponding qubits are  $q_0$ ,  $q_1$ ,  $q_2$ , and  $q_3$ .

applying a random unitary matrix  $U$ , drawn at random from the unitary group  $U(2^n \text{qubits})$ , to the ground state  $|0\rangle$  [37].

### C. Inducing readout noise

To study the effect of readout noise on detector tomography and its expectation values, we induce noise that does not affect state preparation and takes effect only after the quantum state of interest has been prepared. The dominant noise sources we investigate stem from suboptimal (i.e., weaker or stronger than ideal) readout power of the resonators and insufficient amplification in the readout chain.

The amplitude of the readout pulses can be adjusted individually or simultaneously for multiple qubits. Generally, if the readout amplitude is too low, the measurement may not be fully projective, and reliable state distinguishability is not possible. In the case of very high readout amplitude, we excite states outside the qubit manifold, biasing the data [13].

We sweep the frequency and amplitude of the continuous pump tone of the parametric amplifier. This results in varying amplification and therefore modifies the overlap of basis states in the IQ (in phase and quadrature signal) plane.

## IV. READOUT-ERROR-MITIGATED MULTIQUBIT STATE TOMOGRAPHY

We apply multiqubit readout-error mitigation to the system described in the previous section. We use a generalization of the experimental protocol depicted in Fig. 1 of Ref. [23] to multiqubit systems. We study the rate of information extraction from readout-error-mitigated quantum-state

tomography experiments. The rate of information extraction serves a proxy for the total readout noise in the system, which includes both correlated and coherent errors. We use this quantity as a quality metric for the readout, which we compare to assignment fidelity. Assignment fidelity is the standard readout quality metric, but it captures only readout errors in the computational basis. In particular, we measure the number of experimental shots required to reach a given reconstruction infidelity threshold in Sec. IV B. We then discuss the optimal distribution of a fixed shot budget between the different stages of the scheme in Sec. IV C. Finally, we apply the protocol to full three-qubit state tomography to assess the scalability of the approach (Sec. IV B 1).

### A. Extracting infidelity threshold crossings from Bayesian-mean estimated state tomography

We quantify the rate of information extraction by how many shots are required to reach a certain infidelity threshold in a state reconstruction experiment. To motivate the use of readout-error mitigation instead of unmitigated quantum-state tomography, we investigate the infidelity curves of both cases in our two-qubit experiment. Figure 2 shows four mean infidelity curves, two quantum readout-error mitigated (QREM, blue) and two unmitigated (std. QST, red), as functions of the number of shots used for state tomography. Each curve is averaged over 16 products of single-qubit Haar random states [38]. The crossing points of these curves with a given constant infidelity (0.20, 0.15, or 0.10) are extracted using a Bayesian-mean estimator [23]. This allows one to identify the exact number of shots at which the infidelity curve crosses a specified infidelity threshold value. We note that the unmitigated curves saturate at high infidelity values. In fact, at high levels of noise, they do not even cross these thresholds (green horizontal lines). This makes them inappropriate for precisely benchmarking readout quality. If one uses the readout-error-mitigated infidelity curves (depicted in blue) instead, the infidelity thresholds are crossed in the steady exponential regime (specifically, in the linear segment of the log-log plot) of the curves, making them suitable for benchmarking. Furthermore, infidelity curves are most accurately obtained when employing a Bayesian-mean estimator, highlighting an advantage over maximum-likelihood estimators in the reconstruction process. Bayesian-mean estimators can be efficiently computed for two qubits, after which, maximum-likelihood estimators gain an additional advantage in speed.

### B. Readout noise affecting information-extraction rate

To investigate how the rate of quantum information extraction is influenced by readout noise, we perform readout-error-mitigated two-qubit quantum-state tomography experiments on System A (see Fig. 1) with induced

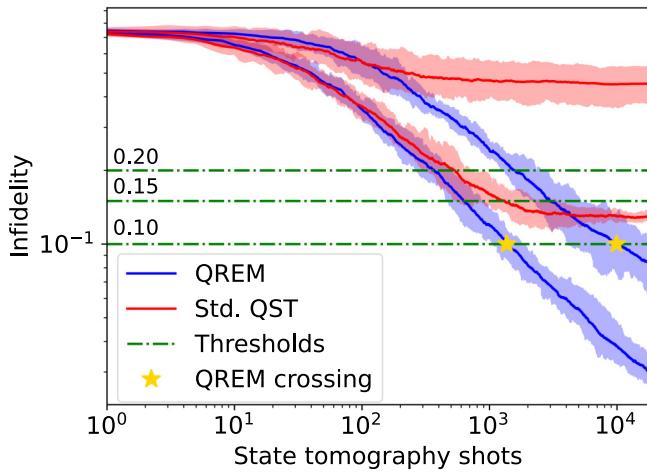


FIG. 2. Mean infidelity curves crossing different threshold values (corresponding to infidelity values of 0.10, 0.15, and 0.20). Two different readout parameters are examined, where blue curves are with and red curves are without quantum readout-error mitigation (QREM). The curves are created using a Bayesian-mean estimator of the reconstructed quantum states. Only readout-error-mitigated infidelities reliably cross the given thresholds, where the lowest threshold was only crossed by the QREM curves. Each curve is averaged over 16 Haar random initial states, and the shaded area indicates the interquartile range.

readout noise. We compare the results with a corresponding analysis using assignment fidelity as a readout quality metric. Since assignment fidelity depends exclusively on computational basis states and does not fully benchmark readout fidelity, our analysis aims to confirm the suitability of using assignment fidelity as a measurement standard.

### 1. Varying readout power

We vary the readout power for both resonators simultaneously, so that they are equal at all times. This leads to the possible existence of multiple optima for information extraction. The upper figure in Fig. 3 illustrates the number of shots necessary to achieve the infidelity thresholds. The plot shows one optimum for information-extraction rates at 0.48 (arb. units) readout amplitude, illustrated by a red dashed line. This is in agreement with both the product and the sum of single-qubit assignment fidelities (see the bottom panel in Fig. 3). The optimum obtained for infidelity thresholding is statistically significant, whereas the measured assignment fidelities are also consistent with a plateau between readout amplitudes of 0.35 and 0.48. This is due to the relatively large error bars, limited by the number of computational basis states measured. It is expected that repeating the experiments with more shots would also highlight one clear optimum at 0.48 (arb. units) readout power. The general features of the readout parameter landscape are explained by low state distinguishability at low

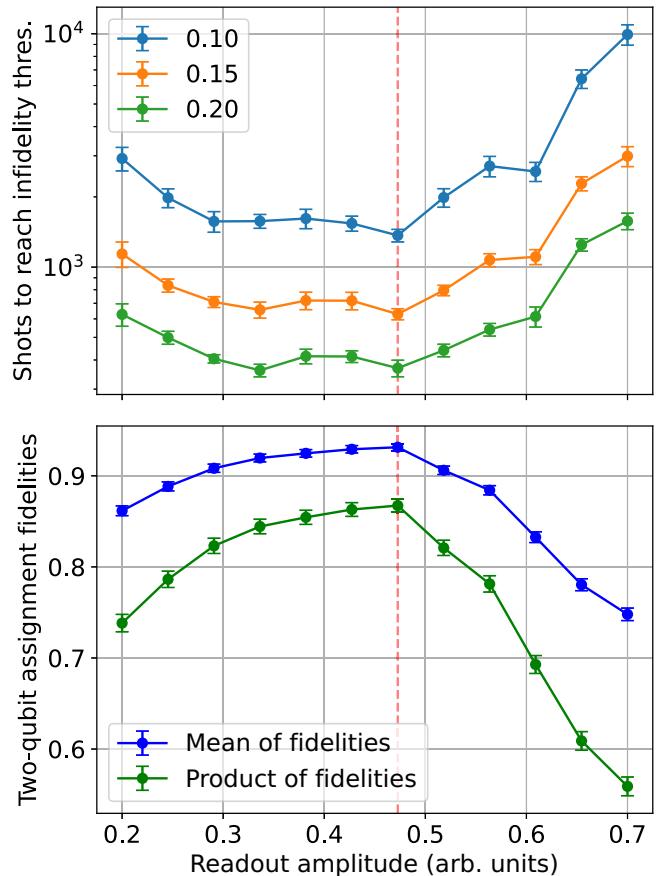


FIG. 3. Comparison of quantum-state-tomography-based readout benchmark with assignment fidelities for varying readout amplitudes. The vertical dashed red line indicates optimal readout amplitude. Top: number of two-qubit state tomography shots required to reach infidelity thresholds of 0.10, 0.15, and 0.20 for various readout amplitudes. The error bars are bootstrapped standard deviations. For very low readout powers, the distinguishability of the basis states is significantly reduced. For very high readout amplitudes, the measurement process excites higher states in the system, which the state classifier cannot correctly take into account. Each point uses 558 000 shots. Bottom: mean and product of single-shot assignment fidelities. Each point uses 2000 shots.

readout powers and state leakage at higher readout powers, as also observed, for example, in Ref. [13]. Both the assignment fidelity and infidelity convergence landscapes are similar and well explained by the above reasoning. The fraction of leaked states at higher readout power is plotted in Appendix D. In summary, the experiment demonstrates that the more comprehensive infidelity thresholding aligns with assignment fidelities at the optimal readout power for two-qubit multiplexed readout. Given the higher cost associated with infidelity thresholding without a corresponding increase in accuracy, assignment fidelity is preferred in this setup.

## 2. Parametric amplification

The resonators of System A are detuned by about 115 MHz from each other; therefore, it is expected that they could have different optimal operating parameters on the shared TWPA. This is due to the nontrivial frequency dependence of the gain, as measured in Ref. [27]. Thus, it is not sufficient to test the TWPA on a single qubit. Performing state tomography on this system (see Fig. 4), we see that information extraction is fastest at around 6.0 dBm pump power. This is in good accordance with the maximum of both the sum and product of single-shot assignment fidelities. Once the TWPA is oversaturated in pump power, a significant slowing of information extraction by about a factor of 6 compared to the optimum is observed (for a more detailed measurement of this saturation, see Appendix E). At low pump powers, the information-extraction rate is slowed because of insufficient distinguishability of the computational basis states at readout, which can also be observed in the corresponding assignment fidelities.

Similar to varying the readout power, both infidelity thresholding and assignment fidelity agree on the optimal pump power.

## C. Optimal shot budget distribution

Experiments often have to be limited in run time. This can be due to experimental drift or simply resource sharing. This raises the question of how one should optimally allocate a fixed measurement budget between detector tomography and state tomography to minimize the resulting infidelity in the reconstruction of a quantum state. We define the shot ratio as

$$r = \frac{\text{QDT shots}}{\text{QDT shots} + \text{QST shots}}, \quad (7)$$

i.e., the ratio of the shots used for detector tomography to the total number of shots performed in the reconstruction of a single state.

To find the optimal ratio for our experiment, we performed single- and two-qubit tomography with a fixed number of total shots for different values of  $r$ . For a representative convergence of the state reconstruction, an average of the same 40 Haar random states was used for each value of  $r$  for single-qubit reconstruction. For two-qubit shot budgeting, ten products of single-qubit Haar random states were used.

The results of the experiments can be seen in Fig. 5. For both single- and two-qubit readout-error-mitigated state tomography, spending half the shot budget on calibration (QDT) is a reasonable choice. In the case of very few shots being spent on calibration, the resulting infidelity is limited by sample fluctuations in the estimated measurement operator. On the upper end, if too few shots remain for

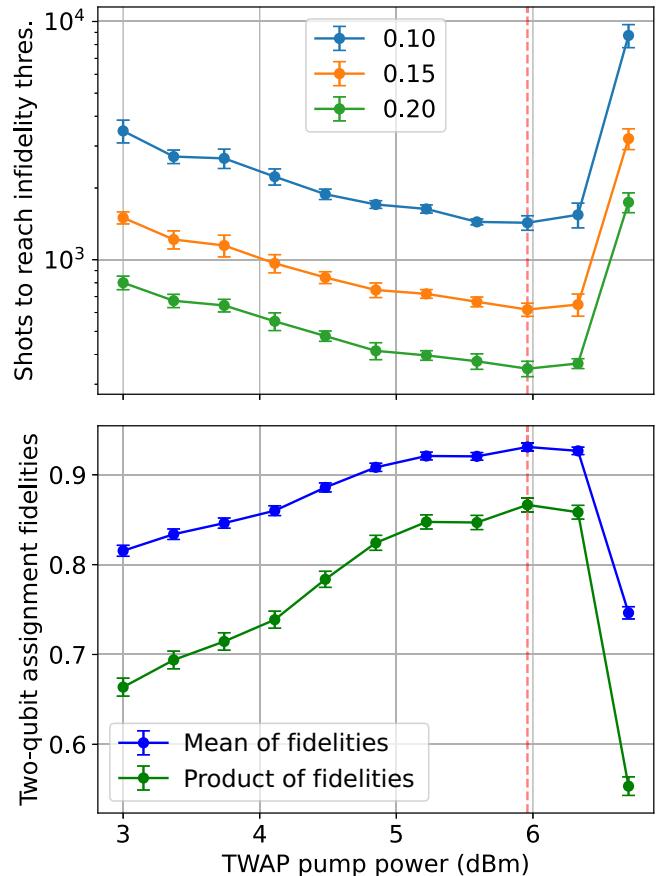


FIG. 4. Comparison of quantum-state-tomography-based readout benchmark with assignment fidelities for varying amplification. The vertical dashed red line indicates optimal pump power. Top: number of shots in state tomography needed to reach 0.10, 0.15, and 0.20 reconstruction infidelity. The error bars show bootstrapped standard deviations. Each point uses 558 000 shots. Bottom: sum and product of assignment fidelities for the performed experiments. State separation increases with pump power up to approximately 6.5 dBm, after which it drops sharply. The error bars show propagated standard deviations from the assignment fidelity. Each point uses 2000 shots.

state tomography, the experiment will not reach the lowest infidelities. The ideal shot ratio ranges are highlighted in green in Fig. 5. From these plots, we can estimate the optimal value of  $r$ , as defined in Eq. (7), to be

$$r_1 = 0.4 \pm 0.2$$

for a single qubit and

$$r_2 = 0.6 \pm 0.3$$

for two-qubit experiments. In Appendix F, we consider a simple theoretical model based on the number of independent parameters to be estimated. Our predicted optima are at larger values than the observed ones, indicating that

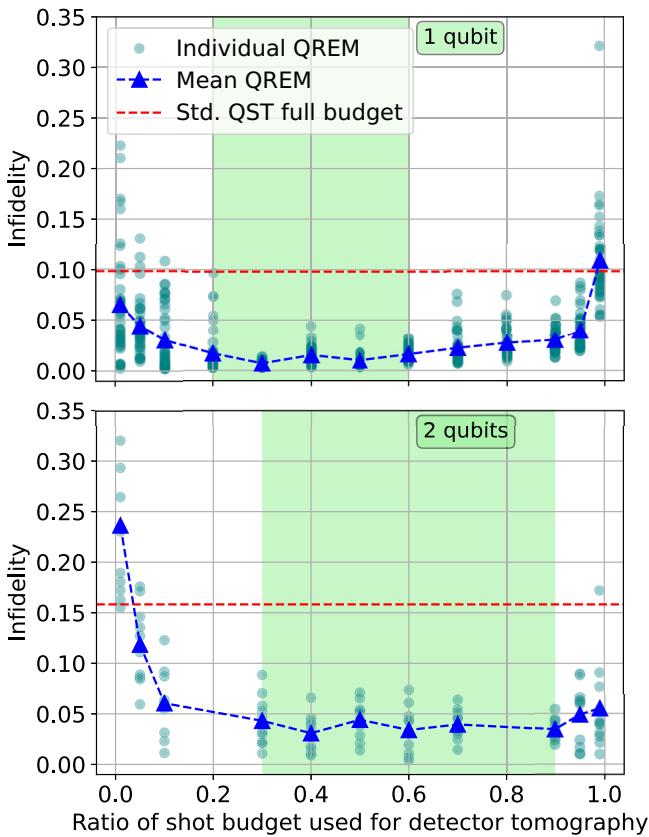


FIG. 5. Reachable infidelities for different distributions of a given shot budget in readout-error-mitigated state reconstruction. The plotted infidelities refer to the final infidelity at the end of quantum-state reconstruction. The blue triangles correspond to the mean infidelities, while the spread in each case is shown with teal circles. The red horizontal line indicates the mean final infidelity at which the full budget was used for standard state tomography. The identified optimal shot ratio is marked in green. Top: single qubit, shot budget of 10 000, averaged over 40 Haar random states. Bottom: two-qubit reconstruction with a shot budget of 40 000, averaged over ten Haar random states.

there is a significant amount of structure, which likely simplifies the statistical estimation task.

#### D. Three-qubit state tomography

In light of the renewed interest in three-qubit gates, noted for their potentially higher fidelity, shorter duration, and ability to replace multiple two-qubit gates [28,29,39], we perform three-qubit quantum-state tomography with and without readout-error mitigation. This examination aims to evaluate the scalability of the protocol introduced in Ref. [23] to more qubits. The experiment was repeated on the same prepared state with an increasing number of shots per basis measurement in both detector tomography and state tomography. The results are plotted in Fig. 6. Infidelities below 0.01 were reached, with an outlier data point at 12 000 shots per basis measurement likely

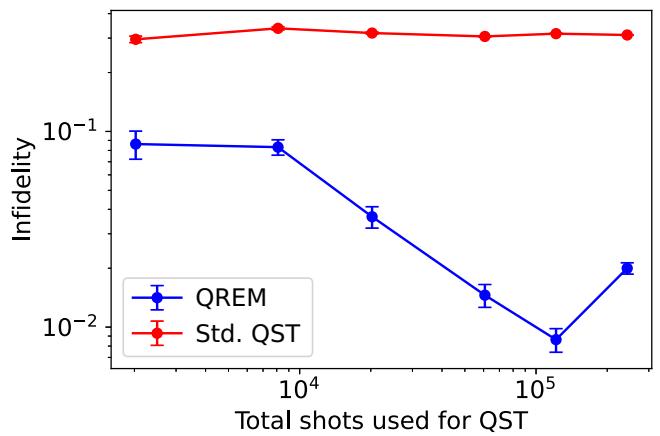


FIG. 6. Reconstruction of a three-qubit density matrix with and without quantum readout-error mitigation (QREM) as a function of the total number of shots used for QST. The error bars indicate the standard deviation of the estimator obtained from 1000 bootstrap resamplings.

attributable to measurement drift, which we investigate further in Appendix G. A comparison of the theoretically expected density matrix to the reconstructed one with and without readout-error mitigation can be found in Appendix H Fig. 12. We observe an infidelity improvement of around a factor of 30 when comparing state tomography with and without readout-error mitigation. Therefore, we conclude that the method introduced in Ref. [23] is scalable to three superconducting qubits and is capable of significantly reducing readout errors. This provides a way to benchmark and understand three-qubit gates, even in the presence of coherent readout errors in the system. Complete state reconstruction is intrinsically not a scalable method; however, with the established applicability to three-qubit benchmarking, the technique can be applied to characterize few-qubit subsystems within larger qubit systems, which is typically sufficient for mitigation of correlated errors [12,21,25].

### V. TWO-QUBIT CORRELATIONS FROM DETECTOR TOMOGRAPHY

As multiplexed readout has become widespread in superconducting qubit devices of growing size, it is important to reliably quantify potential readout correlations that exist due to limited resonance frequency spacing. In this section, we use correlation coefficients extracted from detector tomography to investigate our qubit system. In addition, we analyze the dependence of these correlation coefficients on the amplitude of the readout pulse.

#### A. Correlation measurements for two-qubit subsystems

To investigate the effect of resonator frequency spacing, we performed detector tomography on Systems A

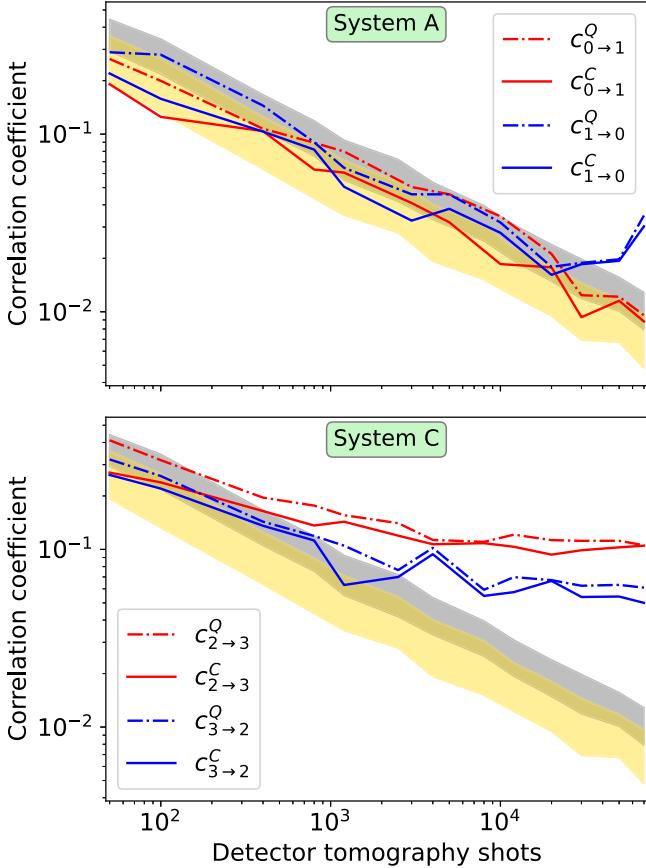


FIG. 7. Correlation coefficients plotted as a function the total number of detector tomography shots used for reconstructing the measurement operators. The silver and gold shaded areas represent the standard deviations of quantum and classical correlation coefficients, respectively, derived from numerical simulations of measurements without correlation. Curves that fall within these regions indicate compatibility with uncorrelated readout. Top: in System A, the readout correlations do not exhibit any readout correlations within the bounds of the experimental constraints. Bottom: System C displays nonzero asymmetric correlations, which suggests that readouts are indeed correlated within the system; however, no coherent correlations were observed within the limitations of the experiment.

and C (see Sec. III A) and extracted both quantum and classical correlation coefficients as functions of detector tomography shots. The results are plotted in Fig. 7. At low numbers of shots, the correlation coefficients are compatible with uncorrelated POVMs (shaded silver and gold areas), indicating that the coefficients are dominated by sample fluctuations. System A is always compatible with uncorrelated readout, while System C shows strong asymmetric readout correlations. At high numbers of shots for System A, the  $0 \rightarrow 1$  correlation coefficients move outside the standard deviation of uncorrelated POVMs, but due to low statistics, we consider this to be consistent

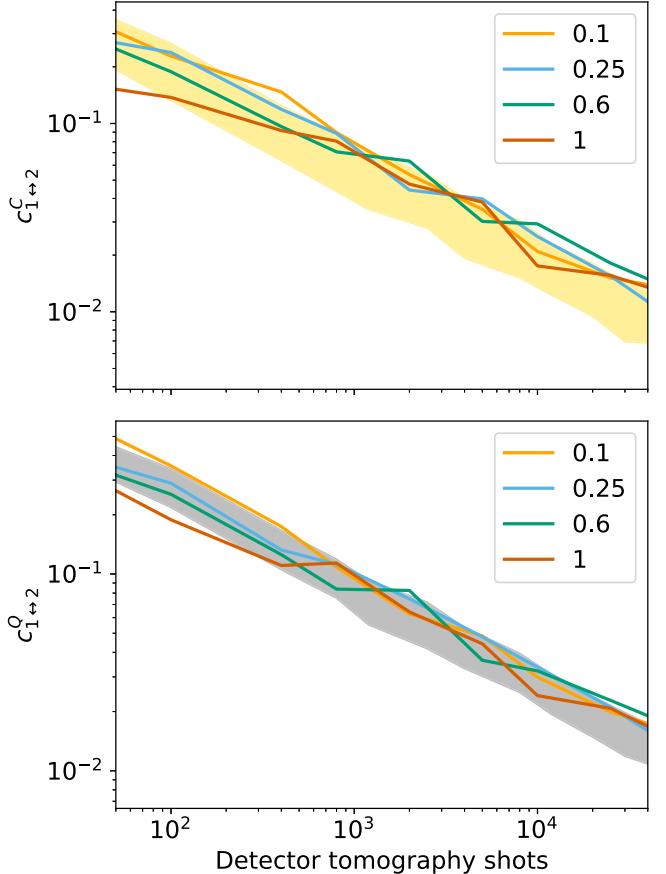


FIG. 8. Correlation coefficients in System A as a function of the number of detector tomography shots for lowered relative readout amplitudes. The silver and gold shaded areas represent the standard deviations of quantum and classical correlation coefficients, respectively, derived from numerical simulations of measurements without correlation. Top: classical correlations decrease independently of the relative readout amplitude, remaining consistent with uncorrelated readout. Bottom: quantum correlations have larger values but are always consistent with uncorrelated readout.

with uncorrelated POVMs. These observations are consistent with the observed frequency spacing of resonators in the two systems:  $\delta_B = 100$  MHz vs  $\delta_C = 8$  MHz. Neither system exhibits significant correlated coherent readout errors. In System C, the correlation coefficients are of the order of 0.1, comparable to those reported for the Rigetti Aspen-M-3 in Ref. [12]. Reference [21] presents numerical simulations of noisy readout using both experimental and synthetic systems. Fitting the probabilistic *i*SWAP channel  $\mathcal{E}(M_i) = (1 - k)M_i + kU_{i\text{SWAP}}M_iU_{i\text{SWAP}}^\dagger$  to the observed correlation coefficients in System C is consistent with a noise strength of  $k \approx 0.1$ .

We therefore conclude that the correlation coefficients can be used as a precise tool to test whether resonators are sufficiently far from each other both spatially on the chip

and in the frequency domain. As correlations are dependent on *in situ* readout variables such as readout power, it is possible to optimize these parameters to maximize the rate of information extraction while simultaneously minimizing the correlations between different qubit readouts. Our results therefore open the way for a more comprehensive optimization of multiqubit readout.

### B. Two-qubit correlation dependence on readout noise strength

To better understand how the correlation coefficients behave under additional noise sources, we investigated how the coefficients change when lowering the distinguishability of computational basis states using readout pulses with varying amplitudes. Lowered state distinguishability can be modeled as depolarizing-type noise on each qubit, which is not expected to impact the correlation coefficients; see Appendix J for more details.

We define the relative readout amplitude as the ratio between the used amplitude of the readout tone and its ideal amplitude. To avoid leakage to higher levels, we restrict ourselves to decreased readout amplitudes, corresponding to relative readout amplitudes not larger than 1. *Ad extreamum*, the single-shot assignment fidelity measured with a relative readout amplitude of 0.1 is only about 55% compared to around 95% at a readout ratio of 1.

The results of the experiments can be seen in Fig. 8. Within the bounds of the experiment, both quantum and classical correlations are independent of the strength of the readout noise, even at very high levels of noise, confirming our theoretical expectations.

## VI. CONCLUSIONS AND OUTLOOK

We investigated a generalized method for readout-error mitigation based on detector tomography and explored its utility as a benchmarking tool for superconducting qubits. The protocol was applied to various two- and three-qubit state reconstruction experiments with superconducting transmon qubits. We introduced a benchmark for qubit readout quality that goes beyond conventional techniques focused on the fidelity of computational basis state assignments, which primarily detect classical errors. Our approach uses quantum infidelities between prepared states and readout-error-mitigated reconstructed states as a proxy for the information-extraction rate. Using reconstruction infidelity characterizes both coherent and correlated readout errors, making it a more comprehensive benchmark than assignment fidelity. We thus propose using the rate of infidelity threshold crossings as a validation tool for assignment fidelity and as a suitable comprehensive figure of merit for certifying single- and multiqubit readout processes. In particular, when optimizing the readout chain, assignment fidelity fails to properly capture correlated

readout errors, while infidelity-thresholding-based benchmarking is able to quantify the correlations present in the measurement process. Therefore, if assignment fidelity falls short, this new figure of merit can be used to find optimal operating parameters for readout chains, such as fine-tuning amplifiers for ideal readout performance. We studied the impact of different readout parameters on the rate of information extraction. In particular, the dependence on the pump power of the parametric amplifier and the amplitude of the readout resonator drive tone were investigated. The assignment fidelity- and infidelity-convergence-based methods produced very similar results in the investigated parameter range, validating the use of assignment fidelity in the investigated system.

We investigated how the budgeting of shots between detector tomography and state tomography influenced the state reconstruction infidelity and thus identified the optimal distribution of shots. To address the increasing interest in two- and three-qubit gates for superconducting qubits, we demonstrated that beyond-classical readout-error mitigation is suitable for these system sizes.

By performing readout correlation measurements, we showed that readout correlations are significant in the weakly detuned system. We observed no correlated coherent readout errors in this system. By adjusting the readout amplitude of the resonators, we showed that even high noise levels do not induce readout correlations, validating correlation coefficients as an additional useful benchmark for the multiplexed readout of superconducting qubits. Our comprehensive benchmark for quantifying qubit readout precision can be applied to test innovative qubit architectures [40] and readout methods [41]. The optimization of readout parameters should not only focus on improving the performance of individual qubits, but also on keeping readout correlations in multiplexed readout to a minimum. It would be interesting to perform optimization experiments on weakly detuned resonator systems to test this limit of qubit readout optimization. Importantly, we encourage the convergence of reconstruction infidelity to be used in upcoming superconducting qubit experiments that aim to push the boundaries of readout accuracy. Furthermore, it would be interesting to compare the infidelity-thresholding approach to a similar approach that uses QREM that only mitigates correlated classical errors, as these approaches require fewer shots for calibration and could be a more economical solution if coherent errors can be neglected.

## ACKNOWLEDGMENTS

The authors thank Andy Ding for providing the qubit chip used in the study. We also thank W. Oliver and G. Calusine for providing the parametric amplifier. This work was partially financed by the Baden-Württemberg Stiftung gGmbH.

A.D.G. performed the experiments with supervision from H.R., J.L., and A.U. The theory and numerics were implemented by A.S.A. with supervision from M.G. A.D.G. and A.S.A. performed the data analysis. The first manuscript was prepared by A.D.G. and A.S.A. All authors contributed to the finalization of the manuscript.

The authors declare no competing interests.

## DATA AVAILABILITY

The data that support the findings of this article are openly available [42]; embargo periods may apply.

## APPENDIX A: DETAILS OF CORRELATION COEFFICIENTS

References [12,18] introduced a metric for quantifying how the state of one qubit impacts the state of another. This is based on efficiently measuring and reconstructing all possible two-qubit POVMs with detector overlapping tomography [18,43]. A readout correlation measure can be obtained as follows. For each two-qubit POVM, trace out half of the POVM conditioned on a state being measured in the traced-out subsystem. Optimize over all possible states present on the traced-out subsystems and select the largest distance between two traced-out POVMs. These correlation coefficients can be written explicitly as

$$c_{j \rightarrow i} = \sup_{\rho_j, \sigma_j} \|(M_0^{i, \rho_j} - M_0^{i, \sigma_j})\|_2, \quad (\text{A1})$$

where  $M_{x_i}^{i, \sigma_j} = \sum_{x_j} \text{Tr}_j(M_{x_i x_j}^{ij} (\mathbb{1}_i \otimes \sigma_j))$ ,  $x_i$  and  $x_j$  label the outcomes on the two-qubit system, and  $\|A\|_2$  refers to the spectral norm (i.e., the largest singular value of  $A$ ). This corresponds to the correlation coefficients employed in Ref. [12] using the “worst-case” distance, with the difference that in that work, the norm applied is the infinity operator norm,  $\|A\|_\infty = \max_{1 \leq k \leq d} \sum_{l=1}^d |a_{kl}|$ , where  $d = \dim(A)$ . The coefficients quantify the impact of the state of qubit  $j$  on the effective POVM measured on qubit  $i$ . Due to the asymmetry of these coefficients, it is useful to define the symmetric correlation coefficient  $c_{i \leftrightarrow j} = (c_{j \rightarrow i} + c_{i \rightarrow j})/2$ .

If the optimization in Eq. (A1) is restricted to only the computational basis states, i.e., products of eigenstates of the  $\sigma_z$  operator  $\{|0\rangle, |1\rangle\}^{\otimes N}$ , the coefficients are called classical correlation coefficients  $c^C$ . If the optimization is performed over all pure states [12], the resulting coefficients are called quantum correlation coefficients  $c^Q$ . The quantum correlation coefficients cannot be smaller than the classical ones and are only equal if there are no coherent errors present in the readout process (up to statistical fluctuations).

TABLE I. Qubit names and resonator and qubit frequencies of the quantum device. The qubits are named in order from left to right on the physical chip, starting from q0. Qubit and resonator frequencies typically varied up to a few MHz between cooldowns.

Qubit	$\omega_{\text{res}}$ (GHz)	$\omega_{01}$ (GHz)
q0	8.716	4.455
q1	8.831	5.409
q2	8.931	4.127
q3	8.939	4.926

## APPENDIX B: QUBIT AND RESONATOR PARAMETERS

Qubit and resonator parameters varied slightly between different cooldowns. Some representative values are listed in Tables I and II.

## APPENDIX C: READOUT CHAIN

To understand the origin of readout errors, we provide a brief review of the readout chain used for typical dispersive-shift measurements of superconducting qubits. After reaching the transmission line on the chip, the readout signals couple into the resonators that match their frequency. The frequency of the resonator depends on the collapsed qubit state through the dispersive shift [44]. This signal is amplified first on the mixing chamber stage with a Josephson traveling-wave parametric amplifier (JTWPA), driven by a continuous microwave pump tone, which is also channeled into the remaining chain. Some signal amplitude is lost before and after the JTWPA due to non-superconducting cabling and connectors. The resulting signal enters two high-electron-mobility transistor (HEMT) amplifiers, which add noise photons to the signal. External mixers leak sidebands into the signal before it reaches the digitizer and integration module, finally yielding a measured value of 0 or 1.

TABLE II. Typical coherence times and single-shot readout fidelities of the individual qubits. The assignment-fidelity values show the best possible single-qubit readout assignment fidelity. Multiqubit assignment fidelities were observed to be lower in some experiments.

Qubit	Assignment fidelity (%)	$T_1$ (μs)	$T_2$ (μs)
q0	96	118	67
q1	93	22	30
q2	92	27	56
q3	80	23	ca. 15

## APPENDIX D: LEAKAGE AND ASSIGNMENT FIDELITY

In this section, we discuss the comparison of infidelity thresholding and assignment fidelity for varying readout powers in more detail (see Sec. IV B 1).

We treat the transmons as three-level quantum systems (qutrits) and distinguish three regions in the IQ plane, corresponding to the ground and excited states and the third (leaked) state. We prepare an equal number of  $|0\rangle$  and  $|1\rangle$  states and apply readout pulses with growing amplitudes to each qubit. The resulting distributions in the IQ plane are grouped into three clusters (ground, excited, and leaked), and a linear sum assignment is performed to relabel the clusters in maximum accordance with the known (but sometimes incorrect) labels 0 or 1. Note that the lowest readout power at which we can reliably perform the leakage estimate in this way is  $1.5 \times$  the optimal readout power. This is because at lower readout powers, it is

more difficult to reliably estimate the leakage ratio, as the dataset becomes very unbalanced, leading to misinterpretation (overfitting). This is a known problem for  $N$ -clustering algorithms. The results are shown in Fig. 9. Although  $q_1$  is more prone to leakage, its assignment fidelity is less sensitive to this effect at high readout powers. This effect can be attributed to the position of states in the IQ (in phase-quadrature) plane.

## APPENDIX E: TWPA POWER SATURATION

Pumping the parametric amplifier with increasing amplitude saturates the amplification, and the signal quality drops off sharply at around 7.3 dBm pump power (see Fig. 10). This measurement was conducted with a different set of microwave cables; therefore, there is a constant offset in power compared to Fig. 4.

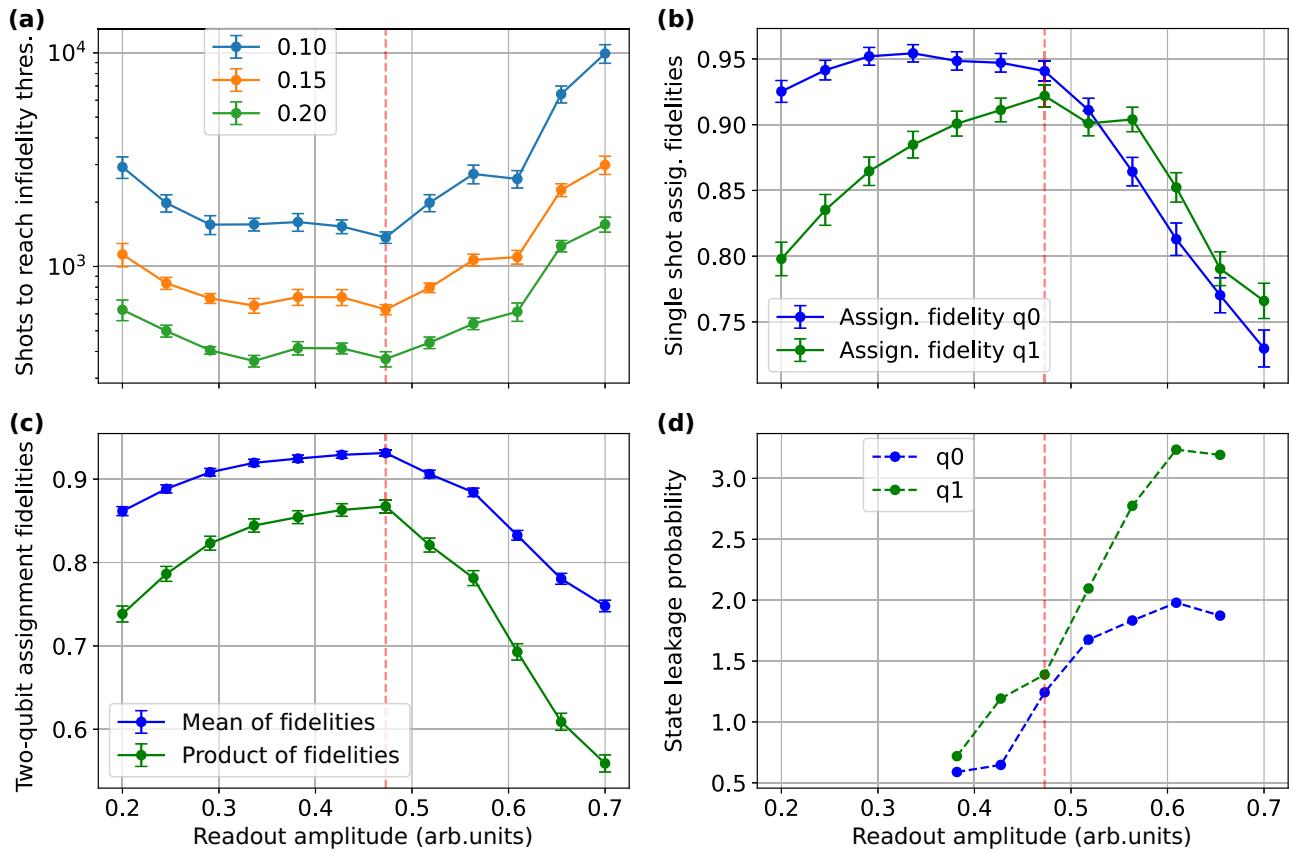


FIG. 9. Comparison of readout-error-mitigated quantum-state-tomography-based readout benchmark with assignment fidelities. The vertical dashed red line indicates the optimal readout amplitude identified in Fig. 3. (a) Number of two-qubit state tomography shots required to reach infidelity thresholds of 0.10, 0.15, and 0.20 for various readout amplitudes. The error bars are bootstrapped standard deviations. For very low readout powers, the distinguishability of the basis states is significantly lowered. For very high readout amplitudes, the measurement process excites higher states in the system, which the state classifier cannot correctly take into account. (b) Single-shot assignment fidelities during the experiments; each point uses 2000 shots. The error bars show one standard deviation. (c) Mean and product of single-shot assignment fidelities from Fig. 3. (d) Probability of measuring a leaked state as a function of readout amplitude.

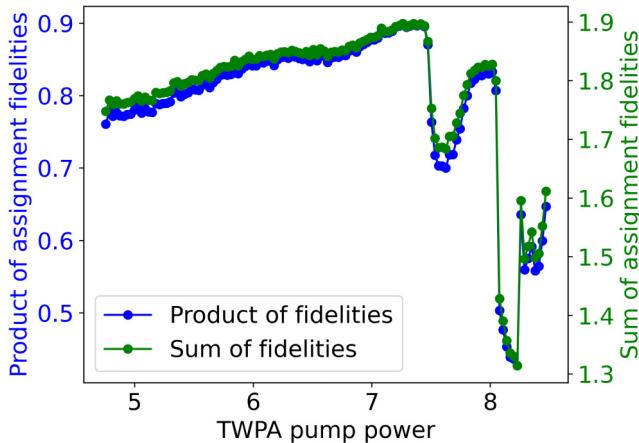


FIG. 10. Sum and product of single-qubit assignment fidelities in System A. A sharp drop in both variables occurs at approximately 7.3 dBm pump power.

## APPENDIX F: THEORETICAL SHOT BUDGETING

In a standard setup of detector tomography followed by state tomography, the distribution of shots should be dictated by the number of parameters to estimate in each part. If there is no knowledge of the quantum states to be estimated, then there are  $4^n - 1$  parameters for each normalized  $n$ -qubit state. An  $m$ -outcome POVM has  $(m - 1)4^n$  parameters that need to be estimated. If the POVM is informationally complete (IC),  $m = 4^n$ . Inserting this into Eq. (7),

$$r_n = \frac{(m - 1)4^n}{(m - 1)4^n + N_s(4^n - 1)} \stackrel{\text{IC}}{=} \frac{4^n}{4^n + N_s}, \quad (\text{F1})$$

where  $N_s$  is the number of states to be estimated with the same reconstructed POVM. Compared with the experiment in Sec. IV C, setting  $N_s = 1$ , we get  $r_1 = 0.80$  and  $r_2 = 0.94$ , significantly higher.

The discrepancy between these two results may arise from the fact that not all the parameters are free and independent in a physical setting. For example, if the reconstructed states are close to pure, there are effectively  $2(2^n - 1)$  parameters in the state (since most parameters are close to zero). In addition, the POVM implemented in the experiment is not an IC POVM but a set of projective measurements on the eigenstates of the Pauli operators, each individually normalized.

If the errors at readout are close to classical (i.e., only redistributing the eigenvalues of the POVM elements), the number of parameters in the POVM is  $(4^n - 1)2^n$ . If we

assume close-to-classical errors, the ratio is

$$r_n \stackrel{\text{mixed}}{=} \frac{2^n}{2^n + N_s},$$

$$\stackrel{\text{pure}}{=} \frac{4^n + 2^n}{4^n + 2^n + 2N_s} \quad (\text{F2})$$

which, compared to the experiment ( $N_s = 1$ ), gives us  $r_1 = (0.67, 0.75)$  and  $r_2 = (0.80, 0.91)$  for (mixed, pure) target states.

## APPENDIX G: MEASUREMENT DRIFT

As both QDT and QST scale exponentially in run time with the number of qubits, the experiment has to run self-consistently over a prolonged period. One limitation placed on the length of experiments is the temporal drift of the operational parameters. To quantitatively estimate the timescales and effects of these drifts on our POVMs, we repeatedly perform two-qubit detector tomography over 11 h while tracking the temperature and humidity of the laboratory. Specifically, we reconstruct the measurement operators corresponding to ZZ- and XX-basis-multiplexed readout of System A and plot their coherent errors in Fig. 11. We define coherent errors in Appendix I. While the coherent errors in the ZZ basis are small and can be explained by statistical errors, the XX basis contains significant coherent errors. This can be explained by the additional rotation gates used, which are not present in ZZ-basis readout. Under stable laboratory conditions, errors are observed to be temporally constant for both. Introducing rapid temperature and humidity changes by turning the laboratory air conditioning off, the amount of coherence in the reconstructed measurement operator of the XX basis drifts significantly. As the ZZ-basis coherences are stable, we identify the likely source of coherence drift as originating from the manipulation of the qubit used for XX-basis measurements.

## APPENDIX H: THREE-QUBIT QUANTUM-STATE TOMOGRAPHY WITH READOUT-ERROR MITIGATION

The three-qubit density matrix obtained with readout-error mitigation has around an order-of-magnitude smaller error in the worst entry of the unmitigated density matrix (see Fig. 12). The nonvanishing matrix entries have the largest errors both with and without readout-error mitigation. The infidelity of reconstruction was calculated from these density matrices.

## APPENDIX I: COHERENT ERRORS

One way to quantify the degree of nonclassicality of readout errors is to measure the distance between the reconstructed POVM and the nearest POVM subject to

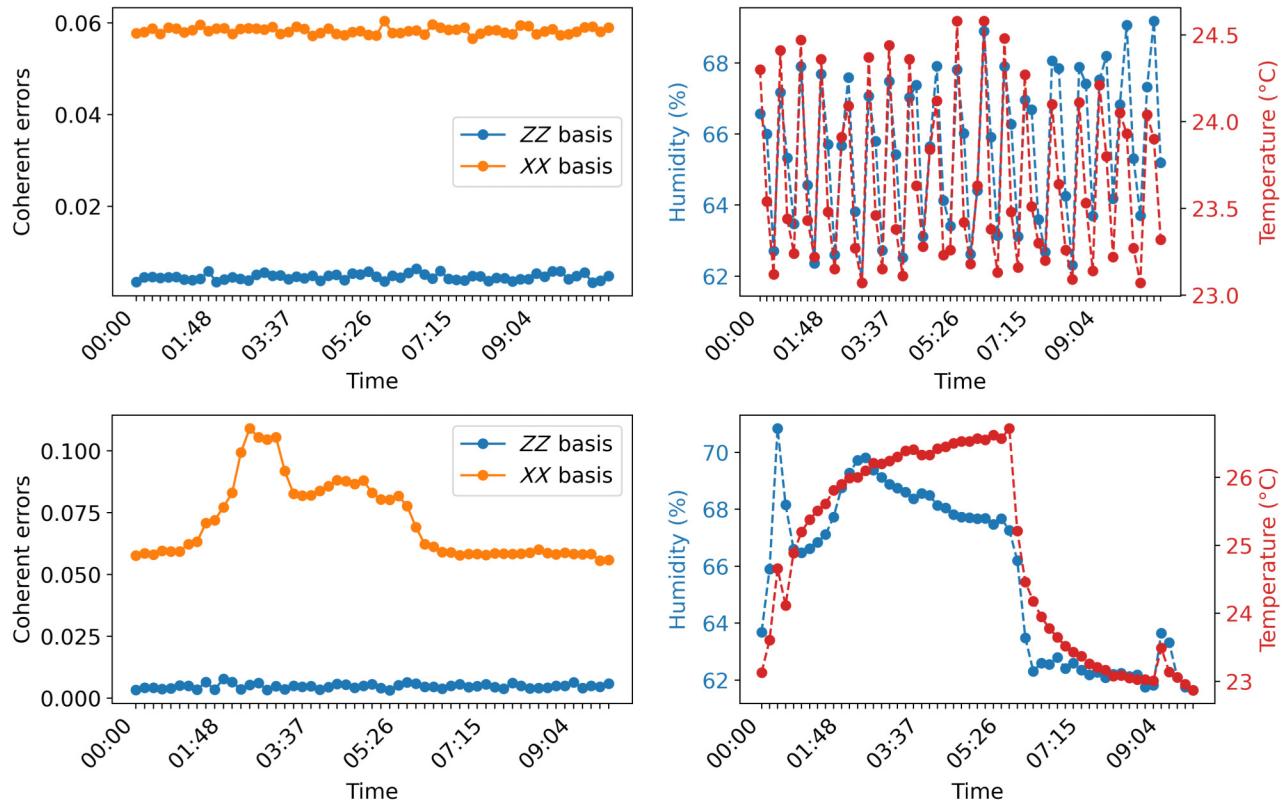


FIG. 11. Stability of coherent readout errors with varied laboratory temperature and humidity. Top left: coherent errors [defined in Eq. (14)] in the reconstructed measurement operators for ZZ- and XX-basis readout under stable laboratory conditions. Top right: temperature and humidity values during the corresponding measurement. The rapid oscillations come from the feedback mechanism of the air conditioning. Bottom left: coherent errors in the reconstructed measurement operators for ZZ- and XX-basis readout under unstable laboratory conditions induced by turning the air conditioning off, then on. Bottom right: temperature and humidity values during the corresponding measurements.

only classical errors. Classical errors are captured by a confusion matrix [17], which involves a statistical redistribution of the probabilities of measurement outcomes in the computational basis. Specifically, it can be described by a matrix  $C$ , which gives the relation between an ideal measurement probability  $p_i$  of outcome  $i$  and the measured frequencies  $\tilde{p}_i = n_i/N$ ,

$$\tilde{p}_i = C_{ij} p_j. \quad (I1)$$

Equation (I1) can be rephrased in terms of a noisy POVM  $\tilde{\mathbf{M}}$  and an ideal projective measurement  $\mathbf{P}$

$$\tilde{M}_i = C_{ij} P_j. \quad (I2)$$

If coherent errors are present, then Eq. (I2) gets an additional term  $\Delta_i$ ,

$$\tilde{M}_i = C_{ij} P_j + \Delta_i. \quad (I3)$$

Since every projective measurement operator  $P_i$  contains only diagonal entries, we can quantify coherent errors by

computing the distance between the reconstructed POVM  $\tilde{\mathbf{M}}$  and the same POVM with only diagonal entries [17],

$$\epsilon_{coh} = \text{dist}(\tilde{\mathbf{M}}, \mathcal{E}_{dep}(\tilde{\mathbf{M}})), \quad (I4)$$

where  $\text{dist}(\cdot, \cdot)$  is a suitable POVM distance and  $\mathcal{E}_{dep}(\tilde{M}_i) = \text{diag}(\tilde{M}_i)$  is the maximally dephasing channel. For coherent errors, we use the average-case distance for POVMs, which is defined as [45]

$$d_{AV}(\mathbf{M}, \mathbf{N}) = \frac{1}{2d} \sum_{i=1}^n \sqrt{\|M_i - N_i\|_{HS}^2 + \text{Tr}(M_i - N_i)^2}, \quad (I5)$$

where  $n$  is the number of outcomes for the POVMs  $\mathbf{M}$  and  $\mathbf{N}$ ,  $d$  is the dimension of the measurement operators, and  $\|A\|_{HS} = \sqrt{\text{Tr}(A^2)}$ .

## APPENDIX J: EFFECT OF DEPOLARIZING NOISE ON CORRELATION COEFFICIENTS

Many common sources of experimental errors, such as reduced computational-state distinguishability, can be

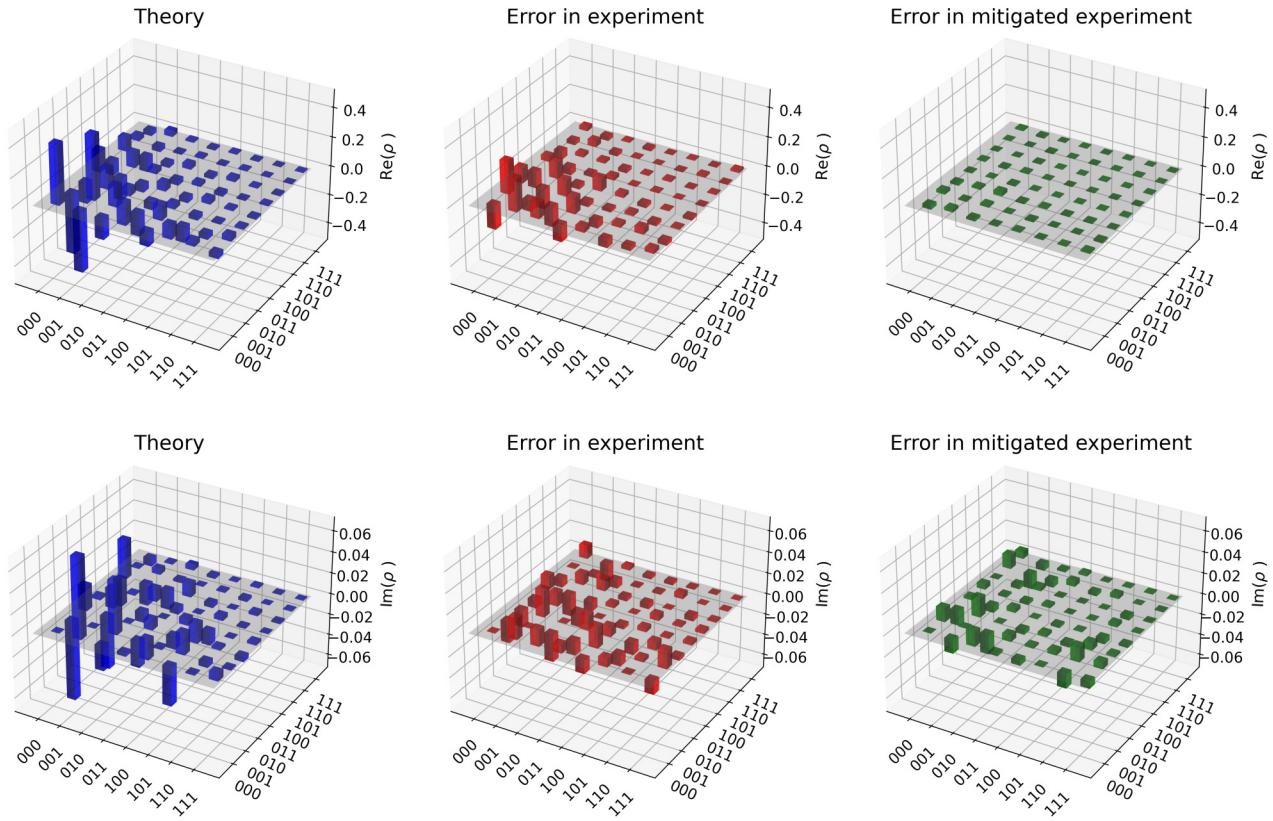


FIG. 12. Comparison of theoretical density matrix of a three-qubit state and its reconstruction with and without readout-error mitigation. The top row shows the real part of the complex eight-dimensional matrices, and the bottom row shows the imaginary part. The quantum states reconstructed with and without readout-error mitigation have infidelities of 0.030 and 0.447, respectively. The different scales for the real and imaginary parts are due to the chosen quantum state. The errors in the imaginary and real parts of the error-mitigated experiments are of the same order of magnitude.

modeled as a depolarizing channel on individual qubits. The single-qubit depolarizing channel can be written as [46]

$$\mathcal{E}_{\text{depol}}(M_i) = \left(1 - \frac{3k}{4}\right) M_i + \frac{k}{4} (X M_i X + Y M_i Y + Z M_i Z), \quad (\text{J1})$$

where  $k$  denotes the strength of depolarization and  $X$ ,  $Y$ , and  $Z$  are the Pauli matrices.

In the experiment discussed in Sec. V B, the distinguishability of the two-qubit computational basis affects each qubit independently. We therefore expect the noise to be separable:

$$\mathcal{E}_{\text{depol}}^2(M_i) = (\mathcal{E}_{\text{depol}} \otimes \mathcal{E}_{\text{depol}})(M_i). \quad (\text{J2})$$

Applying the separable depolarizing channel to the two-qubit measurement operators in the definition of the correlation coefficients, we find that

$$\begin{aligned} M_{x_i, \text{depol}}^{i, \sigma_j} &= \sum_{x_j} \text{Tr}(\mathcal{E}_{\text{depol}}^2(M_{x_i x_j}^{ij})) (1 \otimes \sigma_j) \\ &= \sum_{x_j} \text{Tr}(M_{x_i x_j}^{ij} (\mathbb{1} \otimes \mathcal{E}_{\text{depol}}(\sigma_j))) \\ &= M_{x_i}^{i, \mathcal{E}_{\text{depol}}(\sigma_j)}. \end{aligned} \quad (\text{J3})$$

The correlation coefficients in Eq. (A1) can therefore be written

$$c_{j \rightarrow i} = \sup_{\rho_j, \sigma_j} \|(M_0^{i, \mathcal{E}_{\text{depol}}(\rho_j)} - M_0^{i, \mathcal{E}_{\text{depol}}(\sigma_j)})\|_2. \quad (\text{J4})$$

The depolarizing channel acts on the optimization states  $\rho_j$  and  $\sigma_j$  and effectively reduces the search space for the optimization. Consequently, the correlation coefficients are nonincreasing under single-qubit depolarizing noise.

- [1] R. Acharya *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, *Nature* **614**, 676 (2023).
- [2] R. Acharya, *et al.*, Quantum error correction below the surface code threshold, *Nature* **638**, 920 (2024).
- [3] Y. Li, and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, *Phys. Rev. X* **7**, 024510 (2017).
- [4] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [5] Y. Quek, D. Stilck França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, *Nat. Phys.* **20**, 1648 (2024).
- [6] Z. Li, P. Liu, P. Zhao, Z. Mi, H. Xu, X. Liang, T. Su, W. Sun, G. Xue, J.-N. Zhang, W. Liu, Y. Jin, and H. Yu, Error per single-qubit gate below  $10^{-4}$  in a superconducting qubit, *Npj Quantum Inf.* **9**, 111 (2023).
- [7] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. L. Campbell, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Tunable coupling scheme for implementing high-fidelity two-qubit gates, *Phys. Rev. Appl.* **10**, 054062 (2018).
- [8] M. Werninghaus, D. J. Egger, F. Roy, S. Machnes, F. K. Wilhelm, and S. Filipp, Leakage reduction in fast superconducting qubit gates via optimal control, *Npj Quantum Inf.* **7**, 14 (2021).
- [9] E. Hyppä, A. Vepsäläinen, M. Papič, C. F. Chan, S. Inel, A. Landra, W. Liu, J. Luus, F. Marxer, C. Ockeloen-Korppi, S. Orbell, B. Tarasinski, and J. Heinsoo, Reducing leakage of single-qubit gates for superconducting quantum processors using analytical control pulse envelopes, *PRX Quantum* **5**, 030353 (2024).
- [10] A. Marton, and J. K. Asboth, Coherent errors and readout errors in the surface code, *Quantum* **7**, 1116 (2023).
- [11] C. D. Wilen, S. Abdullah, N. A. Kurinsky, C. Stanford, L. Cardani, G. D'Imperio, C. Tomei, L. Faoro, L. B. Ioffe, C. H. Liu, A. Opremcak, B. G. Christensen, J. L. DuBois, and R. McDermott, Correlated charge noise and relaxation errors in superconducting qubits, *Nature* **594**, 369 (2021).
- [12] J. Tuziemski, F. B. Maciejewski, J. Majsak, O. Słowiak, M. Kotowski, K. Kowalczyk-Murynka, P. Podziemski, and M. Oszmaniec, Efficient reconstruction, benchmarking and validation of cross-talk models in readout noise in near-term quantum devices, [arXiv:2311.10661](https://arxiv.org/abs/2311.10661).
- [13] T. Walter, P. Kurpiers, S. Gasparinetti, P. Magnard, A. Potočnik, Y. Salathé, M. Pechal, M. Mondal, M. Oppliger, C. Eichler, and A. Wallraff, Rapid high-fidelity single-shot dispersive readout of superconducting qubits, *Phys. Rev. Appl.* **7**, 054020 (2017).
- [14] J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočnik, A. Wallraff, and C. Eichler, Rapid high-fidelity multiplexed readout of superconducting qubits, *Phys. Rev. Appl.* **10**, 034040 (2018).
- [15] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, Tomography of quantum detectors, *Nat. Phys.* **5**, 27 (2008).
- [16] Y. Chen, M. Farahzad, S. Yoo, and T.-C. Wei, Detector tomography on IBM quantum computers and mitigation of an imperfect measurement, *Phys. Rev. A* **100**, 052315 (2019).
- [17] F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography, *Quantum* **4**, 257 (2020).
- [18] F. B. Maciejewski, F. Baccari, Z. Zimborás, and M. Oszmaniec, Modeling and mitigation of cross-talk effects in readout noise with applications to the quantum approximate optimization algorithm, *Quantum* **5**, 464 (2021).
- [19] S. Kono, J. Pan, M. Chegnizadeh, X. Wang, A. Youssefi, M. Scigliuzzo, and T. J. Kippenberg, Mechanically induced correlated errors on superconducting qubits with relaxation times exceeding 0.4 ms, *Nat. Commun.* **15**, 3950 (2024).
- [20] X. Li, J. Wang, Y.-Y. Jiang, G.-M. Xue, X. Cai, J. Zhou, M. Gong, Z.-F. Liu, S.-Y. Zheng, D.-K. Ma, M. Chen, W.-J. Sun, S. Yang, F. Yan, Y.-R. Jin, S. P. Zhao, X.-F. Ding, and H.-F. Yu, Cosmic-ray-induced correlated errors in superconducting qubit array, *Nat. Commun.* **16**, 4677 (2025).
- [21] A. S. Aasen, A. Di Giovanni, H. Rotzinger, A. V. Ustinov, and M. Gärttner, Mitigation of correlated readout errors without randomized measurements, *Phys. Rev. A* **112**, 022601 (2025).
- [22] P. C. de Groot, A. F. van Loo, J. Lisenfeld, R. N. Schouten, A. Lupaşcu, C. J. P. M. Harmans, and J. E. Mooij, Low-crosstalk bifurcation detectors for coupled flux qubits, *Appl. Phys. Lett.* **96**, 123508 (2010).
- [23] A. S. Aasen, A. Di Giovanni, H. Rotzinger, A. V. Ustinov, and M. Gärttner, Readout error mitigated quantum state tomography tested on superconducting qubits, *Commun. Phys.* **7**, 301 (2024).
- [24] M. R. Geller, Conditionally rigorous mitigation of multiqubit measurement errors, *Phys. Rev. Lett.* **127**, 090502 (2021).
- [25] M. R. Geller, and M. Sun, Toward efficient correction of multiqubit measurement errors: Pair correlation method, *Quantum Sci. Technol.* **6**, 025009 (2021).
- [26] M. Jerger, S. Poletto, P. Macha, U. Hübner, E. Il'ichev, and A. V. Ustinov, Frequency division multiplexing readout and simultaneous manipulation of an array of flux qubits, *Appl. Phys. Lett.* **101**, 042604 (2012).
- [27] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovsky, X. Zhang, W. D. Oliver, and I. Siddiqi, A near-quantum-limited Josephson traveling-wave parametric amplifier, *Science* **350**, 307 (2015).
- [28] Y. Kim, A. Morvan, L. B. Nguyen, R. K. Naik, C. Jünger, L. Chen, J. M. Kreikebaum, D. I. Santiago, and I. Siddiqi, High-fidelity three-qubit iToffoli gate for fixed-frequency superconducting qubits, *Nat. Phys.* **18**, 783 (2022).
- [29] T. Itoko, M. Malekakhlagh, N. Kanazawa, and M. Takita, Three-qubit parity gate via simultaneous cross-resonance drives, *Phys. Rev. Appl.* **21**, 034018 (2024).
- [30] X.-L. Li, Z. Tao, K. Yi, K. Luo, L. Zhang, Y. Zhou, S. Liu, T. Yan, Y. Chen, and D. Yu, Hardware-efficient and fast three-qubit gate in superconducting quantum circuits, *Front. Phys.* **19**, 51205 (2024).
- [31] T. Thorbeck, A. Eddins, I. Lauer, D. T. McClure, and M. Carroll, Two-level-system dynamics in a superconducting

- qubit due to background ionizing radiation, *PRX Quantum* **4**, 020356 (2023).
- [32] J. Fiurášek, Maximum-likelihood estimation of quantum measurement, *Phys. Rev. A* **64**, 024102 (2001).
- [33] A. I. Lvovsky, Iterative maximum-likelihood reconstruction in quantum homodyne tomography, *J. Opt. B: Quantum Semiclassical Opt.* **6**, S556 (2004).
- [34] G. I. Struchalin, I. A. Pogorelov, S. S. Straupe, K. S. Kravtsov, I. V. Radchenko, and S. P. Kulik, Experimental adaptive quantum tomography of two-qubit states, *Phys. Rev. A* **93**, 012103 (2016).
- [35] B. Lienhard, A. Vepsäläinen, L. C. Govia, C. R. Hoffer, J. Y. Qiu, D. Ristè, M. Ware, D. Kim, R. Winik, A. Melville, B. Niedzielski, J. Yoder, G. J. Ribeill, T. A. Ohki, H. K. Krovi, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Deep-neural-network discrimination of multiplexed superconducting-qubit states, *Phys. Rev. Appl.* **17**, 014024 (2022).
- [36] R. Barends, J. Kelly, A. Megrant, D. Sank, E. Jeffrey, Y. Chen, Y. Yin, B. Chiaro, J. Mutus, C. Neill, P. O’Malley, P. Roushan, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, Coherent Josephson qubit suitable for scalable quantum integrated circuits, *Phys. Rev. Lett.* **111**, 080502 (2013).
- [37] F. Mezzadri, How to generate random matrices from the classical compact groups, *arXiv:math-ph/0609050*.
- [38] K. Życzkowski, K. A. Penson, I. Nechita, and B. Collins, Generating random density matrices, *J Math. Phys.* **52**, 062201 (2011).
- [39] C. W. Warren, J. Fernández-Pendás, S. Ahmed, T. Abad, A. Bengtsson, J. Biznárová, K. Debnath, X. Gu, C. Križan, A. Osman, A. Fadavi Roudsari, P. Delsing, G. Johansson, A. Frisk Kockum, G. Tancredi, and J. Bylander, Extensive characterization and implementation of a family of three-qubit gates at the coherence limit, *Npj Quantum Inf.* **9**, 44 (2023).
- [40] A. Kreuzer, T. Krumrey, H. Tohamy, A. Ionita, H. Rotzinger, and A. V. Ustinov, Stacked Josephson junctions for quantum circuit applications, *arXiv:2503.11437*.
- [41] A. M. Gunyhó, S. Kundu, J. Ma, W. Liu, S. Niemelä, G. Catto, V. Vadimov, V. Vesterinen, P. Singh, Q. Chen, and M. Möttönen, Single-shot readout of a superconducting qubit using a thermal detector, *Nat. Electron.* **7**, 288 (2024).
- [42] A. Aasen, “CREMST.”, 2025, <https://github.com/AdrianAasen/CREMST>
- [43] J. Cotler, and F. Wilczek, Quantum overlapping tomography, *Phys. Rev. Lett.* **124**, 100401 (2020).
- [44] G. Zhu, D. G. Ferguson, V. E. Manucharyan, and J. Koch, Circuit QED with fluxonium qubits: Theory of the dispersive regime, *Phys. Rev. B* **87**, 024510 (2013).
- [45] F. B. Maciejewski, Z. Puchała, and M. Oszmaniec, Operational quantum average-case distances, *Quantum* **7**, 1106 (2023).
- [46] M. A. Nielsen, and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2012).