

# Multiplexed qubit readout quality metric beyond assignment fidelity

Andras Di Giovanni,<sup>1</sup> Adrian Skasberg Aasen,<sup>2,3</sup> Jürgen Lisenfeld,<sup>1</sup>  
Martin Gärttner,<sup>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*

(Dated: June 2024)

The accurate measurement of quantum two-level objects (qubits) is crucial for developing quantum computing hardware. 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 multi-qubit 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 an alternative and more comprehensive benchmark for single- and multi-qubit readout optimization. We find the most effective allocation of a fixed shot budget between detector tomography and state reconstruction in single- and two-qubit experiments. To address the growing interest in three-qubit gates, 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 neither quantum nor classical qubit readout correlations are induced even by very high levels of readout noise. Consequently, correlation coefficients can serve as a valuable tool in qubit readout optimization.

## I. INTRODUCTION

State-of-the-art quantum computing experiments have started pushing the boundary to fault tolerance [1, 2]. Although gate error mitigation 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 pose an important obstacle to general purpose quantum computing as well [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 seeks to mitigate readout errors by reconstructing the experimental measurement operator, instead of assuming that readout is accurately modeled by a  $\hat{z}$  projective computational basis measurements.

We apply a comprehensive readout error mitigation protocol, introduced in Ref. [19], to a multi-qubit quantum state tomography experiment. This protocol goes beyond mitigating classical readout errors 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 [20]. This architectural choice simplifies wiring, but can be prone to significant readout crosstalk, and is hence an ideal test-bed for multi-qubit readout error mitigation. By employing a travelling wave parametric amplifier (TWPA) [21], we can perform single-shot readout of the qubits. number of shots required to reach a given state reconstruction infidelity as a proxy for quantifying how good a given parameter setting is for extracting information from the system. This allows one to test different readout parameters. This data can then be compared to the faster methods of measuring assignment fidelities 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 non-classical readout noise, which opens the way to reduce readout errors in complete benchmarking of two- and three-qubit gates.

An advantage of detector tomography is that it explicitly reconstructs the POVM corresponding to the multi-qubit readout process. This enables extraction of multi-qubit readout correlation coefficients of the system, regardless of their origin, which could be off-resonant drive of resonator, mechanical vibrations [22], or ionizing radiation [23]. 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 for readout quality [24].

A measure for readout correlations should be independent of the state distinguishability in the readout plane. By inducing an additional readout noise, we demonstrate that readout correlation coefficients satisfy this property.

## II. PRELIMINARIES

### A. Generalized measurements

Generalized quantum measurements are a set of operators with 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  $M_i$ . These yield probabilities of outcome  $i$ ,  $p_i$ , occurring during measurement on a quantum state  $\rho$ , through Born's rule

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

This generalized approach to quantum measurements extends beyond projective measurements, enabling the capture of noise within the qubit manifold.

### B. Detector tomography

Detector tomography is the reconstruction of the POVM implemented by a measurement device [15]. The goal of detector tomography is to reconstruct the POVM  $\mathbf{M} = \{M_i\}$  that satisfies 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 reconstruct any given POVM, it is necessary to have an informationally complete set of calibration states, e.g. all possible combinations of the eigenstates of the Pauli operators  $\sigma_x$ ,  $\sigma_y$  and  $\sigma_z$ . With the full set of measurement probabilities in eq. (3) one can estimate the most likely set of operators through a maximum-likelihood estimator, e.g. as outlined in [25].

### 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. [19], and is based on integrating the estimated POVMs directly into a density matrix estimator. In particular, it is based on an adapted likelihood function,

$$\mathcal{L}_{\mathbf{M}^{\text{estm}}}(\rho) \propto \prod_i \text{Tr}(\rho M_i^{\text{estm}})^{n_i} = (\prod_i (p_i)^{\hat{p}_i})^N, \quad (4)$$

enabling its application as either a maximum-likelihood estimator [26] or a Bayesian mean estimator [27].

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

### D. Extracting readout correlations from detector tomography

A useful concept for extracting correlations found in POVMs was introduced in Refs. [12, 18]. It is based on efficiently measuring and reconstructing all possible two-qubit POVMs with detector overlapping tomography [18, 28]. 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 being present on the traced out subsystems and select the largest distance between two traced down POVMs. This gives us the correlation coefficients

$$c_{j \rightarrow i} = \sup_{\rho_j, \sigma_j} \text{dist}(\mathbf{M}^{i, \rho_j}, \mathbf{M}^{i, \sigma_j}), \quad (5)$$

where  $M_{x_i}^{i, \sigma_j} = \sum_{x_j} \text{Tr}_j(M_{x_i x_j}^{ij} (\mathbb{1}_i \otimes \sigma_j))$ ,  $x_i, x_j$  labeling the outcome on the two qubit system and  $\text{dist}(\cdot, \cdot)$  refers to an appropriate distance measure for POVMs. This study utilizes the "worst-case" distance; for further details, refer to Ref. [12]. 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. (5) 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 coefficient are called classical correlation coefficient  $c^C$ . If optimization goes over all pure states [12], it is called the quantum correlation coefficient  $c^Q$ . 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).

### E. Coherent errors

A method to quantify the degree of non-classicality of readout errors is to measure the distance between the reconstructed POVM and the nearest POVM subject to 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. A suitable measure is therefore to compute the distance between the reconstructed POVM and the same POVM with only the diagonal entries [17],

$$\epsilon_{\text{coh}} = \text{dist}(\mathcal{M}, \mathcal{E}_{\text{dep}}(\mathcal{M})), \quad (6)$$

where  $\text{dist}(\cdot, \cdot)$  is a suitable POVM distance, and  $\mathcal{E}_{\text{dep}}(\cdot)$  is the maximally dephasing channel. For coherent errors, we use the average-case distance [29].

### F. Infidelity

We use quantum infidelity,  $I(\rho, \sigma)$ , as a figure of merit for density matrix reconstruction, defined by

$$I(\rho, \sigma) = 1 - F(\rho, \sigma) = 1 - \left[ \text{Tr} \left( \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right) \right]^2, \quad (7)$$

where  $F(\rho, \sigma)$  is the quantum fidelity between a sample state  $\rho$  and a target state  $\sigma$ . A simplification is possible if the target state is pure, yielding

$$I(\rho, \sigma) = 1 - \text{Tr}(\rho\sigma). \quad (8)$$

### G. Multiplexed qubit readout

Multiplexed readout [20] 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 are also amplified and mixed simultaneously after reading out the qubit state. This architectural choice needs less 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 are measured with crosstalk. A readout pulse synthesised for one resonator will also drive a neighbouring 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 [30].

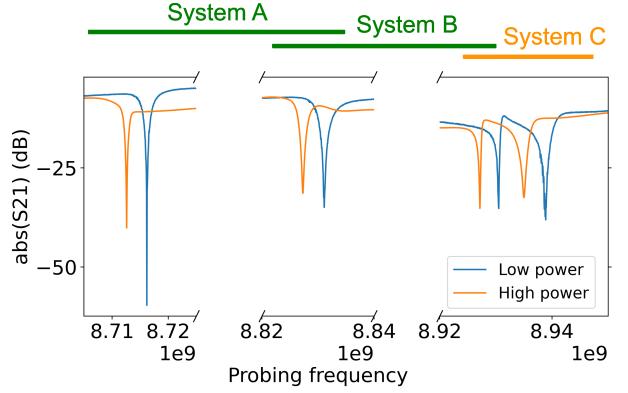


FIG. 1. Resonator spectroscopy of the 4-qubit system. At sufficiently low readout powers, each qubit couples coherently to its resonator, enabling dispersive readout. Systems A, B and C are two-qubit systems of neighbouring resonator-qubit systems. An important feature is the closeness of resonators in System C, symbolized by the orange colour. From left to right, the corresponding qubits are  $q_0, q_1, q_2$  and  $q_3$ .

## III. EXPERIMENTAL REALIZATION

### A. Experimental setup

We investigate a device with four transmon qubits [31], each coupled individually to a resonator for readout. The device is measured below 15 mK in a dry dilution cryostat. Frequencies of the qubits and resonators, as well as typical coherence times can be found in App. A. We define System A as the two-qubit subsystem  $q_0$  and  $q_1$ , System B as subsystem  $q_1$  and  $q_2$ , and finally System C as subsystem  $q_2$  and  $q_3$ , as shown in Fig. 1. A description of the readout chain in the context of readout errors is found in App. B.

### B. 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 frequency and amplitude of the continuous pump tone of the parametric amplifier. This results in

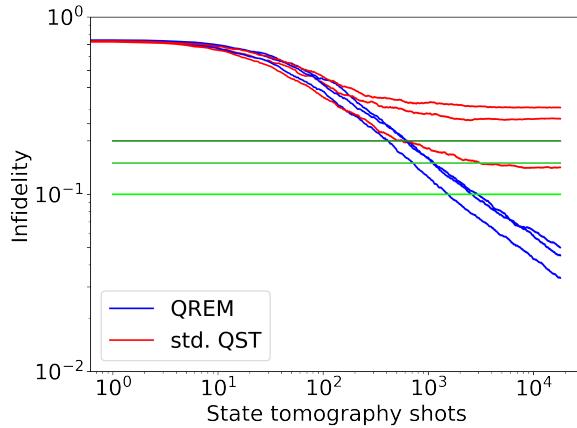


FIG. 2. Mean infidelity curves crossing different threshold values (corresponding to infidelity values of 0.1, 0.15 and 0.2). Six infidelity curves are plotted, each one with (blue) and without (red) readout error mitigation for three different readout parameters. The curves are created using a Bayesian mean estimator of the reconstructed quantum states. Only readout error mitigated infidelities reliably cross the given thresholds.

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 multi-qubit readout error mitigation to the system described in the previous section. We use the experimental protocol depicted in Fig. 1 of [19]. We study the rate at which quantum information can be extracted from the qubit, and show that it is a more comprehensive approach compared to assignment fidelity. In particular, we measure the number of individual 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 briefly motivate the need for readout error mitigation for extracting the crossing of infidelity thresholds in our two-qubit experiment. Fig. 2 shows six mean infidelity curves, 3 readout error mitigated (QREM, blue), and 3 unmitigated (std. QST, red), as a function of shots used for state tomography. Each curve is averaged over 16 products of single-qubit Haar random states [32]. The

crossing points of these curves with a given constant infidelity (0.2, 0.15 and 0.1) are extracted using a Bayesian mean estimator [19]. This allows one to identify the exact shot number at which the infidelity curve crossed 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), corresponding to the thresholds used in Fig. 4. 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 exponent domain (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 2 qubits, after which maximum likelihood estimators gain an additional advantage in speed.

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

We perform readout error mitigated two-qubit quantum state tomography experiments on System A (see Fig. 1) to investigate how the rate of quantum information extraction from the system is influenced by readout noise. Although assignment fidelity is often used to quantify readout quality in superconducting qubits, it is not a complete benchmark of readout fidelity, as only computational basis states are used to determine its value.

###### 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. In Fig. 3 (a) the required number of shots to reach the previously stated infidelity thresholds is plotted. The plot shows one optimum for information extraction rates at 0.48 (a.u.) readout amplitude. This is in agreement with both the product and sum of single-qubit assignment fidelities, see Fig. 3 (b). The optimum as obtained in Fig. 3 (a) is statistically significant, whereas the assignment fidelities measured 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 a.u. readout power. The general features of the readout parameter landscape are explained by low state distinguishability at low readout powers and state leakage at higher readout power, also observed e.g. in Ref.

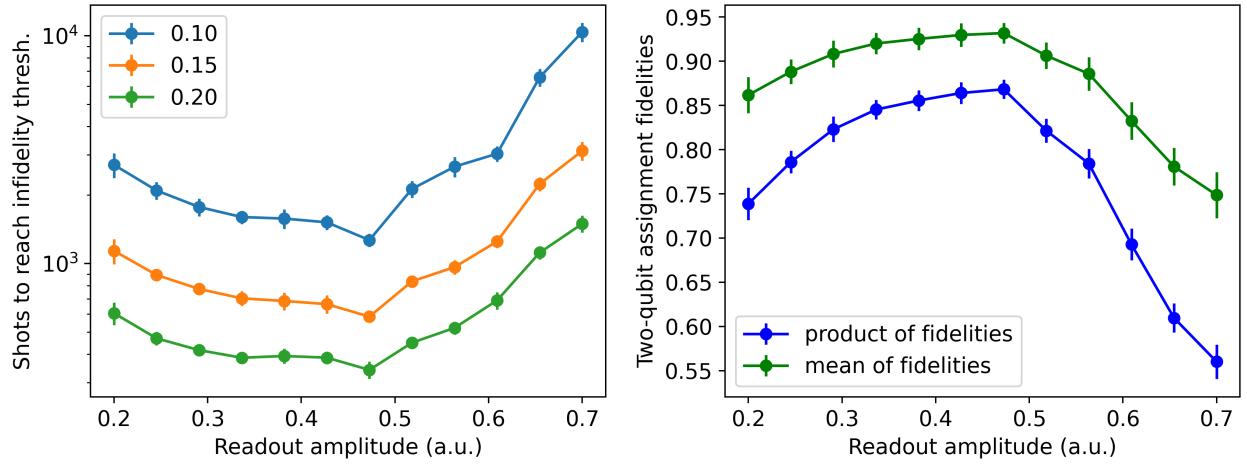


FIG. 3. Comparison of quantum state tomography based readout benchmark with assignment fidelities for varying readout amplitudes. **Left:** The number of two-qubit state tomography shots required to reach infidelity thresholds of 0.1, 0.15 and 0.2 for various readout amplitudes. The error bars are bootstrapped 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 take into account correctly. **Right:** Mean and product of single shot assignment fidelities. The two curves take a very similar shape.

[13]. Both 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 App. C. In conclusion, in the performed experiment, the sum and product of assignment fidelities served as a reliable indicator for how fast information can be extracted from separable states of the two-qubit manifold. The methods agree in the existence and location of a single optimum for the best two-qubit multiplexed readout.

## 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. [21]. Therefore 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 6.3 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 App. D). 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.

## C. Distribution of given shot budget

Experiments often have to be limited in runtime. This can be due to experimental drift or simply resource sharing. This raises the question 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

$$r = \frac{\text{QDT shots}}{\text{QDT shots} + \text{QST shots}} \quad (9)$$

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, 10 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 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 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 es-

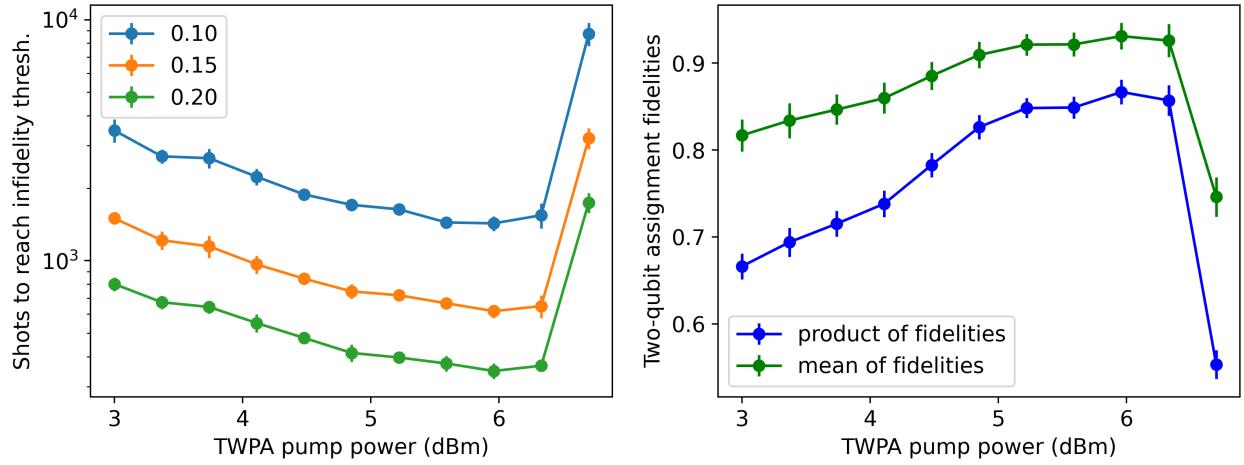


FIG. 4. Comparison of quantum state tomography based readout benchmark with assignment fidelities for varying amplification. **Left:** The number of shots in state tomography needed to reach 0.1, 0.15 and 0.2 reconstruction infidelity. The error bars show bootstrapped standard deviations. **Right:** Sum and product of assignment fidelities for the performed experiments. State separation increases with pump power up to approx. 6.5 dBm, after which it sharply drops. The error bars show propagated standard deviations from the assignment fidelity. Each point uses 1000 shots.

timate the optimal value of  $r$  as defined in eq. (9) 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. For theoretical considerations, see App. E.

#### D. Three-qubit state tomography

Three-qubit quantum state tomography was performed with and without readout error mitigation to test the generalization and scalability of the protocol introduced in Ref. [19] to more qubits. The experiment was repeated on the same prepared state with increasing 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 12000 shots per basis measurement likely attributed to measurement drift, which we investigate further in App. F. A comparison of the theoretically expected density matrix to the reconstructed one with and without readout error mitigation is found in App. G. 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. [19] is scalable to 3 superconducting qubits and is capable to significantly reduce readout errors. This opens a way to benchmark and understand three-qubit gates even if coherent readout errors

are present in the system. This is of particular interest, as three-qubit gates have recently seen a renewed interest in the community [33–35] due to their potentially higher fidelity, shorter duration and ability to replace multiple two-qubit gates.

## V. TWO-QUBIT CORRELATIONS FROM DETECTOR TOMOGRAPHY

In this section we investigate readout correlations with detector tomography. We analyze the dependence of these correlation coefficients on readout noise strength.

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

We perform detector tomography on Systems A and C (see Sec. III A) respectively to extract quantum and classical correlation coefficients as a function of detector tomography shots. The results are plotted in Fig. 7. At low shot numbers, correlations are dominated by sample fluctuations in the reconstructed POVM, which decrease with increasing shots. Therefore, these values are only upper bounds on correlation until they saturate with increasing detector shots. Such saturation is not observed in System A, but is strongly visible in System C. This observation is attributed to the frequency spacing of resonators in the two systems:  $\delta_B = 100$  MHz vs  $\delta_C = 8$  MHz.

Thus detector tomography can be used as a precise tool going beyond measurement basis correlations to test whether resonators are sufficiently far away from each

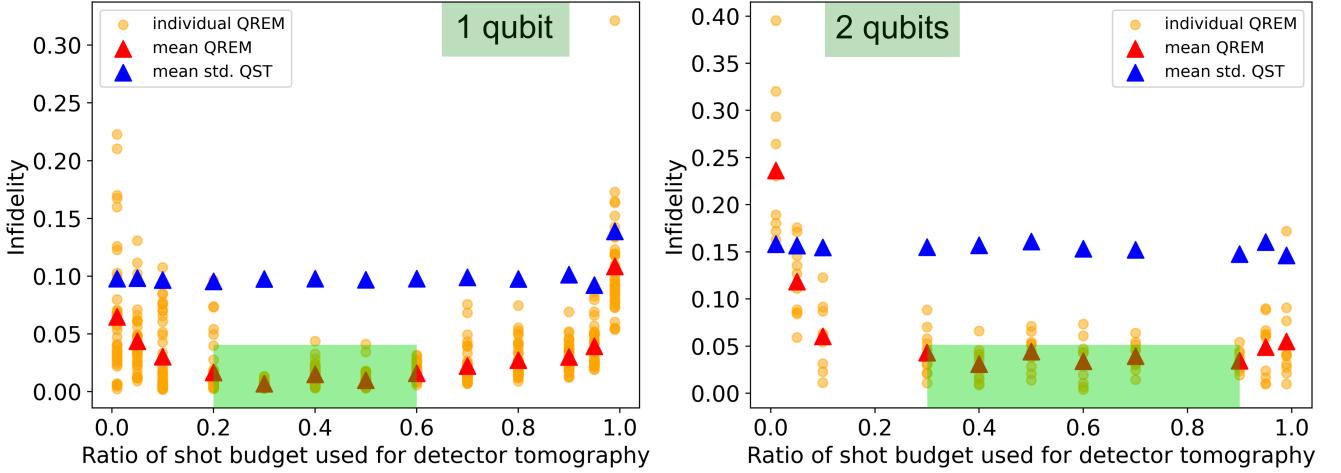


FIG. 5. Reachable infidelity for different distributions of a given shot budget in readout error mitigated state reconstruction. The plotted infidelities refer to the final infidelity at the end of quantum state reconstruction. The blue datapoints (std. QST) corresponds to infidelities of state tomography on the same dataset without using error mitigation. **Left:** Single qubit, shot budget of 10000, averaged over 40 Haar random states. The identified optimal shot ratio is marked in green. **Right:** Two-qubit reconstruction with a shot budget of 40000, averaged over 10 Haar random states. The identified optimal shot ratio is marked in green. Both unmitigated quantum state tomography (std. QST) and readout error mitigated quantum state tomography (QREM) are plotted.

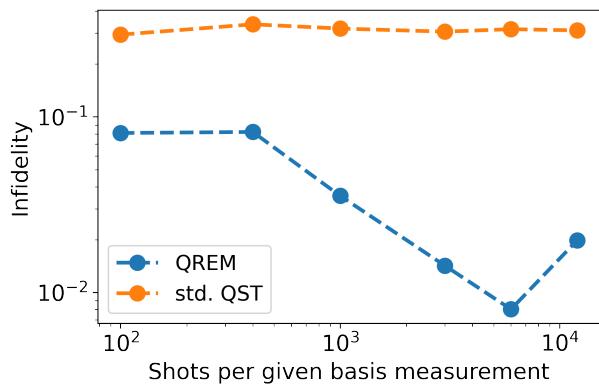


FIG. 6. Reconstruction of a three-qubit density matrix with and without quantum readout error mitigation (QREM) as a function of shots per basis measurement.

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 multi-qubit readout.

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

The dependence of readout correlations on the presence of a readout noise source was studied through lowering the computational basis state distinguishability by using readout pulses with varying amplitudes. 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. At extremum, 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 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.

## VI. CONCLUSIONS AND OUTLOOK

We investigated a recently developed 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 established a benchmark for qubit readout that surpasses traditional methods based on the fidelity of computational basis state assignments by evaluating quantum in-

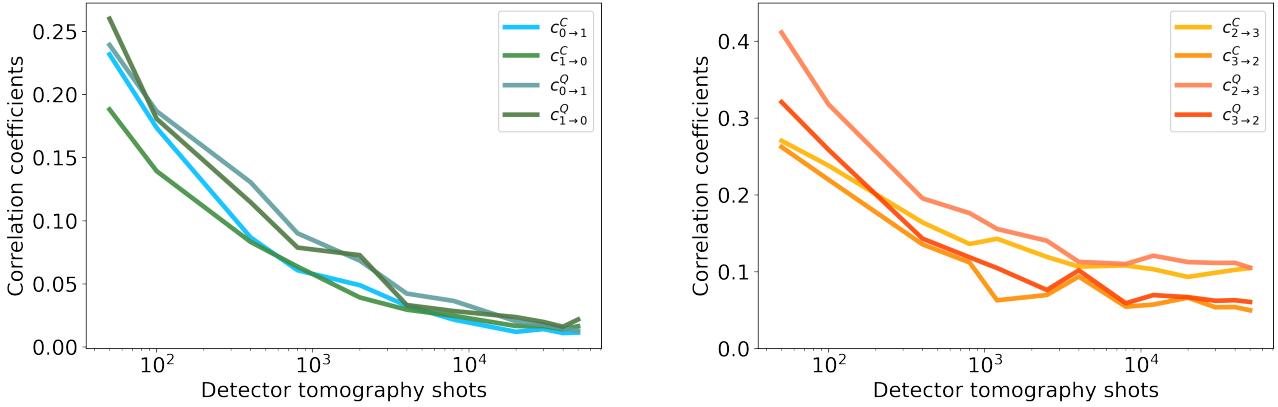


FIG. 7. Correlation coefficients for Systems A and C. **Left:** The correlations in System A do not saturate with increasing detector tomography shots within experimental limitations. **Right:** System C has nonzero asymmetric correlations, converging to around 0.1.

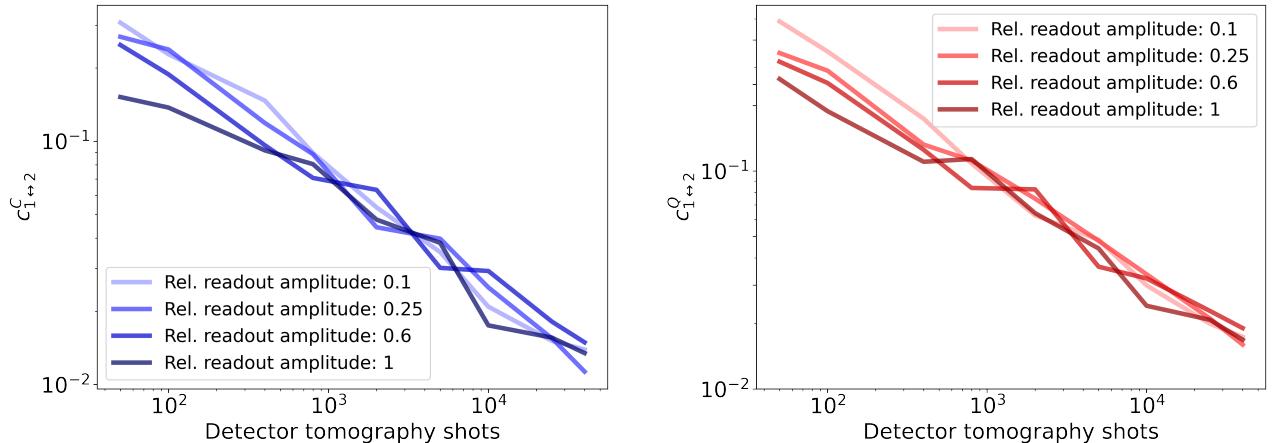


FIG. 8. Convergence of two-qubit symmetric correlation coefficients in system A as a function of detector tomography shots for varying readout noises. **Left:** Classical correlations decrease independently from the relative readout amplitude. **Right:** Quantum correlations have larger values but also decrease independently from the relative readout amplitude.

fidelities between prepared and reconstructed states as a function of the number of measurements performed. We thus propose to use the rate of infidelity convergence as a suitable and comprehensive figure of merit for certifying single- and multi-qubit readout processes. 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 space. However, to fully quantify the efficiency of information extraction on general states, we recommend using the more comprehensive method intro-

duced in this paper.

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 both quantum and classical correlations are significant in the weakly detuned 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 novel

qubit architectures [36] and readout methods [37]. 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.

## DATA AVAILABILITY

The experimental data are available upon request from A. Di Giovanni.

## 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.

## AUTHOR CONTRIBUTIONS

## COMPETING INTERESTS

The authors declare no competing interests.

- 
- [1] A. et al., “Suppressing quantum errors by scaling a surface code logical qubit,” *Nature*, vol. 614, p. 676–681, Feb. 2023.
- [2] Acharya, “Quantum error correction below the surface code threshold,” 2024.
- [3] Y. Li and S. C. Benjamin, “Efficient variational quantum simulator incorporating active error minimization,” *Physical Review X*, vol. 7, June 2017.
- [4] K. Temme, S. Bravyi, and J. M. Gambetta, “Error mitigation for short-depth quantum circuits,” *Physical Review Letters*, vol. 119, Nov. 2017.
- [5] Y. Quek, D. Stilck França, S. Khatri, J. J. Meyer, and J. Eisert, “Exponentially tighter bounds on limitations of quantum error mitigation,” *Nature Physics*, July 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<sup>-4</sup> in a superconducting qubit,” *npj Quantum Information*, vol. 9, Nov. 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,” *Physical Review Applied*, vol. 10, Nov. 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 Information*, vol. 7, Jan. 2021.
- [9] E. Hyypää, 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*, vol. 5, Sept. 2024.
- [10] A. Marton and J. K. Asboth, “Coherent errors and readout errors in the surface code,” *Quantum*, vol. 7, p. 1116, Sept. 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*, vol. 594, p. 369–373, June 2021.
- [12] J. Tuziemski, F. B. Maciejewski, J. Majsa, O. Słowik, 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,” 2023.
- [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,” *Physical Review Applied*, vol. 7, May 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,” *Physical Review Applied*, vol. 10, Sept. 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,” *Nature Physics*, vol. 5, p. 27–30, Nov. 2008.
- [16] Y. Chen, M. Farahzad, S. Yoo, and T.-C. Wei, “Detector tomography on IBM quantum computers and mitigation of an imperfect measurement,” *Physical Review A*, vol. 100, Nov. 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*, vol. 4, p. 257, Apr. 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 approx-

- imate optimization algorithm,” *Quantum*, vol. 5, p. 464, June 2021.
- [19] A. S. Aasen, A. Di Giovanni, H. Rotzinger, A. V. Ustinov, and M. Gärttner, “Readout error mitigated quantum state tomography tested on superconducting qubits,” *Communications Physics*, vol. 7, Sept. 2024.
- [20] 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,” *Applied Physics Letters*, vol. 101, p. 042604, July 2012.
- [21] 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*, vol. 350, p. 307–310, Oct. 2015.
- [22] 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,” *Nature Communications*, vol. 15, May 2024.
- [23] 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*, vol. 4, June 2023.
- [24] 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,” *Applied Physics Letters*, vol. 96, Mar. 2010.
- [25] J. Fiurášek, “Maximum-likelihood estimation of quantum measurement,” *Physical Review A*, vol. 64, July 2001.
- [26] A. I. Lvovsky, “Iterative maximum-likelihood reconstruction in quantum homodyne tomography,” *Journal of Optics B: Quantum and Semiclassical Optics*, vol. 6, pp. S556–S559, May 2004.
- [27] 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,” *Physical Review A*, vol. 93, Jan. 2016.
- [28] J. Cotler and F. Wilczek, “Quantum overlapping tomography,” *Physical Review Letters*, vol. 124, Mar. 2020.
- [29] F. B. Maciejewski, Z. Puchała, and M. Oszmaniec, “Operational quantum average-case distances,” *Quantum*, vol. 7, p. 1106, Sept. 2023.
- [30] B. Lienhard, A. Vepsäläinen, L. C. G. 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,” 2021.
- [31] 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,” *Physical Review Letters*, vol. 111, Aug. 2013.
- [32] K. Życzkowski, K. A. Penson, I. Nechita, and B. Collins, “Generating random density matrices,” *Journal of Mathematical Physics*, vol. 52, June 2011.
- [33] 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,” *Nature Physics*, vol. 18, p. 783–788, May 2022.
- [34] 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 Information*, vol. 9, May 2023.
- [35] T. Itoko, M. Malekakhlagh, N. Kanazawa, and M. Takita, “Three-qubit parity gate via simultaneous cross-resonance drives,” *Physical Review Applied*, vol. 21, Mar. 2024.
- [36] A. Kreuzer, T. Krumrey, H. Tohamy, H. Rotzinger, and A. Ustinov, “Flux qubits with a compact array of vertically stacked josephson junctions,” in preparation.
- [37] 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,” *Nature Electronics*, vol. 7, p. 288–298, Apr. 2024.
- [38] G. Zhu, D. G. Ferguson, V. E. Manucharyan, and J. Koch, “Circuit qed with fluxonium qubits: Theory of the dispersive regime,” *Physical Review B*, vol. 87, Jan. 2013.

## APPENDIX

### Appendix A: Qubit and resonator parameters

Qubit and resonator parameters varied slightly between different cooldown. Some representative values are listed in the two tables below.

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

TABLE I. Qubit names, 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	Assignment fid.	$T_1$	$T_2$
q0	96%	118 $\mu s$	67 $\mu s$
q1	93%	22 $\mu s$	30 $\mu s$
q2	92%	27 $\mu s$	56 $\mu s$
q3	80%	23 $\mu s$	ca. 15 $\mu s$

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. Multi-qubit assignment fidelities were observed to be lower in some experiments.

### Appendix B: Readout chain

To understand the origin of readout errors, we give 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 [38]. This signal is amplified first on the mixing chamber stage with a Josephson travelling 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, yielding finally a measured value of 0 or 1.

### Appendix C: Leakage and assignment fidelity

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 3 regions in the IQ plane, corresponding to ground- and excited states and the third (leaked) state. We prepare an equal amount 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 3 clusters (ground, excited, leaked) and linear sum assignment is performed to relabel the clusters in maximum accordance with the known (but sometimes wrong) labels 0 or 1. Note that the lowest readout power 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 q1 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 D: 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 done with a different set of microwave cables, therefore there is a constant offset in power compared to Fig. 4.

### Appendix E: 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 no knowledge of the quantum states to be estimated is known, then there are  $4^n - 1$  parameters for each normalized  $n$ -qubit state. A  $m$  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. (9)

$$r_n = \frac{(m-1)4^n}{(m-1)4^n + N_s(4^n - 1)} \stackrel{\text{IC}}{=} \frac{4^n}{4^n + N_s}, \quad (\text{E1})$$

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 can arise from the fact that not all the parameters are free and

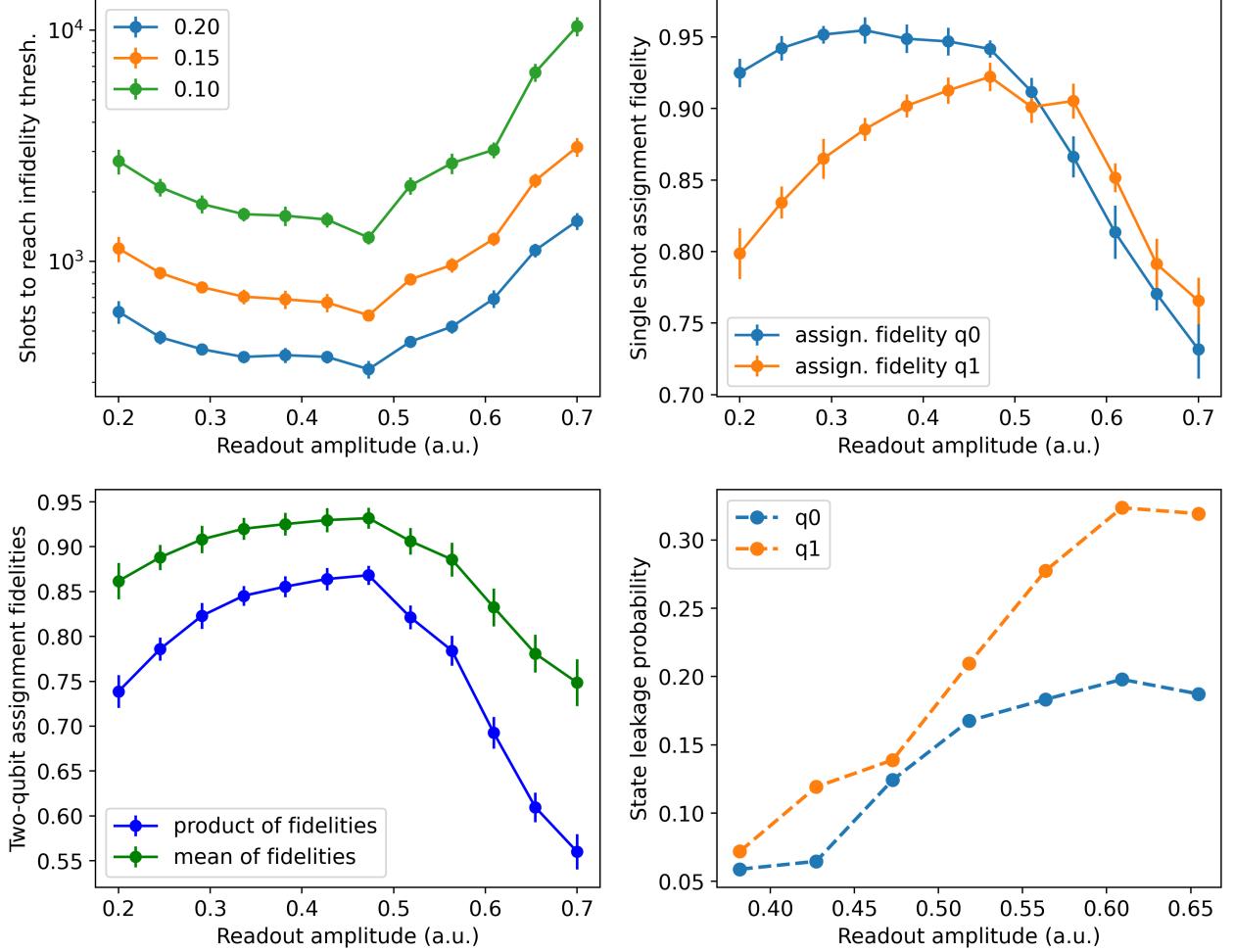


FIG. 9. Comparison of readout error mitigated quantum state tomography based readout benchmark with assignment fidelities. **(a):** The number of two-qubit state tomography shots required to reach infidelity thresholds of 0.1, 0.15 and 0.2 for various readout amplitudes. The error bars are bootstrapped 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 take into account correctly. **(b):** Single shot assignment fidelities during the experiments, each point uses 1000 shots. The error bars show one standard deviation. **(c):** Mean and product of single shot assignment fidelities. The two curves take a remarkably similar shape. The error bars show the propagated standard deviation from **(b)**. **(d):** Probability of measuring a leaked state as a function of readout amplitude.

independent in a physical setting. For example, if the reconstructed states are close to pure, there are effectively  $2^n$  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.

we assume close to classical errors the ratio is

$$r_n \stackrel{\text{mixed}}{=} \frac{2^n}{2^n + N_s}, \quad (E2)$$

$$\stackrel{\text{pure}}{=} \frac{(4^n - 1)}{(4^n - 1) + N_s}$$

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

which, compared to the experiment ( $N_s = 1$ ) gives us  $r_1 = (0.67, 0.75)$  and  $r_2 = (0.80, 0.94)$  for (mixed, pure) target states.

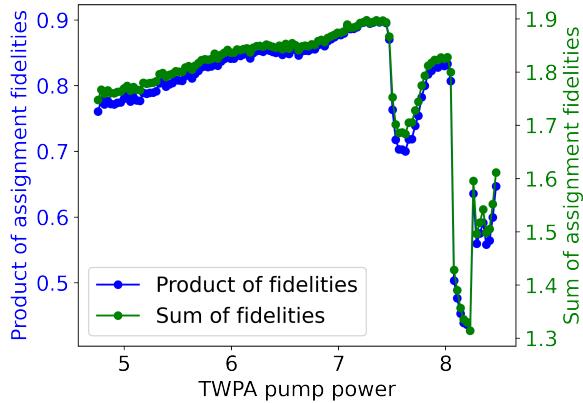


FIG. 10. Sum and product of single qubit assignment fidelities in System A. A sharp drop in both variables occurs at approx. 7.3 dBm pump power.

#### Appendix F: Measurement drift

As both quantum detector- and state tomography scale exponentially in runtime with the number of qubits, the experiment has to run self-consistently over a prolonged time. A 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 during 11 hours, 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. 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 to originate from the manipulation of the qubit, used for XX basis measurements.

#### Appendix G: 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 errors in the worst entry of the unmitigated density matrix (see Fig. 12). The non-vanishing matrix entries have the largest errors both with and without readout error mitigation. From these density matrices the infidelity of reconstruction was calculated.

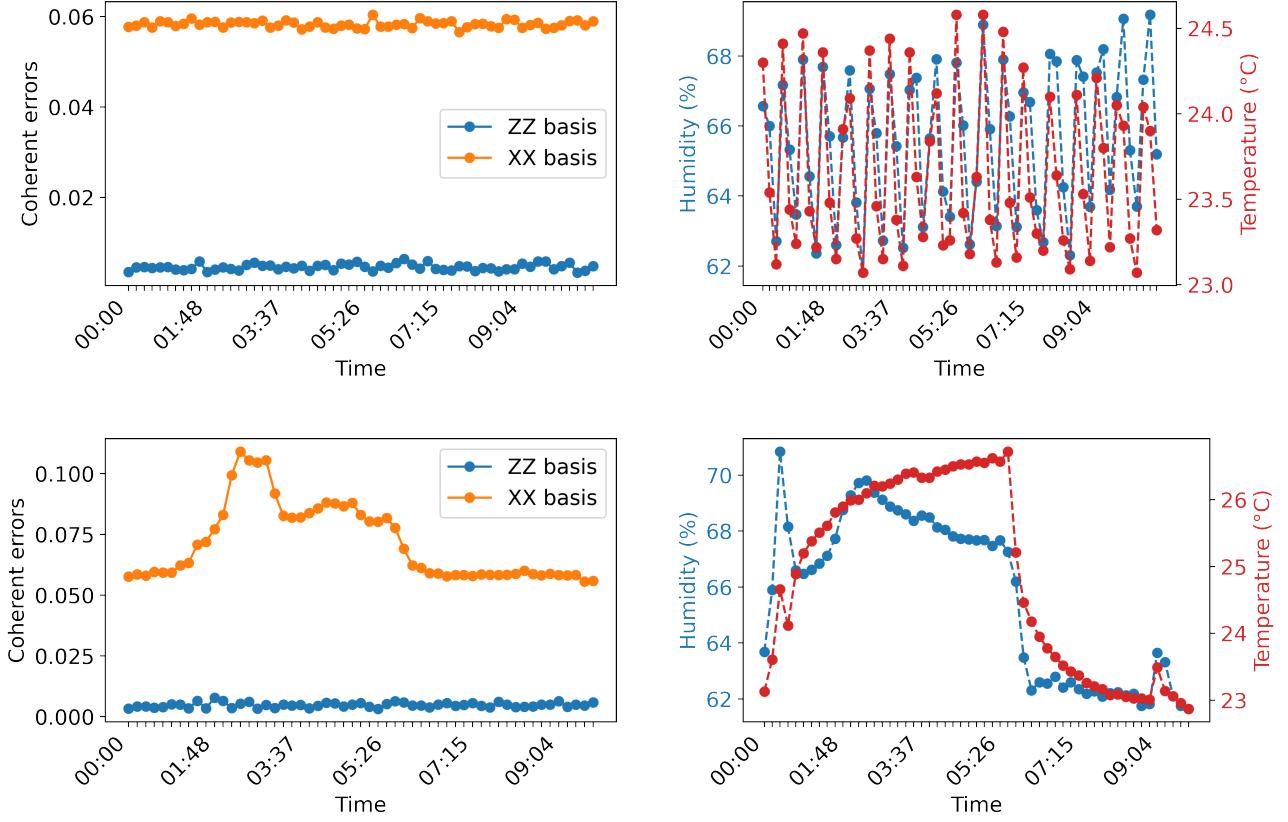


FIG. 11. Stability of coherent readout errors with varied laboratory temperature and humidity. **Top left:** Coherent errors (defined in Eq. 6) in the reconstructed measurement operators for ZZ and XX basis readout under stable laboratory conditions. **Top right:** Temperature and humidity values during 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 measurement.

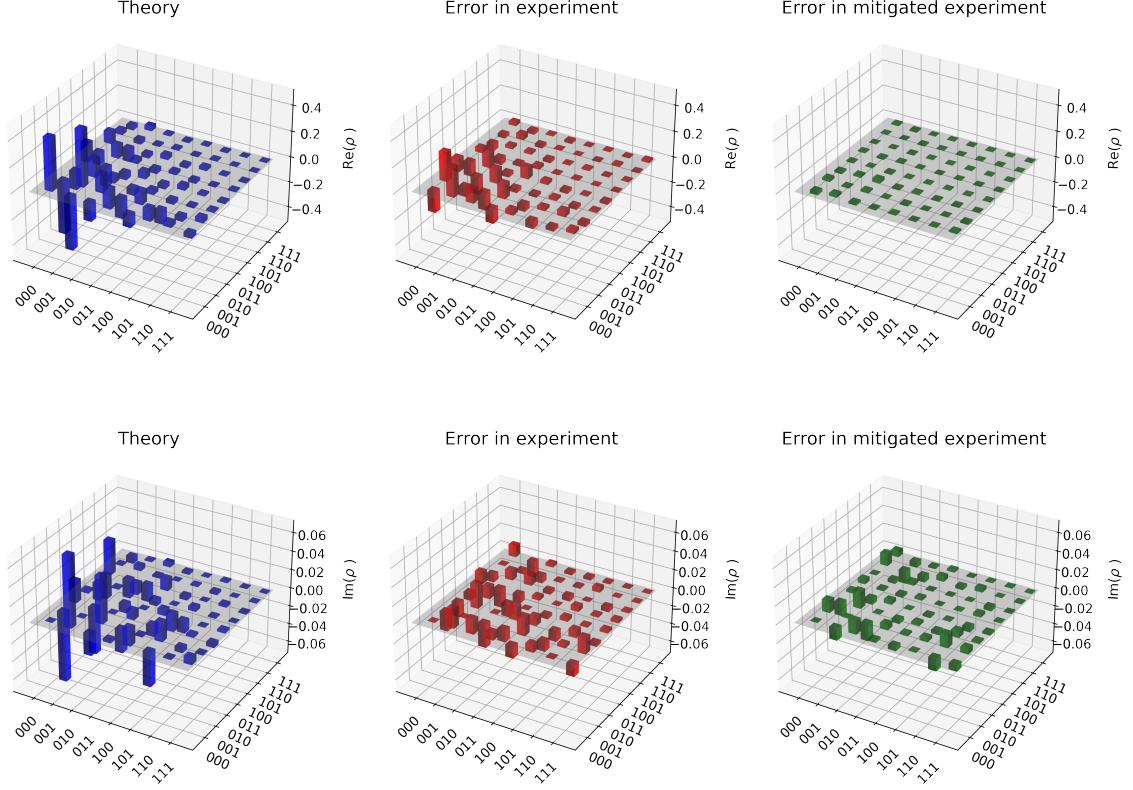


FIG. 12. Comparison of theoretical density matrix of a three-qubit state and its reconstruction with and without readout error mitigation. The top row shows the real part, the bottom row the imaginary part of the complex 8-dimensional matrices. The quantum state reconstructed with and without readout error mitigation have infidelities 0.030 and 0.447 respectively. The different scale in real and imaginary parts is due to the chosen quantum state. The errors in the imaginary and real part of the error mitigated experiments are of the same order of magnitude.