Readout error mitigated quantum state tomography tested on superconducting qubits

Check for updates

Adrian Skasberg Aasen [1,2,5], Andras Di Giovanni [3,5], Hannes Rotzinger [3,4], Alexey V. Ustinov[3,4] & Martin Gärttner [2] ✉

Quantum technologies rely heavily on accurate control and reliable readout of quantum systems. Current experiments are limited by numerous sources of noise that can only be partially captured by simple analytical models and additional characterization of the noise sources is required. We test the ability of readout error mitigation to correct noise found in systems composed of quantum two-level objects (qubits). To probe the limit of such methods, we designed a beyond-classical readout error mitigation protocol based on quantum state tomography (QST), which estimates the density matrix of a quantum system, and quantum detector tomography (QDT), which characterizes the measurement procedure. By treating readout error mitigation in the context of state tomography the method becomes largely readout mode-, architecture-, noise source-, and quantum state-independent. We implement this method on a superconducting qubit and evaluate the increase in reconstruction fidelity for QST. We characterize the performance of the method by varying important noise sources, such as suboptimal readout signal amplification, insufficient resonator photon population, off-resonant qubit drive, and effectively shortened T_1 and T_2 coherence. As a result, we identified noise sources for which readout error mitigation worked well, and observed decreases in readout infidelity by a factor of up to 30.

Building quantum machines capable of harnessing superposition and entanglement promises advances in many fields ranging from cryptography[1,2], material simulation[3], and drug discovery[4,5] to finance[6–8] and route optimizations[9]. While these applications assume a large-scale fault-tolerant quantum computer, currently, we are still in the era of noisy intermediate-scale quantum (NISQ) devices[10], where noise and imperfection of qubits fundamentally limit the applicability of quantum algorithms. In this work, we refer to noise as the physical phenomenon which causes variable operation of the experiment, and error is the deviation of the measured quantities with respect to their theoretical expectation values. To harness the power of quantum computation and quantum simulation today, we need methods that are suitable for working with noisy hardware.

There are predominantly two approaches to combating errors in qubit systems. On the one hand, one can study the exact sources of noise and errors, in order to find new materials or techniques to eliminate them. For superconducting qubits, for instance, parasitic two-level systems coupling to

qubits or resonators have been shown to be a limiting factor for qubit coherence[11], leaving room for further hardware improvements.

On the other hand, instead of reducing the noise inherent in the hardware, one can apply algorithmic methods to handle and reduce the errors. There are primarily two approaches to algorithmic error correction[12], of which the first and most prominent is quantum error correction. The core idea here is to encode logical qubits in multiple noisy physical qubits. Adaptive corrective operations are then performed based on syndrome measurements[13]. These methods require low enough gate errors such that the qubit number overhead does not blow up. Current gate fidelities are too low for general applicability, however, promising results have been achieved by using surface codes[14–17].

One prominent source of errors not captured in error correction are readout errors, a subset of state preparation and measurement (SPAM)-errors[18,19]. Readout errors occur in the process of reading out the state of the qubit, e.g., spin projection measurements. Experimentally, one can

[1]Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany. [2]Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany. [3]Physikalisches Institut, Karlsruher Institut für Technologie, Kaiserstraße 12, 76131 Karlsruhe, Germany. [4]Institut für QuantenMaterialien und Technologien, Karlsruher Institut für Technologie, Kaiserstraße 12, 76131 Karlsruhe, Germany. [5]These authors contributed equally: Adrian Skasberg Aasen, Andras Di Giovanni. ✉e-mail: martin.gaerttner@uni-jena.de

understand the importance of such errors through the example of a superconducting resonator measurement of a transmon. When using only a few photons to read out the resonator, there is a nonzero probability of an incorrect readout. However, by increasing the number of photons, higher qubit levels will be excited, leading to a sharp drop in readout fidelity[20].

Readout errors are better handled by the second approach, quantum error mitigation. These kinds of methods do not correct the errors in execution of the quantum algorithm, but rather reduce the errors by post-processing the measured data. There exists a broad spectrum of error mitigation methods[12,21,22], a large class of which focuses on readout errors, such as unfolding[23,24], T-matrix inversion[25], noise model fitting[26,27], and detector tomography-based methods[28]. Other promising results have recently been shown on an NISQ processor[29] using zero-noise extrapolation[30,31].

Our proposed method falls into the category of detector tomography-based mitigation, similar to ref. 28, where the focus is on the correction of readout probability vectors. Detector tomography-based methods use a set of calibration states, e.g., the eigenstates of the Pauli matrices, to characterize the measurement outcomes realized by the measurement device using the positive operator-valued measure (POVM) formalism. By looking at the reconstructed measurements, one gains information about the noise present in the experiment[32], which can then be leveraged for error mitigation. We emphasize that readout error mitigation operates on the statistics of output distributions collected from multiple runs of the experiment, and is not able to correct the results of an individual projective measurement.

Our approach generalizes the ideas of previous protocols by using a complete characterization of a quantum system. This is done by quantum state tomography, which directly estimates the density matrix of the system[33,34]. The question of QST in error mitigation was also considered in ref. 28, but restricted to only classical errors, i.e., errors that can be described as a stochastic redistribution of the outcome statistics of single basis measurements. Reference 32 does perform a general reconstruction of the detectors, and suggests it can be used in QST, but does not provide any protocol that goes beyond correction of marginals with numerical inversion, mostly focusing on classical errors in their applications. Similar ideas were compared to data pattern tomography in refs. 35,36 with linear inversion. In ref. 37, some of the error mitigation methods mentioned above were compared with simulated depolarizing noise in the context of state reconstruction. The central aspect of our approach is the direct integration of QDT with QST, which leads to an error mitigation scheme that does not require matrix inversions or numerical optimizations. On top of this the protocol makes no assumptions about the error channel that represents the readout noise or about the type of quantum state (e.g., entangled, separable, pure, mixed) that it applies to. As we do not restrict the set of allowed POVMs in the detector tomography in any way, the protocol is also agnostic to the architecture (e.g., superconducting qubits, photons, ultracold-atoms, neutral atoms) and the type of readout mode (e.g., single projective measurement (photons), resonator readout (superconducting qubits)).

With this general scheme, we test the limits of readout error mitigation by subjecting the protocol to common noise sources. Previous works have been extensively tested on publicly accessible IBM quantum computers, which only allow for very limited experimental control over noise sources. In this paper, we evaluate our proposed method on a chip, where we have full control over the experiment. This allows us to systematically induce and vary important noise sources, which leads to a better understanding of the strengths and limitations of the protocol.

Results and discussion
Preliminaries
This section reviews the relevant theoretical concepts to describe general quantum measurements and state reconstruction[13,38,39]. Readers familiar with these concepts may want to skip this section.

Notation used in this paper unless stated otherwise: Objects with tilde, e.g., $\tilde{p}$, indicate that it has been subject to readout noise. Objects with a hat, e.g., $\hat{p}$, are experimentally measured quantities, which contain sample fluctuations and noise. The hat is also used to denote estimators, which will be clear from the context. Operators and objects without either accent will be considered theoretically ideal objects. The two eigenstates of the Pauli operators $\sigma_x, \sigma_y,$ and σ_z, will be denoted by $|0_i\rangle$ and $|1_i\rangle$, where $i \in \{x, y, z\}$. If no subscript is given, the z-eigenstates are implied.

Generalized measurements. Measurements on quantum systems are described by expectation values of operators, given by Born's rule[13,38],

$$\langle O \rangle = \mathrm{Tr}(\rho O), \tag{1}$$

where ρ defines the quantum system and O represents a measurable quantity. Experiments typically perform projective measurements onto a set of basis states. The simplest example is a measurement on a computational basis, with two possible outcomes. The projective operators take on the form $P_0 = |0\rangle\langle0|$ and $P_1 = |1\rangle\langle1|$, where $p_i = \mathrm{Tr}(\rho P_i)$ gives the probability of obtaining the outcome i.

In this work, we are interested in noisy measurements. To be able to capture realistic readout we need to move to generalized quantum measurements, described by positive operator-valued measures (POVMs). A POVM is a particular set of operators $\{M_i\}$, sometimes called effects, that have the interpretation of yielding a probability distribution over measurement outcomes through their expectation values. In particular, any POVM has the three following properties[13]:

$$M_i^\dagger = M_i, \quad M_i \geq 0, \quad \text{and} \quad \sum_i M_i = \mathbb{1}, \tag{2}$$

where the first property guarantees that the resulting expectation values are real, the second guarantees positive expectation values, and the third guarantees that the expectation values sum to unity. Altogether, we have a natural correspondence between each operator M_i and the probability of getting outcome i of a random process,

$$p_i = \langle M_i \rangle = \mathrm{Tr}(\rho M_i). \tag{3}$$

Through this general interpretation of the effects M_i, one can assign each effect to a possible outcome from a measurement device.

We are particularly interested in POVMs that form a complete basis in their respective Hilbert space. Such a POVM is called informationally complete (IC). Using an IC POVM allows one to decompose any operator O in terms of the set $\{M_{\mathrm{IC},i}\}$,

$$O = \sum_i c_i M_{\mathrm{IC},i}. \tag{4}$$

If the POVM is minimal, meaning that its elements are linearly independent, the coefficients c_i are unique. For qubit systems to be informationally complete, we need 4^n linearly independent POVM effects, where n is the number of qubits in the system. An example of an IC POVM is the so-called Pauli-6 POVM. Experimentally, such a measurement can be performed by randomly selecting either the $x-$, $y-$, or $z-$ basis and then performing a projective measurement in that basis. This POVM is defined by the set $\left\{\frac{1}{3}|0_i\rangle\langle0_i|, \frac{1}{3}|1_i\rangle\langle1_i|\right\}$, where $i \in \{x, y, z\}$. Note that the Pauli-6 POVM is not minimal, and its elements are not projectors, due to the prefactor acquired by the random basis selection.

Representation of readout noise. Measurements of quantum objects suffer from imperfections caused by the errors introduced by the measurement device. This means that when intending to perform the POVM $\{M_i\}$, we actually perform the erroneous POVM $\{\tilde{M}_i\}$. In reading out the quantum state ρ we observe the distribution dictated by Born's rule

$$\tilde{p}_i = \mathrm{Tr}(\rho \tilde{M}_i), \tag{5}$$

which we will dub a passive picture of noise. This runs counter to the more common view, where the noise is applied to the quantum state rather than the readout operator, hereafter referred to as the active picture of noise,

$$\tilde{p}_i = \text{Tr}(\tilde{\rho} M_i), \tag{6}$$

where $\tilde{\rho} = \mathcal{E}(\rho)$ is the ideal quantum state passed through a noise channel[13]. These two views are equivalent, since we can only access the probabilities $\tilde{p}_i$. In the remainder of this work, we will be working on the passive noise picture.

Quantum state tomography. The objective of quantum state tomography is to reconstruct an arbitrary quantum state from only the measurement results. The most basic approach to QST can be framed as a linear inversion problem. To uniquely identify the quantum state, we use the decomposition provided by an IC POVM in eq. (4),

$$\rho = \sum_i a_i M_i, \tag{7}$$

where we have dropped the subscript "IC". The task of linear inversion is to determine the real coefficients a_i through the recorded measurement outcomes. Linear inversion is often considered bad practice as it can yield unphysical estimates[40]. It is advisable to use estimators such as the likelihood-based Bayesian mean estimator (BME)[40,41], or maximal likelihood estimator (MLE)[42], which relies directly on the likelihood function

$$\mathcal{L}_M(\rho) \propto \Pi_i \text{Tr}(\rho M_i)^{n_i} = \left(\Pi_i \text{Tr}(\rho M_i)^{\hat{p}_i}\right)^N, \tag{8}$$

which represents the likelihood that the prepared state of the system was ρ, given that outcome i has been observed n_i times, or, equivalently with outcome frequency $\hat{p}_i = \frac{n_i}{N}$. In our notation, the subscript M indicates what POVM was used for the measurement. The estimated state is the integrated mean

$$\rho_{\text{BME}} \propto \int d\rho \, \rho \mathcal{L}_M(\rho) \tag{9}$$

for BME, and the maximizer

$$\rho_{\text{MLE}} = \text{argmax}_\rho \, \mathcal{L}_M(\rho) \tag{10}$$

for MLE.

Noise-mitigated quantum state tomography by detector tomography

In contrast to previous works, which view error mitigation as the correction of outcome statistics (see, e.g., ref. 28), our approach integrates the error mitigation procedure directly into the quantum state tomography framework. This comes with crucial advantages, as we will see, because it allows for the mitigation of a broader class of readout errors and intrinsically yield physical reconstruction without any additional corrective steps. We stress that this and similar protocols only work on mitigation of the statistical behavior of the quantum system, and is not suited for mitigation on the level of single-shot readout.

Quantum detector tomography. QDT[43] can be summarized as reconstructing the effects associated with a given measurement outcome, based on observed outcomes from a set of calibration states. Another, perhaps more instructive, view is that QDT finds a map between the ideal and actual POVM $\{M\} \rightarrow \{\tilde{M}\}$. Such a map could be used to extract noise parameters or learn the general behavior of the measurement device. However, our protocol avoids explicitly reconstructing such a map, and does not even require the knowledge of the ideal POVM $\{M\}$.

QDT starts out by preparing a complete set of calibration states, which span the space of all quantum states, and repeatedly performs measurements on these states. An example is the (over)complete set of Pauli density operators, $\{|0_i\rangle\langle 0_i|, |1_i\rangle\langle 1_i|\}$, where $i \in \{x, y, z\}$, which spans the space of single qubit operators. With the outcomes from all of these measurements, we can set up a linear set of equations given by Born's rule,

$$\frac{n_{is}}{N} = p_{is} = \text{Tr}(\rho_s M_i), \tag{11}$$

where the subscript s iterates over the set of calibration states, e.g., $\{\rho_1 = |0_x\rangle\langle 0_x|, \rho_2 = |1_x\rangle\langle 1_x|, \dots\}$. Equation (11) represents a set of $I \times S$ constraints on $\{M_i\}$ we seek to reconstruct, where S is the total number of unique calibration states prepared and I is the number of POVM elements. In addition, one has the normalization constraint on the effects, $\sum_i M_i = \mathbb{1}$. Altogether, this yields a statistical estimation problem of a very similar nature to the one outlined in state tomography. It can be solved, for example, by using a maximal likelihood estimator, such as the one outlined in ref. 44, which guarantees physical POVM reconstructions.

Readout error-mitigated tomography. The core idea of our protocol is to use QDT as a calibration step before the state reconstruction. Using the information gathered from reconstructing the measurement effects, we modify the standard state estimator using the passive picture of noise, eq. (5). In this way, the estimator is "aware" of the noise and corresponding errors present in the measurement device.

A schematic overview of the protocol is given in Fig. 1. The first step is to reconstruct the POVM of the measurement device using detector tomography. This gives us access to an estimated noisy POVM $\{\tilde{M}^{\text{estm}}_i\}$. This noisy POVM is fed into the quantum state estimator, giving us the likelihood function

$$\mathcal{L}_{\tilde{M}^{\text{estm}}}(\rho) \propto \Pi_i \text{Tr}\left(\rho \tilde{M}^{\text{estm}}_i\right)^{n_i} = \left(\Pi_i (\tilde{p}_i)^{\hat{p}_i}\right)^N, \tag{12}$$

where we have used that $\tilde{p}_i = \text{Tr}(\rho \tilde{M}^{\text{estm}}_i)$. This does indeed give us an estimator that converges to the noiseless state ρ, see Supplementary Note 1 for more information. We highlight that working in the passive noise picture comes with benefits not enjoyed by other estimators. Firstly, no inversion of a noise channel is required, and secondly, the reconstructed state is guaranteed to be physical. Everything is handled internally by the estimator.

Our protocol assumes the following experimental capabilities:
- Access to an IC POVM.
- Perfect state preparation.

The first capability is required by any full state reconstruction method, which can be done by any quantum device that has the ability to perform single qubit rotations and readout. By single qubit rotations, we mean any operation on a quantum device that can be decomposed into tensor products of single qubit rotations acting on each qubit. Similarly, for readout, one only needs tensor products of single qubit POVMs. This is also important for the noise mitigation to be general, since an unambiguous characterization of the erroneous state transformation inevitably requires an IC POVM. The second assumption is more problematic, as this is not strictly fulfilled in any real experiment. To circumvent this problem, one needs to consider state preparation and readout errors in a unified framework which is an active field of research. For this work it suffices to make sure that the state preparation errors are small compared to the readout errors. We discuss this further in subsection "Protocol limitations". For the explicit implementation of the protocol used in the remainder of the paper, see "Explicit protocol realization" in the Methods section.

Experimental results

We present the results of an experimental implementation of the readout error mitigation protocol on a superconducting qubit device. Important

Fig. 1 | Protocol schematic overview. a Detector tomography: A complete set of basis states (e.g., the Pauli states) are prepared and measured repeatedly by the experiment. Based on the outcomes of the measurements, a positive operator-valued measure (POVM) is reconstructed. In a sense, it associates the measured outcome (here visualized as spin up/down measurements) with a measurement effect $\tilde{M}_i$. **b** State tomography: Using the reconstructed POVM, the modified likelihood function is endowed with knowledge of the operation of the measurement device. The system of interest is then prepared and measured repeatedly with the desired number of shots.
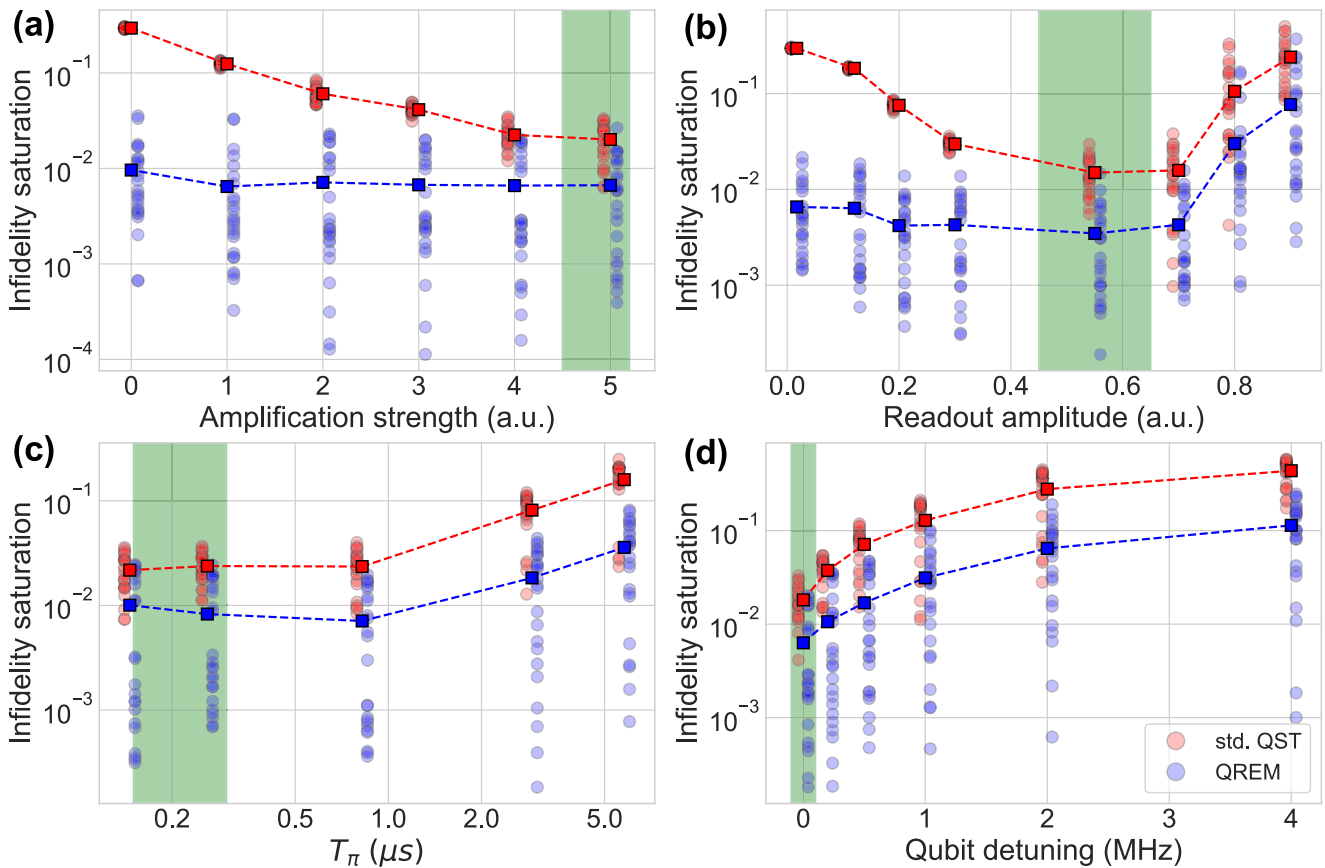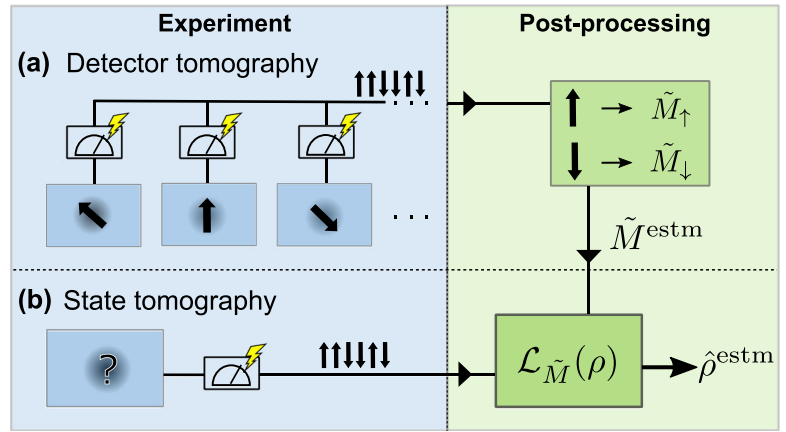
Fig. 2 | Characterizing error mitigation by detector tomography for quantum state tomography (QST) for different noise sources. Infidelity saturation refers to the last infidelity point measured over 240k single-shot measurements. The infidelity saturation for each individual run is plotted in translucent circles and shifted off center, to the left for standard QST and to the right for quantum readout error-mitigated (QREM) QST, for better visibility. The solid colored squares are the average infidelity saturation, connected by dotted lines for guidance. The green highlighted areas indicate the optimal experimental parameters. **a** Decreasing parametric amplification has a significant effect on QST through lesser distinguishability of the two states. Such errors can be mitigated very effectively by detector

tomography. Zero amplification strength corresponds to having turned off the amplifier. **b** An incorrectly set readout amplitude of the resonator leads to increased infidelity in both the too weak and too strong readout regimes. Mitigation fails at higher powers, because higher levels are excited that are not taken into account. **c** Increasing manipulation timescales (T_π being the pulse time required for a bitflip) leads to more T_1 and T_2 decay events. This can be mitigated to some extent by the protocol, as seen by the smaller gradient of the mitigated infidelity curve. **d** Detuning between the qubit transition and drive frequencies can be efficiently mitigated. The mitigated infidelities rise proportionally to the unmitigated ones.

noise sources were varied to study their effect on state reconstruction accuracy. For more information on the induced noise sources and the experimental setup, see the Methods section. Quantum infidelity was used as a measure for state reconstruction accuracy. To make sure infidelity is a reliable performance metric, the infidelity is averaged over Haar-random

target-states, shown as translucent circles in Fig. 2, see Methods for more information.

Insufficient readout amplification. We study the influence of insufficient signal amplification by the wideband traveling-wave parametric

amplifier, in the following called amplifier. All other amplifiers operate with their optimal operational settings. We tune the amplification through the power of the microwave drive of the amplifier. This results in a reduced distinguishability of the two states. The measured saturation values of infidelity for different amplifications is given in Fig. 2a. Lowering the amplification reduces distinguishability in the IQ plane, resulting in high infidelities of around 0.3 with the amplifier turned off. The readout error mitigation protocol performs very well against this type of noise. When error mitigation is used, the reconstruction infidelity does not increase unless the amplifier is turned off completely, indicating that readout error mitigation is effective against this type of noise.

Readout resonator photon number. Ideally, superconducting qubits are read out with low resonator populations. Higher power readout will excite the qubit, potentially also to higher states outside the qubit manifold. Low readout power also enables non-demolition experiments, e.g., for active reset. If the readout power is too low, the IQ-plane state separation is not sufficient for single-shot readout, increasing the infidelity, an effect that our protocol can mitigate, as can be seen in Fig. 2b. This effectively constitutes a lower signal-to-noise ratio, similar to a low parametric amplification. At higher readout power, the higher transmon states excited by the strong readout are not captured in the two-level quantum system simulation, resulting in increased infidelities for both mitigated and unmitigated reconstruction. We can also see a higher spread of unmitigated infidelities when using readout amplitudes above and including 0.55. A possible explanation for this lies in the fact that the $0 \rightarrow 2$, $0 \rightarrow 1 \rightarrow 2$ and $1 \rightarrow 2$ processes have different frequencies, hence these transitions will be induced spontaneously by readout signals with a probability dependent on the current quantum state. This results in a state-dependence of higher state excitations, and therefore increased infidelity. We note that the infidelity ratio between standard QST and QREM becomes constant at large readout amplitudes.

In the following, we study noise sources that do not exclusively manifest as readout errors, as they also introduce significant errors in state preparation, which breaks one of the core assumptions of the protocol. It is nonetheless insightful to investigate the protocol's performance with such noise sources. Further experiments are required to determine the efficacy of the protocol under such conditions.

Shorter T_1 and T_2 times. Decoherence is modeled by exponential decay in fidelity in the z (T_1) and azimuthal (T_2) directions of the Bloch sphere. Thus, we can increase the number of decay events by increasing the length of manipulation pulses, used for state preparation and readout rotations. Experimentally, this is done by decreasing the manipulation power.

Figure 2c shows the results for varying the π-pulse lengths. At increased manipulation lengths, the qubit has more time to decay both through dephasing and energy loss, resulting in a larger error probability. Both unmitigated and mitigated state reconstruction gets progressively less accurate with increased manipulation lengths, with a noise-strength independent factor of 3–4 between them. The combined effect of decoherence on all three basis measurements is expected to be highly state-dependent, which can explain the large infidelity saturation spreads at a given noise strength.

Longer readout pulses lead to state decay already in the state preparation stages, hence the studied noise source is not only a readout error, but falls into the broader SPAM error category. One could restrict induced decay to purely affect readout, e.g. by only decreasing the manipulation power at the readout stage, but this would arguably not correspond to a realistic experimental scenario. Instead, consistently decreasing the manipulation power presents a more relevant experimental scenario.

Qubit detuning. If we consistently apply pulses detuned in frequency, we observe an apparent infidelity reduction by a factor of 4 by the protocol across all detunings, as shown in Fig. 2d. One can also see, that even a small detuning of 0.1 MHz results in lower infidelity than standard QST with perfect frequency-matching. We note that the infidelity saturation of QREM grows immediately with detuning, meaning that the protocol is not able to completely mitigate even small detunings. This can be attributed to the detector tomography step also suffering severely from effects of state preparation inaccuracies. As in the case of increased manipulation times, one could conduct an experiment where the detuning only affects the readout stage by applying resonant pulses for state preparation. As this also does not correspond to a realistic experimental scenario, the more relevant approach is to apply consistently detuned pulses.

At 4 MHz, detuning a pulse of 200 ns will result in an accumulated phase offset of 0.8, effectively scrambling σ_x and σ_y measurement outcomes. Both mitigated and conventional QST infidelity saturation depend roughly linearly on detuning, flattening out when approaching an infidelity of 0.5. The large spreads of the measured infidelity bounds stem from the fact that the effect of the noise depends on the state.

Noise that effectively manifests as a large frequency detuning can be induced, for example, by spontaneous iSWAP "operations" with neighboring two-level systems (TLS). Microwave sources may also suffer from frequency drifts. The errors coming from microwave source offsets are much smaller than the considered detunings. However, TLS-enabled jumps of more than 10 MHz have been observed on the same timescale as one run of QST takes for our experiment[45], making it a potential noise source corresponding to larger qubit detunings. It is important that these jumps do not occur after the device tomography stage, otherwise the reconstructed POVMs will not reflect the jump.

We remark that for all the shown cases where readout error mitigation does not work optimally, specifically in (c), (d), and the right part of (b) in Fig. 2, the ratio between standard QST and QREM infidelities is approximately constant. For noise sources that do not exclusively affect the readout process, a deterioration of the noise-mitigated reconstruction fidelity is expected as the noise strength increases. However, the observation of proportionality between the mitigated and unmitigated infidelities requires further investigation.

Combining noise sources. Several of the previously studied noise sources can be combined in different ways to simulate a noisy experiment. To demonstrate the efficiency of the mitigation ad extremum, we devise an experiment with comparably high levels of noise, by measuring with the off-resonant drive of $\delta\omega = 0.5$ MHz, decreased parametric amplification, and with decreased readout amplitude. While conventional QST saturates very early in the reconstruction, few-percent infidelities are possible by employing error mitigation, thereby opening the way for precise measurements with very noisy readouts, see Fig. 3. As expected (see subsection "Infidelity" in Methods), one can observe a linear dependence of infidelity on shot number on a log-log scale, corresponding to a power-law convergence.

Protocol limitations

We present an analysis of the potential limitations to the performance and reliability of the error mitigation protocol. Note that most of the discussed limitations will affect any estimator, and are not unique to our protocol. It is still important to understand their impact on protocol performance.

Sample fluctuations. A central limiting factor to the precision of any estimator are sample fluctuations, i.e., statistical fluctuations in the number of samples per effect. These fluctuations can manifest themselves in two parts of the protocol, QDT and QST, and cannot be mitigated.

QDT acts as a calibration step and is a one-time cost in terms of samples. Fluctuations in the POVM reconstruction can be viewed as a bias introduced into the state reconstruction. Therefore, it is important that the POVM reconstruction does not impose a bias larger than the expected sample fluctuation in the QST itself. In Fig. 4, we demonstrate that using less than 0.5% of experimental shots for detector tomography lowers the reconstruction infidelity to half its value. The lowest infidelity is achieved by using ca. 10% of shots for QDT, after which no meaningful further
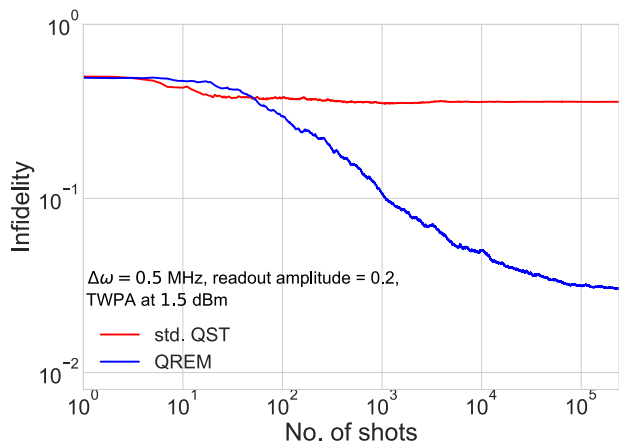
Fig. 3 | Combination of multiple strong noise sources. Conventional quantum state tomography (std. QST) saturates at around 100 shots, while using quantum readout error mitigation (QREM) enables state reconstruction with a factor of 30 lower infidelity. The infidelity is averaged over 25 Haar-random states, to see expected infidelity spread of each Haar-random state, see Supplementary Fig. 1. In this experiment, the readout amplitude was decreased compared to optimal, and the traveling-wave parametric amplifier (TWPA) was pumped with a decreased tone and the qubit was driven with a control pulse detuned in frequency by $\Delta\omega$.
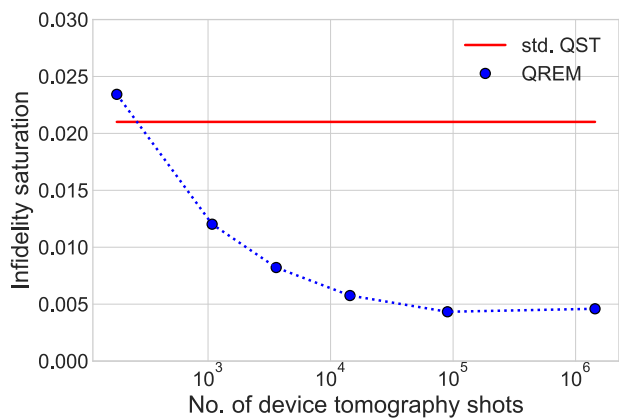
Fig. 4 | Averaged infidelity saturation in quantum state tomography (QST) as a function of calibration shots used in detector tomography. The quantum readout error-mitigated (QREM) measurements are compared to the baseline (std. QST). 240k shots were used for each QST reconstruction, averaged over 25 Haar-random states. Using 1k shots for calibration already results in a decrease in infidelity by a factor of ≈ 2. The curve flattens out at around 100k calibration shots.

improvement was seen in our experiment. We give the recommendation of using half the number of shots used for a single QST for QDT. When averaging over e.g., 25 quantum states for a representative benchmark, this becomes a small overhead.

Error-mitigated estimators, such as the one in eq. (12), aim to reduce the bias with respect to the prepared state caused by a noisy readout process. In reducing the bias, we are, in effect, taking into account that the state has passed through a noise channel. Since noise channels reduce the distinguishability between states, the variance of the modified likelihood function is increased with respect to the unmodified likelihood function, an effect known as the bias-variance trade-off [12,46]. In Fig. 5, we give an artistic rendition of the bias-variance trade-off under mitigation of single qubit depolarizing noise, given by the channel.

$$\mathcal{E}(\rho) = \frac{p}{2}\mathbb{1} + (1-p)\rho. \tag{13}$$
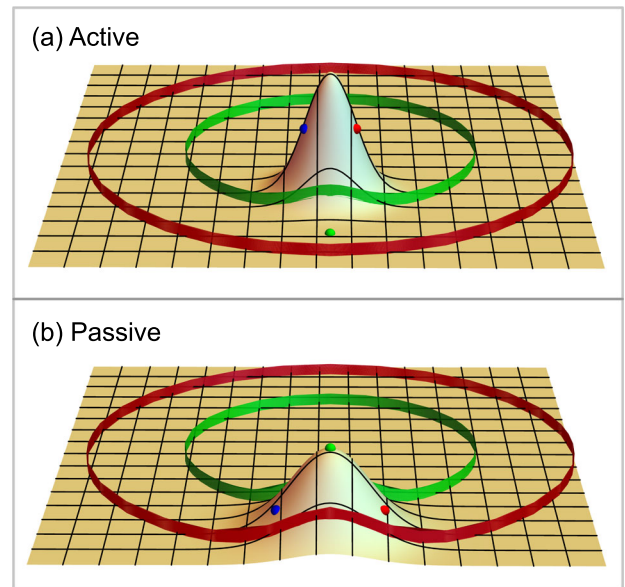

Fig. 5 | Sketch of the likelihood function on the Bloch disk. The Bloch disk represents an intersecting plane through the center of the Bloch sphere, e.g., the x-y-plane of the Bloch sphere. The red ribbon indicates the boundary of the Bloch disk, whereas the green ribbon indicates the boundary of the Bloch disk under the effect of a depolarizing channel. The green point is the true prepared state, which, due to the depolarizing readout noise, produces noisy measurement data. The red and blue points are arbitrary reference states in their active/passive noise representations. **a** An unmitigated likelihood function in the active representation of the noise (see eq. (6)). **b** An error-mitigated likelihood function in the passive representation of noise (see eq. (5)).

We compare (a) an unmitigated estimator in the active noise picture, to (b) the error-mitigated estimator in the passive noise picture. In the error-mitigated estimator the bias with respect to the true state (green point) is removed at a cost of higher variance of the likelihood function. We note that the bias due to the statistical fluctuations in the estimator remains, making it consistent with always-physical state estimators [47].

The increased variance can be combated by increasing the number of shots used for estimation. The additional number of shots required for the error-mitigated estimator to obtain the same confidence (variance of the likelihood) with respect to the true state as the unmitigated estimator with respect to the biased state is called the sampling overhead. The overhead relates to the distinguishability of quantum states and the data-processing inequality [46]. It tells us that the stronger the noise-induced distortion is, the larger the overhead. In QST, the sample overhead manifests itself as a shift in the infidelity curve. In Fig. 6 we have simulated varying depolarizing strengths and applied error mitigation. We see a clear shift as the noise strength increases. Note that such a shift would be present for any error mitigation schemes, including an inversion of the noise channel.

State preparation errors. A generic problem of QDT based methods is that they assume perfect preparation of calibration states, which is impossible in experiments. A common argument when using QDT is that the state preparation has a small error compared to the readout error itself. There are, however, methods developed to deal with the potential systematic error introduced by QDT, see e.g., refs. 41,48.

Another approach is to use the fact that preparation of the calibration states used in QDT only requires single qubit gates. If one has access to error estimates from single qubit gates, one could in principle also correct for state preparation errors in a similar manner to what is done in the protocol in this work (see e.g. randomized benchmarking [49] for gate characterization).

Despite our protocol requiring perfect state preparation, we have investigated noise sources that also affect state preparation, such as qubit
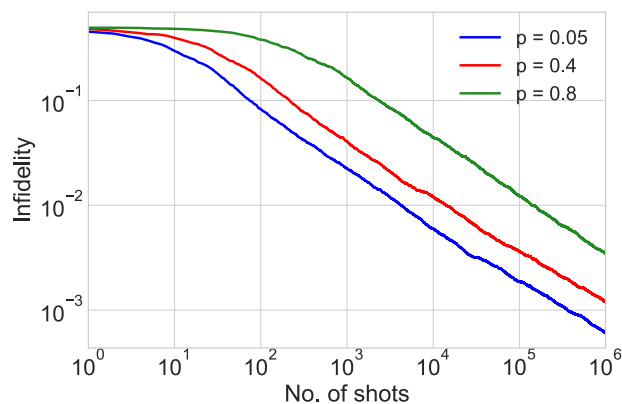
Fig. 6 | Reconstruction infidelity for mitigated simulated depolarizing noise. The various colored curves correspond to different noise strength p. The infidelity is averaged over 100 Haar-random pure states. The depolarizing channel is given in eq. (13).
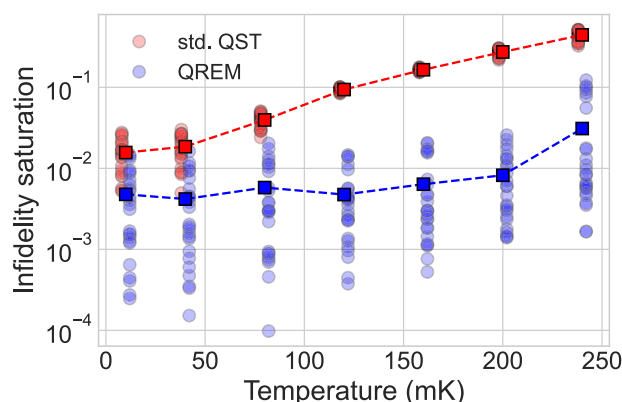

Fig. 7 | Quantum readout error mitigation (QREM) applied to increased qubit temperatures. Infidelity saturation refers to the last infidelity point measured over 240k single-shot measurements. The infidelity saturation for each individual run is plotted in translucent circles and shifted off center, to the left for standard quantum state tomography (std. QST) and to the right for QREM QST, for better visibility. The solid colored squares are the average infidelity saturation, connected by dotted lines for guidance. QREM seems to be successful, but the assumption of perfect state preparation is broken. Averaged over 25 Haar-random pure states with otherwise optimal experimental parameters.

detuning and increased T_1 and T_2 manipulation times. We will add to this investigation an experiment performed at higher qubit temperatures. We systematically increase the temperature of 10 mK up to over 200 mK, where the state in thermal equilibrium will have non-negligible contributions from the excited state. For example, a qubit temperature of 40 mK corresponds to an excited state population of 0.05 %, and a temperature of 120 mK corresponds to a 7.3 % excited state population. Hence, for our experiment without active feedback, it is not possible to reliably prepare a pure calibration state.

The results in Fig. 7 suggest that the mitigated QST is resilient to higher temperatures. The reason why we seemingly obtained a successful mitigation is that our benchmarking method does not allow us to distinguish state preparation errors from readout errors. Since our benchmarking method does not contain any additional gates between preparation and readout, the errors acquired in state preparation are interpreted as readout errors, and the protocol manages to mitigate the errors accordingly. We emphasize, however, that this is only strictly true if the combined effect of state preparation and readout errors can be viewed as a single effective error channel that is independent of the prepared state. While this is the case for finite qubit temperature, which can be modeled as depolarizing noise, it is not the case in general. In particular, for the cases of qubit detuning and

increased T_1 and T_2 time, the putative effective noise channel becomes dependent on the prepared state, leading to a deterioration of the mitigation efficiency. We discuss this in more detail in Supplementary Note 2.

Experimental drift. After performing QDT, our protocol assumes that the POVM stays fixed for the remainder of the measurement sequence. However, since experimental parameters drift over time, the physically realized POVM may change with respect to the POVM used for state reconstruction. Due to this drift, one would ideally recalibrate the POVM before each reconstruction. This entails a relatively large overhead, and is not feasible.

The drift present in our experiment was small and we decided not to warrant any additional corrective measures. This is not necessarily the case in general. In these cases, we propose to add an additional step of periodically performing drift measurements and recalibration to the measurement protocol (see "Measurement protocol" in Methods for more information).

Drift measurements amount to measuring a set of well-known states, e.g. Pauli states, and reconstruct their density matrix. If the reconstruction infidelity goes beyond an acceptable threshold $I(\rho^{\text{estm}}, \rho) \geq \epsilon$, one can perform a recalibration of the measurement device, i.e. repeat the QDT step. This would ensure that the accuracy of the protocol does not degrade.

Conclusion

We have presented a comprehensive scheme for readout error mitigation in the framework of quantum state tomography. It introduces quantum detector tomography as an additional calibration step, with a small overhead cost in the number of experimental samples. After calibration, our method is able to mitigate any errors acquired at readout. Furthermore, it does not require the inversion of any error channels and guarantees that the final state estimate is physical. Comparing to most previously discussed QREM methods, our protocol is able to correct beyond-classical errors. To confirm that such errors indeed make up a significant part of the errors in our experiment, we present a selection of reconstructed POVM elements from the experiment in Supplementary Note 3. The significant off-diagonal contributions in this analysis confirm that non-classical errors are always present and often on the same order of magnitude as the diagonal classical redistribution errors.

To probe the limits of readout error mitigation, we applied our protocol to a superconducting qubit system. We experimentally subjected the qubit to several noise sources and investigated the protocol's ability to mitigate them. We observed an improvement in the readout quality by decreasing infidelity by a factor of 5 to 30 depending on the type of readout noise. The protocol was particularly effective for lowered signal amplification and decreased resonator readout power compared to standard QST. We combined multiple noise sources in an experiment where conventional state reconstruction saturated early on, whereas our method was able to precisely reconstruct the quantum state. This opens up new possibilities for systems with noisy readouts where accurate knowledge of the quantum state is required. For noise sources which do not exclusively affect the readout stage, the protocol did not perform optimally, and we observed a constant ratio between the infidelity of mitigated and unmitigated state reconstruction.

Potential limitations of the scheme were investigated and we presented prescriptions on how to overcome them. Overall, we found that, by using readout error mitigation, one obtains accurate state estimates even under significantly degraded experimental conditions, making the readout more robust.

While being limited by exponential scaling in both memory and the required number of measurements, we expect the protocol to be feasible for up to 5–6 qubits if one replaces BME with MLE in the state reconstruction. This represents an interesting domain for error mitigation in quantum simulation as low-order correlators of a larger system offer relevant information such as correlation propagation[50] and phase estimation[51]. Recent developments in scalable approaches involving overlapping tomography[52,53] could provide a framework for a scalable version of this protocol to large qubit numbers, which we intend to
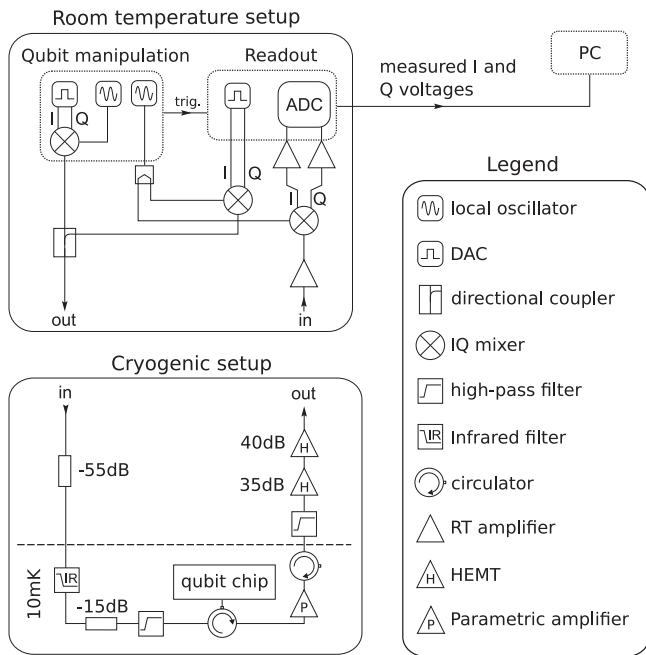
Fig. 8 | Microwave measurement setup. The local oscillator (LO) and phase-shifted pulsed IQ signals are mixed to provide the readout signal for the dispersive measurement of the qubit. A separate ω_{01} pulsed signal manipulates the qubit. The attenuators on the mixing chamber stage are carefully thermalized to mK temperatures to mitigate extra thermal noise. A wideband Josephson traveling-wave parametric amplifier (TWPA) and the two HEMT-based amplifiers provide quantum-limited readout.

explore. In future work, it could be interesting to perform a similar experiment on multiqubit systems. Implementing the protocol on a different qubit architecture with different sources of noise would be a topic of further interest. Another option of interest is to investigate adaptive noise-conscious strategies within this framework[54,55].

Methods
Experimental setup
We implemented the protocol introduced in the "Results and discussion" section on a fixed-frequency transmon qubit[56,57] with frequency $\omega_{01} = 6.3$ GHz coupled to a resonator with frequency $\omega_r = 8.5$ GHz in the dispersive readout scheme. Typical coherence times were observed to be $T_1 = 30 \pm 5\,\mu s$ and $T_2 = 28 \pm 6\,\mu s$. For more information about different types of superconducting qubits, in particular their operation and noise sources affecting them, we refer to ref. 58.

A schematic diagram of the setup is shown in Fig. 8. All measurements without an explicitly stated temperature were performed at 10 mK, using a dry dilution refrigerator. At room temperature we use a time-domain setup with superheterodyne mixing of readout pulses triggered by a high-frequency arbitrary wave generator (AWG) responsible for local oscillator readout and pulsed manipulation tones. Signal lines are attenuated by 70 dB distributed over various stages, as seen in Fig. 8. The measured signals are converted into complex numbers in the so-called IQ plane, where I and Q abbreviate in-phase and quadrature components of the integrated microwave signals. These two signals can be transformed equivalently into amplitude and phase. With optimized experimental calibrations and no induced noises, we can distinguish between two non-overlapping Gaussian distributions corresponding to the two possible outcomes of the measurement, with the qubit state being projected to either ground state $|0\rangle$ or excited state $|1\rangle$. For examples of how these distributions change in the presence of noise, see Supplementary Fig. 3 in Supplementary Note 4. Unless otherwise stated, we perform measurements at the best separation settings for the classification model. This usually corresponds to approx. 98%

distinguishability, meaning that 98% of shots can be correctly categorized as spin up or spin down.

Measurement protocol
or each noise source we study, we set up and execute our experiment as follows:

1. Estimation of π-pulse length (T_π) by Rabi-oscillations close to the qubit frequency.
2. Refined measurement of ω_{01} by a Ramsey experiment.
3. Measurement of T_π from a Rabi-measurement with updated ω_{01} from 2. by fitting a decaying sine function to the data.
4. Ground, $|0\rangle$, and excited states, $|1\rangle$, are measured with a single-shot readout. The location of the states in the IQ plane is learned by a supervised classification algorithm. For more information, see Supplementary Note 4.
5. QDT is performed by preparing each of the six Pauli states, and measuring them in the three bases σ_x, σ_y, and σ_z, which is described by the Pauli-6 POVM. The first two of these measurements are done as a combination of qubit rotation and subsequent σ_z readout. Quantum states are prepared using virtual Z-gates.
6. QST is performed and averaged over 25 random quantum states $U|0\rangle$, where U is a random unitary (Haar-random). Each state is measured in the three Pauli bases.

Performing this experiment for various strengths of the noise allows us to benchmark the ability of the protocol to mitigate the given noise source. For a schematic diagram of the measurement pipelines, see Fig. 9.

The outcome of two example experimental runs are shown in Fig. 10. Both standard and mitigated QST infidelities are extracted on a shot-by-shot basis. Typical features include a priori infidelity of roughly 0.5, with a rapid power-law decay, saturating at a given infidelity level for unmitigated QST. After this saturation is reached, further measurements will not improve the quantum state estimate because the measurements performed on the system are noisy. By using quantum readout error mitigation (labeled as QREM), we can significantly lower infidelity, enabling more precise state reconstruction. For all averaged experiments, the mitigated QST infidelities are consistently below the unmitigated QST infidelities.

Error sources
To probe the generality and reliability of our error-mitigation protocol, we artificially induce a set of noise sources, which introduce readout errors. In particular, we study:

1. Errors introduced by insufficient readout amplification: study through variation of the amplification of the parametric amplifier.
2. Errors introduced by low resonator photon number: study through variation of photon population in resonator by a variation of the readout amplitude.

In addition, we investigate two noise sources which manifest not only as readout errors, but also as state preparation errors. The implications of state preparation errors are discussed in "Protocol limitations" in Results and Discussion.

3. Errors introduced by energy (T_1) and phase relaxation (T_2): study through variation of the drive amplitude.
4. Errors introduced by qubit detuning: study through the variation of the applied manipulation pulse frequency.

Lastly, we combine multiple error sources to simulate a very noisy experiment, where conventional reconstruction methods fail.

Since the infidelity scaling is state-dependent, each experiment is averaged over 25 Haar-random states to get a reliable average performance (see "Explicit protocol realization" for more information). For QDT, we perform a total of $6 \times 3 \times 80,000$ (six states, three measurement bases) shots, for the following QST $25 \times 3 \times 80,000$ (25 states, three measurement bases, 80,000 shots each).

Infidelity

Quantum infidelity is well-suited as a figure of merit for successful quantum state reconstruction. We seek to minimize the infidelity, defined as

$$I(\rho, \sigma) = 1 - F(\rho, \sigma) = 1 - \left[\mathrm{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right) \right]^2, \qquad (14)$$
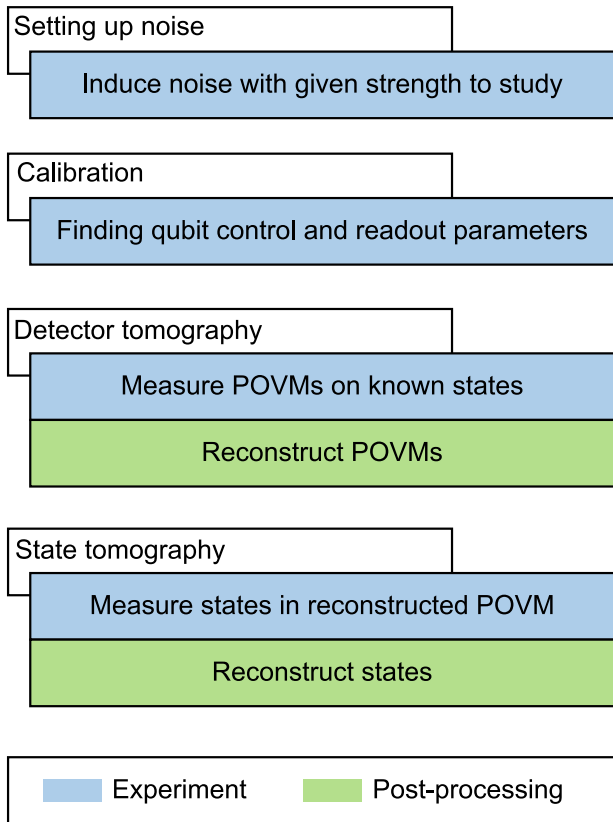

Setting up noise
    Induce noise with given strength to study

Calibration
    Finding qubit control and readout parameters

Detector tomography
    Measure POVMs on known states
    Reconstruct POVMs

State tomography
    Measure states in reconstructed POVM
    Reconstruct states

    Experiment   Post-processing

Fig. 9 | Schematic of the experimental pipeline of the protocol. In the first step, a given noise is induced with a specific strength. The experimental readout is calibrated for this noise. Detector tomography is performed, which reconstructs the noisy Pauli POVM $\{\tilde{M}_i\}$. Finally, quantum state tomography is executed, and reconstruction infidelity is evaluated and averaged over a set of randomly chosen target states.

where $F(\rho, \sigma)$ is the quantum fidelity[59]. We use the standard definition of $\sqrt{\rho}$, which is the square root of the eigenvalues of ρ in an eigendecomposition $\sqrt{\rho} = V\sqrt{D}V^{-1}$, where $\sqrt{D} = \mathrm{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n} \right)$. It acts as a pseudo-distance measure, and is asymptotically close to the Bures distance when $1 - F(\rho, \sigma) \ll 1$[60,61].

In this paper, we only consider the reconstruction of pure target states ρ, for which the infidelity simplifies to

$$I_{\mathrm{pure}}(\rho, \sigma) = 1 - \mathrm{Tr}(\rho\sigma). \qquad (15)$$

The infidelity between a sampled target state ρ and the reconstructed state σ is expected to decrease with $I(\rho, \sigma) \propto N^{-\alpha}$ where N is the number of shots performed, and α is an asymptotic scaling coefficient that depends on features of the estimation problem[55,62,63], such as the deviation from measuring in the target states eigenbasis, and the purity of the target state.

Explicit protocol realization

We present the explicit implementation of the protocol described in the Results and discussion. A pseudo-code outline of the whole measurement and readout error mitigation is presented in Supplementary Algorithm 1. For QDT, we use the maximum likelihood estimator described in ref. 44. We follow the prescription described in the subsection "Quantum Detector Tomography" in Results and discussion, and use all of the Pauli states as calibration states. The number of times each Pauli state is measured equals the maximal number of shots used for a single spin measurement in the state reconstruction, such that the dominant source of shot noise is not QDT. For QST, we use a BME, in particular we use the implementation described in ref. 64. The bank particles are generated from the Hilbert-Schmidt measure[65]. This QST is also equipped with adaptive measurement strategies, which is a possible future extension to the current protocol.

For both QDT and QST, we use the Pauli-6 POVM, which is a static measurement strategy. It is known that the asymptotic scaling of such strategies depends on the proximity of the true state to one of the projective measurements[66,67]. To counteract this, an average over Haar-random states[68] is performed to get a robust performance estimate. In this way, we get the expected performance given that no prior information about the true state is available. We emphasize that for all random state reconstructions, the same QDT calibration is used.

Data availability

The data used for Figs. 4–7, 9, 10 and Supplementary Fig. 1 can be found on the KIT research data repository: https://radar.kit.edu/radar/en/dataset/RNbuograoVUFQNBB. For more information on the dataset, see
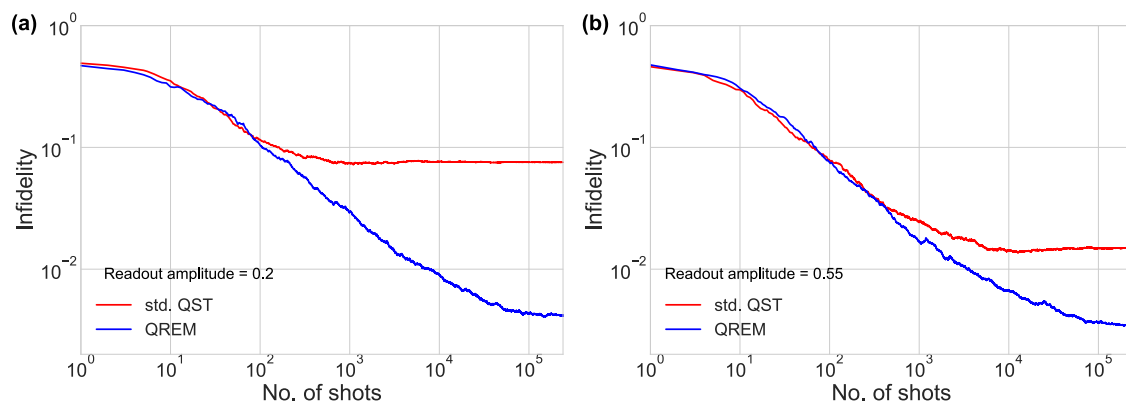

Fig. 10 | Example run of quantum state reconstruction using a Bayesian mean estimator. Mean quantum reconstruction infidelity is plotted as a function of number of shots for **a** optimal readout powers and **b** weak readout. The red lines represent unmitigated, standard quantum state tomography (std. QST), and the blue lines are obtained with quantum readout error mitigation (QREM). Each curve is an average of over 25 Haar-random states. For an indication of the spread of each Haar-random state, see Supplementary Fig. 1. QREM reaches a lower infidelity saturation value than standard reconstruction in both cases. At low readout power, error-mitigated QST can still reconstruct the state with similar accuracy, while standard QST saturates at a significantly higher value.

Supplementary Note 5. The remaining experimental data were available upon request from A. Di Giovanni.

Code availability

The code developed for this project is available on GitHub: https://github.com/AdrianAasen/EMQST. A short tutorial notebook is provided with examples of how to run the software. It can be interfaced with an experiment, or run as a simulation.

References

1. Bennett, C. H. & Brassard, G. Quantum cryptography: public key distribution and coin tossing. *Theor. Comput. Sci.* **560**, 7–11 (2014).
2. Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**, 1484–1509 (1997).
3. Ma, H., Govoni, M. & Galli, G. Quantum simulations of materials on near-term quantum computers. *npj Comput. Mater.* https://doi.org/10.1038/s41524-020-00353-z (2020).
4. Pyrkov, A. et al. Quantum computing for near-term applications in generative chemistry and drug discovery. *Drug Discov. Today* **28**, 103675 (2023).
5. Zinner, M. et al. Toward the institutionalization of quantum computing in pharmaceutical research. *Drug Discov. Today* **27**, 378–383 (2022).
6. Orús, R., Mugel, S. & Lizaso, E. Quantum computing for finance: overview and prospects. *Rev. Phys.* **4**, 100028 (2019).
7. Egger, D. J., Gutierrez, R. G., Mestre, J. C. & Woerner, S. Credit risk analysis using quantum computers. *IEEE Trans. Comput.* **70**, 2136–2145 (2021).
8. Dri, E., Giusto, E., Aita, A. & Montrucchio, B. Towards practical quantum credit risk analysis. *J. Phys. Conf. Ser.* **2416**, 012002 (2022).
9. Harwood, S. et al. Formulating and solving routing problems on quantum computers. *IEEE Trans. Quantum Eng.* **2**, 1–17 (2021).
10. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
11. Lisenfeld, J., Bilmes, A. & Ustinov, A. V. Enhancing the coherence of superconducting quantum bits with electric fields. *npj Quantum Inform* https://doi.org/10.1038/s41534-023-00678-9 (2023).
12. Cai, Z. et al. Quantum error mitigation. *Rev. Mod. Phys.* **95**, 045005 (2023).
13. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2012).
14. Krinner, S. et al. Realizing repeated quantum error correction in a distance-three surface code. *Nature* **605**, 669–674 (2022).
15. Kitaev, A. Fault-tolerant quantum computation by anyons. *Ann. Phys.* **303**, 2–30 (2003).
16. Dennis, E., Kitaev, A., Landahl, A. & Preskill, J. Topological quantum memory. *J. Math. Phys.* **43**, 4452–4505 (2002).
17. Raussendorf, R. & Harrington, J. Fault-tolerant quantum computation with high threshold in two dimensions. *Phys. Rev. Lett.* https://doi.org/10.1103/physrevlett.98.190504 (2007).
18. Greenbaum, D. Introduction to quantum gate set tomography. Preprint at arXiv https://arxiv.org/abs/1509.02921 (2015).
19. Geller, M. R. & Sun, M. Toward efficient correction of multiqubit measurement errors: pair correlation method. *Quantum Sci. Technol.* **6**, 025009 (2021).
20. Walter, T. et al. Rapid high-fidelity single-shot dispersive readout of superconducting qubits. *Phys. Rev. Appl.* https://doi.org/10.1103/physrevapplied.7.054020 (2017).
21. Qin, D., Xu, X. & Li, Y. An overview of quantum error mitigation formulas. *Chin. Phys. B* **31**, 090306 (2022).
22. Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
23. Nachman, B., Urbanek, M., de Jong, W. A. & Bauer, C. W. Unfolding quantum computer readout noise. *npj Quantum Inform.* https://doi.org/10.1038/s41534-020-00309-7 (2020).
24. Pokharel, B., Srinivasan, S., Quiroz, G. & Boots, B. Scalable measurement error mitigation via iterative bayesian unfolding. *Phys. Rev. Research* **6**, 013187 (2024).
25. Geller, M. R. Rigorous measurement error correction. *Quantum Sci. Technol.* **5**, 03LT01 (2020).
26. Bravyi, S., Sheldon, S., Kandala, A., Mckay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Phys. Rev.* https://doi.org/10.1103/physreva.103.042605 (2021).
27. Kwon, H. & Bae, J. A hybrid quantum-classical approach to mitigating measurement errors in quantum algorithms. *IEEE Trans. Comput.* **70**, 1401–1411 (2021).
28. Maciejewski, F. B., Zimborás, Z. & Oszmaniec, M. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum* **4**, 257 (2020).
29. Kim, Y. et al. Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505 (2023).
30. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Revi. Lett.* https://doi.org/10.1103/physrevlett.119.180509 (2017).
31. Li, Y. & Benjamin, S. C. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X* https://doi.org/10.1103/physrevx.7.021050 (2017).
32. Chen, Y., Farahzad, M., Yoo, S. & Wei, T.-C. Detector tomography on IBM quantum computers and mitigation of an imperfect measurement. *Phys. Rev. A* https://doi.org/10.1103/physreva.100.052315 (2019).
33. Smithey, D. T., Beck, M., Raymer, M. G. & Faridani, A. Measurement of the wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum. *Phys. Rev. Lett.* **70**, 1244–1247 (1993).
34. xi Liu, Y., Wei, L. F. & Nori, F. Tomographic measurements on superconducting qubit states. *Phys. Rev. B* https://doi.org/10.1103/physrevb.72.014547 (2005).
35. Motka, L., Paúr, M., R^eháček, J., Hradil, Z. & Sánchez-Soto, L. L. Efficient tomography with unknown detectors. *Quantum Sci. Technol.* **2**, 035003 (2017).
36. Motka, L., Paúr, M., Řeháček, J., Hradil, Z. & Sánchez-Soto, L. L. When quantum state tomography benefits from willful ignorance. *N. J. Phys.* **23**, 073033 (2021).
37. Ramadhani, S., Rehman, J. U. & Shin, H. Quantum error mitigation for quantum state tomography. *IEEE Access* **9**, 107955–107964 (2021).
38. Sakurai, J. J. & Napolitano, J. *Modern Quantum Mechanics* (Cambridge Univ. Press, 2017).
39. Paris, M. & Reháček, J. (eds.) *Quantum State Estimation* (Springer, 2004).
40. Blume-Kohout, R. Optimal, reliable estimation of quantum states. *N. J. Phys.* **12**, 043034 (2010).
41. Gebhart, V. et al. Learning quantum systems. *Nat. Rev. Phys.* **5**, 141–156 (2023).
42. Lvovsky, A. I. Iterative maximum-likelihood reconstruction in quantum homodyne tomography. *J. Opt. B Quantum Semiclassical Opt.* **6**, S556–S559 (2004).
43. Lundeen, J. S. et al. Tomography of quantum detectors. *Nat. Phys.* **5**, 27–30 (2008).
44. Fiurášek, J. Maximum-likelihood estimation of quantum measurement. *Phys. Rev. A* https://doi.org/10.1103/physreva.64.024102 (2001).
45. Meißner, S. M., Seiler, A., Lisenfeld, J., Ustinov, A. V. & Weiss, G. Probing individual tunneling fluctuators with coherently controlled tunneling systems. *Phys. Rev. B* https://doi.org/10.1103/physrevb.97.180505 (2018).

46. Takagi, R., Endo, S., Minagawa, S. & Gu, M. Fundamental limits of quantum error mitigation. *npj Quantum Inform* https://doi.org/10.1038/s41534-022-00618-z (2022).

47. Schwemmer, C. et al. Systematic errors in current quantum state tomography tools. *Phys. Rev. Lett.* https://doi.org/10.1103/PhysRevLett.114.080403 (2015).

48. Zhang, A. et al. Experimental self-characterization of quantum measurements. *Phys. Rev. Lett.* https://doi.org/10.1103/physrevlett.124.040402 (2020).

49. Knill, E. et al. Randomized benchmarking of quantum gates. *Phys. Rev. A* https://doi.org/10.1103/physreva.77.012307 (2008).

50. Richerme, P. et al. Non-local propagation of correlations in quantum systems with long-range interactions. *Nature* **511**, 198–201 (2014).

51. Ebadi, S. et al. Quantum phases of matter on a 256-atom programmable quantum simulator. *Nature* **595**, 227–232 (2021).

52. Cotler, J. & Wilczek, F. Quantum overlapping tomography. *Phys. Rev. Lett.* https://doi.org/10.1103/PhysRevLett.124.100401 (2020).

53. Tuziemski, J. et al. Efficient reconstruction, benchmarking and validation of cross-talk models in readout noise in near-term quantum devices. Preprint at https://arxiv.org/abs/2311.10661 (2023).

54. Ivanova-Rohling, V. N., Rohling, N. & Burkard, G. Optimal quantum state tomography with noisy gates. *EPJ Quantum Technol.* https://doi.org/10.1140/epjqt/s40507-023-00181-2 (2023).

55. Huszár, F. & Houlsby, N. M. T. Adaptive Bayesian quantum tomography. *Phys. Rev. A* https://doi.org/10.1103/physreva.85.052120 (2012).

56. Koch, J. et al. Charge-insensitive qubit design derived from the cooper pair box. *Phys. Rev. A* https://doi.org/10.1103/physreva.76.042319, https://doi.org/10.1103/physreva.76.042319 (2007).

57. Kjaergaard, M. et al. Superconducting qubits: current state of play. *Ann. Rev. Condens. Matter Phys.* **11**, 369–395 (2020).

58. Krantz, P. et al. A quantum engineer's guide to superconducting qubits. *Appl. Phys. Rev.* https://doi.org/10.1063/1.5089550 (2019).

59. Jozsa, R. Fidelity for mixed quantum states. *J. Modern Opt.* **41**, 2315–2323 (1994).

60. Hübner, M. Explicit computation of the bures distance for density matrices. *Phys. Lett. A* **163**, 239–242 (1992).

61. Uhlmann, A. The "transition probability" in the state space of a *-algebra. *Rep. Math. Phys.* **9**, 273–279 (1976).

62. Massar, S. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).

63. Bagan, E., Ballester, M. A., Gill, R. D., Monras, A. & Muñoz-Tapia, R. Optimal full estimation of qubit mixed states. *Phys. Rev. A* https://doi.org/10.1103/physreva.73.032301 (2006).

64. Struchalin, G. I. et al. Experimental adaptive quantum tomography of two-qubit states. *Phys. Rev. A* https://doi.org/10.1103/physreva.93.012103 (2016).

65. Zyczkowski, K. & Sommers, H.-J. Induced measures in the space of mixed quantum states. *J. Phys. A Math. Gen.* **34**, 7111–7125 (2001).

66. Bagan, E., Ballester, M. A., Gill, R. D., Muñoz-Tapia, R. & Romero-Isart, O. Separable measurement estimation of density matrices and its fidelity gap with collective protocols. *Phys. Rev. Lett.* https://doi.org/10.1103/physrevlett.97.130501 (2006).

67. Struchalin, G. I., Kovlakov, E. V., Straupe, S. S. & Kulik, S. P. Adaptive quantum tomography of high-dimensional bipartite systems. *Phys. Rev. A* https://doi.org/10.1103/physreva.98.032330 (2018).

68. Mezzadri, F. How to generate random matrices from the classical compact groups. *Notices of the American Mathematical Society, vol. 54, no. 5, pp. 592 - 604*. Preprint at https://arxiv.org/abs/math-ph/0609050 (2007).

Author contributions
Development of the protocol and software was done by A. Aasen with supervision from M. Gärttner. The experimental realization and data generation was done by A. Di Giovanni with supervision from H. Rotzinger and A. Ustinov. A. Aasen and A. Di Giovanni prepared the draft for the manuscript. All authors contributed to the finalization of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-024-01790-8.

Correspondence and requests for materials should be addressed to Martin Gärttner.

Peer review information *Communications Physics* thanks Jacob Blumoff and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.