# Limitations for adaptive quantum state tomography in the presence of detector noise

Adrian Skasberg Aasen [1,2,*] and Martin Gärttner [1,†]

[1] *Institut für Festkörpertheorie und-optik, Friedrich-Schiller-Universität Jena, Max-Wien-Platz 1, 07743 Jena, Germany*

[2] *Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany*

(Received 8 January 2026; accepted 12 June 2026; published 6 July 2026)

Assumption-free reconstruction of quantum states from measurements is essential for benchmarking and certifying quantum devices, but it remains difficult due to the extensive measurement statistics and experimental resources it demands. An approach to alleviating these demands is provided by adaptive measurement strategies, which can yield up to a quadratic improvement in reconstruction accuracy for pure states by dynamically optimizing measurement settings during data acquisition. A key open question is whether these asymptotic advantages remain in realistic experiments, where readout is inevitably noisy. In this work, we analyze the impact of readout noise on adaptive quantum state tomography with readout-error mitigation, focusing on the challenging regime of reconstructing pure states using mixed-state estimators. Using analytical arguments based on Fisher information optimization and extensive numerical simulations using Bayesian inference, we show that any nonzero readout noise eliminates the asymptotic quadratic scaling advantage of adaptive strategies. We numerically investigate the behavior for finite measurement statistics for single- and two-qubit systems with exact readout-error mitigation and find a gradual transition from ideal to suboptimal scaling. We furthermore investigate scenarios where detector tomography is performed with a limited number of state copies for calibration, showing that insufficient detector characterization leads to estimator bias and limited reconstruction accuracy. Although our result imposes an upper bound on the reconstruction accuracy that can be achieved with adaptive strategies, we observe numerically a constant-factor gain in reconstruction accuracy, which becomes larger as the readout noise decreases. This indicates potential practical benefits in using adaptive measurement strategies in well-calibrated experiments.

DOI: 10.1103/3q13-xpc8

## I. INTRODUCTION

Quantum state tomography is the most measurement intensive characterization and benchmarking tool for quantum systems. Consequently, extensive research efforts have been devoted to identifying optimal measurement strategies that use fewer state copies to achieve the desired reconstruction accuracy. Collective measurement, where $N$ copies of a state are jointly measured, can achieve optimal asymptotic scaling for reconstruction infidelity of $1/N$ for arbitrary quantum states [1–4]. Among static (nonadaptive) copy-by-copy measurement schemes, one near-optimal approach is to measure mutually unbiased bases [5–7]. Static approaches are particularly effective for strongly mixed states, where no eigenvalues are close to zero, generally achieving an asymptotic infidelity that scales as $1/N$ [8]. Throughout this work, we focus exclusively on copy-by-copy measurement schemes, and therefore we will use the terms state copies and measurements interchangeably.

Pure and close-to-pure states are particularly difficult to reconstruct because they are close to the boundary of the estimation domain [3,6,9,10]. These states are of practical concern, as most quantum experiments intend to produce pure states. Compared to strongly mixed states, pure states show a quadratically poorer asymptotic infidelity scaling for static copy-by-copy measurement schemes, going from $1/N \to 1/\sqrt{N}$ [3,8–10]. Adaptive measurement strategies mend this discrepancy by introducing classical communication between measurements, which allows for higher information extraction per state copy. If projective measurements can be performed on multiqubit entangled bases, optimal reconstruction infidelity is recovered with a single adaptive step in the middle of the measurement process [6,9]. If restricted to more practical separable projective measurement, continuous adaptive optimizations are generally required to achieve the best asymptotic scaling for arbitrary pure states [9,11–14]. The current state-of-the-art adaptive quantum state tomography offers a comprehensive selection of practical and numerically fast estimation schemes allowing precise estimation of mixed and pure quantum states [15–26]. Adaptive quantum state reconstruction strategies have been adopted in various experimental qubit platforms, including photonic systems [9,12,18,27–29] and superconducting architectures [30].

In standard state reconstruction protocols, detectors are typically assumed to be ideal, which makes it straightforward to identify the optimal measurements. In real experiments, however, detectors are noisy, which both restricts the

*Contact author: adrian.aasen@uni-jena.de

†Contact author: martin.gaerttner@uni-jena.de

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

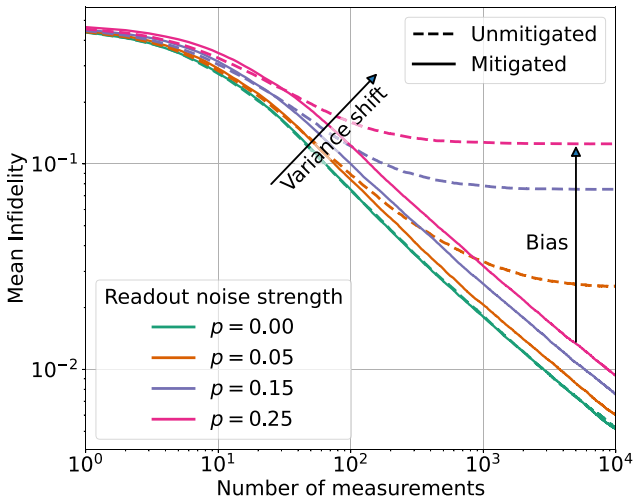



FIG. 1. Simulated single-qubit state reconstruction with and without readout-error mitigation for various levels of depolarizing noise $p$. Without readout-error mitigation (dashed lines), the reconstruction infidelity eventually saturates at a bias value set by the distance to the effectively measured noisy state. When exact readout-error mitigation is applied (solid lines) the mean infidelity curves keep decreasing but shift further to the right with increasing readout noise strength. Each curve is averaged over 2000 Haar-random pure states. Both axes are plotted on a logarithmic scale.

reachable reconstruction infidelity and introduces a noise-induced bias in the estimator. With readout noise, the effective measured state undergoes a transformation through a readout noise channel, denoted $\tilde{\rho} = \mathcal{E}(\rho)$. Without modifications to the estimator, the infidelity will at some point level off at a biased value, determined by the distance between the target state and the effectively measured state (see Fig. 1). This bias can be removed by integrating readout-error mitigation into the state reconstruction process [31–33], often by performing detector tomography [34,35]. In general, error mitigation methods trade estimator bias for increased estimator variance [36], where stronger noise typically induces a higher variance penalty, as illustrated in Fig. 1. If the noisy detection process can be characterized exactly, the bias vanishes, and only an asymptotically decaying variance remains. For static, copy-by-copy measurement strategies, pure and close-to-pure states achieve the same asymptotic $1/\sqrt{N}$ infidelity scaling with and without readout noise in the system [31,37]. While these results may suggest that readout-error-mitigated measurement strategies are robust against readout noise, we find that introducing readout noise fundamentally alters the asymptotic behavior of adaptive measurement strategies. In this work, we demonstrate that even an arbitrarily small amount of readout noise eliminates the asymptotic scaling advantage for adaptive strategies. Furthermore, we observe that the bias-variance tradeoff also appears in adaptive measurement strategies.

Other works have investigated readout noise in the context of optimal quantum state tomography. The decrease in reconstruction infidelity for static measurement schemes caused by noise was investigated in Refs. [38,39]. In Ref. [40], the authors introduced a new optimal static measurement approach that considers readout noise modeled as a noisy unitary gate applied before an ideal readout. In Ref. [30] the authors perform noise-aware adaptive measurement optimization with lowered state distinguishability in superconducting circuits, but without any readout-error mitigation. Further works have looked at including noise in optimal adaptive measurement strategies for parameter estimation tasks [41,42].

In this work, we investigate three different aspects of how noise impacts adaptive measurement strategies. First, in Sec. III, we present two different adaptive measurement strategies and derive the conditions required for a measurement strategy to be optimal. Then, in Sec. IV, we show that imperfect measurements place a nontrivial upper bound on the recoverable infidelity of adaptive measurement strategies for pure and close-to-pure states using Fisher information optimization [9,10]. This bound causes optimal asymptotic infidelity scaling to be lost with any amount of readout noise in the system. Secondly, in Sec. V, we verify this claim for single- and two-qubit systems numerically by using Bayesian mean estimation (BME) with minimization of the Bayesian posterior entropy [11]. Beyond confirming the loss of optimal scaling, the simulations reveal a bias-variance tradeoff similar to that seen in static measurement schemes. We examine the transient behavior and observe a gradual crossover from optimal to suboptimal scaling. Despite the fundamental limitations encountered in the asymptotic regime, our results show that adaptive strategies consistently provide a multiplicative improvement in reconstruction accuracy. Thirdly, in Sec. VI, we numerically investigate scenarios in which a limited number of measurements can be used for detector tomography. We find, as expected, that when very few state copies are used for detector tomography compared to state tomography the infidelity curve saturates to a biased value, similar to unmitigated state tomography in Fig. 1. For the type of noise simulated in this work, detector tomography requires at least five times as many state copies as were used for a single state reconstruction, although this factor typically needs to be adjusted individually for each experiment [37].

## II. PRELIMINARIES

### A. Generalized quantum measurements

A generalized quantum measurement, described by a positive operator-valued measure (POVM), consists of Hermitian operators $\mathbf{M} = \{M_\gamma\}$ that satisfy the following properties:

$$M_\gamma \geqslant 0, \quad M_\gamma^\dagger = M_\gamma, \quad \text{and} \quad \sum_\gamma M_\gamma = \mathbb{1}. \tag{1}$$

Each POVM element $M_\gamma$ is associated with an outcome $\gamma$ of a measurement process. The Born rule provides the probabilities for the different possible outcomes when measuring a quantum state $\rho$:

$$\mathrm{Tr}(\rho M_\gamma) = \langle M_\gamma \rangle = p_\gamma. \tag{2}$$

One benefit of working with POVMs rather than just projective measurements is that they allow us to capture the process of noise at readout.

In this work, we consider single- and two-qubit states, where we use the Pauli-6 POVM unless specified otherwise:

$$\mathcal{M}_{\text{Pauli-6}} = \left\{ \tfrac{1}{3}|0_i\rangle\langle 0_i|, \tfrac{1}{3}|1_i\rangle\langle 1_i| \right\} \quad \text{for } i \in \{x, y, z\}, \tag{3}$$

where $|0_i\rangle$ and $|1_i\rangle$ are the two eigenstates of the corresponding Pauli operators. The two-qubit Pauli-6 POVM is created by taking the tensor product of all possible combinations of the elements of two single-qubit Pauli-6 POVMs.

### B. Readout-error-mitigated state tomography

In order to reliably reconstruct quantum states despite the presence of readout noise in the system, it is necessary to perform readout-error mitigation. This work uses the protocol presented in Ref. [31] that addresses the mitigation of beyond-classical readout errors in quantum state reconstruction. The basic outline of the approach is as follows: quantum detector tomography is used as a calibration step, where the actual noisy POVM $\mathcal{M} = \{M_\gamma\}$ implemented by the measurement device is reconstructed [43]. The reconstructed POVM is then integrated into the likelihood function

$$\mathcal{L}(\rho) \propto \Pi_\gamma \mathrm{Tr}(\rho M_\gamma)^{n_\gamma}, \tag{4}$$

where $n_\gamma$ is the number of occurrences of outcome $\gamma$. Using a Bayesian mean estimator [44] or a maximum likelihood estimator [45], a most probable state can be found that is guaranteed to be physical.

In Secs. IV and V, we assume that the noisy POVM is determined exactly, i.e. that we have access to the exact measurement operators implemented by the detector for readout-error mitigation. This is necessary to study the asymptotic behavior of adaptive measurement strategies but is not experimentally realistic. To amend this, we include the detector tomography step in the numerical simulations in Sec. VI.

### C. Depolarizing noise

Depolarizing noise represents a general and typical type of readout noise, and it is the main form of noise that we examine in this work. It is characterized by a lowered state distinguishability, and can be viewed as a redistribution of measurement outcomes. The depolarizing channel acting on some operator $A$ in a $n$-qubit Hilbert space with strength $p$ is defined as [46,47]

$$\begin{aligned}\mathcal{E}^p_{\mathrm{depol}}(A) &= (1-p)A + p\frac{\mathbb{1}}{2^n}\mathrm{Tr}(A) \\ &\overset{n=1}{=} \left(1-\frac{3p}{4}\right)A + \frac{p}{4}\sum_{i=1}^{3} K_i A K_i^\dagger \end{aligned} \tag{5}$$

where $(K_1, K_2, K_3) = (X, Y, Z)$ are the Pauli matrices. The second form in Eq. (5) is the operator sum form for a single-qubit depolarizing channel. The operator sum form can be generalized to $n$-qubit channels [47,48].

### D. Quantum infidelity

The reconstruction quality metric we use is the quantum infidelity, defined as

$$I(\rho, \sigma) = 1 - F(\rho, \sigma), \tag{6}$$

where $F(\rho, \sigma)$ is the quantum fidelity [49],

$$F(\rho, \sigma) = \left[\mathrm{Tr}\left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right)\right]^2, \tag{7}$$

where $\sqrt{\rho}$ should be understood as the square root of the eigenvalues $\lambda_i$ of $\rho$ in an eigendecomposition $\sqrt{\rho} = V\sqrt{D}V^{-1}$, where $\sqrt{D} = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$.

The quantum fidelity simplifies in some important cases, most notably for when one state is pure,

$$F_{\mathrm{Pure}}(|\psi_\rho\rangle, \sigma) = \langle\psi_\rho|\sigma|\psi_\rho\rangle = \mathrm{Tr}(\rho\sigma), \tag{8}$$

and for qubit states

$$F(\rho, \sigma) = \mathrm{Tr}(\rho\sigma) + 2\sqrt{\det(\rho)\det(\sigma)}. \tag{9}$$

### E. Fisher information matrix

The Fisher information is formally defined as the variance of the score function

$$\mathcal{I}(\theta) = \mathrm{Var}[s(\theta)], \tag{10}$$

where the score function is

$$s(\theta) = \frac{\partial \ln[\mathcal{L}(\theta)]}{\partial\theta}. \tag{11}$$

The Fisher information quantifies how sensitive the likelihood function is to changes in the parameter $\theta$. A high Fisher information indicates that small variations in the measurement data lead to large changes in the inferred parameter, while a low Fisher information implies that the parameter remains relatively insensitive to additional measurements. For multiple parameters $\{\theta_i\}$ the Fisher information generalizes to a matrix with two equivalent formulations:

$$\begin{aligned}[\mathcal{I}(\theta)]_{ij} &= \mathbb{E}\left[\frac{\partial \ln[\mathcal{L}(\theta)]}{\partial\theta_i}\frac{\partial \ln[\mathcal{L}(\theta)]}{\partial\theta_j}\right] \\ &= -\mathbb{E}\left[\frac{\partial^2 \ln[\mathcal{L}(\theta)]}{\partial\theta_i\partial\theta_j}\right], \end{aligned} \tag{12}$$

where the $i$, $j$ entry quantifies the information shared between the parameters $\theta_i$ and $\theta_j$.

## III. METHODS

### A. Defining the estimation problem

When estimating quantum states, there are two sources of error that contribute to the infidelity, the bias of the estimator and the estimator variance. The bias is intrinsic to the estimator and remains constant regardless of the number of measurements performed. In contrast, the variance decreases as the number of measurements increases. We aim to investigate how adaptive measurement strategies influence the variance of the estimator, and, except in Sec. VI, we assume that any estimator bias has been removed using exact readout-error mitigation.

Most standard measurement strategies have a worst-case infidelity scaling of $\propto 1/\sqrt{N}$ [6,10]. According to the Gill-Massar bound, there is an upper limit to the achievable infidelity scaling of any estimator, which is proportional to $1/N$ [50]. Consequently, an estimator is considered asymptotically optimal if it obtains the scaling $1/N$ as $N$ approaches infinity, despite the possibility of further improvements to the scaling coefficient being possible.

In most experiments, the target states are pure states, and it would therefore seem natural to use a pure-state estimator.

Such a rank-appropriate estimator would recover optimal $1/N$ infidelity scaling without any adaptive measurements at all [9,10,51]. The suboptimal $1/\sqrt{N}$ scaling, as will be elaborated in the following sections, arises from a mismatch between the rank of the estimator and that of the target state, and once the two are matched, even static measurements achieve $1/N$. The question of recovering optimal scaling through adaptivity is thus only meaningful for high-rank mixed-state estimators, which overparametrize the target and place its pure-state component on the boundary of the estimator domain (see Appendix A for an intuitive explanation of the resulting degradation). There are, in addition, two major drawbacks in restricting a state estimator to only pure states. First, the space of pure states is not convex, and it is not guaranteed that the likelihood function is monoidal in this space [44]. Secondly, since it is impossible to prepare perfect pure states in any experimental setup, the prepared state is always mixed and would be outside the estimator support. For these reasons our focus will be exclusively on mixed-state estimators. To ensure that the estimators perform optimally on average, the target states will be randomly chosen from a set of Haar-random pure states. In Secs. V and VI we restrict our numerical analysis to pure states only, ensuring that the $1/\sqrt{N}$ scaling we observe at large $N$ is due to detector noise alone. For close-to-pure but mixed target states, the nonadaptive infidelity would eventually cross over to $1/N$ scaling once $N$ is large enough to resolve the target states' finite distance from the pure-state boundary, which would complicate the comparison.

In summary, our investigation mainly focuses on mixed-state estimators with pure target states using exact readout-error mitigation. The quality measure is the asymptotic scaling of the quantum infidelity in the number of measurements. The quantum infidelity is averaged over Haar-random target states.

### B. Adaptive measurement strategies

We examine two kinds of adaptive strategies, those that update the measurement setting a single time and those that update it multiple times.

#### 1. Two-step strategies

Single update methods are the more economical of the two approaches, since updating measurement settings can be very slow compared to continuing measurements in the same setting. If the target state is known in advance, the best strategy is to perform a projective measurement in the eigenbasis of the state. This measurement scheme can, in principle, surpass the Gill-Massar bound [18]. Nonetheless, such an estimator is of limited practical relevance once the target state is already known. A state-agnostic version of this strategy can be created by splitting the estimation process into two steps [6,13,19,50].

(1) Perform an initial-state estimate $\rho_{\text{init}}$ based on an informationally complete measurement, such as the $\mathcal{M}_{\text{Pauli-6}}$ measurements.

(2) Find the eigenstate $|\lambda_1\rangle$ with the largest eigenvalue $\lambda_1$ of $\rho_{\text{init}}$, and perform the remaining measurements using the generalized measurement $U\mathcal{M}_{\text{Pauli-6}}U^\dagger$, where $U$ is a unitary rotation matrix that rotates the all-zero qubit state to this eigenstate of $\rho_{\text{init}}$, $U|0\ldots0\rangle = |\lambda_1\rangle$.

The best way to divide the total measurement budget between the two steps has been shown to be to allocate half to each step [6]. It was also found to be sensitive to the prior used to initialize the estimator [13].

Although these methods were found to be optimal, the eigenbasis of multiqubit systems is very often entangled, which typically causes additional overhead to implement experimentally. The underlying mechanism was found to be related to the singular values of the Fisher information matrix [9,10], allowing Ref. [9] to develop a multistep strategy that uses only the tensor products of measurements on single qubits. This approach is optimal for some multiqubit states, but it is not generally optimal for all states.

#### 2. Multistep strategies

Multistep strategies are a form of adaptive experimental design, where the estimator continuously updates the measurement strategy based on prior measurement outcomes $D$. A prominent example is Bayesian adaptive tomography, where Bayesian inference is used to incorporate the measurement outcomes into a continuously updated posterior distribution $\pi_f(\rho)$. A new measurement setting $\alpha$ is chosen by maximizing the expected value of a utility function $U(\alpha, \pi_f(\rho))$ based on the current posterior distribution $\pi_f(\rho)$ [13]:

$$\alpha_n = \arg\max_\alpha \sum_\gamma \int d\rho\, \pi_f(\rho)\mathrm{Tr}\big(\rho M_\gamma^\alpha\big)U(\alpha, \pi_f(\rho)). \tag{13}$$

There are two common approaches to choosing the utility function, optimizing for a specific parameter, e.g. the infidelity, or maximizing information gain. References [11,52] demonstrated that using the information gain approach significantly simplifies the computation of the optimization objective in Eq. (13). The optimization can be rewritten as

$$\alpha_n = \arg\max_\alpha \left[ H\left[\int d\rho\, \pi_f(\rho)\mathrm{Tr}\big(\rho M_\gamma^\alpha\big)\right] - \int d\rho\, \pi_f(\rho) H\big[\mathrm{Tr}\big(\rho M_\gamma^\alpha\big)\big]\right], \tag{14}$$

where $H[p_\gamma] = -\sum_\gamma p_\gamma \log(p_\gamma)$ denotes the Shannon entropy of the discrete probability distribution $p$. The advantage of this reformulation is that it relies solely on the current posterior distribution $\pi_f(\rho)$, and requires only the computation of discrete Shannon entropies. The terms in Eq. (14) can also be understood intuitively: the first term promotes measurements along the measurement axis with the highest uncertainty (high entropy), while the second term favors measurements where the most likely states have low uncertainty (low entropy). Combined, the optimization seeks the measurement setting $\alpha$ that targets regions of high uncertainty in the outcome distribution while remaining consistent with the most probable candidate states. In Fig. 2, we show that the information-gain strategy achieves optimal asymptotic infidelity scaling for both single- and two-qubit state reconstruction. Notably, in the two-qubit case optimal scaling is reached using only separable measurement operators, whereas the two-step approach can generally not achieve optimal scaling without measurements in entangled bases [9–11].

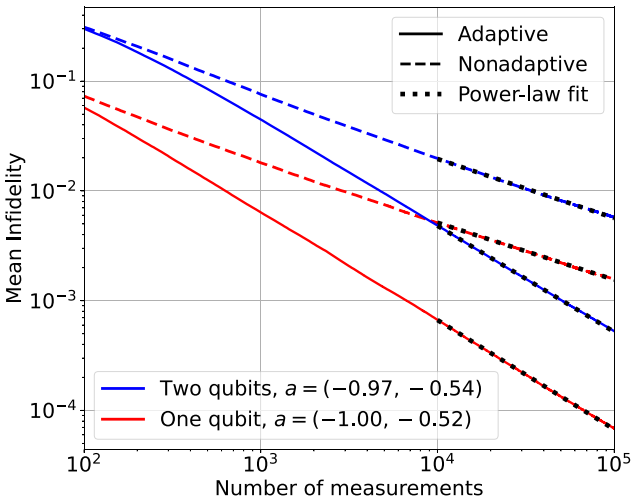



FIG. 2. Comparison of mean infidelity scaling between adaptive and nonadaptive Bayesian measurement strategies for single- and two-qubit state reconstruction. A power-law curve $N^a$ is fitted between $10^4$ and $10^5$ measurements, where the scaling exponent $a$ is displayed for (adaptive, nonadaptive) in the legend. The single-qubit nonadaptive approach uses static Pauli-6 measurements. Both adaptive and nonadaptive two-qubit strategies use factorized measurements, i.e. $\mathcal{M} = \mathcal{N}_1 \otimes \mathcal{N}_2$. The nonadaptive approach uses static factorized Pauli-6 measurements, i.e. $\mathcal{M} = \mathcal{M}_{\text{Pauli-6}} \otimes \mathcal{M}_{\text{Pauli-6}}$. The single- and two-qubit curves are averaged over 2000 and 500 Haar-random states respectively. Both axes are plotted on a logarithmic scale.

### C. Mechanism behind optimality

Here we present a summary of the analytical derivation of the conditions that an optimal adaptive measurement strategy must satisfy, formulated in terms of the Fisher information matrix. To maintain conciseness, we focus solely on the primary quantities relevant to the subsequent sections. For a more comprehensive discussion, see Appendix B.

The basis of this analysis is the mechanism derived in Ref. [10], and refined in Ref. [9]. The starting point of these works is that the state reconstruction infidelity scales, on average, with the singular values $\{\sigma_i\}$ of the Fisher information matrix $\mathcal{I}$ [defined in Eq. (12)]:

$$\langle 1 - F \rangle = \sum_{i=2}^{\nu+1} \frac{1}{\sigma_i}, \tag{15}$$

where $\nu$ is the number of degrees of freedom in the estimated quantum state. The sum in Eq. (15) starts from $i = 2$ because the singular value $\sigma_1$ is related to the normalization of the state and is therefore fixed. Following our estimation problem outlined in Sec. III A, we only consider full-rank estimators, and therefore $\nu = 4^n - 1$. The asymptotic scaling of the singular values is proportional to $N$, which means that the expected infidelity scales proportionally to $1/N$ [see Eq. (B13) in Appendix B 2]. Equation (15) applies only to singular values that are nonvanishing, as in differing from zero beyond statistical fluctuations of the estimator [2,9,10]. The number of nonvanishing singular values corresponds to the number of degrees of freedom of the target state. For example, if the target state is pure, only $2^n - 1 < \nu$ singular values are nonvanishing. Consequently, the scaling described in Eq. (15) holds only for these nonvanishing singular values, while the remaining ones must be treated separately, leading to contributions that scale as $1/\sqrt{N}$ [2,9]. The suboptimal scaling can be understood by considering that, in an overparametrized estimator, certain estimator parameters can lie along the boundary of the allowable parameter space. As a concrete example, consider a single-qubit state represented on the Bloch sphere, with a pure target state. In this case, the true value of the radial component lies exactly on the boundary of its parameter space. We show in Appendix A that, near such a boundary, the infidelity exhibits a square-root degradation in its scaling compared to the case where the true parameter lies in the interior of the parameter space.

The singular values of the Fisher information matrix are related to the eigenvalues of the density matrix. The rows and columns of $\mathcal{I}$ are proportional to the vector [9]

$$\omega_\gamma^\alpha = \begin{pmatrix} \sqrt{\lambda_1} M_\gamma^\alpha |\lambda_1\rangle \\ \sqrt{\lambda_2} M_\gamma^\alpha |\lambda_2\rangle \\ \vdots \\ \sqrt{\lambda_{2^n}} M_\gamma^\alpha |\lambda_{2^n}\rangle \end{pmatrix}, \quad \mathcal{I} \propto N \sum_\alpha \sum_\gamma \frac{1}{p_\gamma^\alpha} \omega_\gamma^\alpha (\omega_\gamma^\alpha)^\dagger, \tag{16}$$

where $\lambda_i$ are the eigenvalues of the target state with the corresponding eigenstates $|\lambda_i\rangle$, and $\gamma$ labels the outcome of measurement setting $\alpha$ with respective probability $p_\gamma^\alpha$:

$$p_\gamma^\alpha = \frac{1}{\tau_\gamma^\alpha} \sum_i \left|\left| \sqrt{\lambda_i} M_\gamma^\alpha |\lambda_i\rangle \right|\right|_2^2, \tag{17}$$

where $\tau_\gamma^\alpha = \text{Tr}(M_\gamma^\alpha)$ and $|||\psi\rangle||_2 = \sqrt{\langle\psi||\psi\rangle}$. From Eq. (16), we see that if any eigenvalues $\lambda_i$ are small (compared to the statistical uncertainty of the estimator), entire rows and columns of $\mathcal{I}$ become suppressed. As a result, the Fisher information matrix effectively behaves as rank deficient, reflecting a mismatch between the degrees of freedom of the estimator and those of the target state. To recover optimal scaling, one should choose a measurement strategy that makes the estimator effectively behave as if it were full rank.

This can be achieved by amplifying the suppressed rows and columns of $\mathcal{I}$ through the factor $1/p_\gamma^\alpha$. In particular, one can choose a measurement setting $\mathcal{M}^\alpha$ containing elements $M_\gamma^\alpha$ that are orthogonal to the eigenvectors associated with the nonvanishing eigenvalues, i.e., $M_\gamma^\alpha |\lambda_i\rangle = 0$ for all $\lambda_i \gg 0$. Such measurements isolate the components of $\omega_\gamma^\alpha (\omega_\gamma^\alpha)^\dagger$ corresponding to the small eigenvalues. The associated probability $p_\gamma^\alpha$ is itself small and proportional to the suppressed entries of $\omega_\gamma^\alpha (\omega_\gamma^\alpha)^\dagger$,

$$p_\gamma^\alpha \propto \left|\left| \sqrt{\lambda_i} M_\gamma^\alpha |\lambda_i\rangle \right|\right|_2^2 \approx \left|\left| \sqrt{\lambda_j} M_\gamma^\alpha |\lambda_j\rangle \right|\right|_2^2, \tag{18}$$

for all $\lambda_i \approx \lambda_j \approx 0$. Consequently, the prefactor $1/p_\gamma^\alpha$ compensates for this suppression, producing a new nonvanishing singular value in the Fisher information matrix. This procedure can be repeated for multiple measurement operators $M_\gamma^\alpha$ until the Fisher information matrix becomes full rank.

We can summarize the above discussion succinctly. To recover optimal asymptotic scaling, the measurement strategy

must ensure that all singular values of the Fisher information matrix are nonvanishing. This is achieved by selecting measurement operators that are orthogonal to the eigenvectors associated with nonvanishing eigenvalues of the target state. In other words, one needs to find measurement operators of the form

$$M_\gamma^\alpha = \tau_\gamma^\alpha |\psi_\gamma^\alpha\rangle\langle\psi_\gamma^\alpha| \text{ such that } \langle\psi_\gamma^\alpha|\lambda_k\rangle = 0 \quad \text{for all } k \quad \text{with} \quad \lambda_k \neq 0, \tag{19}$$

and the number of such measurement operators required for $n$-qubit states is given by [9]

$$N_{\mathcal{M}} = 4^n - 1 - (R2^{n+1} - R^2 - 1), \tag{20}$$

where $R$ is the number of nonvanishing eigenvalues in the target state.

The two adaptive approaches discussed in Secs. III B 1 and III B 2 achieve optimal infidelity scaling in slightly different ways. The two-step adaptive strategy achieves equalization of the terms in Eq. (18) by realigning the measurement operators to be orthogonal to the nonvanishing eigenstates of an intermediate-state estimate $\rho_{\text{init}}$ [9,10]. These intermediate eigenstates are often entangled in the multiqubit case, which means that optimal scaling cannot always be achieved with separable measurements with a single measurement setting. However, optimal scaling can be achieved using multiple settings [9]. Multistep adaptive Bayesian strategies do not explicitly optimize for orthogonal measurement operators, but rather continuously update the measurement strategy based on the uncertainty of the posterior distribution $\pi_f(\rho)$ [11,12]. The different measurement settings are amplifying the small singular values in a comparable way, within the current estimator uncertainty, and as additional measurements are performed, the measurement operators are progressively adjusted to become more orthogonal to the eigenstates corresponding to nonvanishing eigenvalues.

## IV. LOSS OF OPTIMAL ASYMPTOTIC SCALING

To investigate how adaptive measurement strategies perform in the presence of detector noise, we examine how the mechanism described in Sec. III C is affected when depolarizing readout noise is introduced. Depolarizing noise (defined in Sec. II C) serves as a natural stand-in for experimental noise, as most readout imperfections manifest as a statistical redistribution of measurement outcomes, which can be modeled as a type of depolarizing channel. In addition to being an effective noise model commonly observed in experiments [33,53], Pauli twirling can transform many nontrivial noise channels into effective depolarizing channels [36].

With this simple assumption, it is possible to put an upper bound on the recoverable asymptotic infidelity for adaptive measurement strategies. Using the readout-error-mitigated likelihood function in Eq. (4) and rederiving the vector $\omega_\gamma^\alpha$, the ideal measurement operators in Eq. (16) are replaced by their depolarized counterparts:

$$M_\gamma^\alpha \to \mathcal{E}_{\text{depol}}^p\left(M_\gamma^\alpha\right) = \tilde{M}_\gamma^\alpha, \tag{21}$$

where we have used the depolarizing channel from Eq. (5). The depolarized measurement operator $\tilde{M}_\gamma^\alpha$ contains an identity component $\mathbb{1}p\tau_\gamma^\alpha/2^n$, which contributes a state-independent baseline to every outcome probability. Consequently, expanding $p_\gamma^\alpha = \text{Tr}(\rho\tilde{M}_\gamma^\alpha)$ and using $\lambda_i|\langle\psi_\gamma^\alpha|\lambda_i\rangle|^2 \geqslant 0$, one obtains the lower bound

$$p_\gamma^\alpha \geqslant \tau_\gamma^\alpha \frac{p}{2^n}, \tag{22}$$

which limits how much the factor $1/p_\gamma^\alpha$ in Eq. (16) can amplify the suppressed rows and columns in the Fisher information matrix. Note that choosing $\tau_\gamma^\alpha \to 0$ does not help since $\frac{1}{p_\gamma^\alpha}\omega_\gamma^\alpha(\omega_\gamma^\alpha)^\dagger \propto \tau_\gamma^\alpha$. It follows that no multistep adaptive protocol can recover the optimal asymptotic infidelity scaling in the presence of depolarizing readout noise. This constitutes the main analytical result of this work, with the full derivation of Eq. (22) provided in Appendix B. Importantly, this conclusion concerns only the asymptotic regime under exact readout-error mitigation and does not exclude potential advantages of adaptive measurement strategies in transient, finite-sample regimes, which we discuss next. Note that this bound concerns full-rank estimators. A rank-appropriate estimator would not require amplification of suppressed singular values and retains $1/N$ scaling even with readout noise (see Sec. III A).

It is furthermore clear that although we only consider depolarizing noise here, the results remain valid for other classes of noise channel. In particular, all channels that can be decomposed into a depolarizing channel plus another nontrivial channel will have a nontrivial lower bound given by the depolarizing channel. Other common noise channels such as bit-flip, phase-flip, and qubit decay only have two or a single state which remains unaffected by the noise channel, and their contributions under the Haar measure vanish.

## V. NUMERICAL TRANSIENT BEHAVIOR

While the result presented in Sec. IV is conclusive for both two-step and multistep adaptive methods, we continue to explore the transient behavior of the multistep method here. For these simulations, we use Bayesian mean estimation. Implementation details are given in Appendix C. For adaptive measurement optimization we use the multistep information gain maximization described in Sec. III B 2. Specifically, we use the reformulated objective function in Eq. (14).

### A. Single- and two-qubit transient infidelity

In Fig. 3 we show the mean infidelity curves for multistep adaptive Bayesian mean estimates under different strengths of depolarizing readout noise. In both the single- and two-qubit case, the noise-free scenario exhibits the expected optimal $\approx 1/N$ scaling, while even small amounts of depolarizing readout noise degrade the scaling exponents. A further increase in readout noise strength leads to an increased bias-variance shift associated with performing readout-error mitigation at higher noise levels but leaves the scaling behavior unchanged. For mixed target states, no deterioration is found by introducing readout noise (see Appendix D 1).

A generic feature of the low-noise cases is that they initially follow the ideal asymptotic scaling and then gradually deteriorate as the estimation accuracy improves, eventually settling into the suboptimal $1/\sqrt{N}$ scaling characteristic of

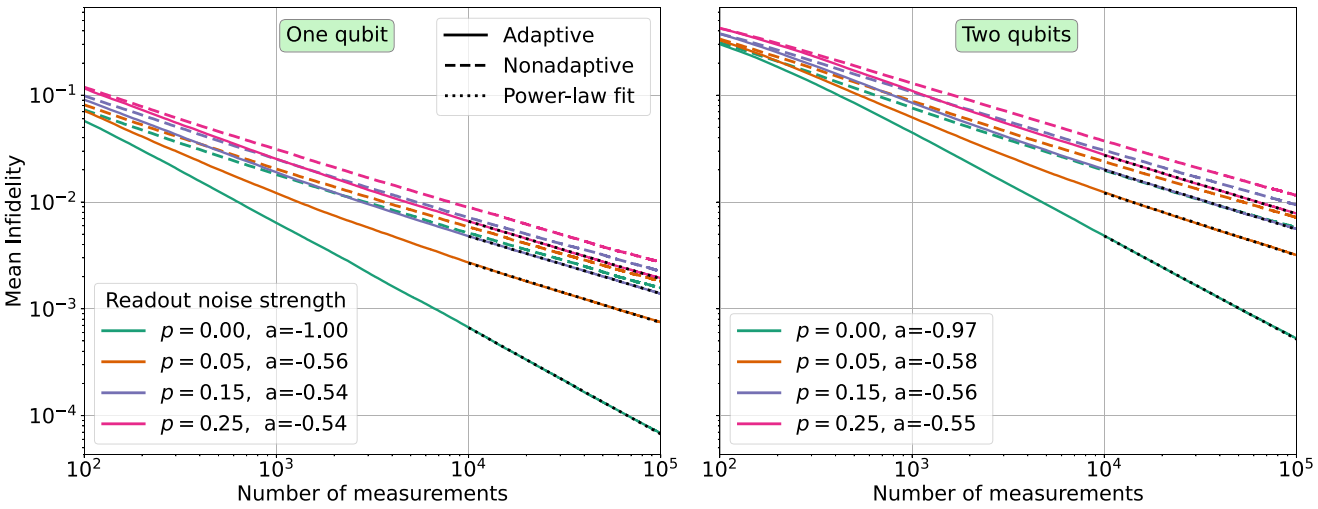



FIG. 3. Single- and two-qubit mean infidelities with different amount of depolarizing readout noise. A power law $N^a$ is fitted to all adaptive curves within the range of $10^4$ to $10^5$ measurements. The resulting scaling exponents $a$ are displayed in the legend. Both axes are plotted on a logarithmic scale. Left: Average over 2000 Haar-random single-qubit states. Right: Average over 500 Haar-random two-qubit states.

nonadaptive approaches. Notably, a residual accuracy advantage of adaptive strategies persists even in the presence of depolarizing readout noise. In Fig. 4, we show the reduction factor in infidelity between adaptive and nonadaptive strategies, $I_{\text{ada}}/I_{\text{nonada}}$, and the scaling exponent extracted from a power-law fit in a rolling window of the number of measurements for the adaptive measurement strategies. The reduction factor levels off for the noisy cases while the noiseless case keeps decreasing the reduction factor. The value at which the reduction factor levels off depends on the strength of the readout noise. For low readout noise, a significant accuracy gain can be obtained by using adaptive measurement strategies. This indicates that for well-calibrated systems, practical benefits can be expected from using adaptive measurement strategies. The rolling fit of the scaling exponent for both single- and two-qubits decreases towards $a = -0.5$ as the number of measurements increases and decreases earlier for higher readout noise. This indicates that the loss of optimal scaling is related to both the accuracy of the estimator and the information-scrambling effect of readout noise. Interquartile ranges for Fig. 3 can be found in Appendix D 2, and estimator uncertainty is discussed in Appendix E.

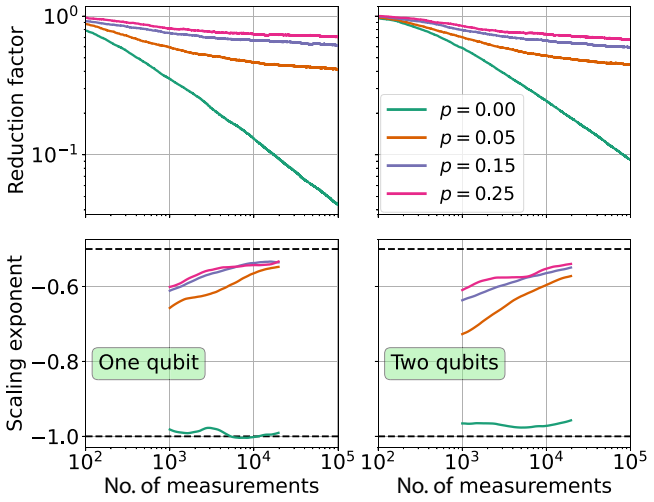



FIG. 4. Top: Reduction factor $I_{\text{ada}}/I_{\text{nonada}}$ between adaptive and nonadaptive method. The zero-noise case continuously decreases while noisy cases level off towards a constant reduction factor. The stronger the readout noise is, the larger the improvement factor becomes. Both axes are plotted on a logarithmic scale. Bottom: Rolling power-law fit $I \propto N^a$ applied to the adaptive curves in Fig. 3, where $a$ denotes the scaling exponent. The rolling power-law fit is applied in the range $[x, 5x]$, where $x$ is the measurement number. Only ranges that can be filled are plotted. The black dashed horizontal lines are for guidance, showing expected scaling for adaptive and nonadaptive curves. The asymptotic scaling decays faster for stronger readout noise. The $x$ axis is plotted on a logarithmic scale.

### B. Infidelity distribution at a given number of measurements

In Fig. 5, we present the binned distribution of the infidelity reached for a fixed number of measurements, which corresponds to a vertical cross sections of all curves in Fig. 3. We see that the adaptive distributions are slightly narrower than their nonadaptive counterparts. The noiseless adaptive case separates from the rest, while the relative distances between the noisy adaptive and nonadaptive distribution peaks remain the same. A similar binned distribution for the number of state copies required to reach a specified infidelity is presented in Appendix D 3.

## VI. FINITE-SAMPLE READOUT-ERROR MITIGATION

In the previous sections, we have assumed that the noisy measurement operators $M_\gamma$ required for the readout-error-mitigated state estimator were known exactly. Here, we turn to a scenario in which one must first perform quantum detector tomography with finite number of samples to obtain estimates of these noisy measurement operators. Similarly to how state tomography reconstructs an unknown state with known

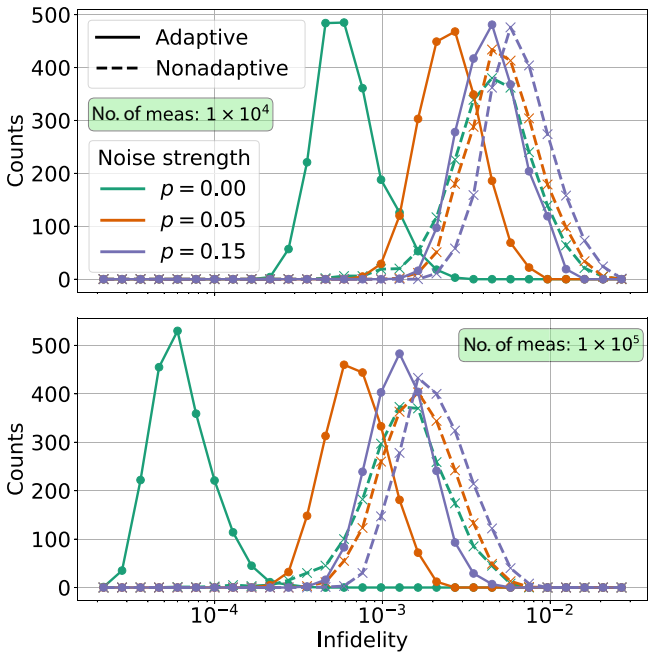



FIG. 5. Distribution of the infidelities reached at a given number of measurements for single-qubit curves in Fig. 3. The noiseless adaptive cases separate from the remaining curves which stay relatively close to each other. The $x$ axis is plotted on a logarithmic scale.

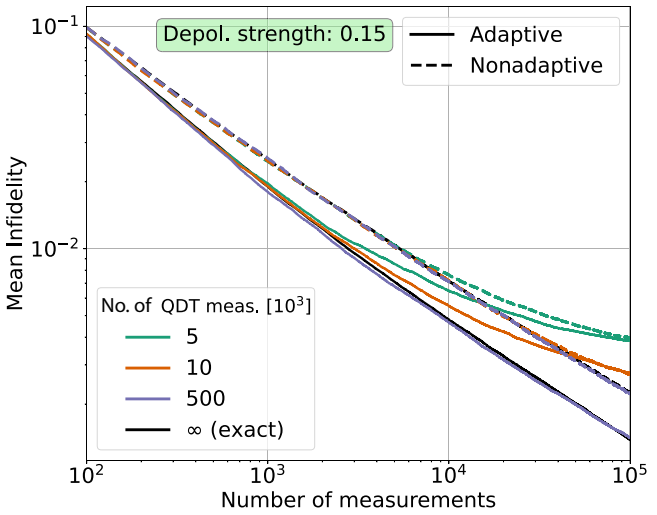



FIG. 6. Mean infidelity curves with different number of state copies used for quantum detector tomography (QDT). The simulated measurement operators are depolarized with strength $p = 0.15$. The curves are averaged over 1000 Haar-random target states, where for each target state detector tomography has been independently performed. The black curves use the exact depolarized measurement operators, and are equivalent to the curves used in Fig. 3. Both axes are plotted on a logarithmic scale.

measurement operators, detector tomography reconstructs unknown measurement operators with known calibration states [34,35]. Our calibration states are the tetrahedron states

$$\begin{aligned}
|\psi_1\rangle &= |0\rangle, \\
|\psi_2\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}|1\rangle, \\
|\psi_3\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}e^{i\frac{2\pi}{3}}|1\rangle, \\
|\psi_4\rangle &= \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}e^{i\frac{4\pi}{3}}|1\rangle.
\end{aligned} \tag{23}$$

Each tetrahedron state is prepared the same number of times for each set of measurement operators. The measurement operators are reconstructed by solving the linear set of equations using the maximum-likelihood estimator [43]

$$f_{s\gamma} = \mathrm{Tr}(\rho_s M_\gamma), \tag{24}$$

where $f_{s\gamma} = n_{s\gamma}/N$ are the observed frequencies for outcome $\gamma$ and calibration states $\rho_s = |\psi_s\rangle\langle\psi_s|$, with $s \in \{1, 2, 3, 4\}$. To compare between adaptive and nonadaptive measurement strategies, we only calibrate the diagonal entries of the measurement operators in the computational basis. This corresponds to only calibrating computational basis measurements, and discarding any phase information. This means that the Pauli-6 POVM used for the nonadaptive measurement strategy is simulated as three separate computational basis measurements with perfect unitary rotations beforehand. Similarly for the adaptive strategy, we assume that each requested setting can be implemented by perfect unitary rotations followed by a depolarized computational basis measurement.

In Fig. 6 we show simulations of state reconstruction where a varying number of state copies have been used for detector tomography. We see that when detector tomography uses fewer state copies, the reconstruction accuracy saturates to a finite value. This behavior is expected because imperfect detector calibration introduces a bias in the estimator. To investigate whether over- and underestimated noise contribute differently to the averaged infidelity, we performed additional simulations in which the reconstructed measurement operators have a constant depolarizing offset relative to the POVM used to generate the data, which is presented in Appendix F. We observed that underestimated noise dominates for pure states and reproduces the characteristics shown in Fig. 6, while overestimated noise yields better reconstruction accuracy due to overestimation of pure states. It is worth noting that when very few state copies are used for detector tomography (here less than state copies used for state tomography), the estimator does not converge properly and should only be used to indicate infidelity saturation (see Appendix E for more information).

Figure 6 indicates that, to make sure that accuracy of the reconstructed measurement operators does not limit the quality of the state reconstruction, one should allocate at least five times as many state copies to detector tomography as to state tomography. Depending on the exact structure of the readout noise, one might need more or less state copies and should be tuned for each experiment [37]. For the detector tomography simulations, 2000 states were used in the particle swarm to have better convergence for the Bayesian estimation (see Appendix C for more information).

## VII. CONCLUSION AND OUTLOOK

We investigated how detector noise affects adaptive measurement strategies for quantum state reconstruction with exact readout-error mitigation. We focused specifically on pure- and close-to-pure-state reconstruction in mixed-state estimators, which is a particularly challenging estimation problem [12,44]. Adaptive measurement strategies have the potential to improve the asymptotic scaling of reconstruction infidelity from $1/\sqrt{N}$ to $1/N$. We showed that with any amount of depolarizing readout noise, the asymptotic benefit gained from performing adaptive measurements vanishes in the nontrivial regime where the detector noise is characterized exactly, thereby removing any noise-induced bias. This breakdown is derived analytically using Fisher information optimization for the estimator, showing that a suboptimal contribution to the averaged infidelity cannot be removed when readout noise is present. Furthermore, we analyzed numerically the transient regimes with exact readout-error mitigation and found a gradual decrease of the scaling exponent as the number of measurements increases. Our results show that adaptive measurement strategies maintain a constant-factor gain $I_{\text{ada}}/I_{\text{nonada}}$, while both $I_{\text{ada/non-ada}} \propto 1/\sqrt{N}$. Both the constant factor gain and the transient scaling improve as readout noise decreases. This indicates that a practical advantage with adaptive measurement strategies can be gained with well-calibrated experiments. We investigated numerically the performance in scenarios where a finite number of state copies are used for detector characterization, by performing quantum detector tomography before performing adaptive quantum state tomography. As expected, not using enough state copies to calibrate the measurement operators causes the state reconstruction accuracy to saturate. For the depolarizing noise simulated in this work, we found that one should use at least five times the number of state copies used for a single state reconstruction for detector tomography.

These findings constrain how useful adaptive strategies can be in practical state reconstruction tasks, since noise can never be entirely removed on a hardware level. Nevertheless, we find that if the readout noise is low, a significant constant-factor improvement can be gained by using adaptive strategies. Our investigation focused solely on depolarizing noise, which is sufficiently general to cover most relevant cases [33,36,53]. However, it is possible that device-specific noise models might reveal methods capable of maintaining better asymptotic scaling despite having readout noise. It is still unclear whether these results extend to more general measurement strategies, such as collective measurements that use quantum communication [2]. The next step would be to verify this property experimentally, as current adaptive experiments do not perform readout-error mitigation, allowing only the extraction of transient scaling [9,12,27]. It would be interesting to see if the implementation of noise-conscious measurement strategies would allow some recovery of the benefit of adaptive measurements [40]. Furthermore, it would be interesting to investigate whether it is possible to derive a direct connection between the Fisher information adaptive strategy [9,10] and the Bayesian utility function approach [11,12]. Another open question is the relationship between the transient scaling exponents and readout noise strength with increasing number of measurements. It would be interesting to investigate the dependence of the improvement factor on the detector noise strength and target state purity. Similar factor improvements are observed for mixed target states (Appendix D 1), indicating that this is a general feature of adaptive measurement strategies for mixed states. A recent work [54] introduced a unified framework for performing optimal adaptive measurement strategists for state, process, and detector tomography, with associated generalized infidelity metrics for process and detector tomography, with a different approach than Refs. [2,12]. It would be interesting to investigate whether a similar approach as the one presented in this manuscript would generalize to process and detector tomography in the unified framework of Ref. [54].

## ACKNOWLEDGMENTS

The authors thank Etienne Stock for helping to derive the mechanism behind optimality. Some PYTHON packages were instrumental in this project: NUMPY [55], SCIPY [56], JOBLIB [57], and MATPLOTLIB [58]. We acknowledge support by the Federal Ministry of Research, Technology and Space (BMFTR) under project BeRyQC. This research is supported by funding from the German Research Foundation under Grant No. 398816777-SFB 1375 (NOA) and Grant No. 550495627-FOR 5919 (MLCQS).

A.S.A. developed the protocol and software under the supervision of M.G., and A.S.A. and M.G. wrote the manuscript.

## DATA AVAILABILITY

The code used to generate and analyze the results presented can be found, together with a tutorial notebook and the compiled dataset, in Ref. [59]. The raw dataset used for this project is available in Ref. [60].

## APPENDIX A: PURE STATES IN MIXED-STATE ESTIMATION

Estimating pure states proves to be more challenging than estimating mixed states when using mixed-state estimators. Given the central role that pure states play in quantum technologies and information theory, it is essential to understand their distinct behavior.

We illustrate this issue by considering two single-qubit states

$$\rho = \tfrac{1}{2}(\mathbb{1} + \vec{r}\cdot\vec{\sigma}) \quad \text{and} \quad \tau = \tfrac{1}{2}(\mathbb{1} + \vec{t}\cdot\vec{\sigma}), \tag{A1}$$

where $\tau$ denotes the target state and $\rho$ denotes the estimated state where each vector component is estimated using $N$ samples. In the Bloch representation, the infidelity can be rewritten as

$$I(\rho,\tau) = \tfrac{1}{2}(1 - \vec{r}\cdot\vec{t} - \sqrt{1-r^2}\sqrt{1-t^2}), \tag{A2}$$

where we have used the identity

$$\det(\vec{r}\cdot\vec{\sigma}) = -r^2, \tag{A3}$$

where $r = |\vec{r}|$ and $t = |\vec{t}|$. For simplicity, we will assume that the two Bloch vectors are coaxial, $\vec{r} = r\hat{n}$ and $\vec{t} = t\hat{n}$, for some unit vector $\hat{n}$. The variance of the radial component is

$\mathrm{Var}(r) = \sigma_r^2/N$, where $\sigma_r^2$ is the variance of individual measurements of the Bloch-vector components.

First, we will consider the case where the target state is mixed, $t \sim r < 1$ and $\sigma_r^2/N \ll 1 - t$. We can therefore relate the lengths of the two vectors through

$$r = t + \delta, \tag{A4}$$

where $\delta \propto 1/\sqrt{N}$ is a random variable drawn from a Gaussian distribution, $\delta \sim \mathcal{N}(0, \sigma_r^2/N)$. Using the fidelity in Eq. (A2) and the expansion

$$\sqrt{1-(t+\delta)^2} = \sqrt{1-t^2} - \frac{t\delta}{\sqrt{1-t^2}} - \frac{\delta^2}{2(1-t^2)^{3/2}} + O(\delta^3), \tag{A5}$$

we get

$$I(\rho, \tau) = \frac{\delta^2}{4\sqrt{1-t^2}} + O(\delta^3). \tag{A6}$$

The infidelity therefore scales as $1/N$ for strongly mixed states.

Second, consider the case where the target state is nearly pure and cannot be distinguished from pure states within the estimator's uncertainty, $1 - t \ll \delta \ll 1$, and we constrain the estimated Bloch vector to remain physical, $r \leqslant 1$. We can write

$$r = 1 - \xi, \tag{A7}$$

where $\xi \propto 1/\sqrt{N}$ is a random variable from some clipped Gaussian distribution. The infidelity simplifies and we get

$$I(\rho, \tau) \approx \xi/2. \tag{A8}$$

Thus, we conclude that when estimating pure states, or states closer to the surface of the Bloch sphere than the statistical error interval of the estimator, the scaling is expected to be square-root worse compared to mixed states, specifically

$$I(\rho, \sigma) \propto \begin{cases} \frac{1}{\sqrt{N}} & \text{for pure states ,} \\ \frac{1}{N} & \text{for mixed states.} \end{cases} \tag{A9}$$

## APPENDIX B: SCALING BEHAVIOR OF RANK-DEFICIENT ESTIMATORS

In this section, we provide details of the connecting steps between the results presented in Sec. IV and the original work, Ref. [10], and the refined explanation found in Ref. [9].

Our goal is to identify conditions on the generalized measurement $\mathcal{M} = \{M_\gamma\}$, under which the singular values $\sigma_i$ of the Fisher information matrix scale optimally. We have dropped the $\alpha$ superscript for convenience; it can always be reabsorbed into $\gamma$. For singular values that are nonvanishing relative to the uncertainty of the estimator, the reconstruction error scales inversely to $\sigma_i$. In this regime, the mean infidelity is given by Eq. (15), which we restate here for convenience:

$$\langle 1 - F \rangle = \sum_{i=2}^{\nu+1} \frac{1}{\sigma_i}. \tag{B1}$$

If some singular values vanish, they correspond to directions at the boundary of the parameter space and must be treated separately. In the limit $\sigma_i \to 0$, such singular values generically scale as $1/\sqrt{N}$ rather than $N$, reflecting the boundary nature of the estimation problem [9] (see also Appendix A).

We therefore seek conditions on $M_\gamma$ such that all relevant singular values are nonvanishing and scale linearly with the total number of measurements, $\sigma_i \propto N$.

### 1. Defining notation

To begin with, we need to define the notation for our specific problem. The target states $\rho$ are pure or close to pure in a $d$-dimensional Hilbert space. The target state can be written in terms of an eigendecomposition (keeping the small eigenvalues for later convenience),

$$\rho = \sum_{k=1}^{d} \lambda_k |\lambda_k\rangle\langle\lambda_k|, \tag{B2}$$

where $\{|\lambda_k\rangle\}$ forms a basis in the $d$-dimensional Hilbert space. We still want to use full-rank estimators, which means that there are $d^2 - 1$ parameters that must be estimated, and $\nu = d^2 - 1$ in Eq. (B1). We will purify this system by introducing an auxiliary Hilbert space of the same dimension, totaling $\mathcal{H}_d \otimes \mathcal{H}_d$. This purification allows us to represent the density matrices as state vectors. In addition to the purification, we will also express the purified complex state vector and the measurement operators in a space containing only real components, allowing us to write

$$\begin{aligned} \rho_{d\times d} &\overset{\text{Purification}}{\to} \sum_{k=1}^{d} \sqrt{\lambda_k}|k\rangle_d \otimes |\lambda_k\rangle_d = |\Psi\rangle_{d^2} \\ &\overset{\text{Real}}{\to} v = \begin{pmatrix} \mathcal{R}(|\Psi\rangle) \\ \mathcal{I}(|\Psi\rangle) \end{pmatrix}, \\ M_{\gamma,d\times d} &\overset{\text{Purification}}{\to} \mathbb{1}_{d\times d} \otimes M_{\gamma,d\times d} = A_\gamma \\ &\overset{\text{Real}}{\to} O_\gamma = \begin{pmatrix} \mathcal{R}(A_\gamma) & -\mathcal{I}(A_\gamma) \\ \mathcal{I}(A_\gamma) & \mathcal{R}(A_\gamma) \end{pmatrix}. \end{aligned} \tag{B3}$$

Using this notation, we can rewrite the Born rule as

$$p_\gamma = \mathrm{Tr}(M_\gamma \rho) = v^T O_\gamma v, \tag{B4}$$

where written out explicitly,

$$|\Psi\rangle = \begin{pmatrix} \sqrt{\lambda_1}|\lambda_1\rangle \\ \vdots \\ \sqrt{\lambda_d}|\lambda_d\rangle \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1}\langle 1|\lambda_1\rangle \\ \vdots \\ \sqrt{\lambda_1}\langle d|\lambda_1\rangle \\ \sqrt{\lambda_2}\langle 1|\lambda_2\rangle \\ \vdots \\ \sqrt{\lambda_d}\langle d|\lambda_d\rangle \end{pmatrix}$$

$$\text{and} \quad A_\gamma = \begin{pmatrix} M_\gamma & 0 & \cdots & 0 \\ 0 & M_\gamma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_\gamma \end{pmatrix}. \tag{B5}$$

Using this notation, we next compute the Fisher information matrix.

#### 2. Fisher information matrix for a multinomial distribution

Our measurement follows the multinomial distribution, and the likelihood functions can be written in the purified notation as

$$\mathcal{L}(v;n_\gamma) \propto \Pi_\gamma (v^T O_\gamma v)^{n_\gamma}, \tag{B6}$$

where to simplify the notation, we limit ourselves to one measurement setting, which leads us to omit the index $\alpha$ that was previously used in Sec. III C. The score function is therefore (dropping $n_\gamma$ as an argument)

$$s(v) = \ln[\mathcal{L}(v)] = \sum_\gamma n_\gamma \ln[v^T O_\gamma v]. \tag{B7}$$

We compute the Fisher information matrix using the second formulation in Eq. (12), recalling that $p_\gamma = v^T O_\gamma v$,

$$\begin{aligned}\frac{\ln[\mathcal{L}(v)]}{\partial v_i} &= \sum_\gamma \frac{n_\gamma}{p_\gamma}\frac{\partial p_\gamma}{\partial v_i},\\ \frac{\partial^2 \ln[\mathcal{L}(v)]}{\partial v_i \partial v_j} &= -\sum_\gamma \frac{n_\gamma}{p_\gamma^2}\frac{\partial p_\gamma}{\partial v_i}\frac{\partial p_\gamma}{\partial v_j} + \sum_\gamma \frac{n_\gamma}{p_\gamma}\frac{\partial^2 p_\gamma}{\partial v_i \partial v_j},\end{aligned} \tag{B8}$$

where

$$\frac{\partial p_\gamma}{\partial v_i} = 2\sum_a (O_\gamma)_{ia} v_a. \tag{B9}$$

Combining everything, we get the overall result

$$[\mathcal{I}(v)]_{ij} = \left\langle \sum_\gamma \frac{4n_\gamma}{p_\gamma^2}(O_\gamma v)_i (v^T O_\gamma)_j - \sum_\gamma \frac{n_\gamma}{p_\gamma}\frac{\partial^2 p_\gamma}{\partial v_i \partial v_j}\right\rangle. \tag{B10}$$

We compute the expectation value noting that for a multinomial distribution we have $\langle n_\gamma \rangle = N p_\gamma$, which gives

$$[\mathcal{I}(v)]_{ij} = N\left(\sum_\gamma \frac{4}{p_\gamma}(O_\gamma v)_i (v^T O_\gamma)_j - \sum_\gamma \frac{\partial^2 p_\gamma}{\partial v_i \partial v_j}\right). \tag{B11}$$

The last term vanishes because the probabilities have to be conserved $\sum_\gamma p_\gamma = 1$, which means that any sum of derivative must vanish, $\sum_\gamma \partial p_\gamma/\partial v_i = 0$. This leads to the final version of the Fisher information matrix,

$$\mathcal{I}(v) = N\sum_\gamma \frac{4}{p_\gamma} O_\gamma v v^T O_\gamma, \tag{B12}$$

and in the complex form

$$\mathcal{I}(|\Psi\rangle) = N\sum_\gamma \frac{4}{p_\gamma} A_\gamma |\Psi\rangle\langle\Psi| A_\gamma. \tag{B13}$$

#### 3. Rank-deficient estimator

From the Fisher information matrix in Eqs. (B13) and (B12), it is clear that the singular values $\sigma_i$ are generally proportional to $N$. However, this is not the case when the target state has fewer relevant degrees of freedom compared to the estimated state. When discussing the rank of the Fisher information matrix, we refer to the formulation in Eq. (B12). The full Fisher information matrix in Eq. (B12) has rank $2d^2$. Half of these dimensions come from purification and are not physically relevant. The remaining dimensions are related to the physical degrees of freedom of the estimated state resulting in $d^2 - 1$ relevant dimensions, subtracting 1 for normalization. In the estimation problem outlined in Sec. III A, the target states are pure and therefore have $2(d-1)$ degrees of freedom. The rank deficiency of the estimation problem is the difference between the number of degrees of freedom in the estimator and the target state [9]:

$$\begin{aligned}\text{rank deficiency} &= d^2 - 1 - 2(d-1)\\ &\overset{d=2^n}{=} 4^n - 1 - 2(2^n - 1).\end{aligned} \tag{B14}$$

As an illustrative example, we will explicitly consider a two-qubit state. The target state can be written as

$$\rho = \sum_{i=1}^4 \lambda_i |\lambda_i\rangle\langle\lambda_i| \tag{B15}$$

with eigenvalues $\lambda_1 \approx 1$ and $\lambda_1 \gg \lambda_2 \approx \lambda_3 \approx \lambda_4 \approx 0$. This results in the terms of Eq. (B13) generically taking the form

$$A_\gamma |\Psi\rangle\langle\Psi| A_\gamma \approx \begin{pmatrix} M_\gamma|\lambda_1\rangle\langle\lambda_1|M_\gamma & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}. \tag{B16}$$

Therefore, due to the close-to-pure target state, the Fisher information matrix has rows and columns that are suppressed when all $p_\gamma \neq 0$. As a result, Eq. (B1) is not valid for the singular values that are small compared to the estimator uncertainty, and the reconstruction infidelity is dominated by contributions that scale suboptimally as $1/\sqrt{N}$.

#### 4. Optimal adaptive measurement strategies

Adaptive strategies recover optimal asymptotic scaling by amplifying the suppressed columns and rows of $\mathcal{I}(|\Psi\rangle)$ by finding measurement operators $M_\gamma$ that satisfy $M_\gamma|\lambda_k\rangle = 0$ for state vectors $|\lambda_k\rangle$ with nonvanishing eigenvalues. This effectively amplifies the suppressed rows and columns by a factor $1/p_\gamma$. To see why this is the case, consider again the two-qubit example with a measurement operator of the form $M_\gamma = \tau_\gamma |\psi_\gamma\rangle\langle\psi_\gamma|$, where $\tau_\gamma = \mathrm{Tr}(M_\gamma)$, and we define

$$|\psi_\gamma\rangle = \alpha_{1,\gamma}|\lambda_1\rangle + \alpha_{2,\gamma}|\lambda_2\rangle + \alpha_{3,\gamma}|\lambda_3\rangle + \alpha_{4,\gamma}|\lambda_4\rangle, \tag{B17}$$

where $\sum_i |\alpha_{i,\gamma}|^2 = 1$. Using this form, we can write the outcome probabilities

$$p_\gamma = \tau_\gamma \sum_i \lambda_i |\alpha_{i,\gamma}|^2 = \langle\Psi|A_\gamma|\Psi\rangle, \tag{B18}$$

and the Fisher information matrix elements starting from

$$A_\gamma|\Psi\rangle = \begin{pmatrix}\sqrt{\lambda_1}M_\gamma|\lambda_1\rangle\\ \sqrt{\lambda_2}M_\gamma|\lambda_2\rangle\\ \sqrt{\lambda_3}M_\gamma|\lambda_3\rangle\\ \sqrt{\lambda_4}M_\gamma|\lambda_4\rangle\end{pmatrix} = \tau_\gamma \begin{pmatrix}\sqrt{\lambda_1}\alpha^*_{1,\gamma}|\psi_\gamma\rangle\\ \sqrt{\lambda_2}\alpha^*_{2,\gamma}|\psi_\gamma\rangle\\ \sqrt{\lambda_3}\alpha^*_{3,\gamma}|\psi_\gamma\rangle\\ \sqrt{\lambda_4}\alpha^*_{4,\gamma}|\psi_\gamma\rangle\end{pmatrix}, \tag{B19}$$

which results in

$$(A_\gamma|\Psi\rangle)_i(\langle\Psi|A_\gamma)_j = \sqrt{\lambda_i\lambda_j}\alpha_{i,\gamma}\alpha^*_{j,\gamma}\tau_\gamma^2|\psi_\gamma\rangle\langle\psi_\gamma|. \tag{B20}$$

$$\frac{1}{p_\gamma}A_\gamma|\Psi\rangle\langle\Psi|A_\gamma = \frac{1}{|\alpha_{2,\gamma}|^2+|\alpha_{3,\gamma}|^2+|\alpha_{4,\gamma}|^2}\begin{pmatrix}0&0&0&0\\0&|\alpha_{2,\gamma}|^2&\alpha_{2,\gamma}\alpha^*_{3,\gamma}&\alpha_{2,\gamma}\alpha^*_{4,\gamma}\\0&\alpha^*_{2,\gamma}\alpha_{3,\gamma}&|\alpha_{3,\gamma}|^2&\alpha_{3,\gamma}\alpha^*_{4,\gamma}\\0&\alpha^*_{2,\gamma}\alpha_{4,\gamma}&\alpha^*_{3,\gamma}\alpha_{4,\gamma}&|\alpha_{4,\gamma}|^2\end{pmatrix}\otimes M_\gamma, \tag{B21}$$

where the small target state eigenvalues $\lambda$ were canceled by a corresponding factor in $p_\gamma$. This matrix introduces a new nonvanishing singular value to the Fisher information matrix $\mathcal{I}(v)$, since both $M_\gamma$ and the lower $3\times 3$ matrix are outer products of state vectors and therefore have rank 1. For a two-qubit system, the rank deficiency [Eq. (B14)] is $4^n-1-2(2^n-1)\overset{n=2}{=}9$, and the lower $3\times 3$ block can form up to nine linearly independent matrices. One such set is given by

$$\begin{aligned}
|\psi_1\rangle &= |\lambda_2\rangle,\\
|\psi_2\rangle &= |\lambda_3\rangle,\\
|\psi_3\rangle &= |\lambda_4\rangle,\\
|\psi_4\rangle &= \frac{1}{\sqrt{2}}(|\lambda_2\rangle+|\lambda_3\rangle),\\
|\psi_5\rangle &= \frac{1}{\sqrt{2}}(|\lambda_2\rangle+i\,|\lambda_3\rangle),\\
|\psi_6\rangle &= \frac{1}{\sqrt{2}}(|\lambda_2\rangle+|\lambda_4\rangle),\\
|\psi_7\rangle &= \frac{1}{\sqrt{2}}(|\lambda_2\rangle+i\,|\lambda_4\rangle),\\
|\psi_8\rangle &= \frac{1}{\sqrt{2}}(|\lambda_3\rangle+|\lambda_4\rangle),\\
|\psi_9\rangle &= \frac{1}{\sqrt{2}}(|\lambda_3\rangle+i\,|\lambda_4\rangle).
\end{aligned} \tag{B22}$$

Because all of these measurement operators are orthogonal to $|\lambda_1\rangle$ and yield linearly independent matrices, an adaptive measurement strategy can recover optimal asymptotic infidelity scaling.

This mechanism extends naturally to the higher-dimensional cases. In general, the requirement for adaptive measurements to achieve optimal scaling is that the measurement operators must be orthogonal to all eigenvectors corresponding to nonvanishing eigenvalues of the target state. Specifically, the condition is

$$\begin{aligned}
&M_\gamma = \tau_\gamma|\psi_\gamma\rangle\langle\psi_\gamma| \text{ such that } \langle\psi_\gamma|\lambda_k\rangle = 0\\
&\text{for all } k \text{ with } \quad \lambda_k \neq 0,
\end{aligned} \tag{B23}$$

and the number of such measurement operators required is given by [9]

$$N_{\mathcal{M}} = 4^n - 1 - (R2^{n+1} - R^2 - 1), \tag{B24}$$

for target states with $R$ nonvanishing eigenvalues.

In Eq. (B16) only the upper left block remains nonvanishing when $\lambda_1 \gg \lambda_2 \approx \lambda_3 \approx \lambda_4 := \lambda$. If we select a measurement operator $M_\gamma$ where $\alpha_{1,\gamma} = 0$, we get

#### 5. Optimal adaptive measurement strategies with noise

The condition described in Eq. (B23) explains how using multiple measurement settings can recover optimal asymptotic infidelity scaling for rank-deficient estimators. We can now investigate how readout noise influences the optimality of adaptive strategies. We will consider the depolarizing channel introduced in Eq. (5), which is restated here for convenience:

$$\mathcal{E}^p_{\text{depol}}(A) = p\frac{\mathbb{1}}{2^n}\text{Tr}(A) + (1-p)A. \tag{B25}$$

When this is applied to measurement operators of the form $M_\gamma = \tau_\gamma|\psi_\gamma\rangle\langle\psi_\gamma|$, it gives

$$\tilde{M}_\gamma = \mathcal{E}_{\text{depol}}(M_\gamma) = p\frac{\mathbb{1}}{2^n}\tau_\gamma + (1-p)\tau_\gamma|\psi_\gamma\rangle\langle\psi_\gamma|. \tag{B26}$$

Computing the Fisher information matrix again using $\tilde{M}_\gamma$ in Eq. (B26), we have

$$A_\gamma|\Psi\rangle = \tau_\gamma\begin{pmatrix}\sqrt{\lambda_1}\left(\frac{p}{2^n}|\lambda_1\rangle + (1-p)\langle\psi_\gamma|\lambda_1\rangle|\psi_\gamma\rangle\right)\\ \vdots\\ \sqrt{\lambda_{2^n}}\left(\frac{p}{2^n}|\lambda_{2^n}\rangle + (1-p)\langle\psi_\gamma|\lambda_{2^n}\rangle|\psi_\gamma\rangle\right)\end{pmatrix}, \tag{B27}$$

and

$$\begin{aligned}
(\mathcal{I}(|\Psi\rangle))_{ij} &= N\sum_\gamma \frac{4}{p_\gamma}(A_\gamma|\Psi\rangle)_i(\langle\Psi|A_\gamma)_j\\
&= N\sum_\gamma \frac{4}{p_\gamma}\tau_\gamma^2\sqrt{\lambda_i\lambda_j}a_{ij,\gamma}
\end{aligned} \tag{B28}$$

where $i,j\in\{1,\dots,2^n\}$ and

$$\begin{aligned}
a_{ij,\gamma} = &\left(\frac{p}{2^n}|\lambda_i\rangle + (1-p)\langle\psi_\gamma|\lambda_i\rangle|\psi_\gamma\rangle\right)\\
&\times\left(\frac{p}{2^n}\langle\lambda_j| + (1-p)\langle\lambda_j|\psi_\gamma\rangle\langle\psi_\gamma|\right).
\end{aligned} \tag{B29}$$

The outcome probabilities are lower bounded by

$$p_\gamma = \tau_\gamma\frac{p}{2^n} + \tau_\gamma(1-p)\sum_{i=1}^{2^n}\lambda_i|\langle\psi_\gamma|\lambda_i\rangle|^2 \geqslant \tau_\gamma\frac{p}{2^n}, \tag{B30}$$

where we have used $\lambda_i|\langle\psi_\gamma|\lambda_i\rangle|^2 \geqslant 0$. The finite lower bound on all $p_\gamma$ prevents the prefactor $1/p_\gamma$ in Eq. (B28) from taking on sufficiently large values to recover suppressed rows and columns associated with small eigenvalues, regardless of the chosen measurement operator $M_\gamma$. Note that choosing a $\tau_\gamma \to 0$ will not help because $\frac{1}{p_\gamma}A_\gamma|\Psi\rangle\langle\Psi|A_\gamma \propto \tau_\gamma$.

We emphasize that this bound is specific to full-rank estimators. The nonzero rank deficiency, Eq. (B14), forces the suppressed rows and columns of $\mathcal{I}(|\Psi\rangle)$ to be recovered through the amplification factor $1/p_\gamma$ that the depolarizing floor $p_\gamma \geqslant \tau_\gamma p/2^n$ then caps. A rank-appropriate estimator ($R_e = R_s$) has no suppressed singular values to amplify, and the optimal $1/N$ scaling is retained even without adaptive measurements.

## APPENDIX C: IMPLEMENTATION DETAILS OF BAYESIAN MEAN ESTIMATION

The implementation of BME is primarily based on the structure suggested in Refs. [12,44]. The core idea of BME is to represent the likelihood function $\mathcal{L}(\rho)$ as an integrable posterior probability distribution $\pi_f(\rho)$, so that a mean estimate can be performed:

$$\hat{\rho}_{\mathrm{BME}} = \int d\rho\, \rho\, \pi_f(\rho). \tag{C1}$$

This is made computationally tractable by introducing a particle swarm, where the posterior distribution $\pi_f(\rho)$ is discretized and approximated by a set of states $\rho_i$ ("particles") and associated weights $w_i$:

$$\pi_f(\rho) \approx \sum_i \rho_i w_i. \tag{C2}$$

This particle swarm representation forms the basis for implementing the Bayesian update cycle, which incorporates a dataset $D = \{o_{f-1}, \ldots, o_1, o_0\}$ into the prior $\pi_0$, where $o_i$ is the index of the outcome of measurement $i$.

### 1. Bayesian update cycle

The prior distribution $\pi_0(\rho)$ is initialized by drawing $n_{\mathrm{part}}$ random particle states $\rho_i$ from the Hilbert-Schmidt distribution of mixed states [61] and the weights equalized, $w_i = 1/n_{\mathrm{part}}$. The data are integrated one by one through the Bayesian update rule, expressed in discretized form as

$$w_i^{n+1} = \frac{w_i^n \mathrm{Tr}\left(\rho_i M_{o_{n+1}}\right)}{\sum_j w_j^n \mathrm{Tr}\left(\rho_j M_{o_{n+1}}\right)}, \tag{C3}$$

where $M_{o_{n+1}}$ corresponds to the measurement operator for the $(n+1)$th measurement outcome.

This update procedure is repeated until the majority of the weight is contained in only a few particles. When this happens, the approximation becomes unreliable, and the swarm must be resampled to better approximate the actual posterior distribution. To evaluate whether the distribution is too concentrated, we have used the following quantity:

$$s = \frac{1}{\sum_i w_i^2}, \tag{C4}$$

where resampling occurs if $s < s_{\mathrm{thres}} = \tau \times n_{\mathrm{part}}$, where $\tau$ is a threshold scaling. If this threshold is crossed, the distribution is resampled using the Metropolis-Hastings (MH) algorithm.

### 2. Metropolis-Hastings resampling

Resampling refines the particle swarm to better approximate the posterior distribution during updates. The key idea is to perturb existing particles so that they cluster more densely near regions of high posterior probability. The resampling follows the procedure from Appendix C of Ref. [12], summarized here.

TABLE I. BME parameters that differ between single- and two-qubit cases.

| Parameter | One qubit | Two qubits |
|---|---|---|
| $n_{\mathrm{part}}$ | 200 | 1000 |
| $\tau$ | 0.1 | 0.05 |
| $k$ | 0.15 | 0.4 |
| $n_{\mathrm{MH}}$ | 100 | 200 |

We randomly select a particle $\rho$ from the particle swarm, where the probability of selecting a particle $\rho_i$ is $w_i$. Then we uniformly perturb the particle $\rho \to \tilde{\rho}$, and accept the perturbation with probability $p$ using the following rule:

$$p = \begin{cases} 1 & \text{when } \mathcal{L}(\tilde{\rho}) \geqslant \mathcal{L}(\rho), \\ \mathcal{L}(\tilde{\rho})/\mathcal{L}(\rho), & \text{when } \mathcal{L}(\tilde{\rho}) < \mathcal{L}(\rho). \end{cases} \tag{C5}$$

If accepted, the particle state $\rho$ is updated to the perturbed state $\tilde{\rho}$. To achieve a uniform perturbation, we use a purification scheme in which the states are embedded in a larger Hilbert space $\rho_{d\times d} \to |\psi\rangle_{d^2}$, where $d$ is the dimension of the original Hilbert space. We will drop the index in the following. In the purified space, the perturbed state vector $|\tilde{\psi}\rangle$ is obtained as follows:

$$|\tilde{\psi}\rangle = a|\psi\rangle + b\frac{|g\rangle - |\psi\rangle\langle\psi|g\rangle}{\||g\rangle - |\psi\rangle\langle\psi|g\rangle\|}, \tag{C6}$$

where $a = 1 - c^2/2$, $b = \sqrt{1-a^2}$, $c$ is drawn at random from the normal distribution $\mathcal{N}(0,\sigma)$ where $\sigma$ is proportional to the width of the current posterior distribution, and $\sigma = 2k\sum_i [1 - F(\rho_i, \bar{\rho})]w_i$ where $\bar{\rho} = \sum_i w_i\rho_i$ and $k$ is a variance-strength modifier. $|g\rangle$ is a state vector with real and imaginary entries sampled from a normal distribution $\mathcal{N}(0,1)$. The purified perturbed state $|\tilde{\psi}\rangle$ is then traced back down to the original Hilbert space, where the step is accepted or rejected. The perturbation is then repeated $n_{\mathrm{MH}}$ more times to the same state, before finally being inserted into a new particle swarm, where all weights are equalized, $w_i = 1/n_{\mathrm{part}}$. This resampling procedure is repeated until $n_{\mathrm{part}}$ new particles are inserted into the new particle swarm.

Explicit parameters used for the single- and two-qubit BME implementations are provided in Table I unless stated otherwise (the exception is Fig. 6, which uses $n_{\mathrm{part}} = 2000$ for single-qubit estimation).

## APPENDIX D: SUPPLEMENTARY NUMERICAL RESULTS

### 1. Mixed target states

In Fig. 7 we show the reconstruction infidelity for mixed target states with varying readout noise strength. The target states are drawn from the Hilbert-Schmidt distribution [61]. Since mixed states do not suffer from the quadratic accuracy penalty of pure states, even nonadaptive methods achieve the ideal $1/N$ scaling. Performing adaptive measurements still improves the accuracy by an overall factor, but the

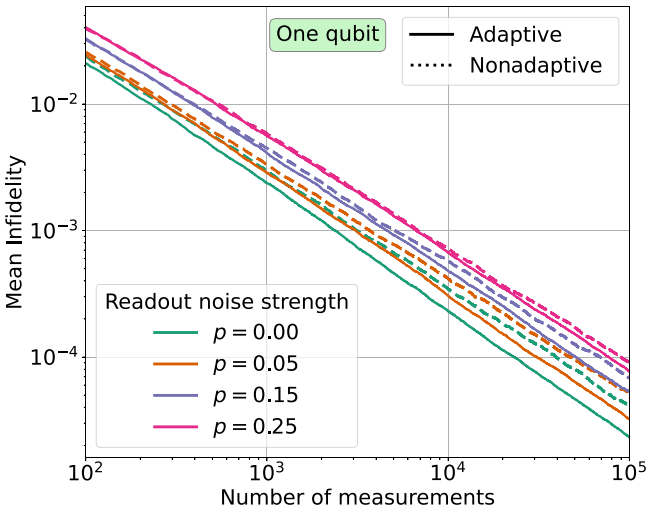



FIG. 7. Mean reconstruction infidelity averaged over 1500 single-qubit Hilbert-Schmidt distributed states for varying depolarizing readout noise strength. All curves retain an approximate $1/N$ scaling over the entire range. Both axes are plotted on a logarithmic scale.

improvement becomes smaller with stronger readout noise. At very high depolarizing strength adaptive and nonadaptive strategies become indistinguishable.

### 2. Interquartile ranges

In Fig. 8 we show the interquartile ranges of the single- and two-qubit curves with depolarizing readout strength $p = 0.05$ from Fig. 3. The interquartile ranges are computed by sorting the infidelity at each measurement number and taking the center 50% of the sorted points. We have consistent spread for both single- and two-qubit nonadaptive strategies, which is due to the random distribution of pure target states, and not to finite measurement statistics. When the target state is closer to the measurement basis, the state estimate performs better than when it is farther away, causing a natural spread [6].

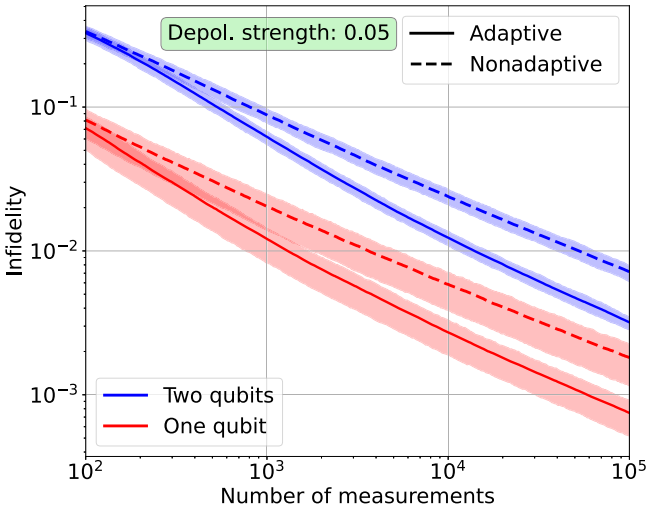



FIG. 8. Interquartile range of single- and two-qubit infidelity curves with depolarizing readout strength 0.05 from Fig. 3. The shaded area denotes the range of the center 50% of all curves at that measurement number. Both axes are plotted on a logarithmic scale.

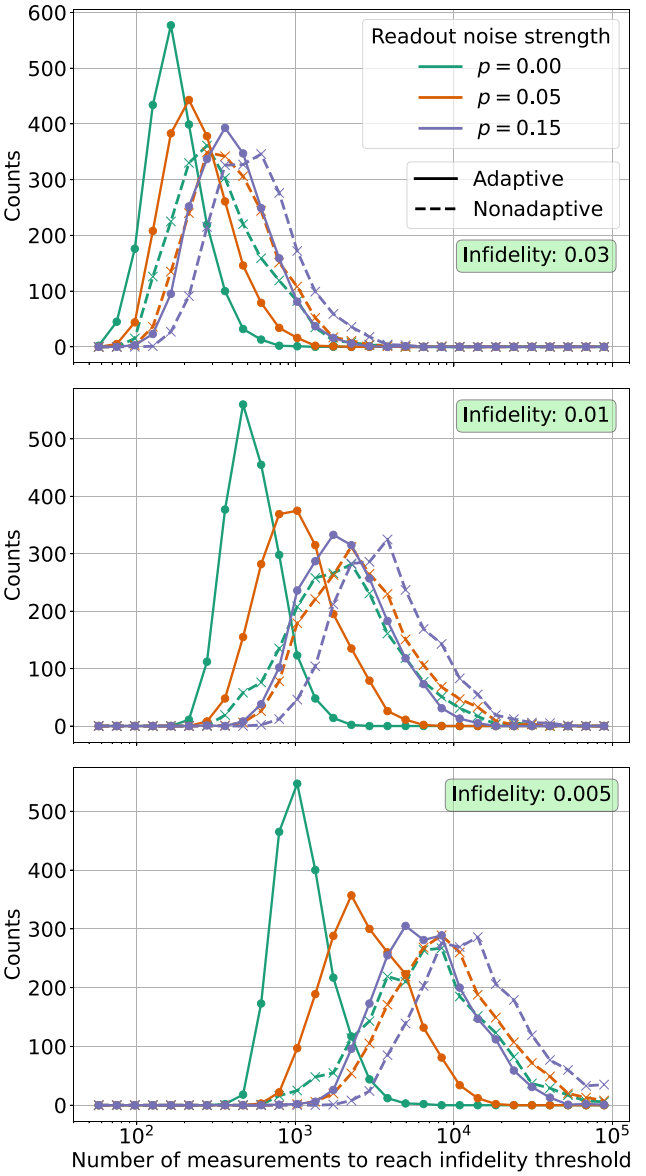



FIG. 9. Binned distribution of the number of measurements required to reach a specified infidelity for the single-qubit curves in Fig. 3. The $x$ axis is plotted on a logarithmic scale.

### 3. Number of measurements to reach specified infidelity

In Fig. 9, we show the binned distribution of the number of measurements required to reach a specified infidelity, corresponding to a vertical slice of Fig. 3. The results are qualitatively similar to the binned distribution at a constant number of measurements in Fig. 5, which instead corresponds to a horizontal intersection of Fig. 3. The noiseless adaptive distribution remains relatively peaked for all intersections, while the noisy adaptive cases broaden with increasing infidelity. The nonadaptive distributions are largely overlayered with slightly shifted peaks corresponding to the variance shift caused by readout-error mitigation.

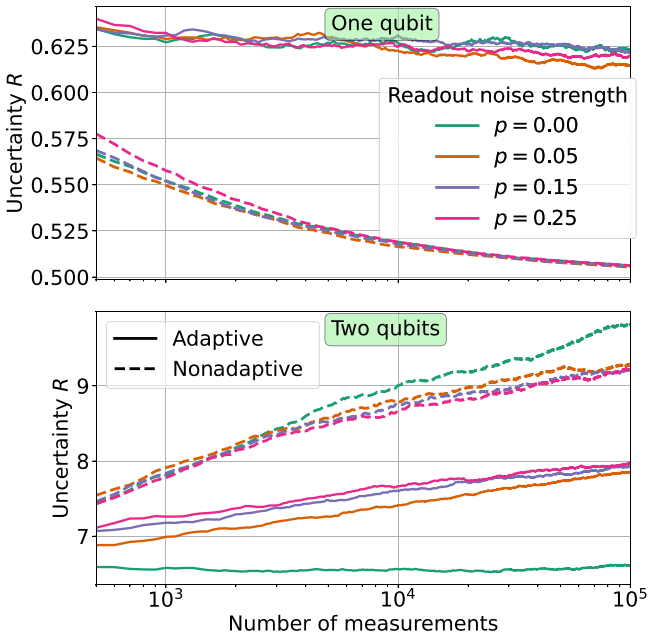



FIG. 10. Single- and two-qubit estimator uncertainty. The uncertainty $R$ is computed according to Eq. (E1). The $x$ axis is plotted on a logarithmic scale.

## APPENDIX E: ESTIMATOR UNCERTAINTY

To ensure that sufficiently many particles are used to estimate the Bayesian posterior distribution, we compute the relative estimator uncertainty,

$$R = \frac{d_B(\hat{\rho}, \rho_T)}{\sum_i w_i d_B(\rho_i, \rho_T)}, \tag{E1}$$

where $\rho_T$ is the target state, $\hat{\rho}$ is the current Bayesian mean estimate, and $\rho_i$ is the particle swarm with associated weights $w_i$, and

$$d_B(\rho, \sigma) = 2[1 - \sqrt{F(\rho, \sigma)}] \tag{E2}$$

is the Bures distance. The relative uncertainty is not a complete measure of how reliable an estimator is, because it only gives a coarse sense of how far the distribution lies, relative to its spread, from the target state. The uncertainty should be about constant to ensure that the estimator converges correctly [12]. In Fig. 10 single- and two-qubit uncertainties are plotted for the curves in Fig. 3. In the single-qubit case the uncertainty is approximately constant for all noise levels. In the two-qubit case the noiseless adaptive uncertainty is approximately constant, while the others scale approximately logarithmically with the number of measurements. The growth of the uncertainty is not large, and is likely due to the shape of the likelihood function close to the estimation boundary. Similar measurements can be found in Ref. [12], where the authors check the uncertainty of estimators with Bures-random mixed target states, which we can reproduce with our estimator using the setting for the two-qubit BME in Fig. 3, listed in Table I.

In Fig. 11 the uncertainties for the estimates in Fig. 6 are plotted. We can see that the uncertainties remain about constant for all cases using more than $10^4$ state copies for detector tomography. When less state copies are used, the uncertainty grows superlogarithmically, which means that the estimator does not converge properly, and the estimate should not be trusted in these cases. The estimator shown in Fig. 11 uses a particle swarm with $n_{\text{part}} = 2000$ states to have better convergence, in contrast to the standard choice of $n_{\text{part}} = 200$ used in all other single-qubit estimators.

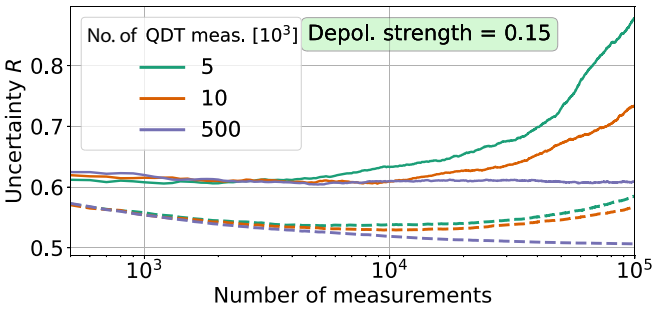



FIG. 11. Estimator uncertainty from Fig. 6 where a finite number of state copies were used for detector tomography. The curves using $5 \times 10^3$ and $10 \times 10^3$ grow superlogarithmically for the adaptive curves, and are therefore not robust estimators with these parameters. The uncertainty $R$ is computed with respect to Eq. (E1). The $x$ axis is plotted on a logarithmic scale.

## APPENDIX F: OVER- AND UNDERESTIMATED DEPOLARIZING NOISE IN DETECTOR TOMOGRAPHY

In Fig. 12, we show readout-error-mitigated state tomography in which the measurement operators used for reconstruction include a constant depolarizing offset relative to those used to generate the data. Two distinct behaviors appear depending on whether the depolarizing strength is

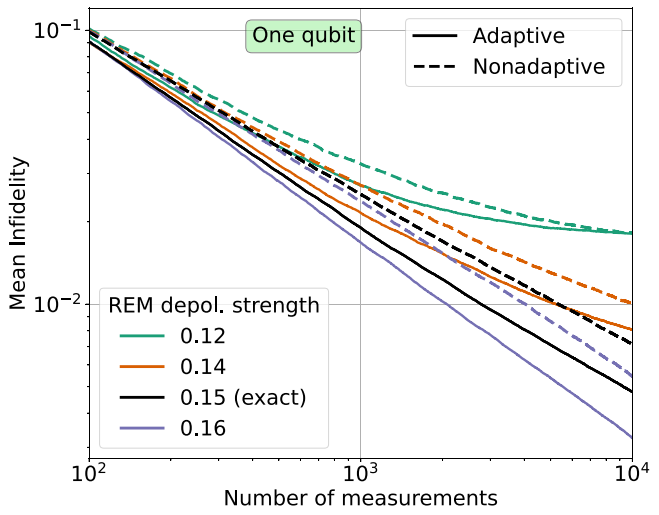



FIG. 12. Mean infidelity curves with different depolarizing noise strengths used for the measurement operators in readout-error-mitigated (REM) state tomography. The simulated target state is depolarized with strength $p = 0.15$. The over- and underestimated cases are averaged over 1000 Haar-random states. The black curves use the exact depolarized measurement operators, and are equivalent to the curves used in Fig. 3. Both axes are plotted on a logarithmic scale.

over- or underestimated. When the noise level is underestimated, the estimator converges to a state in the bulk of the estimator domain, introducing a constant bias for both adaptive and nonadaptive strategies. This is the general behavior expected from estimators with noise, such as the example in Fig. 1, where the unmitigated curves effectively underestimate the readout noise.

In contrast, overestimating the depolarizing strength counterintuitively leads to better performance than using the exact error-mitigation parameters. This likely occurs because the target states are pure and lie on the boundary of the estimator domain, and when the readout noise is overestimated, it pushes the maximum of the likelihood function outside the physical domain, causing the estimator weight to accumulate rapidly on the boundary where the true state resides. This effect should disappear if the states are sufficiently mixed compared to the accuracy of the reconstructed measurement operators. The effect of overestimating readout noise is not visible in Fig. 6 because underestimated noise contributes significantly more to the infidelity averages.

Before discussing the behavior under overestimated noise in more detail, we emphasize that when measurement operators are poorly calibrated, a readout-error-mitigated estimator is not suitable for reliable state reconstruction. Our intent here is only to explain the qualitative features of the curves, not to advocate using such estimators in this regime. When the readout noise is overestimated, the likelihood function is localized outside the estimator domain when overestimating the readout noise and the estimator does not converge well for these states. This is especially pronounced for the adaptive strategy, which saturates earlier than the nonadaptive strategy. This is likely because when the measurement setting keeps changing, the likelihood function is pushed outside the estimator domain in multiple directions, creating an inconsistent dataset for the estimator.


[1] S. Massar and S. Popescu, Optimal extraction of information from finite quantum ensembles, Phys. Rev. Lett. **74**, 1259 (1995).

[2] E. Bagan, M. A. Ballester, R. D. Gill, A. Monras, and R. Muñoz-Tapia, Optimal full estimation of qubit mixed states, Phys. Rev. A **73**, 032301 (2006).

[3] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16 (Association for Computing Machinery, New York, 2016), pp. 913–925.

[4] A. Pelecanos, J. Spilecki, E. Tang, and J. Wright, Mixed state tomography reduces to pure state tomography, arXiv:2511.15806.

[5] W. K. Wootters and B. D. Fields, Optimal state-determination by mutually unbiased measurements, Ann. Phys. (NY) **191**, 363 (1989).

[6] D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, Adaptive quantum state tomography improves accuracy quadratically, Phys. Rev. Lett. **111**, 183601 (2013).

[7] R. B. A. Adamson and A. M. Steinberg, Improving quantum state estimation with mutually unbiased bases, Phys. Rev. Lett. **105**, 030406 (2010).

[8] J. Kahn and M. Gută, Local asymptotic normality for finite dimensional quantum systems, Commun. Math. Phys. **289**, 597 (2009).

[9] G. I. Struchalin, E. V. Kovlakov, S. S. Straupe, and S. P. Kulik, Adaptive quantum tomography of high-dimensional bipartite systems, Phys. Rev. A **98**, 032330 (2018).

[10] Y. I. Bogdanov, Unified statistical method for reconstructing quantum states by purification, J. Exp. Theor. Phys. **108**, 928 (2009).

[11] F. Huszár and N. M. T. Houlsby, Adaptive Bayesian quantum tomography, Phys. Rev. A **85**, 052120 (2012).

[12] G. I. Struchalin, I. A. Pogorelov, S. S. Straupe, K. S. Kravtsov, I. V. Radchenko, and S. P. Kulik, Experimental adaptive quantum tomography of two-qubit states, Phys. Rev. A **93**, 012103 (2016).

[13] S. S. Straupe, Adaptive quantum tomography, JETP Lett. **104**, 510 (2016).

[14] T. Sugiyama, P. S. Turner, and M. Murao, Adaptive experimental design for one-qubit state estimation with finite data based on a statistical update criterion, Phys. Rev. A **85**, 052107 (2012).

[15] C. Ferrie, Self-guided quantum tomography, Phys. Rev. Lett. **113**, 190404 (2014).

[16] C. Granade, C. Ferrie, and S. T. Flammia, Practical adaptive quantum tomography, New J. Phys. **19**, 113017 (2017).

[17] Y. Quek, S. Fort, and H. K. Ng, Adaptive quantum state tomography with neural networks, npj Quantum Inf. **7**, 105 (2021).

[18] B. Qi, Z. Hou, Y. Wang, D. Dong, H.-S. Zhong, L. Li, G.-Y. Xiang, H. M. Wiseman, C.-F. Li, and G.-C. Guo, Adaptive quantum state tomography via linear regression estimation: Theory and two-qubit experiment, npj Quantum Inf. **3**, 19 (2017).

[19] O. E. Barndorff-Nielsen and R. D. Gill, Fisher information in quantum statistics, J. Phys. A **33**, 4481 (2000).

[20] A. Kalev and I. Hen, Fidelity-optimized quantum state estimation, New J. Phys. **17**, 093008 (2015).

[21] L. Zambrano, L. Pereira, S. Niklitschek, and A. Delgado, Estimation of pure quantum states in high dimension at the limit of quantum accuracy through complex optimization and statistical inference, Sci. Rep. **10**, 12781 (2020).

[22] S. T. Ahmad, A. Farooq, and H. Shin, Self-guided quantum state tomography for limited resources, Sci. Rep. **12**, 5092 (2022).

[23] C. Vargas, L. Pereira, and A. Delgado, Near-optimal pure state estimation with adaptive Fisher-symmetric measurements, arXiv:2412.04555.

[24] S. M. Kazim, A. Farooq, J. ur Rehman, and H. Shin, Adaptive quantum state tomography with iterative particle filtering, Quantum Info. Proc. **20**, 348 (2021).

[25] S. Mondal and A. K. Dutta, A Bayesian quantum state tomography along with adaptive frameworks based on linear minimum mean square error criterion, New J. Phys. **25**, 123001 (2023).

[26] A. Farooq, M. A. Ullah, J. ur Rehman, K. Lee, and H. Shin, Self-guided quantum state learning for mixed states, Quantum Info. Proc. **21**, 243 (2022).

[27] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlsby, F. Huszár, and S. P. Kulik, Experimental adaptive Bayesian tomography, Phys. Rev. A **87**, 062122 (2013).

[28] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, Experimental demonstration of adaptive quantum state estimation, Phys. Rev. Lett. **109**, 130404 (2012).

[29] Z. Hou, H. Zhu, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Achieving quantum precision limit in adaptive qubit state tomography, npj Quantum Inf. **2**, 16001 (2016).

[30] H. Hwang, J. K. Choi, and E. Kim, Adaptive quantum tomography in an indistinct measurement system with superconducting circuits, Phys. Rev. Appl. **20**, 064007 (2023).

[31] A. S. Aasen, A. Di Giovanni, H. Rotzinger, A. V. Ustinov, and M. Gärttner, Readout error mitigated quantum state tomography tested on superconducting qubits, Commun. Phys. **7**, 301 (2024).

[32] A. Arrasmith, A. Patterson, A. Boughton, and M. Paini, Development and demonstration of an efficient readout error mitigation technique for use in NISQ algorithms, arXiv:2303.17741.

[33] F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography, Quantum **4**, 257 (2020).

[34] A. Luis and L. L. Sánchez-Soto, Complete characterization of arbitrary quantum measurement processes, Phys. Rev. Lett. **83**, 3573 (1999).

[35] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, Tomography of quantum detectors, Nat. Phys. **5**, 27 (2009).

[36] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Rev. Mod. Phys. **95**, 045005 (2023).

[37] A. Di Giovanni, A. S. Aasen, J. Lisenfeld, M. Gärttner, H. Rotzinger, and A. V. Ustinov, Benchmarking the quality of multiplexed qubit readout beyond assignment fidelity, Phys. Rev. Appl. **24**, 044043 (2025).

[38] Y. I. Bogdanov, B. I. Bantysh, N. A. Bogdanova, A. B. Kvasnyy, and V. F. Lukichev, Quantum states tomography with noisy measurement channels, Proc. SPIE **10224**, 102242O (2016).

[39] Y. I. Bogdanov, N. A. Bogdanova, Y. A. Kuznetsov, K. B. Koksharov, and V. F. Lukichev, Precise tomography of qudits, Russ. Microelectron. **52**, 135 (2023).

[40] V. N. Ivanova-Rohling, N. Rohling, and G. Burkard, Optimal quantum state tomography with noisy gates, EPJ Quantum Technol. **10**, 25 (2023).

[41] R. Demkowicz-Dobrzański, J. Czajkowski, and P. Sekatski, Adaptive quantum metrology under general Markovian noise, Phys. Rev. X **7**, 041009 (2017).

[42] K. Oshio, Y. Suzuki, K. Wada, K. Hisanaga, S. Uno, and N. Yamamoto, Adaptive measurement strategy for noisy quantum amplitude estimation with variational quantum circuits, Phys. Rev. A **110**, 062423 (2024).

[43] J. Fiurášek, Maximum-likelihood estimation of quantum measurement, Phys. Rev. A **64**, 024102 (2001).

[44] R. Blume-Kohout, Optimal, reliable estimation of quantum states, New J. Phys. **12**, 043034 (2010).

[45] A. I. Lvovsky, Iterative maximum-likelihood reconstruction in quantum homodyne tomography, J. Opt. B **6**, S556 (2004).

[46] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, Completely positive dynamical semigroups of n-level systems, J. Math. Phys. **17**, 821 (1976).

[47] A. S. Holevo and V. Giovannetti, Quantum channels and their entropic characteristics, Rep. Prog. Phys. **75**, 046001 (2012).

[48] E. Magesan, Depolarizing behavior of quantum channels in higher dimensions, Quantum Inf. Comput. **11**, 466 (2011).

[49] C. A. Fuchs and C. M. Caves, Ensemble-dependent bounds for accessible information in quantum mechanics, Phys. Rev. Lett. **73**, 3047 (1994).

[50] R. D. Gill and S. Massar, State estimation for large ensembles, Phys. Rev. A **61**, 042312 (2000).

[51] S. Xiao, Y. Wang, J. Zhang, D. Dong, and H. Yonezawa, Quantum state and detector tomography with known rank, IFAC-PapersOnLine **56**, 5881 (2023).

[52] M. K. Patra, Quantum state determination: Estimates for information gain and some exact calculations, J. Phys. A **40**, 10887 (2007).

[53] A. W. R. Smith, K. E. Khosla, C. N. Self, and M. S. Kim, Qubit readout error mitigation with bit-flip averaging, Sci. Adv. **7**, eabi8009 (2021).

[54] S. Xiao, X. Wang, Y. Wang, Z. Hou, J. Zhang, I. R. Petersen, W.-Z. Yan, H. Yonezawa, F. Nori, G.-Y. Xiang, and D. Dong, Unified formalism and adaptive algorithms for optimal quantum state, detector and process tomography, arXiv:2509.05988.

[55] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, *et al.*, Array programming with NumPy, Nature (London) **585**, 357 (2020).

[56] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).

[57] Joblib Development Team Joblib: Running Python functions as pipeline jobs (2021), https://joblib.readthedocs.io/en/stable/.

[58] J. D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[59] A. Aasen, CREMST, Zenodo, 2026, http://doi.org/10.5281/zenodo.18076190.

[60] A. Aasen, Dataset for "Limitations for adaptive quantum state tomography in the presence of detector noise," Zenodo, 2026, http://doi.org/10.5281/zenodo.18076122.

[61] K. Życzkowski, K. A. Penson, I. Nechita, and B. Collins, Generating random density matrices, J. Math. Phys. **52**, 062201 (2011).