

Fast quantum measurement tomography with optimal error bounds

Leonardo Zambrano¹, Sergi Ramos-Calderer^{2,3}, and Richard Kueng⁴

¹ICFO – Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain

²Centre for Quantum Technologies, National University of Singapore, Singapore.

³Quantum Research Center, Technology Innovation Institute, Abu Dhabi, UAE

⁴Department of Quantum Information and Computation at Kepler (QUICK), Johannes Kepler University Linz, 4040 Linz, Austria

We present a two-step protocol for quantum measurement tomography that is light on classical co-processing cost and still achieves optimal sample complexity. Given measurement data from a known probe state ensemble, we first apply least-squares estimation to produce an unconstrained approximation of the POVM, and then project this estimate onto the set of valid quantum measurements. For a POVM with L outcomes acting on a d -dimensional system, we show that the protocol requires $\mathcal{O}((d^3 + d^2L)/\epsilon^2)$ samples to achieve error ϵ in worst-case distance, and $\mathcal{O}(d^2L/\epsilon^2)$ samples in average-case distance. We further establish two matching sample complexity lower bounds of $\Omega((d^3 + d^2L)/\epsilon^2)$ and $\Omega(d^2L/\epsilon^2)$ for any non-adaptive, single-copy POVM tomography protocol. Hence, our projected least squares POVM tomography is sample-optimal in both the dimension and the number of outcomes for both distances. Our method admits an analytic form when using global or local 2-designs as probe ensembles and enables rigorous non-asymptotic error guarantees. Finally, we also complement our findings with empirical performance studies carried out on a noisy superconducting quantum computer with flux-tunable transmon qubits.

1 Introduction

A significant limitation to the practical utility of quantum computers is the presence of errors and decoherence. Without active mitigation or error correction, these effects quickly accumulate and destroy the fragile quantum coherence needed for reliable computation [1–3]. Among the different noise sources, measurement errors play a particularly important role, especially in superconducting qubit platforms where readout is often noisier than gate operations [4]. Reliable characterization of detector noise is therefore crucial, not only for benchmarking hardware but also for enabling error-mitigation strategies.

Accurate detector characterization has already proven useful in several error-mitigation strategies. One prominent approach is to use the information about the detector to correct outcome probabilities or expectation values of observables, thereby reducing the impact of readout noise on subsequent experiments [1, 5–11]. Detector characterization is also important in more general estimation frameworks, such as classical shadows, where the ideal measurement channel assumed in the protocol can be replaced by a noisy one obtained from the characterization [12–15]. In both cases, reliable knowledge of the detectors enables more accurate data processing and extends the practical capabilities of noisy quantum devices.

Quantum measurement tomography (QMT), also known as detector tomography, provides the standard framework for detector characterization by reconstructing the unknown measurement, modeled as a positive operator-valued measure (POVM), from experimental data. Several approaches to QMT have been proposed

Leonardo Zambrano: leonardo.zambrano@icfo.eu

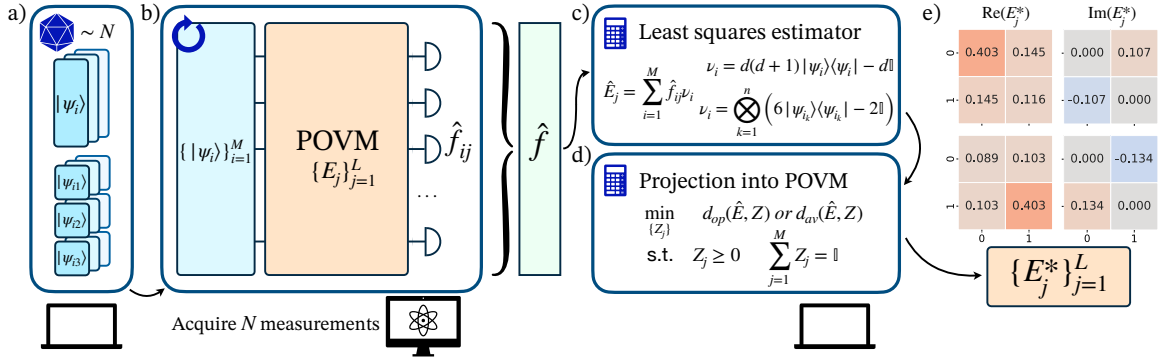


Figure 1: Sketch of the quantum measurement tomography protocol. a) Uniformly sample a set of input states from an IC ensemble $\{|\psi_i\rangle\}$. b) A quantum device is initialized in each randomized input state and measured using the POVM. The resulting frequencies $\hat{f}_{ij}$ for input state $|\psi_i\rangle$ and POVM outcome E_j are stored in a vector $\hat{f}$. c) A least-squares estimator for the POVM is computed using the measured frequencies $\hat{f}$ and the operators ν_i . This step depends on the choice of ensemble, such as a global or local 2-design. d) The raw least squares estimate is projected into the physical subspace of all POVMs via convex optimization. This produces the reconstructed POVM $(E_1^*, \dots, E_L^*)$ as output. e) Example of the first two elements of a single qubit Symmetric Informationally Complete (SIC-)POVM reconstructed from an experimental implementation on two flux-tunable transmon qubits.

and implemented over the years [16–23]. These methods collect statistics from a set of known probe states and then infer a POVM via classical co-processing. A common strategy is to fit the POVM by solving a constrained optimization problem that minimizes the discrepancy between the observed frequencies and those predicted by the estimated POVM [16–21]. While such techniques are widely used and often accurate, they become computationally expensive as the system dimension grows. Additionally, existing convergence guarantees for the estimated POVM are only valid in the asymptotic regime, where the central limit theorem applies. Note that this means that for finite sample sizes, error bars are not necessarily grounded in valid theory. Least-squares estimation (LSE) combined with projection methods offers a less computationally demanding alternative [22–25], and allows for a simpler derivation of reconstruction error bounds. Importantly, these error bounds also hold in the non-asymptotic setting.

Determining the sample complexity of QMT, that is, the number of POVM uses required to reconstruct a measurement with a given accuracy, is a fundamental challenge. Similar to quantum state tomography [26–34], such analysis enables the development of protocols that minimize the use of physical resources. This is particularly important in scenarios where time, memory, or access to quantum devices is limited. Therefore, studying sample complexity supports the design

of optimized protocols, supporting practical applications in quantum computing and the experimental verification of quantum devices.

Given the rapid advances in controllable quantum architectures, it is desirable to develop methods that are optimal in terms of sample complexity and also efficient in classical co-processing cost (think: memory and runtime). In this paper, we propose a two-step protocol that separates the estimation and constraint satisfaction processes within QMT to obtain an overall more computationally efficient reconstruction protocol. Following data acquisition, we use LSE to obtain an initial estimate of a POVM. LSE provides a computationally simple method to extract matrices that approximate the POVM, but without enforcing the required physical constraints. In a second step, we project this estimate onto the set of valid POVMs, ensuring that the final reconstructed operator satisfies all necessary conditions.

Our main contribution lies in analyzing the sample complexity of this two-step protocol and proving that its scaling matches fundamental lower bounds from quantum information theory. More precisely, for a POVM on a d -dimensional system with L outcomes, we show that our protocol requires $\mathcal{O}((d^3 + d^2L)/\epsilon^2)$ samples to achieve error ϵ in a worst-case distance, and $\mathcal{O}(d^2L/\epsilon^2)$ in an average-case distance. Conversely, we prove that *any* protocol based on independent, non-adaptive measurements must use the POVM at least $\Omega((d^3 + d^2L)/\epsilon^2)$ times to achieve the

same error in worst-case distance, and $\Omega(d^2 L/\epsilon^2)$ in average-case distance. Thus, our protocol achieves optimal sample complexity for both distances.

To evaluate the reconstruction error, we consider two natural distance measures between POVMs. The *operational distance* quantifies the worst-case distinguishability of measurement outcomes and is commonly used in discrimination and error-mitigation tasks [7, 35–38]. In contrast, the *average-case distance* captures the expected total variation distance between output distributions when input states are drawn from a random ensemble [37]. This measure reflects the typical performance of the POVM in practical scenarios. Importantly, having rigorous guarantees on how these errors scale with the number of samples provides a solid theoretical foundation for error-mitigation strategies [1, 5–11, 13–15], which often assume precise characterization of the measurement noise without explicitly quantifying the estimation error introduced by finite sample sizes. We make this explicit by showing how our tomographic bounds directly control the residual error of a classical noise-mitigation scheme [7].

Furthermore, we investigate the practical scalability of our method through numerical simulations, benchmarking it against standard Maximum Likelihood Estimation (MLE) measurement tomography. We show that using an alternating projection algorithm [39], our protocol matches the reconstruction accuracy of MLE while maintaining computational efficiency for system sizes where standard optimization solvers become intractable.

Finally, we demonstrate the performance of our protocol on a noisy quantum system, successfully recovering experimentally implemented POVMs on two flux-tunable transmon qubits. By establishing the sample-optimality of our method and validating it in real experimental settings, we provide both rigorous theoretical guarantees and a practical tool for quantum measurement characterization.

2 Preliminaries

In quantum mechanics, measurements are commonly described using the framework of positive operator-valued measures, or POVMs for short. For a quantum system of dimension d , a POVM

E is a collection of linear operators $\{E_j\}_{j=1}^L$ acting on $\mathbb{C}^d$ that satisfy the conditions

$$\sum_{j=1}^L E_j = \mathbf{1} \quad \text{and} \quad E_j \geq 0 \quad (1)$$

for every $j = 1, 2, \dots, L$. Here, $E_j \geq 0$ requires that each operator is positive semidefinite, i.e. $E_j^\dagger = E_j$ (Hermitian) and every eigenvalue is non-negative. When a quantum state ρ is measured using E , the probability of outcome j is given by the Born rule $p_j = \text{tr}(\rho E_j)$. The conditions in Eq. (1) ensure that this is a valid probability distribution, regardless of the quantum state ρ . Note that these probabilities depend on both the state and the POVM in question. Hence, they can be used to perform quantum measurement tomography. In this task, the goal is to reconstruct an unknown POVM using the measurement statistics obtained by applying it to a set of known quantum states $\{\rho_i\}_{i=1}^N$.

To assess the accuracy of tomography, we need a way to quantify the difference between POVMs. One natural choice is the operational distance d_{op} , defined as the largest total variation distance between the probability distributions that two POVMs can produce when applied to the same quantum state [7, 35–37].

Definition 1 (Operational distance). Let E and F be two L -outcome POVMs acting on a Hilbert space of dimension d . We define the operational distance d_{op} as

$$d_{\text{op}}(E, F) = \max_{\rho \in \mathcal{D}(\mathcal{H})} \frac{1}{2} \sum_{j=1}^L |\text{tr}(E_j \rho) - \text{tr}(F_j \rho)|. \quad (2)$$

This is equivalent to

$$d_{\text{op}}(E, F) = \max_{x \in \mathcal{P}(L)} \left\| \sum_{k \in x} (E_k - F_k) \right\|, \quad (3)$$

where $\mathcal{P}(L)$ is the power set of $\{1, 2, \dots, L\}$ and $\|A\|$ is the spectral norm of A , i.e. the largest matrix eigenvalue in modulus. The number $d_{\text{op}}(E, F) \in [0, 1]$ is related to the optimal probability of distinguishing between measurements E and F without using entanglement with a larger auxiliary system.

In our proofs, we will use

$$d_{\text{ext}}(X, Y) = \frac{1}{2} \sup_{a \in \{\pm 1\}^L} \left\| \sum_{j=1}^L a_j (X_j - Y_j) \right\|. \quad (4)$$

where $X = \{X_j\}_{j=1}^L$ and $Y = \{Y_j\}_{j=1}^L$ for X_j and Y_j Hermitian. If E and F are POVMs, then $d_{\text{ext}}(E, F) = d_{\text{op}}(E, F)$.

The operational distance is a worst-case distance measure. In practical scenarios, however, it is often more relevant to compare the average-case behavior of quantum measurements. To capture this, we can consider the expected Total Variation (TV) distance between the measurement statistics of two POVMs, E and F , when applied to pure states $|\psi\rangle$ drawn from a random ensemble ν

$$\text{TV}_{\text{av}}(E, F) = \mathbb{E}_{|\psi\rangle} \left[\frac{1}{2} \sum_{i=1}^L |\langle \psi | E_i | \psi \rangle - \langle \psi | F_i | \psi \rangle| \right]. \quad (5)$$

While calculating this expected TV distance directly is computationally challenging, if the ensemble ν forms at least an approximate state 4-design, $\text{TV}_{\text{av}}(E, F)$ is proportional to the following compact expression [37, Theorem 2]:

Definition 2 (Average case distance). Let E and F be two L -outcome POVMs acting on a Hilbert space of dimension d . We define the average-case distance d_{av} as

$$d_{\text{av}}(E, F) = \frac{1}{2d} \sum_{i=1}^L \sqrt{\|E_i - F_i\|_F^2 + (\text{tr}[E_i - F_i])^2}, \quad (6)$$

where $\|A\|_F = \sqrt{\text{tr}[A^\dagger A]}$ denotes the Frobenius norm.

This distance is related to the operational distance by $a d_{\text{av}}(E, F) \leq d_{\text{op}}(E, F) \leq d d_{\text{av}}(E, F)$, where $a = 0.31$ [37].

Finally, we base our unconditional lower bounds on the sample complexity required for any quantum measurement protocol on Fano's inequality. More precisely, we use the following consequence which is displayed in [32, Corollary 2.7].

Lemma 1. *Let X , Y , and $\hat{X}$ be discrete random variables forming a Markov chain $X \rightarrow Y \rightarrow \hat{X}$, where X takes values in $\mathcal{X}$. Suppose Alice sends a message $X \sim \text{Unif}(\mathcal{X})$ and Bob is able to decode the message with constant probability of success using $\hat{X}$. Then, it must hold that*

$$I(X : Y) = \Omega(\log |\mathcal{X}|), \quad (7)$$

where I denotes the mutual information

3 Projected least-squares quantum measurement tomography

Here, we present a protocol to reconstruct any L -outcome POVM on a d -dimensional system with a sample complexity of $N \in \mathcal{O}\left(\frac{(d^3 + d^2 L)}{\epsilon^2}\right)$ in the operational distance d_{op} (worst case), and $N \in \mathcal{O}\left(\frac{d^2 L}{\epsilon^2}\right)$ in the average distance d_{av} . A sketch of this protocol, and a sample of reconstructed empirical data are shown in Fig. 1.

3.1 Protocol

Our aim is to characterize a POVM E with L elements $\{E_1, E_2, \dots, E_L\}$ acting on a Hilbert space of dimension d . To this end, we collect measurement statistics from a known set of quantum states $\{|\psi_i\rangle\}_{i=1}^M$. This set is chosen to be informationally complete, meaning that for any POVM element E_j , the probabilities $\langle \psi_i | E_j | \psi_i \rangle$ uniquely identify E_j . We then perform least-squares estimation to obtain preliminary estimators for the POVM elements, which are subsequently projected onto the set of physical POVMs.

It is worthwhile to point out that the least-squares estimation step can be computed analytically if the ensemble of probe states $\{|\psi_i\rangle\}_{i=1}^M$ either forms a global 2-design or is a tensor product of single-qubit 2-designs. This 2-design probe ensemble is distinct from the 4-design ensemble used to motivate d_{av} in Sec. 2. While relating d_{av} to the average TV distance requires a 4-design, the tomographic reconstruction itself only requires a 2-design to analytically invert the measurement map and bound the estimator's variance. If the ensemble is a tensor product of single-qubit 2-designs, each state is of the form $|\psi_i\rangle = |\psi_{i_1}\rangle \otimes |\psi_{i_2}\rangle \otimes \dots \otimes |\psi_{i_n}\rangle$, where the states $\{|\psi_{i_k}\rangle\}$ of each qubit $k = 1, \dots, n$ form a 2-design. Examples of 2-designs include stabilizer states [40], states from a symmetric informationally complete POVM (SIC-POVM) [41], and states forming a maximal set of mutually unbiased bases (MUBs) [42]. We propose the following estimation protocol:

1. *Data collection.* We uniformly sample a state from the set $\{|\psi_i\rangle\}_{i=1}^M$, prepare it on the quantum device and then measure it using the POVM. This procedure is repeated N times. Let N_{ij} denote the number of shots in

which the sampled input state was $|\psi_i\rangle$ and the observed POVM outcome was j . We define $\hat{f}_{ij} = N_{ij}/N$. Then $\hat{f}_{ij}$ is an unbiased estimator of the joint probability

$$p_{ij} = \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle. \quad (8)$$

2. *Least squares estimator.* We compute empirical estimators $\hat{E}_j$ of each POVM element separately as

$$\hat{E}_j = \sum_{i=1}^M \hat{f}_{ij} \nu_i, \quad (9)$$

where

$$\nu_i = d(d+1)|\psi_i\rangle\langle\psi_i| - d\mathbb{1} \quad (10)$$

if the set of states $\{|\psi_i\rangle\}_{i=1}^M$ forms a 2-design (see App. A) or

$$\nu_i = \bigotimes_{k=1}^n (6|\psi_{i_k}\rangle\langle\psi_{i_k}| - 2\mathbb{1}) \quad (11)$$

if the set of states is a tensor product of single-qubit 2-designs (see App. B). The elements from Eq. (9) form an unbiased estimator for the POVM, denoted as $\hat{E}$.

3. *Projection onto a POVM.* The bare least squares estimator $\hat{E} = \{\hat{E}_1, \dots, \hat{E}_L\}$ will typically not meet the conditions of a physical POVM. To obtain a physical object we solve the following convex optimization problem [43]

$$\begin{aligned} & \underset{\{Z_1, \dots, Z_L\} \subset \mathbb{C}^{d \times d}}{\text{minimize}} && d_{\text{ext}}(\hat{E}, Z) \text{ or } d_{\text{av}}(\hat{E}, Z) \\ & \text{subject to} && Z_j^\dagger = Z_j, \quad Z_j \geq 0 \quad (12) \\ & && \sum_{j=1}^L Z_j = \mathbb{1}. \end{aligned}$$

We output the optimal solution (argmin) $\{E_j^*\}_{j=1}^L$ of this optimization problem as the estimator of the POVM in the corresponding distance.

The LSE in Eqs. (10) and (11) are related to single-shot estimators in quantum state tomography and classical shadows [12, 29]. Due to the symmetry of the Born rule, probing an unknown POVM with random states is the dual of probing an unknown state with random measurements. Consequently, our estimators ν_i are exactly d times the shadow estimators. This factor of d accounts for the differing trace constraint, $\text{tr}(\rho) = 1$ versus $\sum_{j=1}^L \text{tr}(E_j) = d$.

3.2 Error analysis for operational distance d_{op}

Now, we will show that the worst-case error $d_{\text{op}}(E, E^*)$ can be controlled by the sample size N .

Theorem 2. *Assume an L -outcome POVM E acting on a Hilbert space of dimension d , and an IC set of states $\{|\psi_i\rangle\}_{i=1}^M$. A reconstruction error $d_{\text{op}}(E, E^*) \leq \epsilon$ with probability $1 - \delta$ can be achieved using the protocol described above with sample size*

$$N \geq \frac{8(d^2 + \epsilon(d^2 + 1)/6)}{\epsilon^2} \ln \left(\frac{2^{L+1} 9^{2d}}{\delta} \right) \quad (13)$$

when $\{|\psi_i\rangle\}_{i=1}^M$ forms a 2-design, or

$$N \geq \frac{8(6^n + \frac{\epsilon}{6}(4^n + 1))}{\epsilon^2} \ln \left(\frac{2^{L+1} 9^{2d}}{\delta} \right). \quad (14)$$

if the POVM acts on an n -qubit system and $\{|\psi_i\rangle\}_{i=1}^M$ is a tensor product of single-qubit 2-designs.

These bounds imply that the sample complexity of our protocol scales as $\mathcal{O}((d^3 + d^2 L)/\epsilon^2)$ when using global 2-designs, and as $\mathcal{O}((12^n + 6^n L)/\epsilon^2)$ for tensor product of 2-designs. The number of degrees of freedom in an L -outcome POVM on a d -dimensional Hilbert space is $\Omega(d^2 L)$. Thus the bound matches the parameter-count scaling whenever $L \gtrsim d$, while for few-outcome measurements it contains an additional d^3 term.

Proof. We first control the error in least-squares estimator. Fix a sign vector $a = (a_1, \dots, a_L) \in \{-1, +1\}^L$ and define the signed fluctuation

$$A_a = \sum_{j=1}^L a_j (\hat{E}_j - E_j). \quad (15)$$

The vector a is fixed throughout the following scalar concentration argument. Later we take a union bound over all 2^L choices of a .

Fix a unit vector $|u\rangle \in \mathbb{C}^d$ and $a \in \{\pm 1\}^L$. Let

$$g_i(u) = \langle u | \nu_i | u \rangle. \quad (16)$$

For this fixed u and fixed a , define the real-valued random variable $Z_{u,a}$ such that

$$Z_{u,a}(i, j) = a_j g_i(u). \quad (17)$$

with probability $p_{ij} = \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle$.

The empirical mean of $Z_{u,a}$ is

$$\begin{aligned} \sum_{i=1}^M \sum_{j=1}^L \hat{f}_{ij} Z_{u,a}(i, j) &= \sum_{i=1}^M \sum_{j=1}^L \hat{f}_{ij} a_j \langle u | \nu_i | u \rangle \\ &= \langle u | \left(\sum_{j=1}^L a_j \hat{E}_j \right) | u \rangle. \end{aligned} \quad (18)$$

The expectation of $Z_{u,a}$ is

$$\begin{aligned} \mathbb{E}[Z_{u,a}] &= \sum_{i=1}^M \sum_{j=1}^L p_{ij} a_j \langle u | \nu_i | u \rangle \\ &= \sum_{j=1}^L a_j \langle u | \left(\frac{1}{M} \sum_{i=1}^M \langle \psi_i | E_j | \psi_i \rangle \nu_i \right) | u \rangle \\ &= \langle u | \left(\sum_{j=1}^L a_j E_j \right) | u \rangle. \end{aligned} \quad (19)$$

Hence, the difference between these two quantities is

$$\langle u | A_a | u \rangle = \sum_{i=1}^M \sum_{j=1}^L (\hat{f}_{ij} - p_{ij}) a_j g_i(u). \quad (20)$$

By scalar Bernstein's inequality, for every fixed unit vector u and every fixed sign vector a ,

$$\Pr(|\langle u | A_a | u \rangle| \geq t) \leq 2 \exp \left(- \frac{Nt^2}{2(v^2 + Kt/3)} \right). \quad (21)$$

The Bernstein parameters for $Z_{u,a}$ are computed in Appendices A.2 and B.2. Namely,

$$\text{Var}(Z_{u,a}) \leq v^2, \quad |Z_{u,a} - \mathbb{E}Z_{u,a}| \leq K, \quad (22)$$

with $(v^2, K) = (d^2, d^2 + 1)$ for a global 2-design and $(v^2, K) = (6^n, 4^n + 1)$ for a tensor product of single-qubit 2-designs.

We now pass to operator norm. Let $\mathcal{N}$ be a $1/4$ -net of the unit sphere of $\mathbb{C}^d$. We choose it such that $|\mathcal{N}| \leq 9^{2d}$ [44]. For every Hermitian matrix B , the standard net estimate gives [44]

$$\|B\| \leq 2 \max_{u \in \mathcal{N}} |\langle u | B | u \rangle|. \quad (23)$$

Apply Eq. (21) with $t = \epsilon/2$, and take a union bound over all $u \in \mathcal{N}$ and all $a \in \{-1, +1\}^L$. The failure probability is at most

$$2^{L+1} 9^{2d} \exp \left(- \frac{N\epsilon^2}{8(v^2 + K\epsilon/6)} \right). \quad (24)$$

The sample-size assumptions in Eqs. (13) and (14) make this quantity at most δ . Hence, with probability at least $1 - \delta$,

$$\max_{u \in \mathcal{N}} |\langle u | A_a | u \rangle| \leq \frac{\epsilon}{2} \quad (25)$$

simultaneously for every sign vector $a \in \{-1, +1\}^L$. By Eq. (23), this implies

$$d_{\text{ext}}(E, \hat{E}) = \frac{1}{2} \max_{a \in \{-1, +1\}^L} \left\| \sum_{j=1}^L a_j (\hat{E}_j - E_j) \right\| \leq \frac{\epsilon}{2}. \quad (26)$$

It remains to use the projection step. Since E is itself a feasible POVM, the definition of E^* gives $d_{\text{ext}}(E^*, \hat{E}) \leq d_{\text{ext}}(E, \hat{E})$. Using the triangle inequality,

$$\begin{aligned} d_{\text{ext}}(E, E^*) &\leq d_{\text{ext}}(E^*, \hat{E}) + d_{\text{ext}}(\hat{E}, E) \\ &\leq 2d_{\text{ext}}(\hat{E}, E) \leq \epsilon. \end{aligned} \quad (27)$$

Finally, E and E^* are POVMs. Therefore the extended operational norm coincides with the operational distance:

$$d_{\text{op}}(E, E^*) = d_{\text{ext}}(E, E^*) \leq \epsilon. \quad (28)$$

This proves the theorem. $\square$

Projecting $\hat{E}$ onto the set of POVMs using d_{ext} may become computationally demanding when L is large. As a practical alternative, one can perform the projection using any distance that upper bounds d_{ext} [23, 39]. A natural choice is $d_{\infty} = \frac{1}{2} \min_{Z \in \text{POVM}} \sum_{i=1}^L \|\hat{E}_i - Z_i\|$. Then,

$$d_{\text{op}}(E, E_{\infty}^*) \leq d_{\text{ext}}(E, \hat{E}) + d_{\text{ext}}(E_{\infty}^*, \hat{E}) \quad (29)$$

$$\leq \frac{\epsilon}{2} + d_{\infty}. \quad (30)$$

d_{∞} can be calculated after the protocol is run, and in practice is of the order of ϵ .

3.3 Error analysis for average distance d_{av}

A similar calculation can be done to bound the error in the average distance d_{av} :

Theorem 3. *Assume an L -outcome POVM E acting on a Hilbert space of dimension d and an IC set of states $\{|\psi_i\rangle\}_{i=1}^M$. A reconstruction error $d_{\text{av}}(E, E^*) \leq \epsilon$ with probability $1 - \delta$ can be*

achieved using the protocol described above with sample size

$$N \geq \frac{8L(d^2 + d + d\epsilon/3\sqrt{L})}{\epsilon^2} \ln\left(\frac{4}{\delta}\right) \quad (31)$$

when $\{|\psi_i\rangle\}_{i=1}^M$ forms a 2-design, or

$$N \geq \frac{8L(5^n + 5^{n/2}\epsilon/3\sqrt{L})}{\epsilon^2} \ln\left(\frac{4}{\delta}\right) \quad (32)$$

if the POVM acts on an n -qubit system and $\{|\psi_i\rangle\}_{i=1}^M$ is a tensor product of single-qubit 2-designs.

These sample complexity guarantees scale as order $\mathcal{O}(d^2L/\epsilon^2)$ in the global 2-design setting, and as $\mathcal{O}(5^nL/\epsilon^2)$ when using local 2-designs. Compared to the worst-case (operational) distance, this represents a reduction in sample complexity by a factor of d . The detailed proof is given in App. C.

4 Minimum sample complexity

Here, we provide a lower bound for the sample complexity of non-adaptive measurement tomography considering the worst-case distance d_{op} . The proof is based on a discretization of the problem which allows us to relate quantum measurement tomography to the problem of discrimination of well-separated POVMs. We first construct a set of $R \in \exp(\Omega(d^2))$ POVMs on dimension d that are $\epsilon/8$ apart in operational distance from each other. We then encode a random message using this set and decode it using measurement tomography with sufficient precision. From Fano's inequality, this gives us a lower bound $\Omega(d^2)$ for the mutual information between the encoder and decoder. Additionally, we obtain an upper bound $\mathcal{O}(N\epsilon^2/d)$ for the mutual information between the parties after N uses of the POVM. Using these two results, we derive a bound $N \geq \Omega(d^3/\epsilon^2)$ for the sample complexity of any non-adaptive tomographic procedure. These techniques have been previously employed in quantum information theory and are standard tools in the statistical estimation literature [12, 27, 32, 45–47].

4.1 Construction of an ϵ -packing

Here, we will demonstrate the existence of a set of $R \in \exp(\Omega(d^2))$ POVMs that are at least $\epsilon/8$

apart in operational distance. Each POVM E_{U_x} in this set has $L + 2$ elements of the form

$$\begin{aligned} E_{U_x}^k &= \frac{1}{2L} \mathbb{1} \quad \text{for } k = 1, \dots, L \\ E_{U_x}^{L+1} &= \frac{(1+\epsilon)}{4} \mathbb{1} - \frac{\epsilon}{2} U_x P U_x^\dagger \\ E_{U_x}^{L+2} &= \frac{(1-\epsilon)}{4} \mathbb{1} + \frac{\epsilon}{2} U_x P U_x^\dagger, \end{aligned} \quad (33)$$

where U_x is a unitary operator, P a projector of rank $d/2$ and $0 \leq \epsilon \leq 1/2$.

To construct this packing, we require the following set of unitaries [32]:

Lemma 4. *There exists a set of $R \in \exp(\Omega(d^2))$ unitaries such that for any $U_i \neq U_j$ in the set,*

$$\frac{1}{d} \|U_i P U_i^\dagger - U_j P U_j^\dagger\|_1 \geq \frac{1}{4}, \quad (34)$$

where P is a fixed projector of rank $d/2$.

We will use this to prove the following

Lemma 5. *There exists a set of $R \in \exp(\Omega(d^2))$ POVMs that are $\epsilon/8$ apart in operational distance d_{op} from each other.*

Proof. From Lemma 4 we fix a set of $R \in \exp(\Omega(d^2))$ unitaries to construct a set of POVMs in $\exp(\Omega(d^2))$ of the form of Eq. (33). Then all the POVMs are at least $\epsilon/8$ apart from each other:

$$\begin{aligned} d_{\text{op}}(E_{U_i}, E_{U_j}) &= \frac{\epsilon}{2} \|U_i P U_i^\dagger - U_j P U_j^\dagger\| \\ &\geq \frac{\epsilon}{2d} \|U_i P U_i^\dagger - U_j P U_j^\dagger\|_1 \\ &\geq \frac{\epsilon}{8}. \end{aligned} \quad (35)$$

□

In the proof of the lower bound on the sample complexity, each POVM E_{U_x} of the form of Eq. (33) will encode a uniform random message X that will be decoded using the information given by the measurement outcomes Y .

4.2 Upper bound on mutual information

To prove the upper bound for the mutual information, we will first bound it in terms of the χ^2 -divergence [32]:

Lemma 6. Let X be an arbitrary random variable and Z a discrete random variable. Then,

$$I(X : Z) \leq \frac{1}{\ln(2)} \left(\sum_z \mathbb{E}_{X \sim p_X} \frac{p_{Z|X}(z)^2}{p_Z(z)} - 1 \right), \quad (36)$$

where p_Z is the probability distribution of Z and $p_{Z|X}$ is the probability distribution of Z conditioned on the event $X = x$.

Measuring a set of states $\{\rho_i\}_{i=1}^N$ using a random POVM E_{U_X} in our set of POVMs defines a random variable $Y = (Y_1, \dots, Y_N)$ such that the probability of outcome $y_i \in [L+2]$ in round i is given by $p(Y_i = y_i) = \text{tr}(E_{U_X}^{y_i} \rho_i)$. Computing the mutual information $I(X : Y)$ is challenging because we do not have an explicit description of the POVMs in our construction. The next result, adapted from Proposition 4.3 in Ref. [32], allows us to replace our specific ensemble of POVMs with one defined using Haar-random unitaries in the analysis of mutual information:

Lemma 7. Let $X \sim \text{Unif}[R]$, U be a Haar random unitary and $\{\rho_i\}_{i=1}^N$ a set of N quantum states. Then, there exists a set of $R \in \exp(\Omega(d^2))$ POVMs of the form of Eq. (33) which forms an $\epsilon/8$ packing and satisfies

$$I(X : Y) \leq I(U : Z), \quad (37)$$

where $Y = (Y_1, \dots, Y_N)$ is the outcome of measuring $\{\rho_i\}_{i=1}^N$ with a random POVM E_{U_X} and $Z = (Z_1, \dots, Z_N)$ is the outcome of measuring $\{\rho_i\}_{i=1}^N$ with a random POVM E_U .

This result allows us to upper bound the mutual information by averaging over Haar-random unitaries, instead of computing it directly from the specific set of unitaries used in the $\epsilon/8$ -packing.

Using these results and following Ref. [32] we obtain an upper bound for the mutual information:

Lemma 8. Let $X \sim \text{Unif}([R])$ and $Y = (Y_1, Y_2, \dots, Y_N)$ be the outcome of the measurement of a random POVM E_{U_X} of the form of Eq. (33) over N quantum states. Then,

$$I(X : Y) \leq \frac{N}{\ln(2)} \frac{\epsilon^2}{(d+1)} \quad (38)$$

Proof. We have

$$\begin{aligned} I(X : Y) &\leq I(U : Z) \\ &= \sum_{i=1}^N I(U : Z_i | Z_{i-1}, \dots, Z_1) \\ &\leq \sum_{i=1}^N I(U : Z_i) \\ &\leq \frac{N}{\ln(2)} \mathbb{E}_{U \sim \text{Haar}} \left(\sum_{z_i} \frac{p_{Z_i|U}(z_i)^2}{p_{Z_i}(z_i)} - 1 \right), \end{aligned} \quad (39)$$

where the first line follows from Lemma 7, the second one from the chain rule for mutual information, the third one from the independence of $Z_1, \dots, Z_N$ given U and the final one from Lemma 6. Then, we have to calculate the values of $p_{Z_i}(z_i)$ and $\mathbb{E}_U p_{Z_i|U}(z_i)^2$.

For POVM elements $L+1$ and $L+2$, the probability $p_{Z_i}(z_i) = \mathbb{E}_U p_{Z_i|U}(z_i)$ is of the form

$$p_{Z_i}(z_i) = \mathbb{E}_U \text{tr} \left(\left(\frac{(1 \pm \epsilon)}{4} \mathbb{1} \mp \frac{\epsilon}{2} U P U^\dagger \right) \rho \right) = \frac{1}{4}. \quad (40)$$

For $i = 1 \dots L$ we have $p_{Z_i}(z_i) = 1/(2L)$. Next, we have for POVM elements $L+1$ and $L+2$:

$$\begin{aligned} \mathbb{E}_U p_{Z_i|U}(z_i)^2 &= \mathbb{E}_U \text{tr} \left(\left(\frac{(1 \pm \epsilon)}{4} \mathbb{1} \mp \frac{\epsilon}{2} U P U^\dagger \right) \rho \right)^2 \\ &\leq \frac{1}{16} \left(1 + \frac{2\epsilon^2}{(d+1)} \right) \end{aligned} \quad (41)$$

and for $i = 1 \dots L$ we have $p_{Z_i}(z_i)^2 = 1/(4L^2)$. Then, collecting these results and substituting in Eq. (39), we obtain

$$\begin{aligned} I(X : Y) &\leq \frac{N}{\ln(2)} \left(\sum_{z_i=1}^L \frac{1}{2L} + \left(\frac{1}{2} + \frac{\epsilon^2}{(d+1)} \right) - 1 \right) \\ &= \frac{N}{\ln(2)} \frac{\epsilon^2}{(d+1)} \end{aligned} \quad (42)$$

□

4.3 Lower bound on the sample complexity

Now we show a lower bound on the sample complexity

Theorem 9. Any procedure for quantum measurement tomography of a POVM on a d -dimensional Hilbert space that is $\epsilon/16$ accurate

in operational distance using nonadaptive, single-copy measurements on known input states requires

$$N \in \Omega\left(\frac{d^3 + d^2 L}{\epsilon^2}\right) \quad (43)$$

uses of the unknown POVM.

Proof. Assume a random message is encoded in N copies of a POVM E_{U_X} , where this POVM is uniformly sampled from the $\epsilon/8$ -packing defined by Lemma 5. Let $Y = (Y_1, \dots, Y_N)$ denote the outcomes from measuring N known quantum states using the POVM. Since each POVM in the packing is separated by at least $\epsilon/8$, a tomography algorithm that takes this data and produces an estimate of the POVM within $\epsilon/16$ precision in d_{op} (with some constant probability) is sufficient to decode the message. Then, tomography must have a sample complexity at least as large as the communication problem. From Lemma 8, the mutual information between the measurements used for tomography and the message is upper bounded by

$$I(X : Y) \leq \frac{N}{\ln(2)} \frac{\epsilon^2}{(d+1)}. \quad (44)$$

Using Lemma 1, we also have

$$I(X : Y) \geq \Omega(d^2). \quad (45)$$

Combining these results, we obtain

$$\frac{N}{\ln(2)} \frac{\epsilon^2}{(d+1)} \geq I(X : Y) \geq \Omega(d^2). \quad (46)$$

Thus, we conclude that

$$N \geq \Omega\left(\frac{d^3}{\epsilon^2}\right). \quad (47)$$

The packing above gives the dimension-dependent contribution $N = \Omega(d^3/\epsilon^2)$. The outcome-dependent contribution $N = \Omega(d^2 L/\epsilon^2)$ follows from the average-distance packing in Appendix D, together with $0.31 d_{\text{av}} \leq d_{\text{op}}$. Combining the two independent lower bounds gives

$$N = \Omega\left(\frac{d^3 + d^2 L}{\epsilon^2}\right). \quad (48)$$

□

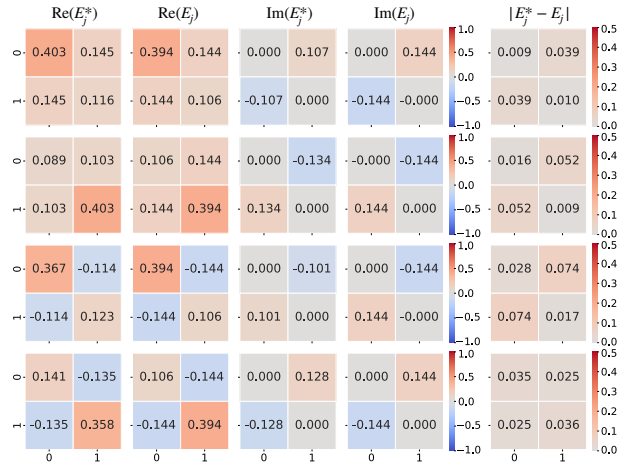


Figure 2: Results for the reconstruction of a noisy one-qubit Symmetric Informationally Complete (SIC-)POVM using an auxiliary qubit for the generalized measurement, implemented on a two-qubit flux-tunable transmon device using a budget of $1.66 \cdot 10^5$ random initial states. The reconstructed noisy POVM is compared to its expected noiseless counterpart, with the absolute difference displayed. Though noisy, the reconstruction follows the target POVM closely. The deviations from the expected values are due to the noisy device implementation.

5 Application to error mitigation

In experiments, implemented measurements often deviate from the ideal model due to both classical and non-classical noise. Classical noise can be modeled by a column-stochastic matrix Λ and mitigated by applying Λ^{-1} to the measured probabilities [7]. Non-classical noise, arising from coherent imperfections in the POVM elements, cannot be corrected by classical post-processing and leads to residual errors. Here we show that combining error mitigation with our tomographic bounds allows one to quantify and correct these errors.

Assume we have an ideal POVM M_{id} , an implemented POVM M_{exp} , and a tomographic estimate M_{est} of M_{exp} . For any quantum state ρ , we denote the corresponding outcome probability vectors by $p_{\text{id}}(\rho)$, $p_{\text{exp}}(\rho)$, and $p_{\text{est}}(\rho)$. Suppose the estimated POVM admits the decomposition $M_{\text{est}} = \Lambda M_{\text{id}} + \Delta$, where Λ is a column-stochastic matrix describing classical outcome mixing, and Δ collects all remaining (non-classical) imperfections. Our goal is to bound the deviation between the ideal statistics and the classically corrected experimental ones using Λ , uniformly over all states ρ .

Using the triangle inequality,

$$\|p_{\text{id}} - \Lambda^{-1}p_{\text{exp}}\|_1 \leq \|p_{\text{id}} - \Lambda^{-1}p_{\text{est}}\|_1 + \|\Lambda^{-1}(p_{\text{est}} - p_{\text{exp}})\|_1. \quad (49)$$

Applying the inequality $\|Ax\|_1 \leq \|A\|_{1 \rightarrow 1}\|x\|_1$ and taking the maximum over all ρ gives

$$\begin{aligned} & \max_{\rho} \|p_{\text{id}} - \Lambda^{-1}p_{\text{exp}}\|_1 \\ & \leq \|\Lambda^{-1}\|_{1 \rightarrow 1} \max_{\rho} (\|\Lambda p_{\text{id}} - p_{\text{est}}\|_1 + \|p_{\text{est}} - p_{\text{exp}}\|_1). \end{aligned} \quad (50)$$

Then, from the definition of the operational distance, we obtain

$$\begin{aligned} & \sup_{\rho} \|p_{\text{id}} - \Lambda^{-1}p_{\text{exp}}\|_1 \\ & \leq 2\|\Lambda^{-1}\|_{1 \rightarrow 1} [d_{\text{op}}(\Lambda M_{\text{id}}, M_{\text{est}}) + d_{\text{op}}(M_{\text{est}}, M_{\text{exp}})]. \end{aligned} \quad (51)$$

The first term quantifies how well the classical noise model ΛM_{id} fits the tomographic reconstruction M_{est} , and can be calculated classically [7], since both terms are known, while the second term represents the tomography error, which can be estimated using the tomographic protocol presented here.

6 Computational scalability and Numerical Simulations

To evaluate the computational scalability of the proposed protocol, we benchmark our two-step method against standard MLE. Although our approach separates statistical estimation from constraint enforcement, the exact projection onto the POVM set via semidefinite programming (SDP) can still become computationally demanding. For the operational distance d_{op} , the complexity scales exponentially with the number of outcomes L , while for the average-case distance d_{av} it matches the complexity of MLE. Here, we show that replacing the exact SDP projection with Dykstra's alternating projection method avoids these bottlenecks without loss of reconstruction accuracy.

6.1 Computational Scalability

Consider an L -outcome POVM acting on a d -dimensional Hilbert space, which is described by

$\mathcal{O}(d^2L)$ independent real parameters. We compare the computational cost of two reconstruction strategies.

The first is standard MLE with semidefinite constraints [16], which maximizes the log-likelihood of the observed frequencies subject to the POVM constraints. Using interior-point methods (for example, CVXPY with the MOSEK solver), the dominant cost arises from assembling and inverting the Hessian at each iteration, yielding a time complexity of $\mathcal{O}(d^6L^3)$ and a memory complexity of $\mathcal{O}(d^4L^2)$ [43].

The second strategy combines LSE with a heuristic projection step in Frobenius norm implemented using Dykstra's algorithm. Instead of solving an exact SDP, this method iteratively projects between the positive semidefinite cone and the sum-to-identity hyperplane [39]. The dominant operation at each iteration is the eigen-decomposition of L matrices of size $d \times d$, resulting in a per-iteration time complexity of $\mathcal{O}(d^3L)$ and space complexity of $\mathcal{O}(d^2L)$.

6.2 Numerical Simulations

We evaluate the protocol in two scaling regimes, using a total measurement budget of $N = 10^7$ shots. As probe states, we use a local 2-design formed by tensor products of the six single-qubit Pauli eigenstates. First, we fix the system size to $n = 3$ qubits and vary the number of outcomes $L \in \{4, 8, 16, 32\}$. Second, we fix the number of outcomes to $L = 8$ and scale the system dimension from $n = 1$ to $n = 6$ qubits.

For each scenario, we generate random target POVMs as follows. We sample L complex matrices G_j with standard normal entries and define positive semidefinite operators $E'_j = G_j G_j^\dagger$. Imposing the completeness relation through $S = \sum_{j=1}^L E'_j$, we construct the POVM elements $E_j = S^{-1/2} E'_j S^{-1/2}$ [23, 39]. Each shot consists of first sampling a probe state ρ_i uniformly from the local 2-design and then measuring the unknown POVM. Thus the joint probabilities are $p_{ij} = \frac{1}{M} \text{tr}(\rho_i E_j)$, where $M = 6^n$ is the number of local 2-design states. The observed counts N_{ij} are sampled from the multinomial distribution with probabilities p_{ij} , and we define the joint empirical frequencies $\hat{f}_{ij} = \frac{N_{ij}}{N}$. Theoretical upper bounds are computed using the analytical expressions derived in Thms. 2 and 3 at a confidence level of

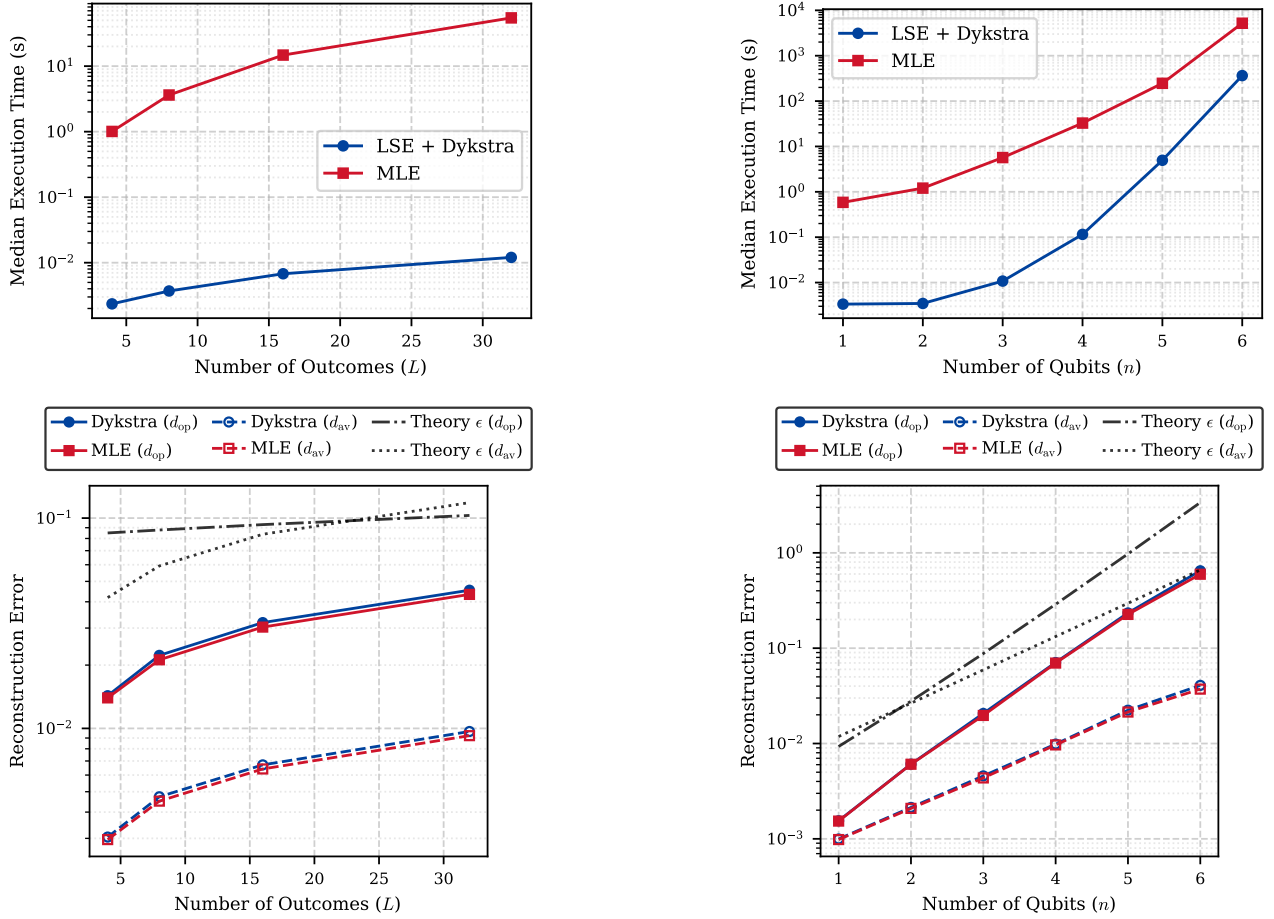


Figure 3: Computational efficiency and reconstruction accuracy of the PLS method. (Top row) Median classical execution time of our LSE+Dykstra approach compared to standard MLE, scaling with the number of POVM outcomes L for $d = 8$ (left), and scaling with the number of qubits n for $L = 8$ (right). (Bottom row) Reconstruction errors under the operational distance proxy (d_{op} , solid lines) and average-case distance (d_{av} , dashed lines) as a function of outcomes L (left) and qubits n (right). Theoretical upper bounds from Thm. 2 and Thm. 3 in black. All simulations use $N = 10^7$ shots.

$1 - \delta = 0.95$.

Figure 3 (top row) displays the median classical runtime over multiple trials as a function of outcomes L and qubits n . Standard MLE exhibits steep polynomial scaling, becoming computationally costly for large L and dimension. In contrast, the LSE + Dykstra method is orders of magnitude faster over the same parameter ranges.

Figure 3 (bottom row) compares the empirical reconstruction errors of both methods under d_{op} and d_{av} . Since evaluating the exact operational distance d_{op} requires an intractable maximization over all 2^L outcome subsets, we instead use the more efficiently computable upper bound $d_{\infty} = \frac{1}{2} \sum_{j=1}^L \|\hat{E}_j - E_j\|$. For all evaluated values of L and n , Dykstra’s algorithm achieves the same reconstruction quality as full MLE while significantly reducing the computational overhead. The

theoretical bounds for both distances are plotted alongside the numerical data, confirming that the observed errors remain consistently below the predicted bounds.

7 Empirical performance evaluations

Here we complement our theoretical findings with empirical performance studies on real quantum hardware. Target POVMs are prepared using noisy quantum gates and readout. Then, quantum measurement tomography is used to recover the POVM performed on the noisy device. In this section, we present POVMs recovered from a noisy experiment, we showcase the performance of the method in a real scenario and we discuss its potential applications.

The experiment is performed on two superconducting flux-tunable transmon qubits. The device is controlled via microwave-based arbitrary waveform generators orchestrated and calibrated by open-source middleware [48, 49]. For further details, we refer to App. E.

The experiment consists of a two-pronged approach, where data acquisition and analysis are performed separately. Random initial states are generated from the ensemble of single-qubit 2-designs. In this case, the eigenstates of the Pauli operators are used. After N random states are drawn, the outcome statistics of the chosen POVM are aggregated in a vector of relative frequencies $\hat{f}$. These probe states are known and, in the presented experimental setting, efficiently prepared via single qubit pulses acting on the target qubits. With the data collected, post-processing starts by a least-squares estimator of each element of the POVM. This estimate, does not generally satisfy the POVM constraints, therefore a projection is needed to obtain the physical object we are interested in.

We choose a set of representative POVMs to showcase the capabilities of the algorithm. In this section we present results on a single-qubit Symmetric Informationally Complete (SIC-)POVM that has been realized using one additional auxiliary qubit [50, 51]. That is, the setup involves two qubits in total. Appendix F contains additional performance studies for the two qubit Identity, Hadamard, Bell and Haar random measurements, as well as details on the implementation of the SIC-POVM.

The reconstruction of the experimental SIC-POVM is shown in Fig. 2. This has been performed with a budget of $N = 1.66 \cdot 10^5$ initial states, randomly sampled from a single qubit two-design. The reconstruction closely follows the target POVM. We also notice that there is a non-negligible difference between the experimental and target results. This difference is considerably larger than the precision of the reconstruction algorithm. This discrepancy is due to the experimental setup having inherent error in the gate application and readout that our protocol is precise enough to characterize.

This estimated POVM can be used to construct, for example, the half-sided noisy measurement channel. This channel combines the noisy estimation of the POVM with the ideal one. It is

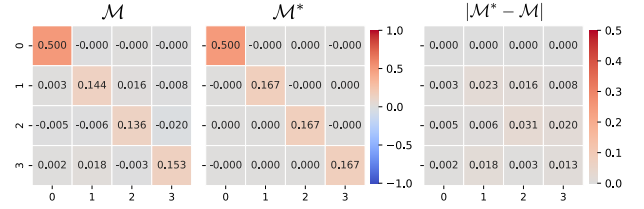


Figure 4: Comparison of the half-sided noisy measurement channel with the expected channel for a one-qubit SIC-POVM. The channels are represented in Pauli basis. The absolute difference between them is also shown on the right. The half-sided noisy channel only slightly deviates from the expected results, proportional to the depolarizing channel. This reconstruction of the channel actually implemented in the device can be used to improve results executed on the experimental device.

computed by using the noisy and exact POVM reconstructions, $\{E_j^*\}_{j=1}^L$ and $\{E_j\}_{j=1}^L$ respectively, in

$$\mathcal{M}^* = \sum_{j=1}^L |E_j\rangle\langle E_j^*|. \quad (52)$$

Here the round bra-ket follows the standard Dirac notation with respect to the Hilbert-Schmidt inner product $(A|B) = \text{tr}[A^\dagger B]$ with A, B linear operators [15]. This half-sided measurement channel $\mathcal{M}^*$ for the SIC-POVM is shown in Fig. 4, compared to the exact channel. We can see the effects of the noisy implementation on the recovered measurement channel. However, the difference between the expected and experimental results is small, and the proportionality to the effective depolarizing channel is still apparent. This channel is the key to error mitigation protocols for *classical shadows* [7, 12–15]. More precisely, the classical shadow protocol [12] relies on the inversion of the measurement channel of a spherical two design, which is proportional to the effective depolarizing channel. Of course, when implementing this POVM measurements on a noisy device, the resulting channel will not be the expected one. Therefore, the mismatch of the implemented channel and the one inverted will introduce some error on the expectation values of interest. The presented protocol, however, can estimate the noisy experimental POVM, which can then be used in place of the depolarizing channel [13]. Several works have tackled this issue by estimating the noisy POVM implementation with different techniques [7, 9, 15], but our protocol offers both a practical and theoretical advancement on the previous methods, as it targets the

noisy POVM implemented directly, making it a clear choice for the implementation of these robust classical shadow methods.

Furthermore, as shown in the previous section, this protocol can be incorporated in calibration and characterization pipelines for quantum devices [7]. Understanding the error sources of these early devices is crucial to perform useful computations. These can be used to design error mitigation schemes and develop targeted error correction frameworks. Readout error is often decoupled from the rest of the process, our protocol instead provides a complete picture for complex measurements, that includes logical operations and auxiliary space.

8 Synopsis

This work addresses the problem of quantum measurement tomography, which aims to reconstruct an unknown POVM based on the statistics obtained from a known set of input states. We propose and analyze a two-step reconstruction protocol that is both computationally fast and provably accurate. The protocol first constructs a linear least-squares estimator of the unknown POVM based on measurement statistics, and then projects this estimator onto the space of valid POVMs. This approach resembles projected least-squares quantum state and process tomography, but is adapted to the measurement problem [29, 30].

A key strength of our method is that the least-squares estimator can be computed analytically when the probe states are drawn from specific ensembles. We focus on two prominent choices: global 2-designs and tensor products of local 2-designs. For both ensembles, we provide rigorous, non-asymptotic performance guarantees for our estimator in terms of its average-case and worst-case reconstruction error.

In particular, for a L -outcome POVM acting on a d -dimensional system, our method requires $\mathcal{O}((d^3 + d^2L)/\epsilon^2)$ samples to achieve error ϵ in operational (worst-case) distance with high probability when using global 2-designs, and $\mathcal{O}((12^n + 6^nL)/\epsilon^2)$ samples when using tensor products of single-qubit 2-designs. For average-case distance, the sample complexity is $\mathcal{O}(d^2L/\epsilon^2)$ and $\mathcal{O}(5^nL/\epsilon^2)$, respectively.

In principle, a POVM can be viewed as a

quantum-to-classical channel, meaning full reconstruction could be achieved via standard quantum process tomography using the diamond norm. However, estimating a quantum channel to accuracy ϵ in diamond distance requires $\mathcal{O}(d_{\text{in}}^3 L_{\text{out}}^3/\epsilon^2)$ samples, which is substantially larger than our results. The choice between our two proposed metrics presents a trade-off depending on the application. The operational distance d_{op} requires more samples but provides strict guarantees for all possible states, making it the necessary choice for rigorous error mitigation (as shown in Sec. 5). On the other hand, the average-case distance (d_{av}) requires fewer resources and is much better suited for randomized or near-term protocols (such as variational algorithms or classical shadows). In these typical experiments, worst-case bounds might overestimate the errors, making d_{av} the most efficient and natural choice.

To assess the optimality of our approach, we also derive lower bounds on the sample complexity of any non-collective, non-adaptive QMT protocol. Our proof combines techniques from quantum information theory and classical statistics, such as discretization arguments and Fano's inequality. We show that any protocol aiming to reconstruct an unknown POVM up to error ϵ must use at least $\Omega((d^3 + d^2L)/\epsilon^2)$ samples for the operational distance and $\Omega(d^2L/\epsilon^2)$ samples for the average distance. These bounds match the scaling of our upper bounds for both distances, achieving strict optimality in both d and L .

Beyond theoretical guarantees, our numerical results confirm that the protocol is practically viable for near-term devices. We show that the empirical reconstruction errors across various system dimensions ($n \leq 6$) and outcome counts ($L \leq 32$) consistently remain below our non-asymptotic upper bounds. Crucially, our approach exhibits stable runtime profiles, offering a computationally superior alternative to standard MLE tomography without a penalty in precision.

Finally, we complement our theoretical results with empirical studies on actual quantum hardware. Our experiments validate the practical performance of our protocol and demonstrate its robustness under real noise conditions. Moreover, we discuss the use of the protocol in the context of error mitigation and characterization of contemporary devices.

Acknowledgments

LZ was supported by the Government of Spain (Severo Ochoa CEX2019-000910-S, FUNQIP and European Union NextGenerationEU PRTR-C17.I1), Fundació Cellex, Fundació Mir-Puig, Generalitat de Catalunya (CERCA program), the EU Quanterra project Veriqtas and the EU and Spanish AEI project QEC4QEA. SRC would like to acknowledge the calibration team at Technology Innovation Institute (TII) for assistance with the experimental setup. RK acknowledges financial support from the Austrian Science Fund (FWF) via the START award q-shadows (713361001) and the SFB BeyondC (10.55776/FG7) (research group 7).

References

- [1] K. Temme, S. Bravyi and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [2] J. Preskill, Quantum Computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
- [3] S. Endo, Z. Cai, S. C. Benjamin and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, *J. Phys. Soc. Jpn.* **90**, 032001 (2021).
- [4] Y. Kim et al., Evidence for the utility of quantum computing before fault tolerance, *Nature* **618**, 500 (2023).
- [5] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, *Phys. Rev. X* **7**, 021050 (2017).
- [6] Y. Chen, M. Farahzad, S. Yoo and T.-C. Wei, Detector tomography on IBM quantum computers and mitigation of an imperfect measurement, *Phys. Rev. A* **100**, 052315 (2019).
- [7] 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).
- [8] S. Bravyi, S. Sheldon, A. Kandala, D. C. McKay and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, *Phys. Rev. A* **103**, 042605 (2021).
- [9] E. van den Berg, Z. K. Mineev and K. Temme, Model-free readout-error mitigation for quantum expectation values, *Phys. Rev. A* **105**, 032620 (2022).
- [10] L. Funcke et al., Measurement error mitigation in quantum computers through classical bit-flip correction, *Phys. Rev. A* **105**, 062404 (2022).
- [11] K. Korhonen et al., Practical techniques for high-precision measurements on near-term quantum hardware and applications in molecular energy estimation, *npj Quantum Inf.* **11**, 110 (2025).
- [12] H.-Y. Huang, R. Kueng and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [13] S. Chen, W. Yu, P. Zeng and S. T. Flammia, Robust shadow estimation, *PRX Quantum* **2**, 030348 (2021).
- [14] D. E. Koh and S. Grewal, Classical Shadows With Noise, *Quantum* **6**, 776 (2022).
- [15] R. Brieger, I. Roth and M. Kliesch, Compressive gate set tomography, *PRX Quantum* **4**, 010325 (2023).
- [16] J. Fiurášek, Maximum-likelihood estimation of quantum measurement, *Phys. Rev. A* **64**, 024102 (2001).
- [17] G. M. D’Ariano, L. Maccone and P. L. Presti, Quantum Calibration of Measurement Instrumentation, *Phys. Rev. Lett.* **93**, 250407 (2004).
- [18] J. S. Lundeen et al., Tomography of quantum detectors, *Nat. Phys.* **5**, 27 (2009).
- [19] A. Feito et al., Measuring measurement: theory and practice, *New J. Phys.* **11**, 093038 (2009).
- [20] L. Zhang et al., Recursive quantum detector tomography, *New J. Phys.* **14**, 115005 (2012).
- [21] M. Cattaneo et al., Self-consistent quantum measurement tomography based on semidefinite programming, *Phys. Rev. Res.* **5**, 033154 (2023).
- [22] S. Grandi, A. Zavatta, M. Bellini and M. G. Paris, Experimental quantum tomography of a homodyne detector, *New J. Phys.* **19**, 053015 (2017).
- [23] Y. Wang et al., Two-stage estimation for quantum detector tomography: Error anal-

- ysis, numerical and experimental results, *IEEE Trans. Inf. Theory* **67**, 2293 (2021).
- [24] S. Xiao et al., On the regularization and optimization in quantum detector tomography, *Automatica* **155**, 111124 (2023).
- [25] E. Nielsen et al., Gate set tomography, *Quantum* **5**, 557 (2021).
- [26] R. O’Donnell and J. Wright, Efficient quantum tomography, in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC ’16 (Association for Computing Machinery, New York, NY, USA, 2016) p. 899–912.
- [27] J. Haah, A. W. Harrow, Z. Ji, X. Wu and N. Yu, Sample-Optimal Tomography of Quantum States, *IEEE Trans. Inf. Theory* **63**, 5628 (2017).
- [28] R. Kueng, H. Rauhut and U. Terstiege, Low rank matrix recovery from rank one measurements, *Appl. Comput. Harmon. Anal.* **42**, 88 (2017).
- [29] M. Guță, J. Kahn, R. Kueng and J. A. Tropp, Fast state tomography with optimal error bounds, *J. Phys. A: Math. Theor* **53**, 204001 (2020).
- [30] T. Surawy-Stepney, J. Kahn, R. Kueng and M. Guta, Projected least-squares quantum process tomography, *Quantum* **6**, 844 (2022).
- [31] S. Chen, J. Li, B. Huang and A. Liu, Tight bounds for quantum state certification with incoherent measurements, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2022) pp. 1205–1213.
- [32] A. Lowe and A. Nayak, Lower bounds for learning quantum states with single-copy measurements, *arXiv:2207.14438* (2022).
- [33] A. Anshu and S. Arunachalam, A survey on the complexity of learning quantum states, *Nat. Rev. Phys.* **6**, 59 (2024).
- [34] L. Zambrano, Quantum tomography for non-iid sources, *arXiv:2602.22057* (2026).
- [35] M. Navascués and S. Popescu, How Energy Conservation Limits Our Measurements, *Phys. Rev. Lett.* **112**, 140502 (2014).
- [36] Z. Puchała, L. Paweł, A. Krawiec and R. Kukulski, Strategies for optimal single-shot discrimination of quantum measurements, *Phys. Rev. A* **98**, 042103 (2018).
- [37] F. B. Maciejewski, Z. Puchała and M. Oszmaniec, Exploring Quantum Average-Case Distances: Proofs, Properties, and Examples, *IEEE Trans. Inf. Theory* **69**, 4600 (2023).
- [38] L. Zambrano, T. Parella-Dilmé, A. Acín and D. Farina, Certification of quantum properties with imperfect measurements, *arXiv:2601.16570* (2026).
- [39] J. Barberà-Rodríguez, L. Zambrano, A. Acín and D. Farina, Boosting projective methods for quantum process and detector tomography, *Physical Review Research* **7**, 013208 (2025).
- [40] C. Dankert, R. Cleve, J. Emerson and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
- [41] J. M. Renes, R. Blume-Kohout, A. J. Scott and C. M. Caves, Symmetric informationally complete quantum measurements, *J. Math. Phys.* **45**, 2171 (2004).
- [42] A. Klappenecker and M. Rotteler, Mutually unbiased bases are complex projective 2-designs, in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.* (IEEE, 2005) pp. 1740–1744.
- [43] L. Vandenberghe and S. Boyd, Semidefinite programming, *SIAM Rev.* **38**, 49 (1996).
- [44] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science* (Cambridge University Press, 2018).
- [45] J. Scarlett and V. Cevher, An introductory guide to Fano’s inequality with applications in statistical estimation, *arXiv:1901.00555* (2019).
- [46] S. T. Flammia, D. Gross, Y.-K. Liu and J. Eisert, Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators, *New J. Phys.* **14**, 095022 (2012).
- [47] I. Roth et al., Recovering quantum gates from few average gate fidelities, *Phys. Rev. Lett.* **121**, 170502 (2018).
- [48] S. Efthymiou et al., Qibolab: an open-source hybrid quantum operating system, *Quantum* **8**, 1247 (2024).
- [49] A. Pasquale et al., Qibocal: an open-source framework for calibration of self-hosted quantum devices, *arXiv:2410.00101* (2024).
- [50] Z. Jiang, A. Kalev, W. Mruczkiewicz and H. Neven, Optimal fermion-to-qubit map-

ping via ternary trees with applications to reduced quantum states learning, [Quantum 4, 276 \(2020\)](#).

- [51] Z. You, Q. Liu and Y. Zhou, Circuit optimization of qubit IC-POVMs for shadow estimation, [arXiv:2409.05676 \(2024\)](#).
- [52] I. Pinelis, Optimum bounds for the distributions of martingales in Banach spaces, [Ann. Probab. , 1679 \(1994\)](#).
- [53] D. Martinez-Taboada and A. Ramdas, Empirical Bernstein in smooth Banach spaces, [arXiv:2409.06060 \(2024\)](#).
- [54] E. Meckes and M. Meckes, Spectral measures of powers of random matrices, [ECP 18, 1 \(2013\)](#).

A Estimator using 2-designs

A.1 Least-squares estimator using 2-designs

The measurement step in the protocol defines a linear map $\mathcal{M} : M(\mathbb{C}^d) \rightarrow \mathbb{R}^M$ from POVM elements to probabilities. The map is defined as

$$[\mathcal{M}(X)]_i = \frac{1}{M} \langle \psi_i | X | \psi_i \rangle, \quad (53)$$

where $\{|\psi_i\rangle\}_{i=1}^M$ is a set of states that forms a 2-design. In the experiment, we uniformly sample a state from the set, prepare it, and then measure it using a POVM E with elements $(E_1, E_2, \dots, E_L)$. After performing several rounds of this procedure, the collected data consists of a vector of relative frequencies $\hat{f} \in \mathbb{R}^{M \times L}$, where $\hat{f}_{ij} = \frac{N_{ij}}{N}$ and N_{ij} is the number of shots in which the input state $|\psi_i\rangle$ was sampled and outcome j was observed. The relation between the frequencies and the POVM elements is given by

$$\hat{f}_j = \mathcal{M}(E_j) + \varepsilon, \quad (54)$$

where $\hat{f}_j = (\hat{f}_{1j}, \hat{f}_{2j}, \dots, \hat{f}_{Mj})$ are all the frequencies associated to E_j and ε represents statistical noise due to the finite sample size.

The task of POVM estimation can be formulated as a least-squares problem. The goal is to find estimators $\hat{E}_j$ that minimize the difference between the observed data $\hat{f}_j$ and the model $\mathcal{M}(E_j)$. The least-squares solution is given by

$$\hat{E}_j = (\mathcal{M}^\dagger \mathcal{M})^{-1} (\mathcal{M}^\dagger (\hat{f}_j)), \quad (55)$$

where $\mathcal{M}^\dagger : \mathbb{R}^M \rightarrow M(\mathbb{C}^d)$ is the adjoint map of $\mathcal{M}$.

To compute the solution, we leverage the 2-design property of the states $\{|\psi_i\rangle\}_{i=1}^M$. Specifically, it can be shown that the map $\mathcal{M}^\dagger \mathcal{M}$ satisfies [29]

$$\begin{aligned} \mathcal{M}^\dagger \mathcal{M}(X) &= \frac{1}{M^2} \sum_{i=1}^M \langle \psi_i | X | \psi_i \rangle |\psi_i\rangle \langle \psi_i| \\ &= \frac{X + \text{tr}(X) \mathbf{1}}{d(d+1)M}. \end{aligned} \quad (56)$$

This result simplifies the inversion of $\mathcal{M}^\dagger \mathcal{M}$. The inverse channel is given by

$$\left(\mathcal{M}^\dagger \mathcal{M} \right)^{-1} (X) = dM \left((d+1)X - \text{tr}(X) \mathbf{1} \right). \quad (57)$$

Substituting this into the least-squares solution, (55), we find that the estimators $\hat{E}_j$ take the form

$$\begin{aligned} \hat{E}_j &= \left(\mathcal{M}^\dagger \mathcal{M} \right)^{-1} (\mathcal{M}^\dagger (\hat{f}_j)) \\ &= \left(\mathcal{M}^\dagger \mathcal{M} \right)^{-1} \left(\frac{1}{M} \sum_{i=1}^M \hat{f}_{ij} |\psi_i\rangle \langle \psi_i| \right) \\ &= \sum_{i=1}^M \hat{f}_{ij} (d(d+1) |\psi_i\rangle \langle \psi_i| - d \mathbf{1}). \end{aligned} \quad (58)$$

Thus, the least-squares estimator provides a closed-form reconstruction of the POVM elements from the observed data.

A.2 Scalar Bernstein parameters for global 2-designs

Here we compute the range and variance parameters used there for global complex projective 2-design probe ensembles.

Fix a unit vector $|u\rangle \in \mathbb{C}^d$ and a sign vector $a = (a_1, \dots, a_L) \in \{\pm 1\}^L$. Define

$$g_i(u) = \langle u | \nu_i | u \rangle, \quad \nu_i = d(d+1) |\psi_i\rangle \langle \psi_i| - d \mathbf{1}, \quad (59)$$

and let $Z_{u,a}$ be the real-valued random variable such that

$$Z_{u,a}(i, j) = a_j g_i(u), \quad (60)$$

with probability $p_{ij} = \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle$.

Using $\frac{1}{M} \sum_{i=1}^M \langle \psi_i | X | \psi_i \rangle \nu_i = X$, we have

$$\begin{aligned} \mathbb{E}[Z_{u,a}] &= \sum_{i=1}^M \sum_{j=1}^L \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle a_j \langle u | \nu_i | u \rangle \\ &= \sum_{j=1}^L a_j \langle u | E_j | u \rangle. \end{aligned} \quad (61)$$

Since E is a POVM,

$$|\mathbb{E}Z_{u,a}| \leq \sum_{j=1}^L \langle u | E_j | u \rangle = 1. \quad (62)$$

We now compute the second moment of $g_i(u)$. Set $x_i = |\langle u | \psi_i \rangle|^2$. Then

$$g_i(u) = d(d+1)x_i - d. \quad (63)$$

Because $\{|\psi_i\rangle\}_{i=1}^M$ is a complex projective 2-design,

$$\frac{1}{M} \sum_{i=1}^M x_i = \frac{1}{d}, \quad \frac{1}{M} \sum_{i=1}^M x_i^2 = \frac{2}{d(d+1)}. \quad (64)$$

Therefore

$$\begin{aligned} \frac{1}{M} \sum_{i=1}^M g_i(u)^2 &= d^2(d+1)^2 \frac{1}{M} \sum_{i=1}^M x_i^2 \\ &\quad - 2d^2(d+1) \frac{1}{M} \sum_{i=1}^M x_i + d^2 \\ &= d^2. \end{aligned} \quad (65)$$

Moreover, since $0 \leq x_i \leq 1$, $|g_i(u)| \leq d^2$. Then, the centered random variable satisfies

$$|Z_{u,a} - \mathbb{E}Z_{u,a}| \leq d^2 + 1. \quad (66)$$

The variance of $Z_{u,a}$ satisfies

$$\begin{aligned} \text{Var}(Z_{u,a}) &\leq \mathbb{E}Z_{u,a}^2 \\ &= \sum_{i=1}^M \sum_{j=1}^L \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle a_j^2 g_i(u)^2 \\ &= \frac{1}{M} \sum_{i=1}^M g_i(u)^2 \\ &= d^2, \end{aligned} \quad (67)$$

where we used $a_j^2 = 1$ and $\sum_j E_j = \mathbb{1}$.

Thus, for global 2-design probe states, the scalar Bernstein parameters used in the operational-distance proof are

$$v^2 = d^2, \quad K = d^2 + 1. \quad (68)$$

B Estimator using single-qubit 2-designs

B.1 Least-squares estimator using single-qubit 2-designs

We now assume we have an n -qubit system and that the set of states $\{|\psi_i\rangle\}_{i=1}^M$ we measure is a tensor product of n single-qubit sets $\{|\psi_{i_k}\rangle\}_{i_k=1}^m$ that form a 2-design. Then, we have $M = m^n$ elements of the form

$$|\psi_i\rangle = |\psi_{i_1}\psi_{i_2}\dots\psi_{i_n}\rangle, \quad (69)$$

Since for each qubit we have a local 2-design, we can define reduced channels $\mathcal{M}_k : M(\mathbb{C}^2) \rightarrow \mathbb{R}^m$ for each qubit k . Then, the map $\mathcal{M}_k^\dagger \mathcal{M}_k$ can be calculated as

$$\mathcal{M}_k^\dagger \mathcal{M}_k(X) = \frac{X + \text{tr}(X)\mathbb{1}}{6m}, \quad (70)$$

with inverse

$$(\mathcal{M}_k^\dagger \mathcal{M}_k)^{-1}(X) = 2m(3X - \text{tr}(X)\mathbb{1}). \quad (71)$$

Following Ref. [29], the total channel is just a tensor product of the local channels,

$$\mathcal{M}^\dagger \mathcal{M}(X) = \bigotimes_{k=1}^n \frac{X + \text{tr}(X)\mathbb{1}}{6m} \quad (72)$$

with inverse

$$\begin{aligned} (\mathcal{M}^\dagger \mathcal{M})^{-1}(X) &= \bigotimes_{k=1}^n 2m(3X - \text{tr}(X)\mathbb{1}) \\ &= M \bigotimes_{k=1}^n (6X - 2\text{tr}(X)\mathbb{1}). \end{aligned} \quad (73)$$

Applying the formula for the least squares estimator, we obtain

$$\begin{aligned} \hat{E}_j &= (\mathcal{M}^\dagger \mathcal{M})^{-1}(\mathcal{M}^\dagger(\hat{f}_j)) \\ &= (\mathcal{M}^\dagger \mathcal{M})^{-1}\left(\frac{1}{M} \sum_{i=1}^M \hat{f}_{ij} |\psi_i\rangle \langle \psi_i|\right) \\ &= \sum_{i=1}^M \hat{f}_{ij} \bigotimes_{k=1}^n (6|\psi_{i_k}\rangle \langle \psi_{i_k}| - 2\mathbb{1}). \end{aligned} \quad (74)$$

B.2 Scalar Bernstein parameters for tensor products of single-qubit 2-designs

We now compute the corresponding scalar Bernstein parameters when the probe ensemble is a tensor product of single-qubit 2-designs.

As before, fix a unit vector $|u\rangle$ and a sign vector $a \in \{\pm 1\}^L$. Define

$$g_i(u) = \langle u | \nu_i | u \rangle, \quad \nu_i = \bigotimes_{k=1}^n (6|\psi_{i_k}\rangle\langle\psi_{i_k}| - 2\mathbb{1}). \quad (75)$$

and let $Z_{u,a}$ be the real-valued random variable such that

$$Z_{u,a}(i, j) = a_j g_i(u) \quad (76)$$

with probability $p_{ij} = \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle$.

The expectation is

$$\mathbb{E}[Z_{u,a}] = \sum_{j=1}^L a_j \langle u | E_j | u \rangle, \quad (77)$$

and therefore $|\mathbb{E}Z_{u,a}| \leq 1$.

We first bound the range of $g_i(u)$. For each qubit,

$$\|6|\psi_{i_k}\rangle\langle\psi_{i_k}| - 2\mathbb{1}\| = 4. \quad (78)$$

Hence

$$|g_i(u)| \leq \|\nu_i\| \leq 4^n. \quad (79)$$

This gives the centered range bound

$$|Z_{u,a} - \mathbb{E}Z_{u,a}| \leq 4^n + 1. \quad (80)$$

It remains to compute a useful second-moment bound for $g_i(u)$. Let $\mathcal{P}_n$ denote the n -qubit Pauli strings. Then, we can expand

$$|u\rangle\langle u| = \frac{1}{d} \sum_{P \in \mathcal{P}_n} r_P P, \quad r_P = \text{tr}(P|u\rangle\langle u|). \quad (81)$$

Since $|u\rangle\langle u|$ is pure $\sum_{P \in \mathcal{P}_n} r_P^2 = d$ Expand

$$\nu_i = \sum_{P \in \mathcal{P}_n} c_P(i) P. \quad (82)$$

Then

$$g_i(u) = \text{tr}(|u\rangle\langle u| \nu_i) = \sum_{P \in \mathcal{P}_n} r_P c_P(i). \quad (83)$$

For a single qubit, if the local probe state is a Pauli eigenstate $|\psi_{\alpha,s}\rangle$ with $\alpha \in \{x, y, z\}$ and $s \in \{\pm 1\}$, then

$$6|\psi_{\alpha,s}\rangle\langle\psi_{\alpha,s}| - 2\mathbb{1} = \mathbb{1} + 3s\sigma_\alpha. \quad (84)$$

Averaging uniformly over the six states gives

$$\mathbb{E}[c_{\mathbb{1}}^2] = 1, \quad \mathbb{E}[c_{\sigma_\alpha} c_{\sigma_\beta}] = 3\delta_{\alpha,\beta}, \quad \mathbb{E}[c_{\mathbb{1}} c_{\sigma_\alpha}] = 0. \quad (85)$$

Tensorizing over n qubits yields

$$\frac{1}{M} \sum_{i=1}^M c_P(i) c_Q(i) = \delta_{P,Q} 3^{w(P)}, \quad (86)$$

where $w(P)$ is the Pauli weight of P .

Using Eqs. (83) and (86), we obtain

$$\begin{aligned} \frac{1}{M} \sum_{i=1}^M g_i(u)^2 &= \frac{1}{M} \sum_{i=1}^M \left(\sum_{P \in \mathcal{P}_n} r_P c_P(i) \right)^2 \\ &= \sum_{P, Q \in \mathcal{P}_n} r_P r_Q \frac{1}{M} \sum_{i=1}^M c_P(i) c_Q(i) \\ &= \sum_{P \in \mathcal{P}_n} 3^{w(P)} r_P^2 \\ &\leq 3^n \sum_{P \in \mathcal{P}_n} r_P^2 \\ &= 3^n d = 6^n. \end{aligned} \quad (87)$$

The variance of $Z_{u,a}$ is then bounded as

$$\begin{aligned} \text{Var}(Z_{u,a}) &\leq \mathbb{E}Z_{u,a}^2 \\ &= \sum_{i=1}^M \sum_{j=1}^L \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle a_j^2 g_i(u)^2 \\ &= \frac{1}{M} \sum_{i=1}^M g_i(u)^2 \\ &\leq 6^n. \end{aligned} \quad (88)$$

Therefore, for tensor products of single-qubit 2-designs, the scalar Bernstein parameters used in the operational-distance proof are

$$v^2 = 6^n, \quad K = 4^n + 1. \quad (89)$$

C Proof of Thm. 3

In the proof, we will need the following theorem:

Lemma 10 (Vector Bernstein Inequality [52, 53]). *Let $\vec{Z}_1, \dots, \vec{Z}_N$ be independent, zero-mean random vectors in a Hilbert space $\mathcal{H}$. Suppose that $\|\vec{Z}_k\|_{\mathcal{H}} \leq K$ and $\frac{1}{N} \sum_{k=1}^N \mathbb{E}\|\vec{Z}_k\|_{\mathcal{H}}^2 \leq \sigma^2$. Then for any $t > 0$,*

$$\Pr \left(\left\| \frac{1}{N} \sum_{k=1}^N \vec{Z}_k \right\|_{\mathcal{H}} \geq t \right) \leq 2 \exp \left(- \frac{Nt^2/2}{\sigma^2 + Kt/3} \right). \quad (90)$$

Then, we can prove the following:

Theorem 3. Assume an L -outcome POVM E acting on a Hilbert space of dimension d and an IC set of states $\{|\psi_i\rangle\}_{i=1}^M$. A reconstruction error $d_{\text{av}}(E, E^*) \leq \epsilon$ with probability $1 - \delta$ can be achieved using the protocol described above with sample size

$$N \geq \frac{8L \left(d^2 + d + \frac{d\epsilon}{3\sqrt{L}} \right)}{\epsilon^2} \ln \left(\frac{4}{\delta} \right) \quad (91)$$

when $\{|\psi_i\rangle\}_{i=1}^M$ forms a 2-design, or

$$N \geq \frac{8L \left(5^n + \frac{5^{n/2}\epsilon}{3\sqrt{L}} \right)}{\epsilon^2} \ln \left(\frac{4}{\delta} \right) \quad (92)$$

if the POVM acts on an n -qubit system ($d = 2^n$) and $\{|\psi_i\rangle\}_{i=1}^M$ is a tensor product of single-qubit 2-designs.

Proof. Let $W_j = \hat{E}_j - E_j$. We first control the error of the unconstrained least-squares estimator. Using the definition of d_{av} , extended to arbitrary Hermitian L -tuples, we have

$$\begin{aligned} d_{\text{av}}(E, \hat{E}) &= \frac{1}{2d} \sum_{j=1}^L \sqrt{\|W_j\|_F^2 + (\text{tr } W_j)^2} \\ &\leq \frac{1}{2d} \sum_{j=1}^L \|W_j\|_F + \frac{1}{2d} \sum_{j=1}^L |\text{tr } W_j|. \end{aligned} \quad (93)$$

Let $T_1 = \frac{1}{2d} \sum_{j=1}^L \|W_j\|_F$ and $T_2 = \frac{1}{2d} \sum_{j=1}^L |\text{tr } W_j|$. We will show that, with probability at least $1 - \delta$, both terms are at most $\epsilon/4$. This implies $d_{\text{av}}(E, \hat{E}) \leq \epsilon/2$.

By Cauchy-Schwarz inequality, we have

$$T_1 \leq \frac{\sqrt{L}}{2d} \sqrt{\sum_{j=1}^L \|W_j\|_F^2} = \frac{\sqrt{L}}{2d} \|\vec{W}\|_{\mathcal{H}}, \quad (94)$$

where $\vec{W} = (\hat{E}_1 - E_1, \dots, \hat{E}_L - E_L)$ is treated as a single vector in the Hilbert space $\mathcal{H} = (\mathbb{C}^{d \times d})^{\oplus L}$ with the norm $\|\vec{W}\|_{\mathcal{H}}^2 = \sum_{j=1}^L \|W_j\|_F^2$.

The least-squares estimator can be written as $\hat{E} = \frac{1}{N} \sum_{k=1}^N \vec{X}_k$. For each shot k , we sample an input state $|\psi_{i_k}\rangle$ and observe an outcome j_k . The random vector $\vec{X}_k \in \mathcal{H}$ has a single nonzero block, namely $(X_k)_j = \nu_{i_k} \mathbb{1}\{j_k = j\}$. By construction, $\mathbb{E}[\vec{X}_k] = \vec{E}$, where $\vec{E} = (E_1, \dots, E_L)$. Hence

$$\vec{W} = \frac{1}{N} \sum_{k=1}^N \vec{Z}_k, \quad \vec{Z}_k := \vec{X}_k - \vec{E}, \quad (95)$$

and the vectors $\vec{Z}_k$ are independent and mean-zero.

For a global 2-design, $\|\vec{X}_k\|_{\mathcal{H}}^2 = \|\nu_{i_k}\|_F^2 = d^4 + d^3 - d^2$. Moreover,

$$\|\vec{E}\|_{\mathcal{H}}^2 \leq \sum_{j=1}^L (\text{tr } E_j)^2 \leq \left(\sum_{j=1}^L \text{tr } E_j \right)^2 = d^2. \quad (96)$$

Consequently,

$$\|\vec{Z}_k\|_{\mathcal{H}} \leq \|\vec{X}_k\|_{\mathcal{H}} + \|\vec{E}\|_{\mathcal{H}} \leq 2d^2, \quad (97)$$

for $d \geq 2$. Thus, we may take $K = 2d^2$ in vector Bernstein.

The variance parameter satisfies

$$\begin{aligned} \mathbb{E}\|\vec{Z}_k\|_{\mathcal{H}}^2 &\leq \mathbb{E}\|\vec{X}_k\|_{\mathcal{H}}^2 \\ &= \sum_{i=1}^M \sum_{j=1}^L \frac{1}{M} \langle \psi_i | E_j | \psi_i \rangle \|\nu_i\|_F^2 \\ &\leq d^4 + d^3. \end{aligned} \quad (98)$$

Here we used $\sum_j E_j = \mathbb{1}$. Therefore $\sigma^2 = d^4 + d^3$.

To guarantee $T_1 \leq \epsilon/4$, it is enough by Eq. (94) to require $\|\vec{W}\|_{\mathcal{H}} \leq \frac{d\epsilon}{2\sqrt{L}}$. Applying vector Bernstein gives

$$\Pr \left(T_1 > \frac{\epsilon}{4} \right) \leq 2 \exp \left(\frac{-Nd^2\epsilon^2/(8L)}{d^4 + d^3 + d^3\epsilon/(3\sqrt{L})} \right). \quad (99)$$

Thus $T_1 \leq \epsilon/4$ with probability at least $1 - \delta/2$ provided

$$N \geq \frac{8L \left(d^2 + d + \frac{d\epsilon}{3\sqrt{L}} \right)}{\epsilon^2} \ln \left(\frac{4}{\delta} \right). \quad (100)$$

We now bound T_2 . For both global and local 2-design frames, $\text{tr } \nu_i = d$. Therefore

$$\text{tr } \hat{E}_j = \frac{d}{N} \sum_{k=1}^N \mathbb{1}\{j_k = j\} = d\hat{p}_j, \quad (101)$$

where $\hat{p}_j$ is the empirical frequency of outcome j . Similarly, $p_j = \frac{\text{tr } E_j}{d}$ is the corresponding marginal outcome probability, obtained by averaging the input state over the 2-design. Hence

$$T_2 = \frac{1}{2} \sum_{j=1}^L |\hat{p}_j - p_j|. \quad (102)$$

Using Cauchy-Schwarz, $T_2 \leq \frac{\sqrt{L}}{2} \|\hat{\vec{p}} - \vec{p}\|_2$. Let $\vec{v}_k = \vec{e}_{j_k} - \vec{p}$. Then $\mathbb{E}[\vec{v}_k] = 0$,

$$\|\vec{v}_k\|_2 \leq 2, \quad \mathbb{E}\|\vec{v}_k\|_2^2 \leq 1. \quad (103)$$

Applying vector Bernstein in $\mathbb{R}^L$ shows that $T_2 \leq \epsilon/4$ with probability at least $1 - \delta/2$ provided

$$N \geq \frac{8L \left(1 + \frac{\epsilon}{3\sqrt{L}}\right)}{\epsilon^2} \ln \left(\frac{4}{\delta}\right). \quad (104)$$

Taking the maximum of the constraints for T_1 and T_2 yields the bound $\epsilon/2$ for the global 2-design.

An analogous derivation holds for local 2-designs. The local frame elements satisfy $\|\nu_i\|_F^2 = 5^n d^2 = \sigma^2$ and $\|\bar{Z}_k\|_{\mathcal{H}} \leq 20^{n/2} + 2^n \leq 2 \cdot 5^{n/2} d = K$. Substituting these values into the Vector Bernstein bound gives the required sample complexity for tensor products of single-qubit designs.

Finally, due to the operational properties of the projection step, we have $d_{\text{av}}(E, E^*) \leq 2d_{\text{av}}(E, \hat{E})$. Therefore, ensuring $d_{\text{av}}(E, \hat{E}) \leq \epsilon/2$ guarantees $d_{\text{av}}(E, E^*) \leq \epsilon$, concluding the proof. $\square$

D Minimum sample complexity for d_{av}

Here, we provide a lower bound for the sample complexity of non-adaptive measurement tomography considering d_{av} . The proof is based on the discretization of the problem, reducing quantum measurement tomography to the problem of discrimination of well-separated POVMs. We first construct a set of $R \in \exp(\Omega(d^2 L))$ POVMs on dimension d that are $\epsilon/4$ apart in $\sqrt{d} d_{\text{av}}$ from each other. We then encode a random message using this set and decode it using measurement tomography with sufficient precision. From Fano's inequality, this gives us a lower bound $\Omega(d^2 L)$ for the mutual information between the encoder and decoder. Additionally, we obtain an upper bound $\mathcal{O}(N\epsilon^2/d)$ for the mutual information between the parties after N uses of the POVM. Using these two results, we derive a bound $N \geq \Omega(d^2 L/\epsilon^2)$ for the sample complexity of any non-adaptive tomographic procedure using d_{av} .

D.1 Construction of an ϵ -packing

Let $\{U_j\}_{j=1}^{L/2}$ be a set of Haar-random unitaries, P a rank- $d/2$ projector and $E_{\{U_j\}}$ a POVM with

L elements given by

$$\begin{aligned} E^j &= \frac{(1 - \epsilon)}{L} \mathbb{1} + \frac{2\epsilon}{L} U_j P U_j^\dagger \\ E^{j+L/2} &= \frac{(1 + \epsilon)}{L} \mathbb{1} - \frac{2\epsilon}{L} U_j P U_j^\dagger \end{aligned} \quad (105)$$

We want to construct a large set of POVMs of this form such that $\sqrt{d} d_{\text{av}}(E_{\{U_j\}}, E_{\{V_j\}}) \geq \epsilon/4$. To do this, we will use the following [54]:

Theorem 11. *Let k and d be positive integers and $M = U_1(d) \times \cdots \times U_k(d)$ a space equipped with the L_2 -sum of Hilbert-Schmidt metrics. Suppose $F : M \rightarrow \mathbb{R}$ is κ -Lipschitz, and that $\{U_j \in U_j(d)\}$ are independent, Haar distributed unitary matrices. Then, for each $t > 0$,*

$$\begin{aligned} \Pr(F(U_1, \dots, U_k) \geq \mathbb{E}[F(U_1, \dots, U_k)] + t) \\ \leq \exp\left(-\frac{dt^2}{12\kappa^2}\right). \end{aligned} \quad (106)$$

This theorem gives us a concentration inequality for d_{av} . Then, if we choose our set of POVMs randomly, it will be ϵ -separated with high probability.

Lemma 12. *There exists a set of $R \in \exp(\Omega(d^2 L))$ POVMs that are $\epsilon/4$ apart in the distance $\sqrt{d} d_{\text{av}}$ from each other.*

Proof. We want to use Theorem 11 for

$$\begin{aligned} F &= \frac{1}{2d} \sum_{i=1}^L \|E_{\{U_j\}}^i - E_{\{V_j\}}^i\|_F \\ &= \frac{\epsilon}{dL} \sum_{j=1}^L \|U_j P U_j^\dagger - V_j P V_j^\dagger\|_F, \end{aligned} \quad (107)$$

since this is a lower bound for $d_{\text{av}}(E_{\{U_j\}}, E_{\{V_j\}})$. Defining $f_j = \|U_j P U_j^\dagger - V_j P V_j^\dagger\|_F$, we need to calculate the Lipschitz constant κ (from the relation $|F - F'| \leq \kappa \|(U_i, V_i) - (U'_i, V'_i)\|$) and $\mathbb{E}\left[\frac{\epsilon}{dL} \sum_{j=1}^L f_j\right]$. We start with the former. Using the reverse triangle inequality, we have

$$\begin{aligned} |f_j - f'_j| & \\ &\leq \|U_j P U_j^\dagger - V_j P V_j^\dagger - U'_j P U'_j{}^\dagger + V'_j P V'_j{}^\dagger\|_F \\ &\leq \|U_j P U_j^\dagger - U'_j P U'_j{}^\dagger\|_F + \|V_j P V_j^\dagger - V'_j P V'_j{}^\dagger\|_F. \end{aligned} \quad (108)$$

Then, since for unitaries W and W' we have the relation $\|W_j P W_j^\dagger - W'_j P W'_j{}^\dagger\|_F \leq 2 \|W_j - W'_j\|_F$

and the unitaries satisfy $W_j = W_{j+L/2}$ for $j = 1, \dots, L/2$, we obtain

$$\left| \sum_{j=1}^L f_j - f'_j \right| \leq 4\sqrt{L} \left(\sum_{j=1}^{L/2} \|U_j - U'_j\|_F^2 + \|V_j - V'_j\|_F^2 \right)^{\frac{1}{2}}. \quad (109)$$

This gives us an upper bound for the Lipschitz constant, $\kappa \leq \frac{4\epsilon}{d\sqrt{L}}$.

Now, we find a lower bound for the average $\mathbb{E} \left[\frac{1}{L} \sum_{j=1}^L f_j \right]$ using [37]

$$\mathbb{E}_{U,V \sim \text{Haar}} f_j \geq \frac{(\mathbb{E}_{U,V} f_j^2)^{3/2}}{(\mathbb{E}_{U,V} f_j^4)^{1/2}}. \quad (110)$$

We have

$$\begin{aligned} \mathbb{E} f_j^2 &= \mathbb{E} \text{tr} \left((U_j P U_j^\dagger - V_j P V_j^\dagger)^2 \right) \\ &= 2\mathbb{E} \text{tr}(P^2) - 2\mathbb{E} \text{tr}(U_j P U_j^\dagger V_j P V_j^\dagger) \\ &= 2 \left(\text{tr}(P^2) - \frac{\text{tr}(P)^2}{d} \right). \end{aligned} \quad (111)$$

Also,

$$\begin{aligned} \mathbb{E} f_j^4 &= 4\mathbb{E} \left(\text{tr} \left((U_j P U_j^\dagger - V_j P V_j^\dagger)^2 \right) \right)^2 \\ &= 4\text{tr}(P^2)^2 - 8 \frac{\text{tr}(P^2) \text{tr}(P)^2}{d} \\ &\quad + 4 \frac{\text{tr}(P)^4 + \text{tr}(P^2)^2}{d^2 - 1} - 8 \frac{\text{tr}(P^2) \text{tr}(P)^2}{d(d^2 - 1)}. \end{aligned} \quad (112)$$

Since the projector P has rank $d/2$, we get

$$\mathbb{E} f_j^2 = \frac{d}{2}, \quad \mathbb{E} f_j^4 = \frac{d^4}{4(d^2 - 1)}. \quad (113)$$

Then, $\mathbb{E} \left(\frac{\epsilon}{dL} \sum_{j=1}^L \|U_j P U_j^\dagger - V_j P V_j^\dagger\|_F \right) \geq \frac{\epsilon}{2\sqrt{d}}$. Using Theorem 11, we have

$$\begin{aligned} \Pr \left(\frac{\epsilon}{dL} \sum_{j=1}^L \|U_j P U_j^\dagger - V_j P V_j^\dagger\|_F \leq \frac{\epsilon}{4\sqrt{d}} \right) \\ \leq \exp \left(-\frac{d^2 L}{12 \times 64 \times 16} \right). \end{aligned} \quad (114)$$

Then, using probabilistic arguments, we can build a set of $R \in \exp(\Omega(d^2 L))$ POVMs

$$\begin{aligned} E^j &= \frac{(1 - \epsilon)}{L} \mathbb{1} + \frac{2\epsilon}{L} U_j P U_j^\dagger \\ E^{j+L/2} &= \frac{(1 + \epsilon)}{L} \mathbb{1} - \frac{2\epsilon}{L} U_j P U_j^\dagger \end{aligned} \quad (115)$$

that satisfy

$$\sqrt{d} d_{\text{av}}(E, F) \geq \frac{\epsilon}{4}. \quad (116)$$

□

D.2 Upper bound on mutual information for the average distance

Lemma 13. *Let $X \sim \text{Unif}[R]$, $\{U_j\}_{j=1}^{L/2}$ be a set of independent Haar random unitaries and $\{\rho_i\}_{i=1}^N$ a set of N quantum states. Then, there exists a set of $R \in \exp(\Omega(d^2 L))$ POVMs of the form of Eq. (105) which forms an $\epsilon/4$ packing for $\sqrt{d} d_{\text{av}}$ and satisfies*

$$I(X : Y) \leq I(\{U_j\} : Z), \quad (117)$$

where $Y = (Y_1, \dots, Y_N)$ is the outcome of measuring $\{\rho_i\}_{i=1}^N$ with a random POVM $E_{\{U_j^X\}}$ and $Z = (Z_1, \dots, Z_N)$ is the outcome of measuring $\{\rho_i\}_{i=1}^N$ with a random POVM $E_{\{U_j\}}$.

Proof. We closely follow Proposition 4.3 in Ref. [32]. Let X label a POVM of the form of Eq. (105). The average distance for two POVMs of this form of is invariant under the replacement $U_j^X \rightarrow W_j U_j^X$ for arbitrary unitary operators $\{W_j\}_{j=1}^{L/2}$, since for two unitaries U and V , we have $\|U P U^\dagger - V P V^\dagger\|_F = \|W U P U^\dagger W^\dagger - W V P V^\dagger W^\dagger\|_F$. Then, this replacement maps an arbitrary initial $\epsilon/4$ -packing to another one.

Define $Y_W = (Y_1, \dots, Y_N)$, with $Y_i \in \{1, \dots, L\}$, as the outcome random variable obtained by measuring the N copies $\{\rho_i\}_{i=1}^N$ with $E_{\{W_j U_j^X\}}$. We claim that

$$\mathbb{E}_{W_j \sim \text{Haar}} I(X : Y_W) \leq I(\{U_j\} : Z) \quad (118)$$

for a set of independent, Haar distributed unitaries $\{W_j\}_{j=1}^{L/2}$. Let $p_{Y|W,X}$ be the distribution of Y_W given $\{U_j^X\}$, with $p_{Y|W,X}(Y = y) =$

$\prod_{i=1}^N \text{tr} \left(E_{W_{y_i} U_{y_i}^X} \rho_i \right)$. We have

$$\begin{aligned} & \mathbb{E}_{\{W_j \sim \text{Haar}\}} I(X : Y_W) \\ &= \mathbb{E}_{W_j} H \left(\mathbb{E}_{X \sim [R]} p_{Y|W,X} \right) - \mathbb{E}_{W_j} \mathbb{E}_{X \sim [R]} H(p_{Y|W,X}) \\ &\leq H \left(\mathbb{E}_{X \sim [R]} \mathbb{E}_{W_j} p_{Y|W,X} \right) - \mathbb{E}_{X \sim [R]} \mathbb{E}_{W_j} H(p_{Y|W,X}). \end{aligned} \quad (119)$$

The first line is just the definition of the mutual information, and the second one follows from the concavity of entropy. By the invariance of the Haar measure, we have that $\mathbb{E}_{\{W_j \sim \text{Haar}\}} p_{Y|W,X} = \mathbb{E}_{\{W_j \sim \text{Haar}\}} p_{Y|W}$ and $\mathbb{E}_{\{W_j \sim \text{Haar}\}} H(p_{Y|W,X}) = \mathbb{E}_{\{W_j \sim \text{Haar}\}} H(p_{Y|W})$. The probability distribution of the outcomes of the random POVM $E_{\{U_j\}}$ is defined as $p_Z = \mathbb{E}_{\{U_j \sim \text{Haar}\}} p_{Y|U} = \mathbb{E}_{\{W_j \sim \text{Haar}\}} p_{Y|W}$. Using Eq. (119), this means that

$$\mathbb{E}_{\{W_j \sim \text{Haar}\}} I(X : Y_W) \leq I(\{U_j\} : Z). \quad (120)$$

Since the expectation of $I(X : Y_W)$ over all sets of unitary operators $\{W_j\}_{j=1}^{L/2}$ is at most $I(\{U_j\} : Z)$, there exists at least one set of unitary operators $\{V_j\}_{j=1}^{L/2}$ for which the inequality $I(X : Y_V) \leq I(\{U_j\} : Z)$ holds. Then we consider the set of POVMs $\{E^{\{V_i U_i^X\}}\}$. $\square$

This means that in the calculation of the mutual information we can replace the average over the unitaries that define the $\epsilon/4$ -packing by an average over Haar random unitaries to obtain an upper bound. This is given by [32]

Lemma 14. *Let $X \sim \text{Unif}([R])$ and $Y = (Y_1, Y_2, \dots, Y_N)$ be the outcome of the measurement of a random POVM $E_{\{U_X\}}$ of the form of Eq. (105) over N quantum states. Then,*

$$I(X : Y) \leq \frac{2N}{\ln(2)} \frac{\epsilon^2}{(d+1)} \quad (121)$$

Proof. We have

$$I(X : Y) \leq I(\{U_j\} : Z) \quad (122)$$

$$= \sum_{i=1}^N I(\{U_j\} : Z_i | Z_{i-1}, \dots, Z_1) \quad (123)$$

$$\leq \sum_{i=1}^N I(\{U_j\} : Z_i) \quad (124)$$

$$\leq \frac{N}{\ln(2)} \mathbb{E}_{\{U_j\}} \left(\sum_{z_i} \frac{p_{Z_i|U_{z_i}}(z_i)^2}{p_{Z_i}(z_i)} - 1 \right), \quad (125)$$

where the first inequality follows from Lemma 13, the second one from the chain rule for mutual information, the third one from the independence of $Z_1, \dots, Z_N$ given $\{U_j\}$ and the final one from Lemma 6 and the independence of $\{U_j\}$.

The probability $p_{Z_i}(z_i) = \mathbb{E}_{\{U_j\}} p_{Z_i|\{U_{z_i}\}}$ is of the form

$$\begin{aligned} p_{Z_i}(z_i) &= \mathbb{E}_{U_i} \text{tr} \left(\left(\frac{(1 \pm \epsilon)}{L} \mathbb{1} \mp \frac{2\epsilon}{L} U_i P U_i^\dagger \right) \rho \right) \\ &= \frac{1}{L} \end{aligned} \quad (126)$$

For the other term, we have

$$\begin{aligned} \mathbb{E}_{U_i} p_{Z_i|U_i}(z_i)^2 &= \mathbb{E}_{U_i} \text{tr} \left(\left(\frac{(1 \pm \epsilon)}{L} \mathbb{1} \mp \frac{2\epsilon}{L} U_i P U_i^\dagger \right) \rho \right)^2 \\ &\leq \frac{1}{L^2} \left(1 + \frac{2\epsilon^2}{(d+1)} \right). \end{aligned} \quad (127)$$

Then, using Lemma 6, we obtain

$$\begin{aligned} I(X : Y) &\leq \frac{N}{\ln(2)} \left(\sum_{z_i=1}^L \frac{1}{L} \left(1 + \frac{2\epsilon^2}{(d+1)} \right) - 1 \right) \\ &= \frac{2N}{\ln(2)} \frac{\epsilon^2}{(d+1)}. \end{aligned} \quad (128)$$

$\square$

D.3 Lower bound on the sample complexity of d_{av}

Theorem 15. *Any procedure for quantum measurement tomography of a POVM on a d -dimensional Hilbert space that is $\epsilon/8$ accurate in average distance using nonadaptive, single-copy measurements on known input states requires*

$$N \in \Omega \left(\frac{d^2 L}{\epsilon^2} \right) \quad (129)$$

uses of the unknown POVM.

Proof. Similar to the calculation of the lower bound for d_{op} , assume a random message is encoded in N copies of a POVM $E_{\{U_j^x\}}$, where this POVM is uniformly sampled from the $\epsilon/4$ -packing defined by Lemma 12. Let $Y = (Y_1, \dots, Y_N)$ denote the outcomes from measuring the random POVM over N quantum states. Since each POVM in the packing is separated by at least $\epsilon/4$, a tomography algorithm that takes this data and produces an estimate of the POVM within $\epsilon/8$ precision in $\sqrt{d}d_{\text{av}}$ (with some constant probability) is sufficient to decode the message. Then, tomography must have a sample complexity at least as large as the discrimination problem. From the previous section, the mutual information between the measurements used for tomography and the message is upper bounded by

$$I(X : Y) \leq \frac{2N}{\ln(2)} \frac{\epsilon^2}{(d+1)}. \quad (130)$$

Using Lemma 1, we also have

$$I(X : Y) \geq \Omega(d^2 L). \quad (131)$$

Combining these results, we obtain

$$\frac{2N}{\ln(2)} \frac{\epsilon^2}{(d+1)} \geq I(X : Y) \geq \Omega(d^2 L). \quad (132)$$

Thus, we conclude that

$$N \geq \Omega\left(\frac{d^3 L}{\epsilon^2}\right) \quad (133)$$

to obtain precision ϵ in $\sqrt{d}d_{\text{av}}$. Then, to obtain precision ϵ in d_{av} , we need at least

$$N \geq \Omega\left(\frac{d^2 L}{\epsilon^2}\right). \quad (134)$$

This matches the upper bound. $\square$

E Experimental details

The experiment is deployed on a subset of two connected qubits within a 5 qubit chip. It is a star topology flux-tunable superconducting transmon device manufactured by Quantware and hosted in the quantum computing lab of Technology Innovation Institute. The device is driven via microwave pulses generated via a Quantum Machines OPX system, orchestrated by the open

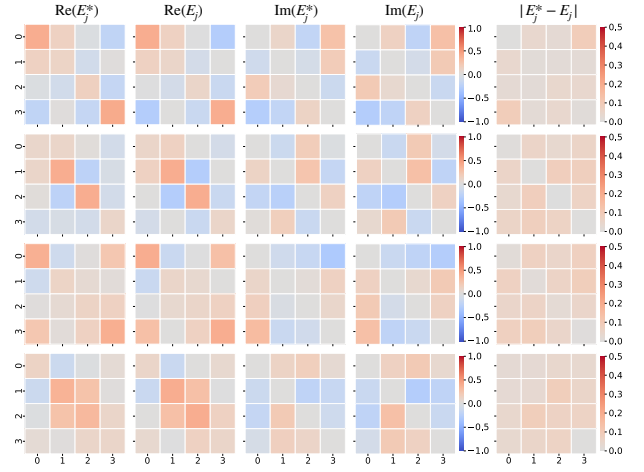


Figure 5: Results for the reconstruction of the full system implementing a noisy two-qubit Symmetric Informationally Complete (SIC-)POVM, implemented on a two-qubit flux-tunable transmon device using a budget of 10^6 random initial states. The reconstructed noisy POVM is compared to its expected noiseless counterpart, with the absolute difference displayed. Though noisy, the reconstruction follows the target POVM closely. The deviations from the expected values are due to the noisy device implementation.

source library Qibolab [48]. The system is calibrated using the available routines on the open source Qibocal [49] package.

The two qubits are calibrated up to usable specifications, as a reduced amount of noise is desirable for our presented algorithms to reconstruct the noisy POVMs. The two qubits are calibrated before launching the experiment up to the following parameters: the relaxation times for the first and second qubits are $T_1^{(1)} = 33.(6) \mu s$ and $T_1^{(2)} = 30.(6) \mu s$ respectively. Equally for the decoherence times $T_2^{(1)} = 13.(0) \mu s$ and $T_2^{(2)} = 9.(1) \mu s$. Assignment fidelity for a computational basis readout of $F_{ro}^{(1)} = 96.(9) \%$ and $F_{ro}^{(2)} = 98.(0) \%$. Lastly, the single qubit native pulse infidelity, as extracted from a randomized benchmarking experiment, of $g_{inf}^{(1)} = 1.3(7) \cdot 10^{-3}$ and $g_{inf}^{(2)} = 6.(7) \cdot 10^{-3}$ for native pulse length of $40 ns$.

The experiments are designed directly in Qibolab language to leverage the lower level logic of the device. Everything is compiled down to the native pulses and interactions of the target quantum system. Code to reproduce the experiment is available upon reasonable request.

F Further results

Here we expand on the reconstruction of noisy experimental POVM implementation on a two-qubit superconducting flux-tunable transmon system. First we give details on the implementation of the SIC-POVM presented in the main text, such as circuits and further reconstructions. After, we present results on other POVMs, such as the Identity, a Hadamard gates or Haar random gates on both qubits, as well as a Bell measurement.

We create the SIC-POVM with a quantum circuit of the form shown in Fig. 6. The data qubit, where the initial state will be prepared is the second qubit in Fig. 6. The parameters for the single qubit rotations are taken as $\theta = \arccos(1/\sqrt{3})$ and $\phi = 3\pi/4$ [50] to generate a SIC-POVM [51]. Furthermore, we also characterize this POVM as a two-qubit measurement, instead of a projection into a single qubit system. In Fig. 5 we showcase the reconstruction. The results are similar to the ones in the main text, but in an extended space. Indeed, the single qubit POVM is located in the top left quadrant of the shown data.

In Fig 7 we collect all the data from the other POVM settings. First, on the top left are the results for the Identity POVM, that is, a direct measurement in the computational basis. While a very straightforward POVM, this reconstruction provides a great deal of information about the readout process of the device. For example, the non-symmetry in readout error when measuring different final states can be extracted from the data shown in this figure. On the top right, we plot the reconstruction results of a POVM measurement with a Hadamard gate before the readout, that is a measurement in the X basis — where we consider the usual convention of the computational basis being the Z direction. The bottom-left panel showcases a similar scenario,

but with a single qubit Haar random gate before the readout. Unlike the other instances, in this one we notice how the reconstructed POVM is randomly spread across the real and imaginary space, without a directly apparent structure. Finally, at the bottom right corner of Fig. 7 the results of the reconstruction of a noisy Bell POVM are shown.

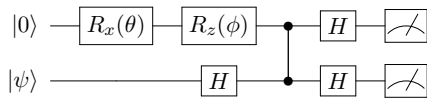


Figure 6: Quantum circuit depicting the implementation of the SIC-POVM. When the single gate rotation parameters are $\theta = \arccos(1/\sqrt{3})$ and $\phi = 3\pi/4$, this results in a Symmetric Informationally Complete measurement. The choice of a CZ gate as the entangling operation is due to the native interactions of the experimental device where this protocol is deployed.

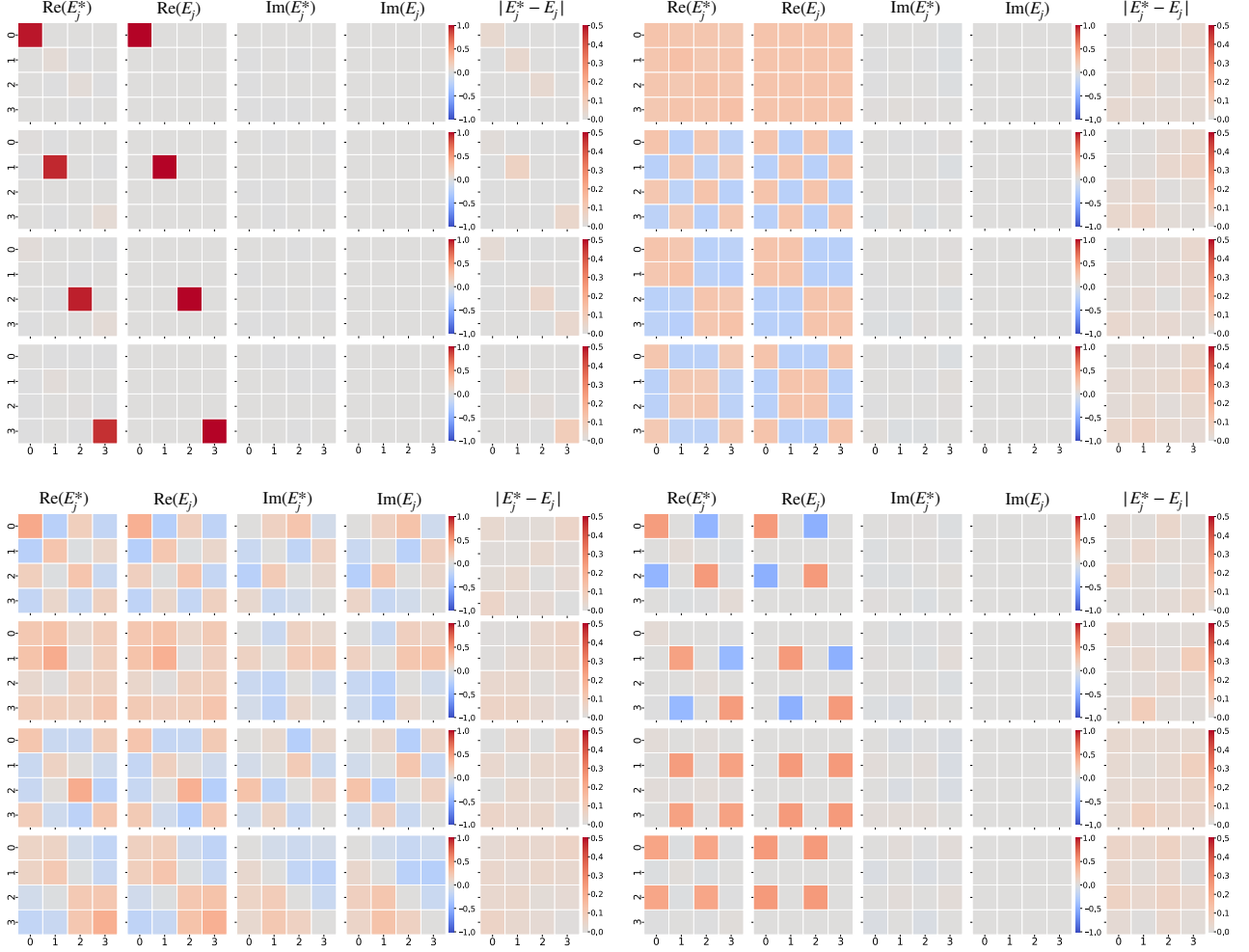


Figure 7: Results for the reconstruction of various noisy two-qubit POVMs, implemented on a two-qubit flux-tunable transmon device using a budget of 10^6 random initial states. The reconstructed noisy POVMs are compared to its expected noiseless counterpart, with the absolute difference displayed. Though noisy, the reconstruction follows the target POVMs closely. The deviations from the expected values are due to the noisy device implementation. More precisely, (top left) depicts the Identity POVM, where only a measurement in the computational basis is performed on both qubits. (top right) depicts a POVM where a Hadamard gate is applied to each qubit before measurement, that is, a change of basis from Z to X measurement basis. (bottom left) POVM where a Haar random gate is implemented before the measurement gate, we notice the how even with a more complex measurement the noisy implementation follows closely the noiseless simulation. Finally, on (bottom right) a Bell type POVM is reconstructed, where unlike the previous three POVMs depicted in this figure, an entangling gate is performed between the two qubits.