Universal framework for simultaneous tomography of quantum states and SPAM noise

Abhijith Jayakumar[1], Stefano Chessa[1,2,3], Carleton Coffrin[4], Andrey Y. Lokhov[1], Marc Vuffray[1], and Sidhant Misra[1]

[1]Theoretical Division, Los Alamos National Laboratory, 87545, NM, USA
[2]NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56126, Pisa, Italy
[3]Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, 61801, IL, USA
[4]Los Alamos National Laboratory, Los Alamos, 87545, NM, USA

We present a general denoising algorithm for performing *simultaneous tomography* of quantum states and measurement noise. This algorithm allows us to fully characterize state preparation and measurement (SPAM) errors present in any quantum system. Our method is based on the analysis of the properties of the linear operator space induced by unitary operations. Given any quantum system with a noisy measurement apparatus, our method can output the quantum state and the noise matrix of the detector up to a single gauge degree of freedom. We show that this gauge freedom is unavoidable in the general case, but this degeneracy can be generally broken using prior knowledge on the state or noise properties, thus fixing the gauge for several types of state-noise combinations with no assumptions about noise strength. Such combinations include pure quantum states with arbitrarily correlated errors, and arbitrary states with block independent errors. This framework can further use available prior information about the setting to systematically reduce the number of observations and measurements required for state and noise detection. Our method effectively generalizes existing approaches to the problem, and includes as special cases common settings considered in the literature requiring an uncorrelated or invertible noise matrix, or specific probe states.

1 Introduction

Quantum computing promises to have the potential to solve complex problems that are beyond the reach of classical computers [1, 25, 54, 71], but realizing this full potential requires overcoming the various challenges posed by noise [5, 13, 57]. These errors can arise from a number of sources, including noise in the specific hardware architectures [6, 7, 10, 28, 33, 37, 61, 63], inaccuracies in control and limitations in the operations that can actually be performed on such systems [2, 17, 19, 58, 67].

To tame these errors, researchers have developed various and still growing number of approaches and strategies that can be included in the macro-categories of quantum error correction [9, 14, 16, 22, 35, 39, 56, 59, 62], quantum error mitigation [8, 21, 31, 64, 65], and noise learning, which includes specific techniques such as, among others, quantum process tomography [15, 47, 48, 55], gate set tomography [24, 51] and randomized benchmarking [27, 34, 43, 44].

Among these sources of noise, state preparation and measurement (SPAM) errors can prove to be particularly significant. As an example for the current best superconducting qubit-based devices, they can be in the range 1-3%, see e.g. [3, 20, 53]. These errors occur when the initial state of a quantum system and/or the measurement of its final state are not precisely known or controlled. SPAM errors can result in systematic biases that can greatly impact the accuracy of quantum information processing in noisy devices both in quantum error correction and in the so-called "noisy intermediate scale quantum" (NISQ) tasks see e.g. [3, 60, 70]. SPAM errors are the focus of this paper, specifically, we address the issue of the simultaneous correct identification of the (possibly arbitrarily correlated and of arbitrary strength) noise affecting detectors after the preparation of a state ρ *and* the correct identification of ρ itself. Despite these two tasks being some of the most fundamental operations one could imagine for quantum information processing, their simultaneous realization is hindered by the fact that state preparation and measurement noise matrix can be determined only up to a gauge transformation [4, 30, 40]. This fact presents a severe limitation for state tomography and noise characterization as the knowledge of the real underlying noise process is essential for diagnostics and the optimization of the device. To address this in recent years some attempts

Abhijith Jayakumar: abhijithj@lanl.gov
Stefano Chessa: schessa@illinois.edu
Sidhant Misra: sidhant@lanl.gov

have been made to develop techniques that resolve these kinds of gauge degeneracy [38, 41, 42, 46].

Our work presents a significant contribution in this direction: we provide a general framework for identifying conditions under which noise models and prepared states can break the gauge freedom, which includes as special cases many previously proposed approaches. We achieve this by introducing a denoising algorithm that can simultaneously estimate both the state of the system and detector noise up to a single gauge parameter. The output of this algorithm gives a complete characterization of the SPAM errors in the system as it gives the maximum possible information about the true state prepared in the system and the stochastic matrix governing the measurement noise.

The main contributions of this work are as follows: First, we completely characterize the gauge freedom in our problem, and prove that the simultaneous characterization of state and noise is only hindered by a single gauge parameter. Next, we give a general algorithm to simultaneously estimate a quantum state and any stochastic matrix characterizing SPAM errors in a quantum system, up to this unavoidable gauge parameter. We also outline methods using which this gauge can be fixed given many forms of prior information about the state or the noise matrix, including practically relevant cases, such as states with known purity, independent ancilla qubits, and known expectation values. To address more practical settings, we devise a randomized version of our algorithm that uses computational basis measurements that only involves the application of Clifford circuits. Finally, we also provide a sample complexity analysis of our algorithm and show that the number of samples required depends naturally on the distance of the state and the noise matrix from a maximally mixed case. The paper is structured as follows:

- In Sec. 2: we state the problem of noise-state *simultaneous tomography*, set the notation, and discuss the gauge freedom intrinsic in the problem. Here, we prove that the problem has only a single gauge degree of freedom.

- In Sec. 3: we outline the noise-state simultaneous tomography algorithm for any POVM. We also show the special case of the algorithm using computational basis measurements and derive the sample complexity of the randomized version. We support our analysis of the randomized algorithm using numerical results that show the tightness of our analysis.

- In Sec. 4: we show how prior knowledge about the system can be used to fix the gauge and also to improve the algorithm in terms of resource efficiency.

- In Sec. 5: we draw the conclusions and discuss the perspectives of this work.

The summary of this structure and our approach is provided in Fig. 1.

2 Problem statement, setting, and notation

2.1 The problem of simultaneous tomography

We consider the problem of fully characterizing persistent errors as well as recovering the underlying quantum state in a quantum system affected by imperfect state preparation protocols or measurement errors. The quantum system consisting of n qubits is prepared in a state ρ and measured using a general *Positive Operator Valued Measure* (POVM) [52]. Given the $n-$qubit POVM, $\{M_k | \ k \in [D], \ M_k \succcurlyeq 0, \ \sum_k M_k = I\}$, we define the measurement probabilities obtained after applying a unitary transformation (U) to the quantum state as follows,

$$y_k(U) := \mathrm{Tr}(U\rho U^\dagger M_k), \quad k = 1, \ldots, D. \tag{1}$$

Now we model the measurement noise in the quantum system as a general stochastic matrix, A, acting on the probability distributions defined in (1).

$$\tilde{y}_k(U) := \sum_{k' \in [D]} A_{k\,k'} \ y_{k'}(U). \tag{2}$$

We will simplify the notation to y_k and $\tilde{y}_k$ in the case where U is just the identity transformation. This transition matrix (A) model is quite universal as it models a general measurement error one can have assuming that this error is independent of the operations performed on the computer before measurement. This is a commonly assumed simplification in the SPAM literature [30, 38, 42]. We also assume that any unitary U in the above equations can be applied without any errors.

The main aim of this work will be to use these types of noisy measurements to fully characterize the system and the noise. We refer to this task as *simultaneous tomography*, which can be defined as the problem of designing a set of unitary operators $U_1, \ldots, U_l$ that are efficiently implementable on a given quantum system, and a procedure that uses the noisy measurements $\tilde{y}_k(U_1), \ldots, \tilde{y}_k(U_l)$ along with prior information about the system to estimate the state ρ and the matrix of measurement noise A.

Simultaneous tomography is directly related to SPAM error characterization. The recovered state ρ can be compared with the state that was intended to be prepared and the state preparation error rates can be computed from their difference [42]. While the noise matrix A represents the errors in measurement

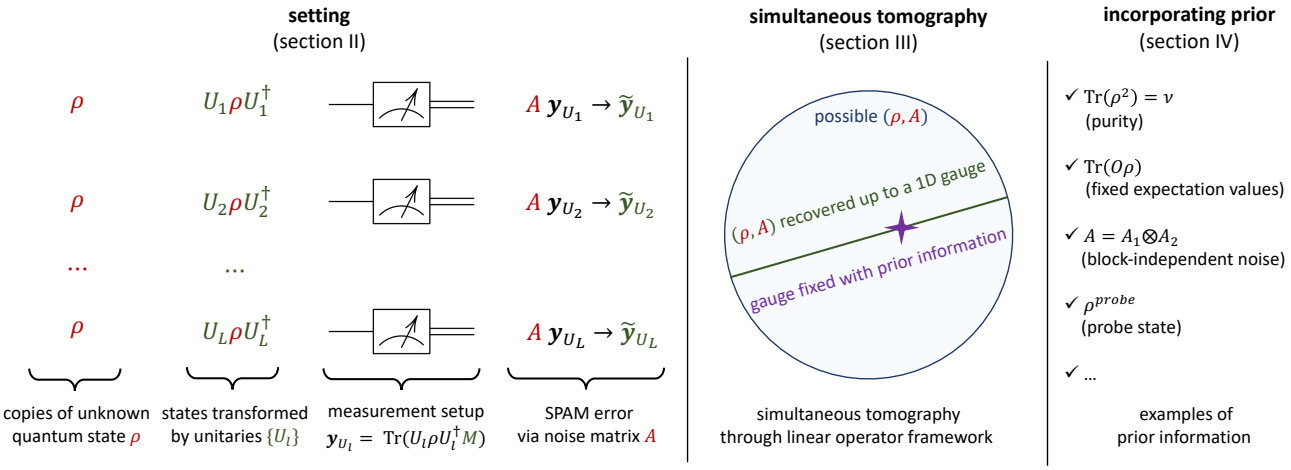

Figure 1: **Summary of our approach to simultaneous tomography of quantum states and SPAM noise.** In section II of the paper, we state the problem of *simultaneous tomography*: the process of estimating the quantum state (ρ) and the noise matrix associated with measurement errors (A) using a unified set of measurements. In section III, we introduce a universal algorithm for performing the simultaneous tomography in full generality up to the fundamental and unavoidable gauge ambiguity, prove that this degeneracy is one-dimensional, and discuss the sample-complexity of our algorithm. Finally, in section IV, we provide many examples of prior information about either the state or the noise that allows one to unambiguously recover the quantum state and the noise matrix. These examples include many settings considered in previous work.

device.

The number of gates, measurements, and classical processing required to perform simultaneous tomography is expected to be larger than those for noiseless state tomography. The exact overhead depends on the structure of the noise, the underlying state, and access to prior knowledge of the state and noise model. In practice, the best choice of gates $U_1, \dots U_L$ will depend on which gates are native and least noisy for the specific quantum system in consideration.

2.2 Linear operator framework

The process of simultaneous tomography consists of two steps: (i) implementing a set of chosen unitaries on the quantum system and obtaining the corresponding noisy measurements, and (ii) performing a set of classical post-processing computations on the measurements to obtain the estimates of the state and measurement noise. By considering only linear classical post-processing, the overall procedure can be viewed as a linear transformation on the underlying state which we describe below.

In a quantum system, the action of any unitary on a state ($\rho \to U\rho U^\dagger$) can be represented by a linear *superoperator*. To demarcate between operators and superoperators, we use the standard notation of $|\rho\rangle\rangle$ for the 4^n-dimensional vector representing ρ in the space acted on by superoperators [51]. Other objects, such as the $2^n \times 2^n$ identity matrix I or POVM operators, can be represented by a 4^n-dimensional vector in a similar way. Naturally, for two operators P and Q the inner-product $\langle\langle P|Q\rangle\rangle$ is defined as $\langle\langle P|Q\rangle\rangle = \mathrm{Tr}(P^\dagger Q)$.

It is advantageous to isolate the action of a unitary on the traceless subspace of operators,

$$U\rho U^\dagger = \Phi(U)\,|\rho\rangle\rangle = \frac{|\hat{I}\rangle\rangle}{2^{n/2}} + \phi(U)\,|\bar{\rho}\rangle\rangle\,. \tag{3}$$

Here, $\Phi(U)$ is the complete superoperator corresponding to the action of U, $\phi(U)$ is the traceless part of this superoperator. We also define $\hat{I} = I/2^{n/2}$, the normalized identity matrix, as well as $|\bar{\rho}\rangle\rangle = |\rho\rangle\rangle - |\hat{I}\rangle\rangle/2^{n/2}$ for the 4^n-dimensional vector representing the traceless part of ρ.

In this notation, the noisy measurements take the form,

$$\tilde{y}_k(U) = \sum_{k' \in [D]} A_{k\,k'} \left(\frac{\langle\langle M_{k'}|\hat{I}\rangle\rangle}{2^{n/2}} + \langle\langle M_{k'}|\,\phi(U)\,|\bar{\rho}\rangle\rangle \right). \tag{4}$$

In general, this expression can be expanded using any basis in the traceless subspace. Let $\mathcal{B}_L$ and $\mathcal{B}_R$ be two sets of traceless and hermitian operators. Further, assume that $\mathcal{B}_R$ is a normalized set of operators (i.e. $\langle\langle P|P\rangle\rangle = 1$) whose linear span is the space of all traceless hermitian operators. Also, assume that the measurement operators lie in the linear span of $\mathcal{B}_L$. Then we can always expand the state in one of the basis sets, and the measurement operators in the other as follows:

$$|\bar{\rho}\rangle\rangle = \sum_{P \in \mathcal{B}_R} s_P\,|P\rangle\rangle\,, \tag{5}$$

$$m_{k\,I} = \langle\langle M_k|\hat{I}\rangle\rangle, \quad m_{k\,Q} = \langle\langle M_k|Q\rangle\rangle\,,\ Q \in \mathcal{B}_L\,. \tag{6}$$

Using these relations we can expand (4) in this specific basis as,

$$\tilde{y}_k(U) = \sum_{k' \in [D]} \frac{A_{k\,k'} m_{k'I}}{2^{n/2}} + \sum_{\substack{k' \in [D], \\ P \in \mathcal{B}_R, Q \in \mathcal{B}_L}} s_P \, \phi(U)_{P\,Q} \, A_{k\,k'} m_{k'\,Q} \, . \quad (7)$$

Here $\phi(U)\,|P\rangle\rangle = \sum_{Q \in \mathcal{B}_L} \phi(U)_{P\,Q}\,|Q\rangle\rangle + |b_L^{\perp}\rangle\rangle$, where $|b_L^{\perp}\rangle\rangle$ is an operator orthogonal to every operator in $\mathcal{B}_L$.

As an example of the setup described above, take the set of all n-qubit normalized Pauli strings, $\hat{\mathcal{P}} \equiv \{\frac{I}{\sqrt{2}}, \frac{X}{\sqrt{2}}, \frac{Y}{\sqrt{2}}, \frac{Z}{\sqrt{2}}\}^{\otimes n}$. We can take the basis sets to be the traceless operators in this set, $\mathcal{B}_L = \mathcal{B}_R = \hat{\mathcal{P}} \setminus \{\hat{I}\}$. In this case, the matrix $\phi(U) \in \mathbb{R}^{4^n-1 \times 4^n-1}$ will just be the well-known Pauli Transfer Matrix representation for U. And s_P would simply be expectation values of the state with the Pauli strings [42].

Running example: To illustrate the ideas in this paper we will use an example of a noisy two-qubit system. The same system will be used throughout the paper at various points as a pedagogical tool.

Consider a 2-qubit system with $\rho = |01\rangle\langle 01| = \frac{(I+Z)\otimes(I-Z)}{4}$. If $\mathcal{B}_R$ is the normalized Pauli basis, then the non-zero coefficients are $s_{I\otimes I}, s_{Z\otimes I} = \frac{1}{2}$, $s_{I\otimes Z}, s_{Z\otimes Z} = -\frac{1}{2}$. For measurements in the computational basis the noise matrix is a 4×4 matrix which we take to be $A = (0.9I + 0.1X)^{\otimes 2}$.

To perform simultaneous tomography, the noisy measurements $\tilde{y}_k(U)$ are passed through *linear classical post-processing* where we compute linear combinations

$$z_k = \sum_l c_l \tilde{y}_k(U_l) \, , \quad \text{where } \sum_l c_l = 1 \, . \quad (8)$$

Using (4) the quantities z_k can be expressed in a basis independent fashion as,

$$z_k = \sum_{k' \in [D]} \frac{A_{k\,k'} m_{k'\,I}}{2^n} + \sum_{k' \in [D]} A_{k\,k'} \langle\langle M_{k'}| \left(\sum_l c_l \phi(U_l) \right) |\vec{\rho}\rangle\rangle. \quad (9)$$

Thus computing the quantities z_k can be viewed as applying the *effective* non-unitary linear transformation

$$\Phi = \sum_l c_l \Phi(U_l) \quad (10)$$

on the state ρ and then obtaining noisy measurements. The affine constraint ($\sum_l c_l = 1$) makes these transformations trace-preserving.

In the context of simultaneous tomography, the following two points are important regarding these linear operators. First, which of these linear transformations are sufficient for successfully performing simultaneous tomography? Second, what set of unitaries is required for efficiently implementing the linear transformations (9)?

To this end, let $\mathcal{U}(2^n)$ be the unitary group on n qubits. For a chosen subset, $\mathcal{S} = \{U_1, \ldots, U_l\} \subseteq \mathcal{U}(2^n)$, the overall computational power of implementing them on the quantum system and performing classical linear post-processing can be summarized by the linear operator space defined by.

$$\mathcal{L}(\mathcal{S}) = \left\{ \sum_l c_l \Phi(U_l) \,\middle|\, U_l \in \mathcal{S}, \; \sum_l c_l = 1 \right\}. \quad (11)$$

Using the linear operator space allows us to view the requirements of a given simultaneous tomography task in a general way without reference to a chosen basis or a given set of gates. If it is determined that a such subset of unitaries is sufficient for a given simultaneous tomography task, then depending on the quantum system there may be multiple ways of realizing this subset. Note that $\mathcal{L}$ does not correspond to any single quantum channel, rather it represents a set of quantum channels.

We call a set of unitaries $S_g \subseteq \mathcal{U}(2^n)$ the *generator set* for a given subset $S \subseteq \mathcal{U}(2^n)$ if S_g has fewer elements than S, and $\mathcal{L}(S) \subseteq \mathcal{L}(S_g)$. Notice that there might be multiple ways to choose S_g given S. The appropriate choice will depend for example on what set of gates are least noisy and natively available on a given quantum architecture and how much classical post-processing power is available. Working with the super-operator space $\mathcal{L}(S)$ allows us to separate *what* is needed to perform simultaneous tomography and *how* to realize it with a given quantum device and classical processing resource.

For instance, we will show that using a complete set of superoperators $\mathcal{L}(\mathcal{U}(2^n))$ is sufficient to perform simultaneous tomography. But for performing simultaneous tomography in the computational basis, we can also use a smaller subset of the Clifford group as a generator set for this super-operator space.Using only CNOT and arbitrary single qubit gates, this generator set can be implemented with circuits of linear depth [45]. We will also discuss some cases of performing this task in the presence of prior information where a limited subset $\mathcal{L}(S) \subset \mathcal{L}(\mathcal{U}(2^n))$ is sufficient.

2.3 Identifiability for the simultaneous tomography problem

Given noisy observations of the form in (7), the pertinent question is whether simultaneous tomography of ρ and A is even possible without any additional information? Interestingly, the answer is "no" in the most general case. To see this, consider a one-parameter family of transformations on the state and noise channel defined as follows,

$$A_{k\,k'} \to A'_{k\,k'}(\alpha) = \alpha A_{k\,k'} + (1-\alpha)\frac{\sum_{j\in[D]} A_{k\,j} m_{j\,I}}{2^{n/2}}\,,$$
$$\rho \to \rho'(\alpha) = \frac{\rho}{\alpha} + \left(1 - \frac{1}{\alpha}\right)\frac{I}{2^n}\,, \quad \alpha \in \mathbb{R}\setminus\{0\}\,. \quad (12)$$

By simple algebra, we can check that this simultaneous transformation of the state and noise will leave the noisy outputs in (7) invariant thus leaving us with no means to distinguish between them. In literature, this kind of invariance has been called *gauge freedom* [8, 30, 40]. The gauge freedom implies that any simultaneous tomography method will have at least a one-parameter ambiguity. These gauge transformations represent a one-parameter manifold in the (ρ, A) space. While the transformations are mathematically well defined for any non-zero α, the set of physically allowed α will be those such that $\rho'(\alpha), A'(\alpha)$ are respectively valid density and stochastic matrices. But even these physical constraints cannot unambiguously fix α in general.

The gauge freedom can be viewed as the inability to separate whether the randomness in the observations comes from the random nature of quantum measurements or if it is a product of classical noise. An extreme example is as follows; suppose we are given a single qubit state and a noisy measurement apparatus. Suppose we also observe that when this qubit is measured in the computational basis after applying any U, both 1 and 0 are seen with equal probability. Given such a system there is no way to distinguish whether the state is maximally mixed or whether it is the measurement device that is completely noisy. However, if we have prior information (confidence about the state preparation itself) that the state is pure, we can ascertain that the randomness came from the measurement device. The gauge freedom in the simultaneous tomography problem generalizes this inherent ambiguity in the problem.

The question remains whether there are other transformations that also leave (7) invariant. The theorem below shows that the transformation in (12) represents the only possible ambiguity in the problem.

Theorem 1. Gauge freedom is the only ambiguity.
Let $\tilde{\mathbf{y}}_{A,\rho}(U)$ be the noisy measurement distribution produced by the quantum state $U\rho U^\dagger$, with the noise characterized by A, as in (7). If for another system in a state ρ' with noisy measurements characterized by A', it is given that $\tilde{\mathbf{y}}_{A,\rho}(U) = \tilde{\mathbf{y}}_{A',\rho'}(U),\ \forall\ U \in \mathcal{U}(2^n)$, then there must exist a gauge parameter $\alpha \in \mathbb{R}\setminus\{0\}$ such that (12) holds.

The proof of this theorem rests on the fact that when we have access to all possible unitary gates in $\mathcal{U}(2^n)$, the induced linear operator space $\mathcal{L}(\mathcal{U}(2^n))$ defined in (11) is *complete*. The precise statement of this completeness result is given in Appendix A. The full proof of the theorem can be found in Appendix B.

This gauge ambiguity can be overcome if we have some prior information about the system that uniquely identifies the correct ρ and A from the one-parameter family in (12). In Sec. 4, we show multiple, physically and operationally relevant cases of prior information that can fix this gauge.

3 Simultaneous tomography: conditions and algorithm

In this section, we will demonstrate how noisy measurements generated according to (7), can be used to reconstruct both ρ and A up to a single gauge parameter. As in the case of noiseless tomography, simultaneous tomography can also be performed by using measurement outcomes produced by observing the state after rotating it using a set of pre-defined unitary operations.

3.1 Sufficient conditions for simultaneous tomography

Beyond the gauge degree of freedom, few edge cases can make simultaneous tomography impossible. For instance, if the A matrix always outputs the uniform distribution in D dimensions, then we can never recover the exact state ρ from the noisy measurement outcomes. To avoid these types of pathological cases we assume that the output of A always has some correlation with the input:

Condition 1. *$\exists k, i, j \in [D]$, such that $A_{ki} \neq A_{kj}$.*

If this condition does not hold, then the probability of observing a certain output conditioned on an input, $Pr(k|k') = A_{k\,k'}$, would be independent of the input. We call such an A the *erasure channel*.

Similarly, simultaneous tomography is impossible if ρ is a *maximally mixed state* ($\rho \propto I$). This would imply that $U\rho U^\dagger = \rho$, and full information about A would not be recoverable from noisy measurements defined in (7). To avoid this case we must assume that the state has some non-zero overlap with the space of traceless operators i.e. at least one of the s_P coefficients is non-zero

Condition 2. $\exists P \in \mathcal{B}_R$, *such that*, $s_P \neq 0$.

Additionally, we also require the set of measurement operators to be linearly independent:

Condition 3. $\{M_i | i \in [D]\}$ *are linearly independent.*

If this condition is not satisfied, then the definition of the measurement operators are non-unique and it is impossible to reconstruct A. But given a linearly dependent POVM, we can always construct a reduced set from it such that this new POVM is linearly independent (see Appendix C).

To describe our algorithm, in 3.2, we assume that the noisy measurement probabilities are directly available to us, i.e., for any U the distribution $\tilde{\mathbf{y}}(U)$ is fully specified. We will discuss the more practical variant of our algorithm with finite measurement shots and randomized measurements in 3.3.

3.2 Simultaneous tomography algorithm

The algorithm relies on the *completeness* of the linear operator space used in the proof of Theorem 1. The proof is a constructive one and naturally leads to the algorithm described in this section.

While simultaneous tomography can be performed on any basis in the operator space, we find that the presentation of the algorithm simplifies considerably if we fix $\mathcal{B}_L$ to be the traceless POVM operators,

$$\mathcal{B}_L = \{\bar{M}_i | i \in [D]\}, \tag{13}$$

where, $\bar{M}_i = M_i - \langle\langle M_i | I \rangle\rangle \frac{I}{D}$.

If $\mathcal{B}_L$ does not span the space of traceless operators, there will be a space orthogonal to it which is unobservable by the POVM. We denote this orthogonal space by $\mathcal{B}_L^\perp$. As an example, if $\mathcal{B}_L$ is given by the traceless computational basis measurement operators, then $\mathcal{B}_L^\perp$ will span the space of all off-diagonal operators:

$$\mathcal{B}_L^\perp = \{Q | \mathrm{Tr}(Q) = 0,\ \langle\langle Q | Q' \rangle\rangle = 0\ \forall Q' \in \mathcal{B}_L,\ Q = Q^\dagger\}. \tag{14}$$

Notice that while $\mathcal{B}_L$ is a basis set, $\mathcal{B}_L^\perp$ is a vector space.

Running example: For the two qubit system measured in the computational basis $\mathcal{B}_L = \{|00\rangle\langle 00| - \frac{I}{4},\ |01\rangle\langle 01| - \frac{I}{4},\ |10\rangle\langle 10| - \frac{I}{4},\ |11\rangle\langle 11| - \frac{I}{4}\}$. Since this spans all tracelss diagonal operators, $\mathcal{B}_L^\perp$ is the set of all off-diagonal 2-qubit operators.

A key step in the algorithm is the construction of a set of canonical super-operators. The first one is E_I, which is a trace-preserving superoperator that effectively eliminates all operators in $\mathcal{B}_R$

$$E_I |I\rangle\rangle = |I\rangle\rangle,\ E_I |P'\rangle\rangle \in \mathcal{B}_L^\perp \quad \forall P' \in \mathcal{B}_R\ . \tag{15}$$

Then we define a set of trace-preserving canonical super-operators that effectively maps a specific operator in $\mathcal{B}_R$ to a specific operator in $\mathcal{B}_L$. For any $P \in \mathcal{B}_R$ and $\bar{M}_i \in \mathcal{B}_L$

$$\begin{aligned} &E_{P,i} |I\rangle\rangle = |I\rangle\rangle,\ E_{P,i} |P\rangle\rangle - |\bar{M}_i\rangle\rangle \in \mathcal{B}_L^\perp, \\ &E_{P,i} |P'\rangle\rangle \in \mathcal{B}_L^\perp \quad \forall P' \in \mathcal{B}_R \setminus \{P\}\ . \end{aligned} \tag{16}$$

These mappings are "effective", as they always have some component in $\mathcal{B}_L^\perp$, which we have left uncharacterized in the above definitions. But this is inconsequential as these components are not observed by the POVM. In what follows, we refer to E_I and $E_{P,i}$ as to the *eliminator operators*, or *eliminators*.

Running example: Since the unobservable part is left unspecified, the definition of the eliminators are not unique. For the two-qubit example we take $\mathcal{B}_R$ as all normalized, traceless Pauli strings. In that case we can always take, $E_I = \frac{1}{4}(|I\rangle\rangle\langle\langle I| + |X\rangle\rangle\langle\langle X|)^{\otimes 2}$.
From this definition,

$$\begin{aligned} E_I |I\rangle\rangle = \frac{1}{4}(&\mathrm{Tr}(I)\ I + \mathrm{Tr}(X \otimes I)\ X \otimes I \\ &+ \mathrm{Tr}(I \otimes X)\ I \otimes X + \mathrm{Tr}(X \otimes X) X \otimes X) = I \end{aligned} \tag{17}$$

Similarly we can check that this eliminates all the diagonal Pauli strings owing to the anti-commutation relation between Z and X. This will not eliminate Pauli strings with only X for instance. But these are off-diagonal operators which lie in $\mathcal{B}_L^\perp$. Similarly other eliminators can also be constructed for this case. The general formula for these constructions in the computational basis is given in Section 3.3.

Now to perform simultaneous tomography, we need to apply these canonical operators to the state by aggregating measurement outcomes as described in (9). To do this, it is sufficient to have a set of unitary operators such that these canonical operators lie in their span. We call such a set of unitaries *tomographically complete* and use noisy measurement outcomes generated by these unitaries, as in (7), to perform simultaneous tomography.

Definition 1 (Tomographically complete set). *We call a set of unitary operators* $\mathcal{U}_{tom} = \{U_1 \ldots U_L\}$ ***tomographically complete*** *if for all* $P \in \mathcal{B}_R$, $i \in [D]$, *we have* $E_{P,i} \in \mathcal{L}(\mathcal{U}_{tom})$ *and* $E_I \in \mathcal{L}(\mathcal{U}_{tom})$.

This implies that if the set $\mathcal{U}_{tom}$ is tomographically complete, then there exists coefficients $c_l^{P,i}$ such that $\sum_l c_l^{P,i} = 1$ and

$$\sum_{l=1}^{L} c_l^{P,i} \Phi(U_l) = E_{P,i}. \tag{18}$$

Further, there exist coefficients, c_l^I, such that $\sum_l c_l^I = 1$ and,

$$\sum_{l=1}^{L} c_l^I \Phi(U_l) = E_I. \tag{19}$$

The definition of this set of unitaries and the corresponding coefficients for constructing the eliminators will obviously depend on the basis sets $\mathcal{B}_R$ and $\mathcal{B}_L$. For the special case of computational basis measurements with $\mathcal{B}_R$ taken to be the Pauli operator basis, we can show that this set is a subset of the Clifford group on $n-$qubits (see (31)).

Now using these coefficients in (8) we can aggregate the noisy measurement outcomes to effectively apply the canonical operators to the state,

$$z_k^I := \sum_{l=1}^{L} c_l^I \tilde{y}(U_l)_k \ , \tag{20}$$

$$z_k^{P,i} := \sum_{l=1}^{L} c_l^{P,i} \tilde{y}(U_l)_k \ , \quad \forall P \in \mathcal{B}_R,\ i,k \in [D] \ . \tag{21}$$

Now from the definition of the eliminators and (9) we can connect the z values to ρ and A.

$$z_k^I = \sum_{k'} \frac{A_{k\,k'} m_{k'\,I}}{2^{n/2}} \ , \tag{22}$$

$$z_k^{P,i} = z_k^I + s_P \sum_{k'} A_{k\,k'} C_{k'\,i} \ , \tag{23}$$

where C is the *covariance matrix* associated with the POVM,

$$C_{ij} := \langle\langle M_i | \bar{M}_j \rangle\rangle = \mathrm{Tr}(M_i M_j) - \frac{\mathrm{Tr}(M_i)\mathrm{Tr}(M_j)}{D} \ . \tag{24}$$

To emphasize, the $z-$values are obtainable from measurements. Our aim is to invert the relations in (22) and (23) to find ρ and A up to the unknown gauge.

Given the ability to obtain these z values, the simultaneous tomography algorithm can be broken down into three steps; finding the positions of the non-zero coefficients of ρ in the $\mathcal{B}_R$ basis, computing A up to a gauge, and computing the other state coefficients of ρ up to gauge. Below we will give a brief description of each of these steps. The full algorithm is given in Algorithm 1. Full technical details of the algorithm can be found in Appendix D.

Running example:
We see from (17) that $E_I = \frac{1}{4}\sum_{U\in\{I,X\}^{\otimes 2}} \Phi(U)$. So from this explicit construction, we get the c_l^I values and we can compute z_k^I from this.
We can group these $z-$values by the k and i indices into 4 dimensional vectors and matrices. For the 2-qubit example we compute these values to be

$$\mathbf{z}^I = [0.25, 0.25, 0.25, 0.25] \ ,$$

$$\mathbf{z}^{I\otimes Z} = \mathbf{z}^{Z\otimes Z} = \begin{pmatrix} -0.03 & 0.33 & 0.33 & 0.37 \\ 0.33 & -0.03 & 0.37 & 0.33 \\ 0.33 & 0.37 & -0.03 & 0.33 \\ 0.37 & 0.33 & 0.33 & -0.03 \end{pmatrix},$$

$$\mathbf{z}^{Z\otimes I} = \begin{pmatrix} 0.53 & 0.17 & 0.17 & 0.13 \\ 0.17 & 0.53 & 0.13 & 0.17 \\ 0.17 & 0.13 & 0.53 & 0.17 \\ 0.13 & 0.17 & 0.17 & 0.53 \end{pmatrix} .$$

We get the uniform stochastic matrix for the other cases where $s_P = 0$.

Step 1: Finding non-zero coefficients
To find a non-zero state coefficient, we first have to isolate all the rows of A that are not all zeros. From (22), this can be clearly done by finding all $l \in [D]$ such that $z_l^I \neq 0$. Now for one such l and for $j \in [D]$, if $z_l^{P,j} - z_l^I \neq 0$, then s_P must be non-zero. On the other hand if for all $j \in [D]$ if $z_l^{P,j} - z_l^I = 0$, then s_P must be zero.

We can repeat this step for the same l for each $P \in \mathcal{B}_R$ to find every non-zero s_P. We also store one particular (j,l), obtained from $z_l^{P,j} - z_l^I \neq 0$ for any P, to use in the final step.

Running example: For $s_P = 0$, we will get $\mathbf{z}^P$ to be the matrix of all 0.25. Only $\mathbf{z}^{I\otimes Z}, \mathbf{z}^{Z\otimes I}$, and $\mathbf{z}^{Z\otimes Z}$ will differ from this and we can identify these with the non-zero coefficients of the state.

Step 2: Finding noise matrix up to gauge
At this step, we choose $R \in \mathcal{B}_R$ such that $s_R \neq 0$. To work around the gauge problem we have to choose one noise matrix from the one-parameter family described by (12). We make this choice by taking $\alpha = s_R$. This makes the explicit s_R dependence vanish from (23). In terms of the gauge transformed noise matrix, (22) and (23) can be expressed as follows.

$$z_k^I = \sum_{k'} \frac{A'(s_R)_{k\,k'} m_{k'\,I}}{2^{n/2}} \ , \tag{25}$$

$$z_k^{R,i} = z_k^I + \sum_{k'} A'(s_R)_{k\,k'} C_{k'\,i} \ , \quad \forall\ i,k \in [D] \ . \tag{26}$$

Once we obtain the $z-$values by aggregating the measurements; we can invert the system of linear

equations to find $A'(s_R)$. This inversion step is always possible if the POVM is linearly independent, i.e., if the Condition 3 holds.

Running example: Choose the gauge to be $s_{Z\otimes I}$. Covariance matrix for computational measurements is $C_{ij} = \delta_{ij} - 0.25$. Now by plugging $z-$values in (23) we can compute

$$A'(s_{Z\otimes I}) = \begin{pmatrix} 0.53 & 0.17 & 0.17 & 0.13 \\ 0.17 & 0.53 & 0.13 & 0.17 \\ 0.17 & 0.13 & 0.53 & 0.17 \\ 0.13 & 0.17 & 0.17 & 0.53 \end{pmatrix} .$$

We can check that this matrix is indeed equal to $s_{Z\otimes I}A + (1 - s_{Z\otimes I})0.25$.

Step 3: Finding state up to gauge
In this step, we exploit the gauge transformation to find the ratio of every non-zero state coefficient with s_R. From (22) and (23) the following relation holds,

$$\frac{s_P}{s_R} = \frac{\sum_{k'} A'_{k\,k'}(s_P)C_{k'\,i}}{\sum_{k'} A'_{k\,k'}(s_R)C_{k'\,i}} = \frac{z_l^{P,j} - z_l^I}{z_l^{R,j} - z_l^I} . \quad (27)$$

Our choice of (j, l) in *Step 1* ensures that the denominator in this expression is always non-zero.

Running example: Choose $j, l = 1$. This choice is made so that $z_j^{Z\otimes I,l} \neq z_l^{I\otimes I}$. From the expression given above,

$$\frac{s_{I\otimes Z}}{s_{Z\otimes I}} = \frac{\mathbf{z}_{1,1}^{I\otimes Z} - \mathbf{z}_1^{I\otimes I}}{\mathbf{z}_{1,1}^{Z\otimes I} - \mathbf{z}_1^{I\otimes I}} = \frac{-0.03 - 0.25}{0.53 - 0.25} = -1 ,$$

Similarly we get, $\frac{s_{Z\otimes Z}}{s_{Z\otimes I}} = -1$.

After these steps, we will know ρ and A up to the unknown parameter s_R. This unknown has to be fixed from prior information, and we will describe various ways of fixing this gauge in Section 4. In the next subsection, we will specialize to the case of computational basis measurements, and analyze the number of measurement shots required to implement this algorithm.

Algorithm 1: Simultaneous tomography up to gauge degree of freedom

```
// Step 1. Find non-zero coefficients of the state
Compute z_k^I  ∀ k ∈ [D] using (22)
K ← {k ∈ [D] | z_k^I ≠ 0}
C ← {} // Empty set
S ← [D] × K // Search space of index tuples
for P ∈ B_R do
    s_P ← 0
    for (i, k) ∈ S do
        Compute z_k^{P,i} using (23)
        if z_k^{P,i} ≠ z_k^I then
            S ← {(i, k)}
            // Replace index set with a single tuple
            C ← C ⋃ {P}
            continue // To the next P
        end
    end
end
// Step 2. Find A up to gauge symmetry
choose R ∈ C
Compute z_k^{R,i} for all k ∈ K, i ∈ [D]
for k ∈ [D] do
    if k ∈ K then
        Solve for A'_{kk'} in
        Σ_{k'} A'_{kk'} C_{k'i} = z_k^{R,i} − z_k^I  ∀i ∈ [D]
        Σ_{k'} A'_{kk'} m_{k'I} = 2^{n/2} z_k^I
        else
            A'_{kk'} ← 0  ∀k' ∈ [D]
        end
    end
end
// Step 3. Find other state coefficients up to a multiplicative constant
{(j, l)} ← S
for P ∈ C do
    s_P / s_R ← (z_l^{P,j} − z_l^I) / (z_l^{R,j} − z_l^I)
end
return {(P, s_P/s_R) | P ∈ C},  A'
```

3.3 Simultaneous tomography with randomized measurements and shot error

The simultaneous tomography algorithm, as described, does not consider the fact that every $\tilde{y}_k(U)$ has to be estimated using a finite number of measurement outcomes. In this section, we will specialize the algorithm to the case where measurements are made in the computational basis and analyze the number of measurement shots required to estimate the state and noise in the system up to gauge. Additionally, we will also use a randomized measurement procedure to estimate the z values required for tomography. The sample complexity bounds in Theorem 2 specify the number of such randomized measurements required for simultaneous tomography. This randomized measurement method can significantly reduce the overhead of simultaneous tomography as the number of operators in the tomographically complete set can be exponentially large in the system size.

We will present our results exclusively for the case

of computational basis measurements, as this is the most pertinent case for practical applications.

3.3.1 Simultaneous tomography in the computational basis

For computational measurements the POVM is simply $\{|k\rangle\langle k| \, | k \in [2^n]\}$, and the covariance operator takes a simple form,

$$C_{ik} = \delta_{ik} - \frac{1}{2^n} \, . \tag{28}$$

For these types of measurements, the natural choice for the right basis set $\mathcal{B}_R$ is the set of all normalized Pauli strings,

$$\mathcal{B}_R = \left\{ \frac{I}{\sqrt{2}}, \frac{X}{\sqrt{2}}, \frac{Y}{\sqrt{2}}, \frac{Z}{\sqrt{2}} \right\}^{\otimes n} \setminus \left\{ \frac{I}{2^{n/2}} \right\} \, . \tag{29}$$

Remarkably for this choice of $\mathcal{B}_R$, the effective elimination operators, defined in (15) and (16), can be constructed using only Clifford operations.

Let $\mathcal{P}_X$ (or $\mathcal{P}_Z$) be the set of Pauli strings composed of only X (or Z) and I. Now define $H_{lQ} = \langle l|Q|l\rangle / 2^{n/2}, \ \forall Q \in \mathcal{P}_Z$. Then we can show that,

$$E_I = \frac{1}{2^n} \sum_{P \in \mathcal{P}_X} \Phi(P) \, , \tag{30}$$

$$E_{Pi} = (1 - \sum_{Q \neq I} H_{iQ}) E_I + \frac{2}{2^n} \sum_{Q \neq I} H_{iQ} \sum_{\substack{Q' \in \mathcal{P}_X \\ [Q',Q]=0}} \Phi(Q' U_{PQ}) \, . \tag{31}$$

Here U_{PQ} is a member of the $n-$qubit Clifford group that maps P to Q. Proof of this construction is given in Appendix E. The overhead of applying these eliminators can be decreased significantly by using a randomized measurement scheme (see Appendix F).

Due to the simplified nature of the covariance matrix and the POVM, the z values in this setting take the following simple forms,

$$z_k^I = \frac{\sum_{k'} A_{kk'}}{2^n} \, , \tag{32}$$

$$z_k^{P,i} = z_k^I + s_p (A_{ik} - z_k^I) \, . \tag{33}$$

This means that if the gauge is fixed to s_R, we get the following simple relation between the noise matrix and gauge s_R,

$$A'_{ik}(s_R) = z_k^{Ri} \, . \tag{34}$$

So for the case of computational basis measurements, the linear inversion in the second step of Algorithm 1 is unnecessary.

3.3.2 Sample complexity

The number of measurements required to estimate ρ and A up to a certain error depends on how far they are from violating the sufficient conditions 1 and 2. The third condition is automatically satisfied as computational basis measurements form a POVM that is linearly independent. More measurements are required for simultaneous tomography the closer ρ is to a maximally mixed state and the closer A is to erasure channel. To measure the distance from these pathological cases, we define the following metrics for ρ and A.

$$||\rho||_{\text{mix}} \equiv \max_{P \in \mathcal{B}_R \setminus I} |s_P| \, , \tag{35}$$

$$||A||_{\text{uni}} \equiv \max_{k,k' \in [2^n]} \left| A_{kk'} - \frac{\sum_i A_{ki}}{2^n} \right| \, . \tag{36}$$

The sufficient conditions 1 and 2 are just non-zero lower bounds on these metrics. These metrics allow us to state the sample complexity for the simultaneous tomography algorithm:

Theorem 2. ***Complexity in computational basis*** *Given an $n-$qubit quantum system such that $||\rho||_{mix} > 0$, $||A||_{uni} > 0$. Choose a threshold parameter $0 < \beta < ||\rho||_{mix}/2$. Then the three main steps of the simultaneous tomography algorithm can be implemented using randomized measurements in the computational basis with the following complexities,*

1. *Using $O(8^n \frac{cn+log(1/\delta)}{\beta^2 ||A||^2_{uni}})$ randomized measurements, we can identify a non-empty subset $\mathcal{C} \subset \mathcal{B}_R$ such that with probability $1-\delta$ the following implications hold,*
$$P \in \mathcal{C} \implies |s_P| \geq \beta \, ,$$
$$|s_P| \geq 1.01\beta \implies P \in \mathcal{C} \, .$$

2. *Given $R \in \mathcal{C}$, using $O(2^n \frac{cn+log(1/\delta)}{\epsilon^2})$ randomized measurements, we can give an estimate $\hat{A}'(s_R)$ for the noise matrix up to gauge such that,*
$$Pr\left(\max_{i,j \in [D]} |\hat{A}'_{i,j}(s_R) - A'_{i,j}(s_R)| > \epsilon \right) \leq \delta \, .$$

3. *Let $\epsilon < \beta/2$. Then for a fixed $R \in \mathcal{C}$ and for every $P \in \mathcal{C}$, we can compute an estimate $\widehat{\frac{s_P}{s_R}}$ using a total of $O(2^n |\mathcal{C}| \frac{cn+\log(1/\delta)}{\epsilon^2 \beta^2 ||A||^2_{uni}})$ such that,*
$$Pr\left(\max_{P \in \mathcal{C}} \frac{\left| \widehat{\frac{s_P}{s_R}} - \frac{s_P}{s_R} \right|}{\left| \frac{s_P}{s_R} \right|} > \epsilon \right) \leq \delta \, .$$

See Appendix F for details on the randomized measurement framework used and the proof of this theorem

The three parts of this theorem correspond to the three steps of Algorithm 1. In the first step of the algorithm our aim is to find the positions of the non zero coefficients of the state. The finite shot version of this step is the construction of the set $\mathcal{C}$

which is guaranteed (with high probability) to contain every operator in $\mathcal{B}_R$ whose overlap with the state is greater than or equal to 1.01β in absolute value. Moreover it is also guaranteed with high probability that $\mathcal{C}$ will only have operators such that their overlap with the state is guaranteed to be greater than or equal to β in absolute value. Now if we choose $\beta = \frac{1}{1.01}\min_{P:s_p\neq 0}|s_P|$, then $\mathcal{C}$ will exactly contain all the positions of the non-zero coefficients. On the other hand, if we are only concerned about recovering the noise matrix, we can take $\beta = \frac{||\rho||_{\text{mix}}}{2}$, which guarantees that $\mathcal{C}$ is non-empty. This will give us at least one non-zero coefficient to set the gauge for the problem.

The second part of the theorem concerns the estimation of the noise matrix up to gauge in the finite shot setting. This is straightforward in the computational basis as we do not have to perform any linear inversion. Notice that the sample complexity of this step is independent of $||\rho||_{\text{mix}}$ and $||A||_{\text{uni}}$. This gives a considerable sample complexity advantage in the setting where we are only interested in recovering A.

Corollary 1. Estimating measurement noise (M error)
Fixing $\beta = ||\rho||_{mix}/2$ in Theorem 2, we can recover the noise matrix up to gauge with the error ϵ, with high probability using $\tilde{O}\left(\frac{8^n}{||\rho||^2_{mix}||A||^2_{uni}} + \frac{2^n}{\epsilon^2}\right)$ randomized measurements.

Here we have used the $\tilde{O}$ notation to hide linear factors in n and $\log(1/\delta)$ for the sake of readability.

The third part of the theorem concerns the estimation of state coefficients up to gauge. We only do this for $P \in \mathcal{C}$ and we set $s_P = 0$ for $P \notin \mathcal{C}$. Thus the threshold parameter (β) implicitly fixes the error in estimating these coefficients. The sample complexity of this step depends on $|\mathcal{C}|$ and hence can be considerably low if the state is sparse in the Pauli basis.

Suppose our aim is to fully characterize the state preparation error. To estimate every element of the state with an *additive* error of ϵ we show the following Corollary of Theorem 2 in Appendix F.3.

Corollary 2. Estimating prepared state (SP error) *Every coefficient of the state up to gauge can be estimated with additive error of $\epsilon \leq ||\rho||_{mix}/2$ with high probability using a total of $\tilde{O}\left(\frac{8^n}{\epsilon^4||A||^2_{uni}}\right)$ randomized measurements*

The exponential dependence on n in these sample complexities is unavoidable because we are attempting to estimate an exponential number of independent, unknown quantities in the most general case. But this dependence can be possibly improved for special cases, like for unentangled states or binary symmetric noise channels. We leave the analysis of such special cases for future work.

3.4 Numerical results

The tomography procedure outlined here uses independent state copies and does not use coherent measurement. In this setting, recent results have shown a sample complexity of $\Theta(8^n)$ for noiseless tomography [12]. Hence we do not expect a substantial improvement in the n dependence in Theorem 2. More interesting is the dependence of the sample complexity w.r.t to $||A||_{\text{uni}}$ and $||\rho||_{\text{mix}}$ (via the β parameter). To check the tightness of our analysis w.r.t these quantities, we numerically study the sample complexity scaling for a few different 2−qubit examples. The results are given in Figure 2. In these experiments, we estimate the sample complexity of Steps 1 and 3 while varying β and $||A||_{\text{uni}}$ independently for a family of two-qubit states. In all the cases, we empirically observe that the number of measurement shots scale as the inverse square of these quantities, corroborating Theorem 2.

4 Incorporating prior information: Gauge fixing and efficiency improvements

We have shown that the gauge freedom in (12) is the only obstacle in performing simultaneous tomography. This gauge can be fixed if we have access to prior information about the state and measurement noise or access to additional measurements. Further, especially from a practical point of view, prior information can be used to significantly reduce the number of measurements and classical post-processing required to perform simultaneous tomography. The linear operator framework provides a natural way to incorporate several types of prior information. We also find that many of the example priors we use correspond to assumptions made in the error mitigation literature previously [26, 42, 49, 50, 66]

4.1 Using prior information to fix the gauge

We start with examples of prior information about ρ and A that are sufficient to fix the gauge. Each of these conditions imply that no two pairs (ρ, A) and (ρ', A') can satisfy the conditions enforced by the prior information and lie in the one-dimensional gauge manifold (12).

Block independent noise:
Suppose that the POVM is described by $M_{kl} = M^1_k \otimes M^2_l$ where $k \in [D_1]$ and $l \in [D_2]$ with $D_1 D_2 = D$. This can refer to a partitioning of a set of binary valued outcomes into two parts. Suppose that the noise acts independently on the two parts such that

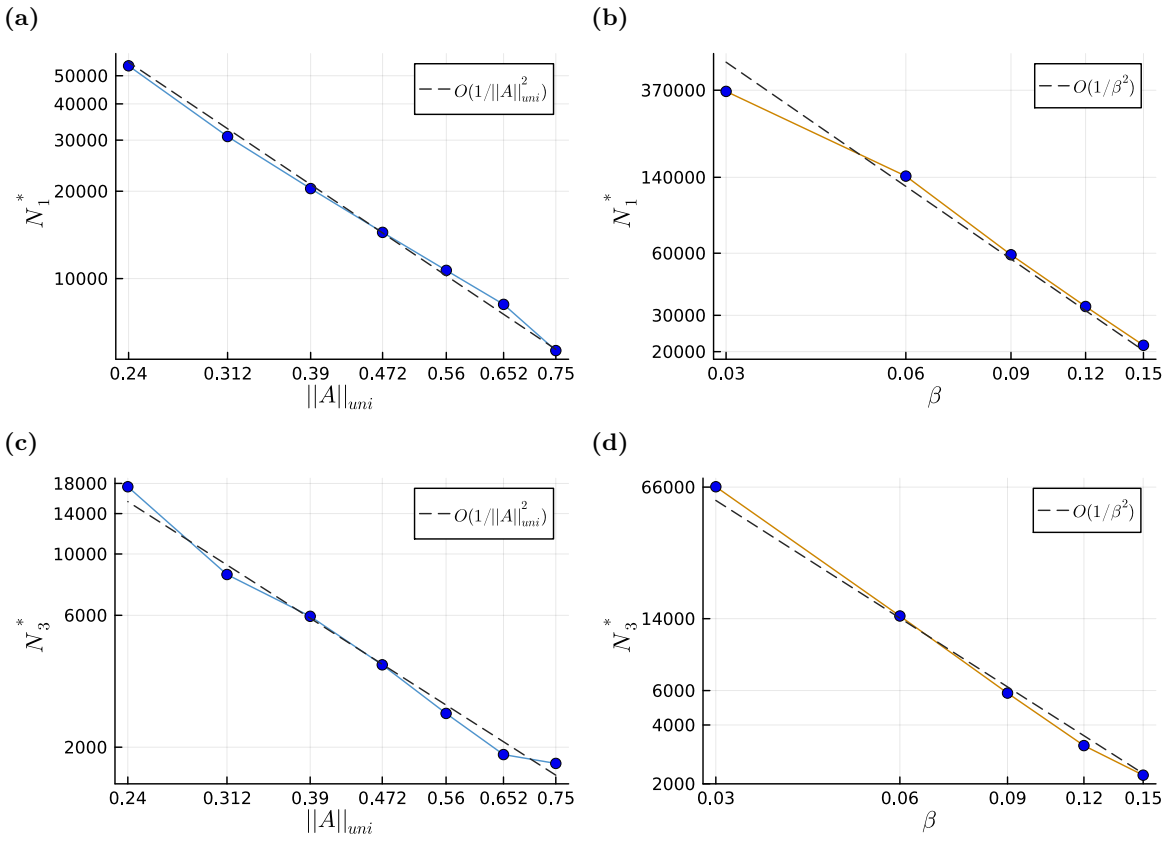

Figure 2: **Simultaneous tomography with randomized measurements in 2-qubit systems. (a)**, **(b)** We study the scaling of Step 1 in Theorem 2 for a two qubit system. N_1^* here is the number of measurements that was required to find the correct positions of the non-zero coefficients in the Pauli basis with a success rate of at least 90%. **(c), (d)** We study the scaling of Step 3 in Theorem 2. Here N_3^* is the number of measurement shots that were required to compute every state coefficient up to gauge with a multiplicative error of at most $\epsilon = 1/3$. In **(a)** and **(c)** , the state is fixed to be $\rho = |01\rangle\langle 01|$ and A is chosen from the one-parameter family $A(\tau) = ((1-\tau)I + \tau X)^{\otimes 2}$ to vary $||A||_{\text{uni}}$. We fix $\beta = 1/4$. In **(b)** and **(d)**, the state is chosen from the one-parameter family $\rho(\tau) = \frac{I}{4} + \tau(Y \otimes I + Z \otimes Z)$ and A is fixed to $A(0.05)$. β is taken to be $||\rho(\tau)||_{\text{mix}}/2 = \tau$ and τ is varied to select β. In all cases, N_1^* and N_3^* are found using logistic regression on the empirical success probability. For each value of N the empirical success probability is estimated using 50 random runs.

$A = A^1 \otimes A^2$ where A^1 acts on M^1 and A^2 acts on M^2. We show that the gauge can be uniquely fixed with this information. The details of the proof of uniqueness and the algorithm to find the unique gauge are given in Appendix I.1. Similar uncorrelated noise models have been used as a simplifying assumption in the literature [26, 32, 49, 50, 66]

Information on purity of the state:
Let the state ρ satisfy the following *purity* conditions:

1. $\text{Tr}(\rho^2) = \nu$
2. There exists $|v\rangle$ such that $\langle v|\rho|v\rangle > 2^{-(n-1)}$.

If the state ρ satisfies these purity conditions then the gauge can be fixed in any system of at least two qubits. In general if the purity of ρ is known, and if its min-entropy [11, 69] is less than $n-1$, then the gauge can be fixed. As an important special case, we note that any *pure state* satisfies the purity conditions with $\nu = 1$ and there exists an eigenvalue equal to one.

We can use this purity information to find the gauge as follows. Algorithm 1 returns a set of state coefficients s'_P up to the gauge freedom such that the actual state coefficients s_P are related to the ones computed by the algorithm by

$$s_P = \alpha s'_P, \quad \forall P \neq I. \tag{37}$$

Using the first purity condition $\text{Tr}(\rho^2) = \nu$, we get

$$\alpha^2 = \frac{\nu - 2^{-n}}{\sum_{P,P' \neq I} s'_P s'_{P'} \langle\langle P|P'\rangle\rangle} . \tag{38}$$

This specifies the state up to a sign

$$\rho = \frac{I}{2^n} \pm \alpha \sum_{P \neq I} s'_P P . \tag{39}$$

Denote the two candidates by ρ_+ and ρ_- with $\rho_+ + \rho_- = \frac{I}{2^{n-1}}$. Using the second purity condition, there exists a state $|v\rangle$ such that $\langle v|\rho_+|v\rangle > 2^{-(n-1)}$. This gives that $\langle v|\rho_-|v\rangle < 0$ which in turn implies

that ρ_- is not a positive semi-definite matrix. Hence one of the two candidate states will not be a valid quantum state and can be used to pick the correct sign and fix the gauge.

Probe state:
The ability to have a known state (for instance $|0\rangle^{\otimes n}$) prepared can be used to fix the gauge as in this case α can be directly inferred from (12). This type of prior information is used in [66], along with Pauli twirling for the purpose of error mitigation.

Suppose that ρ^{pb} is a known probe state that is measured using the M_k to give

$$y_k^{probe} = \sum_{k' \in [D]} A_{kk'} \mathrm{Tr}(\rho^{pb} M_k). \tag{40}$$

Since Algorithm 1 outputs a candidate $A'(\alpha)$ up to the gauge degeneracy, we have

$$y_k^{probe} = \frac{1}{\alpha} \sum_{k'} A'_{kk'}(\alpha) \mathrm{Tr}(\rho^{pb} M_k) + (1 - \frac{1}{\alpha}) \frac{\sum_j A'(\alpha)_{k\,j} m_{j\,I}}{2^{n/2}} \sum_{k'} \mathrm{Tr}(\rho^{pb} M_{k'}). \tag{41}$$

As α is the only unknown quantity, it can be easily computed from the above equation.

4.2 Using prior information for computational improvements

In this section, we list a set of prior information that can both fix the gauge and can be used within the linear operator framework to obtain reductions in number of measurements and post-processing.

Linearly represented prior information:
We consider a set of linearly represented prior information available on the state and the noise matrix. For the state ρ this refers to a set of known expectation values that may for example correspond to known physical properties of the unknown state. Using a basis we can represent this information as follows.

$$\sum_{P \in \mathcal{B}_R} s_P b^i_{S,P} = d^i_S, \quad i = 1, \ldots, N_s \tag{42}$$

Further assume that $d^1_S \neq 0$ which will allow us to fix the gauge. For the noise matrix A, linear prior information refers to the action of A on known vectors. This for example, can be used to represent knowledge about A from previous experiments or from the use of multiple probe states. We denote these known quantities by

$$A b^i_A = d^i_A, \quad i = 1, \ldots, N_A. \tag{43}$$

With access to the information in (42) and (43), we will need access to only a subspace of superoperators $\mathcal{L}^{lin} \subset \mathcal{L}$ to perform simultaneous tomography. Essentially we need to access information in the orthogonal subspace to those given in (42) and (43) by using appropriate operators from the linear operator space. The details are given in Appendix I.2.

Denoising and hierarchical tomography:
We consider a special case of linearly represented prior information on ρ that provides backward compatibility with some previously designed noise-free tomography method. Suppose that a set of unitaries $\mathcal{U}_{nf} = \{U_1, \ldots, U_L\}$ have been designed to perform tomography on an unknown state ρ in the absence of measurement noise. The set $\mathcal{U}_{nf}$ essentially encodes the prior information on ρ in a linear way. This is because we assume that the state can be uniquely specified, in the noiseless setting, from the set of coefficients $\{\mathrm{Tr}(U\rho U^\dagger M_i) | U \in \mathcal{U}_{nf}, i \in [D]\}$.

Our goal is to utilize the set $\mathcal{U}_{nf}$ and provide a method to perform simultaneous tomography in the presence of measurement noise.

In this setting, we can use $\mathcal{B}_L = \mathcal{B}_R = \mathcal{B} = \{\bar{M}_i \mid i \in [D]\}$ and our goal is to find the noise matrix A and the coefficients of $U\rho U^\dagger$ in the basis $\mathcal{B}$ for all $U \in \mathcal{U}_{nf}$. For this, we need access to a subset $\mathcal{L}^{den} \subset \mathcal{L}$ of linear operators given by

$$\mathcal{L}^{den} = \{E_{ij}, \quad i, j \in [D]\}\,, \tag{44}$$

where E_{ij} is defined as

$$E_{ij} |\bar{M}_i\rangle\rangle - |\bar{M}_j\rangle\rangle \in \mathcal{B}_L^\perp\,, \quad E_{ij} |\bar{M}_{i'}\rangle\rangle \in \mathcal{B}_L^\perp \quad \forall i' \neq i\,,$$
$$E_{ij} |Q\rangle\rangle \in \mathcal{B}_L^\perp \quad \forall Q \in \mathcal{B}_L^\perp\,. \tag{45}$$

The set of operators in (45) are sufficient to *denoise* each of the original measurements. If $\mathcal{U}^{den}$ is a generator set for $\mathcal{L}^{den}$ then the generator set for performing simultaneous tomography in the hierarchical setting is given by

$$\mathcal{U}^{hier} = \{U_1 U_2 \mid U_1 \in \mathcal{U}^{den}, U_2 \in \mathcal{U}^{nf}\}. \tag{46}$$

Essentially, we are first using the previously designed tomography gate set $\mathcal{U}^{nf}$ to prepare the states $U_1\rho U_1^\dagger \ldots U_L \rho U_L^\dagger$, and then use a simultaneous tomography procedure to estimate the projection of these states into the subspace defined by the POVM.

Running example:
Suppose we have prior information that ρ is diagonal in the computational basis. So if we use the computational basis measurements, $\mathcal{U}_{nf} = \{I\}$. As another example, if we know that the state, when expanded in the Pauli basis, has only X terms (e.g. $\rho \propto I + X \otimes X$), then $\mathcal{U}_{nf}$ would contain the global rotation operator that takes X to Z.

Binary symmetric channel:
In this part, we consider a special case of hierarchical tomography where the measurements are along the computational basis. The binary symmetric channel refers to the case where each of the n binary observables is flipped with a given probability independent of the rest. Let the bit flip probabilities be $p_1, \ldots, p_n$ where $p_i \neq 1/2$. Then the output noise matrix has the special form

$$A = \bigotimes_{i=1}^{n} A_i^{sym}, \quad \text{where } A_i^{sym} = \begin{bmatrix} 1-p_i & p_i \\ p_i & 1-p_i \end{bmatrix} \tag{47}$$

By Theorem 4, the family of binary symmetric channel noise matrices allows us to fix the gauge degree of freedom. Although this is a special case of block independent noise matrtices, the extra structure can be exploited to obtain a simpler denoising algorithm. The required subset $L^{BSC} \subseteq \mathcal{L}$ for performing this task is relatively small and the corresponding generator set can be realized by depth n circuits consisting of $CNOT$ and $SWAP$ gates. The details of the algorithm is given in Appendix I.4.

Independent ancilla:
We consider the availability of a set of ancilla qubits where the state and measurement noise are independent of the rest. The presence of such an independent ancilla has been assumed in the work [42]. We can view this as a special case of uncorrelated noise and therefore we can fix the gauge. The POVM is the set $\{M_i^{anc} \otimes M_j \mid i \in [D^a], j \in [D^r]\}$. The overall state and the corresponding suitable choice of basis for the traceless subspaces can be written as

$$\rho = \rho^a \otimes \rho^r, \quad \mathcal{B}_L = \mathcal{B}_L^a \otimes \mathcal{B}_L^r, \quad \mathcal{B}_R = \mathcal{B}_R^a \otimes \mathcal{B}_R^r, \tag{48}$$

where as before we choose $\mathcal{B}_L^a = \{M_i^a \mid i \in [D^a]\}$ and $\mathcal{B}_L^r = \{M_i^r \mid i \in [D^r]\}$. By independence of noise on the ancilla qubits, we can decompose the noise matrix as

$$A = A^a \otimes A^r. \tag{49}$$

This special structure also allows us to significantly reduce the set of operators $\mathcal{L}^{anc} \subseteq \mathcal{L}$ that we need to perform tomography on the state ρ. The specification of these operators and the corresponding algorithm is given in Appendix I.3. This serves as a partial generalization of the construction in [42].
We defer a complete analysis and optimization of all the priors discussed in this section, including sample complexities and construction of eliminators, for future work.

5 Discussion

In this work, we have introduced a general framework for simultaneous tomography. We have completely characterized the gauge ambiguity inherent to this problem and have shown many different ways to get around this limitation. The various scenarios discussed in this context also subsume many assumptions made in prior literature to solve this problem.

There are several directions along which this work can be extended. Like any method attempting to perform full tomography on a quantum system, we find that our method also has exponential complexity in the number of qubits. Recent advances in classical shadows have given more practical methods that help in estimating accessible, but limited information from quantum states [29]. Developing a similar technique for simultaneous tomography would help us extract useful information from ρ and A. Recent works on classical shadows in the presence of noise [36] show promise in this direction.

The ideas presented here can also be extended to the simultaneous characterization of ρ along with more general forms of physical transformations acting on it. Gauge ambiguities also exist in such general cases and the same type of priors discussed here might not be able to fix the gauge. For example, the nature of the gauge transformations will change when trying to estimate ρ and a CPTP map Φ given the ability to measure states of the form of $\Phi(U\rho U^\dagger)$ [40]. And in the general case, the gauge group can have a much more complicated structure than the one-parameter case discussed here. We anticipate that the present work can be possibly extended to study a much richer class of problems that naturally arise in the study of quantum systems.

Acknowledgements

The authors acknowledge support from the Laboratory Directed Research and Development program of Los Alamos National Laboratory under Projects 20220545CR-NLS, 20210114ER, 20230338ER, and 20240032DR. SC was supported by the U.S. Department of Energy (DOE) through a quantum computing program sponsored by the Los Alamos National Laborator Information Science and Technology Institute. AYL was partially supported by the U.S. DOE/SC Advanced Scientific Computing Research Program.

References

[1] Scott Aaronson and Lijie Chen. Complexity-theoretic foundations of quantum supremacy experiments. In *Proceedings of the 32nd Computational Complexity Conference*, CCC '17, Dagstuhl, DEU, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 9783959770408. DOI: 10.48550/arXiv.1612.05903.

[2] C. G. Almudever, L. Lao, X. Fu, N. Khammassi, I. Ashraf, D. Iorga, S. Varsamopoulos, C. Eichler, A. Wallraff, L. Geck, A. Kruth, J. Knoch, H. Bluhm, and K Bertels. The engineering challenges in quantum computing. In *Design, Automation & Test in Europe Conference & Exhibition, 2017*, pages 836–845, 2017. DOI: 10.23919/DATE.2017.7927104.

[3] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, Oct 2019. ISSN 1476-4687. DOI: 10.1038/s41586-019-1666-5.

[4] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Jonathan Mizrahi, Jonathan D. Sterk, and Peter Maunz. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. 2013. DOI: 10.48550/ARXIV.1310.4492.

[5] Adam Bouland, Bill Fefferman, Zeph Landau, and Yunchao Liu. Noise and the frontier of quantum supremacy. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1308–1317, 2022. DOI: 10.1109/FOCS52979.2021.00127.

[6] Colin D. Bruzewicz, John Chiaverini, Robert McConnell, and Jeremy M. Sage. Trapped-ion quantum computing: Progress and challenges. *Applied Physics Reviews*, 6(2):021314, 2019. DOI: 10.1063/1.5088164.

[7] Iulia Buluta, Sahel Ashhab, and Franco Nori. Natural and artificial atoms for quantum computation. *Reports on Progress in Physics*, 74(10):104401, sep 2011. DOI: 10.1088/0034-4885/74/10/104401.

[8] Zhenyu Cai, Ryan Babbush, Simon C Benjamin, Suguru Endo, William J Huggins, Ying Li, Jarrod R McClean, and Thomas E O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, 95(4):045005, 2023. DOI: 10.1103/RevModPhys.95.045005.

[9] Earl T. Campbell, Barbara M. Terhal, and Christophe Vuillot. Roads towards fault-tolerant universal quantum computation. *Nature*, 549 (7671):172–179, Sep 2017. ISSN 1476-4687. DOI: 10.1038/nature23460.

[10] Rohit Chaurasiya and Devanshi Arora. *Photonic Quantum Computing*, pages 127–156. Springer International Publishing, Cham, 2022. ISBN 978-3-031-04613-1. DOI: 10.1007/978-3-031-04613-1˙4.

[11] S. S Chehade and A. Vershynina. Quantum entropies. *Scholarpedia*, 14(2):53131, 2019. DOI: 10.4249/scholarpedia.53131. revision #197083.

[12] Sitan Chen, Jerry Li, Brice Huang, and Allen Liu. Tight bounds for quantum state certification with incoherent measurements. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1205–1213. IEEE, 2022. DOI: 10.1109/FOCS54457.2022.00118.

[13] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of nisq. *Nature Communications*, 14(1):6001, 2023. DOI: 10.1038/s41467-023-41217-6.

[14] J. Chiaverini, D. Leibfried, T. Schaetz, M. D. Barrett, R. B. Blakestad, J. Britton, W. M. Itano, J. D. Jost, E. Knill, C. Langer, R. Ozeri, and D. J. Wineland. Realization of quantum error correction. *Nature*, 432(7017):602–605, Dec 2004. ISSN 1476-4687. DOI: 10.1038/nature03074.

[15] Isaac L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, 1997. DOI: 10.1080/09500349708231894.

[16] D. G. Cory, M. D. Price, W. Maas, E. Knill, R. Laflamme, W. H. Zurek, T. F. Havel, and S. S. Somaroo. Experimental quantum error correction. *Phys. Rev. Lett.*, 81:2152–2155, Sep 1998. DOI: 10.1103/PhysRevLett.81.2152.

[17] Antonio D. Córcoles, Abhinav Kandala, Ali Javadi-Abhari, Douglas T. McClure, Andrew W. Cross, Kristan Temme, Paul D. Nation, Matthias Steffen, and Jay M. Gambetta. Challenges and opportunities of near-term quantum computing systems. *Proceedings of the IEEE*, 108(8):1338–1352, 2020. DOI: 10.1109/JPROC.2019.2954005.

[18] Christoph Dankert, Richard Cleve, Joseph

Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80:012304, Jul 2009. DOI: 10.1103/PhysRevA.80.012304.

[19] Nathalie P. de Leon, Kohei M. Itoh, Dohun Kim, Karan K. Mehta, Tracy E. Northup, Hanhee Paik, B. S. Palmer, N. Samarth, Sorawis Sangtawesin, and D. W. Steuerman. Materials challenges and opportunities for quantum computing hardware. *Science*, 372(6539): eabb2823, 2021. DOI: 10.1126/science.abb2823.

[20] Salvatore S. Elder, Christopher S. Wang, Philip Reinhold, Connor T. Hann, Kevin S. Chou, Brian J. Lester, Serge Rosenblum, Luigi Frunzio, Liang Jiang, and Robert J. Schoelkopf. High-fidelity measurement of qubits encoded in multilevel superconducting circuits. *Phys. Rev. X*, 10:011001, Jan 2020. DOI: 10.1103/PhysRevX.10.011001.

[21] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Phys. Rev. X*, 8:031027, Jul 2018. DOI: 10.1103/PhysRevX.8.031027.

[22] Daniel Gottesman. Stabilizer codes and quantum error correction. 1997. DOI: 10.48550/ARXIV.QUANT-PH/9705052. URL https://arxiv.org/abs/quant-ph/9705052.

[23] Daniel Gottesman. Theory of fault-tolerant quantum computation. *Phys. Rev. A*, 57:127–137, Jan 1998. DOI: 10.1103/PhysRevA.57.127.

[24] Daniel Greenbaum. Introduction to quantum gate set tomography. 2015. DOI: 10.48550/ARXIV.1509.02921. URL https://arxiv.org/abs/1509.02921.

[25] Aram W. Harrow and Ashley Montanaro. Quantum computational supremacy. *Nature*, 549 (7671):203–209, Sep 2017. ISSN 1476-4687. URL https://doi.org/10.1038/nature23458.

[26] Johannes Heinsoo, Christian Kraglund Andersen, Ants Remm, Sebastian Krinner, Theodore Walter, Yves Salathé, Simone Gasparinetti, Jean-Claude Besse, Anton Potočnik, Andreas Wallraff, and Christopher Eichler. Rapid high-fidelity multiplexed readout of superconducting qubits. *Phys. Rev. Appl.*, 10:034040, Sep 2018. DOI: 10.1103/PhysRevApplied.10.034040.

[27] J. Helsen, I. Roth, E. Onorati, A.H. Werner, and J. Eisert. General framework for randomized benchmarking. *PRX Quantum*, 3:020357, Jun 2022. DOI: 10.1103/PRXQuantum.3.020357.

[28] Loïc Henriet, Lucas Beguin, Adrien Signoles, Thierry Lahaye, Antoine Browaeys, Georges-Olivier Reymond, and Christophe Jurczak. Quantum computing with neutral atoms. *Quantum*, 4:327, sep 2020. DOI: 10.22331/q-2020-09-21-327.

[29] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, Oct 2020. ISSN 1745-2481. DOI: 10.1038/s41567-020-0932-7.

[30] Christopher Jackson and S. J. van Enk. Detecting correlated errors in state-preparation-and-measurement tomography. *Phys. Rev. A*, 92:042312, Oct 2015. DOI: 10.1103/PhysRevA.92.042312.

[31] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, Mar 2019. ISSN 1476-4687. DOI: 10.1038/s41586-019-1040-7.

[32] Adam C Keith, Charles H Baldwin, Scott Glancy, and Emanuel Knill. Joint quantum-state and measurement tomography with incomplete measurements. *Physical Review A*, 98(4):042318, 2018. DOI: 10.1103/PhysRevA.98.042318.

[33] Adam Kinos, David Hunger, Roman Kolesov, Klaus Mølmer, Hugues de Riedmatten, Philippe Goldner, Alexandre Tallaire, Loic Morvan, Perrine Berger, Sacha Welinski, Khaled Karrai, Lars Rippe, Stefan Kröll, and Andreas Walther. Roadmap for rare-earth quantum computing. 2021. DOI: 10.48550/ARXIV.2103.15743. URL https://arxiv.org/abs/2103.15743.

[34] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77:012307, Jan 2008. DOI: 10.1103/PhysRevA.77.012307.

[35] Emanuel Knill, Raymond Laflamme, and Lorenza Viola. Theory of quantum error correction for general noise. *Phys. Rev. Lett.*, 84:2525–2528, Mar 2000. DOI: 10.1103/PhysRevLett.84.2525.

[36] Dax Enshan Koh and Sabee Grewal. Classical shadows with noise. *Quantum*, 6:776, 2022. DOI: 10.48550/arXiv.2011.11580.

[37] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien. Quantum computers. *Nature*, 464(7285): 45–53, Mar 2010. ISSN 1476-4687. DOI: 10.1038/nature08812.

[38] Raymond Laflamme, Junan Lin, and Tal Mor. Algorithmic cooling for resolving state preparation and measurement errors in quantum computing. *Phys. Rev. A*, 106:012439, Jul 2022. DOI: 10.1103/PhysRevA.106.012439.

[39] Daniel A. Lidar and Todd A. Brun. *Quantum Error Correction*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139034807.

[40] Junan Lin, Brandon Buonacorsi, Raymond

Laflamme, and Joel J Wallman. On the freedom in representing quantum operations. *New Journal of Physics*, 21(2):023006, feb 2019. DOI: 10.1088/1367-2630/ab075a.

[41] Junan Lin, Brandon Buonacorsi, Raymond Laflamme, and Joel J Wallman. On the freedom in representing quantum operations. *New Journal of Physics*, 21(2):023006, 2019. DOI: 10.1088/1367-2630/ab075a.

[42] Junan Lin, Joel J. Wallman, Ian Hincks, and Raymond Laflamme. Independent state and measurement characterization for quantum computers. *Phys. Rev. Research*, 3:033285, Sep 2021. DOI: 10.1103/PhysRevResearch.3.033285.

[43] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.*, 106:180504, May 2011. DOI: 10.1103/PhysRevLett.106.180504.

[44] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, Apr 2012. DOI: 10.1103/PhysRevA.85.042311.

[45] Dmitri Maslov and Martin Roetteler. Shorter stabilizer circuits via bruhat decomposition and quantum circuit transformations. *IEEE Transactions on Information Theory*, 64(7):4729–4738, 2018. DOI: 10.1109/TIT.2018.2825602.

[46] Olivia Di Matteo, John Gamble, Chris Granade, Kenneth Rudinger, and Nathan Wiebe. Operational, gauge-free quantum tomography. *Quantum*, 4:364, nov 2020. DOI: 10.22331/q-2020-11-17-364.

[47] Seth T. Merkel, Jay M. Gambetta, John A. Smolin, Stefano Poletto, Antonio D. Córcoles, Blake R. Johnson, Colm A. Ryan, and Matthias Steffen. Self-consistent quantum process tomography. *Phys. Rev. A*, 87:062119, Jun 2013. DOI: 10.1103/PhysRevA.87.062119.

[48] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Phys. Rev. A*, 77:032322, Mar 2008. DOI: 10.1103/PhysRevA.77.032322.

[49] Benjamin Nachman and Michael R. Geller. Categorizing readout error correlations on near term quantum computers. 2021. DOI: 10.48550/arXiv.2104.04607.

[50] Paul D. Nation, Hwajung Kang, Neereja Sundaresan, and Jay M. Gambetta. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum*, 2:040326, Nov 2021. DOI: 10.1103/PRXQuantum.2.040326.

[51] Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout. Gate set tomography. *Quantum*, 5:557, oct 2021. DOI: 10.22331/q-2021-10-05-557.

[52] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511976667.

[53] A. Opremcak, C. H. Liu, C. Wilen, K. Okubo, B. G. Christensen, D. Sank, T. C. White, A. Vainsencher, M. Giustina, A. Megrant, B. Burkett, B. L. T. Plourde, and R. McDermott. High-fidelity measurement of a superconducting qubit using an on-chip microwave photon counter. *Phys. Rev. X*, 11:011027, Feb 2021. DOI: 10.1103/PhysRevX.11.011027.

[54] Anargyros Papageorgiou and Joseph F. Traub. Measures of quantum computing speedup. *Phys. Rev. A*, 88:022316, Aug 2013. DOI: 10.1103/PhysRevA.88.022316.

[55] J. F. Poyatos, J. I. Cirac, and P. Zoller. Complete characterization of a quantum process: The two-bit quantum gate. *Phys. Rev. Lett.*, 78:390–393, Jan 1997. DOI: 10.1103/PhysRevLett.78.390.

[56] John Preskill. *Fault-Tolerant Quantum Computation*, pages 213–269. 1998. DOI: 10.1142/9789812385253˙0008.

[57] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, aug 2018. DOI: 10.22331/q-2018-08-06-79.

[58] D. J. Reilly. Challenges in scaling-up the control interface of a quantum computer. In *2019 IEEE International Electron Devices Meeting (IEDM)*, pages 31.7.1–31.7.6, 2019. DOI: 10.1109/IEDM19573.2019.8993497.

[59] Joschka Roffe. Quantum error correction: an introductory guide. *Contemporary Physics*, 60(3):226–245, 2019. DOI: 10.1080/00107514.2019.1667078.

[60] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz. Realization of real-time fault-tolerant quantum error correction. *Phys. Rev. X*, 11:041058, Dec 2021. DOI: 10.1103/PhysRevX.11.041058.

[61] M Saffman. Quantum computing with atomic qubits and rydberg interactions: progress and challenges. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 49 (20):202001, oct 2016. DOI: 10.1088/0953-4075/49/20/202001.

[62] P. Shor. Fault-tolerant quantum computation. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, page 56, Los Alamitos, CA, USA, oct 1996. IEEE Computer Society. DOI: 10.1109/SFCS.1996.548464.

[63] Sergei Slussarenko and Geoff J. Pryde. Photonic

quantum information processing: A concise review. *Applied Physics Reviews*, 6(4):041303, 2019. DOI: 10.1063/1.5115814.

[64] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. Fundamental limits of quantum error mitigation. *npj Quantum Information*, 8(1):114, Sep 2022. ISSN 2056-6387. DOI: 10.1038/s41534-022-00618-z.

[65] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.*, 119:180509, Nov 2017. DOI: 10.1103/PhysRevLett.119.180509.

[66] Ewout van den Berg, Zlatko K. Minev, and Kristan Temme. Model-free readout-error mitigation for quantum expectation values. *Phys. Rev. A*, 105:032620, Mar 2022. DOI: 10.1103/PhysRevA.105.032620.

[67] Lieven Vandersypen and Antoni van Leeuwenhoek. 1.4 quantum computing - the next challenge in circuit and system design. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 24–29, 2017. DOI: 10.1109/ISSCC.2017.7870244.

[68] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. DOI: 10.1017/9781108231596.

[69] A. Wehrl. The many facets of entropy. *Reports on Mathematical Physics*, 30(1):119–129, 1991. ISSN 0034-4877. DOI: 10.1016/0034-4877(91)90045-O.

[70] K. Wright, K. M. Beck, S. Debnath, J. M. Amini, Y. Nam, N. Grzesiak, J.-S. Chen, N. C. Pisenti, M. Chmielewski, C. Collins, K. M. Hudek, J. Mizrahi, J. D. Wong-Campos, S. Allen, J. Apisdorf, P. Solomon, M. Williams, A. M. Ducore, A. Blinov, S. M. Kreikemeier, V. Chaplin, M. Keesan, C. Monroe, and J. Kim. Benchmarking an 11-qubit quantum computer. *Nature Communications*, 10(1):5464, Nov 2019. ISSN 2041-1723. DOI: 10.1038/s41467-019-13534-2.

[71] Man-Hong Yung. Quantum supremacy: some fundamental concepts. *National Science Review*, 6(1):22–23, 07 2018. ISSN 2095-5138. DOI: 10.1093/nsr/nwy072.

A Completeness of Linear operator space

In this section we show that the linear operator space induced by hybrid unitary gate operations and linear classical post-processing (8) is complete.

Theorem 3. Completeness *Let $\mathcal{U}(2^n)$ be the unitary group on n-qubits. For any $\phi \in \mathbb{R}^{4^n-1\times 4^n-1}$ the matrix*

$$\begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} \in \mathcal{L}(\mathcal{U}(2^n)). \tag{50}$$

To prove the theorem, we will use a basis representation where $\mathcal{B}_L, \mathcal{B}_R$ are the Pauli basis. However, once the completeness is proved, the result carries over to any pair of basis of the traceless space. The proof of the theorem relies on the existence of two constituent families of linear operators. We call the first the *eliminator* operators.

Lemma 1 (Eliminator operators)**.** *For any $P \in \mathcal{P}$ there exist operators $E_P \in \mathcal{L}(\mathcal{U}(2^n))$ such that*

$$\begin{aligned} E_P \left|I\right\rangle\!\right\rangle = \left|I\right\rangle\!\right\rangle,\ E_P \left|P\right\rangle\!\right\rangle = \left|P\right\rangle\!\right\rangle, \\ E_P \left|Q\right\rangle\!\right\rangle = 0 \text{ for all } Q \in \mathcal{P} \setminus \{P, I\}. \end{aligned} \tag{51}$$

The second corresponds to unitary transformations that take one Pauli operator to another. We call these *permutation* operators.

Lemma 2 (Permutation operators)**.** *For any $P, Q \in \mathcal{P}$ with $P \neq Q$, there exists a unitary U_{PQ} such that $\Phi(U_{PQ})\left|P\right\rangle\!\right\rangle = \left|Q\right\rangle\!\right\rangle$. Moreover U_{PQ} can be implemented using a circuit of at most $O(n)$ depth.*

Proof of Theorem 3. We prove the theorem by constructing a set of canonical linear operators. For $P, Q \in \mathcal{P}$ let $e_{PQ} \in \mathbb{R}^{4^n-1\times 4^n-1}$ such that $[e_{PQ}]_{P,Q} = 1$ and $[e_{PQ}]_{P',Q'} = 0$ whenever $(P', Q') \neq (P, Q)$. Let $\alpha \in \mathbb{R}$ and define the linear operator E^{α}_{PQ} given by

$$E^{\alpha}_{PQ} = \begin{bmatrix} 1 & 0 \\ 0 & \alpha e_{PQ} \end{bmatrix} \tag{52}$$

We will show that $E^{\alpha}_{PQ} \in \mathcal{L}$ using an explicit construction. Showing this is sufficient to prove the theorem as every super-operator that preserves $\left|I\right\rangle\!\right\rangle$ can be written as a linear combination of these eliminators, such that the coefficients sum to unity.

Using Lemma 2 and since U_{PQ} is unitary, we get $\phi(U_{PQ})\,|P\rangle\rangle = |Q\rangle\rangle$ and for any $P' \neq P$ the orthogonality condition $\mathrm{Tr}(\phi(U_{PQ})\,|P'\rangle\rangle, Q) = 0$. Therefore, using the eliminator operators in Lemma 1 we get $E_Q\phi(U_{PQ})\,|P'\rangle\rangle = 0$ for all $P' \neq P$. So the operator E^1_{PQ} can be explicitly constructed as

$$E^1_{PQ} = E_Q\phi(U_{PQ}). \tag{53}$$

By closedness under composition from Lemma 14, we get $E^1_{PQ} \in \mathcal{L}$. For arbitrary $\alpha \in \mathbb{R}$, we can construct

$$E^\alpha_{PQ} := \alpha(E_{PQ} - E_I) + E_I. \tag{54}$$

It follows that $E^\alpha_{PQ} \in \mathcal{L}$ from closedness under linear combination in Lemma 13. Finally, any operators can be constructed as

$$\begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} = \sum_{P,Q\in\mathcal{P}\setminus I} \frac{1}{(4^n-1)^2} E_{PQ}^{(4^n-1)^2\phi_{PQ}}. \tag{55}$$

□

We now establish the existence of the *eliminators* and *permutation* operators. For $P \neq I$, let $F_P = E_P - E_I$. These are rank-1 projectors to the operator P. For consistency, we will also take $F_I = E_I$.

Proof of Lemma 1. For the case of one qubit, we can easily verify that the operators

$$F^1_I = E^1_I = \frac{1}{4}(\phi(I) + \phi(X) + \phi(Y) + \phi(Z)), \tag{56}$$

$$F^1_X = \frac{1}{4}(\phi(I) + \phi(X) - \phi(Y) - \phi(Z)), \tag{57}$$

$$F^1_Y = \frac{1}{4}(\phi(I) + \phi(Y) - \phi(X) - \phi(Z)), \tag{58}$$

$$F^1_Z = \frac{1}{4}(\phi(I) + \phi(Z) - \phi(X) - \phi(Y)), \tag{59}$$

satisfy (51). For the n-qubit case when $P = \otimes_{i=1}^n P_i$, we can construct the n-qubit projector operator as $F_P = \otimes_{i=1}^n F^1_{P_i}$. Now $E_P = F_I + F_P = \frac{1}{4^n}\sum_{P'} \phi(P') + F_P$. It is clear that the expansion of F_P in terms of superoperators $\Phi(P')$ will have positive coefficients if and only if P' commutes with P. This is because for P' and P to commute they should have either zero or an even number of anti-commuting pairs of single qubit operators when they are matched qubit-wise. From this argument, it is clear that $E_P = \frac{2}{4^n}\sum_{P':[P',P]=0} \phi(P')$ □

This is akin to what is known as *twirling* in the literature [18].

Proof of Lemma 2. Circuits that map between Pauli strings from the Clifford group are well studied in the literature. These circuits are examples of stabilizer circuits, i.e. they can be constructed using only CNOT, Hadamard, and Phase gates [23]. Any such stabilizer circuit can be constructed using only $O(n)$ depth [45]. □

B Identifiability: Proof of Theorem 1

Let s_P, s'_P, with $P \in \mathcal{P}^n$, denote the coefficients of ρ and ρ' in the Pauli basis. Then using the assertion of the theorem and (7), we have for all $k \in [D]$,

$$\begin{aligned} 2^{-n/2} \sum_{k'\in[D]} (A_{kk'} - A'_{kk'})m_{k'I} + \\ \sum_{\substack{k'\in[D],\\ P,Q\in\mathcal{P}\setminus I}} (s_P A_{kk'} - s'_P A'_{kk'})\phi(U)_{PQ} m_{k'Q} = 0, \end{aligned} \tag{60}$$

for all $U \in \mathcal{U}$. Thus for any set of unitary operators $U_1, \ldots, U_L$ and scalars $c_1, \ldots, c_L$ such that $\sum_l c_l = 1$,

$$\begin{aligned} 2^{-n/2} \sum_{k'\in[D]} (A_{kk'} - A'_{kk'})m_{k'I} + \\ \sum_{\substack{k'\in[D],\\ P,Q\in\mathcal{P}\setminus I}} (s_P A_{kk'} - s'_P A'_{kk'})\phi_{PQ} m_{k'Q} = 0, \end{aligned} \tag{61}$$

where $\phi = \sum_l c_l \phi(U_l) \in \mathcal{L}(\mathcal{U})$. Since (61) holds for any $\phi \in \mathcal{L}(\mathcal{U})$, and by Theorem 3, the linear operator $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{L}(\mathcal{U})$, we must have

$$\sum_{k'} A_{kk'} m_{k'I} = \sum_{k'} A'_{kk'} m_{k'I}. \tag{62}$$

Similarly, using Theorem 3 we get that $\begin{bmatrix} 1 & 0 \\ 0 & e_{PQ} \end{bmatrix} \in \mathcal{L}^n$ for all $P \in \mathcal{B}_R$ and $Q \in \mathcal{B}_L$. An identical argument yields for all $k \in [D]$,

$$\sum_{k'} (s_P A_{kk'} - s'_P A'_{kk'}) m_{k'Q} = 0, \tag{63}$$

or equivalently in matrix form,

$$(s_P A - s'_P A')\mathbf{m}_{\setminus I} = 0, \quad \forall P \in \mathcal{B}_R. \tag{64}$$

Using this with Lemma 12, we can see that the matrix $(s_P A - s'_P A')$ must have the following form,

$$s_P A_{kk'} - s'_P A'_{kk'} = -d_k. \tag{65}$$

Here d_k is as of yet undetermined. In the following steps, we will fix the value of d_k from (62). Defining $s_P/s'_P = \alpha$, we get

$$A' = \alpha A + \frac{\mathrm{diag}(d)}{s'_P}\mathbb{1}. \tag{66}$$

Combining with (62) and using the fact that $\sum_k m_{kI} = 2^{n/2}$, we get,

$$\sum_{k'} A'_{kk'} m_{k'I} = \alpha \sum_{k'} A_{kk'} m_{k'I} + \frac{d_k}{s'_P} 2^{n/2}, \tag{67}$$

which gives

$$d_k = \frac{s'_P(1-\alpha)\sum_{k'} A_{kk'} m_{k'I}}{2^{n/2}}. \tag{68}$$

This completes the proof of Theorem 1.

C Making a POVM linearly independent

Lemma 3. *Given a D-outcome POVM such that the linear span of this POVM has only dimension r, then we can always construct an r− outcome , linearly independent POVM by taking linear combinations of the original POVM elements.*

Proof. Our aim is to construct a new $r-$ outcome POVM,

$$M'_j = \sum_{i=1}^{D} p_{ji} M_i, \quad j \in [r]. \tag{69}$$

If the matrix p is full rank (rank r). Then it is easy to check that M'_j is satisfies condition 3. Moreover the outcomes probabilities corresponding to the new POVM can calculated from the outcome probabilities of the old POVM.

Because of the linear dependence between the POVM elements, there exists $D-r$ independent vectors $(\vec{c}_i)$ [1] in $\mathbb{R}^D$, such that $\sum_k c_i^k M_k = 0$. Now take some PSD matrix O with nonzero overlap with all the POVM operators. Define $u^k = \mathrm{Tr}(M_k O)$: it follows $\langle \vec{u}, \vec{c}_i \rangle = 0$. Thus there exists one vector $(\vec{u})$ with positive coefficients that is orthogonal to every $\vec{c}_i$.

Now let F be the r dimensional subspace of $\mathbb{R}^D$ consisting of all the vectors orthogonal to the set $\{\vec{c}_i | i = 1, \ldots, D-r\}$. We know that $\vec{u} \in F$. We now claim that F can be spanned by a set of r linearly independent, *positive vectors*.

[1] $\vec{u}$ notation hides the superscript index which runs from 1 to D.

Suppose $\vec{u}, \vec{v_1} \dots \vec{v}_{r-1}$ is a linearly independent set that spans F. We can always chose a positive α_i such that $\vec{v}'_i := \vec{v}_i + \alpha_i \vec{u}$ is positive. Now we need to prove that these newly defined positive vectors are linearly independent.

Suppose there was some linear dependence between the new vectors, such that $\vec{u} + \sum_i \vec{v}'_i a_i = 0$. This would imply a linear relation between the original $\vec{v}_i$ and $\vec{u}$. This would obviously contradict the initial statement that this is a linearly independent set that spans F.

Thus we can always construct a set of linearly independent, positive vectors $\{\vec{u}, \vec{v}'_1, \dots \vec{v}'_{r-1}\}$ that span F. For purely notational convenience define,

$$\vec{\mu}_1 := \vec{u} \ ,$$
$$\vec{\mu}_j := \vec{v}'_{j-1} \ , \quad j = 2 \dots r \ .$$

With these vectors as coefficients we can recombine the old POVM operators to construct an r-outcome POVM.,

$$M'_j = \sum_{i=1}^{D} \mu^i_j \frac{M_i}{\sum_l \mu^i_l} \ , \quad j \in [r] \ . \tag{70}$$

The positivity of $\vec{\mu}_i$ ensures that these new operators are positive. The normalization used in this relation ensures that the new operators sum to identity. Now comparing (69) and (70), we see that the desired p matrix is, $p_{ji} = \frac{\mu^i_j}{\sum_l \mu^i_l}$. By construction the matrix of μ^i_j has rank r. Matrix p is obtained by normalizing all the columns of the μ matrix. Since this cannot change the number of independent columns, p must also have rank r. Which in turn implies that $\{M'_j | j = 1, \dots, r\}$ satisfies the condition 3. □

D Detailed description of Algorithm 1

In Step 1, we want to find a non-zero state coefficient to set this as the gauge for the entire problem. From (12), it is clear that s_P must be non-zero to act as a valid gauge parameter. To this end we must first find all rows of A that are not all zeros. This is done by checking that $z^I_k \neq 0$, as this will imply that the $k-$th row of A is non-zero. Using this we can construct a set $\mathcal{K}$ that holds the location of all non-zero rows. Since,

$$z^{P,i}_k - z^I_k = s_P \sum_{k'} A_{kk'} C_{k'i} \ , \quad \forall P \in \mathcal{B}_R \ , \ i, k \in [D] \ ,$$

if $z^{P,i}_k - z^I_k = 0$, that can either be due to $s_p = 0$ or $\sum_{k'} A_{kk'} C_{k'i} = 0$.

Now to eliminate the second case, we check if $z^{P,i}_k - z^I_k = 0$ for all values of i and $k \in \mathcal{K}$. If this is the case and $s_P \neq 0$; then that is only possible if all the non-zero rows of A are in the null space of C. But if the POVM is independent then C has a rank of $D-1$ (see Lemma 4). It is easy to check from the definition of C that the null space consists only of the uniform vector. So if all the non-zero rows of A are uniform vectors, then A is an erasure channel and it violates Condition 1, necessary for simultaneous tomography. So assuming that this condition holds, $s_P \neq 0$ must imply that there exist some j, l such that $z^{P,j}_l - z^I_l \neq 0$.. Note that this pair j, l is such that $\sum_{k'} A_{lk'} C_{k'j} \neq 0$. Once found for some P can be reused to check other state-coefficients as it is state-independent. We also store this j, l for the last step of the algorithm.

In Step 2, we pick an s_R from $R \in \mathcal{C}$ as the unknown gauge for our problem. In terms of $A'(s_R)$, we have the equations,

$$z^{R,i}_k - z^I_k = \sum_{k'} A'(s_R)_{kk'} C_{k'i}, \quad \forall P \in \mathcal{B}_R, \ i, k \in [D] \ ,$$

and

$$z^I_k = \sum_{k'} \frac{A'(s_R)_{kk'} m_{k'I}}{2^{n/2}} \ .$$

Now since the columns of C along with the uniform vector form a full rank system in $\mathbb{R}^D$, we can invert this system of equations to find $A'(s_R)$.

In Step 3, we use the gauge transform equations (12) again to find relations between state coefficients

$$A'_{kk'}(s_R) = \frac{s_R}{s_P} A'_{kk'}(s_P) + (1 - \frac{s_R}{s_P}) \frac{\sum_{k'} A'_{ki}(s_P) m_{iI}}{2^{n/2}} \ .$$

Lemma 4. *The $D \times D$ covariance matrix has rank $D-1$ if the POVM is linearly independent. And the null space of the covariance matrix is spanned by the uniform vector.*

Proof. The Covariance matrix can be equivalently expressed as,

$$C_{ij} = \langle\langle \bar{M}_i | \bar{M}_j \rangle\rangle \, . \tag{71}$$

This is the Gram matrix associated with the set of traceless measurement operators, $C = W^\dagger W$, where $W = \sum_i |\bar{M}_i\rangle\langle i|$. From the definition of Gram matrix, it is clear that the rank of C will be equal to the dimension of the subspace spanned by the traceless measurement operators. Now the space spanned by the measurement operators has dimension D, by the independence assumption. The normalization constraint on the measurement operators gives,

$$\sum_i M_i = I \quad \implies \sum_i \bar{M}_i = 0 \, . \tag{72}$$

This implies that the subspace spanned by the traceless operators has dimension $D-1$ or less. Suppose that there was some other set of coefficients c_i such that $\sum_i c_i \bar{M}_i = 0$. This would then imply,

$$\sum_i c_i M_i - \left(\sum_j c_j \langle\langle M_j | I \rangle\rangle \right) \frac{I}{D} = 0 \tag{73}$$

$$\implies \sum_i \left(c_i - \frac{\sum_j c_j \langle\langle M_j | I \rangle\rangle}{D} \right) M_i = 0 \, . \tag{74}$$

This is only possible if all the c_i are the same. So the only possible linear relation between the traceless measurements is $\sum_i \bar{M}_i = 0$. This fixes the dimension of their span, and the rank of the covariance matrix to be $D-1$.

The null space of C is spanned by the uniform vector as $\sum_j C_{ij} = 0$.

□

E Elimination operators for the Pauli basis and computational measurements

For computational basis measurements, we take $\mathcal{B}_R$ to be the set of all normalized Pauli operators on $n-$qubits, $\hat{\mathcal{P}}^n = \{I/\sqrt{2}, X/\sqrt{2}, Y/\sqrt{2}, Z/\sqrt{2}\}^{\otimes n}$

$$\mathcal{B}_R = \hat{\mathcal{P}}^n \setminus \{I/\sqrt{2^n}\} \, . \tag{75}$$

For $\mathcal{B}_L$ we take the traceless part of the computational basis POVM

$$\mathcal{B}_L = \{\bar{M}_k = |k\rangle\langle k| - \frac{I}{2^n} \mid k \in [2^n]\} \, . \tag{76}$$

From this definition, it is clear that $\mathcal{B}_L^\perp$ will be the space of all fully off-diagonal operators in the computational basis.

We also denote by $\mathcal{P}_X$ ($\mathcal{P}_Z$), the set of all Pauli strings consisting of only $X(Z)$ operators and I. As before the normalized matrices are given by $\hat{Q} = Q/\sqrt{2^n}$.

Now using these definitions we prove the relations in (30) and (31).

Lemma 5. *Let $\mathcal{B}_R$ and $\mathcal{B}_L$ be as defined in (75) and (76). Let E_I be the $n-$qubit superoperator defined as in (15). Then,*

$$E_I = \frac{1}{2^n} \sum_{P \in \mathcal{P}_X} \Phi(P) \, ,$$

where $\mathcal{P}_X = \{I, X\}^{\otimes n}$.

Proof. First let us look at the one-qubit version of this superoperator, $E_I^1 = \frac{\Phi(I)+\Phi(X)}{2}$. It is clear that this superoperator satisfies all the conditions outlined in (15). It leaves the identity invariant and maps the Pauli operators to either zero or an off-diagonal operator.

$$E_I^1 |I\rangle\rangle = |I\rangle\rangle \, , \quad E_I^1 |X\rangle\rangle = |X\rangle\rangle \, ,$$
$$E_I^1 |Z\rangle\rangle = 0 \, , \quad E_I^1 |Y\rangle\rangle = 0 \, .$$

Now by elementary linear algebra, the following relation holds

$$\Phi(A) \otimes \Phi(B) = \Phi(A \otimes B) \ . \tag{77}$$

This is because for any operators V and W

$$\begin{aligned} \Phi(A) \otimes \Phi(B) \left|V\right\rangle\rangle \otimes \left|W\right\rangle\rangle &= (AVA^\dagger) \otimes (BWB^\dagger) \\ &= (A \otimes B)(V \otimes W)(A \otimes B)^\dagger \\ &= \Phi(A \otimes B) \left|V\right\rangle\rangle \otimes \left|W\right\rangle\rangle \ . \end{aligned}$$

From this elementary fact, $E_I = \frac{1}{2^n} \sum_{P \in \mathcal{P}_X} \Phi(P) = \left(\frac{\Phi(I)+\Phi(X)}{2}\right)^{\otimes n} = (E_I^1)^{\otimes n}$. This implies that E_I acting on any operator in $\mathcal{B}_R$ will either give zero or an operator in $\mathcal{P}_X$. On the other hand, it is clear that E_I leaves the identity invariant. Thus $E_I = \frac{1}{2^n} \sum_{P \in \mathcal{P}_X} \Phi(P)$ is consistent with the definition in (15). □

Now, to prove the same for E_{Pi} eliminators defined in (16), first we define some auxiliary eliminators. These eliminators effectively map between normalized Pauli strings. Let $P \in \mathcal{B}_R$ and $Q \in \mathcal{P}_Z \setminus \{I\}$.

$$\begin{aligned} &E_{PQ} \left|I\right\rangle\rangle = \left|I\right\rangle\rangle, \ E_{PQ} \left|P\right\rangle\rangle - \left|\hat{Q}\right\rangle\rangle \in \mathcal{B}_L^\perp \ , \\ &E_{PQ} \left|P'\right\rangle\rangle \in \mathcal{B}_L^\perp \quad \forall P' \in \mathcal{B}_R \setminus \{P\} \ . \end{aligned} \tag{78}$$

Using these we can write E_{Pi} as follows,

$$E_{Pi} = (1 - \sum_{Q \in \mathcal{P}_Z \setminus \{I\}} H_{iQ}) E_I + \sum_{Q \in \mathcal{P}_Z \setminus \{I\}} H_{iQ} E_{PQ} \ , \tag{79}$$

where $H_{iQ} = \frac{\langle i|Q|i\rangle}{\sqrt{2^n}} = \mathrm{Tr}(\bar{M}_i \hat{Q})$. Since $\mathcal{P}_Z \setminus I$ is an orthogonal basis for traceless diagonal matrices, we have $\bar{M}_i = \sum_{Q \in \mathcal{P}_Z \setminus \{I\}} H_{iQ} \hat{Q}$. From this, it is easy to check that (79) holds.
Now, if U_{PQ} is a member of the Clifford group on n-qubits that maps from P to Q, then $E_{PQ} = E_{QQ} \Phi(U_{PQ})$. So to show the relation in (31), we only need to show that $E_{QQ} = \frac{2}{2^n} \sum_{\substack{Q' \in \mathcal{P}_X \\ Q':[Q,Q']=0}} \Phi(Q')$.

Lemma 6. *Let $\mathcal{B}_R$ and $\mathcal{B}_L$ be as defined in (75) and (76). Let E_{QQ} be the $n-$qubit superoperator defined as in (78) for all $Q \in \mathcal{P}_Z \setminus I$. Then,*

$$E_{QQ} = \frac{2}{2^n} \sum_{\substack{Q' \in \mathcal{P}_X \\ Q':[Q,Q']=0}} \Phi(Q') \ ,$$

where $\mathcal{P}_X = \{I, X\}^{\otimes n}$.

Proof. The proof idea here is similar to that used in Lemma 1.

For $n = 1$ we can easily verify that $E_{ZZ} = \Phi(I)$. Let $F_{QQ} = E_{QQ} - E_I$. For the single qubit case we get $F_{ZZ} = \frac{1}{2}(\Phi(I) - \Phi(X))$. For consistency we will denote $F_{II} = E_I$.
Now, for any $n-$qubit Z string $Q = Q_1 \otimes \ldots Q_n$, we can easily check that $F_{QQ} = \otimes_{i=1}^n F_{Q_i Q_i}$. Since $E_{QQ} = E_I + F_{QQ}$, the operator strings that remain in the expansion for E_{QQ} will be those which have an even number of single qubit X operators at positions where Z operators are present in Q. In the positions where Z does not exist in Q, the operators in the expansion can either be X or I. This precisely describes all the operator strings in $\mathcal{P}_X$ that commute with Q. □

F Sample complexity of simultaneous tomography

F.1 Randomized version of simultaneous tomography in the computational basis.

First we describe how to use a finite number of measurement shots to estimate the $z-$values.

Finite sample estimator for z_k^I
Choose an n qubit Pauli string uniformly at random from $\mathcal{P}_X$. Now apply this operator to the quantum state and measure the outcome. Repeat this procedure N times using independent copies of the state and record the N measurement outcomes.

Let now $X_1^l, \dots X_N^l$ be binary random variables such that X_i^l records whether the $i-$the measurement outcome is l. We then define the estimate

$$\hat{z}_l^I := \frac{1}{N} \sum_i X_i^l \, . \tag{80}$$

This is an unbiased estimate for z_l^I

$$\begin{aligned} \mathbb{E} X_i^l = Pr(X_i^l = 1) &= \frac{1}{2^n} \sum_{P \in \mathcal{P}_X} Pr(X_i^l = 1 | P) \\ &= \frac{1}{2^n} \sum_{P \in \mathcal{P}_X} \tilde{y}_l(P) = z_l^I \, . \end{aligned}$$

Notice that the same N measurement outcomes can be processed in different ways to get estimates z_l^I for all $l \in [2^n]$.

Finite sample estimator for $z_k^{P,i}$

Let us define

$$z_l^{PQ} := \frac{2}{2^n} \sum_{\substack{Q' \in \mathcal{P}_X \\ [Q',Q]=0}} \tilde{y}_l(Q' U_{PQ}) \, , \tag{81}$$

where U_{PQ} is a member of the Clifford group that maps P to Q. From (31)

$$z_l^{P,i} = (1 - \sum_{Q \in \mathcal{P}_Z \setminus I} H_{iQ}) z_l^I + \sum_{Q \in \mathcal{P}_Z \setminus I} H_{iQ} z_l^{PQ} \, .$$

We can estimate z_l^{PQ} in the following way. Choose an n qubit Pauli string Q' uniformly at random from with Q. Since half of $\mathcal{P}$ commutes with Q this can be done with constant overhead. Now apply $Q' U_{PQ}$ to the quantum state and measure the outcome. Repeat this procedure N' times using independent copies of the state and record the N' measurement outcomes.

Let $X_1^l, \dots X_{N'}^l$ be binary random variables such that X_i^l records whether the $i-$the measurement outcome is l. Then we can define the following estimates

$$\hat{z}_l^{PQ} := \frac{1}{N'} \sum_i X_i^l \, , \tag{82}$$

$$\hat{z}_l^{P,i} := (1 - \sum_{Q \in \mathcal{P}_Z \setminus I} H_{iQ}) \hat{z}_l^I + \sum_{Q \in \mathcal{P}_Z \setminus I} H_{iQ} \hat{z}_l^{PQ} \, . \tag{83}$$

These are also unbiased estimates since

$$\mathbb{E} X_i^l = Pr(X_i^l = 1) = \frac{2}{2^n} \sum_{\substack{Q' \in \mathcal{P}_X \\ [Q',Q]=0}} Pr(X_i = 1 | Q') \tag{84}$$

$$= \frac{2}{2^n} \sum_{\substack{Q' \in \mathcal{P}_X \\ [Q',Q]=0}} \tilde{y}_l(Q' U_{PQ}) = z_l^{PQ} \, . \tag{85}$$

The unbiasedness of $\hat{z}_l^{P,i}$ hence follows from linearity of expectation. Notice that the same N' measurement outcomes can be processed in different ways to get estimates $\hat{z}_l^{PQ}$ for all $l \in [2^n]$. These estimates can then be combined to get $\hat{z}_l^{P,i}$.

The number of shots required to guarantee a certain error in these estimates is given in Lemma 7 and 8. The proofs use the Hoeffding's inequality and properties of subgaussian random variables. In Algorithm 2 we describe how this estimates can be used to perform simultaneous tomography.

Algorithm 2: Finite-shot simultaneous tomography in computational basis

Data: $\beta, ||A||_{\text{uni}}$

// Refer Theorem 2 for the number of shots required for each step

// Step 1. Find non-zero coefficients of the state

$\mathcal{C} \leftarrow \{\}$ // Empty set

Estimate $\{\hat{z}_k^I | k \in [2^n]\}$

for $P \in \mathcal{B}_R$ **do**

  Estimate $\{\hat{z}_k^{P,i} | k, i \in [2^n]\}$

  $s_P \leftarrow 0$

  if $\max_{k,i\in[2^n]} |\hat{z}_k^{P,i} - \hat{z}_k^I| \geq 1.005\beta||A||_{uni}$ **then**

    $\mathcal{C} \leftarrow \mathcal{C} \bigcup \{P\}$

  end

end

// Step 2. Find A up to gauge symmetry

choose $R \in \mathcal{C}$

for $(k, i) \in [2^n] \times [2^n]$ **do**

  $A'_{ki} \leftarrow \hat{z}_k^{R,i}$

end

// Step 3. Find other state coefficients up to a multiplicative constant

Re-estimate $\{\hat{z}_k^{R,i} | k, i \in [2^n]\}$ with $\epsilon = \frac{\beta||A||_{\text{uni}}}{4}$ // see Lemma 8

Re-estimate $\{\hat{z}_k^I | k \in [2^n]\}$ with $\epsilon = \frac{\beta||A||_{\text{uni}}}{4}$ // see Lemma 7

$i', l' \leftarrow \text{argmax}_{i,l} |\hat{z}_l^{R,i} - \hat{z}_l^I|$

for $P \in \mathcal{C}$ **do**

  // In this step use the $\hat{z}$ estimates that were compupted from the most amount of shots

  $\frac{s_P}{s_R} \leftarrow \frac{\hat{z}_{l'}^{P,i'} - \hat{z}_{l'}^I}{\hat{z}_{l'}^{R,i'} - \hat{z}_{l'}^I}$

end

return $\{(P, \frac{s_P}{s_R}) | P \in \mathcal{C}\}, \quad A'$

F.2 Proof of Theorem 2

Proof. We will use randomized measurements to estimate the z values required for the algorithm.

The definitions in (30) and (31) show that the z values for performing simultaneous tomography can be estimated by using simple Monte-Carlo estimates as defined in (80) and (83).
For instance, from the definition of E_I in (30) we see that z_k^I can be estimated by choosing a random unitary from $\mathcal{P}_X$, applying it to the state and recording the noisy measurement outcomes. The number of shots required to guarantee a certain accuracy in these estimates with high probability are given by Lemma 7 and Lemma 8.

The covariance matrix for computational basis measurements is given by

$$C_{ik} = \delta_{ik} - \frac{1}{2^n} .$$

From (23), given any P, we have

$$z_k^{P,i} - z_k^I = s_P(A_{ki} - z_k^I) . \tag{86}$$

The first step of Algorithm 1 is finding the non-zero state coefficients. To perform the analogous step in the randomized setting we set a positive parameter $0 < \beta < ||\rho||_{\text{mix}}$ and construct a set $\mathcal{C} \subset \mathcal{P}$ such that for every $P \in C$ we can guarantee with high probability that $|s_P| > \beta$. Similarly, we can also show that with high probability, if $P \notin \mathcal{C}$ then $|s_P| < 1.01\beta$. We show in Lemma 9 that such a set can be constructed using $O(8^n \frac{cn+\log(1/\delta)}{\beta^2||A||_{\text{uni}}^2})$ shots.
From this set $\mathcal{C}$ we can choose R such that s_R can be used as the unknown gauge in the problem, which gives us the noise matrix upto s_R

$$A'_{l\,i}(s_R) = z_l^{R,i} . \tag{87}$$

Due to the simple form of the covariance matrix, the estimation error in the noise matrix is also given by Lemma 8.

For the final phase of the algorithm, we first have to find an element of the noise matrix that is sufficiently bounded away from its corresponding row average. To this end let us first estimate $|\hat{z}_l^{R,i} - \hat{z}_l^I|$ up to a max error of ϵ with high probability using Lemmas 8 and 7. Now let $i^*, l^* = \text{argmax}_{i,l}|z_l^{R,i} - z_l^I|$. From (86) $\max_{i,l}|z_l^{R,i} - z_l^I| = |s_R|||A||_{\text{uni}}$. So $|\hat{z}_{l^*}^{R,i^*} - \hat{z}_{l^*}^I|$ must be ϵ close to $|s_R|||A||_{\text{uni}}$ w.h.p. Let $i', l' = \text{argmax}_{i,l}|\hat{z}_l^{R,i} - \hat{z}_l^I|$, then with high probability the following relations hold

$$|z_{l'}^{R,i'} - z_{l'}^I| \geq |\hat{z}_{l'}^{R,i'} - \hat{z}_{l'}^I| - \epsilon \,, \tag{88}$$

$$\geq |\hat{z}_{l^*}^{R,i^*} - \hat{z}_{l^*}^I| - \epsilon \,, \tag{89}$$

$$\geq |s_R|||A||_{\text{uni}} - 2\epsilon \,. \tag{90}$$

Substituting (86) in the LHS of the above relation gives

$$|A_{i'l'} - z_{l'}^I| \geq ||A||_{\text{uni}} - \frac{2\epsilon}{|s_R|} \,. \tag{91}$$

Since $|s_R| \in \mathcal{C}$, we have that $|s_R| \geq \beta$. So choosing $\epsilon = \frac{||A||_{\text{uni}}\beta}{4}$ would give us $|A_{i'l'} - z_{l'}^I| \geq ||A||_{\text{uni}}/2$ w.h.p..

According to Lemma 8, the number of measurements required to find the indices i' and l' will be $O(2^n \frac{cn+\log(1/\delta)}{\beta^2||A||_{\text{uni}}^2})$. Using these indices, for every $P \in \mathcal{C}$ we can estimate $\frac{s_P}{s_R}$ as

$$\frac{\widehat{s_P}}{s_R} = \frac{|\hat{z}_{l'}^{P,i'} - \hat{z}_{l'}^I|}{|\hat{z}_{l'}^{R,i'} - \hat{z}_{l'}^I|} \,. \tag{92}$$

The key observation here is that the true values of both the numerator and the denominator in the above expression is greater than $\beta||A||_{\text{uni}}/2$ because of the i', l' indices we have chosen. Now using this observation in Lemma 10, we show $O(2^n \frac{cn+\log(1/\delta)}{\epsilon^2\beta^2||A||_{\text{uni}}^2})$ measurements are sufficient to get the above estimate to within ϵ in multiplicative error.

Now we have to repeat this procedure for every $P \in \mathcal{C}$ with the same values of i' and l' to get estimates for the state coefficeints up to gauge.

From the standard union-bound argument we can show that a total of $O(2^n|\mathcal{C}|\frac{cn+\log(1/\delta)}{\epsilon^2\beta^2||A||_{\text{uni}}^2})$ measurements are sufficient to ensure a maximum multiplicative error of ϵ with probability at least $1-\delta$.

□

F.3 Proof of Corollary 2

Proof. The argument here is similar to the proof of Theorem 2. To begin with, perform the first step of the randomized algorithm with a threshold $\beta = \epsilon \leq ||\rho||_{\text{mix}}/2$ to construct a $\mathcal{C}$. We set $\hat{s}_P = 0$ for any $P \notin \mathcal{C}$. This gives an estimate with additive error of ϵ for all $P \notin \mathcal{C}$. From the proof of Lemma 9 we know that this construction requires the estimation of $|z_k^{P,i} - z_k^I|$ with error $\epsilon' = 0.005\epsilon||A||_{\text{uni}}$.

Let

$$R, i', k' = \underset{P\in\mathcal{C}, i,k\in[2^n]}{\text{argmax}} |\hat{z}_k^{P,i} - \hat{z}_k^I| \,, \tag{93}$$

$$P^* = \underset{P\in\mathcal{C}}{\text{argmax}}|z_k^{P,i} - z_k^I| = \underset{P\in\mathcal{C}}{\text{argmax}}|s_P| \,. \tag{94}$$

From the definition of P^*, we have $|z_k^{P^*,i} - z_k^I| = ||\rho||_{\text{mix}}|A_{ik} - z_i^I| \leq ||\rho||_{\text{mix}}||A||_{\text{uni}}$.

For any i, k we have

$$|s_R||A_{ik} - z_i^I| = |z_k^{R,i} - z_k^I| \tag{95}$$

$$\geq |\hat{z}_k^{R,i} - \hat{z}_k^I| - 0.005\epsilon||A||_{\text{uni}} \tag{96}$$

$$\geq |\hat{z}_k^{P^*,i} - \hat{z}_k^I| - 0.005\epsilon||A||_{\text{uni}} \tag{97}$$

$$\geq |z_k^{P^*,i} - z_k^I| - 0.01\epsilon||A||_{\text{uni}} \tag{98}$$

$$= ||\rho||_{\text{mix}}|A_{ik} - z_i^I| - 0.01\epsilon||A||_{\text{uni}} \tag{99}$$

Maximizing this inequality over all $i, k \in [2^n]$ we have

$$|s_R| \geq ||\rho||_{\text{mix}} - 0.01\epsilon \,. \tag{100}$$

From this, with high probability for all $P \in \mathcal{C}$ we have

$$\left|\frac{s_P}{s_R}\right| \leq \frac{||\rho||_{\text{mix}}}{||\rho||_{\text{mix}} - 0.01\epsilon} \leq \frac{1}{0.995} \,. \tag{101}$$

Here we have used the fact that $\epsilon \leq ||\rho||_{\text{mix}}/2$.

Using the above fact in the multiplicative error bound in step 3 of Theorem 2 gives $Pr\left(\max_{P\in\mathcal{C}} \left|\frac{\widehat{s_P}}{s_R} - \frac{s_P}{s_R}\right| > \frac{\epsilon}{0.995}\right) \leq \delta$.

Notice that the error incurred in the last step is slightly higher than ϵ. But this can be rectified by substituting ϵ with a slightly lower value (0.995ϵ) in the above procedure.

The sample complexity for this procedure can be found by adding the sample complexities of the first and third step of Theorem 2.

□

F.4 Technical lemmas for randomized measurements

Lemma 7 (Estimating z^I values)**.** *For an $n-$qubit system, let $N = O(\frac{cn+\log(1/\delta)}{\epsilon^2})$, for a constant $c < 10$. By post-processing N randomized noisy measurement outcomes obtained from applying a random operator in $\mathcal{P}_X$ to the state ρ, we can find $\hat{z}_l^I$ such that*

$$Pr(\max_{l\in[2^n]} |\hat{z}_l^I - z_l^I| > \epsilon) \leq \delta \,. \tag{102}$$

Proof. Consider the estimate $\hat{z}_l^I$ computed as defined in 80. By Hoeffding's inequality, we can obtain a tail bound for these estimates

$$Pr(|\hat{z}_l^I - z_l^I| > \epsilon) \leq 2e^{-2N\epsilon^2} \,. \tag{103}$$

Choosing $N = \frac{c\log(2^n/\delta)}{\epsilon^2}$ for a constant c gives us

$$Pr(|\hat{z}_l^I - z_l^I| > \epsilon) \leq \frac{\delta}{2^n} \,. \tag{104}$$

Let $\mathcal{A}_l$ be the event that $|\hat{z}_l^I - z_l^I| \leq \epsilon$, then using the union bound

$$\begin{aligned} Pr(\max_{l\in[2^n]} |\hat{z}_l^I - z_l^I| \leq \epsilon) &= Pr(\wedge_{l=1}^{2^n} \mathcal{A}_l) && (105)\\ &= 1 - Pr(\vee_{l=1}^{2^n} \bar{\mathcal{A}}_l) && (106)\\ &> 1 - \delta \,. && (107) \end{aligned}$$

□

Lemma 8 (Estimating $z^{P,i}$ values)**.** *For an $n-$qubit system, let $N = O(2^n \frac{cn+\log(1/\delta)}{\epsilon^2})$, for a constant $c < 10$. Given a Pauli string $P \neq I$, by post-processing N randomized noisy measurement outcomes obtained from applying a random unitary from an efficiently characterizable subset of the Clifford group, we can find $\hat{z}_l^{Pi}$ such that*

$$Pr(\max_{l,i\in[2^n]} |\hat{z}_l^{Pi} - z_l^{Pi}| > \epsilon) \leq \delta \,. \tag{108}$$

Proof. Consider the finite sample estimate defined in (82) for z_l^{PQ} using N' measurement shots. Now, let $e_l^{PQ} = \hat{z}_l^{PQ} - z_l^{PQ}$. By Hoeffding's inequality, we find that, e_l^{PQ} is a sub-gaussian random variable [68]:

Definition 2. Subgaussian random variable *A zero-mean random variable X is Subgaussian with a variance proxy of σ^2 if*

$$Pr(|X| > t) \leq 2e^{-\frac{2t^2}{\sigma^2}} \,. \tag{109}$$

We denote this by: $X \sim \texttt{SubG}(\sigma^2)$.

The tail bound from Hoeffding's inequality gives us

$$e_l^{PQ} \sim \texttt{SubG}\left(\frac{1}{N'}\right) \,. \tag{110}$$

In total, estimating all $\hat{z}_l^{PQ}$ for all $Q \in \mathcal{P}_Z \setminus I$ and $l \in 2^n$ requires $N = O(2^n N')$ independent measurement outcomes.

Similarly, we can use N'' randomized measurements to estimate z_l^I as in Lemma 7. We define the error $e_l^I = (1 - \sum_{Q\in\mathcal{P}_Z\setminus I} H_{iQ})(\hat{z}_l^I - z_l^I)$. From (103) we have

$$e_l^I \sim \texttt{SubG}\left(\frac{(1-\sum_{Q\in\mathcal{P}_Z\setminus I} H_{iQ})^2}{N''}\right) . \tag{111}$$

Using these we can compute the following estimates $\hat{z}_l^{P,i} = (1 - \sum_{Q\in\mathcal{P}_Z\setminus I} H_{iQ})\hat{z}_l^I + \sum_{Q\in\mathcal{P}_Z\setminus I} H_{iQ}\hat{z}_l^{PQ}$, for all $i, l \in [2^n]$. With the total number of measurements required being $N = N'' + 2^n N'$.
The error in this estimate can be computed as, $e_l^{P,i} = \hat{z}_l^{P,i} - z_l^{P,i} = e_l^I + \sum_{Q\in\mathcal{P}_Z\setminus I} H_{iQ} e_l^{PQ}$. This error is a linear combination of subgaussian random variables, which is also subgaussian by the following fact [68].

Fact 1. (Theorem 2.6.3 in [68]) *Let $\{X_i \sim \texttt{SubG}(\sigma_i^2) \;\; \forall i \in [D]\}$ be independent, mean zero, subgaussian random variables, and $a = (a_1, \ldots, a_D) \in \mathbb{R}^D$. Then, for every $t \geq 0$, we have*

$$Pr(|\sum_i a_i X_i| > t) \leq 2\exp\left(\frac{-2t^2}{\sigma^2||a||_2^2}\right),$$

where $\sigma^2 = \max_i \sigma_i^2$.

From this we have,

$$e_l^{P,i} \sim \texttt{SubG}\left(\max\left(\frac{(1-\sum_{Q\neq I} H_{iQ})^2}{N''}, \frac{1}{N'}\right)(1 + \sum_{Q\neq I} H_{iQ}^2)\right) . \tag{112}$$

Now, H_{iQ} defined as $\frac{\langle i|Q|i\rangle}{2^{n/2}}$ is just the $n-$qubit Hadamard operator. From its unitarity we have $1+\sum_{Q\neq I} H_{iQ}^2 < 2$. The first row and column of H (corresponding to $|i = 1\rangle = |0\rangle \otimes \ldots \otimes |0\rangle$ and $Q = I$ respectively) is a vector of all $1/\sqrt{2^n}$. Again from unitarity of H we have

$$\sum_{Q\neq I} H_{1Q} = \sqrt{2^n} - \frac{1}{\sqrt{2^n}} , \tag{113}$$

$$\sum_{Q\neq I} H_{iQ} = -\frac{1}{\sqrt{2^n}},\; i \neq 1 . \tag{114}$$

This gives in the worst case $(1 - \sum_{Q\neq I} H_{iQ})^2 < 2^n$.
Combining these upper-bounds in (112) we get

$$e_l^{P,i} \sim \texttt{SubG}\left(2\max\left(\frac{2^n}{N''}, \frac{1}{N'}\right)\right) , \tag{115}$$

and if we choose $N'' = 2^n N'$, we get $e_l^{P,i} \sim \texttt{SubG}\left(\frac{2}{N'}\right)$, with the total number of measurements required being $N = 2^n N' + N'' = O(2^n N')$. From the definition of Subgaussian random variables it follows that

$$Pr(|e_l^{P,i}| > \epsilon) \leq 2\exp\left(-N'\epsilon^2\right) .$$

Choosing then $N' = \frac{c\log(4^n/\delta)}{\epsilon^2}$ for a constant c gives us

$$Pr(|e_l^{P,i}| > \epsilon) \leq \frac{\delta}{4^n} . \tag{116}$$

Using the same union bound argument used in Lemma 7, we can show that

$$Pr(\max_{l,i\in[2^n]} |e_l^{P,i}| > \epsilon) \leq \delta , \tag{117}$$

and the total number of measurements required is $N = O(2^n N') = O(2^n \frac{cn+\log(1/\delta)}{\epsilon^2})$

□

Lemma 9. *Given $0 < \beta < ||\rho||_{mix}/2$, we can construct $\mathcal{C} \subset \mathcal{B}_R$ such that*

1. *For every $P \in \mathcal{C}$ we can guarantee with probability $1 - \delta$ that $|s_P| \geq \beta$.*

2. For every $P \notin \mathcal{C}$ we can guarantee with probability $1-\delta$ that $|s_P| < 1.01\beta$.

The construction of such a $\mathcal{C}$ requires a total of $O(8^n \frac{cn+\log(1/\delta)}{\beta^2 ||A||^2_{uni}})$ randomized measurements.

Proof. From the definition of $||A||_{\text{uni}}$ we know that for every $P \in \mathcal{P}$

$$\max_{i,k\in[2^n]} |z_k^{P,i} - z_k^I| = |s_P| \; ||A||_{\text{uni}} \; . \tag{118}$$

From Lemma 7 and 8, for each P, we can estimate $|\hat{z}_k^{P,i} - \hat{z}_k^I|$ such that, with probability $1-\delta$, the maximum error in the LHS is at most $0.005\beta||A||_{\text{uni}}$ using $O(2^n \frac{cn+\log(1/\delta)}{\beta^2||A||^2_{\text{uni}}})$ measurements.
Once we compute these estimates for every $P \in \mathcal{P}$ by using a total of $O(8^n \frac{cn+\log(1/\delta)}{\beta^2||A||^2_{\text{uni}}})$ measurements, we can define $\mathcal{C}$ such that

$$\mathcal{C} = \{P \mid \max_{k,i\in[2^n]} |\hat{z}_k^{P,i} - \hat{z}_k^I| \geq 1.005\beta||A||_{\text{uni}}\} \; . \tag{119}$$

From the error in the estimates we can guarantee with high probability that for every $P \in \mathcal{C}$ we will have $\max_{k,i\in[2^n]} |z_k^{P,i} - z_k^I| \geq \beta \; ||A||_{\text{uni}}$ and hence $|s_P| \geq \beta$. Similarly, if $P \notin \mathcal{C}$, then $\max_{k,i\in[2^n]} |\hat{z}_k^{P,i} - \hat{z}_k^I| < 1.005\beta||A||_{\text{uni}}$, which gives $\max_{k,i\in[2^n]} |z_k^{P,i} - z_k^I| < 1.01\beta \; ||A||_{\text{uni}}$ with high probability. This in turn implies that $|s_P| < 1.01\beta$.

□

Lemma 10. *Suppose we have $P, R \in \mathcal{P} \setminus I$, such that $|s_P|, |s_R| > \alpha$. And we also know $l, i \in [2^n]$, such that $|A_{il} - z_l^I| > \gamma$. Then we can estimate $\frac{s_P}{s_R}$ such that*

$$Pr(\left|\widehat{\frac{s_P}{s_R}} - \frac{s_P}{s_R}\right| > \epsilon \left|\frac{s_P}{s_R}\right|) \leq \delta \; , \tag{120}$$

using $N = O(2^n \frac{cn+\log(1/\delta)}{\epsilon^2\alpha^2\gamma^2})$ randomized measurements.

Proof. We know that for indices satisfying the assumptions in the lemma

$$\frac{s_P}{s_R} = \frac{z_l^{P,i} - z_l^I}{z_l^{R,i} - z_l^I} \; , \tag{121}$$

So an estimate for the ratio s_P/s_R can be computed using ratios of the appropriate estimates of z values obtained from measurements. From (86) we also know that

$$|z_l^{P,i} - z_l^I| \geq \alpha\gamma \; . \tag{122}$$

From the assumption in the lemma, we know an l, i for which this condition holds. Now using $N = O(2^n \frac{cn+\log(1/\delta)}{\epsilon^2\alpha^2\gamma^2})$ randomized measurements we can find estimates for $z_l^{P,i} - z_l^I$ such that

$$Pr(|(\hat{z}_l^{P,i} - \hat{z}_l^I) - (z_l^{P,i} - z_l^I)| > \frac{\epsilon \; \alpha\gamma}{2}) \leq \delta \; . \tag{123}$$

Using Lemma 11, we have w.h.p,

$$\left|\frac{\hat{z}_l^{P,i} - \hat{z}_l^I}{\hat{z}_l^{R,i} - \hat{z}_l^I} - \frac{s_P}{s_R}\right| \leq \epsilon \left|\frac{s_P}{s_R}\right| \; . \tag{124}$$

□

Lemma 11 (Error in ratios)**.** *Let $\hat{x} = x + \epsilon_x$ and $\hat{y} = y + \epsilon_y$ such that $|x|, |y| \geq \alpha$ and $|\epsilon_x|, |\epsilon_y| \leq \epsilon \leq \frac{\alpha}{2}$. Then*

$$\left|\frac{\hat{x}}{\hat{y}} - \frac{x}{y}\right| \leq O\left(\frac{\epsilon}{\alpha}\right) \left|\frac{x}{y}\right| \; . \tag{125}$$

Proof.

$$\frac{\hat{x}}{\hat{y}} = \frac{x+\epsilon_x}{y+\epsilon_y} = \left(\frac{x}{y}\right) \frac{1+\frac{\epsilon_x}{x}}{1+\frac{\epsilon_y}{y}} \; , \tag{126}$$

so $\frac{1+\frac{\epsilon_x}{x}}{1+\frac{\epsilon_y}{y}}$ is the exact multiplicative error term that we have to bound.
Since $|\epsilon_x/x| < \frac{1}{2}$, we have $\frac{1}{2} < 1 + \frac{\epsilon_x}{x} < \frac{3}{2}$. Also since $|\epsilon_y/y| < \frac{1}{2}$, we have, $1 - \epsilon_y/y \leq \frac{1}{1+\epsilon_y/y} \leq 1 + 2|\epsilon_y/y|$.

Lower bound on error

$$\frac{1+\frac{\epsilon_x}{x}}{1+\frac{\epsilon_y}{y}} \geq (1+\frac{\epsilon_x}{x})(1-\frac{\epsilon_y}{y}) \tag{127}$$

$$= 1+\frac{\epsilon_x}{x}-\frac{\epsilon_y}{y}(1+\frac{\epsilon_x}{x}) \tag{128}$$

$$\geq 1+\frac{\epsilon_x}{x}-c\frac{\epsilon_y}{y} \tag{129}$$

$$\geq 1-\frac{\epsilon}{\alpha}-c\frac{\epsilon}{\alpha} \geq 1-(1+c)\frac{\epsilon}{\alpha}\,. \tag{130}$$

Where c is either $\frac{1}{2}$ or $\frac{3}{2}$ depending on the sign of $\frac{\epsilon_y}{y}$.

Upper bound on error

$$\frac{1+\frac{\epsilon_x}{x}}{1+\frac{\epsilon_y}{y}} \leq (1+\frac{\epsilon_x}{x})(1+2|\frac{\epsilon_y}{y}|) \tag{131}$$

$$\leq 1+\frac{\epsilon_x}{x}+3\left|\frac{\epsilon_y}{y}\right| \tag{132}$$

$$\leq 1+4\frac{\epsilon}{\alpha}\,. \tag{133}$$

Combining these two bounds on the multiplicative error gives us the claimed inequalities in the lemma.

□

G Properties of measurement operators

Recall that the overlap of the POVM on any traceless basis $\mathcal{B}_L$ is defined as

$$m_{kI} = \langle\langle M_k|\hat{I}\rangle\rangle, \quad m_{kQ} = \langle\langle M_k|Q\rangle\rangle,\ Q \in \mathcal{B}_L. \tag{134}$$

Define the $D \times 4^n$ matrix $\mathbf{m}$ with elements given by m_{kQ}, and let $\mathbf{m}_{\setminus I}$ be the submatrix of $\mathbf{m}$ obtained by removing the column corresponding to $Q = I$. Then the following lemma holds.

Lemma 12. *If the POVM is linearly independent, then the matrix* $\mathbf{m}$ *has full row rank* D*. Furthermore, the matrix* $\mathbf{m}_{\setminus I}$ *has row rank* $D-1$ *and the only vector* v *in the left null space that satisfies*

$$v\mathbf{m}_{\setminus I} = 0, \tag{135}$$

is given by $v = \mathbb{1}$ *which is the vector of all ones.*

Proof. The fact that the matrix $\mathbf{m}$ has full row rank D follows directly from linear independence of the POVM. Consequently, the matrix $\mathbf{m}_{\setminus I}$ has row rank $D-1$ and must have a one dimensional left null space. By definition of POVM, $\sum_{k\in[D]} M_k = I$. Therefore, for all $Q \in \mathcal{B}_L$ we get

$$\sum_{k\in[D]} m_{kQ} = \mathrm{Tr}(Q\sum_{k\in[D]} M_k) = \mathrm{Tr}(QI) = 0, \tag{136}$$

since $Q \in \mathcal{B}_L$ is traceless. □

H Induced linear operator space of unitary operators

Recall that for any subset $S \subseteq \mathcal{U}$ of unitary operators on n qubits, the induced linear operators space representing hybrid quantum-classical operations is given by (11)

$$\mathcal{L}(\mathcal{S}) = \left\{\sum_l c_l\Phi(U_l) \mid U_l \in \mathcal{S},\ \sum_l c_l = 1\right\}. \tag{137}$$

These induced operator spaces $\mathcal{L}(S)$ have many natural properties as given below.

Lemma 13 (Closedness under linear combination)**.** *Let $S \subseteq \mathcal{U}$ and let $\Phi_1, \ldots, \Phi_m \in \mathcal{L}(S)$. Then for any $c_1, \ldots, c_m$ such that $\sum_{i=1}^m c_i = 1$, we have $\sum_{i=1}^m c_i \Phi_i \in \mathcal{L}(S)$.*

Proof. Follows directly from the definition of $\mathcal{L}(S)$. $\Phi_i \in \mathcal{L}(S)$ implies that

$$\Phi_i = \sum_{l=1}^{L_i} c_l^i \Phi(U_l^i), \quad \sum_{l=1}^{L_i} c_l^i = 1, \qquad i = 1, 2, \tag{138}$$

where all $U_l^i \in S$. Then $\sum_i c_i \Phi_i = \sum_{i,l} c_i c_l^i \Phi(U_l^i) \in \mathcal{L}(S)$ □

Lemma 14 (Closedness under composition)**.** *Let $S \subseteq \mathcal{U}$ be such that for all $U_1, U_2 \in S$ we have that the unitary $U_1 U_2 \in S$. Let $\Phi_1, \Phi_2 \in \mathcal{L}(S)$. Then the composition of the these operators $\Phi_1 \Phi_2 \in \mathcal{L}(S)$.*

Proof of Lemma 14. By definition of $\mathcal{L}(S)$ we have

$$\Phi_i = \sum_{l=1}^{L_i} c_l^i \Phi(U_l^i), \quad \sum_{l=1}^{L_i} c_l^i = 1, \qquad i = 1, 2, \tag{139}$$

where all $U_l^i \in S$. By using the composition of the two we get

$$\begin{aligned} \Phi_1 \Phi_2 &= \sum_{l_1, l_2} c_{l_1}^1 c_{l_2}^2 \Phi(U_{l_1}^1) \Phi(U_{l_2}^2) \\ &= \sum_{l_1, l_2} c_{l_1}^1 c_{l_2}^2 \Phi(U_{l_1}^1 U_{l_2}^2), \end{aligned} \tag{140}$$

where we have used $\Phi(U_1)\Phi(U_2) = \Phi(U_1 U_2)$ by definition of $\Phi(U)$ for a unitary U. The proof follows from the assertion that $U_{l_1}^1 U_{l_2}^2 \in S$ and because the coefficients sum to one since $\sum_{l_1, l_2} c_{l_1}^1 c_{l_2}^2 = \sum_{l_1} c_{l_1}^1 \sum_{l_2} c_{l_2}^2 = 1$. □

Lemma 15 (Closedness under tensor product)**.** *If $\Phi_1 \in \mathcal{L}(S_1)$ and $\Phi_2 \in \mathcal{L}(S_2)$ then $\Phi_1 \otimes \Phi_2 \in \mathcal{L}(S_1 \otimes S_2)$.*

Proof of Lemma 15. Similar to proof of Lemma 14.

$$\begin{aligned} \Phi_1 \otimes \Phi_2 &= \sum_{l_1, l_2} c_{l_1}^1 c_{l_2}^2 \Phi(U_{l_1}^1) \otimes \Phi(U_{l_2}^2) \\ &= \sum_{l_1, l_2} c_{l_1}^1 c_{l_2}^2 \Phi(U_{l_1}^1 \otimes U_{l_2}^2). \end{aligned} \tag{141}$$

The last implication holds because, by elementary linear algebra, the following relation holds

$$\Phi(A) \otimes \Phi(B) = \Phi(A \otimes B)\,. \tag{142}$$

This is because for any operators V and W

$$\begin{aligned} \Phi(A) \otimes \Phi(B) \left|V\right\rangle\!\rangle \otimes \left|W\right\rangle\!\rangle &= (AVA^\dagger) \otimes (BWB^\dagger) \\ &= (A \otimes B)(V \otimes W)(A \otimes B)^\dagger \\ &= \Phi(A \otimes B) \left|V\right\rangle\!\rangle \otimes \left|W\right\rangle\!\rangle\,. \end{aligned}$$

□

I Incorporating prior information

I.1 Block independent noise

I.1.1 Proof of uniqueness

We first show that under the assumption of block independence we can break the gauge degeneracy.

Theorem 4 (Uniqueness for block-independent noise)**.**

Suppose that the POVM is described by $M_{kl} = M^1_k \otimes M^2_l$ where $k \in [D_1]$ and $l \in [D_2]$ with $D_1 D_2 = D$. Also let the qubit numbers for these two systems be n_1 and n_2, with $n_1 + n_2 = n$. For example, when the POVM is the computational basis, and the outcome of observing each qubit is binary valued as in (76), this refers to a partitioning of n qubits into two parts. Suppose that the noise acts independently on the two parts such that $A = A^1 \otimes A^2$ where A^1 acts on M^1 and A^2 acts on M^2. Let $A' = A'^1 \otimes A'^2$ be another such noise matrix and assume that A^1, A^2, A'^1, A'^2 are not erasure type matrices discussed with Condition 1. Then if A and A' are related by the gauge equivalence (12) we must have $A = A'$.

Proof. Since $A = A^1 \otimes A^2$ and $A' = A'^1 \otimes A'^2$, we will explicitly use double indices to represent each dimension. From (12) we have

$$A' = \alpha A + (1-\alpha)\mathrm{diag}(d)\mathbb{1} \,, \tag{143}$$

where

$$d_{kl} = \frac{\sum_{k',l'} A_{kl,k'l'} m_{k'l',I}}{2^{n/2}} = \frac{\sum_{k',l'} A'_{kl,k'l'} m_{k'l',I}}{2^{n/2}} \,. \tag{144}$$

Since $M_{kl} = M^1_K \otimes M^2_l$ we have $m_{kl,I} = m^1_{K,I} m^2_{l,I}$. Then

$$d_{kl} = \frac{\sum_{k',l'} A_{kl,k'l'} m_{k'l',I}}{2^{n/2}} \tag{145}$$

$$= \frac{\sum_{k'} A^1_{kk'} m^1_{k,I}}{2^{n_1/2}} \frac{\sum_{l'} A^2_{ll'} m^2_{l,I}}{2^{n_2/2}} = d^1_k d^2_l \,, \tag{146}$$

where for $i = 1, 2$ we define $d^i_k = \frac{\sum_{k'} A^i_{kk'} m^1_{k,I}}{2^{n_i/2}}$. Using the independence of the noise in (143), we get

$$A'^1 \otimes A'^2 = \alpha A^1 \otimes A^2 + (1-\alpha)\mathrm{diag}(d^1)\mathbb{1} \otimes \mathrm{diag}(d^2)\mathbb{1} \,. \tag{147}$$

Multiplying (147) on the left by $I \otimes \mathbf{1}^T_{D_1}$ and on the right by $I \otimes \mathbf{1}_{D_2}$, we get

$$A'^1 = \alpha A^1 + (1-\alpha)\mathrm{diag}(d^1)\mathbb{1} \,. \tag{148}$$

A similar computation yields

$$A'^2 = \alpha A^2 + (1-\alpha)\mathrm{diag}(d^2)\mathbb{1} \,. \tag{149}$$

Using $A = A^1 \otimes A^2$ and substituting (148) and (149) into (143) yields

$$\alpha A^1 \otimes A^2 + (1-\alpha)(\mathrm{diag}(d^1)\mathbb{1}_{D_1}) \otimes (\mathrm{diag}(d^2)\mathbb{1}_{D_2}) =$$
$$(\alpha A^1 + (1-\alpha)\mathrm{diag}(d^1)\mathbb{1}_{D_1})(\alpha A^2 + (1-\alpha)\mathrm{diag}(d^2)\mathbb{1}_{D_2}) \,.$$

Rearranging the above gives

$$\alpha(1-\alpha)(A^1 - \mathrm{diag}(d^1)\mathbb{1}_{D_1})(A^2 - \mathrm{diag}(d^2)\mathbb{1}_{D_2}) = 0 \,.$$

Since none of A, A^1, A^2 can be the erasure channel by assertion, we must have $\alpha = 1$ and hence $A = A'$. □

I.1.2 Algorithm to fix the gauge

By Theorem 4, no two noise matrices can be both block independent and be related by the gauge relation (12). Recall that Algorithm 1 returns a candidate noise matrix A' that is related to the true noise matrix A by the relation

$$A = \alpha A' + (1-\alpha)\mathrm{diag}(d')\mathbb{1} \,. \tag{150}$$

Our goal is to find α such that the matrix A decomposes as $A = A^1 \otimes A^2$. Define the operations

$$\mathrm{T}_1[\cdot] = (\mathbf{1}^T_{D_1} \otimes I_{D_1})(\cdot)(\mathbf{1}_{D_1} \otimes I_{D_2}) \,, \tag{151}$$

$$\mathrm{T}_2[\cdot] = (I_{D_1} \otimes \mathbf{1}^T_{D_2})(\cdot)(I_{D_1} \otimes \mathbf{1}_{D_2}) \,. \tag{152}$$

Since $A = A^1 \otimes A^2$ we have

$$\mathrm{T}_2[A] = A^1 = \alpha \mathrm{T}_2[A'] + (1-\alpha)\mathrm{T}_2[\mathrm{diag}(d')\mathbb{1}] \,, \tag{153}$$

$$\mathrm{T}_{S_1}[A] = A^2 = \alpha \mathrm{T}_1[A'] + (1-\alpha)\mathrm{T}_1[\mathrm{diag}(d')\mathbb{1}] \,. \tag{154}$$

Using (153) in (150) we get

$$\begin{aligned} &(\alpha \mathrm{T}_2[A'] + (1-\alpha)\mathrm{T}_2[\mathrm{diag}(d')\mathbb{1}]) \\ &\quad \otimes (\alpha \mathrm{T}_1[A'] + (1-\alpha)\mathrm{T}_1[\mathrm{diag}(d')\mathbb{1}]) \\ &= \alpha \tilde{A} + (1-\alpha)\mathrm{diag}(d)\mathbb{1} \,. \end{aligned} \tag{155, 156}$$

Since A is not the erasure channel, we can assume $\alpha \neq 0$ and get

$$\begin{aligned} &\alpha(\mathrm{T}_2[A'] \otimes \mathrm{T}_1[A'] - \mathrm{T}_2[A'] \otimes \mathrm{T}_1[\mathrm{diag}(d')\mathbb{1}] \\ &\quad - \mathrm{T}_2[\mathrm{diag}(d')\mathbb{1}] \otimes \mathrm{T}_1[A'] + \mathrm{T}_2[A'] \otimes \mathrm{T}_1[\mathrm{diag}(d')\mathbb{1}] \\ &= \mathrm{T}_2[A'] \otimes \mathrm{T}_1[\mathrm{diag}(d')\mathbb{1}] + \mathrm{T}_2[\mathrm{diag}(d')\mathbb{1}] \otimes \mathrm{T}_2[A'] \\ &\quad - \mathrm{diag}(d)\mathbb{1} - A' \,, \end{aligned} \tag{157}$$

The equation above is a matrix equality of the type $\alpha M^1 = M^2$, so we just need to find a matrix element $M^1_{ij} \neq 0$ such that $\alpha = M^2_{ij}/M^1_{ij}$.

I.2 Linearly representable prior information

Let $b^i_S,\ i = N_S+1, \ldots 4^n - 1$ and $b^i_A,\ i = N_A+1, \ldots, D$ be any set of vectors that span the space orthogonal to $b^i_S,\ i = 1 \ldots, N_S$ and $b^i_A,\ i = 1, \ldots, N_A$ respectively. Then simultaneous tomography can be performed by constructing canonical linear operators E^{ij} which we describe below. Let $\mathbf{m}$ be the matrix of coefficients of the POVM given by $[\mathbf{m}]_{k,Q} = m_{kQ}$. By independence of the POVM, the matrix $\mathbf{m}$ has full rank. Therefore, we can construct vectors $\tilde{b}^j$ such that

$$\sum_{Q \in \mathcal{B}_R \cup I} \tilde{b}^j_Q m_{kQ} = b^j_{A,k} \,, \quad k \in [D] \,,\ j = N_A+1, \ldots, D \,. \tag{158}$$

Let E^{ij} be the linear operator such that its matrix representation using the bases $\mathcal{B}_L, \mathcal{B}_R$ is given by

$$E^{ij}_{PQ} = b^i_{S,P} \tilde{b}^j_Q \,, \quad \forall P \in \mathcal{B}_R \,, Q \in \mathcal{B}_L \,. \tag{159}$$

Then, similar to (9) and (23), we can compute the following quantities using linear combinations of observations

$$\begin{aligned} z^{ij}_k &= z^I_k + \sum_{\substack{k' \in [D], \\ P \in \mathcal{B}_R, Q \in \mathcal{B}_L}} s_P E^{ij}_{PQ} A_{kk'} m_{k'Q} \\ &= z^I_k + (\sum_{P \in \mathcal{B}_R} b^i_{S,P} s_P)(\sum_{k' \in [D]} A_{kk'} \sum_{Q \in \mathcal{B}_L} \tilde{b}^j_Q m_{k'Q}) \\ &= z^I_k + (\sum_{P \in \mathcal{B}_R} b^i_{S,P} s_P)(\sum_{k' \in [D]} A_{kk'} (b^j_{A,k'} - \tilde{b}^j_I m_{k'I})) \\ &= z^I_k + (\sum_{P \in \mathcal{B}_R} b^i_{S,P} s_P)(\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k) \,. \end{aligned} \tag{160}$$

We also construct the linear operators required to fix the gauge denoted by E^j and defined as

$$E^j_{PQ} = b^1_{S,P} \tilde{b}^j_Q \,, \quad \forall P \in \mathcal{B}_R \,, Q \in \mathcal{B}_L \,. \tag{161}$$

and the corresponding computable quantity

$$\begin{aligned} z^j_k &= z^I_k + (\sum_{P \in \mathcal{B}_R} b^1_{S,P} s_P)(\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k) \\ &= z^I_k + d^1_S (\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k) \,. \end{aligned} \tag{162}$$

We will need to exploit the fact that A is not the erasure channel to perform simultaneous tomography. This is given in the lemma below.

Lemma 16. *Assume that A is not the erasure channel. Then there exists $k \in [D]$ for which there is a $j \in [D]$ such that*

$$\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k \neq 0 \,. \tag{163}$$

Proof. Summing (158) for $k \in [D]$ we get that

$$\tilde{b}^j_I = \sum_{k \in [D]} b^j_{A,k} \,. \tag{164}$$

Suppose that for a given $k \in [D]$ we have

$$\sum_{k'} A_{kk'} b^j_{A,k'} = \tilde{b}^j_I z^I_k = (\sum_{k' \in [D]} b^j_{A,k'}) z^I_k \,, \quad \forall j \in [D].$$

Since by construction the vectors b^j_A form a complete basis of the D dimensional space, we can invert the above relation to get $A_{kk'} = z^I_k, \ \ \forall k' \in [D]$, implying that A is the erasure channel. Then proof follows from using the fact that A is not the erasure channel.

□

Algorithm 3: Simultaneous tomography with linear prior information

Compute $z^I_k \ \ \forall \ k \in [D]$ using (9)
// Step 1.
for $j \in [M_A]$ **do**
  for $k \in [D]$ **do**
    Compute $\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k = \frac{z^j_k - z^I_k}{d^1_S}$.
// Step 2. Find noise matrix A
for $k \in [D]$ **do**
  Solve the system of equations to obtain $A_{kk'} \ \ \forall k' \in [D]$:
  $\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k = \frac{z^j_k - z^I_k}{d^1_S}, \quad j \in M_A,$
  $\sum_{k' \in [D]} A_{kk'} c^j_{A,k'} = d^j_A, \quad j \in [N_A].$
// Step 3. Find state ρ
Choose $k \in [D]$ and $j \in [D]$ as per Lemma 16 such that $\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k \neq 0$.
for $i \in [M_S]$ **do**
  Compute $\sum_{P \in \mathcal{B}_R} s_P b^i_{S,P} = \frac{z^{ij}_k - z^I_k}{\sum_{k' \in [D]} A_{kk'} b^j_{A,k'} - \tilde{b}^j_I z^I_k}$
Return $\{s_P \mid P \in \mathcal{B}_R, \ \ A\}$.

I.3 Independent ancilla qubits

By running Step 1 of Algorithm 1 on the ancilla qubits, we can identify $P \in \mathcal{B}^a_R$ and $i \in [D^a]$ such that

$$z^{Pi}_k - z^I_k = s_P \sum_{k'} A^a_{kk'} C^a_{k'i} \neq 0 \,. \tag{165}$$

Construct the set of operators $\{E_Q \mid Q \in \mathcal{B}^r_R\}$ given by

$$\begin{aligned} &E_Q \left| P \otimes Q \right\rangle\!\rangle - \bar{M}^a_i \otimes I \in (\mathcal{B}^a_L \otimes I)^\perp \,, \\ &E_Q \left| P' \otimes Q' \right\rangle\!\rangle \in (\mathcal{B}^a_L \otimes I)^\perp \quad \text{if } P' \neq P \text{ or } Q' \neq Q \,. \end{aligned} \tag{166}$$

The above operators can be constructed using linear combinations in (10) since $P \otimes Q$ and $\bar{M}^{anc}_i \otimes I$ are both traceless. Measuring the ancilla qubits independently is equivalent to tracing out the other measurements and

effectively using the measurement operators $\{M_i^a \otimes I \mid i \in [D^{anc}]\}$. Using the operators in (166), we can compute the quantities for all $Q \in \mathcal{B}_R^r$

$$z_k^{PQi} - z_k^I = s_P s_Q \sum_{k'} A_{kk'}^a C_{k'i}^a \,. \tag{167}$$

We can then recover the state coefficients $\{s_Q \mid Q \in \mathcal{B}_R^r\}$ as

$$s_Q = \frac{z_k^{PQi} - z_k^I}{z_k^{Pi} - z_k^I} \,. \tag{168}$$

We now compare to the setting in [42] where they consider the hierarchical setting described in Sec. 4.2, with one ancillary qubit and orthogonal POVMs. In this case, the basis for the ancilla can be chosen to be $\{I, M_1^a - M_2^a\}$ where $M_1^a - M_2^a$ is traceless by definition of POVMs. Since the POVMs are orthogonal, any unitary operator U_i such that

$$U_i((M_1^a - M_2^a) \otimes M_i^r)U_i^\dagger = (M_1^a - M_2^a) \otimes I \,, \tag{169}$$

will satisfy the conditions of the operator described in (166). The construction of this operator when the POVM is the computational Z-basis can be found in [42].

I.4 Denoising the binary symmetric channel

For each $Q \in \mathcal{P}^n$ the vector of measurement operator coefficients $\mathbf{m}_Q$ is an eigenvector of A. Since $\mathbf{m}_I = \mathbb{1}$ we have $Am_I = \mathbf{m}_I$. For every other $Q \in \mathcal{P}_Z^n$ the coefficients are given by $\mathbf{m}_Q = \otimes_{i=1}^n [1, \pm 1]$. Let $S_Q \subset [n]$ be the set of indices for which the component of $\mathbf{m}_Q$ is $[1, -1]$. Then we have $A\mathbf{m}_Q = \lambda_{S_Q}\mathbf{m}_Q$, where $\lambda_S = \prod_{i \in S_Q}(1-2p_i)$. Thus for any operator $\mathbf{\Phi}$ the corresponding measurements are given by

$$\mathbf{y} = \sum_{P,Q} s_P \mathbf{\Phi}_{PQ} A \mathbf{m}_Q = \sum_{P,Q} s_P \mathbf{\Phi}_{PQ} \lambda_{S_Q} \mathbf{m}_Q \,. \tag{170}$$

This allows for the less expensive denoising in Algorithm 4. The generator gate set is given by

$$\mathcal{G}_{BSC} = \{I, SWAP(i,j), CNOT(i,j) \mid i,j \in [n]\}. \tag{171}$$

Algorithm 4: Denoising for binary symmetric output noise.

```
// Step 1: Identifying non-zero state coefficients
```
Initialize $\mathcal{S}_{nz} = \emptyset$;
for $P \in \mathcal{P}_Z^n \setminus I^n$ **do**
  Compute $s_P \lambda_P = [1 \pm 1]^T \mathbf{y}$;
  If $[1 \pm 1]^T \mathbf{y} \neq 0$, update $\mathcal{S}_{nz} \leftarrow \mathcal{S}_{nz} \cup \{P\}$;
end
```
// Step 2: Decode non-zero state coefficients
```
for $P = \bigotimes_{i=1}^n P_i \in \mathcal{S}_{nz}$ **do**
  Identify $S = i \in [n] \mid P_i = Z.$;
  if $|S| = 1$ **then**
    Let $S = \{i\}$ and pick any $j \neq i$;
    Obtain measurements $m_{P,G}$ for each gate $G \in \{I, CNOT(i,j), SWAP(i,j)\}$;
    Decode $s_P \leftarrow \frac{m_{P,I} m_{P,SWAP(i,j)}}{m_{P,CNOT(i,j)}}$;
  else
    Let $S = \{i_1 < i_2, \ldots < i_{|S|}\}$;
    Obtain measurements $m_{P,G}$ for each gate $G \in \{I, CNOT(i_1, i_2), \ldots, CNOT(i_{|S|}, 1)\}$;
    Compute $(1 - 2p_j) = \frac{m_{P,I}}{m_{P,CNOT(i,j)}}$;
    Decode $s_P \leftarrow \frac{m_{P,I}}{\prod_{i \in S}(1-2p_i)}$;
  end
  return $\{s_P \mid P \in \mathcal{P}_Z^n\}$;
end