Refining quantum computing models by removing gauge freedom

Juan J. Gonzalez De Mendoza*
Center for Quantum Information and Control.
Department of Physics and Astronomy.
University of New Mexico.
(Dated: October 24, 2024)

Techniques for quantum device characterization are used to probe the behavior of quantum computational operations, and construct predictive models of them. But these models grow exponentially in complexity with the number of qubits in the system being probed. Reconstructing a fully general model may require an overwhelming, infeasible amount of data. One way to avoid this catastrophe is to identify "reduced" statistical models that have relatively few parameters, yet still accurately describe the observed data. Reduced-model techniques will be essential for characterizing many-qubit systems. Finding reduced models by numerically exploring the space of models is made much harder by a pernicious gauge freedom in the "gate set" model of qubit systems. This gauge freedom occurs because infinitely many distinct yet equivalent models predict the same physical outcomes. I have been developing a new parameterization for gate set models, referred to as first-order gauge-invariant (FOGI), which eliminates most of the barriers to finding reduced models. I will describe the FOGI framework, and then I will introduce an algorithm that I am developing for automated model selection (AMS), which constructs reduced models by finding and removing parameters that are not necessary to capture the noise affecting a quantum device. Finally, I will demonstrate AMS with FOGI models in practice by applying it on simulated data and experimental data from neutral atom qubits operated at Sandia. In experiment, AMS finds a model of the atomic qubit's dynamics that has only half as many parameters as the standard fully general model, but loses virtually no accuracy.

I. INTRODUCTION

Gate-based quantum computing has been shown to hold the potential for a variety of algorithmic speedups over classical computing since the 1980's [1], yet a useful advantage over classical computers has not been observed within four decades since. Even from the early days, the realizability of quantum computers has been met with a justifiable amount of skepticism [2]. Unlike their classical counterpart, qubits are subject to a continuum of control errors, alongside exponentially fast decoherences [3]. Even with tremendous theoretical advances to combat these issues, such as the birth of the field of quantum error-correction, we still have some daunting obstacles on our way to a quantum computer with arbitrarily low error rates. One of which simply being the improvement of physical error rates of quantum logic operations. The collection of operations available to a gate-based quantum computer is known as its gate set, which comprises a collection of state preparations, measurements, and processes. In order to improve the physical error rates of a gate set we employ techniques that allow us to gain insight about how a quantum device is actually performing compared to how it *should* be performing. The field dedicated to developing such techniques is called Quantum Characterization, Verification and Validation (QCVV).

The work presented in this paper builds upon quantum tomography, a family of QCVV protocols for reconstructing models of unknown quantum states and noise processes using empirical data. Tomographic protocols face a common problem: the number of free parameters in the models that they seek to reconstruct grows exponentially with the number of qubits being probed. As a result, existing tomographic protocols cannot feasibly be used to probe (and improve) more than a few qubits at once. However, there is a potential route around this obstacle. A tomographic protocol could in principle be combined with an algorithm that identifies and removes model parameters that are not necessary to describe the experimental data, yielding an efficiently describable 'reduced model' that still fits the data, , a.k.a., automated model selection (AMS). A major barrier for AMS is the existence of gauge freedom within gate sets, where infinitely many different models are physically indistinguishable from each other [4]. AMS is tasked with the identification of parameters that are not necessary to express the noise affecting a quantum device. On the other hand, gauge freedom allows for this noise to be expressed by different parameters, depending on the choice of gauge. Thus, making it impossible to properly quantify how much descriptive value a single parameter can provide for a given data. This problem begs for the construction of a gate set parameterization choice which lacks gauge degrees of freedom.

In this paper I present necessary background for: the description of noisy quantum systems, Hilbert-Schmidt superoperator representation (section II); basic quan-

* Also at Quantum Performance Lab, Sandia National Laboratories

tum tomography (section III); and a basis for quantum channels which is conducive for physical interpretation, the elementary error generators (section IV). Finally, I present our contributions in section V consisting of (1) a novel gate set parameterization which alleviates problems with gauge freedom under the assumption of small errors, and small gauge transformations, called first-order gauge-invariant (FOGI) quantities; and (2) the successful validation of AMS on FOGI parameterized models with simulated data, followed by its deployment on Sandia's neutral atom quantum computer.

II. NOISY OPERATIONS AND HILBERT-SCHMIDT SPACE

The standard curriculum in many undergraduate and graduate quantum mechanics courses covers the description and evolution of *closed* quantum systems. Quantum states that do not interact with anything outside the system in question evolve in a unitary fashion as described by the Schrodinger equation, and are called *pure* states. In the field of QCVV, we are concerned with the modeling of real quantum devices which are not perfectly isolated. As a consequence, we require a richer description of quantum states which can capture interactions with unknown external systems, a.k.a., their *environment*.

The most general description of noisy quantum states is given by density matrices, denoted by ρ. Density matrices comprise a probabilistic mixture of pure states:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{1}$$

These are $d \times d$ positive semidefinite, trace-1 operators. Density matrices can evolve in a richer fashion than pure states, where unitary dynamics is a special case. More generally, $\rho \to G[\rho]$, where G is a linear map known as a quantum *channel* or *process*.

Within the field of quantum information we focus on the treatment of finite-dimensional Hilbert spaces constructed from 2-dimensional subsystems called *qubits* (where $d = 2^n$, and n is the number of qubits). Due to their physical constraints, a density matrix for one qubit only has three degrees of freedom. Because of this, we can represent a single qubit as a 3-dimensional vector within the *Bloch sphere*, as shown in Fig. 1.

It is convenient to embed density matrices into a higher-dimensional (d^2) vector space called *Hilbert-Schmidt space*. This space is equipped with an inner product $\langle A, B\rangle = Tr(A^\dagger B)$. Similar to Dirac notation for Hilbert space, elements of Hilbert-Schmidt space, A, can be represented by column vectors called "superkets" as $|A\rangle\rangle$, and elements of its dual space can be represented by row vectors called "superbras" as $\langle\langle A|$.

The field of quantum computing focuses on measurement operations which yield one outcome out of a finite and discrete set of possibilities. Outcome probabilities are linear functions of states, which allows us to represent these as elements of the dual space such that $P(i, \rho) = \langle\langle E_i|\rho\rangle\rangle = Tr(E_i^\dagger \rho)$ describes the probability of observing outcome i after preparing state ρ. In order to constraint these probabilities to be positive, and add up to one, we require that $\sum_i E_i = \mathbb{1}$ and $E_i \geq 0$. A set of these superbras $\{\langle\langle E_i|\}$ which satisfy the probability constraints describe a single measurement operation, and is called a positive-operator valued measurement (POVM). Every element of a POVM, $\langle\langle E_i|$, corresponds to a measurement outcome, and is called an "effect". Lastly, quantum processes govern the evolution of quantum states. Due to their linearity, they can be represented as matrices that act on superkets $G(\rho) \to G|\rho\rangle\rangle$, these are called "superoperators" or "process matrices".

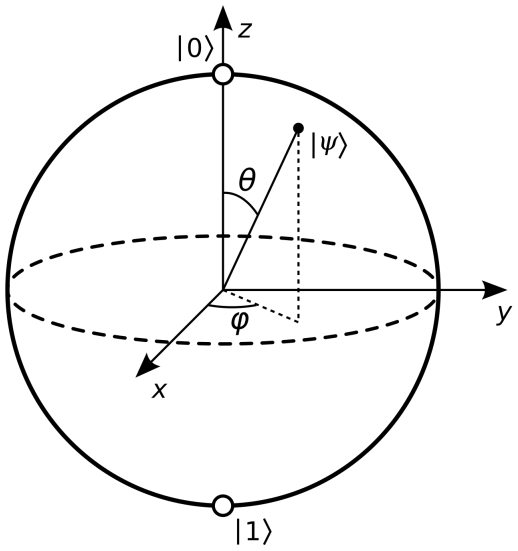

FIG. 1. The Bloch sphere: a geometrical representation of quantum states. The entries of the Bloch vector $\vec{n}$ are given by the expectation values of the quantum state with the non-identity 1-qubit Pauli operators: $\vec{n} = (Tr(X\rho), Tr(Y\rho), Tr(Z\rho))$. Vectors on the boundary of the sphere represent pure states.

My work lies within the context of gate-based quantum computing devices, whose only task is to run *quantum circuits*. A quantum circuit consists of (1) an initial state picked from a finite set of N_ρ available state preparations $\{|\rho_i\rangle\rangle\}_{i=1}^{N_\rho}$, (2) a combination of different quantum processes picked from a finite set of N_G options $\{G_i\}_{i=1}^{N_G}$, and lastly (3) the application of one out of N_M measurements which result in one outcome out of $N_E^{(M)}$ possibilities $\{\langle\langle E_i^{(m)}|\}_{m=1,i=1}^{N_M, N_E^{(m)}}\}$. The collection of all of these operations available to a quantum device is called a *gate set*, denoted by $\mathcal{G}$, where

$$\mathcal{G} = \{\{|\rho_i\rangle\rangle\}_{i=1}^{N_\rho}; \{G_i\}_{i=1}^{N_G}; \{\langle\langle E_i^{(m)}|\}_{m=1,i=1}^{N_M, N_E^{(m)}}\}. \tag{2}$$

The entries of each of these operations in a given basis $\{B_i\}$ can be computed through the inner product as

follows:

$$[|\rho\rangle\rangle]_i = \langle\langle B_i|\rho\rangle\rangle = Tr(B_i^\dagger \rho), \tag{3a}$$

$$[\langle\langle E|]_i = \langle\langle E|B_i\rangle\rangle = Tr(E^\dagger B_i), \tag{3b}$$

$$G_{i,j} = \langle\langle B_i|G|B_i\rangle\rangle = Tr(B_i^\dagger G(B_i)). \tag{3c}$$

Physical constraints for POVM effects have already been discussed, but states and processes must satisfy different conditions. States must (1) be Hermitian, (2) have trace equal to 1, and (3) be positive semi-definite. For processes to be physical, they must output a physical state for all possible physically valid input states. This results in matrices which are (1) completely positive (CP) [5], and (2) trace preserving (TP), typically referred to as CPTP processes. Enforcing condition (2) for both of these operations is quite simple in the superoperator/superket representation if the basis $\{B_i\}$ contains one element proportional to the identity, and the rest of the elements are traceless. This results in all states having one fixed entry $\langle\langle \mathbb{1}|\rho\rangle\rangle = Tr(\mathbb{1}\rho) = Tr(\rho) = 1$. On the other hand, a quantum process G is TP iff $\langle\langle \mathbb{1}|G = \langle\langle \mathbb{1}|$. Thus, similar to states, all processes have a fixed row given by $\langle\langle \mathbb{1}|G|B_i\rangle\rangle = \langle\langle \mathbb{1}|B_i\rangle\rangle = Tr(\mathbb{1}B_i) = Tr(B_i)$, which equals a constant for $B_i = \mathbb{1}$, and 0 otherwise. As a consequence, physical processes are a smaller subspace of $d^2 \times d^2$ matrices. Enforcing the TP constraint removes d^2 degrees of freedom by fixing a row to constant values, while CP constraints correspond to inequalities which only restrict some of the degrees of freedom. This results in $d^2(d^2-1)$ dimensions in CPTP superoperator space.

III. QUANTUM TOMOGRAPHY: A CHARACTERIZATION TECHNIQUE

In this section we introduce prerequisites for tomography of quantum systems following the treatment presented originally in [5].

The goal of quantum tomography is to obtain a complete mathematical description of a gate set, typically with the hope of being able to infer information about a physical device from its mathematical reconstruction. Among the most popular tomography methods are:

- Quantum State/Measurement Tomography: Results in a full description of state-preparations/measurements, while assuming knowledge of measurements/state-preparations that can be performed.
- Quantum Process Tomography: Results in a full description of quantum processes, while assuming knowledge of both state preparations *and* measurements.
- Gate Set Tomography: Results in a full description of quantum processes, measurements and states, without assuming any prior knowledge.

To remove some of the abstraction in these descriptions, let us delve into the basics of each of these.

A. State and Measurement Tomography

The goal of state tomography is to construct a predictive mathematical model for a given state preparation operation. In this example, we use their superket representation $|\rho\rangle\rangle$.

There are two assumptions required to perform tomography, (1) the ability to create many copies of the state to be analyzed, ρ; and the (2) availability of an *informationally complete* (IC) set of perfect measurements to be applied to the state being analyzed. A set of measurements is IC iff, for every ρ, the outcome probabilities given by the Born rule

$$p_{i,m} = Tr(\rho E_i^{(m)}), \tag{4}$$

$$\text{or } p_{i,m} = \langle\langle E_i^{(m)}|\rho\rangle\rangle, \tag{5}$$

, only have one possible ρ that satisfies them, where these probabilities correspond to observing the POVM effect $E_i^{(m)}$ after preparing state ρ. This condition is equivalent to the set $\{\langle\langle E_i^{(m)}|\}$ spanning the dual space. For convenience, we create an arbitrary one-to-one mapping where we list all combinations of indices i and m, $\{i, m\} \mapsto k$. If we stack all the POVM elements into a matrix

$$A = \begin{pmatrix} \langle\langle E_0| \\ \langle\langle E_1| \\ \vdots \\ \langle\langle E_{N_E}| \end{pmatrix}, \tag{6}$$

we can write all probabilities from Eq. (5) in one convenient equation:

$$\vec{p}_\rho = A|\rho\rangle\rangle. \tag{7}$$

The IC constraint ensures A has full column rank. If A is square, we can invert it and solve for $|\rho\rangle\rangle$:

$$A^{-1}\vec{p}_\rho = |\rho\rangle\rangle. \tag{8}$$

We can generalize this for non-squared matrices by using the pseudo-inverse instead:

$$A^T\vec{p}_\rho = A^T A|\rho\rangle\rangle, \tag{9}$$

$$(A^T A)^{-1}A^T\vec{p}_\rho = |\rho\rangle\rangle. \tag{10}$$

Now, notice that on the left-hand side we have $\vec{p}$, which comprises outcome probabilities corresponding to all different measurements being applied to $|\rho\rangle\rangle$, and the matrix A. Matrix A is known as part of our assumptions.

Thus, if we approximate $\vec{p}_\rho$ by setting it equal to observed outcome frequencies, we can estimate $|\rho\rangle\rangle$.

Similarly, measurement tomography works the same way, but all superkets and superbras are interchanged. We get analogous equations for our measurements:

$$\langle\langle E| = \vec{p}_E B^{-1}, \tag{11}$$

$$\text{or } \langle\langle E| = \vec{p}_E (B^T B)^{-1} B^T. \tag{12}$$

Where B is composed of all known state preparations,

$$B = \left(|\rho_1\rangle\rangle \;\; |\rho_2\rangle\rangle \;\; \ldots \;\; |\rho_{N_\rho}\rangle\rangle \right). \tag{13}$$

Which also allows for the reconstruction of $\langle\langle E|$ by estimating $\vec{p}_E$ through observed outcome frequencies.

B. Process Tomography

In quantum process tomography, it is necessary to assume both perfect IC states *and* measurement in order to reconstruct a process matrix G. Similarly to the previous two tomographic methods, creating state ρ_k, applying gate G, and observing outcome E_j has outcome probabilities governed by the Born rule as

$$P_{j,i} = \langle\langle E_j | G | \rho_i \rangle\rangle. \tag{14}$$

Using the matrices A and B defined in Eqs. 6 and 13, we can define a new $N_{E_N} \times N_\rho$ matrix, P, whose entries are $P_{i,j}$ as

$$P = AGB. \tag{15}$$

We can now manipulate our state preparation and measurement (SPAM) matrices to solve for our unknown operation,

$$G = A^{-1} P B^{-1}. \tag{16}$$

We can generalize this for non-invertible A and B matrices by using the pseudo-inverse and obtain

$$G = (A^T A)^{-1} A^T P B^T (B B^T)^{-1}. \tag{17}$$

Just like the previous two methods, we can approximate our probabilities from empirical frequencies in order to estimate our desired quantum operation, G.

C. Gate Set Tomography

All of the tomographic methods above share a common weakness: assuming knowledge of some SPAM operations. Measurement tomography assumes a perfect set of states, therefore state tomography should be performed beforehand to ensure this assumption is valid; but state tomography makes the same assumption about measurements, so maybe instead measurement tomography should be performed first. This results in a circular dependency. In practice, all operations are faulty, and making any of these assumptions can yield unreliable results. In this work, I focus on a tomography protocol that can simultaneously characterize all the operations in a given gate set [1], called *gate set tomography* (GST).

1. Linear Gate Set Tomography (LGST)

Perhaps the simplest GST protocol, linear gate set tomography (LGST), follows from the groundwork laid down in the previous tomography sections. Define a new matrix, called the Gram matrix, as the product of the matrices from Eq. (6) and (13):

$$\tilde{\mathbb{1}} = AB. \tag{18}$$

Its entries are given by inner products of states and measurement effects,

$$\tilde{\mathbb{1}}_{i,j} = \langle\langle E_i | \rho_j \rangle\rangle, \tag{19}$$

corresponding to probabilities of simple prepare-and-measure circuits. Row "i" of this matrix corresponds to performing measurement tomography on the POVM element $\langle\langle E_i|$, while column "j" corresponds to performing state tomography on state $|\rho_j\rangle\rangle$. By approximating each of these probabilities by setting them equal to outcome frequencies, we can estimate every entry of $\tilde{\mathbb{1}}$ through experimental data.

Similar to our previous protocols, we will need to assume the existence of an IC set of states and measurements for the construction of the A and B matrices. It is important to note that this assumption is subtly different from those made in state, measurement and process tomography. We assume the *existence* of IC operations, but we do not assume that we *know* them. This allows for the modeling noisy, and unknown, SPAM operations.

For simplicity, we will make this assumption slightly stricter, and also assume that the states and effects are *minimally* IC, which means they form a basis for superkets and superbras respectively[2]. This implies that A, B, and $\tilde{\mathbb{1}}$ are square, full rank and same dimension. This allows us to invert them and solve for A^{-1}:

$$A^{-1} = B \tilde{\mathbb{1}}^{-1}. \tag{20}$$

---

[1] Protocols with this property, are called self-consistent.

[2] This assumption can be relaxed by including pseudo-inverses, but careful analysis is required, as is done in reference [5].

If we plug this into previous equations describing state, measurement, and process tomography (8, 11, 16) we get

$$|\rho\rangle\rangle = B\tilde{\mathbb{1}}^{-1}\vec{p}_\rho, \tag{21}$$
$$\langle\langle E| = \vec{p}_E B^{-1}, \tag{22}$$
$$\text{and } G = B\tilde{\mathbb{1}}^{-1}PB^{-1}. \tag{23}$$

We have obtained a set of equations, where the only unknowns are the operations we want to estimate, and the matrix B. It may seem that we have reached a dead end, where we do need to assume knowledge of B in order to estimate our desired operations. Yet, it turns out that B is merely a mathematical artifact with no observable features, a matrix of *gauge degrees of freedom*. To see why, consider any circuit outcome probability, and plug in the results from Eqs. 21-23:

$$\begin{aligned} p(E, G\dots G', \rho) &= \langle\langle E|G\dots G'|\rho\rangle\rangle, \\ &= (\vec{p}_E B^{-1})(B\tilde{\mathbb{1}}^{-1}PB^{-1})\dots(B\tilde{\mathbb{1}}^{-1}P'B^{-1})(B\tilde{\mathbb{1}}^{-1}\vec{p}_\rho), \\ &= \vec{p}_E\tilde{\mathbb{1}}^{-1}P\dots P'\tilde{\mathbb{1}}^{-1}\vec{p}_\rho. \end{aligned} \tag{24}$$

Observe that the resulting expression is independent of B. This is true for *all possible* circuit outcome probabilities. B is an unknowable and unmeasurable matrix. GST is able to estimate a gate set only *up to a gauge*. A more detailed discussion on gauge freedom is found in section V 1. For the purposes of GST, we have now achieved our goal. The tomographer has the freedom to choose any B and obtain a gate set model that accurately describes their empirical data.

2. *The GST Workflow*

The LGST protocol described above is not exactly what is used in the contributions of this paper, but it serves the pedagogical purpose of illustrating the basic ideas behind characterizing all gate set operations simultaneously. This paper regards the construction of reduced gate set models, which requires the estimation of said models through GST. For this, I utilize the standard GST implementation in the open-source Python library called pyGSTi. At a high level, This version of GST has two main differences from LGST:

1. Gate set models are obtained through a highly customized optimization algorithm centered around maximum likelihood estimation (MLE). MLE's goal is to obtain the absolute maximum of the likelihood function,

$$\mathcal{L} = P(data|\mathcal{G}(\vec{\theta})), \tag{25}$$

where $\vec{\theta}$ is a list of parameters which completely describe a gate set. By the likelihood principle [6], models obtained through MLE have equal or greater predictive power than their linear inversion counterparts in LGST.

2. Since this protocol does not rely on linear inversion as per point (1), it is able to include data from longer/deeper circuits, where more than one process is applied. This enables a clever circuit list construction, referred to as *germ selection*, which amplifies errors and achieves higher precision.

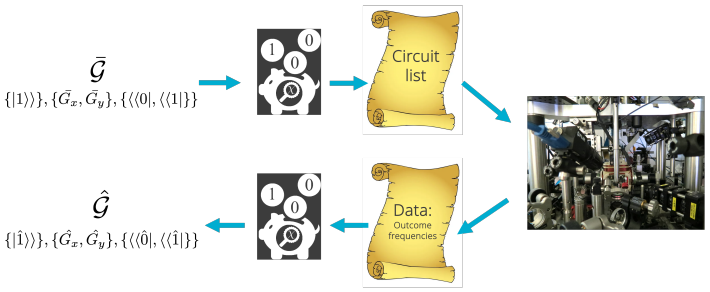

FIG. 2. A diagram of the steps necessary to perform standard GST using the Python open-source library pyGSTi with experimental data. First, an ideal gate set description is fed into pyGSTi to obtain a list of circuits. These circuits are then executed on the quantum computer and their output is recorded in a data file, depicted in this diagram is Sandia's neutral atom device used to collect data in the contributions section. Finally, pyGSTi reads the data file and returns a gate set model which best describes the collected frequencies.

The workflow needed to perform GST using pyGSTi is illustrated in Fig. 2. First, a target gate-set description of the device must be fed into pyGSTi, which is used to construct an *experiment design*, consisting of a circuit list tailored to the given gate set. These circuits are then run on the quantum hardware and their outcomes are recorded into a file which is fed back into pyGSTi. Finally, we obtain a gate set model computed through MLE.

This refined GST protocol is free from linear inversion, opening the door to different gate set descriptions. Given a list of parameters $\vec{\theta}$, it is now possible perform GST as long as a mapping is specified to obtain a gate set model from said parameters $\vec{\theta} \to \mathcal{G}$. The specification of this mapping is referred to as a *parameterization choice*. As a result, tomographers have the freedom to choose parameterizations different from the entries of process matrices, superkets, and superbras. Furthermore, it is an incredibly important choice as it influences both the objective function's landscape and the interpretability of GST results.

IV. GATE SET PARAMATERIZATIONS: PHYSICAL INFERENCE FROM TOMOGRAPHY

An important and necessary part of GST is for the tomographer to choose a parameterization to represent the gate set, but in practice, what does this entail?

As described earlier, a straight-forward parameterization for gate sets is to choose every element of the process matrices, alongside the elements of the superkets and superbras in SPAM, individually as parameters. These elements are dependent on the choice of basis for Hilbert-Schmidt space. To illustrate this, consider a gate set for one qubit with one gate, one state preparation, and one POVM with two elements represented in the Pauli basis $\{\mathbb{1}, X, Y, Z\}$. It is useful to arrange the parameters of a statistical model into a vector, usually denoted by $\vec{\theta}$, where

$$\vec{\theta} = (\rho_x, \rho_y, \rho_z, G_1, G_2, \ldots, G_{12}, E_0, E_1, E_2, E_3). \quad (26)$$

We can reconstruct our gate set operations simply by arranging them into matrices and vectors that satisfy TP constraints as described in section II:

$$\mathcal{G}(\vec{\theta}) = \{|\rho\rangle\rangle = \begin{pmatrix} 1 \\ \rho_x \\ \rho_y \\ \rho_z \end{pmatrix}, G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ G_1 & G_2 & G_3 & G_4 \\ G_5 & G_6 & G_7 & G_8 \\ G_9 & G_{10} & G_{11} & G_{12} \end{pmatrix},$$
$$\langle\langle E_0| = \begin{pmatrix} E_0 \\ E_1 \\ E_2 \\ E_3 \end{pmatrix}^T, \langle\langle E_1| = \begin{pmatrix} 1 - E_0 \\ -E_1 \\ -E_2 \\ -E_3 \end{pmatrix}^T \}.$$

This is a mathematically convenient parameterization, but so far I have ignored what is arguably the most important part of quantum characterization: how does a tomographer infer information from these parameters? I presented these QCVV methods as a means to improve physical devices; however, to achieve this, we must extract information about the underlying physical mechanisms behind the noise. Process matrices are hard to interpret. Constructing a gate set parameterization conducive to physical interpretation was a big incentive behind the development of *elementary error generators* [7].

A. Elementary Error Generators

In this section we summarize the findings of reference [7], which introduces a basis for processes that allows for easier interpretation and construction of reduced models. This framework works under the assumption of *small* and *markovian*[3] errors.

First, we must define a slightly different model for our operations, motivated by the tomographer's goal: understanding how the hardware's operations *differ* from their perfect counterparts. In order to achieve this, we adopt a post-gate error model, where each noisy gate, G, consists of an application of the perfect gate $\bar{G}$, followed by an error channel $\mathcal{E}$:

$$G = \mathcal{E}\bar{G}. \quad (27)$$

Notice that when the gate is error-free, $\mathcal{E} = \mathbb{1}$. We can isolate the deviations of the error channel from the identity by instead computing

$$\log(\mathcal{E}) = (\mathcal{E} - \mathbb{1}) + O((\mathcal{E} - \mathbb{1})^2). \quad (28)$$

Under a small error assumption, $\mathcal{E}$ is close to the identity, which translates to having all eigenvalues close to 1. As a consequence, $\mathcal{E}$ is invertible and only has positive eigenvalues. By Theorem 3.4 in [8], $\mathcal{E}$ has a real logarithm. We denote this quantum-process-generator matrix as L,

$$L = \log(\mathcal{E}), \quad (29)$$
$$G = e^L\bar{G}, \quad (30)$$

whose magnitude directly represents the magnitude of errors in G. Thus, if we manage to find a basis for L, we will have an easier time interpreting how a hardware's gates differ from their ideal counterparts. Similar to how Hamiltonians generate unitary evolutions, L generates errors.

With the motivation for an exponentiated matrix laid down, we can now move on to the construction of a useful basis for L. We will identify four subspaces so that we can write any error generator as a linear combination of them:

$$L = L_{\mathbb{H}} + L_{\mathbb{S}} + L_{\mathbb{C}} + L_{\mathbb{A}}. \quad (31)$$

We now derive each of these subspaces.

1. Hamiltonian Generators

Consider channels consisting of perfect unitary evolution, generated by some Hamiltonian J:

$$\mathcal{E}[\rho] = e^{-iJ}\rho e^{iJ}, \quad (32)$$

where we choose to expand J into the Pauli basis $J = \sum_p h_p P$. Recall the Campbell identity:

$$e^x Y e^{-x} =$$
$$Y + [X, Y] + \frac{1}{2!}[X, [X, Y]] + \frac{1}{3!}[X, [X[X, Y]]] + \ldots,$$

Plugging these two identities into Eq. (32), we obtain

[3] Markovianity is a ubiquitous assumption for quantum evolution in the literature, and is defined as a quantum process which evolves $\rho \to \rho'$, where ρ' can be completely determined by ρ without the need of any other variables.

$$\mathcal{E}[\rho] = \rho - i[J,\rho] + \frac{1}{2!}[J,[J,\rho]] + \dots \tag{33}$$

$$= \rho - i[\sum_p h_p P, \rho] \tag{34}$$

$$- i\frac{1}{2!}[\sum_Q h_Q Q, [\sum_R h_R R, \rho]] + \dots \ .$$

Where P, Q, and R are Pauli operators. It can be shown that the nested commutator equals 0,

$$[\sum_Q h_Q Q, [\sum_R h_R R, \rho]] = 0. \tag{35}$$

Thus, simplifying our unitary channel to

$$\mathcal{E}[\rho] = \rho - i[\sum_P h_P P, \rho]. \tag{36}$$

Furthermore, this can be rewritten as a generated channel as

$$\mathcal{E}[\rho] = e^{\sum_p h_P H_P}[\rho], \tag{37}$$

where

$$H_P[\rho] = -iP\rho + i\rho P = -i[P,\rho]. \tag{38}$$

We call these "Hamiltonian" generators, which correspond to a linearly independent set indexed by d^2-1 non-identity Pauli operators that span $L_{\mathbb{H}}$.

2. Stochastic Generators

Let us now consider a larger set of quantum channels, that includes probabilistic/convex mixtures of unitary channels:

$$\mathcal{E}[\rho] = \sum_k p_k e^{-iJ_k}\rho e^{iJ_k}. \tag{39}$$

Expanding to first order will yield the same Hamiltonian generators derived in the previous section, so let us approximate up to second order in J_k,

$$\mathcal{E}[\rho] \approx \sum_k p_k(\mathbb{1} - iJ_k - \frac{J_k^2}{2})\rho(\mathbb{1} + iJ_k - \frac{J_k^2}{2}) \tag{40}$$

$$= \sum_k p_k(\rho + i(\rho J_k - J_k\rho) + J_k\rho J_k - \frac{1}{2}(J_k^2\rho + \rho J_k^2)), \tag{41}$$

$$= \rho + \sum_k p_k(i[\rho, J_k]) + J_k\rho J_k - \frac{1}{2}\{J_k^2, \rho\}. \tag{42}$$

Once again, we choose to represent each Hamiltonian in the Pauli basis, and combine constants to create new coefficients to obtain:

$$\mathcal{E}[\rho] \approx \rho + H_J[\rho] + \sum_P s_p(P\rho P - \rho)$$

$$+ \sum_{P,Q>P} c_{P,Q}(P\rho Q + Q\rho P - \frac{1}{2}\{\{P,Q\},\rho\}), \tag{43}$$

where H_J are the Hamiltonian error generators $\sum_k p_k H_{J_k}$. Analyzing the resulting equation, we find two new terms scaled by the s_P and $c_{P,Q}$ coefficients. Each of these coefficients can be varied independently. Therefore, allowing us to represent any probabilistic mixture of small magnitude Hamiltonian evolutions through a combination of Hamiltonian generators, "Pauli stochastic" elementary generators

$$S_P[\rho] = P\rho P - \rho, \tag{44}$$

and "Pauli-correlation" elementary generators

$$C_{P,Q}[\rho] = P\rho Q + Q\rho P - \frac{1}{2}\{\{P,Q\},\rho\}. \tag{45}$$

The S and C error generators span the $L_{\mathbb{S}}$ and $L_{\mathbb{C}}$ subspaces, respectively. The S generators, just like the Hamiltonian generators, are indexed by (d^2-1) Pauli operators. The C generators were split off in Eq. (43) under the assumption that $Q \geq P$ (for any ordering of the set of Pauli operators), resulting in only $(d^2-1)(d^2-2)/2$ distinct Pauli indices. The union of these linearly independent error generators form a $((d^2-1)d^2/2)$-dimensional subspace.

3. Active Generators

So far, we have (d^2-1) degrees of freedom from the H generators and $((d^2-1)d^2/2)$ from the S and C generators. On the other hand, TP superoperator space has $d^2(d^2-1)$ degrees of freedom. This leaves us with $(d^2-1)(d^2-2)/2$ dimensions that are not accounted for.

To find generators that span the remaining dimensions, we utilize a well-studied basis, the Choi-sum representation [9]:

$$G[\rho] = \sum_{P,Q} \chi_{P,Q} P\rho Q. \tag{46}$$

The elements of this sum are called Choi units, which span the entire space of d^2 superoperators.

Notice that the H generators are linear combinations of the Choi units $\chi_{\mathbb{1},P}$ and $\chi_{P,\mathbb{1}}$. Similarly S generators are linear combinations of the $\chi_{P,P}$ and $\chi_{\mathbb{1},\mathbb{1}}$ Choi units. On the other hand, C generators contain symmetric linear combinations of the Choi units $\chi_{P,Q}$ and $\chi_{Q,P}$. With

these observations in mind, a new set of elementary generators is constructed with anti-symmetrized Choi-units such that they are linearly independent from the rest. These are called "active" generators:

$$A_{P,Q}[\rho] = i(P\rho Q - Q\rho P + \frac{1}{2}\{[P,Q],\rho\}). \tag{47}$$

With these $(d^2-1)(d^2-2)/2$ degrees of freedom, we are now able to span TP superoperator space.

4. Elementary Error Generators for 1 Qubit

The elementary error generators are:

$$H_P = -i[P,\rho], \tag{48}$$
$$S_P = P\rho P - \rho, \tag{49}$$
$$C_{P,Q} = P\rho Q + Q\rho P - \frac{1}{2}\{\{P,Q\},\rho\}, \tag{50}$$
$$A_{P,Q} = i(P\rho Q - Q\rho P + \frac{1}{2}\{[P,Q],\rho\}). \tag{51}$$

For 1 qubit where $P,Q \in \{X,Y,Z\}$, we can describe a quantum process with the parameter vector

$$\vec{\theta} = (h_x, h_y, h_z, s_x, s_y, s_z, c_{x,y}, c_{x,z}, c_{y,z}, a_{x,y}, a_{x,z}, a_{y,z}), \tag{52}$$

through the mapping:

$$\mathcal{E}[\rho] = e^L \bar{G},$$
$$L = \sum_R h_R H_R + s_R S_R + \sum_{P,Q>P} c_{P,Q} C_{P,Q} + a_{P,Q} A_{P,Q}.$$

Hamiltonian generators cause unitary rotations on the Bloch sphere around the axis defined by the index of the generator, as seen in fig. 3.A. These are used to describe gate over/under-rotations and axis misalignment.

Pauli-Stochastic generators cause shrinking of the Bloch Sphere along orthogonal directions to the axis defined on the generator's index, as shown in figure 3.B. Shrinking of the Bloch sphere along non-principal axes is achieved through the inclusion of Pauli-correlation error generators, as seen in figure 3.C. Stochastic channels are ubiquitous in error-correction frameworks, typically in the form of depolarizing channels.

Finally, active generators result in affine shifts of the Bloch sphere towards the orthogonal axis to its indices as shown in figure 3.D. These are necessary to generate any non-unital channel, such as amplitude damping, a well-studied channel in the literature. When describing multi-qubit dynamics, these generators result in more than just affine shifts. They can be interpreted as the environment causing different unitary rotations which depend on measurement outcomes on the system, known as *active* feedback. Physical interpretations from non-affine effects from these generators are more mysterious than the rest of the elementary generators.

It is important to point out that C and A generators cannot generate a CPTP map on their own, as they map points inside of the Bloch sphere out of it. To create a CPTP map they must be paired with other appropriate S generators that keep states contained within the Bloch sphere. This constraint can be described by two simple inequalities which, if satisfied, ensures the resulting channel is CPTP. The inequalities are:

$$|c_{P,Q}| \leq \sqrt{s_P s_Q}, \tag{53}$$
$$|a_{P,Q}| \leq \sqrt{s_P s_Q}. \tag{54}$$

The elementary generators, although agnostic to hardware platforms, serve as an excellent medium for tomographers to translate GST results into physical mechanisms. It is important to have in consideration that the elementary error generators as modeled in this section can only help infer physical *effects*, instead of literal physical phenomena, due to the post-gate error model. For example, consider a quantum device with one available gate, X. Let there be a bug in its classical control code, which results in an application of a second unitary gate, U, *before* X is applied

$$G = XU. \tag{55}$$

If $[X,U] \neq 0$, this gate is *not* equivalent to X followed by an application of U. Consequently, if we estimate the gate set using the post-gate error model

$$G = e^{L(\vec{\theta})} X, \tag{56}$$

it would result in a set of parameters $\hat{\theta}$ that do not replicate the "real" error, $e^{L(\hat{\theta})} \neq U$. Fortunately, for the sake of interpretability, an error's subspace (within H, S, C or A) does not depend on the choice of pre- or post-gate models. This means that in the previous example, although $e^{L(\hat{\theta})} \neq U$, because U is composed only of H generators (due to its unitary definition), $e^{L(\hat{\theta})}$ must also be unitary.

5. Error Generators for noisy SPAM operations

So far, we have only discussed the use of error generators for processes; however, we still need to parameterize noise in SPAM operations. We could approach this problem in a similar fashion model noise by error channels that either precede or follow the perfect SPAM operation,

$$|\rho\rangle\rangle = e^{L_\rho} |\bar{\rho}\rangle\rangle, \tag{57a}$$
$$\langle\langle E| = \langle\langle \bar{E}| e^{L_E}. \tag{57b}$$

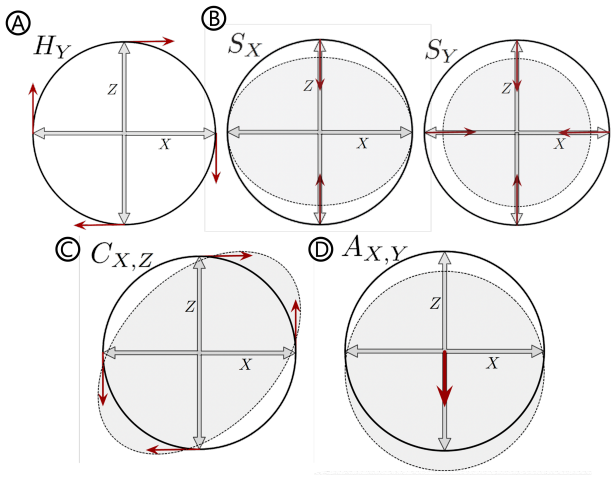

FIG. 3. Effects on different slices of the Bloch Sphere by independent channels generated by a variety of elementary generators. H_y rotates every state along the Y-axis, in unitary fashion. S_x shrinks the X-Z plane along the Z-axis. S_y shrinks all state projections on the X-Z plane. $C_{X,Z}$ expands states along the X+Z axis and shrinks states along the other two orthogonal axes. $A_{X,Y}$ reduces the <Z> value for all states resulting in a downward shift along the Z axis. Figure credit: [7]

A problem with this model is the emergence of additional intrinsic gauge freedom. This gauge space contains all the error channels e^{L_ρ}/e^{L_E} whose eigenspace contains $|\bar{\rho}\rangle\rangle/\langle\langle\bar{E}|$. For example, in a gate set containing a state preparation of $|\bar{\rho}\rangle\rangle = |0\rangle\rangle$, errors that generate rotations along the Z axis will have no effect

$$e^{h_z H_z}|0\rangle\rangle = |0\rangle\rangle. \tag{58}$$

The additional parameters associated with this gauge freedom obscure interpretation of the error generators.

A solution to this is to find the relevant subspace of error generators which does affect the ideal operations. We can do so by, again, considering the first order expansion of our error channel,

$$|\rho\rangle\rangle = e^{L_\rho}|\bar{\rho}\rangle\rangle = (\mathbb{1} + L_\rho)|\bar{\rho}\rangle\rangle = |\bar{\rho}\rangle\rangle + L_\rho|\bar{\rho}\rangle\rangle, \tag{59a}$$
$$|\rho\rangle\rangle = |\bar{\rho}\rangle\rangle + |\rho_N\rangle\rangle. \tag{59b}$$

If we represent Eq. (59b) in the Pauli basis for one qubit, we obtain:

$$|\rho\rangle\rangle = \begin{pmatrix}1\\ \bar{\rho_x}\\ \bar{\rho_y}\\ \bar{\rho_z}\end{pmatrix} + \begin{pmatrix}0\\ \rho_x\\ \rho_y\\ \rho_z\end{pmatrix}. \tag{60}$$

$|\rho_N\rangle\rangle$'s first entry must be 0, due to the trace-preserving property of e^L. This provides a way to represent the errors in a state that only requires three parameters (ρ_x, ρ_y, ρ_z) instead of 12. This effectively removes the intrinsic gauge degrees of freedom. A mapping onto error generators can be found for any given ideal state preparation by finding an expression for the right-hand side of

$$|\rho_N\rangle\rangle = L_\rho|\bar{\rho}\rangle\rangle, \tag{61}$$

which yields the following relationship:

$$\begin{pmatrix}0\\ \rho_x\\ \rho_y\\ \rho_z\end{pmatrix} = \begin{pmatrix}0\\ -2a_{yz} - 2(s_y + s_z)\bar{\rho_x} + 2(c_{xy} - h_z)\bar{\rho_y} + 2(c_{xz} + h_y)\bar{\rho_z}\\ 2a_{xz} + 2(c_{xy} + h_z)\bar{\rho_x} - 2(sx + sz)\bar{\rho_y} + 2(cyz - hx)\bar{\rho_z}\\ -2 * axy + 2(cxz - hy)\bar{\rho_x} + 2(cyz + hx)\bar{\rho_y} - 2(sx + sy)\bar{\rho_z}\end{pmatrix}. \tag{62}$$

Noisy measurements need a more careful treatment. As mentioned in section II, the only conditions for POVMs concern the collection of all its effects instead of individual superbras. As a result, there are as many constraints as there are entries in an effect (4 for one qubit). On the other hand, the number of degrees of freedom in a POVM depends on the number of elements it contains. As a consequence, unlike states and pro-

cesses, the number of parameters needed to construct a comprehensive model of a noisy POVM can vary for a fixed number of qubits. Attempting to use the model described in Eq. (57b) implies that the *same* error channel e^{L_E}, precedes *all* ideal elements of a POVM. This results in only a *constant* number of parameters to characterize noise with a *variable* number of degrees of freedom. By the pigeonhole principle, this is not guaranteed to work for a POVM with too many elements. To circumvent this problem, we choose a similar parameterization as the one used for state preparations. Just as we did in Eq. (59b), we consider the action of a post-gate error channel to first order. For a one qubit system, with a standard POVM (M) with 2 elements ($\langle\langle E_0|, \langle\langle E_1|$) we get:

$$M = \bar{M} + M_N, \tag{63}$$

where

$$\bar{M} = \begin{pmatrix} \langle\langle \bar{E}_0| + \langle\langle \bar{E}_1| \\ \langle\langle \bar{E}_0| - \langle\langle \bar{E}_1| \end{pmatrix} = \begin{pmatrix} \langle\langle \mathbb{1}| \\ \langle\langle \bar{E}_0 - \bar{E}_1| \end{pmatrix}, \tag{64a}$$

$$M_N = \begin{pmatrix} 0 & 0 & 0 & 0 \\ M_1 & M_x & M_y & M_z \end{pmatrix}. \tag{64b}$$

In this case, due to the low number of POVM elements, we can also find a representation with elementary error generators by following the same procedure done above for state preparations if desired.

We have now obtained a parameterization for every operation in a one qubit gate set. The parameters for a gate set consisting of a single state preparation, a single two element POVM, and one gate G can be collected into one vector

$$\vec{\theta} = (\rho_x, \rho_y, \rho_z, M_1, M_x, M_y, M_z, h_x, h_y, h_z, \\ s_x, s_y, s_z, c_{x,y}, c_{x,z}, c_{y,z}, a_{x,y}, a_{x,z}, a_{y,z}).$$

V. AUTOMATED MODEL SELECTION (AMS)

One of the biggest drawbacks of all tomographic methods is the overwhelming amount of information that they output. Even for a mere two qubits, *each* gate needs 240 elementary error generator parameters to build a model of its noise. Regardless of how interpretable a parameterization may be, an infeasible amount of parameters can hinder the tomographer's ability to succeed at inferring physical information. Even though the space of *all* possible noise on a gate set is exponentially large, in practice we expect the details of a device's physics to naturally give rise to a much smaller, physically relevant subspace of errors. In such situations, an accurate noise model can be constructed by setting the physically irrelevant parameters to 0, resulting in a smaller, nested model[4], which we will call a *reduced* model. Unfortunately, identifying descriptive parameters is not an easy task. Unintuitively, the magnitude of a parameter is not always correlated with its descriptive power. Parameters that are small in magnitude can drastically change outcome probabilities of certain circuits if set to 0. Thankfully, the maximum likelihood value is a widely accepted tool to measure the how "good" a model is. If the removal of a parameter significantly decreases the maximum likelihood of the model, then we know that the parameter is necessary to accurately describe the data[10].

Experimental implementations of this concept have already been *manually* deployed in order to obtain interpretable models for two qubit GST[11]. Implementing an automated reduced model finder is a key step toward unlocking GST capabilities for systems with more than two qubits.

The goal is to construct an algorithm that finds a model that has the least number of parameters and still accurately describes experimental data. As motivated earlier, I utilize the maximum likelihood $\mathcal{L}$ of a model to quantify its predictive power. For mathematical convenience, the logarithm of the likelihood value is used instead. Having this in mind, I have implemented a simple greedy algorithm as follows:

Algorithm 1 Greedy AMS Algorithm

```
D ← data collected
T ← assigned by user                          ▷ Threshold
P ← GST(Full Model, data)                     ▷ Parent model
L ← list of 0's                               ▷ reduced models' log(L) list
while min(L) ≤ T do
    X ← log(L(P))
    i = 0
    for R in reducedModels(P) do
        R' ← GST(R, data)                     ▷ pyGSTi's Standard GST
        Y ← log(L(R'))
        L[i] = X − Y                          ▷ Δ log(L)
        i+ = 1
    end for
    P = R' with lowest L/highest Y
end while
```

We start by receiving a threshold T and the full model (also called the parent model) from the user. The threshold identifies how many units of $\log(\mathcal{L})$ the user is willing to sacrifice for each parameter removed. We then build a list of reduced models from the parent model, where each reduced model has a single, and different, parameter removed from the parent model (reducedModels(P)). We estimate the parameter values for each reduced model through GST (with the same initial collected data), and calculate their $\log(\mathcal{L})$ values. The reduced model with the highest $\log(\mathcal{L})$ is selected as the new candidate parent model for the next iteration. This process is repeated

[4] Nested models refers to the construction of a child model whose parameters are a subset of the parent's parameters. By construction, a child model cannot have higher expressive power than its parent.

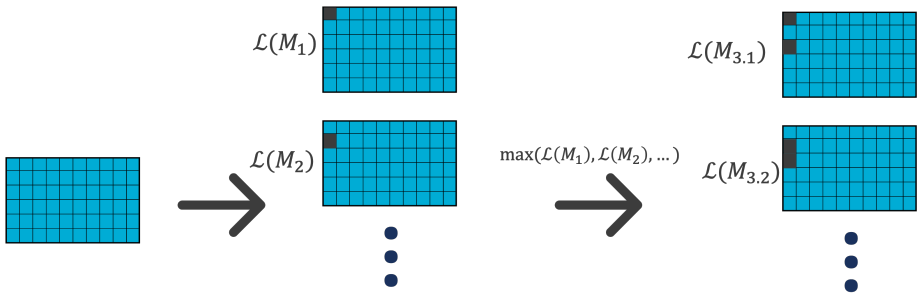

FIG. 4. Greedy AMS algorithm: Every box represents a model, and the squares within a box represent the model's parameters, where a black square depicts a parameter that was removed. Starting on the left, we start with a model containing all of its N parameters, and estimate their values through GST. We then create N reduced models, where each of them have a different parameters removed, run GST on each of them, and calculate their $\log(\mathcal{L})$. Finally, we mark the reduced model with the highest likelihood, use it as the new parent model and repeat from the beginning. This is repeated until the difference in $\log(\mathcal{L})$ between the parent model and the newly marked parent model exceeds a threshold set by the user.

until all possible model reductions result in a decrease in log-likelihood greater than the specified threshold. This algorithm is illustrated in Fig. 4.

An important part of AMS is missing: *how should the user choose the* $\Delta\log(\mathcal{L})$ *loss threshold,* T*?* To answer this we can leverage tools from statistical model selection theory. If the "true" value of the parameter is 0, assuming the statistical noise is normally distributed, then on average the $\log(\mathcal{L})$ decreases by $\frac{1}{2}$ a unit when removed. Any value of $\Delta\log(\mathcal{L}) \leq \frac{1}{2}$ corresponds to a strictly negligible parameter. Furthermore, Akaike's Information Criterion (AIC) suggests a $T \geq 1$ threshold [10].

This algorithm has been tested before, utilizing error generators as parameters, with modest results [4]. It was conjectured that a major obstacle to its success was the existence of gauge freedom the parameterization.

1. Gauge Freedom

In section III C 1, I showed how gauge freedom appears in the derivation of LGST. We also alluded to the fact that this is an intrinsic property of gate sets. To see why this is the case, consider an arbitrary gate set:

$$\mathcal{G} = \{\{|\rho_i\rangle\rangle\}_{i=1}^{N_\rho}; \{G_i\}_{i=1}^{N_G}; \{\langle\langle E_i|\}_{i=1}^{N_E}\}. \tag{65}$$

Just like before, consider the outcome probability of an arbitrary circuit:

$$p(i, \{g_1, \ldots, g_n\}, j) = \langle\langle E_i|G_{g_1} \ldots G_{g_n}|\rho_j\rangle\rangle. \tag{66}$$

This inner product is invariant under the insertion of identity operators between every operation, which can be decomposed into any invertible matrix as follows:

$$p(i, \{g_1, \ldots, g_n\}, j) = \langle\langle E_i|\mathbb{1}G_{g_1}\mathbb{1} \ldots \mathbb{1}G_{g_n}\mathbb{1}|\rho_j\rangle\rangle, \tag{67a}$$
$$= \langle\langle E_i|TT^{-1}G_{g_1}TT^{-1} \ldots TT^{-1}G_{g_n}TT^{-1}|\rho_j\rangle\rangle. \tag{67b}$$

As a consequence, there exists an infinite number of gauge-equivalent gate sets, $\mathcal{G} \xrightarrow{T} \mathcal{G}'$, defined by the gauge transformation described above

$$\mathcal{G}' = \{\{T^{-1}|\rho_i\rangle\rangle\}_{i=1}^{N_\rho}; \{T^{-1}G_iT\}_{i=1}^{N_G}; \{\langle\langle E_i|T\}_{i=1}^{N_E}\}. \tag{68}$$

A set of gate sets which are equivalent up to a gauge transformation, is called a "gauge orbit". Choosing the matrix T corresponds to a choice of gauge, which specifies a point within a gauge orbit.

Gauge choice is quite important for tomographers due to the prevalence of standard gauge-dependent metrics used to benchmark quantum devices, such as fidelity and trace distance. Fundamentally, gauge-dependent metrics, devoid of other context, do not correspond to physically observable quantities. Regardless, these standard metrics can still be useful within the right context, and fully gauge-invariant replacements are not known.

2. Gauge Freedom + AMS = Gauge Problem

In addition to the arguments given above as to why making physical inferences from gauge-dependent metrics is problematic, gauge dependent parameterizations cause additional problems for the AMS algorithm described in this work. To illustrate the problem, let us work through an example.

Consider an arbitrary gauge-dependent gate set parameterization that consists only three parameters $\vec{\theta} = (\theta_\alpha, \theta_\beta, \theta_\gamma)$. Now, let there be a gauge orbit where only

two models contain at least one parameter equal 0, call their parameter vectors $\vec{\theta}_1$ and $\vec{\theta}_2$ such that

$$\mathcal{G}(\vec{\theta}_1) \stackrel{T}{=} \mathcal{G}(\vec{\theta}_2), \tag{69}$$

where

$$\vec{\theta}_1 = (0, a, b), \tag{70a}$$
$$\vec{\theta}_2 = (c, 0, 0). \tag{70b}$$

A model from this gauge orbit is then used to generate a set of noiseless data. In this scenario, it is clear that the ideal reduced model corresponds to $\vec{\theta_I}^R = (c, \times, \times)$, resulting in a single parameter model. Recall that "removing" a parameter, is equivalent to pinning that parameter to 0. As a consequence, the $log(\mathcal{L})$ value for all of these models is equal, and maximal:

$$\log\Big(\mathcal{L}(\vec{\theta_1})\Big) = \log\Big(\mathcal{L}(\vec{\theta_2})\Big) = \max_{\vec{\theta}} \log\Big(\mathcal{L}(\vec{\theta})\Big) \tag{71a}$$
$$= \log\Big(\mathcal{L}(\vec{\theta_I}^R)\Big) \tag{71b}$$

When AMS is run with this data, and input with a full model, its first step would construct three reduced models, each corresponding to a different parameter being removed, which are then fit through GST. Asumming the optimizer within the GST protocol is able to find an absolute maxima, it would result in the following three reduced models:

$$\vec{\theta}_1^R = (\times, a, b),$$
$$\vec{\theta}_2^R = (c, \times, 0),$$
$$\vec{\theta}_3^R = (c, 0, \times).$$

The next step in our algorithm corresponds to choosing the model with the highest $\log(\mathcal{L})$ (lowest decrease in $\log(\mathcal{L})$ from parent). All of these reduced models also lie within the same orbit as the data generating model. Thus, they all have the same $\log(\mathcal{L})$ value, there is no clear choice for AMS to "pick the highest log-likelihood model". Depending on the details of the algorithm implementation (not specified in the pseudo-code above), AMS could choose model $\vec{\theta}_1^R$. The next iteration of reduced models would correspond to $(\times, \times, \theta_\gamma)$ and $(\times, \theta_\beta, \times)$. In this example AMS would be unable to find the smallest, one-parameter model $\vec{\theta}_I^R$.

A naive solution would be to implement a $\log(\mathcal{L})$ tie-breaker condition which chooses the reduced model with the highest number of zero-valued parameters. This solution only works in the example crafted above, in a noiseless sampling scenario. In practice, where gauge landscapes are much more complicated, and noise is present, this fix would still be vulnerable to the same problems. Fundamentally, the issue lies in the gauge-freedom to transfer errors between parameters. Ideally, with a gauge-invariant parameterization, this problem would be avoided.

A. First Order Gauge-Invariant Parameterization

The construction of a non-trivial parameterization that uniquely identifies gauge orbits does not seem to be an easy task. In this section, I present unpublished work in which such a parameterization is derived by restricting the space of gate sets to the special regime of small errors. Under this assumption, we obtain gate set quantities which are approximately gauge-invariant and can help identify gauge orbits. We call these "first order gauge-invariant" (FOGI) quantities.

As motivated earlier, we will focus on gate set *errors*. Processes are parameterized by elementary error generator coefficients, and SPAM operations by affine shift coefficients. Together, these parameters describe gate sets with respect to a target gate set $\bar{\mathcal{G}}$,

$$\mathcal{G}(\vec{\theta}) = \{\{|\bar{\rho}_i\rangle\rangle + |\rho_N(\vec{\theta})\rangle\rangle\}_{i=1}^{N_\rho}; \{e^{L_{G_i}(\vec{\theta})}\bar{G}_i\}_{i=1}^{N_G}; \{\bar{M}_i + M_{Ni}(\vec{\theta})\}_{i=1}^{N_E}\}. \tag{72}$$

Gauge transformations applied to $\mathcal{G}$ have the effect of changing the parameter values of our representation in Eq. (72)

$$\mathcal{G}(\vec{\theta}) \stackrel{T}{\to} \mathcal{G}(\vec{\theta'}), \tag{73a}$$
$$e^{L(\vec{\theta})}\bar{G} \stackrel{T}{\to} e^{L(\vec{\theta'})}\bar{G}, \tag{73b}$$
$$|\bar{\rho}\rangle\rangle + |\rho_N(\vec{\theta})\rangle\rangle \stackrel{T}{\to} |\bar{\rho}\rangle\rangle + |\rho_N(\vec{\theta'})\rangle\rangle, \tag{73c}$$
$$\bar{M} + M_N(\vec{\theta}) \stackrel{T}{\to} \bar{M} + M_N(\vec{\theta'}). \tag{73d}$$

First, consider a scalar function $f : \Theta \to \mathbb{R}$ that takes a gate set's parameters as its input $f(\vec{\theta})$, we call this a "property" of the gate set. If a gate set property has the same value for all gauge-equivalent models, then f is a gauge-invariant property. i.e.,

$$\forall T, \vec{\theta} \text{ s.t. } \mathcal{G}(\vec{\theta'}) \stackrel{T}{\to} \mathcal{G}(\vec{\theta})), f(\vec{\theta'}) = f(\vec{\theta})$$
$$\longleftrightarrow f \text{ is a gauge-invariant property.}$$

Where $\stackrel{T}{\to}$ describes the effect of a gauge transformation as shown in Eq. (68). Plugging this transformation into Eqs. (73 a-d), we obtain the effect of gauge transformations on error parameters for:
processes:

$$T^{-1}e^{L_G(\vec{\theta})}\bar{G}T = e^{L_G(\vec{\theta'})}\bar{G}, \tag{74}$$
$$\to e^{L_G(\vec{\theta'})} = T^{-1}e^{L_G(\vec{\theta})}\bar{G}T\bar{G}^{-1}; \tag{75}$$

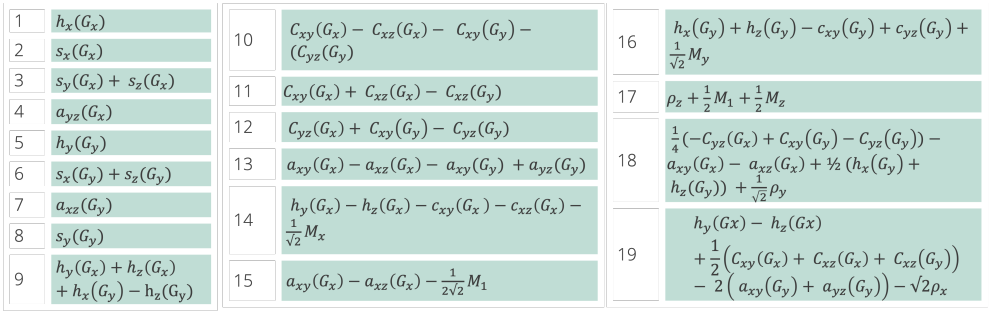

FIG. 5. A set of 20 quantities that span FOGI space for the gate set $\mathcal{G} = \{\{|0\rangle\rangle\}, \{X, Y\}, \{\langle\langle 0|, \langle\langle 1|\}\}$. Found through the co-kernel of the $\mathcal{A}$ map from Eq. (85), and represented in the basis derived in section IV A 5. These quantities represent physical, observable, directions in parameter space, and are invariant under gauge transformations (up to first order). Error generator rates are followed by the corresponding gate they affect inside parenthesis.

states:

$$T^{-1}(|\bar{\rho}\rangle\rangle + |\rho_N(\vec{\theta})\rangle\rangle) = |\bar{\rho}\rangle\rangle + |\rho_N(\vec{\theta'})\rangle\rangle, \tag{76}$$

$$\rightarrow |\rho_N(\vec{\theta'})\rangle\rangle = T^{-1}(|\bar{\rho}\rangle\rangle + |\rho_N(\vec{\theta})\rangle\rangle) - |\bar{\rho}\rangle\rangle; \tag{77}$$

and POVMs:

$$(\bar{M} + M_N(\vec{\theta}))T = \bar{M} + M_N(\vec{\theta'}), \tag{78}$$

$$\rightarrow M_N(\vec{\theta'}) = (\bar{M} + M_N(\vec{\theta}))T - \bar{M}. \tag{79}$$

Now, I formalize the small error assumption by only considering gate sets that are close enough to the target gate set $\bar{G}$, such that these second-order terms can be neglected:

$$|\bar{G} - G|^2 \approx 0. \tag{80}$$

As a consequence, our post-gate error model for generated channels from Eq. (30) can be simplified to:

$$G = e^L \bar{G} \approx (\mathbb{1} + L)\bar{G}. \tag{81}$$

Finally, this assumption implies that for any two gauge-equivalent gate sets within this regime, there must be a gauge transformation T between them such that:

$$T = e^{\epsilon K} = \mathbb{1} + \epsilon K + O(\epsilon^2),$$
$$T \approx \mathbb{1} + \epsilon K. \tag{82}$$

Where ϵ is upper-bounded by $|L|$, and $|K| = 1$. We can plug these assumptions into Eq. (75) to obtain the effects of gauge transformations on a quantum process' parameters:

$$(\mathbb{1} + L_G(\vec{\theta'})) = (\mathbb{1} - \epsilon K)(\mathbb{1} + L_G(\vec{\theta}))\bar{G}(\mathbb{1} + \epsilon K)\bar{G}^{-1}, \tag{83}$$

by simplifying and removing $O(\epsilon^2)$ terms this reduces to

$$L_G(\vec{\theta'}) = L_G(\vec{\theta}) + \epsilon \bar{G} K \bar{G}^{-1}. \tag{84}$$

This shows that, up to first order, gauge transformations cause *affine shifts* in our parameters given by $\epsilon\bar{G}K\bar{G}^{-1}$. This defines a linear map from the matrices that generate gauge transformations onto corresponding parameter shift values. With a similar derivation, we can obtain the same mapping for SPAM operators and write them together as:

$$\mathcal{A}(K) = \begin{cases} \epsilon\bar{G}K\bar{G}^{-}1 & \text{Processes,} \\ -K|\bar{\rho}\rangle\rangle & \text{States,} \\ \bar{M}K & \text{Measurements.} \end{cases} \tag{85}$$

The range of this map contains all points in parameter space that when added to a gate set's parameters, results in a different gauge-equivalent gate set. We refer to this map as a "gauge action" function. Thus, the complement of the range of $\mathcal{A}$ ($\overline{\mathcal{R}(\mathcal{A})}$) corresponds to observable degrees of freedom that are orthogonal or insensitive to gauge transformations (up to first order). We therefore call $\overline{\mathcal{R}(\mathcal{A})}$ FOGI space. One can find an infinite number of ways to span this space by finding a basis for the co-kernel of $\mathcal{A}$. In Fig. 5 we list 19 quantities that span the FOGI space of a gate set $\mathcal{G} = \{\{|0\rangle\rangle\}, \{G_x(\pi/2), G_y(\pi/2)\}, \{\langle\langle 0|, \langle\langle 1|\}\}$.

VI. AUTOMATED MODEL SELECTION WITH FOGI QUANTITIES

It was previously conjectured that the first AMS implementation failed due to gauge-dependent parameterizations. In the last section, I laid the theoretical groundwork to construct gate set models that are resilient to gauge freedom. I now present the results from deploying AMS on FOGI parameterized models for a one-qubit neutral atom device.

First, to validate FOGI-AMS's performance, we begin by studying its performance with simulated data. I constructed noise models from elementary error generators

#FOGI	$s_x(G_x)$ Model	AMS	AMS (Noise)
0	0	×	×
1	0.001	0.00099999998	0.000996
2	0	×	×
3	0	×	×
4	0	×	×
5	0	×	×
6	0	×	×
7	0	×	×
8	0	×	×
9	0	×	×
10	0	×	×
11	0	×	-4.926e-5
12	0	×	×
13	0	×	×
14	0	×	×
15	0	×	×
16	0	×	×
17	0	×	×
18	0	×	×

TABLE I. Output of Automated Model Selection ran on data generated from the $s_x(G_x) = .001$ noise model. Each row represents a different FOGI quantity as enumerated in the first column. Each entry in the $s_x(G_x)$ Model column represents the corresponding FOGI value for the model from which data was simulated. Finally, each row in the AMS columns represents the FOGI value estimated by GST on the smallest model accepted by AMS, where one column includes sampling-noise, and the other does not. AMS correctly identified the optimal model containing a single parameter when no noise is present, and only resulted in one false-positive parameter in the presence of sampling noise.

#FOGI	Full Model GST	AMS
0	-0.0186	-0.0208
1	0.0034	0.0050
2	0.0042	×
3	0.0001	×
4	-0.0227	-0.0257
5	0.0078	0.0100
6	-0.0012	×
7	0.0015	×
8	0.0004	×
9	-0.0049	×
10	0.0046	×
11	-0.0081	×
12	-0.0008	×
13	-0.0362	-0.0340
14	0.0072	0.0076
15	0.0314	0.0296
16	-0.0034	-0.0035
17	-0.0218	-0.0220
18	0.0200	0.0211

TABLE II. Output of Automated Model Selection ran on data collected from a neutral atom device comprised of 1 qubit, G_x & G_y gates, $|1\rangle\rangle$ state preparation, and Z basis measurement. We compare the output of AMS against the best estimate on a full parameter model. AMS found a model with 9 less parameters while only losing a mere total of 1.847 units of log-likelihood.

because it is easier to ensure the resulting gate set contains CPTP processes by satisfying Eqs. 53 & 54. An obvious choice, is to start with a noise model which results in only one non-trivial FOGI quantity. By quick inspection of the set of FOGI quantities, we may notice that the elementary error generator $s_x(G_x)$ only appears in parameter #2. In table I I present the output of AMS ran on the noise model $s_x(G_x) = 0.001$, with and without sampling noise[5], with T = 1. For the noiseless sampling simulation, we successfully obtained the optimal model with a single parameter. On the other hand, when sampling noise is present, AMS returns a single false positive. In this example, any T greater than 1.375 would have achieved the optimal solution for the noisy data. Increasing the value of T comes with the risk of misidentifying signal as noise.

Finally, we present the deployment of AMS on experimental data on table II, collected from Sandia's neutral atom quantum computer with the same gate set as the simulated data above[6]. The experiment consisted of 70 circuits where 200 shots were collected for each one of them. I obtained a model containing half as many parameters which only lost .2 units of log-likelihood per parameter removed.

Every circuit within a GST experiment design is constructed to amplify a specific error [5]. The existence of such a smaller yet almost equally predictive model could

[5] No sampling-error corresponds to circuit outcome frequencies set equal calculated probabilities.

[6] There is one difference between the gate sets. The experiment's state preparation is $|1\rangle\rangle$, instead of $|0\rangle\rangle$. This produces almost identical FOGI quantities as the ones in Fig. 5, except for a couple of sign changes.

enable for much shorter future GST experiments, by removing circuits that amplify errors shown to be irrelevant by AMS. These results show a promising and exciting avenue towards a more efficient characterization workflow for this neutral atom device.

A. Future Work

a. Streamlining the characterization workflow: A significant experimental downside of GST is the insurmountable amount of data that needs to be collected to probe all the physical degrees of freedom within a gate set. As briefly mentioned in the previous section, it is possible to remove certain circuits from an experiment design, if we are able to identify the error they probe as irrelevant. AMS holds the power to uncover these irrelevant errors. It is an open question whether these reduced models will maintain their descriptive power over time under the presence of drift[7].

b. AMS speedup: Currently, every reduced model needs to be run through GST's optimization routine to find the best estimate for its parameters, resulting in incredible computational overhead. Due to the nested nature of the reduced models, maximum likelihood values can be estimated to avoid running GST on child models.

c. AMS from the bottom-up: Our current approach iteratively removes parameters from a fully parameterized model. An opposite approach could work where we add parameters to an empty, zero-parameter model until we do not *gain* enough likelihood from adding any more parameters. This heuristic can be used in parallel to the current one, in order to explore a larger landscape of the models' space.

d. pyGSTi integration: While all the work described in this paper was done within pyGSTi's open-source GitHub repository, it is yet to be included in an official release. Further developments must be made to ensure an intuitive user interface, proper documentation/tutorials, and full integration with the rest of the code base.

e. FOGI GST tools: One of pyGSTi's most popular features is the generation of GST reports, which help tomographers visualize GST results in a well-organized HTML document. Tools based on FOGI models, such as FOGI error metrics, will be added as part of default report construction.

ACKNOWLEDGMENTS

I wish to acknowledge the support and mentorship of Corey Ostrove and Robin Blume-Kohout for countless hours of lectures, discussions, and pair programming; everyone at the Quantum Performance Laboratory; my collaborators at Sandia's neutral atom quantum computing team, Matthew Chow and Bethany Little, for collecting the experimental data; all my colleagues at the Center for Quantum Information and Control at the University of New Mexico; and co-workers Cole Maurer, Mohsin Raza, Basie Seibert, and Chaithanya Rayudu for insightful conversations and suggestions. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.

[1] D. Deutsch, Quantum theory, the church–turing principle and the universal quantum computer, Proc. R. Soc. A **400**, 97 (1985).
[2] S. Haroche and J. Raimond, Quantum Computing: Dream or Nightmare?, Physics Today **49**, 51 (1996), https://pubs.aip.org/physicstoday/article-pdf/49/8/51/11027736/51_1_online.pdf.
[3] I. H. Deutsch, Harnessing the power of the second quantum revolution, PRX Quantum **1**, 020101 (2020).
[4] S. Seritan, K. Rudinger, T. Proctor, and R. Blume-Kohout, Automated Model Selection for Gate Set Tomography, in *APS March Meeting Abstracts*, APS Meeting Abstracts, Vol. 2023 (2023) p. B72.013.
[5] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).
[6] J. Berger and R. Wolpert, *The Likelihood Principle*, Institute of Mathematical Statistics. Lecture notes : monographs series (Institute of Mathematical Statistics, 1988).
[7] R. Blume-Kohout, M. P. da Silva, E. Nielsen, T. Proctor, K. Rudinger, M. Sarovar, and K. Young, A taxonomy of small markovian errors, PRX Quantum **3**, 10.1103/prxquantum.3.020335 (2022).
[8] J. Gallier, Logarithms and square roots of real matrices (2013), arXiv:0805.0245 [math.GM].
[9] G. Homa, A. Ortega, and M. Koniorczyk, Choi representation of completely positive maps: a technical introduction (2024), arXiv:2402.12944 [quant-ph].
[10] K. Burnham and D. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach* (Springer Verlag, 2002).
[11] M. T. Mądzik, S. Asaad, A. Youssry, B. Joecker, K. M. Rudinger, E. Nielsen, K. C. Young, T. J. Proctor, A. D. Baczewski, A. Laucht, V. Schmitt, F. E. Hudson, K. M. Itoh, A. M. Jakob, B. C. Johnson, D. N. Jamieson, A. S. Dzurak, C. Ferrie, R. Blume-Kohout, and A. Morello, Precision tomography of a three-qubit donor quantum processor in silicon, Nature **601**, 348–353 (2022).

[7] Drift refers to the change in the errors affecting a quantum device that occurs naturally with time.