Randomized Linear Gate-Set Tomography

Yanwu Gu[1,2,*] Rajesh Mishra,[1,2] Berthold-Georg Englert,[2,3,4] and Hui Khoon Ng[1,2,4,†]

[1]*Yale-NUS College, Singapore*

[2]*Centre for Quantum Technologies, National University of Singapore, Singapore*

[3]*Department of Physics, National University of Singapore, Singapore*

[4]*MajuLab, CNRS-UCA-SU-NUS-NTU International Joint Research Unit, Singapore*

Characterizingthe noise in the set of gate operations that form the building blocks of a quantum computational device is a necessity for assessing the quality of the device. Here, we introduce *randomized linear gate-set tomography*, an easy-to-implement gate-set-tomography procedure that combines the idea of state-preparation-and-measurement-error-free characterization of standard gate-set tomography with no-design randomized tomographic circuits and computational ease brought about by an appropriate linear approximation. We demonstrate the performance of our scheme through simulated examples as well as experiments done on the IBM Quantum Experience Platform. In each case, we see that the performance of our procedure is comparable with that of standard gate-set tomography, while requiring no complicated tomographic circuit design and taking much less computational time in deducing the estimate of the noise parameters. This allows for straightforward on-the-fly characterization of the gate operations in an experiment.

I. INTRODUCTION

With the recent advent of quantum-computing devices, albeit small-scale ones only, there has been a lot of interest in understanding the nature of imperfections in the quantum gate operations, as a gauge of the quality of the computation as well as the hurdles to overcome in the quest for larger devices. While randomized benchmarking methods [1–4] are the go-to approach for a partial but easy gauge of gate quality, quantum process tomography [5–7] is the conventional method for full characterization of the noise in gate operations. One prepares a specific set of quantum states as inputs to the gate of interest and measures a specific set of observables at the output. A comparison of the measurement results with the expected outputs from an ideal no-noise gate gives an estimate of the noise process of the actual gate operation. To do quantum process tomography well thus requires well-calibrated input preparation and measurement procedures, much better characterized than the noisy gate of interest. The current

experimental reality, however, is such that the preparation and measurement procedures suffer from equally severe, and—more problematically—unstable, noise as the gates of interest, the so-called state preparation and measurement (SPAM) errors. A separate precalibration, prior to the process tomography, is often not good enough to get rid of the effects of SPAM errors in the tomography results.

A more feasible approach is to characterize simultaneously the noise in the preparation-and-measurement procedures, alongside the noise in the gate operation itself. Earlier self-calibration ideas [8–12], for state tomography but in principle applicable to process tomography, seek to deduce parameters of the SPAM procedures from the same tomographic data. More recently, a suite of methods under the heading of gate-set tomography (GST) [13–16] has emerged as a comprehensive approach to characterizing a "gate set"—a specified set of gates that can be implemented in a quantum-information-processing device, for both computational tasks as well as the SPAM procedures—while avoiding the issue of SPAM errors. GST has been employed in various experiments; see, for example, Refs. [17–22].

In short, GST handles the issue of SPAM errors by acknowledging the presence of noise in the SPAM procedures within the estimation itself. The task is no longer the characterization of a specific gate using other precalibrated gates in SPAM, but to characterize all the gates, those used in the computation as well as those in the SPAM

*a0129460@u.nus.edu

†huikhoon.ng@yale-nus.edu.sg

procedures, from the same tomographic data, reminiscent of the early self-calibration work in state tomography. GST uses circuits—sequences of gates taken from the gate set of interest—with the structure of quantum process tomography: an initial segment of gates corresponding to the preparation, from an initial default state, of a tomographic input state; a middle segment comprising the gate(s) of interest; a final segment containing the gates that, together with a final default measurement, implement the measurement of a particular tomographic observable. In the actual GST noise-reconstruction procedure, no difference is made between the gates from the three segments—all gates are considered as noisy, and the noise parameters of the entire gate set are the targets of the reconstruction—but this underlying structure is the basis of the GST circuit design.

In this work, we introduce a simpler alternative to GST, offering speed of reconstruction (seconds compared to half an hour for GST in a two-qubit case) based on a linear-noise regime, easy circuit design (randomized compared to the carefully designed structured GST circuits), and applicability to tomographically incomplete gate sets, not naturally accessible within the GST framework, while yielding results quite comparable to those from GST. GST, in its entirety, can give a more accurate reconstruction of the noise, but if one is looking for an on-the-fly and easy-to-implement approximate characterization of the current noise in a device, our *randomized linear gate-set tomography* (RL-GST) offers a quick and simple solution. Potentially, the results from our RL-GST approach can also be used as the starting point for the optimization procedure in the usual GST, for a more refined noise estimate. The approximation of a linear-noise regime, as we see, restricts RL-GST to short circuit lengths compared with the inverse of the noise strength, but such a requirement is natural anyway for accurate calculations by the quantum devices.

Below, we first describe our RL-GST procedure, explaining the linear-noise regime and the choice of randomized circuits. We follow that with examples, simulated ones as well as experiments on the IBM Quantum Experience Platform [23], to illustrate the performance of RL-GST, in comparison with that of standard GST.

II. TOMOGRAPHY IN THE LINEAR REGIME

A. Preliminaries

Consider a quantum-computing platform capable, ideally, of performing operations from a gate set $\mathfrak{g} \equiv \{g_\gamma\}_{\gamma=1}^{\Gamma}$, preparing (initialization) a default pure input state $|\psi_{\text{in}}\rangle$, and measuring (readout) a default observable $M \equiv \sum_\mu \lambda_\mu P_\mu$, with λ_μs the eigenvalues of M, and P_μs the corresponding eigenoperators, which are projectors not necessarily of rank 1. Here, we assume a discrete set of Γ gates for an n-qubit system. Each g_γ is a unitary map acting on an operator on the n-qubit Hilbert space as $g_\gamma(\cdot) = U_\gamma(\cdot)U_\gamma^\dagger$, where U_γ is a unitary operator.

In a real platform, every component will be noisy. We denote the noisy version of each component by a tilde: $\widetilde{g}_\gamma$ is the noisy implementation of the ideal gate g_γ, $\widetilde{\rho}_{\text{in}}$ is the noisy initialization of the ideal input density operator $\rho_{\text{in}} \equiv |\psi_{\text{in}}\rangle\langle\psi_{\text{in}}|$, and $\widetilde{M} = \sum_\mu \lambda_\mu \widetilde{P}_\mu$ is the noisy version of the ideal readout M. While g_γ is a unitary map, $\widetilde{g}_\gamma$ is assumed to be a general—not necessarily unitary—completely positive (CP) and trace-preserving (TP) linear map. We write the noisy gate as $\widetilde{g}_\gamma = (\mathbb{1} + e_\gamma)g_\gamma$, so that $\mathbb{1} + e_\gamma \equiv \widetilde{g}_\gamma g_\gamma^{-1}$ is the CPTP noise map for the gate g_γ, and e_γ is its deviation from the identity map $\mathbb{1}$. Note that e_γ is not TP; in fact, $\text{tr}\, e_\gamma(\cdot) = 0$ for any input. Neither is e_γ necessarily CP; only $\mathbb{1} + e_\gamma$ is CP. Similarly, we assume that $\widetilde{\rho}_{\text{in}} = (\mathbb{1} + e_{\text{in}})(\rho_{\text{in}})$ and $\widetilde{P}_\mu = (\mathbb{1} + e_{\text{ro},\mu})^\dagger(P_\mu)$, for CPTP $\mathbb{1} + e_{\text{in}}$ and $\mathbb{1} + e_{\text{ro},\mu}$. $\widetilde{\rho}_{\text{in}} = (\mathbb{1} + e_{\text{in}})(\rho_{\text{in}})$ can be read as having the ideal input ρ_{in} followed by the application of the noise map $\mathbb{1} + e_{\text{in}}$; $\widetilde{P}_\mu = (\mathbb{1} + e_{\text{ro},\mu})^\dagger(P_\mu)$ should be, because of the adjoint, read instead as the application of the noise map $\mathbb{1} + e_{\text{ro},\mu}$, following by the ideal projector P_μ. Note that the set of $e_{\text{ro},\mu}$s are not all independent: We must have $\sum_\mu \widetilde{P}_\mu = \mathbb{1}$, for a valid measurement. For the case of weak noise, necessary for a useful quantum-computing platform, we expect $\|e_\gamma\|, \|e_{\text{in}}\|, \|e_{\text{ro},\mu}\| \ll \|\mathbb{1}\|$, for some map norm $\|\cdot\|$. It is convenient to assume $\|\mathbb{1}\| = 1$, and that $\|e_\gamma\|, \|e_{\text{in}}\|, \|e_{\text{ro},\mu}\| \lesssim \epsilon$ for some positive $\epsilon \ll 1$, for all γ; ϵ will be the book-keeping parameter for our perturbative analysis below.

In the operation of the quantum computer, following what is sometimes referred to as the standard circuit model, the default input state is prepared, and a circuit $\mathcal{C}$—a sequence of gates from the gate set $\mathfrak{g}$—is performed on that input state, and the default observable M is measured. $\mathcal{C}$ contains not just the gate sequence that carries out the desired computation on some input state. It includes, as the starting sequence, the gates that prepare the default input state $|\psi_{\text{in}}\rangle$ into the actual desired input state, and, as the ending sequence, the gates that, together with the readout M, implement a desired measurement on the output state:

$$\mathcal{C} = \underbrace{\mathcal{G}_L \mathcal{G}_{L-1} \cdots}_{\text{measurement}} \cdots \underbrace{\mathcal{G}_{k+1} \mathcal{G}_k \mathcal{G}_{k-1}}_{\text{computation}} \cdots \underbrace{\cdots \mathcal{G}_2 \mathcal{G}_1}_{\text{preparation}}, \quad \mathcal{G}_k \in \mathfrak{g} \ \forall k.$$

(1)

In reality, we have the noisy version of the ideal circuit $\mathcal{C}$, written as $\widetilde{\mathcal{C}} = \widetilde{\mathcal{G}}_L \cdots \widetilde{\mathcal{G}}_1$, where $\widetilde{\mathcal{G}}_k$ is the noisy implementation of the ideal gate $\mathcal{G}_k$. As for the noisy version of the gate-set elements, we write $\widetilde{\mathcal{G}}_k = (\mathbb{1} + \mathcal{E}_k)\mathcal{G}_k$, where $\mathbb{1} + \mathcal{E}_k$ is assumed to be CPTP. We assume a sequence-independent noise model, i.e., $\mathcal{E}_k = e_\gamma$ whenever $\mathcal{G}_k = g_\gamma$.

We introduce some convenient notation before proceeding further. We can "vectorize" the operators on the n-qubit Hilbert space, and write $|O\rangle\!\rangle$ as the vectorized operator O, now considered as a vector in the vector space of operators.

The adjoint vector $\langle\!\langle O|$ is defined such that the Hilbert-Schmidt inner product between two operators O_1 and O_2 can be written as $\mathrm{tr} O_1^\dagger O_2 = \langle\!\langle O_1|O_2\rangle\!\rangle$. We write the action of a map $\mathcal{G}$ on an operator O as $\mathcal{G}(O) = \mathcal{G}|O\rangle\!\rangle$, where we use, for conciseness, the same symbol for the map $\mathcal{G}(\cdot)$ on the left, and the (super)operator (matrix) $\mathcal{G}$ representing the map on the right, acting on the vector (column) $|O\rangle\!\rangle$. Then, we can write the entire quantum computation, from preparing the input state ρ_{in}, to implementing the circuit $\mathcal{C}$, to getting the measurement outcome μ, as $\langle\!\langle P_\mu|\mathcal{C}|\rho_{\mathrm{in}}\rangle\!\rangle$. Its noisy version is, correspondingly, $\langle\!\langle \widetilde{P}_\mu|\widetilde{\mathcal{C}}|\widetilde{\rho}_{\mathrm{in}}\rangle\!\rangle$. We also use the shorthand $\mathcal{G}_{m:l}$, for positive integers m and l, to denote the gate sequence $\mathcal{G}_m\mathcal{G}_{m-1}\cdots\mathcal{G}_{l+1}\mathcal{G}_l$ if $m \geq l$, and the $\mathbb{1}$ map if $m < l$; for example, $\mathcal{C} = \mathcal{G}_{L:1}$, $\mathcal{G}_{1:1} = \mathcal{G}_1$, and $\mathcal{G}_{1:2} = \mathbb{1}$. In addition, we write $_\mu\langle\!\langle \cdot \rangle\!\rangle$ as a shorthand for $\langle\!\langle P_\mu| \cdot |\rho_{\mathrm{in}}\rangle\!\rangle$.

B. Linear-noise regime

We are now ready to define the linear regime for the noise. The output of the quantum-computer running circuit $\mathcal{C}$ are the outcome probabilities $\{p_\mu \equiv \langle\!\langle P_\mu|\mathcal{C}|\rho_{\mathrm{in}}\rangle\!\rangle = {}_\mu\langle\!\langle \mathcal{C}\rangle\!\rangle\}$. We write the noisy version $\widetilde{p}_\mu$ as an "error expansion," organized in powers of ϵ,

$$
\begin{aligned}
\widetilde{p}_\mu &= \langle\!\langle \widetilde{P}_\mu|\widetilde{\mathcal{C}}|\widetilde{\rho}_{\mathrm{in}}\rangle\!\rangle \\
&= {}_\mu\langle\!\langle \mathcal{C}\rangle\!\rangle && [\mathcal{O}(1)\text{ terms}] \\
&+ {}_\mu\langle\!\langle \mathcal{C}e_{\mathrm{in}} + \sum_{k=1}^{L}\mathcal{G}_{L:k+1}\mathcal{E}_k\mathcal{G}_{k:1} + e_{\mathrm{ro},\mu}\mathcal{C}\rangle\!\rangle && [\mathcal{O}(\epsilon)\quad\text{terms}] \\
&+ {}_\mu\langle\!\langle \mathcal{G}_{L:2}\mathcal{E}_1\mathcal{G}_1 e_{\mathrm{in}} + \mathcal{G}_{L:3}\mathcal{E}_2\mathcal{G}_2\mathcal{E}_1\mathcal{G}_1 +\ldots\rangle\!\rangle && [\mathcal{O}(\epsilon^2)\text{terms}] + \ldots.
\end{aligned}
\tag{2}
$$

The linear-noise regime is defined as the linear truncation of this error expansion, keeping up to, and including, the $\mathcal{O}(\epsilon)$ terms, and dropping $\mathcal{O}(\epsilon^2)$ and higher-order terms. This corresponds to at most one insertion of the "error" e_γ, $e_{\mathrm{ro},\mu}$, or e_{in} in each term in the error expansion. This linear regime is a good approximation only if

$$
L\epsilon \ll 1,
\tag{3}
$$

so that the neglected higher-order terms give small corrections.

Our goal is to estimate the errors e_γ, $e_{\mathrm{ro},\mu}$, and e_{in}, relevant for a quantum-computing platform with its gate set and its initialization and readout procedures, well enough for making predictions of the outcomes $\{p_\mu\}$ of circuits run on the platform. We do this by running a set of tomographic circuits $\{\mathcal{C}^{(i)}\}$ on the quantum-computing platform and recording the outcome data. From the collected data, our adoption of the linear-noise regime permits a straightforward linear-inversion procedure (see below) for obtaining estimates of the error maps e_γ, $e_{\mathrm{ro},\mu}$, and e_{in}. The linear-noise regime captures the dominant sources of noise in the platform, and the errors estimated within this regime suffice to make accurate predictions of the outcomes of circuits with length $\lesssim 1/\epsilon$.

C. Linear inversion

Let us first rewrite Eq. (2) in the linear regime, for circuit $\mathcal{C}^{(i)}$, as

$$
\begin{aligned}
&\widetilde{p}_\mu^{(i)} - p_\mu^{(i)} \\
&\simeq {}_\mu\langle\!\langle \mathcal{C}^{(i)}e_{\mathrm{in}} + \sum_{k=1}^{L}\mathcal{G}_{L:k+1}^{(i)}\mathcal{E}_k^{(i)}\mathcal{G}_{k:1}^{(i)} + e_{\mathrm{ro},\mu}\mathcal{C}^{(i)}\rangle\!\rangle \\
&= {}_\mu\langle\!\langle \mathcal{C}^{(i)}e_{\mathrm{in}} + \sum_{\gamma=1}^{\Gamma}\sum_{k\in\mathrm{Pos}_\gamma^{(i)}}\mathcal{G}_{L:k+1}^{(i)}e_\gamma g_\gamma \mathcal{G}_{k-1:1}^{(i)} + e_{\mathrm{ro},\mu}\mathcal{C}^{(i)}\rangle\!\rangle,
\end{aligned}
\tag{4}
$$

where $\mathrm{Pos}_\gamma^{(i)}$ denotes the subset of indices k corresponding to the positions in the circuit $\mathcal{C}^{(i)} = \mathcal{G}_L^{(i)}\mathcal{G}_{L-1}^{(i)}\ldots\mathcal{G}_1^{(i)} = \mathcal{G}_{L:1}^{(i)}$, where $\mathcal{G}_k^{(i)} = g_\gamma$. For concreteness, we introduce an orthonormal basis $\{|B_a\rangle\!\rangle\}_{a=1}^{d^2}$ (d is the dimension of the Hilbert space), and write the gate errors as $e_\gamma = \sum_{ab}e_{\gamma;ab}|B_a\rangle\!\rangle\langle\!\langle B_b|$, for complex matrix elements $e_{\gamma;ab} \equiv \langle\!\langle B_a|e_\gamma|B_b\rangle\!\rangle$. In addition, we write the noisy initialization as $e_{\mathrm{in}}|\rho_{\mathrm{in}}\rangle\!\rangle = \sum_a e_{\mathrm{in};a}|B_a\rangle\!\rangle$, for complex scalars $e_{\mathrm{in};a} \equiv \langle\!\langle B_a|e_{\mathrm{in}}|\rho_{\mathrm{in}}\rangle\!\rangle$; the noisy readout is, similarly, $\langle\!\langle P_\mu|e_{\mathrm{ro},\mu} = \sum_a e_{\mathrm{ro},\mu;a}\langle\!\langle B_a|$ for $e_{\mathrm{ro},\mu;a} \equiv \langle\!\langle P_\mu|e_{\mathrm{ro},\mu}|B_a\rangle\!\rangle$. The complex parameters $e_{\gamma;ab}$, $e_{\mathrm{in};a}$, and $e_{\mathrm{ro},\mu;a}$ fully describe the noise in the platform, and are the quantities we want to estimate.

In terms of these noise parameters, inserting the identity $\sum_a |B_a\rangle\rangle\langle\langle B_a|$ at appropriate places, Eq. (4) can be rewritten as

$$\widetilde{p}_\mu^{(i)} - p_\mu^{(i)} \simeq c_{\mu;a}^{(i)} e_{\text{in};a} + c_{\text{in},\mu\gamma;ab}^{(i)} e_{\gamma;ab} + c_{\text{in};a}^{(i)} e_{\text{ro},\mu;a}, \quad (5)$$

with the understanding that repeated indices are summed. Here, the various $c^{(i)}$ coefficients are complex scalars determined solely by the ideal circuit $\mathcal{C}^{(i)}$ and our basis choice: $c_{\text{in};a}^{(i)} \equiv \langle\langle B_a|\mathcal{C}^{(i)}|\rho_{\text{in}}\rangle\rangle$, $c_{\text{in},\mu\gamma;ab}^{(i)} \equiv \sum_{k\in\text{Pos}_\gamma^{(i)}} \langle\langle P_\mu| \mathcal{G}_{L:k+1}^{(i)}|B_a\rangle\rangle \langle\langle B_b|g_\gamma \mathcal{G}_{k-1:1}^{(i)}|\rho_{\text{in}}\rangle\rangle$, and $c_{\mu;a}^{(i)} \equiv \langle\langle P_\mu|C^{(i)}|B_a\rangle\rangle$. Each circuit $\mathcal{C}^{(i)}$ gives a set of equations, Eq. (5) for different μs, for the noise parameters e, with $\widetilde{p}^{(i)}$ estimated from the experimental data, and $p^{(i)}$ and $c^{(i)}$ coefficients computed from the ideal $\mathcal{C}^{(i)}$ and chosen basis. Stacking up the equations, for some choice of ordering of μ and i, we can write down a system of linear equations for e,

$$\widetilde{\mathbf{p}} - \mathbf{p} = \mathbf{C}\mathbf{e}, \quad (6)$$

where $\widetilde{\mathbf{p}}(\mathbf{p})$ is the column of $\widetilde{p}_\mu^{(i)}(p_\mu^{(i)})$, $\mathbf{e}$ is the column of noise parameters, and $\mathbf{C}$ is a (generally nonsquare) matrix of the c coefficients, with each row comprising the cs for the same μ and i. We refer to $\mathbf{C}$ as the *design matrix*, to emphasize its dependence only on our choice of circuits and basis, not on the noise maps of the system. The noise parameters are then estimated by solving Eq. (6) for $\mathbf{e}$. We do this by linear inversion, which essentially amounts to writing $\mathbf{e} = \mathbf{C}^{-1}(\widetilde{\mathbf{p}} - \mathbf{p})$, provided we can make sense of the inverse of $\mathbf{C}$. This is what we discuss next.

A remark, before we discuss the design matrix further: The linear-noise regime is a natural approximation in the weak-noise limit, the expected scenario for any useful quantum-computing platform. It has been used in many situations, including randomized benchmarking and idle tomography [24–26]. In particular, the linear-noise regime was employed previously, with an accompanying linear-inversion procedure, in the precursor, Ref. [13], to the current formulation of GST. There, however, the characterization circuits all follow the SPAM-gate structure of standard GST, with a single application of the gate of interest in between. Here, our characterization circuits, as we see, are instead randomized ones. Our linear gate-set-tomography procedure is also different from that in Ref. [14] carrying the same name: In Ref. [14], the "linear" tag refers solely to the linear inversion employed in the estimation, with the linear structure achieved with special choices of GST-type circuits, without the use of the linear-noise approximation as we do here.

D. Dependence relations

The design matrix $\mathbf{C}$ can be noninvertible for a few reasons. First, $\mathbf{C}$ is a nonsquare matrix, as we can have more equations, from the different choices of $\mathcal{C}^{(i)}$ and different μ values, than there are $\mathbf{e}$ parameters. This in itself is not a problem: We need only to include in $\mathbf{C}$ a linearly independent set of equations found, for example, by a row-reduction procedure on the original $\mathbf{C}$. That is not all, however. That $\mathbb{1} + e_\gamma$, $\mathbb{1} + e_{\text{in}}$, and $\mathbb{1} + e_{\text{ro},\mu}$ are all CP (and hence hermicity preserving) and TP maps translate to equality constraints on the e parameters, so that not all e parameters are independent in the first place. Furthermore, dependence relations between the columns of the design matrix—leading to nonunique solutions for $\mathbf{e}$—arise because of what is sometimes known as "gauge freedom," coupled with our linear regime approximation.

We can easily accommodate the CPTP conditions for n-qubit situations by working in the Pauli-operator basis: the set of Pauli operators together with the identity operator, $\{(1/\sqrt{2})\sigma_a\}_{a=0}^3 = (1/\sqrt{2})\{\mathbb{1}, \sigma_1 = \sigma_x, \sigma_2 = \sigma_y, \sigma_3 = \sigma_z\}$ for a single qubit, and the tensor products $\{[1/(\sqrt{2})^n]\sigma_{a_1} \otimes \sigma_{a_2} \ldots \sigma_{a_n}\}$, for $a_1, a_2, \ldots, a_n = 0, 1, 2, 3$ for n qubits. We write an element of the n-qubit Pauli basis as $[1/(\sqrt{2})^n]\sigma_{\mathbf{a}}$ for $\mathbf{a} \equiv (a_1, a_2, \ldots, a_n)$. Note that $\sigma_{\mathbf{0}}$ is always the element proportional to the identity operator, and is the only basis element with nonvanishing trace.

Now, a TP map cannot take a traceless Pauli operator to the identity with nonvanishing trace. Thus, in the Pauli basis, the TP condition requires $e_{\gamma;\mathbf{0a}} = 0$ for all $\mathbf{a}$ (recall, $\mathbb{1} + e_\gamma$, not e_γ itself, is TP). The CP condition is first a hermicity-preserving condition, which in the Pauli basis, translates into the statement that $e_{\gamma;\mathbf{ab}}$ are all real, not complex, numbers. That it is CP adds only inequality constraints that do not reduce the number of independent noise parameters. Thus, each gate error e_γ is determined by no more than $d^2(d^2 - 1)$ real parameters. For the initialization and readout error maps, we are concerned only with those matrix elements picked out by $|\rho_{\text{in}}\rangle\rangle$ and $\langle\langle P_\mu|$, namely, $e_{\text{in};\mathbf{a}} = \langle\langle \sigma_{\mathbf{a}}|e_{\text{in}}|\rho_{\text{in}}\rangle\rangle$ and $e_{\text{ro},\mu;\mathbf{a}} = \langle\langle P_\mu|e_{\text{ro},\mu}|\sigma_{\mathbf{a}}\rangle\rangle$. That $\mathbb{1} + e_{\text{in}}$ is TP means $e_{\text{in};\mathbf{0}} = \langle\langle \sigma_{\mathbf{0}}|e_{\text{in}}|\rho_{\text{in}}\rangle\rangle = (1/\sqrt{d})\text{tr}e_{\text{in}}(\rho_{\text{in}}) = 0$, for any ρ_{in}. We are thus left with $d^2 - 1$ real (since $\mathbb{1} + e_{\text{in}}$ is CP) parameters to specify the initialization error. For the readout, the TP condition does not impose any additional constraints, and the CP condition translates into a total of d^2 real parameters to specify each $e_{\text{ro},\mu}$, subject to the additional overall constraint $\sum_\mu \widetilde{P}_\mu = \mathbb{1}$.

In addition, we have to consider dependences amongst the columns of the design matrix due to gauge freedom. Gauge freedom in this context, following the terminology in earlier papers [13–15,27], refers to the freedom to attribute the observed noise to the gate imperfections versus the initialization and readout (or SPAM) errors. This stems from our ability to probe the physical device only by sending in a single default input state and getting the readout values for a single default observable; one cannot access the gate-only noise without involving also the initialization and readout errors. In our

linear-noise regime, this gauge freedom is expressible as the fact that we can access only the values of the sum $_\mu\langle\!\langle\mathcal{C}e_{\text{in}}\rangle\!\rangle + \sum_{k=1}^{L} {}_\mu\langle\!\langle\mathcal{G}_{L:k+1}\mathcal{E}_k\mathcal{G}_{k:1}\rangle\!\rangle + {}_\mu\langle\!\langle e_{\text{ro},\mu}\mathcal{C}\rangle\!\rangle$, not the values of the individual terms. Assuming that we have removed the redundancy in **C** arising from the CPTP constraints as well as those from linear dependences in the circuit choices, the presence of this gauge freedom can be phrased, in our linear regime, as the nontrivial null space of **C**. A nontrivial null space translates into the fact that any two solutions **e** and **e′** of Eq. (6) differ by a vector $\boldsymbol{\alpha}$ in the null space. The null space of **C** can easily be found numerically by performing a singular-value decomposition on **C**.

The addition of a null-space vector $\boldsymbol{\alpha}$ to **e** to obtain **e′** can be expressed as a transformation on the original e noise maps. A particular transformation that generates a gauge freedom in our linear-noise regime is

$$e_\gamma \to e_\gamma + \mathcal{Q} - g_\gamma\mathcal{Q}g_\gamma^\dagger,$$
$$e_{\text{in}}|\rho_{\text{in}}\rangle\!\rangle \to e_{\text{in}}|\rho_{\text{in}}\rangle\!\rangle + \mathcal{Q}|\rho_{\text{in}}\rangle\!\rangle, \qquad (7)$$
and $\qquad \langle\!\langle P_\mu|e_{\text{ro},\mu} \to \langle\!\langle P_\mu|e_{\text{ro},\mu} - \langle\!\langle P_\mu|\mathcal{Q},$

for any γ-independent linear map $\mathcal{Q}$. This is a gauge transformation in that it preserves the value of $_\mu\langle\!\langle\mathcal{C}e_{\text{in}}\rangle\!\rangle + \sum_{k=1}^{L} {}_\mu\langle\!\langle\mathcal{G}_{L:k+1}\mathcal{E}_k\mathcal{G}_{k:1}\rangle\!\rangle + {}_\mu\langle\!\langle e_{\text{ro},\mu}\mathcal{C}\rangle\!\rangle$ for all μ, a statement that can be checked explicitly. Every choice of $\mathcal{Q}$ leads to an equivalent set of noise maps, equivalent in terms of their effects on the circuit output within the linear-noise regime. While we cannot, at the moment, show that the form of Eq. (7) fully captures the gauge freedom allowed within our linear-noise regime, the full null spaces in our illustrative examples below can all be described in this manner; see Appendix A.

One should compare our gauge transformation Eq. (7) to the analogous transformation found in the GST papers. In GST, the appropriate gauge transformation is given by [13–15]

$$\widetilde{g}_\gamma \to M\widetilde{g}_\gamma M^{-1},$$
$$|\widetilde{\rho}_{\text{in}}\rangle\!\rangle \to M|\widetilde{\rho}_{\text{in}}\rangle\!\rangle, \qquad (8)$$
and $\qquad \langle\!\langle\widetilde{P}_\mu| \to \langle\!\langle\widetilde{P}_\mu|M^{-1},$

for any invertible M. To compare with Eq. (7), we first note that the first line of Eq. (7) can be rewritten in terms of $\widetilde{g}_\gamma$ as $\widetilde{g}_\gamma \to \widetilde{g}_\gamma + \mathcal{Q}g_\gamma - g_\gamma\mathcal{Q}$. Now, we suppose our $\mathcal{Q}$ can be regarded as a small quantity, and that $M = e^\mathcal{Q} \simeq \mathbb{1} + \mathcal{Q}$. Then, the first line of Eq. (8) can be written as $\widetilde{g}_\gamma \to M\widetilde{g}_\gamma M^{-1} \simeq (\mathbb{1} + \mathcal{Q})(g_\gamma + e_\gamma g_\gamma)(\mathbb{1} - \mathcal{Q}) \simeq \widetilde{g}_\gamma + \mathcal{Q}g_\gamma - g_\gamma\mathcal{Q}$, dropping terms second-order in small quantities ($\mathcal{Q}$ and e_γ), yielding our transformation Eq. (7) for $\widetilde{g}_\gamma$. Similar relations for the transformations on $|\widetilde{\rho}_{\text{in}}\rangle\!\rangle$ and $\langle\!\langle\widetilde{P}_\mu|$ hold as well.

E. The RL-GST procedure

Finally, the procedure for our RL-GST approach is as follows:

1. Choose a set of N_c tomographic circuits $\{\mathcal{C}^{(i)}\}$, with N_c larger than the number of noise parameters. Each circuit $\mathcal{C}^{(i)}$ is built from a uniform random sequence of gates from the gate set $\mathfrak{g}$.
2. Compute **p** from the circuits. In addition, perform the circuits on the actual device, and obtain an estimate of $\widetilde{\mathbf{p}}$ from the gathered data.
3. Calculate the design matrix **C** from the circuit set, for a choice of operator basis. We can elect to use the Pauli basis, for ease in incorporating the CPTP conditions on our noise parameters.
4. Perform singular-value decomposition of $\mathbf{C} = \mathbf{VDW}^\dagger$, where **D** is the diagonal (generally nonsquare) matrix of singular values λ_js, and **V** and **W** are unitary matrices. The columns of **W** corresponding to the zero singular values span the null space of **C**.
5. We construct the (pseudo)inverse of **C** as $\mathbf{C}^{-1} = \mathbf{WD}^{-1}\mathbf{V}^\dagger$ where $\mathbf{D}^{-1} \equiv \text{diag}\{\lambda_1', \lambda_2', \ldots\}$ with $\lambda' \equiv 1/\lambda_j$ if $\lambda_j \neq 0$, and 0 otherwise.
6. Then, an estimate $\widehat{\mathbf{e}}$ for the noise parameters **e** is obtained by direct inversion,

$$\widehat{\mathbf{e}} = \mathbf{C}^{-1}(\widetilde{\mathbf{p}} - \mathbf{p}). \qquad (9)$$

Here, as we are looking for a quick estimate of the noise, and in the tradition of linear-inversion estimators, we do not impose any CP constraints on our estimator. The $\mathbb{1} + \widehat{e}_\gamma$ noise maps obtained from $\widehat{\mathbf{e}}$ will thus generally not be CPTP. If the CPTP nature is crucial, one can further apply a standard procedure to project the obtained estimator to the nearest CPTP map, for example, by the procedure of Ref. [28].

As an added step, one could perform a singular value decomposition of the design matrix before carrying out the tomographic circuits. This gives the number of zero singular values in the design matrix, yielding useful information about our choice of tomographic circuits. In particular, if we find that to be larger than $(d^2 - 1)d^2$, the number of singular values caused by Eq. (7) (see Appendix A), then we should figure out if these extra zero singular values arise from an unlucky sampling of tomographic circuits or from some other kinds of gauge freedom beyond Eq. (7) by sampling more circuits or adding some short circuits (see examples below, where we add "null" circuits).

Note that each circuit $\mathcal{C}^{(i)}$ should have length $L^{(i)} \ll 1/\epsilon$, for the linear regime to hold. ϵ can be estimated, *a priori*, using quick partial characterization methods like randomized benchmarking. $L^{(i)}$ should also generally not be too short—recalling the structure of quantum process tomography, where one should use a variety of input states

and measure a variety of observables on the output, the circuit $\mathcal{C}^{(i)}$ should have enough gates to give enough variety to sense all aspects of the error maps. As we see in our illustrative examples below, the accuracy of the RL-GST procedure is not sensitive to the specific choice of $L^{(i)}$s, and in fact, one can use a variety of values for $L^{(i)}$ within the same circuit set. Furthermore, the circuit set can be preassessed, before doing the experiments, for its effectiveness in sensing the noise parameters: A good choice should have good linear independence between the rows of $\mathcal{C}$, a notion that can be quantified in many ways. We have observed, however, that our randomized gate choice generally yields circuit sets with good linear independence, and have not found it necessary to do such a preassessment in our examples. For further discussion on the accuracy scaling with circuit length, see Appendix B.

Armed with $\widehat{\mathbf{e}}$, we can now make predictions from the device implementations of further "test" circuits, and expect accurate estimates of the outputs of the noisy device, namely, the quantities $p_\mu(\mathcal{C}) \equiv \langle\!\langle \widetilde{P}_\mu | \widetilde{\mathcal{C}} | \widetilde{\rho}_{\text{in}} \rangle\!\rangle$ for test circuit $\mathcal{C}$, as long as the length of the test circuit satisfies $L\epsilon \ll 1$. In our examples below, we measure the prediction error using the statistical distance,

$$\Delta \equiv \frac{1}{2} \sum_\mu |\widehat{p}_\mu - p_\mu|, \qquad (10)$$

where $\widehat{p}_\mu$ is computed from our estimate $\widehat{\mathbf{e}}$, while p_μ is the true value.

Because of the gauge freedom, we cannot say that our estimate $\widehat{\mathbf{e}}$, one representative choice amongst those that differ by a null-space vector, gives the true values of the actual noise parameters. Thus, $\widehat{\mathbf{e}}$ cannot be used to directly infer quantities like the strength of the noise, or the true noise maps, beyond this sandwich structure of $\langle\!\langle P_\mu | \cdot | \rho_{\text{in}} \rangle\!\rangle$. However, if the null-space equivalence can be fully captured by the gauge transformation Eq. (7) (as is the case in our examples below), then we can use $\widehat{\mathbf{e}}$ to estimate a common measure for judging the quality of gates: the average gate-set infidelity (AGSI), defined as

$$\text{AGSI}(\mathfrak{g}) \equiv \frac{1}{\Gamma} \sum_{\gamma=1}^{\Gamma} [1 - \overline{F}(g_\gamma, \widetilde{g}_\gamma)], \qquad (11)$$

with

$$\overline{F}(g,g') \equiv \int d\psi \, \langle\psi|(g^\dagger \circ g')(\psi)|\psi\rangle = \frac{\text{tr} g^\dagger g' + d}{d(d+1)}, \qquad (12)$$

for $d\psi$ taken as the Haar measure. The final expression for $\overline{F}$ is a Haar-average formula familiar in many contexts (e.g., see Refs. [29,30]). $\overline{F}(g_\gamma, \widetilde{g}_\gamma)$ is invariant under the gauge transformation Eq. (7), i.e., $\text{tr} g_\gamma^\dagger (\mathbb{1} + e_\gamma) g_\gamma =$

$\text{tr} g_\gamma^\dagger (\mathbb{1} + e'_\gamma) g_\gamma$ for $e'_\gamma = e_\gamma + \mathcal{Q} - g_\gamma \mathcal{Q} g_\gamma^\dagger$, since $\mathcal{Q} - g_\gamma \mathcal{Q} g_\gamma^\dagger$ is traceless. In our examples below, we thus use, apart from the prediction errors, the AGSI as another figure of merit for judging the performance of our RL-GST procedure. Note that the AGSI is *not* invariant under the GST gauge transformation Eq. (8), as previously noted in Ref. [27].

III. RESULTS

To illustrate our method, we carry out simulated experiments for single- and two-qubit gate sets, as well as use our RL-GST procedure to characterize three five-qubit IBM quantum-computing devices available for free access. We compare the performance of our method against that of standard GST. For implementing the GST protocol, we make use of the PYTHON package "pygsti" from Ref. [31].

A. Single-qubit simulations

To assess the performance of our RL-GST procedure, it is important to begin with a situation where we know the true noise in the gates. We simulate noisy single-qubit gate operations by attaching to an ideal gate a noise map of the form $e(\mathbf{e}) \equiv \text{AD}(e_0) \circ \text{Pauli}(e_1, e_2, e_3) \circ R_x(e_4) \circ R_y(e_5) \circ R_z(e_6)$. Here, $\text{AD}(e_0)$ is the amplitude-damping channel $\text{AD}(e_0)(\cdot) = (|0\rangle\langle 0| + \sqrt{1-e_0}|1\rangle\langle 1|)(\cdot)(|0\rangle\langle 0| + \sqrt{1-e_0}|1\rangle\langle 1|) + e_0|0\rangle\langle 1|(\cdot)|1\rangle\langle 0|$, $\text{Pauli}(e_1, e_2, e_3)$ is the Pauli channel $\text{Pauli}(e_1, e_2, e_3)(\cdot) = [1 - (e_1 + e_2 + e_3)](\cdot) + \sum_{a=1}^{3} e_a \sigma_a(\cdot)\sigma_a$, and $R_{X(Y,Z)}(e_{4(5,6)})$ is the rotation about the $x(y,z)$ axis of the Bloch sphere by angle $e_{4(5,6)}$. For each gate in the gate set, the noise parameters $\mathbf{e}$ are randomly chosen, but they remain fixed for the entire tomography procedure. In addition, we set $\widetilde{\rho}_{\text{in}} \equiv \widetilde{\rho}_{\text{in}}^{(1)} = |0\rangle\langle 0| + (a/\sqrt{2})(X + Y - Z)$, and $\widetilde{P}_0$ is taken to be identical to $\widetilde{\rho}_{\text{in}}$. The a is set as 0.01 in our examples below. Data for the tomography procedure are generated using $N_s = 8192$ shots (or repetitions) per circuit.

Incomplete gate set. We first discuss the case of a tomographically incomplete—or just incomplete for short—gate set, by which we mean a gate set unable to produce a tomographically complete set of states starting with the default $|0\rangle$ input. A *computationally* complete (or universal) gate set is always tomographically complete, but a tomographically complete gate set need not be computationally complete. An example of the latter is the single-qubit gate set $\{\mathbb{1}, X_{\pi/2}, Y_{\pi/2}\}$, where $X_{\pi/2}(Y_{\pi/2})$ is a $\pi/2$ rotation about the $x(y)$ axis of the Bloch sphere. An incomplete gate set means we are unable to sense all aspects of the gate noise (up to the gauge transformation discussed earlier), but we seek only to predict the outcomes of circuits built from that same gate set, for which the inaccessible noise parameters play no role.

An often-encountered incomplete gate set is the set of single-qubit Pauli gates $\{X \equiv \sigma_x, Y \equiv \sigma_y, Z \equiv \sigma_z\}$. For each circuit length $L^{(i)} = 8, 16, 32, 64$, and 128, we generate uniform randomly, from the Pauli gate set, a set of $N_c = 150$ circuits $\{C^{(i)}\}$. Duplicate circuits are discarded, and a null circuit, i.e., one with $L^{(i)} = 0$, is always added as we find it prevents accidental generation of an extra zero singular value in the design matrix. We obtain the estimate $\widehat{\mathbf{e}}$ for the noise parameters as described in the previous section, and use it to make predictions for a further 1000 test circuits, chosen uniform randomly to have length from the list $\{10, 50, 100, 200, 500, 1000\}$ and generated uniform randomly from the Pauli gate set.

Figure 1 shows the prediction error Δ for three different "true" noise strengths, as quantified by the AGSIs computed from the true $\mathbf{e}$ parameters. In all three cases, the prediction errors are small. For the two sets of noise parameters with smaller AGSI values, they grow as the length of the test circuit grows, to be expected not just because of an accumulation of estimation errors, but also because we approach the boundary of $L\epsilon \sim 1$ for which our linear approximation fails. The prediction error for the set of parameters with an AGSI value of 1.1×10^{-3} has very subtle details as shown by the inset of Figure 1. This unusual feature disappears for the parameters with smaller noise strengths (the threshold of AGSI is almost 2×10^{-4}). Right now we do not have a satisfying explanation for this feature, but it does not detract from the fact that the prediction error by our method is very small. We leave this as a future investigation. For this case of incomplete gate set, we show no comparison with standard GST, which does not accommodate such a situation.

Complete gate set. As a second example, we consider characterization of tomographically complete gate sets. The first gate set we consider is $\{\mathbb{1}, g_x \equiv X_{\pi/2}, g_y \equiv Y_{\pi/2}\}$. Here, we explore different ways of choosing the $L^{(i)}$s for our randomized circuits, and also compare with the GST results. For a fair comparison, we need to fix the total resource cost of each option. We opt for the total number of operations—gates, initializations, and readouts—as our measure of resource cost. Since GST is the most restrictive in terms of its requirements on the circuit structure, we use it to set the resource cost. As we explain below, the GST circuits we run for our simulated example require a total of about 48 150 operations. We take this also as the resource target for the variants of RL-GST explored here.

We examine four possibilities for the choices of $L^{(i)}$s in RL-GST: (1) fixed $L^{(i)} = 32$; (2) fixed $L^{(i)} = 64$; (3) fixed $L^{(i)} = 128$; (4) choose $L^{(i)}$ uniform randomly from the list $\{8, 16, 32, 64, 128\}$. To attain the resource cost of about 48 150 operations, we choose 1416, 730, 370, and 187 circuits, respectively, for the four options, together with a null circuit for each case. Altogether, these choices correspond to 48 146, 48 182, 48 102, and 48 248 (expected) operations, respectively. The actual number of circuits and
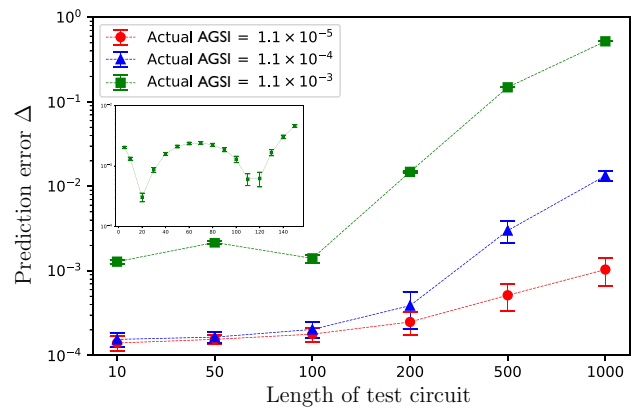

FIG. 1. Prediction errors for the incomplete gate set $\{X, Y, Z\}$ using simulated experiments. Three sets of true noise parameters, corresponding to AGSI values of 1.1×10^{-5}, 1.1×10^{-4}, and 1.1×10^{-3}, are used. Tomographic circuits are generated uniform randomly from the gate set with varying lengths. For each set of noise parameters, the entire simulation is repeated 10 times to generate the error bars. (All error bars in this paper indicate one standard deviation.) The inset is a detailed plot for the set of noise parameters with AGSI=1.1×10^{-3}.

operations used in the simulations are often a little fewer, as we discard, without replacement, any duplicate circuits generated in the random drawing.

The GST comparison is done with help of the pygsti PYTHON package. In GST, each tomographic circuit has the structure $F'_j + \text{germ}^k + F_i$, where F_i and F'_j are fiducial gate sequences corresponding to tomographically complete state preparation and measurement (the i, j indices label which state and which measurement outcome, respectively, while the "germ" is some chosen short gate sequence repeated k times to amplify the effect of the noise). Here, we set $\{F_i\}$ and $\{F'_j\}$ to both be the set $\{\text{null}, g_x, g_y, g_x^2, g_x^3, g_y^3\}$. The set $\{F_i|0\rangle\}$ are then precisely the standard process-tomographic states, $\{|0\rangle, |-i\rangle, |+\rangle, |1\rangle, |i\rangle, |-\rangle\}$, where $|\pm\rangle \equiv (1/\sqrt{2})(|0\rangle + |1\rangle)$ and $|\pm i\rangle \equiv (1/\sqrt{2})(|0\rangle \pm i|1\rangle)$); the set $\{\langle 0|F'_j\}$ are then the adjoints of the standard six-state measurement (i.e., measure the Pauli X, Y, Z observables) outcomes. The germs we make use of are those suggested by the GST PYTHON package: $\mathbb{1}, g_x, g_y, g_y g_x$, and $g_y g_x g_x$; each germ is repeated for k values taking values $2^0, 2^1, \ldots$, and 2^7. The maximum value of 2^7 is chosen to ensure $L\epsilon \ll 1$ for our simulated noise strength, while the powers-of-2 increments are suggested by the PYTHON package. In addition, we also include circuits of the form $F'_j + F_i$ for different i, js, corresponding to taking a "null" germ. Duplicate circuits are removed. Altogether, we have 1120 circuits, involving a total of 48 149 operations.

As an added angle for exploring the importance of the germ design in GST, we retain the F_i and F'_j state preparation and measurement fiducial sequences in the GST

circuits, but replace the middle GST-prescribed germ^k by a uniform randomly chosen sequence, each of fixed length $L^{(i)} = 2^7$, of gates from the gate set. We use ten different random gate sequences altogether. We also include, as before, circuits with the structure $F'_j + F_i$. Altogether, we have 384 circuits (after eliminating duplicates), with a total of 48 133 operations.

Linear inversion, as described in our RL-GST procedure earlier, is used to obtain the estimators for all the variants of RL-GST differing in the choice of **C**. For the GST circuits, we compute the estimators from the optimization carried out by the GST PYTHON package, with or without the CP option (both estimators are TP). For added comparison, we use our linear inversion procedure

for the data obtained from the GST tomographic circuits (labeled L-GST in Fig. 2). The results are given in Figs. 2(a) and 2(b). Figure 2(a) plots the AGSI estimates for the gate set; Fig. 2(b) gives the prediction errors for 1000 test circuits (as in the incomplete gate-set example), randomly chosen from the gate set and of lengths $\{10, 50, 100, 200, 500, 1000\}$. The test circuits in Fig. 2(b) are for a single randomly chosen set of true noise parameters with AGSI $= 1.1 \times 10^{-4}$. From the plots, our four RL-GST variants, with different $L^{(i)}$ choices have comparable performance, though the shorter circuits have larger scatter in the AGSI predictions. The one with varying $L^{(i)}$ lengths has the smallest prediction error, but the difference is small, suggesting a robustness against the specific choice
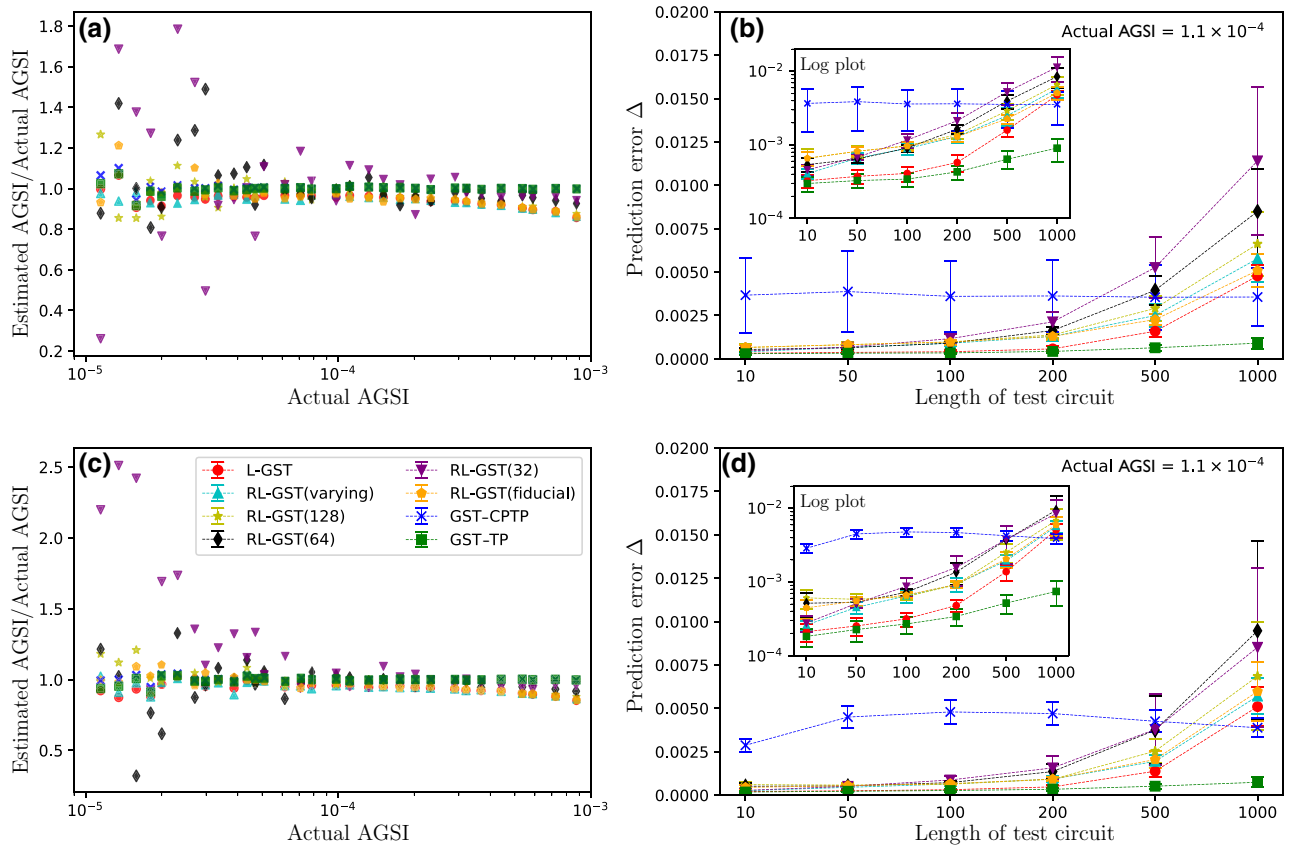

FIG. 2. Simulated experiments for single-qubit gate sets: plots (a),(b) are for $\{I, X_{\pi/2}, Y_{\pi/2}\}$; plots (c),(d) are for $\{I, T, H\}$. The legend is for all four plots. (a),(c) Ratios of the estimated AGSI [Eq. (11)] to the actual AGSI computed from the true parameters are shown for our RL-GST procedure, with different choices of tomographic circuit lengths (32, 64, 128, and varying). In addition, we show the ratios for L-GST, referring to the use of the GST tomographic circuits, and for RL-GST (fiducial), referring to tomographic circuits with the fiducial F_i, F'_j segments but random gates in between. Error bars are generated by repeating simulation for each set of noise parameters 10 times; the sizes of the resulting error bars are similar for all the data. For plot (a), the average—over the different sets of noise parameters—error bars are $\{2.0, 2.3, 7.0, 13.1, 25.1, 4.2, 1.9, 1.9\} \times 10^{-6}$, listed in the order of appearance in the legend (L-GST and RL-GST variants); for plot (c), the average error bars are $\{2.2, 3.1, 7.1, 14.0, 26.5, 4.4, 1.7, 1.7\} \times 10^{-6}$. (b),(d) Prediction errors for the different variants of RL-GST, L-GST, as well as GST procedures are shown, for true noise parameters with actual AGSI $= 1.1 \times 10^{-4}$. The error bars are generated by repeating the simulation for each set of noise parameters 10 times. The inset shows the same plot on a logarithmic vertical scale. Note that the AGSI is invariant for both RL-GST and GST under the weak-noise gauge transformation of Eq. (7).

of $L^{(i)}$. All four do somewhat worse than GST, but one should note the much lower time cost for nearly the same performance: For each run of the protocol (getting the estimate $\widehat{\mathbf{e}}$ from one set of tomographic circuits), RL-GST tomography procedure took under 1 s to complete on a regular laptop. The GST procedure, either the TP or the CPTP options provided by the PYTHON package, took about 30 s on the same laptop for the GST-prescribed circuit set. The time-cost advantage for RL-GST becomes even more pronounced with the multiqubit examples below.

Figures 2(a) and 2(b) also offer lessons about GST. Observe that the variant with the F_i, F'_j structure, but now sandwiching a random sequence, does similarly as our RL-GST options, suggesting that a proper choice of germ is important for GST to work well. Also notable in Fig. 2(b) is the close performance of the linear inversion estimator to that of the GST-optimized estimator, with prediction errors differing more only for longer circuits, a direct consequence of our linear approximation. As the optimization to reconstruct the estimator is the time-expensive component of GST, this suggests that one could opt for a simple linear inversion while retaining the structure of the GST tomographic circuits.

We repeat the entire simulation experiment for a different single-qubit gate set, $\{\mathbb{1}, H, T\}$, where H is the Hadamard gate $H \equiv |+\rangle\langle 0| + |-\rangle\langle 1|$, and $T \equiv |0\rangle\langle 0| + e^{i\pi/4}|1\rangle\langle 1|$. In this case, the GST fiducial sequences are chosen as $F_i \in \{I, H, HT^4H, T^2H, T^4H, T^6H\}$ and $F'_j \in \{I, H, HT^4H, HT^6, HT^4, HT^2\}$. Both again generate the six-state preparation and measurement as used for the $\{\mathbb{1}, g_x, g_y\}$ gate set, but the preparation-and-measurement sequences are different here since $T^\dagger = T^7 \neq T$. The germs, as prescribed by the GST PYTHON package, are I, $T, H, HT^3, TI^4, T^2HI^3, HTHI^3, H^4T, THT^3I$, and HTH^2T^2. As before, we pick our RL-GST random circuits to match the number of operations used in the GST-prescribed set. The results are given in Figs. 2(c) and 2(d).

From the two gate sets, we draw the conclusion that RL-GST offers a quick estimate of the noise, providing comparable—though slightly worse—results as GST, but in a much shorter time. One also need not be too bothered with the choice of tomographic circuits as randomly chosen circuit sequences with different $L^{(i)}$ values work nearly as well as the carefully crafted ones from GST.

B. Two-qubit simulations

Next, we simulate a two-qubit system with crosstalk between the two qubits. To each two-qubit gate, we attach a noise map of the form $e(\mathbf{e}) \equiv [\mathrm{AD}(e_0) \otimes \mathrm{AD}(e_1)] \circ \mathrm{Pauli2}(e_2, e_3, \ldots, e_{16}) \circ \{[R_x(e_{17}) \circ R_y(e_{18}) \circ R_z(e_{19})] \otimes [R_x(e_{20}) \circ R_y(e_{21}) \circ R_z(e_{22})]\}$. Here, AD, R_x, R_y, and R_z are as in the single-qubit situation, while Pauli2 is the two-qubit Pauli map—the source of the two-qubit crosstalk—$\mathrm{Pauli2}(e_2, \ldots, e_{16})(\cdot) = \sum_{i,j=0,x,y,z} q_{ij}\sigma_i \otimes \sigma_j$

$(\cdot)\sigma_i \otimes \sigma_j$, with $q_{0x} = e_2, q_{0y} = e_3, q_{0z} = e_4, q_{x0} = e_5, q_{xx} = e_6$, etc., and q_{00} is determined by $\sum_{i,j=0,x,y,z} q_{ij} = 1$. The noisy input state is the product state $\widetilde{\rho}_{\mathrm{in}} = \widetilde{\rho}_{\mathrm{in}}^{(1)} \otimes \widetilde{\rho}_{\mathrm{in}}^{(1)}$ as used in the single-qubit simulation; the noisy Z measurement on each qubit (no crosstalk) is assumed to comprise the outcomes (positive operator-valued measure) operators $\widetilde{P}_0^{(1)}$ and $\mathbb{1} - \widetilde{P}_0^{(1)}$, where $\widetilde{P}_0^{(1)} = \widetilde{\rho}_0^{(1)}$.

The gates on the two qubits are assumed to be $\mathbb{1}, g_x = X_{\pi/2}, g_y = Y_{\pi/2}$ for each of the qubits, and a two-qubit CNOT gate with one of the qubits (qubit 1) as the control, the other (qubit 2) as the target. With the expectation of crosstalk between the qubits, the gate set we are to characterize comprises all possible two-qubit combinations of the gates, namely, the CNOT together with $\{g_{ij}\}_{i,j=0,x,y}$, where $g_{ij} \equiv g_i \otimes g_j$ with $i, j = 0, x, y$ and $g_0 \equiv \mathbb{1}$.

As in the single-qubit simulations, we impose a common
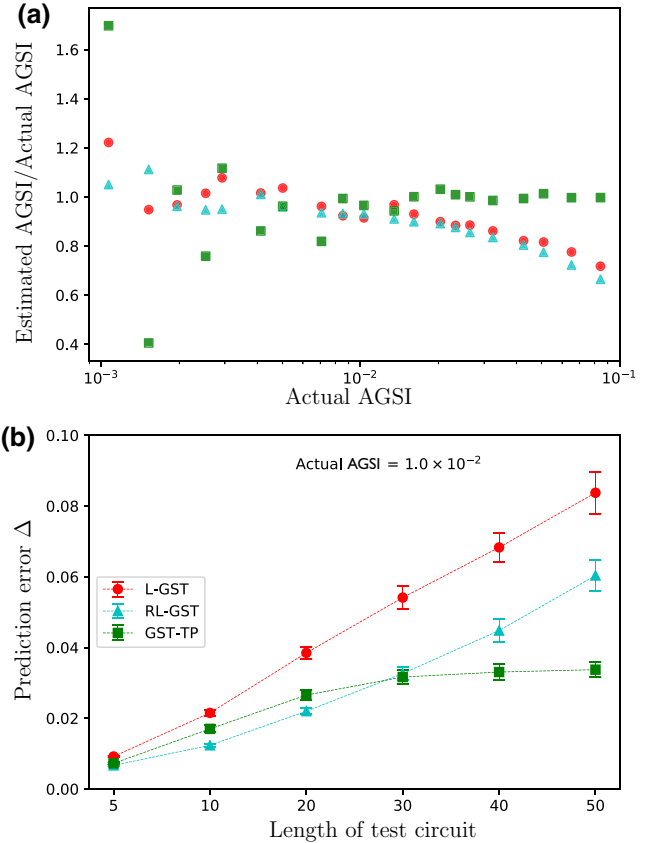

(a)

FIG. 3. Simulation results for a two-qubit gate set. (a) The ratios of the estimated AGSI [Eq. (11)] to the actual AGSI computed from the true parameters are shown for the RL-GST, L-GST, and GST-TP procedures, with error bars of average sizes 2.6×10^{-4}, 7.5×10^{-4}, and 6.7×10^{-4}, respectively. (b) The prediction errors for RL-GST, L-GST, and GST procedures are shown, for true noise parameters with actual AGSI = 1.1×10^{-2}. The error bars are generated by repeating the simulation for each set of noise parameters 10 times. The legend is for both plots (a),(b).
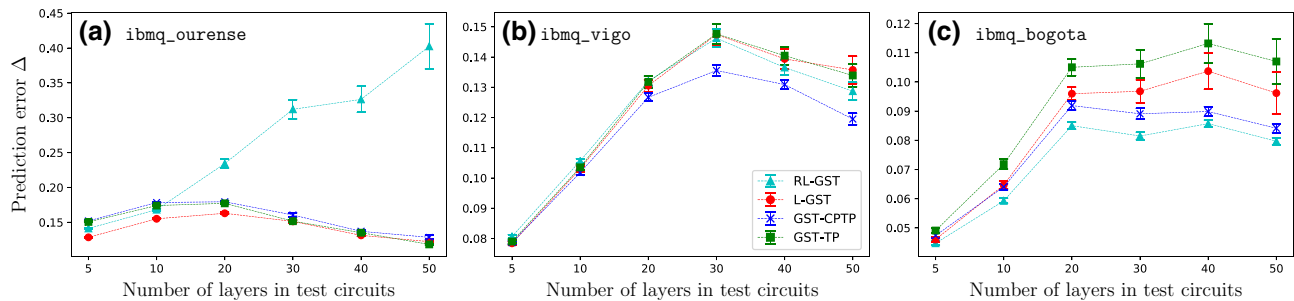
FIG. 4. Prediction errors of five-qubit test circuits on three IBM devices for the different estimation procedures. All experiments were done on 10 September 2020. The error bars indicate our uncertainty in the estimates $\widehat{\epsilon}$, generated by bootstrapping (100 samples) the tomographic experimental data. For `ibmq_vigo` and `ibmq_bogota`, all estimation procedures give similar prediction errors, with RL-GST taking the least computational time and design effort. For `ibmq_ourense`, all methods show worse prediction errors, with RL-GST performing more poorly than other procedures, a likely reason being the time drift in the noise of the device. The goodness-of-fit quantities (see Table I below) for ibmq_ourense are large for all procedures, so the model fit is poor anyway, and one cannot draw reliable conclusions about the performance of the schemes; see main text for further discussion.

resource cost, set by the GST procedure. Because running the full GST algorithm for the two-qubit situation turns out to be very time consuming, we use simpler circuits here. The fiducial sequences $\{F_i\}$ and $\{F_j'\}$ are taken from the set $\{\text{null}, g_{0x}, g_{0y}, g_{0x}g_{0x}, g_{x0}, g_{xx}, g_{xy}, g_{0x}g_{xx}, g_{y0}, g_{yx}, g_{yy}, \quad g_{0x}$ $g_{yx}, g_{x0}g_{x0}, g_{x0}g_{xx}, g_{x0}g_{xy}, g_{xx}g_{xx}\}$. Ideal versions of the fiducial sequences map the input $|0\rangle$ state to the 16 states $\{|0\rangle, |-i\rangle, |+\rangle, |1\rangle\}^{\otimes 2}$. Furthermore, we choose the germs for the GST circuit as individual gates in the gate set, and each germ is applied only once ($k = 1$). There are 2336 GST-type circuits generated this way, with a total of 13 612 operations. For our RL-GST procedure (no F_i–F_j' structure), we choose $\lceil 13612/(8+2)\rceil = 1362$ circuits, with eight gates in each circuit; the null sequence is again added in as an additional circuit.

Figure 3 shows the results for the two-qubit situation. For GST, for shorter runtime, we ran only the case with TP (not necessarily CP) constraints, more comparable to our linear-inversion estimator. One single GST run took about 35 min to complete, versus just 12 s for RL-GST, which achieved similar estimation accuracy as is apparent in Fig. 3. Again, we use the data from the GST circuits to compute the linear-inversion estimator (L-GST); the results are also given in Fig. 3, this time doing more poorly than our RL-GST and the original GST estimator.

C. Characterization of IBM quantum devices

Next, we apply our method to characterize the noise in three of the five-qubit devices publicly available on the IBM Quantum Experience platform [23]. The IBM devices make use of primitive gates $g_x = X_{\pi/2}$ and $g_y = Y_{\pi/2}$ to accomplish a variety of Clifford gate operations. We hence consider a gate set comprising the single-qubit gates $\{\mathbb{1}, g_x, g_y\}$ for each of the five qubits in the device, together with the two-qubit CNOT gates specific to each

device. The default (ideal) input state is $|0\rangle$ for each qubit; Z is measured on each qubit at the end of every circuit.

For a five-qubit device, the most general approach would be to assign a five-qubit noise map for every single gate. For example, the ideal single-qubit g_x gate applied on qubit 0 should be thought of as $g_x \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1} \otimes \mathbb{1}$ with a five-qubit noise map to be characterized. In principle, our RL-GST approach can accommodate such a scenario, as can the GST procedure. The sheer combinatorial variety of gates, and hence the number of noise parameters, that one would have to handle, however, are forbidding, and likely ultimately not useful.

Instead, we opt for a more practical route: First characterize the noise in the single-qubit gates individually, as if they are done on separate single qubits, and then use these single-qubit gates, now with known (estimated) noise, to characterize the noise in the two-qubit gates. This assumes that the crosstalk between qubits in the device is unimportant, which may or may not hold for the device in question. We discuss goodness-of-fit measures at the end to gauge the suitability of our assumed model.

We characterize the noise in three IBM five-qubit devices: `ibmq_ourense` [32], `ibmq_vigo` [33], and `ibmq_bogota` [34]. To characterize the single-qubit gates, we make use of both our randomized tomographic circuits as well as the GST-type circuits. For the GST circuits, we follow the procedure as in our single-qubit simulation for the same $\{\mathbb{1}, g_x, g_y\}$ gate set described earlier, having the structure $F_j' + \text{germ}^k + F_i$, and the fiducial sequences are chosen as before (the six-state tomographic states and measurement). Because of the time drift in the IBM devices, we find it necessary to use tomographic circuits that are not too long. Consequently, the k values are restricted to $k = 1, 2,$ and 4 only. This gives 285 tomographic circuits $\{C^{(i)}\}$. For RL-GST, to maintain the same resource cost as for GST, we randomly choose 83 circuits

TABLE I. Data for single- and two-qubit gates in the IBM devices (`ibmq_ourense`, `ibmq_vigo`, and `ibmq_bogota`). All experiments were performed on 10 September 2020 and the IBM AGSI values were from the calibration cycle closest to the experiment date. For single-qubit gates, the estimated AGSI values are computed from the RL-GST, L-GST, GST-CPTP, GST-TP procedures. Negative infidelity values can arise because we do not impose CP constraints in the estimation. Each single-pulse single-qubit gate takes about 36 ns to complete. For two-qubit gates, cxa_b refers to a CNOT gate between qubits a and b, with qubit a as the control, and qubit b as the target. The RL-GST, L-GST, GST-CPTP, and GST-TP headings refer to the procedure used for estimating the noise parameters for the *single*-qubit gates. In every case, the two-qubit gate-noise parameters are estimated using standard linear inversion. Negative infidelity values for the two-qubit gates can thus arise.

Single-qubit gates

Qubit		AGSI ($\times 10^{-4}$)					Goodness of fit n_σ			
		IBM	RL-GST	L-GST	GST-CPTP	GST-TP	RL-GST	L-GST	GST-CPTP	GST-TP
ibmq_ourense	Q_0	7.0	14.3 ± 0.9	18.3 ± 1.3	18.0 ± 1.1	17.5 ± 1.3	23 ± 3	30 ± 3	30 ± 4	28 ± 3
	Q_1	4.1	13.1 ± 0.9	22.1 ± 1.2	21.5 ± 1.0	20.8 ± 1.2	32 ± 3	35 ± 3	308 ± 63	24 ± 3
	Q_2	1.7	2.9 ± 0.6	-0.03 ± 1.0	2.8 ± 0.4	-0.1 ± 0.9	6 ± 2	4 ± 2	13 ± 20	2 ± 2
	Q_3	4.5	8.4 ± 1.2	126.3 ± 2.3	207.9 ± 1.8	178.2 ± 2.6	1040 ± 20	1582 ± 18	1630 ± 41	1215 ± 15
	Q_4	10.8	20.8 ± 1.4	59.8 ± 2.4	183.2 ± 2.4	170.1 ± 3.6	843 ± 14	1885 ± 22	1961 ± 128	1753 ± 21
ibmq_vigo	Q_0	3.5	1.5 ± 0.6	5.5 ± 1.1	6.8 ± 0.7	6.6 ± 0.9	46 ± 7	22 ± 4	17 ± 4	14 ± 3
	Q_1	4.8	12.1 ± 0.7	11.9 ± 1.0	12.6 ± 0.8	11.6 ± 1.0	78 ± 6	66 ± 6	47 ± 4	43 ± 4
	Q_2	3.9	7.0 ± 0.6	4.5 ± 1.1	5.7 ± 0.5	3.8 ± 0.8	28 ± 4	25 ± 4	21 ± 3	18 ± 2
	Q_3	4.8	7.6 ± 0.7	4.3 ± 1.0	9.1 ± 0.6	4.9 ± 0.9	45 ± 5	44 ± 3	45 ± 7	41 ± 3
	Q_4	7.2	7.4 ± 0.5	6.6 ± 0.9	7.4 ± 0.6	6.4 ± 0.8	30 ± 4	30 ± 4	33 ± 5	28 ± 3
ibmq_bogota	Q_0	2.8	3.3 ± 0.6	5.6 ± 1.1	7.0 ± 0.8	6.8 ± 0.8	4 ± 2	7 ± 2	13 ± 3	6 ± 2
	Q_1	1.9	7.3 ± 0.6	6.6 ± 1.1	8.8 ± 0.9	8.5 ± 1.0	28 ± 4	13 ± 3	13 ± 9	11 ± 2
	Q_2	1.8	3.8 ± 0.8	-1.1 ± 1.2	3.8 ± 0.6	-8.0 ± 1.6	6 ± 2	527 ± 16	524 ± 16	508 ± 13
	Q_3	2.9	6.5 ± 0.7	5.8 ± 1.0	6.7 ± 0.8	6.2 ± 0.9	7 ± 3	11 ± 3	26 ± 5	6 ± 2
	Q_4	2.6	5.3 ± 0.8	5.2 ± 1.3	6.9 ± 0.8	6.3 ± 0.9	9 ± 3	5 ± 3	10 ± 3	2 ± 2

Two-qubit gates

Gates		Gate time (ns)	Average gate infidelity ($\times 10^{-2}$)				
			IBM	RL-GST	L-GST	GST-CPTP	GST-TP
ibmq_ourense	cx0_1	270.2	1.11	-1.53 ± 0.20	-1.14 ± 0.15	-1.75 ± 0.17	-0.98 ± 0.16
	cx1_0	234.7	1.11	-0.68 ± 0.20	-0.29 ± 0.16	-1.00 ± 0.17	-0.39 ± 0.16
	cx1_2	391.1	0.74	-0.38 ± 0.21	-0.48 ± 0.16	-1.21 ± 0.23	-0.47 ± 0.17
	cx2_1	426.7	0.74	0.07 ± 0.21	-0.03 ± 0.14	-0.56 ± 0.23	0.23 ± 0.14
	cx1_3	611.6	1.28	-11.78 ± 0.25	-8.70 ± 0.18	-12.23 ± 0.28	-9.87 ± 0.21
	cx3_1	576.0	1.28	-11.95 ± 0.26	-8.87 ± 0.19	-12.68 ± 0.31	-9.87 ± 0.23
	cx3_4	305.8	0.92	-8.71 ± 0.27	-5.24 ± 0.20	0.60 ± 0.71	-0.37 ± 0.32
	cx4_3	270.2	0.92	-10.58 ± 0.29	-7.10 ± 0.21	-1.52 ± 0.97	-2.53 ± 0.35
ibmq_vigo	cx0_1	519.1	1.13	1.64 ± 0.19	1.38 ± 0.16	1.82 ± 0.17	1.56 ± 0.18
	cx1_0	554.7	1.13	1.23 ± 0.16	0.97 ± 0.15	1.38 ± 0.16	1.13 ± 0.16
	cx1_2	227.6	0.68	3.73 ± 0.21	2.92 ± 0.17	3.35 ± 0.17	3.25 ± 0.18
	cx2_1	263.1	0.68	3.89 ± 0.19	3.08 ± 0.17	3.45 ± 0.17	3.32 ± 0.17
	cx1_3	497.8	1.33	0.58 ± 0.19	0.13 ± 0.16	0.52 ± 0.19	0.21 ± 0.17
	cx3_1	462.2	1.33	0.88 ± 0.19	0.43 ± 0.16	0.82 ± 0.20	0.55 ± 0.18
	cx3_4	270.2	0.78	0.97 ± 0.19	0.87 ± 0.15	1.26 ± 0.17	0.98 ± 0.16
	cx4_3	305.8	0.78	-0.32 ± 0.17	-0.42 ± 0.15	-0.22 ± 0.18	-0.47 ± 0.16
ibmq_bogota	cx0_1	433.8	1.02	3.17 ± 0.17	2.65 ± 0.15	3.26 ± 0.19	3.10 ± 0.16
	cx1_0	398.2	1.02	2.23 ± 0.18	1.71 ± 0.16	2.16 ± 0.19	1.97 ± 0.17
	cx1_2	376.9	0.58	2.73 ± 0.22	0.27 ± 0.18	1.10 ± 0.23	0.29 ± 0.21
	cx2_1	412.4	0.58	1.02 ± 0.21	-1.44 ± 0.16	-0.75 ± 0.23	-1.58 ± 0.18
	cx2_3	625.8	0.91	1.84 ± 0.21	-0.12 ± 0.15	0.99 ± 0.20	-0.07 ± 0.17
	cx3_2	590.2	0.91	2.10 ± 0.20	0.14 ± 0.15	1.20 ± 0.19	0.17 ± 0.17
	cx3_4	369.8	0.96	1.16 ± 0.19	0.69 ± 0.16	1.18 ± 0.18	0.90 ± 0.17
	cx4_3	334.2	0.96	1.16 ± 0.20	0.69 ± 0.16	1.20 ± 0.18	1.00 ± 0.17

each for circuit lengths 8 and 16; a null circuit is again included, as before. The same tomographic circuit is run on each of the five qubits of the device, and submitted together to the IBM platform, to be run simultaneously. The number of shots per circuit is $N_s = 8192$.

Tomographic circuits for the two-qubit gates are constructed simply as $F'_j + \text{CNOT} + F_i$, where the F'_j and F_i are now the fiducial sequences used in our two-qubit simulations above. When doing the reconstruction of the noise parameters for the two-qubit gates, the noise parameters for the single qubit are set to be those estimated from the single-qubit characterization rounds. The two-qubit gate characterization thus looks exactly like what one would do in standard process tomography. For simplicity, we use linear inversion (respecting the TP, but not CP, constraint) to obtain an estimate of the two-qubit noise process.

Figure 4 shows the prediction errors for the three different devices, using our RL-GST as well as the GST procedures (TP and CPTP variants) for the single-qubit gate characterizations; we include also the data for L-GST, where our linear-regime analysis is applied to GST tomographic circuits. The test circuits are five-qubit circuits. Each test circuit is composed of multiple layers. Each layer is randomly chosen (50-50) to comprise either only single-qubit gates or single- and two-qubit gates. In a single-qubit-gate layer, a single-qubit gate is chosen uniform randomly from the gate set $\{\mathbb{1}, g_x, g_y\}$ for each qubit. In a layer chosen to have two-qubit gates, we first uniform randomly pick one of the CNOT gates to apply on the appropriate control and target qubits; the other qubits are each assigned a sequence of randomly chosen single-qubit gates, of a length that occupies the time taken for the CNOT gate in the same layer to be completed. The prediction errors are computed as in Eq. (10), with p_μs set as the relative frequencies obtained from the IBM device itself.

Now, the IBM devices are well known to suffer from time drifts in the gate noise [35–37], a situation not accommodated within our inference model. Crosstalk between qubits is another source of possible model deviation. Given these possible deviations to our assumed model, it is prudent to perform a goodness-of-fit test, to assess the reliability of our estimates for the single-qubit gates. A common approach is to use the χ^2 statistic, defined in this case as

$$\chi^2 = \sum_{i=1}^{N_c} \left(\frac{p_i - \widehat{p}_i}{\Sigma_i} \right)^2. \tag{13}$$

Here, N_c is the number of circuits, p_i and $\widehat{p}_i$, as before, are the measured and predicted (using $\widehat{\mathbf{e}}$) outcome probabilities of noise parameters, and Σ_i is the standard deviation of the measured p_is, approximated as $\Sigma_i = \left[(p_i(1 - p_i)/N_s) \right]^{1/2}$. The quantity χ^2 satisfies a χ^2 distribution with $f \equiv N_c - N_p$ degrees of freedom, where N_p is the number of

free (error) parameters. Following Ref. [15], we quantify goodness of fit by n_σ, defined as

$$n_\sigma \equiv \frac{\chi^2 - f}{\sqrt{2f}}. \tag{14}$$

Table I gives the data for the single- and two-qubit gates for the IBM devices: the estimated AGSI values from the RL-GST, GST, and the single-qubit goodness-of-fit measure n_σ. Comparing the goodness-of-fit numbers with the performance shown in Fig. 4, we see that prediction errors become large when the goodness-of-fit numbers exceed approximately 100. The fit is mostly good for ibmq_vigo and ibmq_bogota, and this is reflected in the lower prediction errors in Figs. 4(b) and 4(c), while the higher prediction errors for ibmq_ourense is mirrored by the large n_σ values in Table I. We suggest not proceeding with the two-qubit gate characterization unless the single-qubit n_σ values are sufficiently small.

The prediction errors for ibmq_vigo and ibmq_bogota are comparable across the RL-GST, L-GST, and GST procedures, with RL-GST taking the least computational time and design effort. RL-GST seems more susceptible to the issue of poor fit, as apparent in the plot for ibmq_ourense, but this could also be due to the fact that the tomographic data for the randomized circuits used in RL-GST were taken earlier in time, and hence with a longer time gap to the test-circuit runs, than the GST circuits. The noise in the test circuits may have drifted further from that captured in the RL-GST data than in the GST data. Another notable feature in Fig. 4 is the slight decrease in prediction errors as the number of layers (> 20) in the test circuits increases. We suspect this is due to the state of the quantum computer stabilizing towards an equilibrium value, regardless of the circuit applied, as observed empirically from the output measurement statistics, as the number of gate operations increases.

IV. CONCLUSIONS

In this work, we introduced the procedure of RL-GST, combining the idea of SPAM-error-free characterization of standard GST with the design ease brought about by our randomized circuit choice and computational speed from our linear-noise approximation. This gave us an easy-to-implement gate-set-tomography approach that can be done on the fly, without complicated tomographic circuit design or long computational wait time, but which yielded similar performance as standard GST, as evidenced by our illustrative examples. The assumption of the linear-noise regime is a very reasonable approximation for devices potentially capable of achieving useful quantum-information processing. We showed how our procedure can be used to characterize the real quantum-computing devices offered by the IBM Quantum Experience Platform, and the estimated noise parameters can be used as inputs via the

standard GST optimization approach, if further refinement is needed.

Looking ahead, one can imagine examining how a time-drift model can be accommodated within the linear-noise regime, for a better description of the actual noise in real devices, overcoming the issue of poor fit observed in the ibmq_ourense experiment. Crosstalk between the qubits can already be accounted for within our current model, but more work can be done to reduce the computational load brought about by the exponential growth in the number of noise parameters. An interesting direction currently under investigation within our group is to put in noise models built from elementary noise channels, each with a small number of noise parameters, but which together can describe a large variety of noise processes seen in real devices. We will report the results in due course.

ACKNOWLEDGMENTS

APPENDIX A: DEPENDENCE RELATIONS

Here, we provide further technical details on dependence relations, and in particular, discuss the class of gauge transformations expressed by Eq. (7). A dependence relation between the columns of the design matrix $\mathbf{C}$ can be expressed as the existence of a nonzero vector $\boldsymbol{\alpha}$ such that

$$\boldsymbol{\alpha} \cdot \mathbf{C}_\ell = 0, \tag{A1}$$

for any row $\mathbf{C}_\ell$ of $\mathbf{C}$. Any vector in the null space of $\mathbf{C}$ gives such a relation, and the dimension of the null space is equal to the number of independent dependence relations.

The transformation in Eq. (7) gives a class of dependence relations. We ask how many independent dependence relations can be generated by such a transformation, as determined by a map $\mathcal{Q}$. We first observe that the first row of $\mathcal{Q}$ is a zero vector for an informationally complete gate set. In particular, because of the trace-preserving constraints of each unitary map g_γ, the first row of $\mathcal{Q} -$

$g_\gamma \mathcal{Q} g_\gamma^\dagger$ should be zero vector, that is,

$$
\begin{aligned}
0 &= \mathcal{Q}_{0,j} - \sum_{m,k} g_{\gamma;0,m} \mathcal{Q}_{m,k} g_{\gamma;k,j}^\dagger \\
&= \mathcal{Q}_{0,j} - \sum_k \mathcal{Q}_{0,k} g_{\gamma;k,j}^\dagger \\
&= \left(\mathcal{Q} - \mathcal{Q} g_\gamma^\dagger \right)_{0,j} = \sum_k \mathcal{Q}_{0,k} \left(\mathcal{I} - g_\gamma^\dagger \right)_{k,j}
\end{aligned} \tag{A2}
$$

holds for any γ and j. The second line results from the fact that g_γ is a unitary map whose first row is the vector $(1, 0, 0, \dots)$ when written in the Pauli operator basis. We get a system of linear equations,

$$\left(\mathcal{I} - g_\gamma^\dagger \right)^T |\mathcal{Q}_0\rangle\!\rangle = 0, \tag{A3}$$

where $|\mathcal{Q}_0\rangle\!\rangle^T$ is the first row of $\mathcal{Q}$. All maps and vectors are now thought of as written in the Pauli operator basis. Because g_γ is a real matrix, we have,

$$g_\gamma |\mathcal{Q}_0\rangle\!\rangle = |\mathcal{Q}_0\rangle\!\rangle. \tag{A4}$$

Thus, $|\mathcal{Q}_0\rangle\!\rangle$ is a simultaneous eigenvector with eigenvalue 1 for all the gates g_γ.

For an informationally complete gate set, we show that $|\mathcal{Q}_0\rangle\!\rangle$ can only be $c|I\rangle\!\rangle = (c, 0, 0, 0, \dots)$ where c is some complex number. We assume that the gate set is a generating set for a group of unitary maps. The space spanned by $|I\rangle\!\rangle$ is an invariant subspace of the group. If there is another $|\mathcal{Q}_0\rangle\!\rangle$ that is orthogonal to $|I\rangle\!\rangle$ and satisfies Eq. (A4), then $|\mathcal{Q}_0\rangle\!\rangle$ spans another invariant space. Because finite-dimensional unitary representations of any group are completely reducible, the complement space of $|I\rangle\!\rangle \langle\!\langle I|$ and $|\mathcal{Q}_0\rangle\!\rangle \langle\!\langle \mathcal{Q}_0|$ is also an invariant subspace whose projector is denoted as $\overline{\mathcal{Q}}$. The input state ρ_{in} can be written as $|\rho_{\text{in}}\rangle\!\rangle = (1/\sqrt{d})|I\rangle\!\rangle + |\mathcal{Q}_0\rangle\!\rangle \langle\!\langle \mathcal{Q}_0|\rho_{\text{in}}\rangle\!\rangle + \overline{\mathcal{Q}}|\rho_{\text{in}}\rangle\!\rangle$. The states that can be prepared by the gate set thus have the form $g_\gamma|\rho_{\text{in}}\rangle\!\rangle = (1/\sqrt{d})|I\rangle\!\rangle + |\mathcal{Q}_0\rangle\!\rangle \langle\!\langle \mathcal{Q}_0|\rho_{\text{in}}\rangle\!\rangle + \overline{\mathcal{Q}}|\rho_\gamma\rangle\!\rangle$ where ρ_γ can be any state. However, this set of states cannot form an operator basis, as an operator $(1/\sqrt{d})|I\rangle\!\rangle - |\mathcal{Q}_0\rangle\!\rangle \langle\!\langle \mathcal{Q}_0|\rho_{\text{in}}\rangle\!\rangle$ cannot be written as a linear combination of this set of states. The gate set must hence be informationally incomplete, violating our initial assumption.

Because the input state ρ_{in} has trace 1, the first element of $\mathcal{Q}|\rho_{\text{in}}\rangle\!\rangle$ is zero. Then we get

$$\langle\!\langle \mathcal{Q}_0|\rho_{\text{in}}\rangle\!\rangle = \frac{1}{\sqrt{d}} c = 0 \quad \Rightarrow \quad c = 0. \tag{A5}$$

Thus the first row of $\mathcal{Q}$ is a zero vector. Then $\mathcal{Q}$ can be considered as a vector in the space spanned by $|\sigma_{\mathbf{a}}\rangle\!\rangle \langle\!\langle \sigma_{\mathbf{a}'}|$, where $\mathbf{a} \neq \mathbf{0}$, which is $(d^2 - 1)d^2$ dimensional in size.

It is easy to see that the transformation in Eq. (7), from a $\mathcal{Q}$ to a $\boldsymbol{\alpha}$, is a a linear map. We can prove that the

kernel of this linear map is only $\mathcal{Q} = 0$. Consider a $\mathcal{Q}$ in the kernel, then we get $\mathcal{Q} - g_\gamma \mathcal{Q} g_\gamma^\dagger = 0$ for any g_γ and $\mathcal{Q}|\rho_{\text{in}}\rangle\rangle = 0$. Thus $\mathcal{Q} g_\gamma |\rho_{\text{in}}\rangle\rangle = g_\gamma \mathcal{Q}|\rho_{\text{in}}\rangle\rangle = 0$ holds for any g_γ. Because $g_\gamma |\rho_{\text{in}}\rangle\rangle$ form a basis for the operator space, then $\mathcal{Q}$ can only be zero map. Thus the dimension of the null space of $\mathbf{C}$ due to the transformation in Eq. (7) is equal to the dimension of the space of all the $\mathcal{Q}$, i.e., equal to $d^2(d^2 - 1)$.

By vectorizing the map $\mathcal{Q}$, the linear map transforming $\mathcal{Q}$ to $\boldsymbol{\alpha}$ can be represented as a matrix. Thus for a given $\boldsymbol{\alpha}$, one can determine whether this $\boldsymbol{\alpha}$ is caused by the transformation by solving a system of linear equations. If there is a solution and the first row of the solution is a zero vector, then the $\boldsymbol{\alpha}$ is generated by Eq. (7). We find all the dependence relations $\boldsymbol{\alpha}$ in our examples are from the transformation in Eq. (7). We conjecture this is true for all informationally complete gate sets but cannot prove this in full at the current juncture.

APPENDIX B: ACCURACY SCALING WITH CIRCUIT LENGTH

Here we analyze the accuracy scaling of our RL-GST method with the length of the tomographic circuits, or more precisely, the maximum length of the circuits L_{max} (see definition below).

We must first determine a figure of merit to measure the accuracy of estimation. Recently Ref. [38] gave a comprehensive analysis of gate-set tomography written by the original authors of GST, where they used the diamond distance between the estimated gates and actual gates as the measure of accuracy. However, the diamond distance is a gauge-dependent quantity. The authors chose to gauge optimize the estimated gates to the actual gates before computing the diamond distance. Rather than having to make a gauge choice, especially since we do not usually know the actual gates, we instead choose a gauge-invariant metric based on the performance of predicting the results of test circuits, as defined by Eq. (10) in the main text. Here we predict the outcome probabilities of 10 000 random circuits with length 10, which is long enough to generate many different circuits and is very easy to compute the actual outcome probabilities. Below we show the accuracy scaling of our method and GST with circuit length using simulated experiments.

Here we simulate the gate set $\{g_i = I, g_x = X_{\pi/2}, g_y = Y_{\pi/2}\}$. We still use the noise model presented in the main text. The noise parameter is set such that the error strength is almost 10^{-3}, more precisely, with AGSI $= 7.2 \times 10^{-4}$. The fiducial sequences for GST circuits are the same as those in the main text, but for the germs, we choose a larger set suggested by [15]

$$\{g_i, g_x, g_y, g_x g_y, g_x g_y g_i, g_x g_i g_y, g_x g_i g_i, g_y g_i g_i, g_x g_x g_i g_y,$$
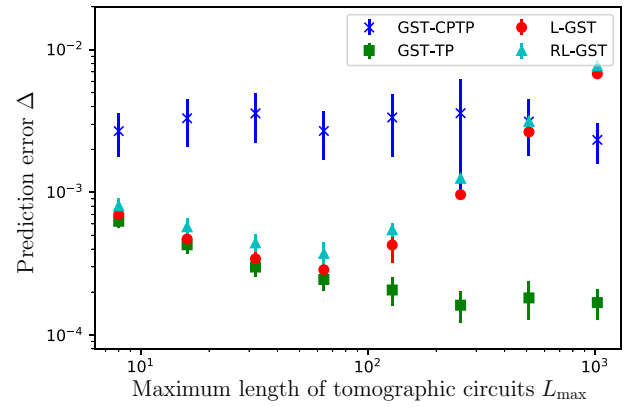
$$g_x g_y g_y g_i, g_x g_x g_y g_x g_y g_y\}. \tag{B1}$$

FIG. 5. Prediction error of test circuits versus the maximum length of tomographic circuits L_{max}.

We simulate GST for circuit lengths $\{2^0, 2^1, \ldots, 2^{\log_2 L_{\text{max}}}\}$, with L_{max} taking values $8, 16, 32, 64, 128, 256, 512,$ and 1024.

For our RL-GST, we choose random circuits with varying lengths. Compared with random circuits with fixed lengths, the random circuits with varying lengths are observed to be more flexible and robust. If we find the noise drifts heavily as circuit length increases or the circuits with large length violate the constraint for linear approximation, we can safely discard data from these long circuits. Because short random circuits do not have enough structure to be tomographically sensitive, the length of the random circuits in our method always starts from 8. The list of lengths of random circuits has the structure $\{2^3, 2^4, \ldots, 2^{\log_2 L_{\text{max}}}\}$, with L_{max} taking the same values as for GST. We sample the same number of random circuits for each length and keep the total resource cost the same as that in GST for each fixed L_{max}.

Figure 5 shows the prediction errors of RL-GST and GST versus the maximum length of tomographic circuits. The prediction error of GST-CPTP stays almost constant, likely due to a much larger estimation error of SPAM compared with the other estimators. The prediction error of GST-TP decreases as L_{max} increases but plateaus at almost $L_{\text{max}} = 256$. For L-GST and RL-GST, the prediction errors first decrease and then increase when $L_{\text{max}} > 64$. This is because the long circuits violate the linear approximation. For good performance, we propose that the maximum length of the circuits used in our method should satisfy $L_{\text{max}}\epsilon \leq 0.1$, or more precisely $(L_{\text{max}}\epsilon)^2 \leq 1/\sqrt{N_s}$ ($N_s = 8192$ here) because the error from linear approximation is swamped by statistical uncertainty. We fit the data points (in log-log scale) of GST-TP, L-GST, RL-GST up to $L_{\text{max}} = 64$, and the obtained slope is $-0.45, -0.43, -0.37$. Up to $L_{\text{max}} = 64$, the accuracy of our RL-GST is comparable with those of GST. For higher accuracy, we can

increase the number of circuits used instead of increasing the length of circuits.

———

[1] 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 (2008).

[2] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, Phys. Rev. Lett. **106**, 180504 (2011).

[3] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).

[4] O. Moussa, M. P. da Silva, C. A. Ryan, and R. Laflamme, Practical Experimental Certification of Computational Quantum Gates Using a Twirling Procedure, Phys. Rev. Lett. **109**, 070504 (2012).

[5] M. Paris and J. Řeháček, eds., *Quantum State Estimation* (Springer-Verlag Berlin Heidelberg, Berlin Heidelberg, 2004).

[6] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).

[7] J. Poyatos, J. I. Cirac, and P. Zoller, Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate, Phys. Rev. Lett. **78**, 390 (1997).

[8] D. Mogilevtsev, J. Řeháček, and Z. Hradil, Relative tomography of an unknown quantum state, Phys. Rev. A **79**, 020101 (2009).

[9] D. Mogilevtsev, Calibration of single-photon detectors using quantum statistics, Phys. Rev. A **82**, 021807 (2010).

[10] A. M. Brańczyk, D. H. Mahler, L. A. Rozema, A. Darabi, A. M. Steinberg, and D. F. V. James, Self-calibrating quantum state tomography, New J. Phys. **14**, 085003 (2012).

[11] N. Quesada, A. M. Brańczyk, and D. F. V. James, Self-calibrating tomography for multidimensional systems, Phys. Rev. A **87**, 062118 (2013).

[12] B. P. Williams and P. Lougovski, Quantum state estimation when qubits are lost: A no-data-left-behind approach, New J. Phys. **19**, 043003 (2017).

[13] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, Phys. Rev. A **87**, 062119 (2013).

[14] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, arXiv:1310.4492 (2013).

[15] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nat. Commun. **8**, 1 (2017).

[16] D. Greenbaum, Introduction to quantum gate set tomography, arXiv:1509.02921 (2015).

[17] D. Kim, D. Ward, C. Simmons, J. K. Gamble, R. Blume-Kohout, E. Nielsen, D. Savage, M. Lagally, M. Friesen, and S. Coppersmith *et al.*, Microwave-driven coherent operation of a semiconductor quantum dot charge qubit, Nat. Nanotechnol. **10**, 243 (2015).

[18] J. P. Dehollain, J. T. Muhonen, R. Blume-Kohout, K. M. Rudinger, J. K. Gamble, E. Nielsen, A. Laucht, S. Simmons, R. Kalra, and A. S. Dzurak *et al.*, Optimization of a solid-state electron spin qubit using gate set tomography, New J. Phys. **18**, 103018 (2016).

[19] M. Rol, C. Bultink, T. O'Brien, S. De Jong, L. Theis, X. Fu, F. Luthi, R. Vermeulen, J. de Sterke, and A. Bruno *et al.*, Restless Tuneup of High-Fidelity Qubit Gates, Phys. Rev. Appl. **7**, 041001 (2017).

[20] Y. Chen, M. Farahzad, S. Yoo, and T.-C. Wei, Detector tomography on ibm quantum computers and mitigation of an imperfect measurement, Phys. Rev. A **100**, 052315 (2019).

[21] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, Quantum computation with universal error mitigation on a superconducting quantum processor, Sci. Adv. **5**, eaaw5686 (2019).

[22] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, Probing Context-Dependent Errors in Quantum Processors, Phys. Rev. X **9**, 021045 (2019).

[23] IBM Quantum Experience, https://quantum-computing.ibm.com.

[24] A. Ash-Saki, M. Alam, and S. Ghosh, Experimental characterization, modeling, and analysis of crosstalk in a quantum computer, IEEE Trans. Quantum Eng. **1**, 1 (2020).

[25] R. Blume-Kohout, E. Nielsen, K. Rudinger, K. Young, M. Sarovar, and T. Proctor, in *APS March Meeting Abstracts* (American Physical Society, Boston, Massachusetts, 2019), Vol. 2019, p. P35.

[26] R. Blume-Kohout, Idle tomography, https://www.osti.gov/biblio/1581878.

[27] T. Proctor, K. Rudinger, K. Young, M. Sarovar, and R. Blume-Kohout, What Randomized Benchmarking Actually Measures, Phys. Rev. Lett. **119**, 130502 (2017).

[28] G. C. Knee, E. Bolduc, J. Leach, and E. M. Gauger, Quantum process tomography via completely positive and trace-preserving projection, Phys. Rev. A **98**, 062336 (2018).

[29] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A **60**, 1888 (1999).

[30] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Phys. Lett. A **303**, 249 (2002).

[31] E. Nielsen, R. J. Blume-Kohout, K. M. Rudinger, T. J. Proctor, L. Saldyt, and USDOE, Python gst implementation (pygsti) v. 0.9, version v. 0.9 (2019).

[32] 5-qubit backend: IBM Q team, "IBM Q 5 Ourense backend specification V1.2.0," (2020). Retrieved from https://quantum-computing.ibm.com.

[33] 5-qubit backend: IBM Q team, "IBM Q 5 Vigo backend specification V1.2.1," (2020). Retrieved from https://quantum-computing.ibm.com.

[34] 5-qubit backend: IBM Q team, "IBM Q 5 Bogota backend specification V1.0.4," (2020). Retrieved from https://quantum-computing.ibm.com.

[35] J. Helsen, F. Battistel, and B. M. Terhal, Spectral quantum tomography, Npj Quantum Inf. **5**, 1 (2019).

[36] Y.-Q. Chen, K.-L. Ma, Y.-C. Zheng, J. Allcock, S. Zhang, and C.-Y. Hsieh, Non-Markovian Noise Characterization with the Transfer Tensor Method, Phys. Rev. Appl. **13**, 034045 (2020).

[37] G. White, C. Hill, F. Pollock, L. Hollenberg, and K. Modi, Demonstration of non-markovian process characterisation and control on a quantum processor, Nat. Commun. **11**, 1 (2020).

[38] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, arXiv:2009.07301 (2020).