

Self-consistent quantum process tomography

Seth T. Merkel,^{1,*} Jay M. Gambetta,¹ John A. Smolin,¹ Stefano Poletto,¹ Antonio D. Córcoles,¹ Blake R. Johnson,² Colm A. Ryan,² and Matthias Steffen¹

¹*IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, USA*

²*Raytheon BBN Technologies, Cambridge, Massachusetts 02138, USA*

(Received 9 November 2012; published 24 June 2013)

Quantum process tomography is a necessary tool for verifying quantum gates and diagnosing faults in architectures and gate design. We show that the standard approach of process tomography is grossly inaccurate in the case where the states and measurement operators used to interrogate the system are generated by gates that have some systematic error, a situation all but unavoidable in any practical setting. These errors in tomography cannot be fully corrected through oversampling or by performing a larger set of experiments. We present an alternative method for tomography to reconstruct an entire library of gates in a self-consistent manner. The essential ingredient is to define a likelihood function that assumes nothing about the gates used for preparation and measurement. In order to make the resulting optimization tractable, we linearize about the target, a reasonable approximation when benchmarking a quantum computer as opposed to probing a black-box function.

DOI: [10.1103/PhysRevA.87.062119](https://doi.org/10.1103/PhysRevA.87.062119)

PACS number(s): 03.65.Wj, 03.67.Ac, 42.50.Pq

I. INTRODUCTION

The design and control of scalable quantum architectures at the level of precision necessary for fault-tolerant computation remains a considerable challenge. There has been remarkable progress in numerous physical systems (e.g., superconducting circuits [1–3], atomic systems [4,5], etc.) to improve metrics such as the gate fidelity [6], and incrementally these values are approaching the fault-tolerant threshold. Diagnosing errors and improving designs of gates and architectures, however, often requires much more information than a single scalar value. Instead we need a full characterization of a quantum process, and that necessitates quantum process tomography (QPT) [7].

The essential idea of QPT is as follows: prepare an initial quantum state, apply the operation we would like to reconstruct, measure the expectation value of some observable, and then repeat for different initial states and observables until it is possible to retrodict the process through some form of matrix inversion. This scales exponentially poorly with the number of qubits n , since the parameters necessary to specify a general map are $O(2^{4n})$, but generally we only need to perform QPT on small subsystems (one or two qubits) of a larger architecture and can then verify that the subsystems are isolated through other means [8]. Additionally, because there will be errors on the measurement outcomes, if for no reason other than finite measurement statistics, the resulting process may not be physical (i.e., a completely positive trace preserving map). When this nonphysicality arises from Gaussian noise on the measurement outcomes, it can be corrected using maximum likelihood estimation or through Bayesian inference techniques [9–16].

In addition to stochastic measurement noise, quantum process tomography should be consistent with respect to systematic noise. Most quantum computing architectures allow only for fixed initial states and measurement operators so that the state preparation and measurement (SPAM) described for tomography involves the application of gates that may have

the same degree of error as the process being interrogated. These errors in QPT cannot be corrected by gathering more measurement statistics or enforcing physicality alone and are known to be problematic in quantum state tomography [17]. Here we show that with standard maximum likelihood techniques, these errors, and in particular, coherent errors, lead to fidelity estimates from QPT that can be very poor. Even more problematic is that the ratio of the QPT to SPAM errors actually can grow as the SPAM error *decreases*, implying that as the gate fidelities approach the threshold it can be even harder to estimate them correctly. SPAM errors can be accounted for in fidelity estimation by using randomized benchmarking [5,18–21], but this does not provide full tomographic information about the gates in order to diagnose and correct errors. We observe a disagreement between benchmarking [21] and tomography [1] experiments on similar samples, which suggests that SPAM errors are a current limitation that needs to be addressed.

In this manuscript we explore techniques to compensate for SPAM errors while still obtaining a full characterization of our quantum gates. One simple technique is to oversample, not by increasing the repetitions of a given experiment, but by increasing the number of gates used for SPAM. We show that this technique shows no improvement for general unitary errors; however, it does yield an improvement for a uniform error such as a global frame transformation. The primary subject of this work is an alternative approach to tomography that places the SPAM gates on the same footing as the gate under investigation, and then retrodicts an entire library of unitary gates in a self-consistent manner. The “cost” of this method in terms of the number of independent measurements and postprocessing is polynomially equivalent to performing standard QPT on each gate in the library being reconstructed. Our approach assumes very little about the gate error model, in contrast to the self-correcting state tomography proposed in [22] or the bootstrap method in [23], which otherwise compensate for the same SPAM issues.

The remainder of the manuscript proceeds as follows. In Sec. II we give a detailed description of QPT and simulate

*Corresponding author: stmerkel@hrl.com

the effects of SPAM errors for a range of error models and system parameters. In Sec. III we show the effects of oversampling using unitary 2-designs, and in Sec. IV we develop self-consistent gate-set tomography. Finally, in Sec. V we apply our self-consistent tomography to experimental data obtained from two independent experiments on a system consisting of a single superconducting qubit.

II. THE TROUBLE WITH TOMOGRAPHY

A general quantum process on a finite d -dimensional Hilbert space \mathcal{H}_d ($d = 2^n$ for qubit systems) is a completely positive trace-preserving (CPTP) map. There are many ways to represent such a map Λ , but in this manuscript we will primarily use two of them: the Choi matrix [24], which is given by

$$\rho_\Lambda = \frac{1}{d} \sum_{ij} E_{ij} \otimes \Lambda(E_{ij}), \quad (1)$$

where $E_{ij} = |i\rangle\langle j|$, and the Pauli transfer matrix (PTM) [1], which is

$$(R_\Lambda)_{ij} = \frac{1}{d} \text{Tr}[P_i \Lambda(P_j)], \quad (2)$$

where P_j denotes the set of Pauli matrices, $\{I, X, Y, Z\}^{\otimes n}$, though one could extend this to qudit systems by considering any Hermitian basis for the P_j values. The transformation between these two representations is the linear mapping,

$$(R_\Lambda)_{ij} = \text{Tr}(\rho_\Lambda P_j^T \otimes P_i), \quad (3)$$

$$\rho_\Lambda = \frac{1}{d^2} \sum_{ij} (R_\Lambda)_{ij} P_j^T \otimes P_i,$$

and each representation has useful properties. The Choi matrix is positive semidefinite if and only if Λ is completely positive. If we define $P_0 = I_d$, then the map is trace preserving if and only if the first row of \mathcal{R}_Λ is the vector $(1, 0, 0, \dots)$ and is unital if the first column has the same structure. Furthermore, in the PTM picture, composition becomes matrix multiplication, $\mathcal{R}_{\Lambda_1 \circ \Lambda_2} = \mathcal{R}_{\Lambda_1} \mathcal{R}_{\Lambda_2}$, since the PTM is the standard superoperator group representation defined over the Pauli basis.

Let us define an experiment by the triple $\{\rho_j, M_j, m_j\}$, which describes preparing the system in state ρ_j , measuring the operator M_j and obtaining an expectation value m_j according to

$$m_j = \text{Tr}[M_j \Lambda(\rho_j)]. \quad (4)$$

We can define a vector $|\rho\rangle\rangle$ whose elements are $\langle\langle j|\rho\rangle\rangle = \text{Tr}(P_j \rho)/d$ as well as $\langle\langle M|j\rangle\rangle = \text{Tr}(M P_j)$ (the omission of the dimensional factor in the measurement is intentional and is a consequence of bounding the values of the \mathcal{R} matrix between ± 1). In this form,

$$m_j = \langle\langle M_j | \mathcal{R}_\Lambda | \rho_j \rangle\rangle = \text{Tr}(|M_j\rangle\rangle \langle\langle \rho_j |^T \mathcal{R}_\Lambda). \quad (5)$$

This final expression is simply another inner product of the matrix \mathcal{R}_Λ with the super-operator $|M_j\rangle\rangle \langle\langle \rho_j|$. For the vectorization of the superoperator \mathcal{R}_Λ we use the notation \mathbf{r}_Λ as the column major, vectorized \mathcal{R}_Λ and \mathbf{s}_j as the vectorized $|M_j\rangle\rangle \langle\langle \rho_j|$, so that

$$m_j = \mathbf{s}_j^T \mathbf{r}_\Lambda. \quad (6)$$

We can now express the entire set of experiments as

$$\mathbf{m} = S^T \mathbf{r}_\Lambda, \quad (7)$$

where the columns of the rectangular matrix S are the vectors \mathbf{s}_j .

The first estimate for our quantum process \mathbf{r}_{bare} arises from linear inversion. While S is typically not square, we can invert SS^T or, in the case of an incomplete set of measurements, find its pseudoinverse (in order to be invertible the columns of S must be linearly independent, necessitating at least d^2 measurements, $d^2 - 1$ if we assume the trace is preserved). This leads to the bare estimate

$$\mathbf{r}_{\text{bare}} = (SS^T)^{-1} S \mathbf{m}. \quad (8)$$

The astute reader may have noticed that at no point have we enforced that our estimate \mathbf{r}_{bare} corresponds to a physical map, and we will find that for many error models it does not.

The dominant source of error, in the tomography literature at least [9–16], is statistical error due to finite measurement statistics. Instead of direct access to m_j we measure

$$m_j = \mathbf{s}_j^T \mathbf{r}_\Lambda + \sqrt{N_j} W_j, \quad (9)$$

where W_j is a Gaussian random variable with mean zero and unit variance and N_j is the noise power. Due to the central limit theorem we can always approximate the error model as Gaussian if we repeat the measurement sufficiently many times, in which case N_j scales as $1/(\text{no. of repetitions of experiment } j)$. The transformation $m'_j = m_j/\sqrt{N_j}$ and $\mathbf{s}'_j = \mathbf{s}_j/\sqrt{N_j}$ simplifies the Gaussian likelihood function to

$$\mathcal{L}(\mathbf{m}'|\mathbf{r}) = \exp\left(-\frac{1}{2} \sum_j |m'_j - \mathbf{s}'_j^T \mathbf{r}|^2\right). \quad (10)$$

For this error model the most rigorous estimate of the gate would be to perform some sort of Bayesian estimation where $\mathbf{r}_{\text{Bayes}} = \int \mathbf{r} \mathcal{L}(\mathbf{m}'|\mathbf{r}) d\mathbf{r} / \int \mathcal{L}(\mathbf{m}'|\mathbf{r}) d\mathbf{r}$, but these integrals are often extremely hard to calculate, especially since there is no uniquely good measure over quantum channels $d\mathbf{r}$. A much simpler technique is to report the maximum of the likelihood function. If the global maximum lies inside the space of physical maps, then it is simply the \mathbf{r}_{bare} from Eq. (8), and if most of the support of $\mathcal{L}(\mathbf{m}'|\mathbf{r})$ is also physical, then the Bayesian estimate and the maximum likelihood estimate will be approximately the same. In the case where the bare estimate is unphysical, maximizing $\mathcal{L}(\mathbf{m}'|\mathbf{r})$ over the space of physical \mathbf{r} is a semidefinite program (SDP) [25] and can be solved for small instances easily with optimizers such as SeDuMi [26]. This is because the likelihood function is of the form of a 2-norm distance between vectors \mathbf{m}' and $S'^T \mathbf{r}$, and the constraints are that the Choi matrix is positive semidefinite. We use the approach outlined in the supplement of [1], which is an extension of the state tomography techniques in [27,28] as well as [12].

While there are issues regarding maximum likelihood estimations for QPT, there is a deeper problem that affects any reconstruction method that utilizes the likelihood function. To calculate the likelihood function we require \mathbf{s}_j , which in turn implies we have a complete characterization of ρ_j and M_j . In any experimental implementation of tomography there

will be SPAM errors, and in many situations the magnitude of the SPAM errors is the same order as the size of the gate errors we are trying to estimate. If the SPAM errors are completely uncorrelated with the preparation and measurement gates, the situation is reducible to the previous case by effectively treating the SPAM errors as additional sampling noise on the measurement. When the errors are systematic, we may write $|\rho_j\rangle \rightarrow \mathcal{R}_{\mathcal{E}_j}|\rho_j\rangle$, $|M_j\rangle \rightarrow \mathcal{R}_{\mathcal{F}_j}|M_j\rangle$ and therefore $\mathbf{s}_j \rightarrow \mathcal{R}_{\mathcal{E}_j} \otimes \mathcal{R}_{\mathcal{F}_j}^T \mathbf{s}_j$.

For the remainder of this manuscript we will consider a slightly less general form of tomography that is applicable to many experimental implementations: a fixed initial state $|\rho_0\rangle$, measurement operator $\langle\langle M_0|$, and a library of gates $G = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_N\}$. In this picture we describe experiments according to the convention

$$m_{ij} = \langle\langle M_0|\mathcal{R}_j\mathcal{R}_\Lambda\mathcal{R}_i|\rho_0\rangle\rangle = \langle\langle M_j|\mathcal{R}_\Lambda|\rho_i\rangle\rangle. \quad (11)$$

Systematic errors are described $G^{(\text{err})} = \{\mathcal{R}_{\mathcal{E}_1}\mathcal{R}_1, \mathcal{R}_{\mathcal{E}_2}\mathcal{R}_2, \dots, \mathcal{R}_{\mathcal{E}_N}\mathcal{R}_N\}$, under the assumption that the initial gate set was composed of unitary maps.

In Fig. 1 we simulate the effects of both systematic and stochastic measurement noise in the case where the measured gate is a perfect identity gate. This test is a good primitive for both theory and experiment, since the identity is the one gate that should be perfectly implementable in any experiment by immediately performing measurement after state preparation (i.e., doing nothing for no time). In this figure the systematic SPAM error comes from a depolarizing channel of varying strengths \mathcal{E}_{dep} . We do not impose the CPTP constraint on the outcome and therefore measure the difference between the reconstructed gate and the identity by a diamond norm distance [29], which is calculated through the semidefinite programming technique in [30]. In this simulation the gates map the $|0\rangle$ state to the four points on a tetrahedron, which is essentially the most symmetric, minimal set of gates. We observe that for large noise powers the error in the estimate decreases exponentially with respect to decreasing noise power (and thus increasing repetitions) until it hits a floor determined by the systematic error in the interrogating gates.

In Fig. 1(b) we look at a similar plot for different libraries of unitary SPAM gates. The first two gate sets are defined in terms of the set of states to which they map the qubit ground state $|0\rangle$: a tetrahedron or the six cardinal directions on the Bloch sphere. The second two groups are unitary 2-designs: a 12-element subgroup of the Clifford group and the entire 24-element Clifford group. Both of these groups correspond to the solid rotations of a cube, with the 12-element subgroup consisting of 180° rotations about the faces and 120° rotations about the corners of the cube. Generically, increasing the number of gates and maximizing the entropy of SS^T speeds convergence to the estimation floor, but the actual limit remains unchanged.

By introducing systematic errors into our interrogating gate set, the maximum of the original estimator Eq. (4) does not yield a faithful reconstruction of the unidentified gate. In fact, the resulting estimate can be highly nonphysical in the absence of any stochastic measurement noise. Semidefinite programming techniques constrain the estimator to the space of physical maps using the covariance matrix of the Gaussian

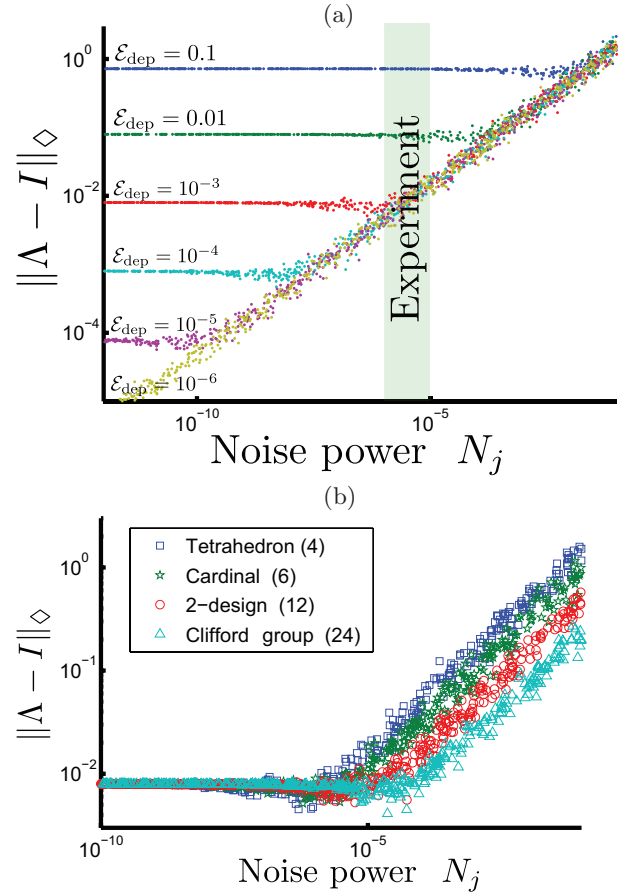


FIG. 1. (Color online) Reconstructing a perfect identity gate with imperfect tomography due to depolarizing errors on the SPAM gates as well as Gaussian random noise on the measurement outcomes. We plot the diamond norm distance between the bare estimate Eq. (8) and the identity versus the noise power. In (a) we vary the strength of the depolarizing error for a fixed gate set, with four elements corresponding to mapping the ground state to the four corners of a tetrahedron. The experimental noise power is obtained from [1]. In (b) we fix the depolarizing strength at $\mathcal{E}_{\text{dep}} = 10^{-3}$ and vary the set of gates used for SPAM over four sets of gates of differing order.

likelihood function as a metric which we use to minimize the distance between a physical state and the unphysical estimate. However, if this likelihood estimator is incorrect in the absence of any stochastic measurement noise, then there is no reason to trust the covariance matrix metric over any other metric on the operator space (e.g., the flat metric). Minimizing the distance over the flat metric can be performed using the methods in [15], which was proven to be optimal. Since there is no gain in using the covariance method for our problem, we use a flat metric in the following simulations due to its ease of calculation.

In Fig. 2 we look at the error in reconstructing the identity, in the absence of stochastic measurement errors, for different models of the systematic errors. We find the closest physical state to the estimate under a flat metric. Since the two gates under comparison are physical, we can compare them with the gate fidelity F , or more precisely, the gate error $1 - F$. It is informative to plot the ratio of the reconstruction error to the average SPAM error on the individual gates versus the average SPAM errors. There is a stark difference between coherent

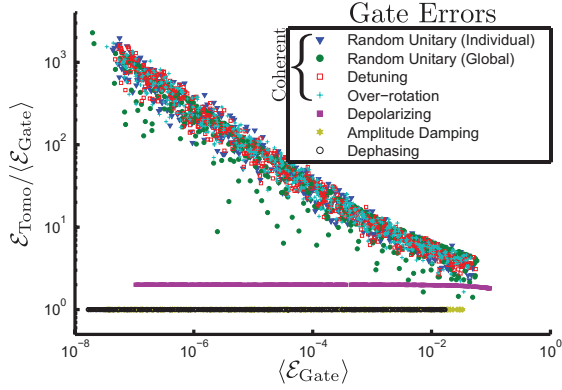


FIG. 2. (Color online) The ratio of the error in reconstructing the identity to the average gate error plotted versus the average gate error for a single qubit using a tetrahedral gate set and no stochastic measurement noise. The estimated operation is the “closest” physical map to the bare reconstruction in a flat metric. Seven different error models are compared, four coherent errors (different random unitary maps applied to each gate, a single random unitary map applied to each gate, a detuning error applied i.i.d. to each gate sampled from a Gaussian distribution, and an over-rotation error applied i.i.d. to each gate sampled from a Gaussian distribution) as well three incoherent processes (amplitude damping, dephasing, and depolarizing errors).

errors (over-rotation, detuning errors, etc.) and incoherent error models (T1 or T2 processes, for example). In the incoherent case, the error in estimating the identity is proportional to the error on the individual gates, while for the coherent model the ratio of the errors grows exponentially poorly as the gate error decreases.

The dependence of the tomography error on the gate errors has dramatic implications for benchmarking quantum gates with QPT. A quantum computing architecture viable for fault-tolerant error correction will require gates with error rates in the range of 10^{-3} to 10^{-5} . From Fig. 2 we see that the *measured* error rates for a fault-tolerant gate set will be more like 10^{-2} to 10^{-3} if the dominant source of errors is coherent. In fact, measured error rates of 10^{-5} are not even present in Fig. 2 and correspond to physical gate error rates smaller than 10^{-8} . In many cases the dominant source of error is T_1 - or T_2 -type errors, in which case this effect will be minimal; however, as coherence times increase, coherent errors will very likely become a bottleneck. From Fig. 1 we see that for the typical noise power in current experiments we expect SPAM errors to dominate the estimate if the gate error is less than about a percent, which is well above the error due solely to T1 or T2.

III. QPT WITH OVERSAMPLING

From the previous section one can observe that to eliminate fully stochastic measurement errors it is sufficient to oversample, repeating experiments many times in order to accurately measure expectation values. Oversampling is also an intuitive approach to dealing with SPAM errors if instead of increasing repetitions we alternatively increase the number of independent experiments (e.g., by increasing the size of the gate library G). In this section we show that this method does *not* compensate for general SPAM errors; however, when the

gate library is a unitary 2-design some uniform SPAM errors are removed.

Starting from Eq. (5), we can derive the bare estimate in a slightly different manner that will be more conducive to the following discussion. First, note that we can multiply both sides by $|M_j\rangle\langle\rho_i|$ and sum over i and j to obtain

$$\sum_{ij} m_{ij} |M_j\rangle\langle\rho_i| = \left(\sum_j |M_j\rangle\langle M_j| \right) \mathcal{R}_\Lambda \left(\sum_i |\rho_i\rangle\langle\rho_i| \right). \quad (12)$$

In this case the bare estimate is

$$\begin{aligned} \mathcal{R}_\Lambda^{(\text{est})} &= \left(\sum_{j'} |M_{j'}\rangle\langle M_{j'}| \right)^{-1} \sum_{ij} m_{ij} |M_j\rangle\langle\rho_i| \\ &\times \left(\sum_{i'} |\rho_{i'}\rangle\langle\rho_{i'}| \right)^{-1}. \end{aligned} \quad (13)$$

If there are SPAM errors the term m_{ij} has the form

$$m_{ij} = \langle\langle M_0 | \mathcal{R}_{\mathcal{E}_j} \mathcal{R}_j \mathcal{R}_\Lambda \mathcal{R}_{\mathcal{E}_i} | \rho_0 \rangle\rangle, \quad (14)$$

which leads to

$$\begin{aligned} \mathcal{R}_\Lambda^{(\text{est})} &= \left(\sum_{j'} |M_{j'}\rangle\langle M_{j'}| \right)^{-1} \sum_j \mathcal{R}_j^T |M_0\rangle\langle\langle M_0 | \mathcal{R}_{\mathcal{E}_j} \mathcal{R}_j \\ &\times \mathcal{R}_\Lambda \sum_i \mathcal{R}_{\mathcal{E}_i} \mathcal{R}_i | \rho_0 \rangle\rangle \langle\langle \rho_0 | \mathcal{R}_i^T \left(\sum_{i'} |\rho_{i'}\rangle\langle\rho_{i'}| \right)^{-1}. \end{aligned} \quad (15)$$

There are some very interesting consequences of the final expression, which we shall examine in a few special cases. If the SPAM error is independent of the gate, then due to our error conventions we get a cancellation of the final term in Eq. (15), yielding

$$\mathcal{R}_\Lambda^{(\text{est})} = \left(\sum_j |M_j\rangle\langle M_j| \right)^{-1} \left(\sum_{j'} |M_{j'}\rangle\langle M_{j'}| \right) \times \mathcal{R}_\Lambda \mathcal{R}_{\mathcal{E}}. \quad (16)$$

We could derive similar expressions if the constant error occurred before the gate or in some combination of pre- and postgate error. If in addition the error commutes with all of the gates in the library (as is the case with depolarizing errors) the result is, for example,

$$\mathcal{R}_\Lambda^{(\text{est})} = \mathcal{R}_{\text{dep}} \mathcal{R}_\Lambda \mathcal{R}_{\text{dep}}, \quad (17)$$

independent of the set of gates used for tomography. This form of the reconstructed map lends some intuition to Fig. 2, where the depolarizing error and other incoherent errors led to a tomography error that had polynomial scaling with respect to the error on the gate set. Qualitatively, errors of this form “commute” with the reconstruction procedure.

If the error does not commute with all of the gates but the library forms a unitary 2-design [31], we can also simplify equation Eq. (16). We start by independently analyzing the two quantities in parenthesis in Eq. (16). These sums are both

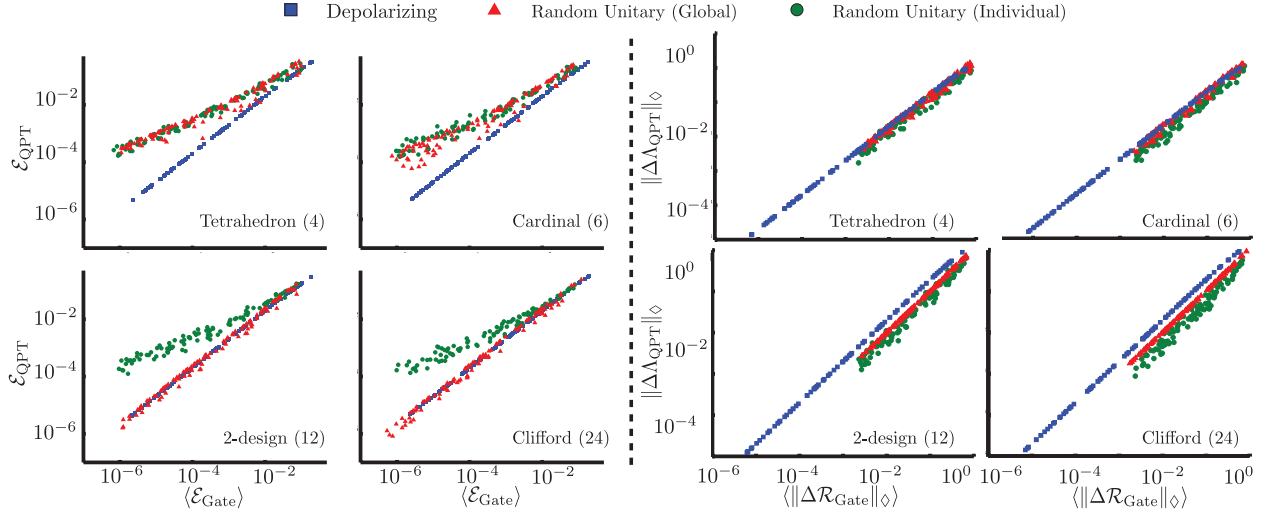


FIG. 3. (Color online) Simulated error in reconstructing the identity versus average SPAM error in terms of the gate fidelity (left) and the diamond norm distance (right). We look at four gate sets and three types of SPAM error.

twirls over a unitary 2-design \mathcal{W} , which has the following generic form:

$$\mathcal{W}(A) = |I\rangle\langle I| \langle I|A|I\rangle + (\mathbb{I} - |I\rangle\langle I|) \frac{\text{Tr}(A) - \langle I|A|I\rangle}{d^2 - 1}, \quad (18)$$

[8]. Therefore the first factor is given by

$$\mathcal{W}(|M_0\rangle\langle M_0|) = |I\rangle\langle I| \langle I|M_0\rangle\langle M_0| + (\mathbb{I} - |I\rangle\langle I|) \frac{\langle M_0|M_0\rangle - \langle I|M_0\rangle\langle M_0|I\rangle}{d^2 - 1} \quad (19)$$

and the second by

$$\begin{aligned} \mathcal{W}(|M_0\rangle\langle M_0| \mathcal{R}_{\mathcal{E}}) &= |I\rangle\langle I| \langle I|M_0\rangle\langle M_0| \mathcal{R}_{\mathcal{E}}|I\rangle + (\mathbb{I} - |I\rangle\langle I|) \\ &\times \frac{\langle M_0|\mathcal{R}_{\mathcal{E}}|M_0\rangle - \langle I|M_0\rangle\langle M_0|\mathcal{R}_{\mathcal{E}}|I\rangle}{d^2 - 1}. \end{aligned} \quad (20)$$

Both of these are proportional to depolarizing channels, which generically have the form $|I\rangle\langle I| + \epsilon(\mathbb{I} - |I\rangle\langle I|)$. The product $\mathcal{W}(|M_0\rangle\langle M_0|) \mathcal{W}(|M_0\rangle\langle M_0| \mathcal{R}_{\mathcal{E}})$ is therefore proportional to a depolarizing channel with a strength

$$\epsilon = \frac{\frac{\langle M_0|\mathcal{R}_{\mathcal{E}}|M_0\rangle}{\langle M_0|\mathcal{R}_{\mathcal{E}}|I\rangle\langle I|M_0\rangle} - 1}{\frac{\langle M_0|M_0\rangle}{\langle I|M_0\rangle^2} - 1} \quad (21)$$

and proportionality constant

$$\alpha = \frac{\langle M_0|\mathcal{R}_{\mathcal{E}}|I\rangle}{\langle I|M_0\rangle}. \quad (22)$$

The bare estimate for a constant error on a gate set composed of a unitary 2-design is thus given by

$$\mathcal{R}_{\Lambda}^{(\text{est})} = \alpha \mathcal{R}_{\epsilon}^{(\text{dep})} \mathcal{R}_{\Lambda} \mathcal{R}_{\mathcal{E}}. \quad (23)$$

The previous theoretical analysis suggests that there may be a fundamental difference when the gate library is a unitary 2-design, and this is confirmed in Fig. 3. When we measure the errors in terms of the gate fidelity, we see a clear difference with

respect to the unitary 2-designs. In that case a global unitary error has the same effect on the fidelity as a depolarizing error, as expected from the previous argument. QPT errors resulting from independent unitaries on each of the gates remains qualitatively the same for all four sets. What we describe as a global unitary error is a strange object and should not be confused for control errors such as over rotation or detuning errors, but is instead more akin to a global frame misalignment.

Interestingly, the errors look very different when measured by the diamond norm (the right and left plots of Fig. 3). The diamond norm distance of the reconstruction scales polynomially with the gate-set diamond norm error for all forms of the SPAM error and gate sets. The fact that the rate of the depolarizing error spans twice as large a domain as the coherent errors in the diamond norm is also curious. These two metrics clearly yield very different results, and it is an important question as to which is appropriate for an experiment to report. For this reason, we consider both for the remainder of this manuscript.

IV. SELF-CONSISTENT TOMOGRAPHY

We have thus far failed at tomography in this manuscript for the simple reason that our methodology is based on the likelihood function, and we have used the wrong one. A proper likelihood function should incorporate the SPAM errors, but characterizing the SPAM errors themselves requires QPT and therefore a likelihood function. This is an impasse for standard QPT methods. The alternative we present in this section is to assume nothing about the SPAM errors (though later we will bound their magnitude) and instead estimate an entire library of gates in a self-consistent manner.

Let us write the set of experimental gates as $G^{(\text{exp})} = \{\mathcal{R}_{\mathcal{E}_1} \mathcal{R}_1, \mathcal{R}_{\mathcal{E}_2} \mathcal{R}_2, \dots, \mathcal{R}_{\mathcal{E}_N} \mathcal{R}_N\}$, which is a faulty version of the library $G^{(\text{ideal})} = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_N\}$. The free parameters of this model are the error terms $\mathcal{R}_{\mathcal{E}_i}$, while the ideal gates are fixed. We begin by performing a series of N^3 experiments (at least d^6) of the form

$$m_{ijk} = \langle M_0 | \mathcal{R}_{\mathcal{E}_k} \mathcal{R}_k \mathcal{R}_{\mathcal{E}_j} \mathcal{R}_j \mathcal{R}_{\mathcal{E}_i} \mathcal{R}_i | \rho_0 \rangle, \quad (24)$$

which in essence consists of the tomography data necessary to reconstruct each element of $G^{(\text{exp})}$.

The likelihood of a trial set of gates $\tilde{G} = \{\tilde{\mathcal{R}}_{\mathcal{E}_1} \mathcal{R}_1, \tilde{\mathcal{R}}_{\mathcal{E}_2} \mathcal{R}_2, \dots, \tilde{\mathcal{R}}_{\mathcal{E}_N} \mathcal{R}_N\}$ is now given by

$$\mathcal{L}(\tilde{G}) = \exp \left(- \sum_{i,j,k} |m_{ijk} - \langle \langle M_0 | \tilde{\mathcal{R}}_{\mathcal{E}_k} \mathcal{R}_k \tilde{\mathcal{R}}_{\mathcal{E}_j} \mathcal{R}_j \tilde{\mathcal{R}}_{\mathcal{E}_i} \mathcal{R}_i | \rho_0 \rangle \rangle|^2 \right). \quad (25)$$

The constraint such that the gate library is physical is that each of the maps $\tilde{\mathcal{R}}_{\mathcal{E}_j}$ are completely positive and trace preserving. Maximizing the likelihood function (25) is equivalent to minimizing the least-squares estimator

$$\text{LSQ}(G) = \sum_{i,j,k} |m_{ijk} - \langle \langle M_0 | \tilde{\mathcal{R}}_{\mathcal{E}_k} \mathcal{R}_k \tilde{\mathcal{R}}_{\mathcal{E}_j} \mathcal{R}_j \tilde{\mathcal{R}}_{\mathcal{E}_i} \mathcal{R}_i | \rho_0 \rangle \rangle|^2 \quad (26)$$

but is highly nontrivial due to the fact that this function is 6th order in the gate set.

We can reduce the order of the estimator by linearizing the evolution about the ideal set of gates. This is a reasonable assumption if we are testing components of a faulty quantum computer as opposed to probing some unknown physics or interaction. In this situation we allow $\tilde{\mathcal{R}}_{\mathcal{E}_j} = \mathcal{R}_j + \tilde{\mathcal{E}}_j$, where $\tilde{\mathcal{E}}_j$ is not necessarily physical but is small, $\|\tilde{\mathcal{E}}_j\| \ll 1$. Expanding the evolution to first order in $\tilde{\mathcal{E}}_j$ we have

$$\tilde{m}_{ijk} = \langle \langle M_0 | \mathcal{R}_k \mathcal{R}_j \mathcal{R}_i + \tilde{\mathcal{E}}_k \mathcal{R}_k \mathcal{R}_j \mathcal{R}_i + \mathcal{R}_k \tilde{\mathcal{E}}_j \mathcal{R}_j \mathcal{R}_i + \mathcal{R}_k \mathcal{R}_j \tilde{\mathcal{E}}_i \mathcal{R}_i | \rho_0 \rangle \rangle, \quad (27)$$

or by transforming back to the physical $\tilde{\mathcal{R}}_{\mathcal{E}_j}$,

$$\tilde{m}_{ijk} = \langle \langle M_0 | -2\mathcal{R}_k \mathcal{R}_j \mathcal{R}_i + \tilde{\mathcal{R}}_{\mathcal{E}_k} \mathcal{R}_k \mathcal{R}_j \mathcal{R}_i + \mathcal{R}_k \tilde{\mathcal{R}}_{\mathcal{E}_j} \mathcal{R}_j \mathcal{R}_i + \mathcal{R}_k \mathcal{R}_j \tilde{\mathcal{R}}_{\mathcal{E}_i} \mathcal{R}_i | \rho_0 \rangle \rangle. \quad (28)$$

Therefore self-consistent reconstruction of a set of gates can be described as the following least-squares estimation:

$$\min_{\tilde{\mathcal{R}}_{\mathcal{E}_1}, \tilde{\mathcal{R}}_{\mathcal{E}_2}, \dots} \sum_{ijk} |m_{ijk} + \langle \langle M_0 | 2\mathcal{R}_k \mathcal{R}_j \mathcal{R}_i - \tilde{\mathcal{R}}_{\mathcal{E}_k} \mathcal{R}_k \mathcal{R}_j \mathcal{R}_i - \mathcal{R}_k \tilde{\mathcal{R}}_{\mathcal{E}_j} \mathcal{R}_j \mathcal{R}_i - \mathcal{R}_k \mathcal{R}_j \tilde{\mathcal{R}}_{\mathcal{E}_i} \mathcal{R}_i | \rho_0 \rangle \rangle|^2, \quad (29)$$

subject to the constraint that the error maps $\tilde{\mathcal{R}}_{\mathcal{E}_j}$ are all physical.

Minimizing this linearized least-squares fit under the physicality constraint is another example of a semidefinite program. In fact, it is isomorphic to the original QPT problem on a larger space where $\mathcal{R}_\Lambda \rightarrow \mathcal{R}_{\mathcal{E}_1} \oplus \mathcal{R}_{\mathcal{E}_2} \oplus \dots \oplus \mathcal{R}_{\mathcal{E}_N}$. There is, however, one major difference, which is that the optimal fit is no longer unique, which implies that the matrix corresponding to SS^T in this space is not full rank. As an example, take any unitary transformation $G^{(\text{ideal})} = \{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_N\} \rightarrow G^{(\text{err})} = \{\mathcal{R}_U^T \mathcal{R}_1 \mathcal{R}_U, \mathcal{R}_U^T \mathcal{R}_2 \mathcal{R}_U, \dots, \mathcal{R}_U^T \mathcal{R}_N \mathcal{R}_U\}$ where $\mathcal{R}_U |\rho_0\rangle = |\rho_0\rangle$. The measurement outcomes are invariant to such a transformation, since

$$\begin{aligned} m_{ijk} &= \text{Tr}(|\rho_0\rangle \langle \langle M_0 | \mathcal{R}_i \mathcal{R}_j \mathcal{R}_k) \\ &\rightarrow \text{Tr}(|\rho_0\rangle \langle \langle M_0 | \mathcal{R}_U^T \mathcal{R}_i \mathcal{R}_U \mathcal{R}_U^T \mathcal{R}_j \mathcal{R}_U \mathcal{R}_U^T \mathcal{R}_k \mathcal{R}_U) \\ &= \text{Tr}(\mathcal{R}_U |\rho_0\rangle \langle \langle M_0 | \mathcal{R}_U^T \mathcal{R}_i \mathcal{R}_j \mathcal{R}_k) = m_{ijk}. \end{aligned} \quad (30)$$

When we prepare and measure in the computational basis, this is equivalent to a frame invariance with respect to a diagonal unitary transformation. This is not analogous to moving to the interaction picture since the resulting frame is stationary, and does not generate an inertial correction to the Hamiltonian. It is more akin to an energy rescaling. Such a frame is physically irrelevant, and in the following simulations we calculate distance between gate sets by optimizing to find the most generous frame.

In Fig. 4 we compare the results of standard QPT and our self-consistent approach. The error model is that of individual random unitary errors, the uncorrectable error from the previous section, and the gate library is the six-element gate set that maps $|0\rangle$ to the cardinal directions on the Bloch sphere. We plot the errors from both the fidelity and the diamond norm of the estimates versus the error of the physical gate set with respect to the ideal target. For each we show the average result for the six gates. Additionally, we also plot the ratio of the error in the estimate to the ideal for both measures.

In both cases there appears to be an exponential improvement in accuracy over standard QPT with respect to decreasing gate error. This is fundamentally different from the results of the previous section, where the fidelity error changed substantially but the diamond norm remained qualitatively the same, and so this method is not an effective depolarization of the error.

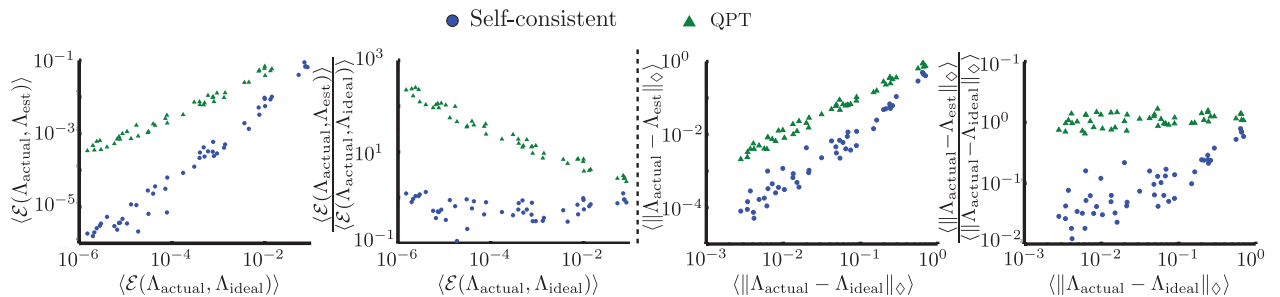


FIG. 4. (Color online) A comparison of standard QPT and the self-consistent method presented in this manuscript for individual random unitary errors for the set of gates mapping to the six cardinal directions. We plot both the diamond norm and fidelity error between the estimates and the actual gates (averaged over the set of gates) plotted versus the equivalent distance of the actual set from the ideal gates. We show the magnitude of the tomography error and the ratio of the error with respect to the actual gate error for both cases.

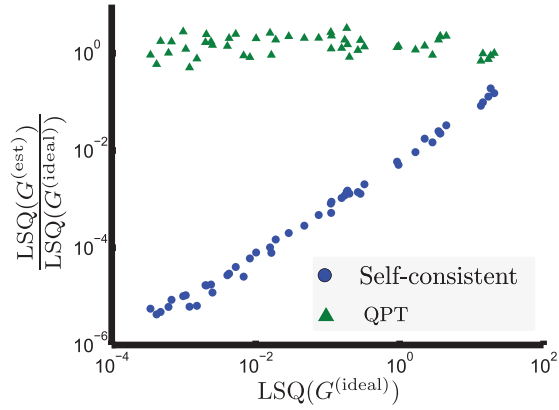


FIG. 5. (Color online) A comparison of least-squares estimator (negative log-likelihood function) for standard QPT and the linearized likelihood estimation method with SPAM errors is generated by random unitary maps of each gate. On the x axis is the estimator of the ideal gate set in the absence of errors with respect to the measurement outcomes and on the y axis the ratio of the estimated gate set's log likelihood to the ideal.

To clarify the differences between the two approaches, we plot the least-squares estimator Eq. (25) for the two estimates in Fig. 5. Again, we look at the ratio of the value of the least-squares fit to the estimated value divided by the fit of the ideal gates to the data. The estimate from QPT is flat with respect to lowering gate error, which corresponds to the fact that QPT is minimizing the wrong quantity. Our estimator does not properly minimize this quantity either, due to the linear approximation, but the scaling is dramatically more favorable.

There are some obvious shortcomings of this method. We have increased the number of experiments and the complexity of the reconstruction algorithm. These increases are both polynomial in the dimension of the system, but since the dimension is already exponential in the number of qubits the resulting overhead is nontrivial. In principle, we can apply these techniques to two- or maybe three-qubit systems, but even by three qubits we are minimizing a $d^6 = 262\,144$ dimensional function.

Additionally, it is numerically hard to reach much smaller errors in our simulation with the default precision of our SDP

solvers (though in this paper we have essentially covered the error range from current to fault tolerant). If we extended the plot in Fig. 5 to the left, we would very quickly see the self-consistent method flatten from the numerical tolerance. The routines for solving semidefinite programs are known to be finicky, and switching solvers from sedum to SDP3 can lead to slightly different answers. This is true not just of self-consistent topography, but of our SDPs to calculate the diamond norm. In fact, calculating the norms, and fitting optimal frames, generally takes longer than the optimization itself.

V. EXPERIMENT

We apply our self-consistent tomography method to data generated from two independent experiments with single-junction transmon (SJT) qubits. Sample A is an SJT capacitively coupled to a coplanar waveguide resonator (used also in Ref. [8] as sample b), while sample B is mounted in a cavity machined from bulk copper. In both cases the qubits are driven by resonant microwaves that are shaped in order to perform optimal qubit gates. For these experiments we use the gate set $\{I, X_\pi, X_{\pi/2}, Y_{\pi/2}\}$, where the notation R_θ denotes a rotation about the axis R of angle θ , and we prepare and measure in the energy eigen-basis (denoted Z). See Refs. [1,32] for the design and construction of these systems.

In sample A(B) the qubit frequency was $\omega_q/2\pi = 4.7610(4.473\,8)$ GHz and was coupled to a resonator of frequency $\omega_r/2\pi = 7.4269(12.118)$ GHz. In sample A there was a second qubit coupled to the resonator with frequency $\omega_q/2\pi = 5.3401$ GHz that played no role in the following experiment since the two qubits were only weakly coupled. The coherence times of the two samples were $T_1 = 8.1(4.0)\,\mu\text{s}$ and $T_2^* = 16.2(7.2)\,\mu\text{s}$. The samples are radiation shielded and cooled to 15 mK in a dilution refrigerator.

In Table I we show the results of both estimation methods as compared to the desired gate. We repeat all measurement ≈ 5000 times in order to have low, but non-negligible, stochastic measurement errors. We can do a similar bootstrapping calculation as in the appendix of [1] to estimate that our statistical errors lead to an uncertainty on the order of 10^{-4} in the fidelity. In both cases we find that QPT overestimates

TABLE I. The gate error from QPT and self-consistent tomography for (A) an SJT capacitively coupled to a coplanar waveguide and (B) an SJT suspended in a copper cavity. We report the fidelity error and diamond norm for both approaches (QPT and self-consistent tomography), and for each gate in the set. The last row in each sample, labeled $\langle U \rangle$, is the average of the previous four as well as the average error obtained from randomized benchmarking.

Sample	Gate	$1 - F_g(\Lambda_{\text{QPT}}, \Lambda_{\text{ideal}})$	$1 - F_g(\Lambda_{\text{SC}}, \Lambda_{\text{ideal}})$	$1 - F_g(\text{RB})$	$\ \Lambda_{\text{QPT}} - \Lambda_{\text{ideal}}\ _\diamond$	$\ \Lambda_{\text{SC}} - \Lambda_{\text{ideal}}\ _\diamond$
A	I	0.0010	0.0013	—	0.020	0.024
A	X_π	0.0083	0.0048	—	0.062	0.044
A	$X_{\pi/2}$	0.0030	0.00023	—	0.038	0.026
A	$Y_{\pi/2}$	0.0050	0.0047	—	0.051	0.069
A	$\langle U \rangle$	0.0043	0.0015	0.0029	0.043	0.041
B	I	0.11	0.0237	—	0.50	0.081
B	X_π	0.0276	0.0013	—	0.26	0.035
B	$X_{\pi/2}$	0.0266	0.016	—	0.31	0.096
B	$Y_{\pi/2}$	0.0181	0.023	—	0.20	0.078
B	$\langle U \rangle$	0.045	0.016	0.0016	0.31	0.072

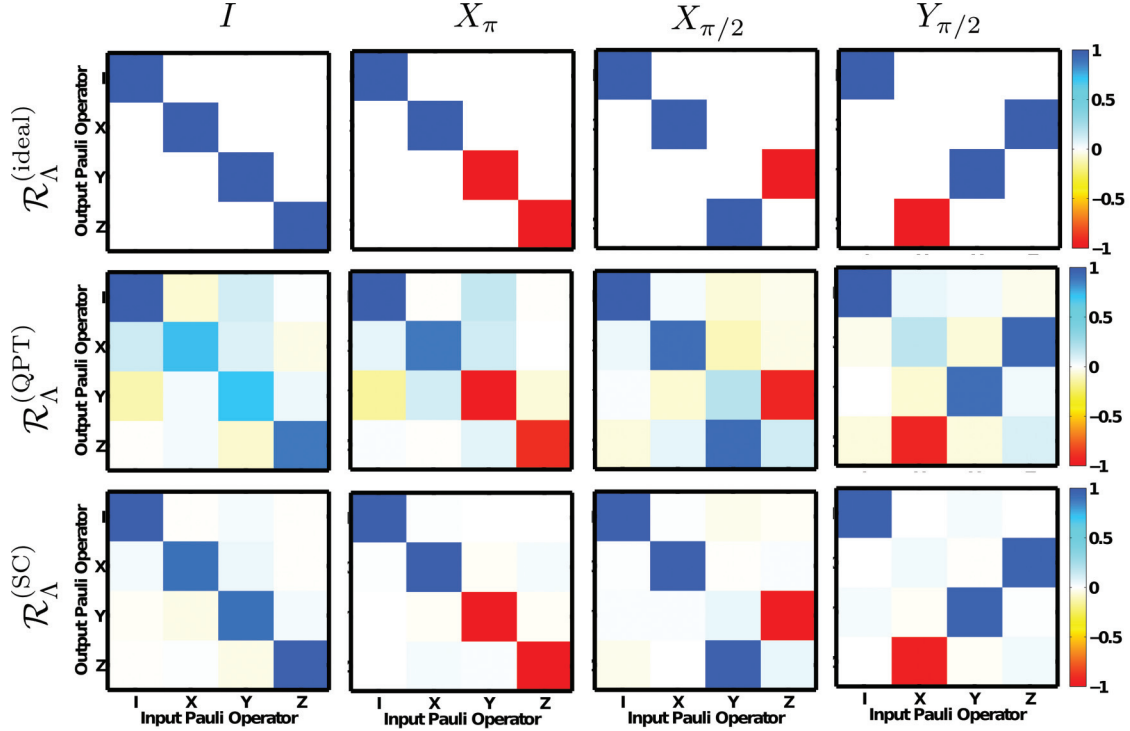


FIG. 6. (Color online) The \mathcal{R} matrices for the ideal gates and the reconstruction from QPT and self-consistent tomography for each of the four measured gates on sample B.

the error in both the diamond norm and the gate fidelity as compared to the self-consistent method, well outside of this uncertainty bound. For sample A the error on the QPT was roughly 1.5 times the self-consistent error, while for sample B it was roughly 3 times as large. We show plots of the Pauli transfer matrices for both estimates and the ideal maps in Fig. 6 for sample B.

Furthermore, since the gates in our library were generators of the Clifford group, we were able to perform randomized benchmarking on the two samples using the methods in [19,20]. The fidelity error from randomized benchmarking was even smaller than the error from the self-consistent reconstruction, dramatically so in the case of sample B. We conjecture that this is due to errors in the measurement operator M_0 , since the calibration of the measurement voltage to the $|0\rangle$ and $|1\rangle$ states has been observed to shift over the course of long experimental runs. Randomized benchmarking is insensitive to this type of error while the methods presented in this manuscript are not. Accounting for errors in the initial state ρ_0 and the fixed observable M_0 are both important future avenues to extend this method, as well as that of dealing with slowly time-varying errors.

VI. CONCLUSION

The methods of tomography are very powerful when using a well-characterized system to probe a black-box operation, but this is rarely the case when verifying quantum computing hardware. Instead, the situation is one in which all components are faulty, and in that scenario we have shown that process tomography can fail dramatically. Instead of separating the preparation and measurement phases of tomography, we

instead must view the experiment as a sequence of faulty gates, and while we have not discussed errors on the fiducial state and fixed measurement operator, those, too, should probably be brought into question in the future. Without incorporating these errors into our model, it is impossible to obtain a more accurate reconstruction through statistical analysis alone.

In this manuscript we have developed a protocol that self-consistently reconstructs a library of quantum gates by modifying the likelihood function of quantum process tomography to incorporate our uncertainty of the state preparation and measurement gates. In simulation, this self-consistent method outperforms standard QPT in terms of the accuracy of estimation, since QPT consistently underestimates gate fidelities in the presence of SPAM errors. This method requires the same number of experiments as QPT and only adds polynomial overhead to the amount of classical postprocessing. In fact, in any case where tomography has been performed on the entire set of gates used for state preparation and measurement, no further experiments are needed to apply our protocol. Colloquially, we have taken to calling this the “overkill” method of tomography, since the number of the experiments borders on the ridiculous for all but one- or two-qubit systems, but this will be the case for any protocol that provides full information about a library of maps. Self-consistently reconstructing a library of gates, as opposed to performing tomography on the individual members, yields more trustworthy and generally higher fidelity estimates in both experiment and simulation, and we expect the difference between standard QPT and the self-consistent method presented here to grow more pronounced as we reach lower and lower gate errors.

ACKNOWLEDGMENTS

We acknowledge contributions from J. Chow and Thomas Ohki and discussions with R. Blume-Kohout and Graeme Smith. We acknowledge support from IARPA under Contract

No. W911NF-10-1-0324. All statements of fact, opinion, or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the US Government.

-
- [1] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, *Phys. Rev. Lett.* **109**, 060501 (2012).
 - [2] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *Nature (London)* **460**, 240 (2009).
 - [3] R. C. Bialczak, M. Ansmann, M. Hofheinz, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, J. Wenner, M. Steffen, A. N. Cleland, and J. M. Martinis, *Nat. Phys.* **6**, 409 (2010).
 - [4] H. Haffner, W. Hansel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Korber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Guhne, W. Dur, and R. Blatt, *Nature (London)* **438**, 643 (2005).
 - [5] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, *Phys. Rev. Lett.* **108**, 260503 (2012).
 - [6] M. A. Nielsen, *Phys. Lett. A* **303**, 249 (2002).
 - [7] I. L. Chuang and M. Nielsen, *J. Mod. Opt.* **44**, 2455 (1997).
 - [8] J. M. Gambetta, A. D. Corcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, *Phys. Rev. Lett.* **109**, 240504 (2012).
 - [9] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, *Phys. Rev. A* **64**, 052312 (2001).
 - [10] Z. Hradil, J. Rehacek, J. Fiurasek, and M. Jezek, *Lect. Notes Phys.* **649**, 59 (2004).
 - [11] A. I. Lvovsky, *J. Opt. B: Quantum Semiclassical Opt.* **6**, S556 (2004).
 - [12] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, *Phys. Rev. A* **78**, 052122 (2008).
 - [13] R. Blume-Kohout, *Phys. Rev. Lett.* **105**, 200504 (2010).
 - [14] R. Blume-Kohout, *New J. Phys.* **12**, 043034 (2010).
 - [15] J. A. Smolin, J. M. Gambetta, and G. Smith, *Phys. Rev. Lett.* **108**, 070502 (2012).
 - [16] M. Christandl and R. Renner, *Phys. Rev. Lett.* **109**, 120403 (2012).
 - [17] D. Rosset, R. Ferretti-Schöbitz, J. Bancal, N. Gisin, and Y. Liang, *arXiv:1203.0911*.
 - [18] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Phys. Rev. A* **77**, 012307 (2008).
 - [19] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. Lett.* **106**, 180504 (2011).
 - [20] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. A* **85**, 042311 (2012).
 - [21] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, *Phys. Rev. Lett.* **109**, 080505 (2012).
 - [22] A. M. Braczyk, D. H. Mahler, L. A. Rozema, A. Darabi, A. M. Steinberg, and D. F. V. James, *New J. Phys.* **14**, 085003 (2012).
 - [23] V. V. Dobrovitski, G. de Lange, D. Ristè, and R. Hanson, *Phys. Rev. Lett.* **105**, 077601 (2010).
 - [24] M.-D. Choi, *Linear Algebra and its Applications* **10**, 285 (1975).
 - [25] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
 - [26] J. F. Sturm, *Optim. Methods Software* **11**, 625 (1999).
 - [27] A. Silberfarb, P. S. Jessen, and I. H. Deutsch, *Phys. Rev. Lett.* **95**, 030402 (2005).
 - [28] C. A. Riofro, P. S. Jessen, and I. H. Deutsch, *J. Phys. B* **44**, 154007 (2011).
 - [29] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Washington, DC, 2002).
 - [30] J. Watrous, *Theory of Computing* **5**, 11 (2009).
 - [31] C. Dankert, R. Cleve, J. Emerson, and E. Livine, *Phys. Rev. A* **80**, 012304 (2009).
 - [32] C. Rigetti, J. M. Gambetta, S. Poletto, B. L. T. Plourde, J. M. Chow, A. D. Córcoles, J. A. Smolin, S. T. Merkel, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, *Phys. Rev. B* **86**, 100506 (2012).