Detector tomography on IBM quantum computers and mitigation of an imperfect measurement

Yanzhu Chen,[1,2,3] Maziar Farahzad,[2] Shinjae Yoo,[4] and Tzu-Chieh Wei[1,2,3]

[1]*C. N. Yang Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3840, USA*
[2]*Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, New York 11794-3800, USA*
[3]*Institute for Advanced Computational Science, State University of New York at Stony Brook, Stony Brook, New York 11794-5250, USA*
[4]*Computational Science Initiative, Brookhaven National Laboratory, P.O. Box 5000, Upton, New York 11973-5000, USA*

(Received 6 May 2019; revised manuscript received 15 September 2019; published 13 November 2019)

We use quantum detector tomography to characterize the qubit readout in terms of measurement positive operator-valued measures (POVMs) on IBM quantum computers IBM Q 5 Tenerife and IBM Q 5 Yorktown. Our results suggest that the characterized detector model deviates from the ideal projectors, ranging from 10 to 40%. This is mostly dominated by classical errors, evident from the shrinkage of arrows from the poles in the corresponding Bloch-vector representations. There are also small deviations that are not "classical," of order 3% or less, represented by the tilt of the arrows from the z axis. Further improvement on this characterization can be made by adopting two- or more-qubit detector models instead of independent single-qubit detectors for all the qubits in one device. We also find evidence indicating correlations in the detector behavior, i.e., the detector characterization is slightly altered (to a few percent) when other qubits and their detectors are in operation. Such peculiar behavior is consistent with characterization from the more sophisticated approach of the gate set tomography. We also discuss how the characterized detectors' POVMs, despite deviation from the ideal projectors, can be used to estimate the ideal detection distribution.

DOI: 10.1103/PhysRevA.100.052315

I. INTRODUCTION

Recent manufacture of 49, 50, and 72 superconducting qubits from companies such as Intel, IBM, and Google gives the prospect of demonstrating quantum advantage in the not distant future. However, these and near-future machines are at best noisy intermediate-scale quantum processors [1]. Therefore, developing and harnessing tools for characterizing noise and error, mitigating them, and verifying quantum processing will be essential in running programs on quantum devices and in further locating the parameter windows in applications towards quantum advantage. Several tools have been developed, including tomographic ones for *quantum states and processes* [2–8]. But these rely on accurate measurement and/or state preparation, which the system may not have, and the methods do not scale favorably with the system size. If one is only concerned with partial characterization, such as the average gate error rate, then the so-called randomized benchmarking [9–11] provides a reliable estimation independent of state preparation and measurement error. Although these tools seem to be standard, there are still some aspects of them not fully explored.

When one speaks of qubit decoherence, there are typically two associated processes: (1) relaxation (with time T_1), usually related to the transition of the excited state(s) back to the ground state or the system returning to thermal equilibrium, and (2) dephasing (with time T_2), related to off-diagonal elements of the density matrix decaying exponentially with time. In reality, a qubit will couple to the environment and such interaction (and with other qubits in an undesired way) will induce relaxation and dephasing, and possibly other forms of decoherence. These will be loosely referred to as noise, and any quantum gate that does not operate as desired is said to have errors. For instance, in the IBM quantum computers, the error rate in measurement readout (2–10%) is comparable to that of two-qubit gates (3–7%) and both rates are greater than that of single-qubit gates (0.1–0.2%) by one order of magnitude. Single-qubit state preparation for short circuits is to some extent of high fidelity, but the computation for longer circuits will inevitably suffer from noise. One tomographic tool that sometimes gets overlooked is the so-called quantum detector tomography (QDT) [12], more recently discussed in photon detectors [13–16], which seems to provide a first tool to improve the readout or detector characterization, via short quantum circuits involving single-qubit gates. Because the measurement error is higher than state preparation (of $|0\rangle$) and single-qubit gates on IBM Q devices, we perform quantum detector tomography to characterize the detectors. The experiments are done via the IBM cloud computing under the framework Qiskit [17]. We point out some behavior revealed by experiments that require further investigation into physical devices, beyond the setting of quantum circuits.

We remark that a more thorough characterization scheme that makes the fewest assumptions is *gate set tomography* (GST) [18,19], where an initial state, a set of quantum gates, and a positive operator-valued measure (POVM) are characterized simultaneously. Such characterization can provide information for error mitigation purpose [20]. Since GST requires a large number of gate sequences, some of which are very long, it is currently limited to single-qubit and two-qubit processes in practice. Another recently proposed scheme that is less costly tries to characterize state preparation and measurement iteratively [21].

In the next section (Sec. II) we briefly review the tomographic tool for detectors. We present our experimental results in Sec. III. An unexpected behavior was seen in the physical

 ©2019 American Physical Society

qubit labeled as qubit 3 of IBM Q 5 Tenerife. Its detector characterization seems to be different when it is done alone (with other qubits being idle) from when it is done when other qubits are also in operation. This can be a consequence of detector crosstalk or qubit operations influencing each other and requires further investigation into the physical process. In Sec. IV, we describe how such characterized readout can be used for mitigation of measurement error, in the sense of inferring ideal measurement statistics. In Sec. V, we also use the gate set tomography and compare its detector characterization with that from the simple detector tomography. We make concluding remarks in Sec. VI. Some experimental data and further results of GST are presented in the Appendices, including QDT for the 14 qubits of IBM Q 16 Melbourne.

II. TOMOGRAPHIC TOOL FOR DETECTORS

A. Quantum detector tomography

In this section we review the tool of quantum detector tomography [12], which was often studied for photon detectors [13–16]. A short review of quantum state tomography (QST) and quantum process tomography (QPT) is included in the Appendices for completeness. In QST one has a set of projectors or more general POVM elements $\{\Pi^{(i)}\}$ (e.g., $\mathcal{A}_1 \equiv \{|0/1\rangle\langle 0/1|, |+/-\rangle\langle +/-|, |\pm i\rangle\langle \pm i|\}$, corresponding to eigenstates of Pauli matrices), and measures them with respect to an unknown state ρ, yielding a set of data $p_{\rho,i} = \mathrm{Tr}(\rho\Pi^{(i)})$. QDT is a dual viewpoint: with a set of known states $\{\rho_f\}$, one is asked to estimate a fixed but unknown set of measurement operators $\{\Pi^{(i)}\}$ characterizing a detector. Here we formulate qubit detectors that are most relevant to realistic measurement in cloud quantum computers, such as IBM Q and Rigetti's. The usual assumption is that the set of states $\{\rho_f\}$'s is well known or at least with much smaller error rates than detection. For the state preparation in $|0\rangle$, the typical ground state of superconducting qubits, it is fairly accurate. Moreover, in these systems the single-qubit gates have higher fidelity (than the measurement and two-qubit gates), and only Z measurement can be implemented. Measurement in other bases needs to be actively made by the users to perform a suitable rotation before the Z measurement. Hence we will consider two measurement operators $\Pi^{(0)}$ and $\Pi^{(1)}$ for a single qubit, which is constrained by the trace-preserving condition that $\Pi_0 + \Pi_1 = \mathbb{1}$. In the ideal case, $|0\rangle\langle 0| = (\mathbb{1} + \sigma_3)/2$ and $|1\rangle\langle 1| = (\mathbb{1} - \sigma_3)/2$.

Let us denote for convenience

$$\Pi_1^{(n)} = \sum_{i=0}^{3} a_i^{(n)} \sigma_i, \tag{1}$$

where the subscript of $\Pi_1^{(n)}$ means the single-qubit detector and (n) denotes the measurement outcome 0 or 1. The Pauli basis is $\sigma_0 = \mathbb{1}, \sigma_1 = \sigma_x, \sigma_2 = \sigma_y, \sigma_3 = \sigma_z$. We can use a vector $\vec{a}^{(n)} = (a_0^{(n)}, a_1^{(n)}, a_2^{(n)}, a_3^{(n)})$ to collectively denote the parameters. There are some constraints: (1) $\vec{a}^{(0)} + \vec{a}^{(1)} = (1, 0, 0, 0)$ due to completeness and (2) $|a_0^{(n)}|^2 \geqslant \sum_{i=1}^{3} |a_i^{(n)}|^2$ in order for $\Pi^{(n)}$ to be non-negative. We choose and prepare ρ from the six-element set $\mathcal{A}_1$ (listed above), and those other than $|0\rangle$ can be prepared from it with relatively high fidelity by single-qubit gates. Then the measurement process accumulates a set of data $P_{\rho_i,n} = \mathrm{Tr}(\rho_i\Pi_n)$, which is a 6×2 matrix for each detector. From this we can find the best fit, under the above constraints, to extract $\vec{a}^{(n)}$ that describes the action of the detector. In adopting this model, we have made the assumption that there is no crosstalk between the qubits in one device so that the detectors are viewed as independent. Relaxing this assumption a little bit, we can have a multiqubit detector model, where a binary string is produced as the measurement result. Just like the single-qubit case, the N-qubit detector model is written as

$$\Pi_N^{(\vec{n})} = \sum_{\vec{i}} c_{\vec{i}}^{(\vec{n})} \sigma_{i_0} \otimes \ldots \otimes \sigma_{i_j} \ldots \otimes \sigma_{i_{N-1}}, \tag{2}$$

where the binary string $\vec{n} = (n_0, \ldots, n_{N-1})$ is the measurement outcome and each component of $\vec{i} = (i_0, \ldots, i_{N-1})$ runs from 0 to 3. It is natural to ask whether measuring only one single qubit in the device gives the same result as measuring all the qubits and tracing out the other irrelevant ones. This question will be addressed in our experiments.

In an experiment, the characterized detectors can be used to perform QST on the resultant state, hence mitigating the effect from detector errors. However, we remark that some correction can be made even without an informationally complete set of measurements on the state; see discussions below in Sec. IV.

B. Maximum likelihood estimation

In this section we summarize the maximum likelihood estimation (MLE) analysis for detectors that we will employ [12]. The log likelihood function is defined as

$$\ln \mathcal{L} = \sum_n \sum_i f_{n,i} \ln \mathrm{Tr}(\Pi^{(n)}\rho_i), \tag{3}$$

where $\{\Pi^{(n)}\}$ is the POVM characterizing the detector and $f_{n,i}$ is the frequency of measuring the state ρ_i and obtaining outcome n. The sum over index i contains an informationally complete set of test states. The normalization constraint $\sum_n \Pi^{(n)} = \mathbb{1}$ is implemented by Lagrange multipliers. Maximization with the constraint leads to the equation

$$\Pi^{(n)} = R^{(n)}\Pi^{(n)}R^{\dagger(n)}. \tag{4}$$

$R^{(n)}$ is determined by the normalization constraint, and is given by

$$R^{(n)} = \sum_i \frac{f_{n,i}}{p_{n,i}} \left(\sum_m \sum_{j,k} \frac{f_{m,j}f_{m,k}}{p_{m,j}p_{m,k}} \rho_j \Pi^{(m)} \rho_k \right)^{-\frac{1}{2}} \rho_i, \tag{5}$$

where $p_{m,j}$ denotes the theoretical probability of measuring the state ρ_j and obtaining outcome m. Note that $R^{(n)}$ is a function of the POVM $\{\Pi^{(m)}\}$, not only through the explicit dependence but also because $p_{m,j} = \mathrm{Tr}(\Pi^{(m)}\rho_j)$. In our analysis, we choose the multiqubit Pauli matrices as the basis to express $\{\Pi^{(m)}\}$ and $\{\rho_j\}$. Each iteration starts with updating $\{\Pi^{(m)}\}$ according to Eq. (4), and ends with calculating $\{R^{(m)}\}$ from Eq. (5) for the next iteration. The termination condition

is set as

$$\sum_n \left|\left|\Pi_t^{(n)} - \Pi_{t+1}^{(n)}\right|\right| < \epsilon, \tag{6}$$

where the subscript denotes the tth and the $(t+1)$th iterations, the norm is taken to be the Frobenius norm, and ϵ is some (arbitrarily chosen) cutoff value. Positivity and normalization are preserved as long as the initial values of $\{\Pi^{(m)}\}$ form a POVM. It is worth mentioning that ϵ should be sufficiently small such that the numerical error introduced by this cutoff would be smaller than the uncertainty in the estimated parameters due to statistical fluctuations.

III. RESULTS OF QUANTUM DETECTOR TOMOGRAPHY

We performed QDT on the two IBM Q 5 devices, Tenerife (ibmqx4) and Yorktown (ibmqx2), and present the results below. Each experiment consists of 100 runs with the setting of 8192 shots for each run on the IBM Q devices. The effective total number of repetitions is 819 200. This gives us effectively 819 200 shots. The test states in $\mathcal{A}_1$ were prepared by first initializing the qubits in $|0\rangle$ (which is the ground state of each qubit) and acting on it by the single-qubit gates Pauli X, Hadamard H, and the S gates, as well as their combinations. The MLE [12] was used for calculating the POVM parameters from measured frequencies, reviewed earlier. The positivity is ensured by construction, e.g., using $\frac{1}{2^N}\mathbb{1}$ as the initial POVM elements for the iteration.

First we adopted the single-qubit detector model. One can carry out the detector tomography procedure for each physical qubit individually, leaving the other qubits in the machine idle, or simultaneously carry out the same procedure for all qubits (or a subset of them). We henceforth refer to these two different ways as "individual measurement" and "parallel measurement," respectively. In principle there should not be any difference except that due to statistical fluctuations between the two, since using the single-qubit detector model we have assumed independence of the qubits. However, in reality we see significant discrepancy between the results obtained from the two types of experiments, which we will describe below.

A next-step generalization would be to adopt the two-qubit detector model. We examined all pairs of qubits in the two machines, and compared the results with those obtained for the single-qubit detector model (both individual and parallel). If the discrepancy we observed is solely due to pairwise influence, this would be captured in the two-qubit detector tomography. However, this is not the case, as we will see in Sec. III B. One can readily generalize this to detector models involving three or more qubits as in Eq. (2). For the five-qubit devices, a five-qubit detector model will be the best to characterize the measurement for the two five-qubit IBM machines. In order to obtain all $2^5 = 32$ operators $\Pi^{(\vec{n})}$ using the aforementioned basis states, $6^5 = 7776$ circuits are required. But some kind of compressed sensing technique may be used to mitigate this, as was done for QST [22]. We would like to point out that to run this list of circuits on the current devices it needs to be separated into smaller lists of jobs, since there is an upper limit on the circuit count for one single submitted job.

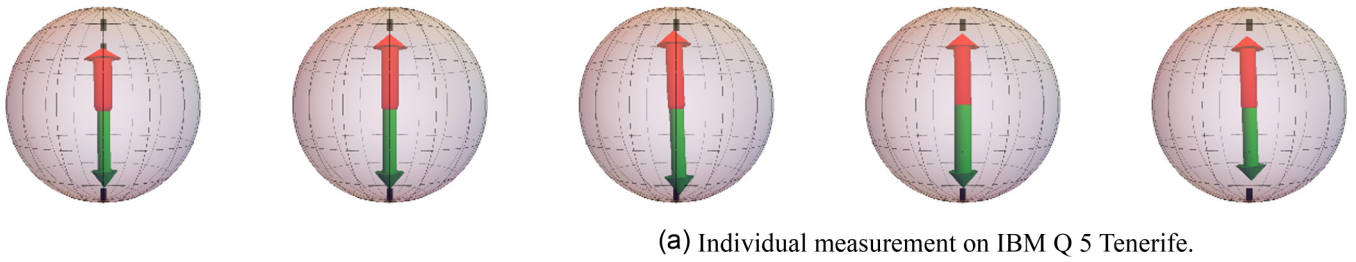
A. Single-qubit detector: Parallel vs individual

The results of QDT are visualized in Fig. 1 using Bloch spheres, with detailed parameters listed in Tables V(a), V(b), VI(a), and VI(b). The three-dimensional arrow represents the vector $\vec{r} = (a_1, a_2, a_3)/a_0$, and should be $(0, 0, \pm 1)$ for ideal detectors $\Pi^{0/1} = (\mathbb{1} \pm \sigma_z)/2$. We use the thickness of the arrow to represent the parameter a_0. Each detector is found to have its axis align mostly with the z axis but behave with some notable difference to the ideal 0 and 1 projectors: (1) shrinkage of the arrows from poles, where the lengths of the arrows represent the purity of the measurement, and the shorter they are the further away are the detectors from perfect projection; (2) tilt of the arrows, where the measurement axis is slightly tilted from the Pauli z basis, but it is only of a few percent or less; and (3) thickness of arrows, which represents the amount of $a_0^{(0)}$ and $a_0^{(1)} = 1 - a_0^{(0)}$. Intuitively, we can consider the conditional probability of measuring 0 given that the state is 0, $P(0|0) = a_0^{(0)} + a_3^{(0)}$, and the conditional probability of measuring 0 given that the state is 1, $P(0|1) = a_0^{(0)} - a_3^{(0)}$ and similarly $P(1|0) = a_0^{(1)} + a_3^{(1)} = 1 - a_0^{(0)} - a_3^{(0)}$, and $P(1|1) = 1 - a_0^{(0)} + a_3^{(0)}$. Given that the tilt is small, the detector errors are mostly classical flips. From the above relations, we have $2a_0^{(0)} = P(0|0) + P(0|1)$ and $2a_0^{(1)} = P(1|0) + P(1|1)$. Given that $|0\rangle$ is the ground state in each qubit, we expect that $a_0^{(0)} \geqslant a_0^{(1)}$. Indeed most of the detectors do satisfy this regardless of the schemes of measurement, parallel or individual, except qubit 3 in IBM Q 5 Tenerife measured in parallel. These features are displayed in Fig. 1.

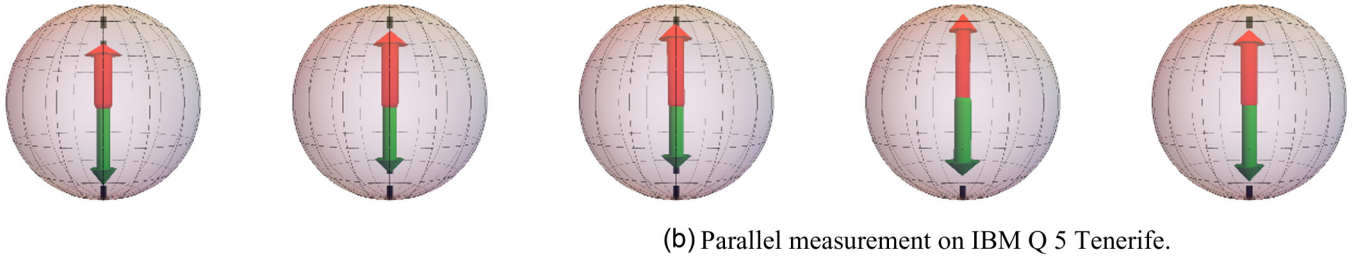
The details of the single-qubit detector results for the two devices, as measured individually for each physical qubit leaving the other qubits idle, are shown in Tables V(a) and VI(a), respectively. And those obtained by carrying out single-qubit detector tomography simultaneously for all five qubits in the machine are presented in Tables V(b) and VI(b). We use the nonparametric bootstrap method [19] to estimate the uncertainty for each parameter, which gives errors typically of the order $O(10^{-4})$, with the largest among them up to 0.003. A detailed discussion is in Sec. III C below. We note that Fig. 5 corresponds to data in Table V, and Fig. 6 corresponds to data in Table VI.

A notable feature is found that for almost all qubits $a_0^{(0)}$ is larger than $a_0^{(1)} = 1 - a_0^{(0)}$, which comes from relaxation to the ground state $|0\rangle$. There is an exception for qubit 3 of the device IBM Q 5 Tenerife, where $a_0^{(0)} < a_0^{(1)}$ when measured together with the other qubits in parallel. This was not seen when qubit 3 was measured alone, which hints at influence from the other qubits.

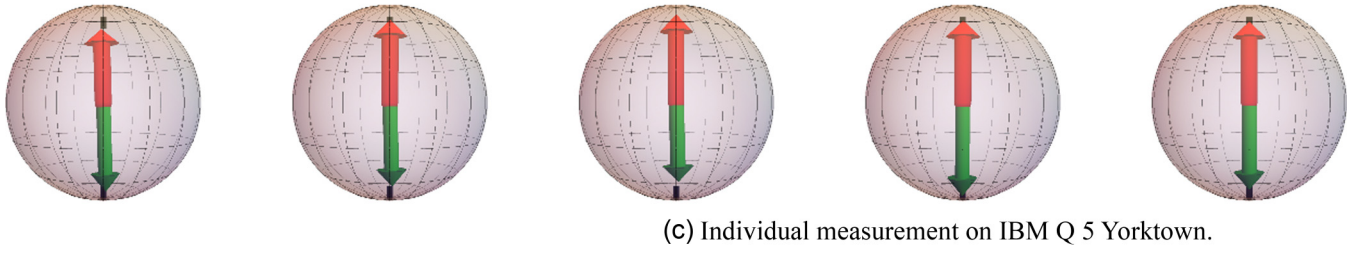
A measure of discrepancy between individual measurement and parallel measurement is the distance between the vectors $\vec{a}^{(0)} = (a_0^{(0)}, a_1^{(0)}, a_2^{(0)}, a_3^{(0)})$ [note that $\vec{a}^{(1)} = (1, 0, 0, 0) - \vec{a}^{(0)}$] obtained in the two different ways. This distance corresponds to the Frobenius norm of the difference between the two $\Pi^{(0)}$ operators up to a factor of 2. These are shown in Table I. It is worth noticing that the statistical fluctuation in the estimated $\vec{a}^{(0)}$ limits the resolution of the distance to an order $O(10^{-3})$. We distinguish this order of magnitude from the magnitude of the fluctuations in the parameters of the detector model, which are typically $O(10^{-4})$. This



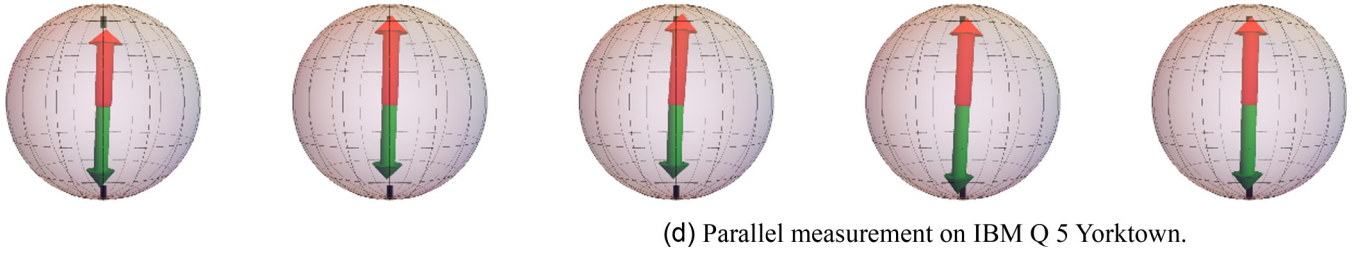
(a) Individual measurement on IBM Q 5 Tenerife.



(b) Parallel measurement on IBM Q 5 Tenerife.



(c) Individual measurement on IBM Q 5 Yorktown.



(d) Parallel measurement on IBM Q 5 Yorktown.

FIG. 1. Detector spheres for qubits 0 to 4 of (a, b) IBM Q 5 Tenerife and (c, d) IBM Q 5 Yorktown. The arrow represents the vector $(a_1, a_2, a_3)/a_0$ from measurement $\tilde{\Pi}^{(n=0,1)} = \vec{a}^{(n)} \cdot \vec{\sigma}$, where the north pole and the south pole correspond to ideal $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, respectively. Positivity is reflected by the length of the arrow being smaller than 1. The width of the arrow represents the weight a_0 in the corresponding POVM element, for which the ideal case is 1/2. These data are in Tables V and VI. For example, individual measurement of qubit 4 of IBM Q 5 Tenerife produces the last sphere in (a), which corresponds to $\Pi^{(0)} = 0.521(1)\mathbb{1} - 0.012(2)\sigma_x - 0.0122(4)\sigma_y + 0.3798(4)\sigma_z$ and $\Pi^{(1)} = 0.479(1)\mathbb{1} + 0.012(2)\sigma_x + 0.0122(4)\sigma_y - 0.3798(4)\sigma_z$.

distance is derived from those parameters and the estimated fluctuations in the distances originate from, but are of lager magnitude than, the fluctuations in the detector parameters. We see that the distance between the two $\vec{a}^{(0)}$ vectors obtained from individual measurement and parallel measurement is one order of magnitude larger, which indicates that there is some correlation due to several qubits being operated and measured simultaneously, visible even in the presence of statistical fluctuations. These differences are visible from Bloch spheres between first and second rows and between third and fourth rows in Fig. 1.

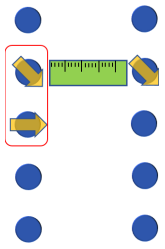
TABLE I. Distance between the single-qubit detector from individual measurement and that from parallel measurement, for IBM Q 5 Tenerife (ibmqx4) and IBM Q 5 Yorktown (ibmqx2).

Device \ Qubit	0	1	2	3	4
ibmqx4	0.011	0.010	0.023	0.087	0.025
ibmqx2	0.042	0.017	0.044	0.031	0.024

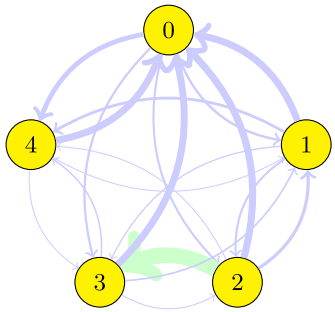
B. Beyond the single-qubit detector

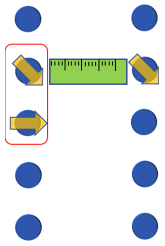
Two-qubit QDT and cross talk

The two-qubit detector model $\Pi_2^{(n_0,n_1)}$ for a pair of qubits is characterized by 64 parameters, which can be organized into four 4×4 matrices $c_{i,j}^{(n_0,n_1)}$ for the four outcomes $(n_0, n_1) =$ (00), (01), (10), and (11), respectively. Imagine we have two uncorrelated systems A and B, where the POVM for the composition is $\{\Pi_{AB}^{(n_A,n_B)} = \Pi_A^{(n_A)} \otimes \Pi_B^{(n_B)}\}$. $\sum_{n_A} \Pi_A^{(n_A)} = \mathbb{1}_A$ and



qubit of interest \ qubit traced out	0	1	2	3	4
0	-	0.086	0.088	0.086	0.086
1	0.030	-	0.040	0.021	0.012
2	0.026	0.026	-	0.013	0.009
3	0.028	0.016	0.195	-	0.012
4	0.060	0.035	0.012	0.018	-





qubit of interest \ qubit traced out	0	1	2	3	4
0	-	0.032	0.044	0.028	0.032
1	0.020	-	0.020	0.022	0.026
2	0.017	0.011	-	0.019	0.008
3	0.003	0.018	0.014	-	0.080
4	0.004	0.026	0.024	0.027	-

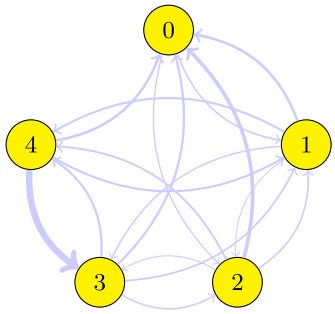


FIG. 2. The left panels explain the comparison scheme: the single-qubit detector on the left is obtained by tracing out another qubit in a two-qubit detector while that on the right is from individual measurement. The tables contain the distances between these two: the entry in the ith row and jth column is the distance between the single-qubit detector of qubit i conditioned on qubit j and that for qubit i obtained from individual measurement. Values between 0.01 and 0.1 are highlighted in blue and those above 0.1 are highlighted in green. The right panels are schematic influence diagrams where each arrow points from the qubit traced out to the qubit of interest and the thickness is proportional to the distance. This comparison is shown for IBM Q 5 Tenerife (upper panels) and IBM Q 5 Yorktown (bottom panels), respectively.

the same condition for B are satisfied independently. When we have no access to system B, we need to sum over all possible outcomes for B to get $\{\Pi_A^{(n_A)} \otimes \mathbb{1}_B = \sum_{n_B} \Pi_{AB}^{(n_A,n_B)}\}$. We can then take the partial trace over B to recover

$$\Pi_A^{(n_A)} = \frac{1}{\dim(B)} \mathrm{Tr}_B \sum_{n_B} \Pi_{AB}^{(n_A,n_B)}, \tag{7}$$

where $\dim(B)$ is the dimension of the Hilbert space for B. In doing so we are assuming that any state of B is equally likely to occur, i.e., no information about B is accessible. To check whether a pair of qubits is separable, we can calculate the singular values of the matrices $c_{i,j}^{(n_0,n_1)}$ for them. If the single-qubit detector assumption holds well, the operators can be decomposed in the following way:

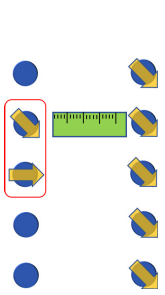
$$\Pi_2^{(n_0,n_1)} = \Pi_1^{(n_0)} \otimes \Pi_1^{(n_1)}, \tag{8}$$

where $\Pi_1^{(n)}$ is a single-qubit detector operator. In this case there will be only one nonzero singular value for any of the four $c_{i,j}^{(n_0,n_1)}$ matrices. This is a direct analogy to characterization of the entanglement of a bipartite system. We can also calculate from the singular values the analogy of entanglement measures, the magnitudes of which give a measure of how badly the assumption of independent single-qubit detectors is violated. We will not present detailed analysis about this here.

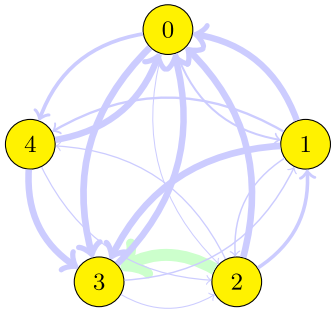
How do we characterize a single-qubit detector reduced from the detector in the presence of other qubits? From the two-qubit detector model [see Eq. (2) applied to two qubits], one can trace out one qubit and obtain a single-qubit detector model for the other qubit. For example, tracing out the second qubit in a pair, we get a single-qubit detector for the first qubit according to

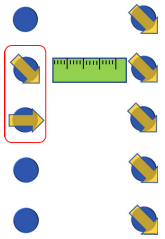
$$\Pi_1^{(0)} = \tfrac{1}{2}\mathrm{Tr}_{2\mathrm{nd}}\big(\Pi_2^{(00)} + \Pi_2^{(01)}\big), \tag{9}$$

where the trace is taken over the second qubit only. Note that by doing this $\Pi_1^{(0)} + \Pi_1^{(1)} = \mathbb{1}_{\mathbb{C}^2}$ is an automatic consequence of $\sum_{n_0,n_1} \Pi_2^{(n_0,n_1)} = \mathbb{1}_{\mathbb{C}^2\times\mathbb{C}^2}$. We call the single-qubit detector obtained by tracing out another qubit in a pair "a single-qubit detector conditioned on another qubit." We calculate such models for each qubit conditioned on any of the other qubits, and compare the result to the single-qubit detector obtained from individual measurement, parallel measurement, and parallel measurement of only those two qubits (henceforth referred to as "pairwise parallel measurement"). In the ideal case where all detectors are independent of each other, these results should agree within statistical uncertainty. If there is influence of only one other qubit on a given qubit, then we expect one of the four conditional single-qubit detector results to coincide with the result obtained from parallel measurement. From this we can also find which qubit is affecting a given qubit. Again we use the distance between two $\vec{a}^{(0)}$ vectors to characterize the agreement between two results. We present the comparison between the single-qubit detector from individual measurement and that conditioned on another qubit in Fig. 2, and leave the other two comparison schemes in the Appendices (Figs. 3 and 4). In these figures we also use schematic diagrams to visualize the crosstalk. In the influence diagrams, the arrows are of thicknesses proportional to the distance between the detectors obtained in the two different ways and point from the qubit traced out to the qubit of interest, indicating the influence of the former to the latter. As



qubit of interest \ qubit traced out	0	1	2	3	4
0	-	0.079	0.082	0.080	0.080
1	0.025	-	0.044	0.018	0.008
2	0.010	0.018	-	0.013	0.015
3	0.083	0.083	0.149	-	0.081
4	0.046	0.029	0.016	0.007	-





qubit of interest \ qubit traced out	0	1	2	3	4
0	-	0.014	0.034	0.034	0.037
1	0.009	-	0.013	0.018	0.022
2	0.029	0.039	-	0.029	0.037
3	0.034	0.040	0.018	-	0.053
4	0.024	0.018	0.010	0.008	-

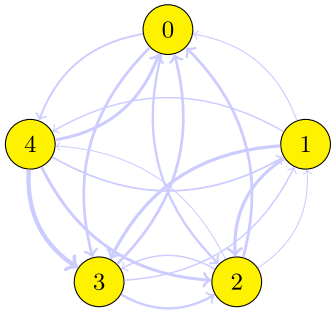


FIG. 3. The left panels explain the comparison scheme: the single-qubit detector on the left is obtained by tracing out another qubit in a two-qubit detector while that on the right is from parallel measurement. The tables contain the distances between these two: the entry in the ith row and jth column is the distance between the single-qubit detector of qubit i conditioned on qubit j and that for qubit i obtained from parallel measurement. Values between 0.01 and 0.1 are highlighted in blue and those above 0.1 are highlighted in green. The right panels are schematic influence diagrams where each arrow points from the qubit traced out to the qubit of interest and the thickness is proportional to the distance. This comparison is shown for IBM Q 5 Tenerife (upper panels) and IBM Q 5 Yorktown (bottom panels), respectively.

mentioned before, due to statistical fluctuations the distance cannot be resolved below the order $O(10^{-3})$. Therefore, it is sensible to compare the entries in the tables to this order of magnitude. For all qubits in both Tenerife and Yorktown, the distances obtained are mostly one order of magnitude larger [of order $O(10^{-2})$], with the largest of order $O(10^{-1})$ for qubit 3 of Tenerife. This suggests that pairwise influence and crosstalk do exist. This is important to take into account when we analyze results of measurement, and this suggests that by adopting the two-qubit detector model the measurement result may be further improved than using just the single-qubit detector model. In the upper panels of Fig. 2 we can see that the result for qubit 3 measured individually differs significantly from that conditioned on qubit 2. This suggests possible influence on qubit 3 by qubit 2. Further discussion on the implications from the other comparison schemes is given in Appendix B.

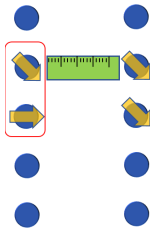
We comment that we also performed three-qubit detector tomography on certain subsets of qubits. These could potentially be used in measurement error mitigation, e.g., in the circuit of Greenberger-Horne-Zeilinger production, similar to the two-qubit detector tomography for the Bell-state circuits below. But we do not present those results here.

C. Error analysis

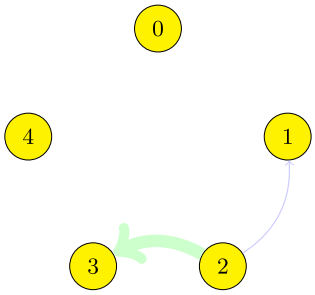
1. Nonparametric bootstrap approach

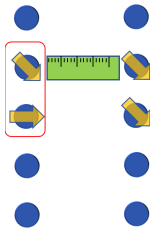
Following the nonparametric bootstrap error analysis [23], we evaluate the uncertainty in the parameters by first obtaining different estimates from resampled data sets of experimental data and calculating statistics of these estimates. In our analysis each experiment was repeated 100 times, each time with 8192 shots on the IBM Q devices. This gives us effectively 819 200 shots, from which the result is calculated using MLE. To evaluate an uncertainty in this result, we resample the set of 100 runs with replacement to obtain new sets of experimental data. These sets are of the same size (100) as our original data set and from each we can calculate a new estimate of the result. The assumption supporting the method of nonparametric bootstrap is that the resampled data will approximate the true probability distribution and their distribution allows one to study the confidence region for the estimated parameters. Since our experiments involve large numbers of repetitions, each of which is assumed independent of the others, by the central limit theorem the distribution of the observed frequency of a given outcome can be approximated by a Gaussian distribution. The fitted detector parameters are linear in the observed frequencies and consequently their distributions will be similar to Gaussian. This enables one to associate the confidence interval of one parameter with the standard error in its estimated value. For a single parameter following Gaussian distribution, the interval within 1.96 standard errors of the estimate contains the true value with probability 95%. The standard deviation across the bootstrap resampled data sets is an estimate of the standard error in the fitted parameter from the original sampled data.

Therefore, the standard deviation of the estimates from resampled data sets gives a scale for the uncertainty in our result. We generated 100 resampled data sets for each experiment. For some selected cases we tried more (up to 1000) resampled data sets, which gave similar standard deviation to that from 100 resampled data sets. Therefore, we believe the



qubit of interest \ qubit traced out	0	1	2	3	4
0	-	0.0006	0.0006	0.0004	0.0003
1	0.0004	-	0.0154	0.0005	0.0001
2	0.0008	0.0016	-	0.0009	0.0023
3	0.0002	0.0006	0.1642	-	0.0028
4	0.0010	0.0007	0.0022	0.0023	-





qubit of interest \ qubit traced out	0	1	2	3	4
0	-	0.0013	0.0036	0.0007	0.0010
1	0.0011	-	0.0116	0.0009	0.0133
2	0.0033	0.0055	-	0.0090	0.0010
3	0.0003	0.0006	0.0044	-	0.0191
4	0.0007	0.0088	0.0020	0.0130	-

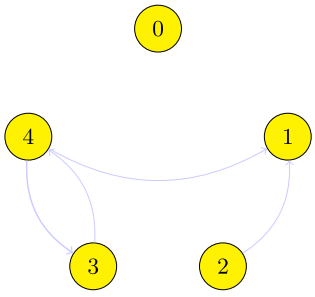


FIG. 4. The left panels explain the comparison scheme: the single-qubit detector on the left is obtained by tracing out another qubit in a two-qubit detector while that on the right is from pairwise parallel measurement. The tables contain the distances between these two: the entry in the ith row and jth column is the distance between the single-qubit detector of qubit i conditioned on qubit j and that for qubit i obtained from pairwise parallel measurement. Values between 0.01 and 0.1 are highlighted in blue and those above 0.1 are highlighted in green. The right panels are schematic influence diagrams where each arrow points from the qubit traced out to the qubit of interest and the thickness is proportional to the distance. This comparison is shown for IBM Q 5 Tenerife (upper panels) and IBM Q 5 Yorktown (bottom panels), respectively.

standard deviation from 100 resampled data sets can represent the fluctuations well. The estimated statistical fluctuations in the parameters of the obtained detector models are typically of order $O(10^{-4})$, with some exceptions of order $O(10^{-3})$ (see Tables V and VI). This would cause fluctuations of order up to $O(10^{-3})$ in the distance between two $\vec{a}^{(0)}$ vectors.

2. *Hessian approach*

In addition to bootstrap, the authors of [19] used likelihood ratio confidence regions [24] to define an error bar in the following way:

$$\Delta f = \sqrt{c(\nabla f)^{\dagger} \cdot H^{-1} \cdot (\nabla f)}, \tag{10}$$

where f is a function of the fitted parameters, H is the Hessian matrix without any redundancy in the parameter space, and the constant c is the argument for the cumulative density function of the normal distribution when its value reaches the desired confidence (e.g., 95%). This definition relies on the Gaussian-like likelihood function and the fact that in GST there are no constraints on the fitted parameters. As discussed previously it is reasonable to assume Gaussian distribution for the detector parameters, but the latter statement is not true in QDT. Specifically the operators in a POVM must be positive semidefinite. When the constraint $|a_0^{(0)}|^2 \geqslant \sum_{i=1}^{3} |a_i^{(0)}|^2$ is close to being saturated, the distribution will be non-Gaussian. Note that normalization is not a constraint if we remove the redundancy in the parameters. For example, in the single-qubit detector case one only considers the parameter space formed by $\vec{a}^{(0)}$ since $\vec{a}^{(1)} = (1, 0, 0, 0) - \vec{a}^{(0)}$. Under the positivity constraint the detector parameters are not completely independent and the distribution may be distorted from Gaussian. However, we argue that if the distribution of each parameter has a narrow peak we may overlook the constraint and still use the definition to quantify the uncertainty in our estimates. We calculate this quantity for the detector parameters and find that most of the time the result agrees in order of magnitude with the fluctuation estimated using the nonparametric bootstrap. In the rarer cases where the nonparametric bootstrap produces fluctuations of order $O(10^{-3})$ instead of $O(10^{-4})$, estimates from this definition are still of order $O(10^{-4})$. This discrepancy is likely to arise from ignoring the positivity constraint.

IV. APPLICATION OF CHARACTERIZED DETECTORS: INFERRING IDEAL DETECTION

Given the characterized detectors, one should be able to infer from the existing measurement data the "correct" joint distribution $P_{(n_0,n_1,\ldots,n_{N-1})}$ of obtaining N-qubit outcomes $(n_0, n_1, \ldots, n_{N-1})$ in the ideal computational basis to some extent. Assuming there is no detector crosstalk,

$$\begin{aligned}\tilde{P}_{(n_0,n_1,\ldots,n_{N-1})} &= \mathrm{Tr}\left(\rho\, \tilde{\Pi}_{[0]}^{n_0} \otimes \tilde{\Pi}_{[1]}^{n_1} \otimes \ldots \otimes \tilde{\Pi}_{[N-1]}^{n_{N-1}}\right) \\ &= \mathrm{Tr}\left[\rho \prod_{j=0}^{N-1} \left(\sum_{q=0}^{3} a_{q,[j]}^{(n_j)} \sigma_{q,[j]}\right)\right],\end{aligned} \tag{11}$$

where $[j]$ denotes the jth physical qubit in the device; we use $\tilde{P}_{(n_0,n_1,\ldots,n_{N-1})}$ to denote the experimental distribution and $P_{(n_0,n_1,\ldots,n_{N-1})}$ to denote the ideal distribution. And when $|a_1^{(n)}|, |a_2^{(n)}| \ll |a_3^{(n)}|$, which is the case in the IBM quantum

computers, Eq. (11) becomes approximately

$$\begin{aligned}\tilde{P}_{(n_0,n_1,\ldots,n_{N-1})} &\approx \mathrm{Tr}\left[\rho \prod_{j=0}^{N-1}\left(a_{0,[j]}+a_{3,[j]}^{(n_j)}\sigma_{3,[j]}\right)\right]\\ &= \sum_{(m_0,m_1,\ldots,m_{N-1})} P_{(m_0,m_1,\ldots,m_{N-1})}\\ &\quad\times \prod_{j=0}^{N-1}\left(a_{0,[j]}^{(n_j)}+(-1)^{m_j}a_{3,[j]}^{(n_j)}\right)\\ &\equiv \sum_{\vec{m}} M_{\vec{n};\vec{m}}P_{\vec{m}}, \end{aligned} \tag{12}$$

where summation over $\vec{m}$ runs through all possible outcomes. Moreover, M is a left-stochastic matrix and its matrix elements are given by

$$M_{\vec{n};\vec{m}} \equiv \prod_{j=0}^{N-1}\left(a_{0,[j]}^{(n_j)}+(-1)^{m_j}a_{3,[j]}^{(n_j)}\right). \tag{13}$$

This can be used to invert the relation to obtain $P_{\vec{m}}$. We remark that the matrix M is equivalent to the transition matrix in [25], where it is estimated by preparing and measuring the classical states instead of using QDT data. A similar idea was proposed by the IBM group [26]. We remark that obtaining this matrix from detector tomography data becomes inefficient for a large number of qubits because the number of experiments needed grows exponentially with the number of qubits.

However, a problem is that the resultant $P_{\vec{m}}$ by direct inversion may have negative components. Similar issues were also addressed in [18], and dealt with by setting a cutoff. In the near-term devices (e.g., IBM Q 5 Yorktown in Sec. IV), this problem is very likely to occur due to statistical fluctuations in measured frequencies. Another way to obtain $P_{\vec{m}}$ circumventing the negativity problem is to minimize the distance squared, $|MP-\tilde{P}|^2$, subject to the constraints of positivity and normalization. This is a quadratic programming problem, the objective function of which is convex (and the solution of which can be found in polynomial time using the ellipsoid method). In Sec. IV we demonstrate this procedure using a built-in function in the PYTHON package SCIPY [27]. We stress that this correction procedure only serves as an easy first-step mitigation, which does not have the full power of QST using characterized detectors. The advantage is that one only needs the measured frequencies for all outcomes in the computational basis and no further experiments are needed.

We remark that this conclusion is based on the assumption that $\tilde{\Pi}_{[j]}^{(n_j)} \approx a_{0,[j]}^{(n_j)}\mathbb{1}+a_{3,[j]}^{(n_j)}\sigma_{3,[j]}$. The situation will be complicated when there are non-negligible components a_1 and a_2, in which case a trick can be used if we can run additional circuits, which are the same as before except with additional Pauli Z gates at the end. This is similar to the idea behind the error mitigation scheme in [28]. The gates added to the end are of the form

$$Z(\vec{K}) \equiv \prod_{i=0}^{N-1}\sigma_{3,[i]}^{K_i}, \tag{14}$$

where $\vec{K}$ is a binary string of length N that denotes whether there is a Pauli Z gate on each qubit in the device. Given a particular $\vec{K}$, the probability is given by

$$\begin{aligned}\tilde{P}_{\vec{n}}(\vec{K}) &= \mathrm{Tr}\left[Z(\vec{K})\rho Z(\vec{K})\prod_{j=0}^{N-1}\left(\sum_{q=0}^{3}a_{q,[j]}^{(n_j)}\sigma_{q,[j]}\right)\right]\\ &= \mathrm{Tr}\left\{\rho\prod_{j=0}^{N-1}\left[\left(a_{0,[j]}^{(n_j)}+a_{3,[j]}^{(n_j)}\sigma_{3,[j]}\right)\right.\right.\\ &\quad\left.\left.+(-1)^{K_j}\left(a_{1,[j]}^{(n_j)}\sigma_{1,[j]}+a_{2,[j]}^{(n_j)}\sigma_{2,[j]}\right)\right]\right\}. \end{aligned} \tag{15}$$

There are 2^N different $\vec{K}$'s, including the original circuit with the probability given by Eq. (11). Adding up $\tilde{P}_{\vec{n}}(\vec{K})$'s cancels the terms involving a_1 and a_2, and their average gives the probability in Eq. (12).

Crosstalk between qubits can further complicate the situation. Let us again make the assumption that in Eq. (2) only coefficients involving Pauli indices $i=0$ and 3 dominate, i.e., $c_{\vec{i}}^{(\vec{n})} \approx 0$ for all the $\vec{i}$'s with any entry equal to 1 or 2. Now the probability is

$$\begin{aligned}\tilde{P}_{\vec{n}} &= \sum_{i_0=0,3}\cdots\sum_{i_{N-1}=0,3} c_{\vec{i}}^{(\vec{n})}\mathrm{Tr}\left(\rho\sigma_{i_0}\otimes\ldots\otimes\sigma_{i_{N-1}}\right)\\ &= \sum_{i_0=0,3}\cdots\sum_{i_{N-1}=0,3} c_{\vec{i}}^{(\vec{n})}\sum_{\vec{m}}(-1)^{\vec{m}\cdot\vec{i}/3}P_{\vec{m}}\\ &= \sum_{\vec{m}}\hat{M}_{\vec{n};\vec{m}}P_{\vec{m}}, \end{aligned} \tag{16}$$

where $\hat{M}$ is given by

$$\hat{M}_{\vec{n};\vec{m}} = \sum_{\vec{I}} c_{\vec{I}}^{(\vec{n})}(-1)^{\vec{m}\cdot\vec{I}/3}. \tag{17}$$

Although $\vec{I}=(i_0,\ldots,i_{N-1})$ has the same expression as $\vec{i}$, we distinguish between them because each component of $\vec{I}$ is equal to 0 or 3. Note the summation only runs through these $\vec{I}$'s. We can use the same procedure to extract the ideal distribution $P_{\vec{m}}$. Moreover, if the assumption that $c_{\vec{i}}^{(\vec{n})} \approx 0$ for all the $\vec{i}$'s with any entry equal to 1 or 2 does not hold, we can still use the average procedure by running additional circuits with gates (14) appended to the original circuits.

Using characterized detectors

We demonstrate how to apply the characterized detectors in a simple real-life experiment for a first-step correction, without carrying out QST, as described in Sec. II. First we applied a Hadamard gate on qubit 3, and then a controlled-NOT (CNOT) gate on qubits 3 (condition) and 4 (target) in IBM Q 5 Yorktown, followed by measurement of all qubits. The resultant state, in the perfect scenario, should be a Bell state between qubits 4 and 3, $(|00\rangle_{43}+|11\rangle_{43})/\sqrt{2}$. The ideal probability distribution will be $P_{00000}=P_{11000}=0.5$ with all other components of $P_{\vec{n}}$ equal to zero. The circuit was repeated for 50 runs, each run with 8192 shots. The largest two components are $\tilde{P}_{00000}=0.466$ and $\tilde{P}_{11000}=0.422$, with the others

TABLE II. Measured frequencies in comparison with those after inversion with cutoff at zero and with those after minimizing $|MP - \tilde{P}|^2$ for (a) a two-qubit Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ on IBM Q 5 Yorktown. The experiment was repeated for 50 runs, each run with 8192 shots. The leftmost column is the qubits operated on and measured.

Qubits of interest	Control-target	Data type	P_{00}	P_{01}	P_{10}	P_{11}	$\|MP - \tilde{P}\|$
0, 1	0-1	Experiment	0.470	0.040	0.054	0.436	
		Inversion	0.489	0.003	0.0	0.509	0.0055
		Optimization	0.490	0.001	0.0	0.509	0.0053
		Uncertainty	0.0018	0.0014	0.0014	0.0016	
3, 4	3-4	Experiment	0.481	0.031	0.041	0.448	
		Inversion	0.480	0.024	0.0	0.497	0.0158
		Optimization	0.483	0.019	0.0	0.498	0.0150
		Uncertainty	0.0013	0.003	0.003	0.0012	

ranging from zero to the order 0.01 ($\tilde{P}_{00001} = 0.013$, $\tilde{P}_{01000} = 0.042$, $\tilde{P}_{10000} = 0.032$, $\tilde{P}_{11001} = 0.011$). Direct inversion gives some negative entries in $P_{\vec{n}}$. An immediate technique is setting any negative entry to zero, and then renormalizing $P_{\vec{n}}$. This results in the two largest components being $\tilde{P}_{00000} = 0.479$ and $\tilde{P}_{11000} = 0.498$, with the biggest among the others of the order 0.01. We then turn to maximizing $|MP - \tilde{P}|^2$ subject to constraints of positivity and normalization. We argue that this method is more desirable because it avoids setting some arbitrary small value as the cutoff. This was done using the optimization function "optimize.minimize" in the PYTHON package SCIPY, with the sequential least-squares programming (SLSQP) method. The tolerance parameter "ftol" was set to 10^{-20} and optimization was typically done after between 300 and 400 iterations. First we apply Eq. (12) with detector parameters obtained from individual measurement. The "corrected" $P_{\vec{n}}$ has two dominant components, $P_{00000} = 0.493$ and $P_{11000} = 0.507$, with all the other components of the order $O(10^{-17})$. We evaluate the uncertainty by combining the uncertainties in the entries of M [as in Eq. (12)] and the statistical fluctuations in the measured probabilities $\tilde{P}_{\vec{n}}$. The estimated uncertainties in the entries of $P_{\vec{n}}$ vary in magnitude, with the largest of the order $O(10^{-3})$ (in particular, the uncertainties in P_{00000} and P_{11000} are 0.002 and 0.0018, respectively). We repeat this analysis using detector parameters obtained from parallel measurement, which gives $P_{00000} = 0.495$ and $P_{11000} = 0.505$ and the other components of the order $O(10^{-17})$. The estimated uncertainties in the entries of $P_{\vec{n}}$ are of the order $O(10^{-3})$ or less (in particular, the uncertainties in P_{00000} and P_{11000} are 0.0012 and 0.0010, respectively). We would like to stress that our experiment involves applying CNOT on a particular state (i.e., $|0+\rangle_{43}$), which does not fully capture the errors that the CNOT gate can cause and therefore does not reflect the overall fidelity of CNOT gate. It would require full QPT to characterize the CNOT gate.

When a number n_{exp} of qubits are operated on in an experiment, the matrix $\hat{M}$ in Eq. (16) used for error mitigation needs to characterize the same qubits involved in the experiment. To obtain this information, without making the assumption of the qubits being independent of each other, one should use the n_{exp}-qubit detector model to carry out the detector tomography. The reason is that in such a tomography procedure exactly the same qubits as those in the experiment are turned on, so that the crosstalk between detectors will be captured. To demonstrate the use of the double-qubit detector model, we prepare a Bell state on two qubits using the Hadamard gate and CNOT gate, followed by measuring only the two qubits involved. It is worth noting that the CNOT gates used respect the connectivity of the qubits in the real machine, so that we know which qubits are actually operated on. We list the measured frequencies in comparison with their corrected versions in Table II. It is clear that with the first-step correction P_{00} and P_{11} are brought closer to the ideal value 0.5, and their difference is reduced. We also note that the direct inversion with cutoff at zero and the optimization using SLSQP give similar results.

V. GATE SET TOMOGRAPHY

Here we provide results of the GST analysis we carry out on the two IBM machines. We begin with a brief review of the GST scheme developed in [18,19]. A gate set in GST is defined as the collection of an unknown initial state ρ, a set of unknown completely positive trace preserving (CPTP) gates $\{G_k\}$, and a two-outcome unknown POVM $\{E, \mathbb{1} - E\}$. The first step is called "linear GST," which broadly speaking is to express the gate set in some arbitrary basis. Taking the Hilbert-Schmidt space of matrices on the original Hilbert space as the new vector space, the expressions are the

TABLE III. Single-qubit detector results by using GST on IBM Q 5 Tenerife, measured for all five qubits in parallel. However, if we perform the same GST only on qubit 3, then we obtain $\vec{a}^{(n_3=0)} = (0.5182, 0.0036, 0.0022, 0.4449)$ and $\vec{a}^{(n_3=1)} = (0.4818, -0.0036, -0.0022, -0.4449)$.

Qubit	Operator	a_0	a_1	a_2	a_3
0	$\Pi^{(0)}$	0.5292	−0.0116	0.0021	0.4707
	$\Pi^{(1)}$	0.4708	0.0116	−0.0021	−0.4707
1	$\Pi^{(0)}$	0.5491	0.0046	0.0058	0.4594
	$\Pi^{(1)}$	0.4509	−0.0046	−0.0058	−0.4594
2	$\Pi^{(0)}$	0.5183	0.0013	−0.0063	0.4816
	$\Pi^{(1)}$	0.4817	−0.0013	0.0063	−0.4816
3	$\Pi^{(0)}$	0.4521	0.0064	0.0061	0.4520
	$\Pi^{(1)}$	0.5478	−0.0064	−0.0061	−0.4520
4	$\Pi^{(0)}$	0.5006	0.0082	0.0079	0.4370
	$\Pi^{(1)}$	0.4994	−0.0082	−0.0079	−0.4370

TABLE IV. Single-qubit detector results by using GST on IBM Q 5 Yorktown, measured for all five qubits in parallel.

Qubit	Operator	a_0	a_1	a_2	a_3
0	$\Pi^{(0)}$	0.5074	−0.0307	−0.0298	0.4847
	$\Pi^{(1)}$	0.4926	0.0307	0.0298	−0.4847
1	$\Pi^{(0)}$	0.5107	0.0078	0.0061	0.4892
	$\Pi^{(1)}$	0.4893	−0.0078	−0.0061	−0.4892
2	$\Pi^{(0)}$	0.5131	0.0230	0.0229	0.4858
	$\Pi^{(1)}$	0.4869	−0.0230	−0.0229	−0.4858
3	$\Pi^{(0)}$	0.5137	0.0170	0.0136	0.4858
	$\Pi^{(1)}$	0.4863	−0.0170	−0.0136	−0.4858
4	$\Pi^{(0)}$	0.5206	0.0171	0.0139	0.4789
	$\Pi^{(1)}$	0.4794	−0.0171	−0.0139	−0.4789

following:

$$\begin{aligned}
\tilde{\mathbb{1}} &= \sum_{j,k} \langle\langle E|F_j F_k|\rho\rangle\rangle, \\
|\hat{\rho}\rangle\rangle &= \tilde{\mathbb{1}}^{-1} \sum_{j} \langle\langle E|F_j|\rho\rangle\rangle, \\
\langle\langle \hat{E}| &= \sum_{k} \langle\langle E|F_k|\rho\rangle\rangle, \\
\hat{G}_i &= \tilde{\mathbb{1}}^{-1} \sum_{j,k} \langle\langle E|F_j G_i F_k|\rho\rangle\rangle,
\end{aligned} \tag{18}$$

$|\rho\rangle\rangle$ and $\langle\langle E|$ are vectorized versions of ρ and E, respectively, and $\{F_j\}$ is a set of gates (acting on vectorized versions of density matrices and POVM) with which the initial state and the measurement will be informationally complete, respectively. The inner products in these equations are obtained from experiments. This can only determine the gate set up to a transformation

$$\begin{aligned}
\rho &= M\hat{\rho}, \\
E &= \hat{E}M^{-1}, \\
G_i &= M\hat{G}_i M^{-1},
\end{aligned} \tag{19}$$

where M is some invertible matrix. This freedom was termed the "gauge" by the authors of [18], and can be removed by trying to match the gate set towards some target. Long sequences of gates are used in experiments so as to capture the amplified gate parameter errors and thus achieve better characterization of a certain subset of gates $\{G_k\}$, called "germs." However, state preparation and measurement (SPAM) parameter errors cannot be amplified. The full GST analysis in [19] can be summarized as follows: (1) linear GST for the first estimate, (2) gauge optimization to match the target gate set, (3) iteratively adding data for χ^2 minimization to avoid local minima, and (4) final MLE analysis and gauge optimization.

Results of single-qubit detector characterization from GST

We also used GST and ran corresponding circuits on IBMQx4, with the standard gate set $\mathbf{G} = \{G_x = R_x(\pi/2), G_y =$

TABLE V. Single-qubit detector results for IBM Q 5 Tenerife (a) measured individually for each physical qubit leaving the other qubits idle and (b) measured for all five qubits in parallel.

(a) Individual measurement.

Qubit	Operator	a_0	a_1	a_2	a_3
0	$\Pi^{(0)}$	0.590(2)	−0.006(3)	−0.0063(4)	0.3562(5)
	$\Pi^{(1)}$	0.410(2)	0.006(3)	0.0063(4)	−0.3562(5)
1	$\Pi^{(0)}$	0.544(1)	0.001(3)	0.0008(3)	0.4059(5)
	$\Pi^{(1)}$	0.456(1)	−0.001(3)	−0.0008(3)	−0.4059(5)
2	$\Pi^{(0)}$	0.5427(5)	−0.0179(9)	−0.0173(5)	0.4294(4)
	$\Pi^{(1)}$	0.4573(5)	0.0179(9)	0.0173(5)	−0.4294(4)
3	$\Pi^{(0)}$	0.5381(5)	−0.003(1)	−0.0030(4)	0.4054(4)
	$\Pi^{(1)}$	0.4619(5)	0.003(1)	0.0030(4)	−0.4054(4)
4	$\Pi^{(0)}$	0.521(1)	−0.012(2)	−0.0122(4)	0.3798(4)
	$\Pi^{(1)}$	0.479(1)	0.012(2)	0.0122(4)	−0.3798(4)

(b) Parallel measurement.

Qubit	Operator	a_0	a_1	a_2	a_3
0	$\Pi^{(0)}$	0.587(2)	−0.000(3)	−0.0001(4)	0.3618(5)
	$\Pi^{(1)}$	0.413(2)	0.000(3)	0.0001(4)	−0.3618(5)
1	$\Pi^{(0)}$	0.5483(8)	0.006(2)	0.0053(4)	0.4116(4)
	$\Pi^{(1)}$	0.4517(8)	−0.006(2)	−0.0053(4)	−0.4116(4)
2	$\Pi^{(0)}$	0.5329(5)	−0.0065(7)	−0.0064(5)	0.4430(5)
	$\Pi^{(1)}$	0.4671(5)	0.0065(7)	0.0064(5)	−0.4430(5)
3	$\Pi^{(0)}$	0.4535(8)	0.002(1)	0.0023(4)	0.4229(5)
	$\Pi^{(1)}$	0.5465(8)	−0.002(1)	−0.0023(4)	−0.4229(5)
4	$\Pi^{(0)}$	0.522(1)	0.000(2)	−0.0002(4)	0.3975(4)
	$\Pi^{(1)}$	0.478(1)	−0.000(2)	0.0002(4)	−0.3975(4)

TABLE VI. Single-qubit detector results for IBM Q 5 Yorktown (a) measured individually for each physical qubit leaving the other qubits idle and (b) measured for all five qubits in parallel.

(a) Individual measurement.

Qubit	Operator	a_0	a_1	a_2	a_3
0	$\Pi^{(0)}$	0.545(2)	−0.013(2)	−0.012(3)	0.424(3)
	$\Pi^{(1)}$	0.455(2)	0.013(2)	0.012(3)	−0.424(3)
1	$\Pi^{(0)}$	0.530(2)	0.003(1)	0.0028(5)	0.4625(5)
	$\Pi^{(1)}$	0.470(2)	−0.003(1)	−0.0028(5)	−0.4625(5)
2	$\Pi^{(0)}$	0.5159(2)	0.0007(3)	0.0005(4)	0.4788(4)
	$\Pi^{(1)}$	0.4841(2)	−0.0007(3)	−0.0005(4)	−0.4788(4)
3	$\Pi^{(0)}$	0.534(1)	0.003(1)	0.0029(4)	0.4600(5)
	$\Pi^{(1)}$	0.466(1)	−0.003(1)	−0.0029(4)	−0.4600(5)
4	$\Pi^{(0)}$	0.5181(6)	0.001(4)	0.0004(4)	0.4417(4)
	$\Pi^{(1)}$	0.4819(6)	−0.001(4)	−0.0004(4)	−0.4417(4)

(b) Parallel measurement.

Qubit	Operator	a_0	a_1	a_2	a_3
0	$\Pi^{(0)}$	0.544(1)	0.016(1)	0.0163(4)	0.4130(4)
	$\Pi^{(1)}$	0.456(1)	−0.016(1)	−0.0163(4)	−0.4130(4)
1	$\Pi^{(0)}$	0.5199(3)	0.0115(3)	0.0109(4)	0.4703(4)
	$\Pi^{(1)}$	0.4801(3)	−0.0115(3)	−0.0109(4)	−0.4703(4)
2	$\Pi^{(0)}$	0.5181(2)	0.0320(3)	0.0318(3)	0.4749(4)
	$\Pi^{(1)}$	0.4819(2)	−0.0320(3)	−0.0318(3)	−0.4749(4)
3	$\Pi^{(0)}$	0.5304(4)	0.0244(3)	0.0250(3)	0.4634(4)
	$\Pi^{(1)}$	0.4696(4)	−0.0244(3)	−0.0250(3)	−0.4634(4)
4	$\Pi^{(0)}$	0.5149(2)	0.0121(1)	0.0121(4)	0.4594(4)
	$\Pi^{(1)}$	0.4851(2)	−0.0121(1)	−0.0121(4)	−0.4594(4)

$R_y(\pi/2), G_I = \mathbb{1}\}$ for the state preparation and measurement fiducials. GST is particularly well suited for characterizing gates but less so for state preparation and measurement. However, it still provides a good comparison for detector tomography without the bias of assuming high-fidelity gate and state preparation. To run the GST with a long sequence requires a large number of different circuits to run. Since we are only interested in the detectors, we only choose $\{G_I\}$ to be the germs and only of six different lengths, i.e., [1,121,241,361,481,601]. (Our other motivation was to capture some simple relaxation or decoherence from the decohered identity operation; see also Appendix E.) Even for such a simple setup, there are 272 different circuits to run (compared to just six for QDT). For each we take 8192 shots to obtain statistics. Ideally, we could have included G_x and G_y in the set of germs, but it will require many more circuits to run.

We use the PYGSTI package (version 0.9.6) of PYTHON to analyze the data. We first run linear-inversion GST without using the long-sequence circuits to obtain an initial estimate. We perform the gauge optimization by setting gate parameters to be trace preserving (TP). If the obtained POVMs for the detectors are not positive, we repeat the analysis by additionally setting a nonzero SPAM penalty factor (typically from 0.3 to 0.5) so as to obtain positive POVMs. We then project the gates (this does not modify characterization of the initial state or measurement POVMs) to be CPTP. Using such an initial estimate, we then run the long-sequence GST (LSGST), which takes into account long-sequence circuits by performing iterative maximum likelihood by including longer sequences successively. Since we are concerned mostly with the detectors' POVMs, if the obtained POVMs after LSGST are not positive, we will repeat gauge optimization by setting the SPAM penalty factor.

The circuits for GST were done in parallel for all five qubits on both IBMQx2 and IBMQx4. The obtained detectors' POVMs characterized by the above GST procedure are shown in Tables III and IV. They agree within a few percent with those using simple detector tomography earlier. The particular qubit 3 of IBMQx4 also displays the unusual behavior that its detector $a_0^{(n=0)} = 0.4521$ is smaller than $a_0^{(n=1)} = 0.5479$ in parallel measurement together with other qubits.

We then performed individual GST procedure only for the qubit 3 (leaving all other qubits idle), and obtain the detectors' characterization $\vec{a}^{(n_3=0)} = (0.5182, 0.0036, 0.0022, 0.4449)$ and $\vec{a}^{(n_3=1)} = (0.4818, -0.0036, -0.0022, -0.4449)$, which is closer to what was obtained by simple detector tomography on the qubit 3 individually with about 4% difference.

VI. CONCLUSION AND DISCUSSION

In summary we performed the standard quantum detector tomography on the two devices IBM Q 5 Tenerife and IBM Q 5 Yorktown, assuming negligible errors in the ground-state preparation and the single-qubit gates used to prepare the eigenstates of the three Pauli operators. Our resultant POVM shows deviation from the ideal projectors $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ and can be used for a first-order correction in experiments.

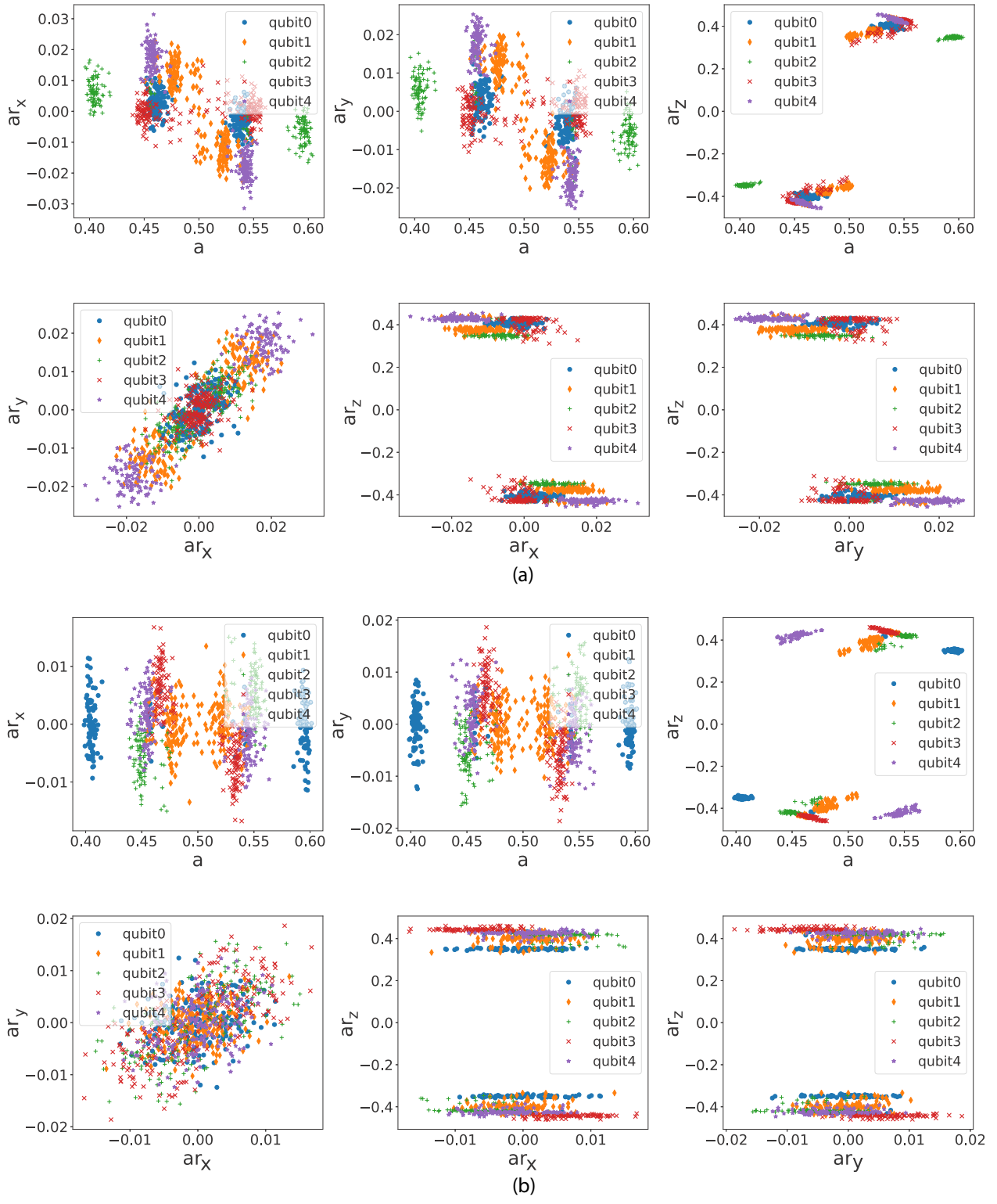


FIG. 5. Visualizing the single-qubit detector parameters for the five qubits in IBM Q 5 Tenerife obtained by (a) individual measurement and (b) parallel measurement. The parameters $ar_{x,y,z} = a_{1,2,3}$. These are displayed for both $\Pi^{(1)}$ and $\Pi^{(0)}$.

We also found evidence of crosstalk between qubits in one device. In particular, discrepancy was seen between individual measurement and parallel measurement. We believe adopting two- or more-qubit detector models can improve the results further compared with assuming an independent single-qubit detector model. To study that, more knowledge about hardware is required. Some peculiar features were observed in the qubit 3 of IBM Q 5 Tenerife that need further investigation.

The peculiar behavior of the qubit 3 from simple QDT agrees with that obtained from using a more sophisticated approach of the GST. This method, in principle, is capable of

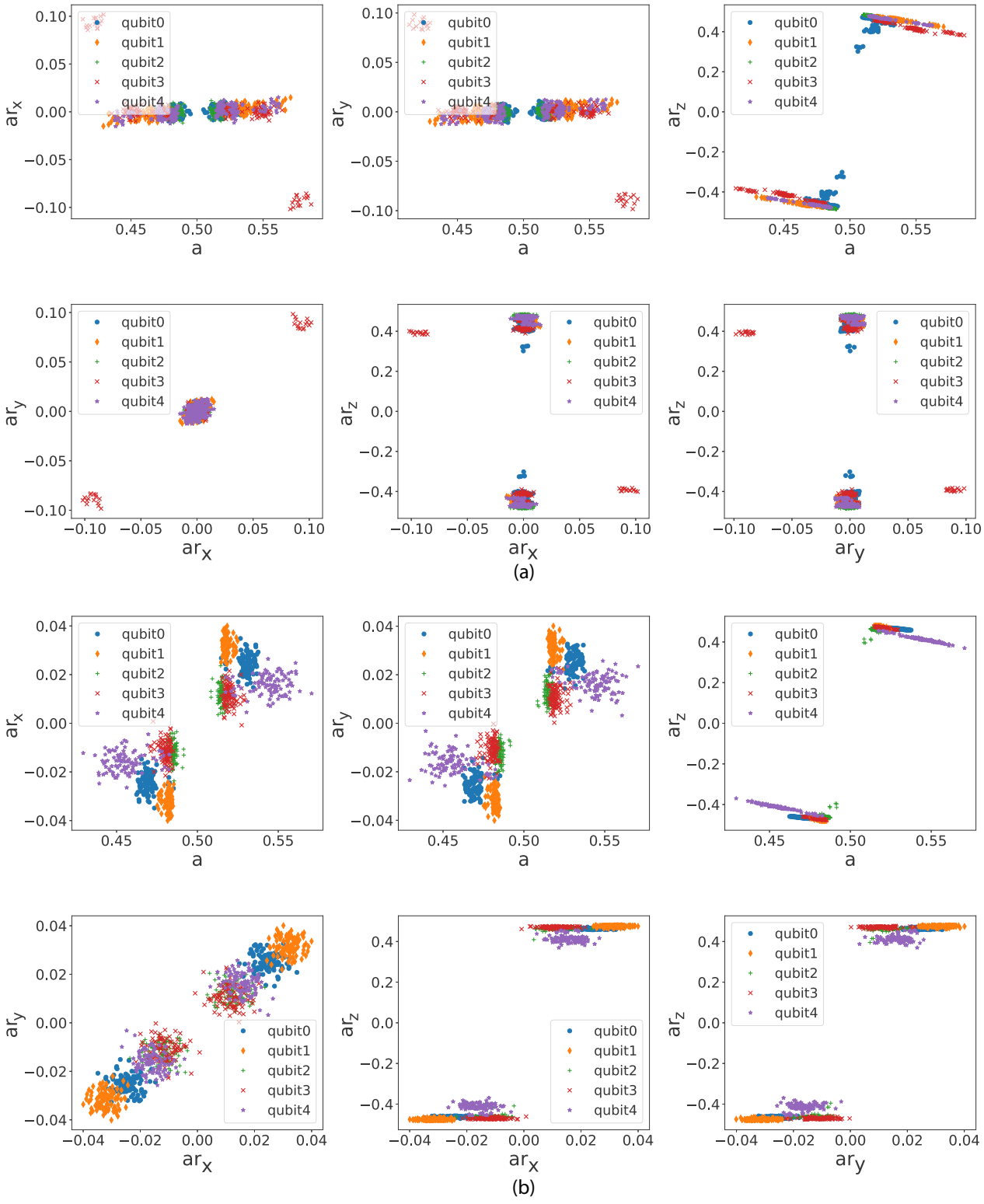


FIG. 6. Visualizing the single-qubit detector parameters for the five qubits in IBM Q 5 Yorktown obtained by (a) individual measurement and (b) parallel measurement. The parameters $ar_{x,y,z} = a_{1,2,3}$. These are displayed for both $\Pi^{(1)}$ and $\Pi^{(0)}$.

deducing the initial state, gate operations, and measurement POVMs in one go, by running various circuit sequences. Our simple QDT relies on the assumption that the detector error rate is higher than that of state preparation and simple single-qubit gates.

We point out some directions for future work. Given that the total number N of qubits can be large, complete detector tomography will not be efficient. One can consider employing compressed sensing, as done in the state tomography [22]. On the other hand, since the

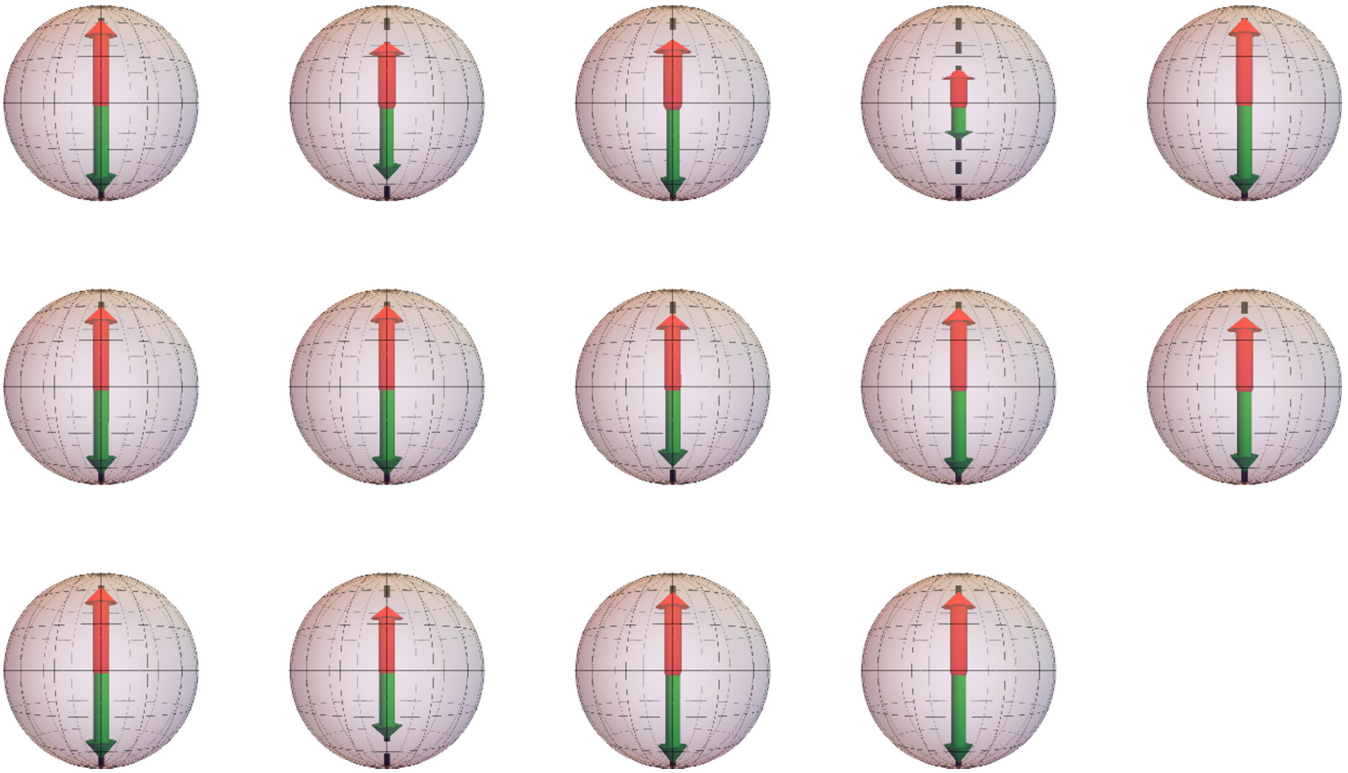


FIG. 7. Detector spheres for qubits 0 to 13 of IBM Q 16 Melbourne (from left to right, the rows corresponding to qubits 0 to 3, 4 to 7, 8 to 11, and 12 and 13, respectively).

characterization of the triad—detector, state, and process—forms a loop, there should be further improvement (to the next order in error) on detector characterization, and then on the state and process tomography, an idea similar to that in [21].

ACKNOWLEDGMENTS

This work was supported by NSF Grants No. PHY 1620252 and No. PHY 1915165, as well as SUNY Center-Scale Proposal Planning and Development Grant Program No. RFP 18-02CSP (Grant No. CSP18035) and Research Subcontract No. 358980 from Bookhaven Science Associates, LLC. We acknowledge the use of IBM quantum computers open to the public. We thank Sergey Bravyi, Anthony DeGennaro, Antonio Mezzacapo, and Ollie Saira for useful discussions. Y.C. would also like to thank Xinzhong Chen for discussion about error analysis.

APPENDIX A: REVIEW OF QST AND QPT

1. Quantum state tomography

The idea of quantum state tomography was proposed [3,29] early in the context of quantum optics using quasiprobability distributions, and it has become a standard procedure in measuring multiple qubits [4,5]. To infer the best estimate of an unknown state ρ, one can choose a (over-)complete set of POVM elements $\{\Pi^{(i)}\}$ and obtain the data $p_{\rho,i} = \mathrm{Tr}(\rho\Pi^{(i)})$ from measurements. The approach was later extended to a "hedged" version [30] and a mean Bayesian version [31] that deal with certain drawbacks of MLE [32,33]. However, QST requires $O(3^n)$ different measurement bases for n qubits, but compressed sensing can be used to ameliorate this [22].

Nevertheless, quantum state tomography remains an indispensable ingredient in characterizing small quantum systems, and even a partial tomography (for some part of a larger system) can also be useful when one is verifying some properties, such as the existence of entanglement, that may not require a complete global wave function. *However, most of the description relies on the assumption that almost perfect projective von Neumann measurements can be performed*. Here, we will consider a more realistic scenario where measurements are not necessarily projective, as in, e.g., IBM quantum computers the measurement errors of which are not negligible, of order 2–5%. (See manufactures' released data for devices' properties, e.g., on IBM Q Experience or Rigetti Computing, but some useful information was listed in the Appendix of Ref. [34].)

2. Quantum process tomography

Related to state tomography is the characterization of a quantum process, which may arise from application of a gate or evolution of a system that possibly couples to its environment. In the latter case, it is commonly considered in the Markovian limit, and one arrives at the so-called master equation for the system state $\rho(t)$ [2]:

$$\frac{d\rho(t)}{dt} = -\frac{i}{\hbar}[H,\rho] + \mathcal{L}(\rho) \tag{A1}$$

$$= -\frac{i}{\hbar}[H,\rho] + \sum\nolimits_j (2L_j\rho L_j^\dagger - \{L_j^\dagger L_j,\rho\}), \tag{A2}$$

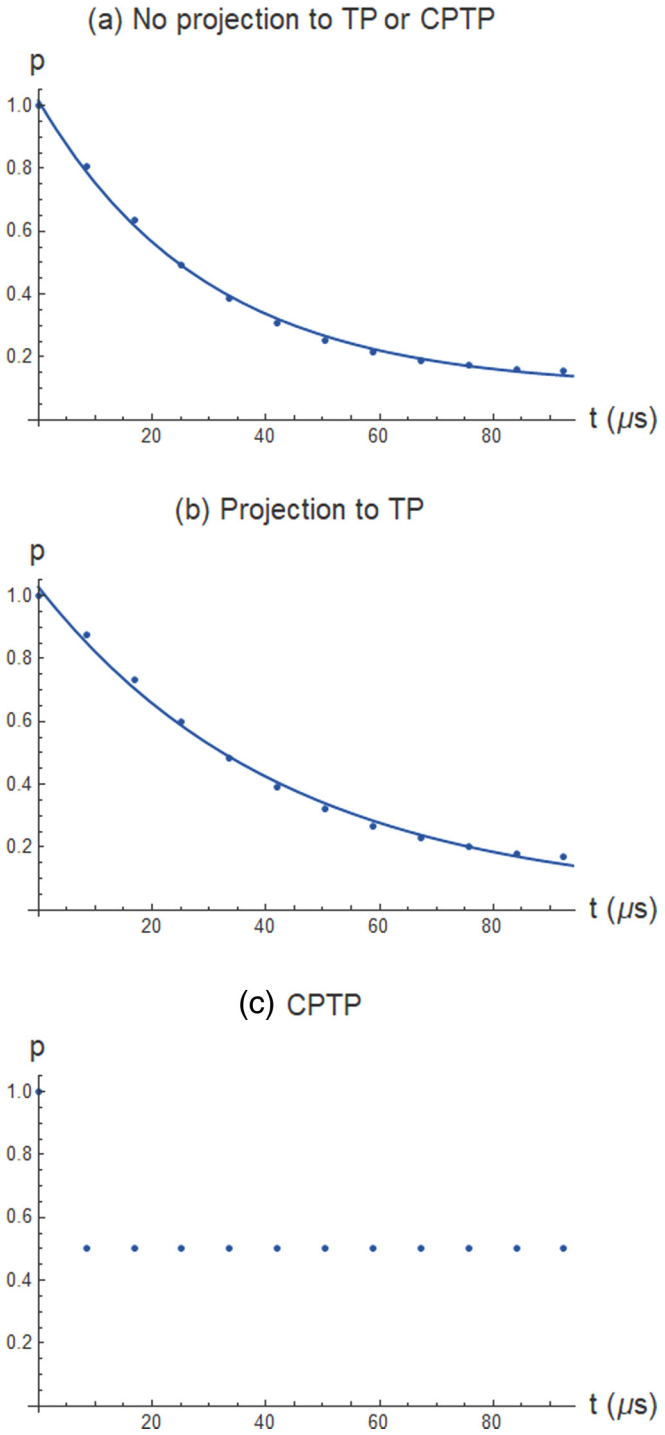


FIG. 8. Qubit relaxation: probability p of measuring $|1\rangle$ vs a time t, calculated using the identity gate from (a) best GST estimate without imposing TP or CPTP condition, (b) projecting the identity gate from (a) to be TP, and (c) projecting the identity gate from (a) to be CPTP.

where H is the system Hamiltonian, and L_j's are the Lindblad operators, representing the effect of coupling to environment. One can describe the change of ρ in a discrete time step Δt as a quantum process, $\rho(t_0) \to \rho(t_0 + \Delta t) = \mathcal{E}(\rho)$. A general quantum process $\mathcal{E}$ can be described by a set of Kraus operators E_j, so that its action on ρ is $\mathcal{E}(\rho) = \sum_j E_j \rho E_j^\dagger$, where without loss of generality we can assume $\mathcal{E}$ to be trace preserving: $\sum_j E_j^\dagger E_j = I$, unless there is some loss or leakage. The procedure to infer $\mathcal{E}$ is called quantum process tomography, which is natural to consider given QST [2,6–8]. It is possible to infer the quantum process because of linearity, and if one applies (unknown) $\mathcal{E}$ to a complete basis of a density matrix, e.g., $|k\rangle\langle l| \to \mathcal{E}(|k\rangle\langle l|)$, then by measuring the output the process can be determined [2]. The matrix element $|k\rangle\langle l|$ can be expressed in terms of a linear combination of different states $|\psi\rangle\psi|$ in, e.g., $\mathcal{A}_1 \equiv \{|0/1\rangle\langle 0/1|, |+/-\rangle\langle +/-|, |\pm i\rangle\langle \pm i|\}$ for one qubit, and thus *quantum process tomography uses quantum state tomography as a subroutine.* Instead of varying the input states over some "complete" (or even overcomplete) set, such as $\mathcal{A}_1$ above, one can also use a bipartite maximally entangled state $|\Psi\rangle_{AB}$, where the party A corresponds to the system that will be acted on by the process $\mathcal{E}_A$, and the party B acts as an ancillary role. Then state tomography on the resulting bipartite system (after A undergoing the process $\mathcal{E}$) gives identical determination of the process $\mathcal{E}$ [35].

However, in currently available small-scale quantum computers, both measurement and state preparation have errors, and if there is some separation of rates in these different types of errors, as in IBM and Rigetti quantum computers, then we can give better individual characterization.

APPENDIX B: TABLES AND FIGURES

We list Tables V and VI containing processed experimental data. The error estimated for each parameter is typically of the order $O(10^{-4})$, with the largest error among them up to 0.003. We adopt the precision according to the estimated errors. To further demonstrate the crosstalk, we present the comparisons between the single-qubit detector obtained by tracing out the other qubit in a double-qubit detector and the single-qubit detector obtained from parallel measurement (Fig. 3) and pairwise parallel measurement (Fig. 4). We can see that by tracing out one qubit in a pair the double-qubit detector result is reduced to the single-qubit result obtained from pairwise parallel measurement within statistical fluctuations for most pairs. In Tenerife, one significant exception is qubit 3's detector under the influence of qubit 2's detector, and in Yorktown qubit 1's detector is influenced by that of qubit 2. This significant discrepancy could be a consequence of basis-dependent influence, since in the two-qubit detector the two qubits are treated independently whereas in pairwise parallel measurement the operations on them are always the same. We have seen in Fig. 2 signs of influence on qubit 3 by qubit 2. If qubit 2 is the single source of crosstalk, one would expect the result for qubit 3 from parallel measurement or pairwise parallel measurement to agree with the result conditioned on qubit 2 (operating only qubits 2 and 3). From the upper panels of Figs. 3 and 4, however, we can see that this is not the case. In fact, the result for qubit 3 measured in parallel with all the other qubits differs from any of the results obtained by tracing out the other qubit from two-qubit detectors. This could imply basis-dependent influence and/or nontrivial correlation when several physical qubits are being operated on.

APPENDIX C: SCATTER PLOTS FOR DETECTOR PARAMETERS

In Figs. 5 and 6, we present scattered plots for detector parameters in IBM Q 5 Tenerife and in IBM Q 5 Yorktown for 100 different runs. These are detector characterization in higher-dimensional parameter space projected onto two-dimensional cross sections. They show signs of separation of the single-qubit detectors for different physical qubits in the parameter space. For some cross sections, data from different detectors seem to cluster to five groups (and their corresponding "inverse"), e.g., panels 1 and 2 in Fig. 5(a) (i.e., those involving parameter a). For others, the data seem to all clump together, e.g., the fourth panel in Figs. 5(a) and 5(b) (i.e., those involving ar_x and ar_y, as these have small values). The parameters of the vectors are the signatures of the detectors. As they are not made identical (e.g., such as atoms), they can differ slightly, despite the experimental effort to make them as much identical to the ideal ones. These may be of interest from the machine learning perspective to learn the detectors, but this is beyond the scope of this paper.

APPENDIX D: QDT FOR IBM Q 16 MELBOURNE

For completeness, we present the detector tomography on all 14 qubits of the IBM Q 16 Melbourne device. The results are presented in the form of Bloch spheres in Fig. 7, done via data taken in parallel for all 14 qubits. As in the other machines, all detectors align pretty much along the vertical z axis. We notice that the POVMs for qubit 3 have arrows that are relatively shorter than all the rest. Its detectors have the largest imperfection. We note that one can repeat individual, pairwise, or triple characterization we discussed in the main text, but we do not present them here.

APPENDIX E: RELAXATION FROM GST

Here we illustrate the relaxation from GST. Since in our GST implementation we have only used the identity gate $G_I = \mathbb{1}$ in the germ set, we can examine whether the identity operation, which is essentially letting the qubit idle, can allow us to extract relaxation of a qubit in the excited state $|1\rangle$. Since the circuits for the GST include a gate sequence such as $(G_I)^m G_x G_x$, we already have the data for relaxation. Note that $(G_x)^2 = i\sigma_x$ flips $|0\rangle$ to the excited state $|1\rangle$ and the m identity gates represent an idling of m units of gate duration, including gate and buffer times. The sequence measures relaxation. Let us use qubit zero of ibmqx4 for illustration. Note that each single-qubit gate duration, including the buffer time, is 70 ns. The identity gate we obtained from GST before imposing the TP or CPTP condition, expressed in the Pauli basis, is

$$G_I = \begin{pmatrix} 0.998699 & 0.001783 & -0.001243 & 0.002354 \\ -0.047276 & 1.044539 & -0.065670 & 0.083989 \\ 0.030418 & -0.068910 & 0.980521 & -0.054349 \\ 0.049455 & -0.093771 & 0.056438 & 0.904232 \end{pmatrix}. \tag{E1}$$

Note that the element $(G_I)_{\alpha,\beta}$ represents the amplitude that σ_β is mapped to σ_α (with $\sigma_0 = \mathbb{1}$), under the idling operation that is supposed to be the identity gate. This allows us to extract a relaxation time $T_1 \approx 29.5\ \mu$s. In Fig. 8(a), we show the probability of obtaining $|1\rangle$ after first applying an ideal σ_x gate to $|0\rangle$ and then applying m (=120, 240, ..., etc.) such discrete identity gates. The curve is an exponential decay fit to the discrete data. Since the obtained identity gate is not positive, using it to simulate other process, such as T_2 decoherence time, we would obtain some probability that is negative, an unphysical result (not shown).

Projecting the gate to be TP, then the identity gate becomes

$$G_I = \begin{pmatrix} 1. & 0. & 0. & 0. \\ -0.047276 & 1.044539 & -0.065670 & 0.083989 \\ 0.030418 & -0.068910 & 0.980521 & -0.054348 \\ 0.049455 & -0.093771 & 0.056438 & 0.904232 \end{pmatrix}. \tag{E2}$$

From simulating the relaxation using this identity gate, we extract a relaxation time $T_1 \approx 43.2\ \mu$s; see Fig. 8(b). This value differs about 50% with one obtained earlier without projecting the identity gate to be TP.

Projecting the gate to be CPTP, then the identity gate becomes

$$G_I = \begin{pmatrix} 1. & 0. & 0. & 0. \\ 0. & 0.9731358 & -0.02961343 & 0.08478270 \\ 0. & -0.03283288 & 0.9433932 & -0.05294062 \\ 0. & -0.08548849 & 0.05312586 & 0.9170383 \end{pmatrix}. \tag{E3}$$

However, from this matrix, we cannot extract a meaningful relaxation time; see Fig. 8(c). Although it is out of the scope of this paper, it will be interesting to see if one can characterize the identity gate (which lets the qubit idle) and obtain its correct CPTP description that can describe relaxation or even dephasing.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University, Cambridge, England, 2000).

[3] U. Leonhardt, Quantum-State Tomography and Discrete Wigner Function, Phys. Rev. Lett. **74**, 4101 (1995).

[4] Z. Hradil, Quantum-state estimation, Phys. Rev. A **55**, R1561 (1997).

[5] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[6] Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, and H. J. Kimble, Measurement of Conditional Phase Shifts for Quantum Logic, Phys. Rev. Lett. **75**, 4710 (1995).

[7] 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).

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

[9] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B **7**, S347 (2005).

[10] 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).

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

[12] J. Fiurášek, Maximum-likelihood estimation of quantum measurement, Phys. Rev. A **64**, 024102 (2001).

[13] J. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. Pregnell, C. Silberhorn, T. Ralph, J. Eisert, M. Plenio, and I. Walmsley, Tomography of quantum detectors, Nat. Phys. **5**, 27 (2009).

[14] A. Feito, J. Lundeen, H. Coldenstrodt-Ronge, J. Eisert, M. Plenio, and I. Walmsley, Measuring measurement: Theory and practice, New J. Phys. **11**, 093038 (2009).

[15] J. J. Renema, G. Frucci, M. J. de Dood, R. Gill, A. Fiore, and M. P. van Exter, Tomography and state reconstruction with superconducting single-photon detectors, Phys. Rev. A **86**, 062113 (2012).

[16] J. Renema, R. Gaudio, Q. Wang, Z. Zhou, A. Gaggero, F. Mattioli, R. Leoni, D. Sahin, M. De Dood, A. Fiore, and M. P. van Exter, Experimental Test of Theories of the Detection Mechanism in A Nanowire Superconducting Single Photon Detector, Phys. Rev. Lett. **112**, 117604 (2014).

[17] H. Abraham, I. Y. Akhalwaya, G. Aleksandrowicz *et al.*, Qiskit: An open-source framework for quantum computing (2019), doi:10.5281/zenodo.2562110.

[18] 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).

[19] 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**, 14485 (2017).

[20] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).

[21] A. C. Keith, C. H. Baldwin, S. Glancy, and E. Knill, Joint quantum-state and measurement tomography with incomplete measurements, Phys. Rev. A **98**, 042318 (2018).

[22] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum State Tomography Via Compressed Sensing, Phys. Rev. Lett. **105**, 150401 (2010).

[23] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, London, 1993).

[24] R. Blume-Kohout, Robust error bars for quantum tomography, arXiv:1202.5270 (2012).

[25] M. Sun and M. R. Geller, Efficient characterization of correlated spam errors, arXiv:1810.10523 (2018).

[26] A. Mezzacapo, 2019 (private communication).

[27] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.* (SciPy 1.0 Contributors), SciPy 1.0-fundamental algorithms for scientific computing in python, arXiv:1907.10121 (2019).

[28] F. A. Calderon-Vargas and J. P. Kestner, Dynamically Correcting a Cnot Gate for Any Systematic Logical Error, Phys. Rev. Lett. **118**, 150502 (2017).

[29] K. Vogel and H. Risken, Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase, Phys. Rev. A **40**, 2847 (1989).

[30] R. Blume-Kohout, Hedged Maximum Likelihood Quantum State Estimation, Phys. Rev. Lett. **105**, 200504 (2010).

[31] R. Blume-Kohout, Optimal, reliable estimation of quantum states, New J. Phys. **12**, 043034 (2010).

[32] C. Ferrie and R. Blume-Kohout, Maximum likelihood quantum state tomography is inadmissible, arXiv:1808.01072 (2018).

[33] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühn, Systematic Errors in Current Quantum State Tomography Tools, Phys. Rev. Lett. **114**, 080403 (2015).

[34] B. Pokharel, N. Anand, B. Fortman, and D. A. Lidar, Demonstration of Fidelity Improvement Using Dynamical Decoupling with Superconducting Qubits, Phys. Rev. Lett. **121**, 220502 (2018).

[35] J. B. Altepeter, D. Branning, E. Jeffrey, T.-C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White, Ancilla-Assisted Quantum Process Tomography, Phys. Rev. Lett. **90**, 193601 (2003).