# Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking

Easwar Magesan,[1,2] Jay M. Gambetta,[3] B. R. Johnson,[4] Colm A. Ryan,[4] Jerry M. Chow,[3] Seth T. Merkel,[3]
Marcus P. da Silva,[4] George A. Keefe,[3] Mary B. Rothwell,[3] Thomas A. Ohki,[4] Mark B. Ketchen,[3] and M. Steffen[3]

[1]*Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[2]*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[3]*IBM T.J. Watson Research Center, Yorktown Heights, New York 10598, USA*
[4]*Raytheon BBN Technologies, Cambridge, Massachusetts 02138, USA*
(Received 1 April 2012; published 24 August 2012)

We describe a scalable experimental protocol for estimating the average error of individual quantum computational gates. This protocol consists of interleaving random Clifford gates between the gate of interest and provides an estimate as well as theoretical bounds for the average error of the gate under test, so long as the average noise variation over all Clifford gates is small. This technique takes into account both state preparation and measurement errors and is scalable in the number of qubits. We apply this protocol to a superconducting qubit system and find a bounded average error of 0.003 [0,0.016] for the single-qubit gates $X_{\pi/2}$ and $Y_{\pi/2}$. These bounded values provide better estimates of the average error than those extracted via quantum process tomography.

Determining how well an operation is implemented on a quantum device is of fundamental importance in quantum information theory. Such a characterization allows a direct comparison between different architectures for computation as well as an understanding of the performance of the building blocks of a quantum computer. The standard method for characterizing a quantum operation is quantum process tomography (QPT) [1,2] which is subject to two significant drawbacks: first, it is not scalable in the number of subsystems (qubits) comprising the system; and second if state-preparation and measurement (SPAM) errors are present, then these errors will contribute to those of the gate being characterized, hence giving an unfaithful estimation of the actual error. In many cases, one does not require the complete knowledge that QPT aims to provide. As a result, various methods for partially characterizing a quantum gate have been developed [3–11]. Ideally such a method should be scalable in the number of qubits, $n$, comprising the system as well as provide a faithful measure of the noise that is independent of SPAM errors.

One particular method for partial noise characterization is "randomized benchmarking" (RB) [4,11,12], with Ref. [11] providing the first scalable RB protocol that satisfies all of the above criteria. The general idea of RB is to implement random sequences of gates that compose to the identity operation, and measure the fidelity of each sequence. Averaging over different realizations results in a fidelity decay versus the sequence length, from which the average error over the full gate set is estimated via fitting the curve to a derived model. The simplicity of this protocol has lead to various experimental implementations of the single-qubit gate protocol presented in Ref. [12], including in atomic ions with different types of traps

[12–14], liquid state nuclear spins [15], superconducting qubits [16–18], and atoms in optical lattices [19].

The multiqubit RB protocol described in Ref. [11] is restricted to benchmark only the full Clifford group on $n$ qubits, Clif$_n$. While this provides a significant step towards scalable benchmarking of a quantum information processor, it is desirable in many cases to benchmark individual gates in Clif$_n$ rather than the entire set. One method for characterizing the fidelity of single Clifford gates has been provided in Ref. [20], proposing an extension of the protocol introduced in Ref. [6]. The main drawback of this method is that it does not account for SPAM errors which can bias estimates of the gate error. Note that benchmarking Clifford gates rather than general elements of the unitary group is not a significant restriction as any unitary gate can be implemented with fault-tolerance using special input states, Clifford elements and computational basis measurements [21]. Additionally, the unitary group can be generated via Clif$_n$ through the addition of a single gate not in the group [22]. Thus, benchmarking Clifford elements provides significant information regarding the reliability of a general quantum gate and is a relevant metric for fault-tolerant thresholds [22–25].

In this Letter, we present a new protocol for benchmarking individual Clifford gates via randomization. Our protocol consists of interleaving random gates between the gate, $\mathcal{C}$, of interest. In the limits of either perfect random gates or that the average error of all gates is depolarizing, our protocol estimates the gate error of $\mathcal{C}$ perfectly. In the completely general case where the random gates have arbitrary errors with small average variation, we provide explicit bounds for the error of $\mathcal{C}$. These bounds give direct information regarding the quality of computational gates and thus useful information about reaching thresholds for

fault-tolerant quantum computation. The method utilizes many of the techniques of Ref. [11] and thus is both scalable (with time-complexity $O(n^4)$) and independent of SPAM errors. Finally, we experimentally demonstrate this protocol on a superconducting qubit, extracting a gate error of 0.003 with theoretical lower or upper bound of [0,0.016] for both $X_{\pi/2}$ and $Y_{\pi/2}$ gates ($U_\theta$ is a rotation of $\theta$ around axis $U$). This error is smaller than the gate errors extracted via QPT ($0.011^{+0.011}_{-0.009}$ and $0.020^{+0.009}_{-0.008}$, respectively). A similar technique was recently employed to benchmark two qubits gates in an ion trap system [26]. Here we provide a general expression for individual gate error for an arbitrary number of qubits.

*Interleaving benchmarking protocol.*—To benchmark the Clifford element $\mathcal{C}$, which has an associated noise operator $\Lambda_\mathcal{C}$, we fix an initial state $|\psi\rangle$ and perform the following steps:

*Step 1*: Implement standard randomized benchmarking [see Fig. 1(a)] which, for completeness, we briefly summarize here (additional details in Refs. [11,27]). For various values of $m$, choose $K$ sequences of random gates where the first $m$ gates are chosen uniformly at random from $\text{Clif}_n$. The $(m + 1)$th gate is chosen to be the inverse of the composition of the first $m$ random gates and can be found efficiently by the Gottesman-Knill theorem [28]. Assuming each Clifford element $\mathcal{C}_{i_j}$ for each step $j$ has some associated error, $\Lambda_{i_j}$, the sequence of gates is modeled by

$$\mathcal{S}_{\mathbf{i_m}} = \Lambda_{i_{m+1}} \circ \mathcal{C}_{i_{m+1}} \circ (\circ_{j=1}^m [\Lambda_{i_j} \circ \mathcal{C}_{i_j}]), \quad (1)$$

where $\circ$ is a composition, $\mathbf{i_m}$ is the $m$-tuple $(i_1, \ldots, i_m)$ and $i_{m+1}$ is uniquely determined by $\mathbf{i_m}$. Next, measure the probability that the initial state is not changed by the sequence, $\text{Tr}[E_\psi \mathcal{S}_{\mathbf{i_m}}(\rho_\psi)]$, which we call the "survival probability." Here, $\rho_\psi$ is a quantum state that takes into account state-preparation errors and $E_\psi$ is the positive operator valued measure element that takes into account measurement errors. In the ideal (noise-free) case the survival probability will be 1 for each sequence. Averaging the survival probability over the $K$ sequences gives the sequence fidelity $F_{\text{seq}}(m, \psi)$ and a fit to either the zeroth or first-order model:
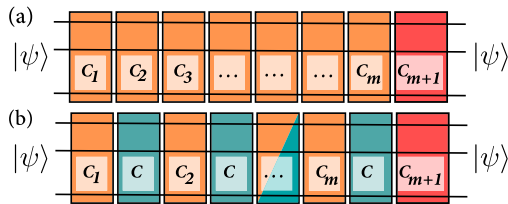


(a)

(b)

FIG. 1 (color online). Randomized benchmarking protocols. (a)–(b) Schemes for the standard and interleaved benchmarking protocols. The target gate, $\mathcal{C}$ (green) is interleaved with random gates $\mathcal{C}_i$ (orange) chosen from $\text{Clif}_n$. A final gate $\mathcal{C}_{m+1}$ (red) is performed to make the total sequence equal to the identity operation.

$$F_{\text{seq}}^{(0)}(m, \psi) = A_0 p^m + B_0,$$
$$F_{\text{seq}}^{(1)}(m, \psi) = A_1 p^m + C_1(m - 1)p^{m-2} + B_1, \quad (2)$$

gives the depolarizing parameter $p$ (the average error rate over all Clifford gates is given by $r = (d - 1)(1 - p)/d$, where $d = 2^n$ is the dimension of the system. The coefficients $A_{1(0)}$, $B_{1(0)}$, and $C_1$ absorb the state preparation and measurement errors as well as the error on the final gate.

*Step 2*: Choose $K$ sequences of Clifford elements where the first Clifford $\mathcal{C}_{i_1}$ in each sequence is chosen uniformly at random from $\text{Clif}_n$, the second is always chosen to be $\mathcal{C}$, and alternate between uniformly random Clifford elements and deterministic $\mathcal{C}$ up to the $m$th random gate [see Fig. 1(b)]. The $(m + 1)$th gate is chosen to be the inverse of the composition of the first $m$ random gates and $m$ interlaced $\mathcal{C}$ gates (we adopt the convention of defining the length of a sequence by the number of random gates). The superoperator representing the sequence is

$$\mathcal{V}_{\mathbf{i_m}} = \Lambda_{i_{m+1}} \circ \mathcal{C}_{i_{m+1}} \circ (\circ_{j=1}^m [\mathcal{C} \circ \Lambda_\mathcal{C} \circ \Lambda_{i_j} \circ \mathcal{C}_{i_j}]). \quad (3)$$

For each of the $K$ sequences, measure the survival probability $\text{Tr}[E_\psi \mathcal{V}_{\mathbf{i_m}}(\rho_\psi)]$ and average over the $K$ random sequences to find the new sequence fidelity $F_{\overline{\text{seq}}}(m, \psi)$. Fit $F_{\overline{\text{seq}}}(m, \psi)$ to one of the new zeroth or first-order models to obtain the depolarizing parameter $p_{\tilde{\mathcal{C}}}$. The expressions for these models are given by Eq. (2) where $p$ is replaced by the new depolarizing parameter $p_{\tilde{\mathcal{C}}}$.

*Step 3*: From the values obtained for $p$ (Step 1) and $p_{\tilde{\mathcal{C}}}$ (Step 2), the gate error of $\Lambda_\mathcal{C}$ (which is exactly given by $r_\mathcal{C} = 1$- average gate fidelity of $\Lambda_\mathcal{C}$) is estimated by

$$r_\mathcal{C}^{\text{est}} = \frac{(d - 1)(1 - p_{\tilde{\mathcal{C}}}/p)}{d}, \quad (4)$$

and must lie in the range $[r_\mathcal{C}^{\text{est}} - E, r_\mathcal{C}^{\text{est}} + E]$ where

$$E = \min \begin{cases} \frac{(d-1)[|p - p_{\tilde{\mathcal{C}}}/p| + (1-p)]}{d} \\ \frac{2(d^2-1)(1-p)}{pd^2} + \frac{4\sqrt{1-p}\sqrt{d^2-1}}{p}. \end{cases} \quad (5)$$

One interpretation of $E$ is that it arises from imperfect random gates. To see this, first note that in the limit of perfect random gates, $p \to 1$, $r_\mathcal{C}^{\text{est}}$ goes to the standard error for a depolarizing channel with strength $p_\mathcal{C}$ (equivalently $r_\mathcal{C}^{\text{est}} \to r_\mathcal{C}$), and $E$ goes to zero. In the more specific case of $\Lambda$ being a Pauli channel, one can replace the second possibility in Eq. (5) with $2(d^2 - 1)(1 - p)/pd^2$ and when $\Lambda$ is depolarizing, $E = 0$. In the typical case, we expect that $\Lambda$ will be close to a depolarizing channel and the above general bounds will over-estimate the gate error.

*Experimental implementation.*—Using the new protocol, we verified the performance of two single-qubit gates on a superconducting transmon qubit [29]. The device is similar to the one described in Ref. [30], but we focus on just a single qubit with $\omega_{01}/2\pi = 5.4493$ GHz, anharmonicity

of $(\omega_{12} - \omega_{01})/2\pi = -228$ MHz, and coherence times of $T_1 = 5.0$ $\mu$s and $T_2^{\text{echo}} = 3.2$ $\mu$s.

Single-qubit control was performed by means of shaped microwave pulses applied to capacitively coupled bias lines that address individual qubits. We used Gaussian shaped pulses with a derivative envelope applied to an orthogonal quadrature to minimize errors due to higher levels of the transmon [31]. The Gaussian width was $\sigma = 5$ ns and the pulses were truncated to have a total duration of $4\sigma = 20$ ns. A pulse calibration procedure was used which employed several sequences of repeated pulses that amplify small rotation angle and phase errors. A Levenberg-Marquardt search provided all calibrated pulse parameters in a few minutes.

To perform standard randomized benchmarking, we chose a Clifford generating set of $s = \{I, X_{\pm\pi/2}, X_\pi, Y_{\pm\pi/2}, Y_\pi\}$. Each Clifford gate in a random sequence is performed by a random choice from the set of minimal length constructions of that gate. For the generating set $s$, a Clifford gate has an average length of 1.875 pulses. To find the average fidelity for sequences of length $N$, we create 32
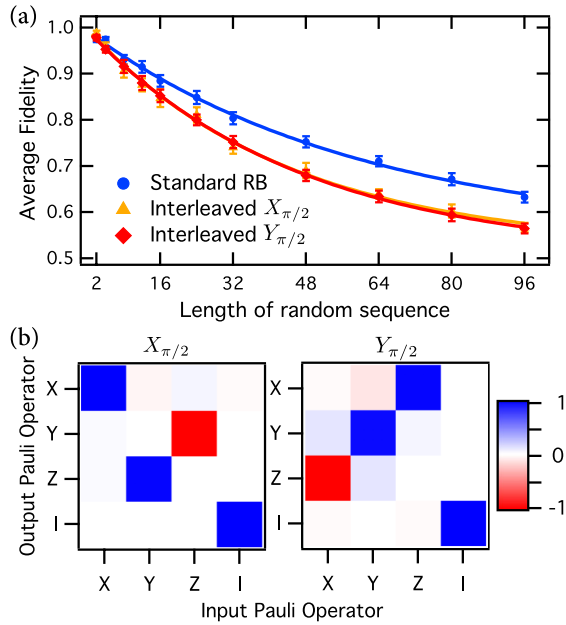


FIG. 2 (color online). Experimental implementation of inter-leaved RB. (a) Measurement of average fidelity over 32 random sequences each of lengths between 2 and 96. The blue data (circles) show the result of the standard RB protocol, while red (triangles) and orange (diamonds) data correspond to interleaved sequences for $X_{\pi/2}$ and $Y_{\pi/2}$ gates, respectively. All data are well described by the first model of Eq. (2), with $p = 0.984 \pm 0.004$ (standard RB), $p_{\bar{c}} = 0.978 \pm 0.005$ (interleaved $X_{\pi/2}$) and $p_{\bar{c}} = 0.979 \pm 0.001$ (interleaved $Y_{\pi/2}$) to 95% confidence. Error bars are the standard error of the mean of each point. (b) Pauli transfer maps from process tomography of the $X_{\pi/2}$ and $Y_{\pi/2}$ gates with corresponding gate errors of $0.011^{+0.011}_{-0.009}$ and $0.020^{+0.009}_{-0.008}$, respectively.

random sequences of $N + 1$ Clifford gates, measure $\langle\sigma_z\rangle$ at the end of each, and then average the results. Figure 2(a) shows the measured average fidelities (blue circles) versus sequence length. The data fit well to the first model of Eq. (2) with $p = 0.984 \pm 0.004$, corresponding to an esti-mated average error rate for the entire Clifford group of $r = (1 - p)/2 = 0.008 \pm 0.002$, which is in reasonable agreement with the expected error of 0.006 from decoherence.

Since the Clifford generating rotations in $s$ can each be implemented with just a single pulse, we expect a lower error rate for such gates than the average over the entire group. We verify this for $X_{\pi/2}$ and $Y_{\pi/2}$ gates with inter-leaved benchmarking. The resulting average fidelities for the $X_{\pi/2}$ and $Y_{\pi/2}$ interleaved sequences are shown in Fig. 2(a) as orange triangles and red diamonds, respec-tively. The fidelities are lower than the standard RB results because of an effective doubling in the number of pulses in the interleaved sequences. These sequences fit to a model with the $p_{\bar{c}} = 0.978 \pm 0.005$ and $0.979 \pm 0.001$. By Eq. (4), this gives our best estimated error rate for $X_{\pi/2}$ and $Y_{\pi/2}$ gates of $r_{\mathcal{C}}^{\text{est}} = 0.003 \pm 0.003$ and Eq. (5) pro-vides bounds of $[0, 0.016]$.

To further test the robustness of the technique, we also intentionally introduce additional error on a target gate to test the sensitivity of the interleaved benchmarking proto-col to calibration errors. These results are summarized in Table I. For the small set of calibration errors introduced, the model reliably tracks the anticipated pulse infidelity.

We compare the interleaved RB result to the standard method of measuring gate performance by performing QPT. The process matrices for the $X_{\pi/2}$ and $Y_{\pi/2}$ gates are shown in Fig. 2(b) in the Pauli basis of the Liouville representation (also known as the Pauli transfer map, see Ref. [30]). To extract these maps we employ maximum likelihood estimation to ensure that the maps are physical (we require the maps to be completely positive, but allow them to be non-trace-preserving because of potential leak-age out of the qubit space). The gate errors extracted from these maps are $1 - \mathcal{F} = 0.011^{+0.011}_{-0.009}$ and $0.020^{+0.009}_{-0.008}$ re-spectively. We attribute the increase in error seen in QPT to SPAM errors. Additionally, the use of maximum like-lihood estimation leads to difficulties in assigning error bars to the fidelities through Monte Carlo bootstrapping. Consequently, interleaved benchmarking provides a more reliable estimate for the performance of Clifford gates.

*Derivation of the fitting models, gate errors, and bounds.*—The main idea behind the derivation of the fitting models is the following "unitary 2-design" property of the Clifford group: If $\Lambda$ is a quantum channel and $\text{Clif}_n = \{\mathcal{C}\}$ then the "twirl" of $\Lambda$, $\mathcal{W}(\Lambda)$, defined by

$$\mathcal{W}(\Lambda) := \frac{1}{|\text{Clif}_n|} \sum_{j=1}^{|\text{Clif}_n|} \mathcal{C}_j \circ \Lambda \circ \mathcal{C}_j^\dagger \qquad (6)$$

is the unique depolarizing channel $\Lambda_d$ with the same average fidelity as $\Lambda$ [3]. The average fidelity of $\Lambda$ is given by

$$\bar{\mathcal{F}}_{\Lambda,I} := \overline{\mathrm{tr}(|\phi\rangle\langle\phi|\Lambda(|\phi\rangle\langle\phi|))}, \quad (7)$$

which is just the average over all pure states $|\phi\rangle\langle\phi|$ of the usual fidelity between the output state $\Lambda(|\phi\rangle\langle\phi|)$ and input state $|\phi\rangle\langle\phi|$. Hence if $\Lambda_d$ is given by $\Lambda_d(\rho) = p\rho + (1-p)\frac{1}{d}$ then $\bar{\mathcal{F}}_{\Lambda,I} = p + \frac{(1-p)}{d}$.

We now provide a brief overview of the derivation of the fitting models. Defining $\mathcal{D}_{i_j} = \mathcal{C}_{i_j} \circ \bigcirc_{s=1}^{j-1}[\mathcal{C} \circ \mathcal{C}_{i_s}]$ allows us to write the interleaving sequence as

$$\mathcal{V}_{\mathbf{i_m}} = (\Lambda_{i_{m+1}}) \circ (\bigcirc_{j=1}^{m}[\mathcal{D}_{i_j}{}^{\dagger} \circ \Lambda_{\mathcal{C}} \circ \Lambda_{i_j} \circ \mathcal{D}_{i_j}]). \quad (8)$$

The zeroth order model corresponds to the noise being independent of the gate; i.e., $\Lambda_{i_j} = \Lambda$ is independent of $\mathcal{D}_{i_j}$ for every $j$. In this case when we average over many sequences in Eq. (8), we obtain a composition of twirls of $\Lambda_{\mathcal{C}} \circ \Lambda$. Hence, we obtain a composition of depolarizing channels $\Lambda_{\bar{\mathcal{C}},d} = (\Lambda_{\mathcal{C}} \circ \Lambda)_d$, where for any state $\rho$, $\Lambda_{\bar{\mathcal{C}},d}(\rho) = p_{\bar{\mathcal{C}}}\rho + (1-p_{\bar{\mathcal{C}}})\frac{1}{d}$. Here, $1 - p_{\bar{\mathcal{C}}}$ corresponds to the depolarizing strength of $\Lambda_{\mathcal{C}} \circ \Lambda$.

The first-order model corresponds to the case where the noise depends on the gate. In this case, we apply a perturbative argument similar in nature to that of Ref. [11] (for more details see Ref. [27]) to derive the fitting model. Each $\Lambda_i$ is perturbed about the average of all the $\Lambda_i$, denoted by $\Lambda$, and provided the average variation of the $\|\delta\Lambda_i\|$, $\gamma := \frac{1}{\mathrm{Clif}_n}\sum_i\|\delta\Lambda_i\|$, is small (i.e., $\gamma^2 \ll 2/[m(m+1)]$) the first-order model is a valid description of the fidelity decay curve. Note that the norm $\|\cdot\|$ can be any norm satisfying certain general properties (see Ref [27] for more detail). One usually chooses the weakest norm satisfying these properties which allows for the largest class of gate-dependent errors. It is important to emphasize that the type of the noise is irrelevant for this sufficient condition, as long as the average of the magnitudes is sufficiently small, the analysis can be terminated at first order.

We now outline how to obtain the expression for the error given by Eq. (4) as well as the various expressions for

$E$ given by Eq. (5) and in the surrounding text. Let us begin by looking at the difference in average fidelity between $\Lambda_{\bar{\mathcal{C}}} = \Lambda_{\mathcal{C}} \circ \Lambda$ and $\Lambda_{\tilde{\mathcal{C}}} := \Lambda_{\mathcal{C}} \circ \Lambda_d$, $|\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - \bar{F}_{\Lambda_{\tilde{\mathcal{C}}},I}|$. Since $\Lambda_d$ is depolarizing,

$$|\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - \bar{F}_{\Lambda_{\tilde{\mathcal{C}}},I}| = \left| \bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - p\bar{F}_{\Lambda_{\mathcal{C}},I} - \frac{1-p}{d} \right|. \quad (9)$$

Hence,

$$\frac{\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I}}{p} - \frac{1-p}{dp} - E \leq \bar{F}_{\Lambda_{\mathcal{C}},I} \leq \frac{\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I}}{p} - \frac{1-p}{dp} + E, \quad (10)$$

where $E$ is an upper bound for $|\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - \bar{F}_{\Lambda_{\tilde{\mathcal{C}}},I}|/p$. Using $r_{\mathcal{C}} := 1 - \bar{F}_{\Lambda_{\mathcal{C}},I}$ and $\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} = p_{\bar{\mathcal{C}}} + \frac{(1-p_{\bar{\mathcal{C}}})}{d}$ we find Eq. (4).

The first expression in Eq. (5) can be obtained by noting that $|\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - \bar{F}_{\Lambda_{\tilde{\mathcal{C}}},I}|$ can be upper bounded by

$$\left| \bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - p\bar{F}_{\Lambda,I} - \frac{(1-p)}{d} \right| + p|\bar{F}_{\Lambda,I} - \bar{F}_{\Lambda_{\mathcal{C}},I}|$$
$$\leq \frac{(d-1)[|p_{\bar{\mathcal{C}}} - p^2| + (p - p^2)]}{d}, \quad (11)$$

where we have used $|\bar{F}_{\Lambda,I} - \bar{F}_{\Lambda_{\mathcal{C}},I}| \leq \frac{(d-1)(1-p)}{d}$.

The second expression in Eq. (5) is obtained by first noting that,

$$|\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - \bar{F}_{\Lambda_{\tilde{\mathcal{C}}},I}| \leq \|\Lambda - \Lambda_d\|_{\diamond}, \quad (12)$$

where $\|\|_{\diamond}$ is the "diamond norm" [32]. By the triangle inequality,

$$|\bar{F}_{\Lambda_{\bar{\mathcal{C}}},I} - \bar{F}_{\Lambda_{\tilde{\mathcal{C}}},I}| \leq \|\Lambda - I\|_{\diamond} + \frac{2(d^2-1)(1-p)}{d^2}, \quad (13)$$

where $\|\Lambda_d - I\|_{\diamond} = \frac{2(d^2-1)(1-p)}{d^2}$ [27]. It can be shown that for arbitrary $\Lambda$ [33],

$$\|\Lambda - I\|_{\diamond} \leq 4\sqrt{1-p}\sqrt{d^2-1}, \quad (14)$$

which gives the second expression in Eq. (5).

In the case of $\Lambda$ being equal to a Pauli channel, $\|\Lambda - I\|_{\diamond} = 2(d^2-1)(1-p)/d^2$ always holds [27]. Lastly, the depolarizing case is obtained by noting that $\Lambda_{\bar{\mathcal{C}}} = \Lambda_{\tilde{\mathcal{C}}}$ and as such $E = 0$ by definition.

*Conclusion.*—We have presented a scalable protocol for benchmarking individual quantum gates. We explicitly derive various bounds for the error of the imperfect gate in terms of parameters that are output from the protocol. The gate error can be estimated exactly in the limit of perfect gates or if the average of the error operators over all gates is depolarizing, which we believe is close to the typical case. The method is scalable in the size of the quantum system and is independent of SPAM errors. We have applied this protocol to a superconducting qubit and shown the gate errors for each of $X_{\pi/2}$ and $Y_{\pi/2}$ rotations to be lower estimates than those obtained using QPT.

TABLE I. Gate errors extracted with interleaved RB for intentional pulse miscalibration errors of $X_{\pi/2}$. The first column is the applied over-rotation about the $X$ axis with predicted $\Lambda_{\mathcal{C}} = \exp[-i\epsilon\sigma_x/2]$, the second column is found using $r_{\mathrm{th}} = 2[1 - \cos^2(\epsilon/2)]/3$, the third is the experimentally extracted gate errors via interleaving with fit uncertainties to 95% confidence, and the fourth is the bounds from Eq. (5).

| Amp. error $\epsilon$ | $r_{\mathrm{th}}$ | $r_{\mathcal{C}}^{\mathrm{est}}$ | $r_{\mathcal{C}}^{\mathrm{est}}$ bound |
|---|---|---|---|
| 0.0 | 0.000 | $0.003 \pm 0.003$ | [0,0.016] |
| $\pi/20$ | 0.004 | $0.011 \pm 0.004$ | [0,0.022] |
| $\pi/10$ | 0.016 | $0.029 \pm 0.008$ | [0,0.058] |

[1] I. Chuang and M. Nielsen, J. Mod. Opt. **44,** 2455 (1997).

[2] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78,** 390 (1997).

[3] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80,** 012304 (2009).

[4] J. Emerson, R. Alicki, and K. Zyczkowski, J. Opt. B **7,** S347 (2005).

[5] B. Levi, C. C. Lopez, J. Emerson, and D. G. Cory, Phys. Rev. A **75,** 022314 (2007).

[6] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh D. G. Cory, and R. Laflamme, Science **317,** 1893 (2007).

[7] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, Phys. Rev. A **78,** 012347 (2008).

[8] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett. **100,** 190403 (2008).

[9] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107,** 210404 (2011).

[10] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106,** 230501 (2011).

[11] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106,** 180504 (2011).

[12] 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).

[13] M. J. Biercuk *et al.*, Quantum Inf. Comput. **9,** 0920 (2009).

[14] K. R. Brown, A. C. Wilson, Y. Colombe, C. Ospelkaus, A. M. Meier, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. A **84,** 030303 (2011).

[15] C. Ryan, M. Laforest, and R. Laflamme, New J. Phys. **11,** 013034 (2009).

[16] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **102,** 090502 (2009).

[17] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **82,** 040305 (2010).

[18] H. Paik *et al.*, Phys. Rev. Lett. **107,** 240501 (2011).

[19] S. Olmschenk, R. Chicireanu, K. D. Nelson, and J. V. Porto, New J. Phys. **12,** 113007 (2010).

[20] O. Moussa, M. P. da Silva, C. A. Ryan, R. Laflamme, arXiv:1112.4505.

[21] S. Bravyi and A. Kitaev, Phys. Rev. A **71,** 022316 (2005).

[22] P. Shor, in *Proceedings of the 37th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE Press, Burlington, Vermont, 1996).

[23] D. Aharonov and M. Ben-Or, in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)* (ACM, New York, 1997).

[24] E. Knill, R. Laflamme, and W. Zurek, Proc. R. Soc. A **454,** 365 (1998).

[25] J. Preskill, arXiv:quant-ph/9712048.

[26] J. P. Gaebler *et al.*, arXiv:1203.3733.

[27] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85,** 042311 (2012).

[28] D. Gottesman, arXiv:quant-ph/9705052.

[29] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **76,** 042319 (2007).

[30] J. M. Chow *et al.*, Phys. Rev. Lett. **109,** 060501 (2012).

[31] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. Wilhelm, Phys. Rev. Lett. **103,** 110501 (2009).

[32] A. Kitaev, Russ. Math. Surv. **52,** 1191 (1997).

[33] B. Salman and R. Koenig, New J. Phys. **13,** 093036 (2011).