

Investigating the limits of randomized benchmarking protocols

Jeffrey M. Epstein, Andrew W. Cross, Easwar Magesan, and Jay M. Gambetta
IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

(Received 13 August 2013; published 18 June 2014)

In this paper, we analyze the performance of randomized benchmarking protocols on gate sets under a variety of realistic error models that include systematic rotations, amplitude damping, leakage to higher levels, and $1/f$ noise. We find that, in almost all cases, benchmarking provides better than a factor-of-2 estimate of average error rate, suggesting that randomized benchmarking protocols are a valuable tool for verification and validation of quantum operations. In addition, we derive models for fidelity decay curves under certain types of non-Markovian noise models such as $1/f$ and leakage errors. We also show that, provided the standard error of the fidelity measurements is small, only a small number of trials are required for high-confidence estimation of gate errors.

DOI: [10.1103/PhysRevA.89.062321](https://doi.org/10.1103/PhysRevA.89.062321)

PACS number(s): 03.67.Ac, 03.67.Pp, 03.65.Wj, 03.67.Lx

I. INTRODUCTION

The advancement of experimental quantum information processing requires a method to benchmark errors on quantum gates. These benchmarks provide straightforward methods for comparing different experimental implementations and also establish compliance with error thresholds for processes such as error correction [1]. The standard method for characterizing errors is quantum process tomography (QPT) [2,3], which provides complete error reconstruction. Implementing QPT comes at a significant price though, since its complexity scales exponentially in order to determine the 16^n real parameters of the n -qubit quantum error process. In addition, QPT is vulnerable to state preparation and measurement (SPAM) errors, which are the errors associated with preparing and measuring different states. Because these errors may be on the same order as the error on the gate of interest, they can cause significant inaccuracies in the reconstructed errors. A recent study found that QPT overestimated small errors by several orders of magnitude [4].

An alternative to QPT is randomized benchmarking (RB) [5–11]. Because this method extracts specific parameters of interest from the noise, as opposed to the complete set of parameters obtained from QPT, it does not suffer from exponential scaling. RB is also impervious to SPAM errors since it examines fidelity decays over random gate sequences. RB protocols have become an important tool for quantum verification and validation, and have been used to benchmark one- and two-qubit gates in atomic ions [5,9,12–14], nuclear magnetic resonance [15], and superconducting qubit [8,10,11,16,17] experiments. RB protocols and their fidelity decay models are provably valid in a wide variety of scenarios; however, assumptions about the form of the fidelity decay under complex noise models may introduce inaccuracies. As well, there have been concerns about the convergence of the estimate from finite-sampling effects since typical experiments are performed with many fewer random sequences than predicted using the Hoeffding bound [6]. For these reasons, it is important to study and develop models in the limits where analytic support for current RB methods may be lacking.

In this paper, we address many of these issues by providing theoretical results on finite-sampling effects and modeling decay curves for realistic types of noise. As well, we test the

performance of RB in different scenarios by implementing numerical simulations of both standard [5,6] and interleaved [8,9] RB protocols under several physically relevant single-qubit error models. Two main classes of noise were tested; Markovian and non-Markovian. Markovian noise is memoryless and as such is history independent, while non-Markovian noise is history dependent, so the noise at one moment may depend on previous gates in the sequence. The first type of Markovian noise we investigated was systematic rotations represented by both random and fixed unitary operators. These reflect the effects of gate calibration errors and control field fluctuations. The second type of Markovian noise was amplitude damping, which can represent the process of spontaneous emission. The non-Markovian noise we considered was $1/f$ noise and leakage to higher levels outside the qubit manifold. $1/f$ noise is ubiquitous in nature [18–20] and is present to some degree in most physical implementations of qubits [21–25], although its relative importance is implementation dependent. Leakage can plague a variety of systems, including transmons [26], phase qubits [27], and quantum dots [28]. We discuss models of the fidelity decay curves for $1/f$ and leakage noise and perform numerical simulations under these models.

The structure of our paper is as follows. In Sec. II we introduce the RB protocols and in Sec. III we describe the simulation methods used throughout the presentation. Section IV provides a theoretical and numerical analysis of finite-sampling effects in RB and results for simulations of standard RB with Markovian noise. In Sec. V we describe the model and simulation of $1/f$ noise and provide the results of RB for this $1/f$ noise model. In Sec. VI we provide a theoretical analysis for modeling fidelity decay under a leakage noise model and present numerical results. In Sec. VII we discuss interleaved RB and present results of different noise simulations. We make concluding remarks in Sec. VIII.

II. RANDOMIZED BENCHMARKING PROTOCOLS

The underlying idea behind an RB protocol is to apply sequences of randomly chosen gates from some group and measure fidelity decay as a function of sequence length [29]. Ideally, this allows one to extract average properties of the errors associated with implementing these gates in real quantum devices. The standard protocol [6,7], which extends

[5], chooses the gates from the Clifford group and gives an estimate of the average gate fidelity F_g , or, equivalently, error rate over the group. Interleaved benchmarking [8,9] extends the standard protocol to estimate the average gate fidelity of an individual gate.

For some group \mathbf{G} of unitary operations, the general RB protocol is as follows [6,7]:

(1) Choose gates from the group $\mathbf{G} = \{U_i\}$ to form K sequences of each length m from some set $\{m\}$ of sequence lengths.

(2) For each sequence U_1, \dots, U_m , determine the gate $U_{m+1} = (U_m \dots U_1)^\dagger$.

(3) Apply each sequence U_1, \dots, U_{m+1} to some initial state ρ_i , measure the output state ρ_f , and repeat several times to determine the survival probability of some output state for each sequence.

(4) Average this survival probability over all sequences of the same length, and fit to a predetermined model.

(5) From this model determine the desired quantities of the map.

Unless otherwise specified, quantum channels will be expressed in the Pauli transfer matrix (PTM) representation [30], in which a matrix corresponding to the quantum channel \mathcal{R} on the space of density operators on the d -dimensional Hilbert space of an n -qubit system is defined such that $\rho' = \sum_{i,j} \mathcal{R}_{ij} P_i \text{tr}(P_j \rho) / d$, where $P_0 = I^{\otimes n}$, $P_1 = I^{\otimes n-1} \otimes X$, $P_2 = I^{\otimes n-1} \otimes Y$, etc. For dimension $d = 2^n$, the representation is of dimension $d^2 \times d^2$, as the density operators on n qubits are spanned by the d^2 n -qubit Pauli operators $\mathbf{P} = \{I, X, Y, Z\}^{\otimes n}$. We will use a calligraphic font to denote a quantum channel (or map) and a standard math font for an operator.

The channel representing the average sequence of length m can be written as

$$\mathcal{S}^{(m)} = \frac{1}{K} \sum_{\mathbf{i}} \mathcal{S}_{\mathbf{i}}^{(m)}, \quad (2.1)$$

where the sum is over the K sequences $\mathbf{i} = (i_1, \dots, i_m)$ with

$$\mathcal{S}_{\mathbf{i}} = \mathcal{E}_{i_{m+1}} \mathcal{U}_{i_{m+1}} \left(\prod_{j=1}^m \mathcal{E}_{i_j} \mathcal{U}_{i_j} \right). \quad (2.2)$$

Here \mathcal{E}_{i_j} is the noise on gate \mathcal{U}_{i_j} implemented at time j with history (i_1, \dots, i_{j-1}) . Since $\{\mathcal{U}_{i_j}\}$ is a group, the sequence can be rewritten as [7]

$$\mathcal{S}_{\mathbf{i}} = \mathcal{E}_{i_{m+1}} \left(\prod_{j=1}^m \tilde{\mathcal{U}}_{i_j}^T \mathcal{E}_{i_j} \tilde{\mathcal{U}}_{i_j} \right), \quad (2.3)$$

where $\tilde{\mathcal{U}}_{i_j}$ is another element of \mathbf{G} and all sequences $\tilde{\mathcal{U}}_1, \dots, \tilde{\mathcal{U}}_m$ are uniformly distributed within the ensemble. In the limit where \mathcal{E}_{i_j} can be approximated by the average error $\bar{\mathcal{E}}$, the average sequence can be represented by [7]

$$\mathcal{S}^{(m)} = \bar{\mathcal{E}} (\bar{\mathcal{E}}_{\mathbf{G}})^m. \quad (2.4)$$

Here $\bar{\mathcal{E}}_{\mathbf{G}}$ represents the twirl over the group \mathbf{G} and is given by

$$\bar{\mathcal{E}}_{\mathbf{G}} = \frac{1}{|\mathbf{G}|} \sum_{\mathcal{U} \in \mathbf{G}} \mathcal{U}^T \bar{\mathcal{E}} \mathcal{U}, \quad (2.5)$$

which is just a group average. Depending on the group \mathbf{G} , this channel can have a simple structure with a small number of parameters which may be determined by fitting the measured fidelities to the fidelity decay model (FDM)

$$F(m) = \text{tr}[E \mathcal{S}^{(m)} \rho] = \tilde{\mathbf{e}}^T (\bar{\mathcal{E}}_{\mathbf{G}})^m \mathbf{p}, \quad (2.6)$$

where $\rho = \mathbf{p}^T \mathbf{P} / d$ represents the initial state and $\tilde{E} = \tilde{\mathbf{e}}^T \mathbf{P}$ represents the measurement (E) and the error in the final gate ($\bar{\mathcal{E}}$).

From the above there are two important assumptions that need to be addressed.

Assumption 1: Finite sampling. The sample average fidelity converges to the average over all possible sequences for small sample sizes. Because of the length of the sequences used, it is infeasible to implement more than a very small fraction of all possible sequences of each length.

Assumption 2: Noise homogeneity. The average variation of the errors is weak so that most errors are close to the average. In practice, this may not always be satisfied since the errors may have strong gate dependence (calibration errors, etc.), time dependence (control field power fluctuations, etc.), or history dependence (leakage to higher levels, $1/f$ noise etc.).

A. Standard Clifford benchmarking

Standard Clifford randomized benchmarking (SRB) estimates the average error rate of the errors on the members of the full n -qubit Clifford group. The gates are chosen from this group and Schur's lemma tells us that Eq. (2.5) gives the depolarizing channel

$$\bar{\mathcal{E}}_{\mathbf{G}} = \begin{pmatrix} 1 & & & \\ & \alpha & & \\ & & \ddots & \\ & & & \alpha \end{pmatrix}, \quad (2.7)$$

where the basis is ordered such that I is first. The system is prepared in any initial state and the FDM, Eq. (2.6), becomes

$$F = A \alpha^m + \tilde{e}_0, \quad (2.8)$$

where the constants $A = \sum_{j \neq 0} \tilde{e}_j p_j$ and \tilde{e}_0 absorb all SPAM errors. In the case that there are no SPAM errors $\tilde{e}_0 = 1/d$ and $A = 1/d$. As shown in Ref. [7] the average error rate is estimated by $\hat{r} = (1 - \alpha)(d - 1)/d$.

B. Interleaved randomized benchmarking

Interleaved randomized benchmarking (IRB) allows estimation of the error on an individual gate U_{int} . The essential idea is to perform two benchmarking experiments; one identical to the standard method described above (which gives the average error depolarizing parameter α for the Clifford gates $\{\mathcal{U}_i\}$), and one in which the gate of interest is inserted (interleaved) between each randomly chosen gate in each sequence to give the depolarizing parameter $\tilde{\alpha}_{\text{int}}$ for the gates $\{\mathcal{U}_{\text{int}} \mathcal{U}_i\}$.

Step 1. Perform Clifford benchmarking as described in the previous section to obtain the average depolarizing parameter α of the errors on the Clifford gates \mathcal{U}_i .

Step 2. Repeat Step 1, but insert the gate of interest after each of the randomly selected Clifford gates. Then the sequences

may be expressed as

$$\mathcal{S}_i = \mathcal{E}_{i_{m+1}} \mathcal{U}_{i_{m+1}} \left(\prod_{j=1}^m \mathcal{U}_{i_{\text{int},j}} \mathcal{E}_{i_j} \mathcal{U}_{i_j} \right) \quad (2.9)$$

for $\mathcal{E}_{i_{\text{int},j}}$ the error on the interleaved gate at time j . As before, the group structure permits the sequences to be rewritten as $\mathcal{S}_i = \mathcal{E}_{i_{m+1}} (\prod_{j=1}^m \tilde{\mathcal{U}}_{i_j}^T \mathcal{E}_{i_{\text{int},j}} \mathcal{E}_{i_j} \tilde{\mathcal{U}}_{i_j})$, which has the same form as Eq. (2.3). The interleave estimate for the depolarizing parameter corresponding to the error on the gate of interest is $\alpha_{\text{int}} = \bar{\alpha}_{\text{int}}/\alpha$. The estimated error rate is $\hat{r}_{\text{int}} = (1 - \alpha_{\text{int}})(d - 1)/d$ [8,9]. Note that this estimate is provably valid under the following assumption.

Assumption 3: Product twirl. On average, the twirl of the product of two channels is well approximated by the product of twirls. This approximation is exactly correct in the case that at least one of the factor gates is depolarizing, but not in general. A pathological case is when the error on the interleaved gate partially inverts the error on the prior Clifford gate, in which case IRB can underestimate the error rate.

We note that, even when Assumption 3 is not satisfied, it is possible to obtain bounds on the gate error of \mathcal{U}_{int} [8,31].

III. SIMULATION METHODS

In the numerics we consider only single-qubit Clifford groups, and in this case there are 24 different Clifford operators. A convenient way to decompose these is to introduce the Pauli group $\mathbf{P} = \{\mathbb{1}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, the exchange group $\mathbf{S} = \{\mathbb{1}, \mathcal{S}, \mathcal{S}^2\}$, and the Hadamard group $\mathbf{H} = \{\mathbb{1}, \mathcal{H}\}$. The Pauli group is represented by the maps

$$\begin{aligned} \mathcal{X} &= \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}, & \mathcal{Y} &= \begin{pmatrix} & & & 1 \\ & & -1 & \\ & & & 1 \\ & & & & -1 \end{pmatrix}, \\ \mathcal{Z} &= \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & 1 \end{pmatrix}, \end{aligned} \quad (3.1)$$

which just correspond to π rotations around the x , y , and z axes, respectively. The exchange-axis group

$$\mathcal{S} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \quad \mathcal{S}^2 = \begin{pmatrix} 1 & & & \\ & & & 1 \\ & & 1 & \\ & & & 1 \end{pmatrix} \quad (3.2)$$

exchanges $(x, y, z) \rightarrow (z, x, y) \rightarrow (y, z, x)$ and the Hadamard group

$$\mathcal{H} = \begin{pmatrix} 1 & & & \\ & & & 1 \\ & & -1 & \\ & & & 1 \end{pmatrix} \quad (3.3)$$

exchanges $(x, y, z) \rightarrow (z, -y, x)$. The single-qubit Clifford group is the group generated by all combinations of elements in \mathbf{H} , \mathbf{P} , and \mathbf{S} , and has size $2 \times 3 \times 4 = 24$. It is worth noting that the group formed by all combinations of elements in \mathbf{P} and \mathbf{S}

TABLE I. A list of the 24 Clifford operators and their decomposition into either physically relevant generators or simple mathematical elements. The operators above the horizontal line form both a group and a two-design. The $-$ signifies application in time and the mean number of physical generators per Clifford group is 1.875.

Clifford elements	Physical decomposition
$\mathbb{1} - \mathbb{1} - \mathbb{1}$	$\mathbb{1}$
$\mathbb{1} - \mathbb{1} - \mathcal{S}$	$\mathcal{Y}_{\pi/2} - \mathcal{X}_{\pi/2}$
$\mathbb{1} - \mathbb{1} - \mathcal{S}^2$	$\mathcal{X}_{-\pi/2} - \mathcal{Y}_{-\pi/2}$
$\mathcal{X} - \mathbb{1} - \mathbb{1}$	\mathcal{X}
$\mathcal{X} - \mathbb{1} - \mathcal{S}$	$\mathcal{Y}_{-\pi/2} - \mathcal{X}_{-\pi/2}$
$\mathcal{X} - \mathbb{1} - \mathcal{S}^2$	$\mathcal{X}_{\pi/2} - \mathcal{Y}_{-\pi/2}$
$\mathcal{Y} - \mathbb{1} - \mathbb{1}$	\mathcal{Y}
$\mathcal{Y} - \mathbb{1} - \mathcal{S}$	$\mathcal{Y}_{-\pi/2} - \mathcal{X}_{\pi/2}$
$\mathcal{Y} - \mathbb{1} - \mathcal{S}^2$	$\mathcal{X}_{\pi/2} - \mathcal{Y}_{\pi/2}$
$\mathcal{Z} - \mathbb{1} - \mathbb{1}$	$\mathcal{X} - \mathcal{Y}$
$\mathcal{Z} - \mathbb{1} - \mathcal{S}$	$\mathcal{Y}_{\pi/2} - \mathcal{X}_{-\pi/2}$
$\mathcal{Z} - \mathbb{1} - \mathcal{S}^2$	$\mathcal{X}_{-\pi/2} - \mathcal{Y}_{\pi/2}$
$\mathbb{1} - \mathcal{H} - \mathbb{1}$	$\mathcal{Y}_{\pi/2} - \mathcal{X}$
$\mathbb{1} - \mathcal{H} - \mathcal{S}$	$\mathcal{X}_{-\pi/2}$
$\mathbb{1} - \mathcal{H} - \mathcal{S}^2$	$\mathcal{X}_{\pi/2} - \mathcal{Y}_{-\pi/2} - \mathcal{X}_{-\pi/2}$
$\mathcal{X} - \mathcal{H} - \mathbb{1}$	$\mathcal{Y}_{-\pi/2}$
$\mathcal{X} - \mathcal{H} - \mathcal{S}$	$\mathcal{X}_{\pi/2}$
$\mathcal{X} - \mathcal{H} - \mathcal{S}^2$	$\mathcal{X}_{\pi/2} - \mathcal{Y}_{\pi/2} - \mathcal{X}_{\pi/2}$
$\mathcal{Y} - \mathcal{H} - \mathbb{1}$	$\mathcal{Y}_{-\pi/2} - \mathcal{X}$
$\mathcal{Y} - \mathcal{H} - \mathcal{S}$	$\mathcal{X}_{\pi/2} - \mathcal{Y}$
$\mathcal{Y} - \mathcal{H} - \mathcal{S}^2$	$\mathcal{X}_{\pi/2} - \mathcal{Y}_{-\pi/2} - \mathcal{X}_{\pi/2}$
$\mathcal{Z} - \mathcal{H} - \mathbb{1}$	$\mathcal{Y}_{\pi/2}$
$\mathcal{Z} - \mathcal{H} - \mathcal{S}$	$\mathcal{X}_{-\pi/2} - \mathcal{Y}$
$\mathcal{Z} - \mathcal{H} - \mathcal{S}^2$	$\mathcal{X}_{\pi/2} - \mathcal{Y}_{\pi/2} - \mathcal{X}_{-\pi/2}$

is a two-design consisting of 12 elements, and is the minimum group that can fully depolarize any quantum operation.

In many experiments the fundamental operations are $\exp[-iX\theta/2]$ or $\exp[-iY\theta/2]$, which represent rotations around the X or Y axis by angle θ . In the PTM representation these are represented by

$$\mathcal{X}_\theta = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \cos(\theta) & -\sin(\theta) \\ & & \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad (3.4)$$

$$\mathcal{Y}_\theta = \begin{pmatrix} 1 & & & \\ & \cos(\theta) & & \sin(\theta) \\ & & 1 & \\ & -\sin(\theta) & & \cos(\theta) \end{pmatrix}. \quad (3.5)$$

Choosing $\theta = \pi$ gives the Pauli maps \mathcal{X} and \mathcal{Y} , respectively, and choosing $\theta = \pm\pi/2$ gives the standard $\mathcal{X}_{\pm\pi/2}$ and $\mathcal{Y}_{\pm\pi/2}$ Clifford generators. Table I lists the decompositions of all 24 Clifford elements in terms of both $\mathbf{H} - \mathbf{P} - \mathbf{S}$ and the simple rotations by π and $\pi/2$.

For the benchmarking simulations, unless explicitly noted, the parameter values used were $K = 10000$ and $M_{\text{max}} = 4096$, with $m \in \{1, 2, 4, \dots, M_{\text{max}}\}$. Exponential fits to the FDM were performed with the MATLAB `nlinfit` function for the model $A\alpha^m + \tilde{\epsilon}_0$, and 90% confidence intervals were found using the Jacobian option of the MATLAB `nlparci` function [32].

Given a set of RB experiments on a gate set with exact average error rate r and estimated average error rate $\{\hat{r}\}$, we can define the RB accuracy by

$$\mu = \log_{10}(\hat{r}/r), \quad (3.6)$$

and the confidence C by the size of 90% confidence intervals for the fits to the FDM. The definition of accuracy incorporates logarithms to symmetrically weight multiplicative, rather than additive, deviations of the estimate from the true value. The average error rate r is defined by

$$r = 1 - F_g = \frac{d^2 - \text{Tr}[\bar{\mathcal{E}}]}{d^2 + d}. \quad (3.7)$$

Here, F_g is the gate fidelity, defined as the average of input-state survival probabilities $P_{\text{survival}}(|\psi\rangle_{\text{in}}) = \text{Tr}[\rho_{\text{out}}|\psi\rangle\langle\psi|_{\text{in}}]$ over all pure input states. The gate fidelity of a gate Λ may then be expressed as

$$F_g = \int d\phi \langle\phi|\Lambda(\phi)|\phi\rangle. \quad (3.8)$$

Changing variables so that the integration is over $\text{SU}(2)$ (i.e., writing $|\phi\rangle = U_\phi|0\rangle$) and defining $\bar{\mathcal{E}}$ to be the PTM of the twirled map of the gate Λ , Eq. (3.7) is recovered.

IV. SRB: MARKOVIAN ERRORS

In this section we consider Markovian errors, which are errors where the error map at each time j is independent of previous gates and errors in the sequence. Markovian error can arise in a wide variety of scenarios, such as gate miscalibration (over- or under-rotation or off-resonant driving), amplitude damping, and control field fluctuations with correlation times much shorter than the individual gate time.

In order to examine the limits of SRB, we work with quantum error maps [completely positive and trace-preserving (CPTP) maps] that are “as far away as possible” from those on which SRB works best. Up to statistical errors resulting from measurement, SRB exactly estimates the error rate when each gate error is a depolarizing error. We consider the diamond norm distance [33–35] between a gate and a depolarizing channel of the same error rate to be a predictor of benchmarking performance. We use a unitary error model since unitary channels are far from depolarizing with respect to the diamond norm distance (see below).

Random unitary channels of size N were generated by choosing $N \times N$ matrices S and T from the Ginibre ensemble [36] in which elements are chosen independently from the normal distribution with mean zero and variance 1. The unitary channel $U = \exp[-iH\epsilon]$ is defined with the normalized Hermitian matrix $H = (G + G^\dagger) / \sqrt{\text{tr}[(G + G^\dagger)^2]}$ for $G = S + iT$ and where ϵ is a parameter (found numerically) that gives U the desired error rate r . Random CPTP maps acting on density matrices of dimension d were generated by creating a random unitary map of size d^3 (with non-normalized Hermitian H) and defining a set of Kraus operators

$$\begin{aligned} K_i(j,k) &= U[d^2(i-1) + j,k], \\ i &= 1, \dots, d^2, \\ j,k &= 1, \dots, d. \end{aligned} \quad (4.1)$$

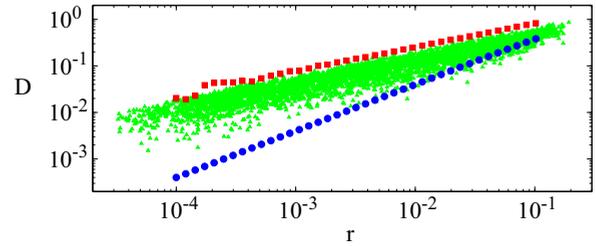


FIG. 1. (Color online) Distance of maps from a depolarizing channel. For a range of error rates, unitary channels (red squares) were farthest from and amplitude-damping channels (blue circles) closest to depolarizing channels of the same error rate. Random maps (green triangles) typically fell between these extremes. The distance was measured using the diamond norm distance from a depolarizing channel, and r is the average error rate.

A large number of random unitary and CPTP maps were generated, as were amplitude damping maps (see Sec. IV B) of the same error rates. For each map with error rate r , the diamond norm distance from the depolarizing channel of error rate r was calculated. As shown in Fig. 1, we found that unitary channels were farthest from the depolarizing channels, whereas amplitude-damping channels were particularly close. Therefore, as mentioned above, we consider unitary channels as an adequate worst-case test of SRB.

A. Finite-sampling effects

As shown in Ref. [7], the Hoeffding bound can be used to obtain an estimate of the required number K of sequences for a good estimate of the fidelity $F(m)$ at each sequence length m . If the trials at each sequence length correspond to independent and identically distributed random variables with range $[a,b]$ then

$$K = \frac{\ln(2/\delta)(b-a)^2}{2\epsilon^2}. \quad (4.2)$$

Here ϵ is the size of the confidence interval and $1 - \delta$ the confidence level. For a 90% confidence level ($\delta = 0.1$) and a confidence interval of $\epsilon = 10^{-4}$, we need as many as $K = 10^8$ trials for each data point. We will show that this value of K is actually much larger than necessary. The reason for this is that estimation of α (and thus r) from a process such as least-squares (LSQ) estimation [37] *simultaneously* uses the information from all data points, whereas the Hoeffding bound analyzes the number of trials for each data point independently. We first provide a simple numerical example from which we see that K can be chosen quite reasonably. Afterwards, we provide a general theoretical result whereby we obtain confidence intervals for α and r using linearization of the nonlinear regression model about the LSQ solution. An implication of this result is that K can be chosen to be significantly smaller than the Hoeffding estimate above of 10^8 .

For our numerical analysis, we considered various time-independent Markovian error models. Since essentially identical results were obtained for all models, we present the results for the case of gate- and time-independent unitary error. Figure 2(a) plots the size of the confidence interval C (at 90% confidence level) on the parameter α for three different error

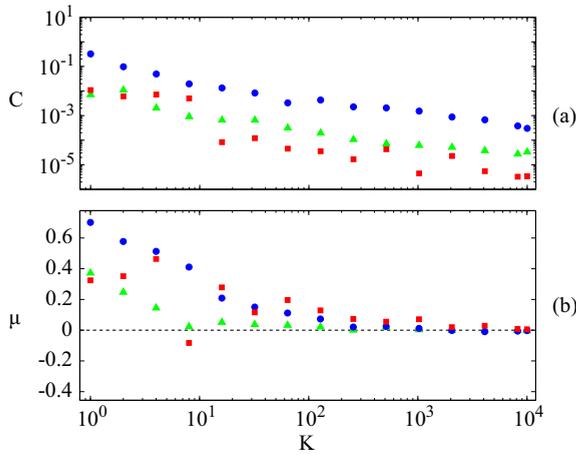


FIG. 2. (Color online) SRB with fixed unitary Markovian noise for error rates $r = 10^{-4}$ (red squares), $r = 10^{-3}$ (green triangles), and $r = 10^{-2}$ (blue circles). (a) Convergence of the confidence C (see text for details). The black line corresponds to the Hoeffding bound explained in the text. (b) Convergence of the accuracy μ . Note the rapid convergence of the estimate to within a factor of 2 of the average error rate by $K = 100$.

rates: $r = 10^{-4}$ (red squares), $r = 10^{-3}$ (green triangles), and $r = 10^{-2}$ (blue circles).

From these results, we see that for $K \sim 10$ – 100 , the size of the confidence interval is on the order of the underlying errors, suggesting that fewer than 100 sequences are sufficient to converge to the actual error rate r . This is further illustrated in Fig. 2(b) where we plot the accuracy μ as a function of K . This shows that \hat{r} converges (at some $K \sim 10$ – 100) to within a factor of 2 of r . Thus, smaller values of K than 10^8 suffice for estimating r .

We now turn to a more general theoretical analysis based on nonlinear regression methods [37]. The FDM Eq. (2.8) is a nonlinear function with parameters $\{\alpha, A, \tilde{e}_0\}$. In the case of linear regression, constructing confidence intervals is exact; however, for nonlinear regression, confidence intervals are typically constructed via approximative methods. One of the most widely utilized methods, and the approach we take here, is to obtain the least-squares solution, linearize the model around this solution, and construct confidence intervals for the linearization.

In our model there is one predictor variable m_i , three parameters we want to estimate $\theta = (\theta_1, \theta_2, \theta_3) = (\alpha, A, \tilde{e}_0)$, and one dependent variable $F(m_i, \theta)$. Let y_i represent the data we acquire so that if \mathbf{Y} represents the vector of y_i values,

$$\mathbf{Y} = \mathbf{F}(\tilde{\theta}) + \tilde{\xi}. \quad (4.3)$$

Here, $\tilde{\xi}$ is the realization of the random noise process ξ that produces the observed data, $\tilde{\theta}$ is the exact value for the parameters, and $F_i(\tilde{\theta}) = F(m_i, \tilde{\theta})$. We assume for simplicity that each ξ_i is normally distributed with variance σ^2/K , where σ is the single-shot standard deviation for estimating the fidelity at each sequence length (for simplicity, we assume σ is independent of the sequence length).

The LSQ estimator of θ is the vector $\hat{\theta}$ that satisfies

$$\hat{\theta} = \operatorname{argmin}[S(\theta)], \quad (4.4)$$

where

$$S(\theta) = [\mathbf{Y} - \mathbf{F}(\theta)]^T [\mathbf{Y} - \mathbf{F}(\theta)]. \quad (4.5)$$

Assuming that the model in Eq. (2.8) is an accurate description of the fidelity decay, a linearization of $F(m_i, \hat{\theta})$ around the LSQ solution $\hat{\theta}$ produces a linear model from which we can obtain confidence intervals.

In order to use a linearized model for computing confidence intervals, we need to compute the covariance matrix of the linearization at the estimator $\hat{\theta}$. This is done by using the Jacobian of the nonlinear model evaluated at $\hat{\theta}$,

$$\hat{V} = s^2 [J(\hat{\theta})^T J(\hat{\theta})]^{-1}. \quad (4.6)$$

Here

$$s^2 = \frac{S(\hat{\theta})}{N - D} \sim \frac{N\sigma^2}{(N - D)K} \quad (4.7)$$

is the average estimated residual variance, D is the number of parameters to be estimated, $N - D$ is the number of degrees of freedom in the model, and $J(\hat{\theta})$ is the Jacobian of $\mathbf{F}(\theta)$ at $\hat{\theta}$ which has entries

$$J_{i,j}(\hat{\theta}) = \left. \frac{\partial F(m_i; \theta)}{\partial \theta_j} \right|_{\hat{\theta}}. \quad (4.8)$$

Now that we have the linearization of the FDM about $\hat{\theta}$, we can use standard methods for constructing exact confidence intervals for linear models. Indeed, for each j , with probability $1 - \delta$, the true value θ_j of the parameter θ_j lies in the interval

$$\tilde{\theta}_j \in [\hat{\theta}_j - \hat{V}_{j,j}^{1/2} t_{N-D, 1-\delta/2}, \hat{\theta}_j + \hat{V}_{j,j}^{1/2} t_{N-D, 1-\delta/2}]. \quad (4.9)$$

Here $t_{N-D, 1-\delta/2}$ is the $(1 - \frac{\delta}{2})$ th quantile of the Student's t distribution with $N - D$ degrees of freedom. The Student's t distribution with k degrees of freedom is the symmetric distribution given by

$$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi} \Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}, \quad (4.10)$$

where $\Gamma(\cdot)$ is the standard Γ function. The $(1 - \frac{\delta}{2})$ th quantile of the Student's t distribution with k degrees of freedom, denoted $t_{k, 1-\delta/2}$, is just the value a that satisfies

$$\Pr[X < a] = 0.95, \quad (4.11)$$

where X is distributed according to the Student's t distribution with k degrees of freedom. Note that by symmetry $\Pr[X < a] = 0.95$ is equivalent to

$$\Pr[-a < X < a] = 0.9. \quad (4.12)$$

Now, $R := J(\hat{\theta})^T J(\hat{\theta})$ is a 3×3 matrix and it is straightforward to calculate

$$Q = R^{-1}, \quad (4.13)$$

where we note that Q depends on the fixed parameters N and $\{m_i\}_{i=1}^N$, and on the estimators \hat{A} , \hat{e}_0 , and $\hat{\alpha}$. Since we are mainly interested in estimating α , let us focus on

$$Q_{1,1} := ([J(\hat{\theta})^T J(\hat{\theta})]^{-1})_{1,1}. \quad (4.14)$$

We have

$$(\widehat{V}_{1,1})^{1/2} = s\sqrt{Q_{1,1}} = \frac{\sigma\sqrt{NQ_{1,1}}}{\sqrt{(N-D)K}}, \quad (4.15)$$

and so, since $D = 3$,

$$\tilde{\alpha} \in \left[\hat{\alpha} - \frac{t_{N-3,1-\delta/2}\sigma\sqrt{NQ_{1,1}}}{\sqrt{(N-3)K}}, \hat{\alpha} + \frac{t_{N-3,1-\delta/2}\sigma\sqrt{NQ_{1,1}}}{\sqrt{(N-3)K}} \right]. \quad (4.16)$$

That is, the confidence interval depends on the standard error σ/\sqrt{K} of the experiment.

Let us now choose a set of parameters that could represent a possible randomized benchmarking experiment. First, suppose we want a 90% confidence so that $\delta = 0.1$. As well, suppose $N = 7$, the set of m correspond to $\{m_i\}_{i=1}^7 = \{2^i\}_{i=1}^7$, and, to calculate $Q_{1,1}$, we take $\hat{\theta} = (0.993, \frac{1}{2}, \frac{1}{2})$. This gives $t_{4,0.05} \sim 2.132$ and $\sqrt{Q_{1,1}} \sim 0.0476$ so

$$\tilde{\alpha} \in \left[\hat{\alpha} - 0.134 \frac{\sigma}{\sqrt{K}}, \hat{\alpha} + 0.134 \frac{\sigma}{\sqrt{K}} \right]. \quad (4.17)$$

We can now see how different values for σ and K affect the confidence interval. Taking $\sigma = 0.004$ and $K = 50$ implies, with confidence 90%,

$$\tilde{\alpha} \in [\hat{\alpha} - 7.59 \times 10^{-5}, \hat{\alpha} + 7.59 \times 10^{-5}]. \quad (4.18)$$

Hence, we can see that small values of K (much smaller than those dictated by Hoeffding bounds for each data point) will still lead to robust estimates of the error rate \hat{r} .

B. Results and discussion

Here, we consider the performance of SRB with respect to various Markovian noise models. The models that we consider are as follows:

Gate-dependent random unitaries. A different random unitary error is applied to each Clifford gate.

Fixed random unitary. The same random unitary error is applied to all Clifford gates.

Generator-dependent unitaries. Each Clifford gate was decomposed into a minimum number of generators $X_{\pm\pi/2}$, $Y_{\pm\pi/2}$, X , and Y (Table I) which were each assigned a random unitary error of strength $r/1.875$. Error maps were determined from the decompositions. Note that the decomposition is not unique but the results do not depend on this choice.

Amplitude damping. The generator gates $X_{\pm\pi/2}$ and $Y_{\pm\pi/2}$ are typically implemented via Rabi rotation about the X or Y axis at rate Ω for time $t_g = \pi/2\Omega$. Amplitude-damping noise of error rate r is characterized by the rate $\gamma = 1/T_1 = 4\Omega \ln[(1 + \sqrt{4 - 6r})/(3 - 6r)]/\pi$ and the noisy generator maps are given by

$$\mathcal{X}_{\pm\pi/2} = \prod_{l=1}^M \begin{pmatrix} 1 & & & \\ & \eta & & \\ & & \eta \cos(\pi/2M) & \mp \eta \sin(\pi/2M) \\ 1 - \eta^2 & & \pm \eta^2 \sin(\pi/2M) & \eta^2 \cos(\pi/2M) \end{pmatrix}, \quad (4.19)$$

$$\mathcal{Y}_{\pm\pi/2} = \prod_{l=1}^M \begin{pmatrix} 1 & & & \\ & \eta \cos(\pi/2M) & & \pm \eta \sin(\pi/2M) \\ & & \eta & \\ 1 - \eta^2 & \mp \eta^2 \sin(\pi/2M) & & \eta^2 \cos(\pi/2M) \end{pmatrix}, \quad (4.20)$$

where $\eta = [1/(\sqrt{4 - 6r} - 1)]^{-1/M}$ (and similarly for \mathcal{X} and \mathcal{Y}). These expressions are precise in the limit $M \rightarrow \infty$, but we take M to be a finite numerical parameter chosen large enough so that further refinement does not significantly improve the approximation and small enough that significant numerical errors do not accumulate in MATLAB. We took $M = 2\,000\,000$. With these approximate generator maps, we used the decomposition given in Table I to approximate the entire noisy Clifford group.

Gaussian noise (fast). The noise on all gates at time j was $V_j = U^{\epsilon_j}$ for some fixed unitary channel U . The time-dependent parameter ϵ_j was chosen such that V_j has error rate r_j that is normally distributed with mean r and standard deviation $r/4$.

Slow drift. This is identical to the fixed unitary case except that the fixed unitary depends on k so that $U_k = U^\epsilon$ for numerically determined ϵ such that r_k increases linearly from $r/2$ at $k = 1$ to $3r/2$ at $k = K$.

For all noise models tested, SRB estimated the error rate to within a factor of 2 of the actual average error rate (Fig. 3, Table II). SRB performed best with amplitude damping noise, supporting our hypothesis that SRB would work well for errors near the depolarizing channel. Generator-dependent noise was estimated most poorly. A possible cause of this, relative to the other Markovian cases, is variation of the error rate over different Clifford gates. This variation may lead to deviations from the exponential FDM, a supposition supported by the larger confidence intervals for this model compared to the others.

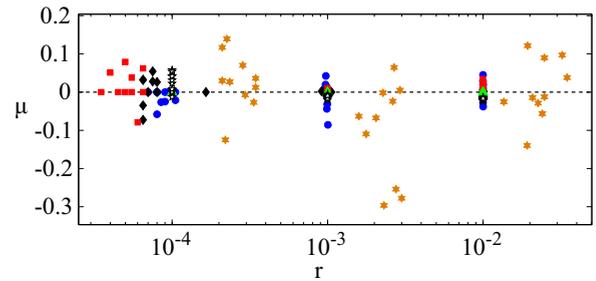


FIG. 3. (Color online) The accuracy of several Markovian models with $K = 10\,000$ sequences. Different random unitaries (blue circles), fixed random unitaries (black diamonds), generator-dependent unitaries (orange six-pointed stars), amplitude damping (green triangles), Gaussian unitaries (red squares), and slow drift (white five-pointed stars). All errors except generator-dependent unitary errors were estimated to within 25% of r , with amplitude-damping noise determined significantly better. Generator-dependent unitary noise was estimated to within 50% of r . The large horizontal spread at low error rates is due simply to the precision of the procedure used for generating random unitary channels of fixed error rate.

TABLE II. For each noise model and each average error rate r , a set of $n = 10$ experiments (with the expectation of $1/f$ noise, for which $n = 1$), each using $K = 10000$ benchmarking sequences was performed. Accuracy $\bar{\mu}$, standard error $s = (\sqrt{\mu^2 - \bar{\mu}^2})/\sqrt{n}$, and average confidence \bar{C} are as defined in the text. The accuracies of the SRB estimates for Markovian errors are in most cases better by roughly an order of magnitude than the accuracies for non-Markovian errors. The precision and fit confidence are not significantly different between the two types of noise, except in the cases of amplitude-damping noise, in the presence of which SRB performs especially well. Square brackets indicate multiplication by a power of 10, i.e., $A[x] = A \times 10^x$.

Error rate	0.0001			0.001			0.01		
	$\bar{\mu}$	s	\bar{C}	$\bar{\mu}$	s	\bar{C}	$\bar{\mu}$	s	\bar{C}
Random unitary	-1.3[-2]	5.9[-3]	4.3[-6]	-7.3[-3]	1.1[-2]	2.2[-5]	2.2[-3]	7.6[-3]	3.9[-4]
Fixed unitary	6.6[-3]	1.1[-2]	4.5[-6]	2.2[-3]	7.5[-4]	2.3[-5]	-1.2[-3]	1.2[-3]	3.5[-4]
Generator dependent	2.7[-2]	2.2[-2]	4.6[-6]	-1.0[-1]	3.9[-2]	5.3[-5]	7.0[-3]	2.4[-2]	1.2[-3]
Gaussian	1.5[-2]	1.3[-2]	1.2[-5]	1.3[-3]	1.9[-3]	2.8[-5]	1.6[-2]	3.0[-3]	5.4[-4]
Slow drift	2.5[-2]	7.6[-3]	4.5[-6]	-1.2[-2]	1.4[-3]	4.0[-5]	-1.9[-2]	1.0[-3]	4.7[-4]
Amplitude damping	1.7[-4]	9.8[-5]	1.2[-7]	6.9[-6]	3.5[-5]	8.9[-7]	-3.8[-5]	1.2[-4]	5.2[-5]
Leakage (random)	1.2[-3]	7.4[-3]	3.6[-6]	-6.6[-3]	5.9[-3]	1.4[-5]	6.0[-4]	4.5[-3]	1.7[-4]
Leakage (fixed)	4.2[-2]	2.7[-2]	2.8[-6]	2.7[-2]	1.3[-2]	2.0[-5]	-2.0[-4]	1.6[-2]	6.0[-4]
1/f	7.1[-2]	N/A	7.9[-6]	-1.3[-2]	N/A	4.8[-5]	-8.4[-2]	N/A	8.5[-4]

V. SRB: 1/f NOISE

We model a one-qubit system subject to semiclassical phase noise by the Hamiltonian

$$H(t)/\hbar = \frac{\Omega_X(t)}{2}X + \frac{\Omega_Y(t)}{2}Y + \xi(t)Z, \quad (5.1)$$

where $\Omega_{X,Y}(t)$ are real control fields and $\xi(t)$ is a realization of a real random noise process. The noise process $\xi(t)$ can be characterized by its power spectral density (PSD)

$$S(f) = \int_{-\infty}^{\infty} C(t)e^{-i2\pi ft} dt, \quad (5.2)$$

where

$$C(t) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \xi(s)\xi(s+t) ds \quad (5.3)$$

is the autocorrelation function. The noise is said to be $1/f$ if its PSD is given by $S(f) = A/f$ for some constant A .

A simple discrete model of $1/f$ noise is obtained by summing a large number of random telegraph noise (RTN) realizations with different switching rates [38]. A two-state telegraph noise signal $s_k(t)$ switches between $\{+1, -1\}$ with constant rate f_k , and interarrival times τ of switching events are exponentially distributed with probability distribution $p(\tau) = f_k e^{-f_k \tau}$. If the density of switching rates is proportional to $1/f$ in the interval $[f_{\min}, f_{\max}]$, then $\xi(t) = A' \sum_k s_k(t)$ has PSD

$$S(f) \propto \frac{1}{\pi f} \left[\arctan\left(\frac{f_{\max}}{\pi f}\right) - \arctan\left(\frac{f_{\min}}{\pi f}\right) \right], \quad (5.4)$$

which is proportional to $1/f$ if $f_{\min} \ll \pi f \ll f_{\max}$ [39]. The noise power is proportional to the square of A' but also depends on the cutoff frequencies and number of RTN signals participating in the sum.

A. Simulated Ramsey experiments

This $1/f$ noise model produces Gaussian decay of coherences [40–42]. By simulating Ramsey experiments, we verify that the model reproduces this type of decay for several values

of noise power and relate the extracted values of T_2^* to average gate fidelities.

Each realization of $1/f$ noise is constructed from 50 RTN signals whose initial state is uniformly random. The low- and high-frequency cutoffs are $f_{\min} = (10N\Delta t)^{-1}$ and $f_{\max} = (2\Delta t)^{-1}$, respectively, where Δt is the smallest time step appearing in the simulation and N is the total number of time steps in any noise realization.

In a simulated Ramsey experiment, each $\xi(t)$ yields a pure-state trajectory $|\psi(t)\rangle = e^{-i2\pi Z \int_0^t \xi(s) ds} |\psi(0)\rangle$ where $|\psi(0)\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Taking the ensemble average over 2000 noise realizations, we obtain the mixed quantum state $\rho(t)$ whose coherence $\sigma(t) = 2|\rho_{12}(t)|$ exhibits Gaussian decay $e^{-(t/T_2^*)^2}$, as shown in Fig. 4(a) (see also Fig. 5). Taking $d = 2$ (one qubit) and the gate Λ to be the phase-damping gate, the gate fidelity defined in Eq. (3.8) is given by $F_g = [2 + \sigma(t_g)]/3$ where t_g is the gate time, here taken to be $20\Delta t$. Provided that $T_2^* > t_g$, $1/f$ noise leads to higher gate fidelities than stochastic dephasing with the same value of T_2^* [see Fig. 4(b)].

B. Results and discussion

To perform a single RB experiment subject to $1/f$ noise, we choose a sequence of random Clifford gates $C_1, C_2, \dots, C_{M_{\max}}$. For each subsequence C_1, \dots, C_m an inverting Clifford gate $\Upsilon_m = (C_m \cdots C_1)^\dagger$ is determined. The subsequences and inverting gates are concatenated to form the total sequence $C_1, \Upsilon_1, C_1, C_2, \Upsilon_2, \dots, C_1, C_2, \dots, C_{M_{\max}}, \Upsilon_{M_{\max}}$ for the RB experiment. Next each Clifford gate is expressed in terms of generators $G_{\pm\pi/2}$ where $G \in \{X, Y\}$, and each generator $G_{\pm\pi/2}$ is realized by a normalized Gaussian pulse with amplitude $\pm\pi/2$ in the corresponding time interval of the control field $\Omega_G(t)$. The duration of the $1/f$ noise, $\xi(t)$, is the same as the duration of the total sequence, i.e., the noise has the correct correlations over the entire duration of the RB experiment.

For each RB experiment, we generate subsequences of lengths $m = 1, 2, \dots, 2^n, \dots, M_{\max}$ up to $M_{\max} = 4096$. A time step Δt was chosen such that each normalized Gaussian pulse was sampled at 20 equally spaced points. For calculating

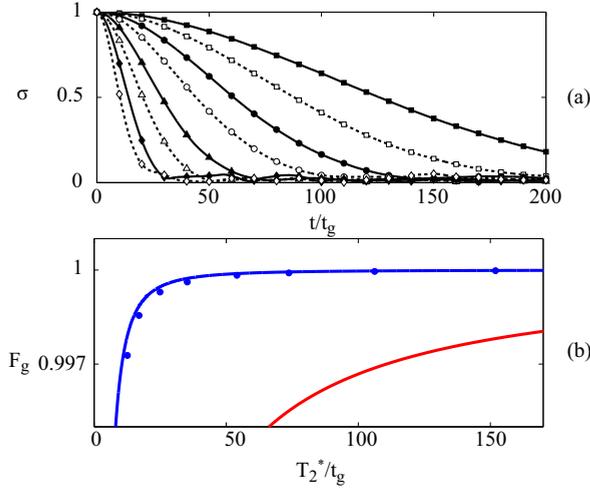


FIG. 4. (Color online) Simulated Ramsey experiments for $1/f$ noise. (a) The coherence $\sigma(t) = 2|\rho_{12}(t)|$ is plotted versus time t/t_g where t_g is the duration of a gate. Each curve corresponds to a different noise power. The dashed lines interpolate the numerical data and every tenth data point is shown. (b) For each decay in (a) we extract t_g as the time at which $\sigma = 1/e$ and plot the corresponding average gate fidelity $F_g = [2 + \sigma(t_g)]/3$ (blue points). The solid red curve is the gate fidelity for stochastic noise with exponential decay $\sigma(t) = e^{-t/T_2^*}$ of the coherences. The solid blue curve is the (analytic) gate fidelity for $1/f$ noise: $F_g = (2 + e^{-(t_g/T_2^*)^2})/3$.

cutoff frequencies, N was taken to be the total number of time steps in the experiment. The amplitude A' of the $1/f$ noise was adjusted to achieve target average gate error rates r of approximately 10^{-4} , 10^{-3} , and 10^{-2} corresponding to T_2^*/t_g values of 94, 30, and 8, respectively. Finally the time evolution was calculated using the time-ordered composition of discrete unitary gates $U(t + \Delta t, t) = \exp[-i2\pi H(t)\Delta t/\hbar]$.

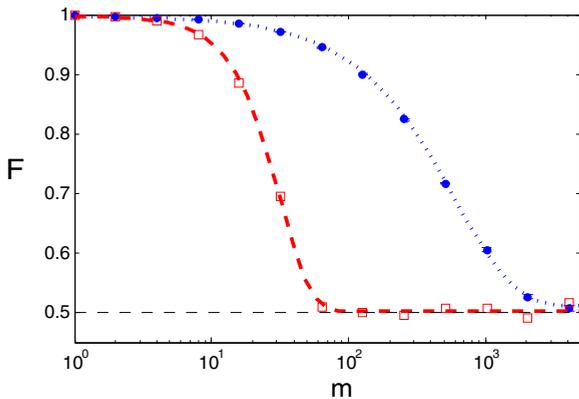


FIG. 5. (Color online) Fidelity decay in the presence of $1/f$ noise ($A' = 2 \times 10^{-8} \rightarrow r \sim 10^{-3}$) with application of random Clifford gates (blue circles, mean of four experiments, $K = 2500$) and identity gates (red squares, $K = 2500$). Also shown are fits to the FDM $F = 0.486(0.998)^m + 0.510$ (blue dotted line) and to the Gaussian model $F = 0.495(0.999)^{m^2} + 0.503$ (red dashed line). The standard error in the data is less than 4.8×10^{-3} and fit parameters are rounded to three digits.

A qubit initialized in the X - Y plane and subjected to $1/f$ phase noise suffers a rapid Gaussian decay of state fidelity, but this decay is dramatically slowed by performing random Clifford gates (Fig. 5). Insofar as a computation may be modeled as a random sequence of Clifford gates, the relevant quantity for discussing computational errors may be the gate fidelity F_g or, equivalently, the error rate r , rather than the dephasing time T_2^* . Supporting this notion is the fact that, for a fixed value of T_2^* , $1/f$ noise has a much higher gate fidelity than does noise that exhibits an exponential Ramsey decay (Fig. 4).

This behavior is potentially due to the depolarizing effect of twirling the $1/f$ noise with Clifford gates. Consider a model with instantaneous Clifford gates followed by noise. We fix a noise realization $\xi(t)$ and average over SRB sequences. The noise gives rise to a sequence of correlated random variables whose samples are $\phi_j = \int_{jt_g}^{(j+1)t_g} \xi(s)ds$, where j labels the Clifford gate in an SRB experiment. Since each ϕ_j is a real number, the noise operator realizations $e^{-i2\pi\phi_j Z}$ are single-axis unitary rotations. These are each twirled independently and by direct calculation become depolarizing channels \mathcal{E}_{ϕ_j} with depolarizing parameter [see Eq. (2.7)]

$$\alpha_j = [1 + 2 \cos(4\pi\phi_j)]/3. \quad (5.5)$$

Defining $\phi_m = (\phi_{m-1}, \phi_{m-2}, \dots, \phi_1)$, a sequence of m Clifford gates therefore produces the channel

$$\int_{\phi_m} p(\phi_m) \mathcal{E}_{\phi_{m-1}} \mathcal{E}_{\phi_{m-2}} \cdots \mathcal{E}_{\phi_1} \mathcal{E}_{\phi_0} d\phi_m, \quad (5.6)$$

where the joint distribution does not factor into $p(\phi_{m-1})p(\phi_{m-2}) \cdots p(\phi_1)$ due to the low-frequency components of the noise.

This behavior foreshadows the result that, for all cases tested, RB provides an estimate within a factor of 2 of the actual average error rate for $K \sim 100$ or greater (Fig. 6). Note, however, that the confidence interval of the RB estimate in the presence of $1/f$ noise seems to saturate as K increases, and accuracy ceases to improve (Fig. 7). This indicates that the exponential model of fidelity decay is not completely accurate,

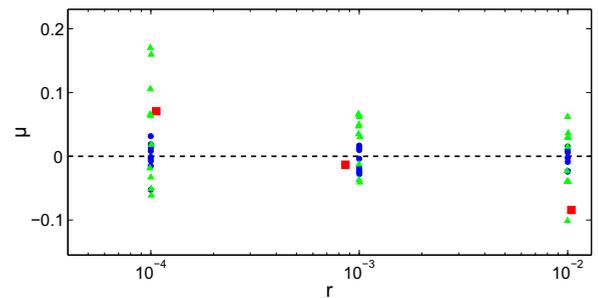


FIG. 6. (Color online) SRB results for several non-Markovian error models and error rates with $K = 10000$ averaged sequences. Leakage, both random and fixed, was nearly always estimated to within a factor of 2, and there was no significant difference between the two types, except for the largest error rate, where random leakage had better accuracy, precision, and fit confidence. $1/f$ noise was estimated in all cases to within 25% of r . Random leakage (blue circles), fixed leakage (green triangles), and $1/f$ (red squares).

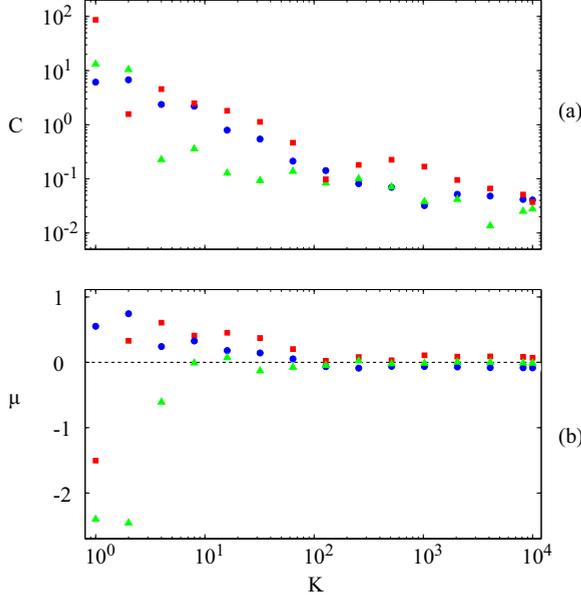


FIG. 7. (Color online) Convergence in K for standard RB with $1/f$ noise of (a) accuracy μ and (b) confidence interval C on a fit of the fidelity decay to the exponential model. Noise strengths are $A' = 2.5 \times 10^{-9}$ ($T_2^*/t_g = 93.95$) (red squares), $A' = 2.0 \times 10^{-8}$ ($T_2^*/t_g = 30.05$) (green triangles), and $A' = 2.5 \times 10^{-7}$ ($T_2^*/t_g = 8.35$) (blue circles).

in contrast to the Markovian case. As a consequence, there is some K_{\max} such that further increases in sample size will not yield a more accurate estimate.

VI. SRB: LEAKAGE

Leakage errors are interesting because they can build coherences in levels above or outside the qubit subspace, leading to highly non-Markovian dynamics. In order to simulate leakage, we extend the Hilbert space of our simulation from a qubit to a qutrit, i.e., a three-level system. The noisy Clifford gates are modeled by the ideal gate acting on the qubit subspace followed by a unitary error gate acting on the full qutrit Hilbert space. Each Clifford gate is assigned a different random qutrit unitary error U_i :

$$C_i^{\text{sim}} = U_i \circ (C_i^{\text{ideal}} \oplus \mathbb{I}). \quad (6.1)$$

Figure 8 illustrates leakage-induced error processes in the qubit subspace, as well as the effect of RB on these processes. When a single leakage error (noisy identity gate) is repeatedly applied, the survival probability of nonstationary states oscillates. In contrast, when RB is performed, the random Clifford sequences average this oscillatory behavior, turning it into a uniform decay.

To understand analytically how leakage errors enter into a benchmarking experiment we start by extending the Pauli operators to act on the three-dimensional Hilbert space by defining the following orthogonal basis of nine operators

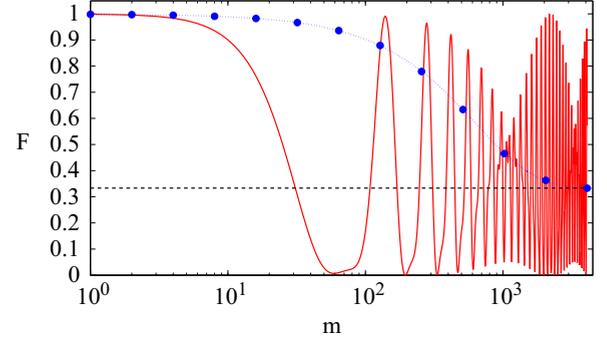


FIG. 8. (Color online) Fidelity decay in the presence of random leakage noise ($r = 10^{-3}$) with repeated application of random noisy Cliffords (blue circles, $K = 10000$) and repeated application of a noisy identity gate, i.e., a fixed unitary qutrit gate (red line). Blue dotted line is $F = .666(.998)^m + .333$, the FDM fit to the RB data. Note the decay to $1/3$ (black dashed line). The standard errors in the data increase monotonically with m from 10^{-6} to approximately 10^{-3} , and fit parameters are rounded to three digits. The solid red line is the survival probability of the excited state $P(m) = |(1|U^m|1)|^2$ for repeated application of a fixed (randomly chosen) three-level unitary error U . If the excited state is nonstationary, $P(m)$ is an oscillatory function that is not periodic unless the stationary states accumulate commensurate phases.

$P_1 = I$:

$$P_2 = \sqrt{3/2} \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix}, \quad P_3 = \sqrt{3/2} \begin{pmatrix} Y & 0 \\ 0 & 0 \end{pmatrix},$$

$$P_4 = \sqrt{3/2} \begin{pmatrix} Z & 0 \\ 0 & 0 \end{pmatrix}, \quad P_5 = \sqrt{1/2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad (6.2)$$

$$P_6 = \sqrt{3/2} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad P_7 = \sqrt{3/2} \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix},$$

$$P_8 = \sqrt{3/2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad P_9 = \sqrt{3/2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \quad (6.3)$$

The first four represent population and correlations in the qubit subspace, the fifth represents population inversion in the third level, and the last four represent correlations between the qubit subspace and the third level, and drive leakage. Let us normalize each element so that the set forms an orthonormal basis. Partitioning the basis into spaces $\{P_1, P_2, P_3, P_4, P_5\}$ and $\{P_6, P_7, P_8, P_9\}$, we find that Eq. (2.5) becomes

$$\bar{\mathcal{E}}_{\mathbb{G}} = \frac{1}{|\mathbb{G}|} \sum_{U \in \mathbb{G}} \begin{pmatrix} U^T \oplus 1 & 0 \\ 0 & U_L^T \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} U \oplus 1 & 0 \\ 0 & U_L \end{pmatrix}$$

$$= \frac{1}{|\mathbb{G}|} \sum_{U \in \mathbb{G}} \begin{pmatrix} (U^T \oplus 1)A(U \oplus 1) & (U \oplus 1)BU_L \\ U_L^T C(U \oplus 1) & U_L^T D U_L \end{pmatrix}, \quad (6.4)$$

where U is the 4×4 PTM representation of the map of the Clifford operator in the qubit subspace and the 4×4 matrix U_L is the effect of the unitary in the leakage space. U_L does

not have a simple form and depends on the relative phase between the qubit subspace map and the higher level. It maps elements from the leakage subspace only to other elements in the leakage subspace.

It is simple to show that the top left element in the block matrix equation Eq. (6.4) becomes the 5×5 matrix

$$\frac{1}{|\mathcal{G}|} \sum_{U \in \mathcal{G}} (U^T \oplus 1) \mathcal{A} (U \oplus 1) = \begin{pmatrix} 1 & & & & \\ & \alpha & & & \\ & & \alpha & & \\ & & & \alpha & \\ \mathcal{A}_{51} & & & & \mathcal{A}_{55} \end{pmatrix}. \quad (6.5)$$

This has a similar form to Eq. (2.7), where the parameter α represented a depolarizing parameter. The other three blocks do not take such a simple form. However, by allowing the phase in the higher level to be both $U \oplus 1$ and $U \oplus (-1)$ for all U in the single-qubit Clifford group (extending the size of the single-qubit Clifford group to 48 elements) then the set of unitaries becomes $\{(U \oplus 1)(\pm \mathcal{U}_L)\}$. With this addition, the off-diagonal matrices of Eq. (6.4) become zero, that is, the effective map is block diagonal. This addition could be simple to implement experimentally, as ideal implementations of $\pm\pi/2$ and $\pm 3\pi/2$ rotations have opposite phases in the leakage levels. Thus, by making the Clifford group from these generators, it is possible to construct all $\{(U \oplus 1)(\pm \mathcal{U}_L)\}$ operators. Since the map is now block diagonal, the FDM is straightforward to calculate. We work in the operator basis $\{P_1, P_2, \dots, P_9\}$. The ground state is given by

$$|0\rangle\langle 0| = \frac{1}{3}P_1 + \sqrt{\frac{3}{18}}P_4 + \sqrt{\frac{1}{18}}P_5. \quad (6.6)$$

Acting with $\bar{\mathcal{E}}_G^m$ on the corresponding vector of coefficients and calculating an expectation produces the FDM

$$\begin{aligned} F &= \text{Tr}(|0\rangle\langle 0| \bar{\mathcal{E}}_G^m |0\rangle\langle 0|) \\ &= \frac{\alpha^m}{2} + \frac{1}{3} + \frac{(\mathcal{A}_{55})^m}{6} + \frac{\mathcal{A}_{51}}{3\sqrt{2}} \sum_{j=0}^{m-1} (\mathcal{A}_{55})^j \\ &= C_1 \alpha^m + C_2 \mathcal{A}_{55}^m + C_3, \end{aligned} \quad (6.7)$$

where $C_1 = 1/2$, $C_2 = 1/6 - D$, $C_3 = 1/3 + D$, and $D = \mathcal{A}_{51}/[3\sqrt{2}(1 - \mathcal{A}_{55})]$. This is a simple sum of exponentials. Initial-state preparation and measurement errors that act only in the qubit subspace change only the constants (not the functional form). Furthermore, if the leakage error is from a unital operation (which includes unitary operations), then $\mathcal{A}_{51} = 0$. From this model we see that leakage causes the fidelity to asymptotically decay to C_3 , which is equal to $1/3$ for unital noise. When there is no leakage ($\mathcal{A}_{55} = 1$ and $\mathcal{A}_{51} = 0$), the decay goes to $1/2$ as expected from standard RB. This implies that the asymptotic fidelity value can be used an indicator of the type of noise present in the system.

Here we numerically investigate this two-phase model of RB with two types of leakage noise: fixed random and different random three-dimensional unitary channels. A typical benchmarking experiment is shown in Fig. 8 as the blue dotted line for the different random unitary errors. It is clear that the

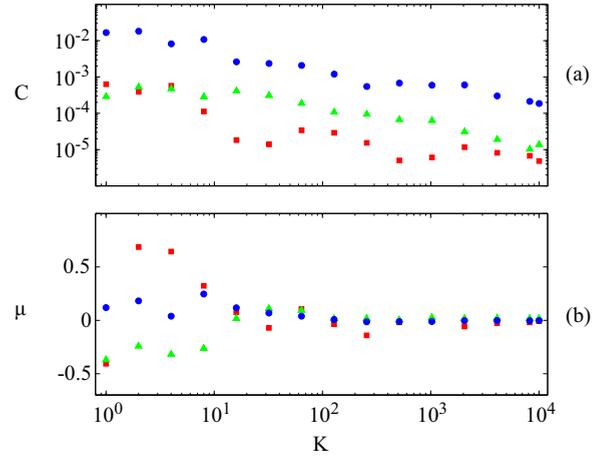


FIG. 9. (Color online) The 90% confidence intervals on the random leakage models are plotted against sample size K for multiple values of $r = 10^{-4}$ (red squares), $r = 10^{-3}$ (green triangles), and $r = 10^{-2}$ (blue circles).

fidelity does not decay to $1/2$ but rather to $1/3$. For this case the confidence interval and μ parameter are shown in Fig. 9. Here again by about 100 samples the estimated error and the actual error agree. Furthermore, we find that both error models are well captured by the benchmarking experiments and predict the correct underlying error (see Fig. 6).

VII. INTERLEAVED RB: MARKOVIAN ERRORS

The goal of IRB is to obtain bounds on the gate fidelity of an individual gate with noise \mathcal{E}_{int} . In the limit where the average noise $\bar{\mathcal{E}}$ is depolarizing, the bounds collapse, and an exact estimate of α_{int} is possible. More precisely, for channels \mathcal{E}_{int} and $\bar{\mathcal{E}}$ with depolarizing parameters α_{int} and $\bar{\alpha}$, if $\bar{\mathcal{E}}$ (or \mathcal{E}_{int}) is depolarizing, then the depolarizing parameter of the composed channel $\mathcal{E}_{\text{int}}\bar{\mathcal{E}}$ is given by the product $\alpha_{\text{int}}\bar{\alpha}$. Typically, one expects that averaging over many sequences implies that $\bar{\mathcal{E}}$ converges to a depolarizing channel and so the depolarizing parameter of $\mathcal{E}_{\text{int}}\bar{\mathcal{E}}$ converges to $\alpha_{\text{int}}\bar{\alpha}$.

We examine the extent to which this is the case by setting $\bar{\mathcal{E}}$ to be an average of L unitary channels. We plot how the diamond norm between $\bar{\mathcal{E}}$ and the depolarizing channel with the same average fidelity scales with L in Fig. 10. We see

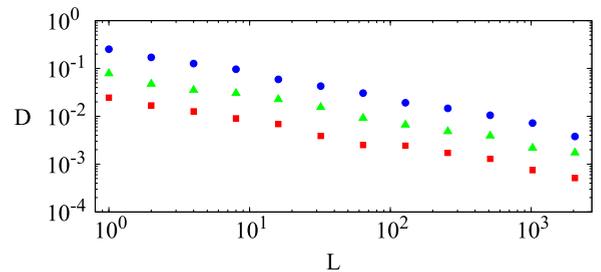


FIG. 10. (Color online) The diamond norm distance D between $\bar{\mathcal{E}}$ and the depolarizing channel of error rate r decreases as a function of L . $r = 10^{-4}$ (red squares), $r = 10^{-3}$ (green triangles), and $r = 10^{-2}$ (blue circles).

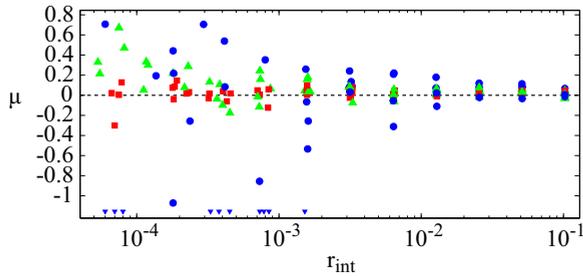


FIG. 11. (Color online) IRB in the presence of random unitary noise for $K = 10000$. \hat{r}_{int} is the estimated interleaved error rate, r_{int} is the true rate, and $\mu = \log_{10}(\hat{r}_{\text{int}}/r_{\text{int}})$. The average Clifford error is $r = 10^{-4}$ (red squares), $r = 10^{-3}$ (green triangles), $r = 10^{-2}$ (blue circles; blue triangles indicate negative IRB estimates).

that the approximation becomes increasingly accurate with increasing L .

IRB was tested using a noise model in which each Clifford gate received a random, gate-dependent unitary error of error rate r , and the interleaved gate received a unitary error gate of error rate r_{int} . The method was tested for three values of r and a wide range of values for r_{int} (Fig. 11). While the IRB estimates were within a factor of 2 of the true interleaved error rate for $r_{\text{int}} \geq r$, they were less accurate as r_{int} became much less than r (Fig. 11). We can see that when r_{int}/r is about 0.1, the estimate begins to diverge from the true value. Thus, IRB can be a reliable tool in most regimes of interest, unless the interleaved gate is significantly better than a typical gate.

VIII. CONCLUSION

In this paper we reviewed randomized benchmarking protocols and numerically investigated their application on a single qubit under various physically realistic and relevant error models. These models included systematic rotations, amplitude damping, leakage to higher levels, and $1/f$ noise. While each randomized benchmarking protocol has a domain

of validity for which it provably gives robust error estimates, we found that, in most cases analyzed, benchmarking provides better than a factor-of-2 estimate of the average error rate. This suggests that RB protocols can be utilized in quite general situations and thus are a valuable tool for verification and validation of quantum operations.

We showed using both numerical and general theoretical results that the number of different random sequences in a benchmarking experiment can be much less than the Hoeffding bound estimates [7]. Our theoretical method consisted of finding the nonlinear least-squares solution, linearizing the nonlinear model around this solution, and constructing exact confidence intervals for the linearized multivariate model. We see that the size of the confidence intervals scales linearly with the standard error.

In the case of $1/f$ noise, we find that randomized benchmarking protocols produce a fidelity decay that can be modeled by a composition of correlated depolarizing channels. The degree of correlation can affect the extent to which a simple exponential decay is valid. For leakage errors, we devised a protocol that allows for the estimation of gate errors under a decay model of a sum of exponentials. The asymptotic behavior of fidelity decays can be used as a measure of the extent to which leakage errors are present in an experiment. Finally we showed that, in practice, the interleaved randomized benchmarking protocol provides bounds that are tighter than those theoretically predicted [8]. Provided the error on the interleaved gate is not much smaller (by a factor of 10) than the average error, the estimated error rate is a reliable quantity.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with Seth T. Merkel, Jerry Chow, and Oliver Dial. We acknowledge support from IARPA under Contract No. W911NF-10-1-0324. All statements of fact, opinion, or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the U.S. Government.

-
- [1] A. W. Cross, D. P. Divincenzo, and B. M. Terhal, *Quantum Inf. Comput.* **9**, 541 (2009).
 - [2] I. L. Chuang and M. A. Nielsen, *J. Mod. Opt.* **44**, 2455 (1997).
 - [3] J. F. Poyatos, J. I. Cirac, and P. Zoller, *Phys. Rev. Lett.* **78**, 390 (1997).
 - [4] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, *Phys. Rev. A* **87**, 062119 (2013).
 - [5] 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).
 - [6] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. Lett.* **106**, 180504 (2011).
 - [7] E. Magesan, J. M. Gambetta, and J. Emerson, *Phys. Rev. A* **85**, 042311 (2012).
 - [8] E. Magesan *et al.*, *Phys. Rev. Lett.* **109**, 080505 (2012).
 - [9] J. P. Gaebler *et al.*, *Phys. Rev. Lett.* **108**, 260503 (2012).
 - [10] J. M. Gambetta *et al.*, *Phys. Rev. Lett.* **109**, 240504 (2012).
 - [11] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. T. Plourde, and M. Steffen, *Phys. Rev. A* **87**, 030301(R) (2013).
 - [12] M. J. Biercuk *et al.*, *Quantum Inf. Comput.* **9**, 0920 (2009).
 - [13] 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(R) (2011).
 - [14] S. Olmschenk *et al.*, *New J. Phys.* **12**, 113007 (2010).
 - [15] C. A. 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(R) (2010).
 - [18] J. B. Johnson, *Phys. Rev.* **26**, 71 (1925).

- [19] W. Schottky, *Phys. Rev.* **28**, 74 (1926).
- [20] D. J. Levitin, P. Chordia, and V. Menon, *Proc. Natl. Acad. Sci. USA* **109**, 3716 (2012).
- [21] F. C. Wellstood, C. Urbina, and J. Clarke, *Appl. Phys. Lett.* **50**, 772 (1987).
- [22] F. Yoshihara, K. Harrabi, A. O. Niskanen, Y. Nakamura, and J. S. Tsai, *Phys. Rev. Lett.* **97**, 167001 (2006).
- [23] D. H. Slichter, R. Vijay, S. J. Weber, S. Boutin, M. Boissonneault, J. M. Gambetta, A. Blais, and I. Siddiqi, *Phys. Rev. Lett.* **109**, 153601 (2012).
- [24] J. M. Taylor, J. R. Petta, A. C. Johnson, A. Yacoby, C. M. Marcus, and M. D. Lukin, *Phys. Rev. B* **76**, 035315 (2007).
- [25] J. J. Pla *et al.*, *Nature (London)* **496**, 334 (2013).
- [26] 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).
- [27] J. M. Martinis, *Quantum Inf. Process.* **8**, 81 (2009).
- [28] J. Medford *et al.*, *Nat. Nanotechnol.* **8**, 654 (2013).
- [29] J. Emerson, R. Alicki, and K. yczkowski, *J. Opt. B: Quantum Semiclass. Opt.* **7**, S347 (2005).
- [30] J. M. Chow *et al.*, *Phys. Rev. Lett.* **109**, 060501 (2012).
- [31] S. Kimmel *et al.*, *Phys. Rev. X* **4**, 011050 (2014).
- [32] Computer code MATLAB, version 8.0 (R2012b) (The MathWorks Inc., Natick, MA, 2012).
- [33] D. Aharonov, A. Kitaev, and N. Nisan, *Proceedings of the 30th Annual ACM Symposium on Theory of Computing Dallas, TX, 1998* (ACM, New York, 1998).
- [34] A. Kitaev, A. H. Shen, and M. N. Vyalii, *Classical and Quantum Computation* (American Mathematical Society, Boston, 2002).
- [35] J. Watrous, *Quantum Inf. Comput.* **5**, 058 (2005).
- [36] J. Ginibre, *J. Math. Phys.* **6**, 1432 (1965).
- [37] A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis* (Wiley, Hoboken, NJ, 2012).
- [38] B. Kaulakys, V. Gontis, and M. Alaburda, *Phys. Rev. E* **71**, 051105 (2005).
- [39] P. Kuopanportti, M. Möttönen, V. Bergholm, O.-P. Saira, J. Zhang, and K. Birgitta Whaley, *Phys. Rev. A* **77**, 032334 (2008).
- [40] Y. M. Galperin, B. L. Altshuler, J. Bergli, and D. V. Shantsev, *Phys. Rev. Lett.* **96**, 097009 (2006).
- [41] Y. M. Galperin, B. L. Altshuler, and D. V. Shantsev, *Fundamental Problems of Mesoscopic Physics*, NATO Science Series II: Mathematics, Physics, and Chemistry Vol. 154 (Springer, Amsterdam, 2004), pp. 141–165.
- [42] A. Shnirman, Yu. Makhlin, and G. Schön, *Phys. Scr. T* **102**, 147 (2002).