

Randomized benchmarking of quantum gates

E. Knill,^{*} D. Leibfried, R. Reichle,[†] J. Britton, R. B. Blakestad, J. D. Jost, C. Langer,[‡] R. Ozeri,[§] S. Seidelin,^{||} and D. J. Wineland

National Institute of Standards and Technology, MC 891, 325 Broadway, Boulder, Colorado 80305, USA

(Received 26 July 2007; published 8 January 2008)

A key requirement for scalable quantum computing is that elementary quantum gates can be implemented with sufficiently low error. One method for determining the error behavior of a gate implementation is to perform process tomography. However, standard process tomography is limited by errors in state preparation, measurement and one-qubit gates. It suffers from inefficient scaling with number of qubits and does not detect adverse error-compounding when gates are composed in long sequences. An additional problem is due to the fact that desirable error probabilities for scalable quantum computing are of the order of 0.0001 or lower. Experimentally proving such low errors is challenging. We describe a randomized benchmarking method that yields estimates of the computationally relevant errors without relying on accurate state preparation and measurement. Since it involves long sequences of randomly chosen gates, it also verifies that error behavior is stable when used in long computations. We implemented randomized benchmarking on trapped atomic ion qubits, establishing a one-qubit error probability per randomized $\pi/2$ pulse of 0.00482(17) in a particular experiment. We expect this error probability to be readily improved with straightforward technical modifications.

DOI: [10.1103/PhysRevA.77.012307](https://doi.org/10.1103/PhysRevA.77.012307)

PACS number(s): 03.67.Lx, 03.67.Pp, 37.10.De, 37.10.Gh

I. INTRODUCTION

In principle, quantum computing can be used to solve computational problems having no known efficient classical solutions, such as factoring and quantum physics simulations, and to significantly speed up unstructured searches and Monte Carlo simulations [1–4]. In order to realize these advantages of quantum computing, we need to coherently control large numbers of qubits for many computational steps. The smallest useful instances of the above-mentioned algorithmic applications require hundreds of qubits and many millions of steps. A quantum computing technology that realistically can be used to implement sufficiently large quantum computations is said to be “scalable.” Current quantum computing technologies that promise to be scalable have demonstrated preparation of nontrivial quantum states of up to eight qubits [5] but it is not yet possible to apply more than a few sequential two-qubit gates without excessive loss of coherence. Although there have been experiments to determine the behavior of isolated gates applied to prepared initial states [5–15], there have been no experiments to determine the noise affecting gates in a general computational context.

An important challenge of quantum computing experiments is to physically realize gates that have low error whenever and wherever they are applied. Studies of fault-tolerant quantum computing suggest that in order to avoid excessive resource overheads, the probability of error per unitary gate should be well below 10^{-2} [16–18]. The current consensus is

that it is a good idea to aim for error probabilities below 10^{-4} . What experiments can be used to verify such low error probabilities? One approach is to use process tomography to establish the complete behavior of a quantum gate. This requires that the one-qubit “analysis” gates employed in the tomography have lower error than the bound to be established on the gate under investigation. If this requirement is met, process tomography gives much useful information about the behavior of the gate, but fails to establish that the gate will work equally well in every context where it may be required. In particular, because process tomography requires preparing known initial states and one-qubit analysis gates, there is no obvious way to determine how well a gate works when used as the k th gate of a computation. To do so would require characterizing all the previous gates and, in some independent fashion, the analysis gates. Process tomography can also be very time consuming as its complexity scales exponentially with the number of qubits.

We propose a randomized benchmarking method to determine the error probability per gate in computational contexts. Randomization has been suggested as a tool for characterizing features of quantum noise in Ref. [19]. The authors propose implementing random unitary operators U followed by their inverses U^{-1} . Under the assumption that the noise model can be represented by a quantum operation acting independently between the implementations of U and U^{-1} , the effect of the randomization is to depolarize the noise. The average fidelity of the process applied to a pure initial state is the same as the average over pure states of the fidelity of the noise operation. (The latter average is known as the average fidelity and is closely related to the entanglement fidelity of an operation [20].) They also show that the average fidelity can be obtained with few random experiments. They then consider self-inverting sequences of random unitary operations of arbitrary length. Assuming that the noise can be represented by quantum operations that do not depend on the choice of unitaries, the fidelity decay of the sequence is

^{*}knill@boulder.nist.gov

[†]Present address: University of Ulm, Ulm, Germany.

[‡]Present address: Lockheed Martin, Huntsville, Alabama.

[§]Present address: Weizmann Institute of Science, Rehovot, Israel.

^{||}Present address: Institut Néel, Grenoble, France.

shown to represent the strength of the noise. Our randomized benchmarking procedure simplifies this procedure by restricting the unitaries to Clifford gates and by not requiring that the sequence is strictly self-inverting. An alternative approach to verifying that sequences of gates realize the desired quantum computation is given in Ref. [21]. In this approach, successively larger parts of quantum networks are verified by making measurements involving their action on entangled states. This “self-testing” strategy is very powerful and provably works under minimal assumptions on gate noise. It is theoretically efficient but requires significantly more resources and multisystem control than randomized benchmarking.

Our randomized benchmarking method involves applying random sequences of gates of varying lengths to a standard initial state. Each sequence ends with a randomized measurement that determines whether the correct final state was obtained. The average computationally relevant error per gate is obtained from the increase in error probability of the final measurements as a function of sequence length. The random gates are taken from the Clifford group [22], which is generated by $\pi/2$ rotations of the form $e^{-i\sigma\pi/4}$ with σ a product of Pauli operators acting on different qubits. The restriction to the Clifford group ensures that the measurements can be of one-qubit Pauli operators that yield at least one deterministic one-bit answer in the absence of errors. The restriction is justified by the fact that typical fault-tolerant architectures (those based on stabilizer codes) are most sensitive to errors in elementary Clifford gates such as the controlled NOT. Provided the errors in these gates are tolerated, other gates needed for universality are readily implemented by purified state preparation [16,23]. Note that the results of Ref. [19] hold if the unitaries are restricted to the Clifford group, because the Clifford group already has the property that noise is depolarized. We believe that randomized benchmarking yields computationally relevant errors even when the noise is induced by, and depends on, the gates, as is the case in practice. The relevant assumptions are given in Sec. V.

Randomized benchmarking as discussed and implemented here gives an overall average fidelity for the noise in gates. To obtain more specific information, the technique needs to be refined. In Ref. [24], randomization by error-free one-qubit unitaries is used to obtain more detailed information about noise acting on a multiqubit system. Randomized benchmarking can be adapted to use similar strategies.

II. RANDOMIZED BENCHMARK OF ONE QUBIT

For one qubit, our randomized benchmarking procedure consists of a large number of experiments, where each experiment consists of a pulse sequence that requires preparing an initial quantum state ρ , applying an alternating sequence of either major axis π pulses or identity operators (“Pauli randomization”) and $\pi/2$ pulses (“computational gates”), and performing a final measurement M . The pulse sequence between state preparation and measurement begins and ends with π pulses. For one qubit, the initial state is $|0\rangle$. Because the major axis π and $\pi/2$ rotations are in the Clifford group, the state is always an eigenstate of a Pauli operator during

the pulse sequence. The Pauli randomization applies unitary operators (“Pauli pulses”) that are (ideally) of the form $e^{\pm i\sigma_b\pi/2}$, where the sign \pm and $b=0,x,y,z$ are chosen uniformly at random and we define σ_0 to be the identity operator. For ideal pulses, the choice of sign determines only a global phase. However, in an implementation, the choice of sign can determine a physical setting that may affect the error behavior. The computational gates are $\pi/2$ pulses of the form $e^{\pm i\sigma_u\pi/4}$, with $u=x,y$. The sign and u are chosen uniformly at random, except for the last $\pi/2$ pulse, where u is chosen so that the final state is an eigenstate of σ_z . The computational gates generate the Clifford group for one qubit. Their choice is motivated by the fact that they are experimentally implementable as simple pulses. The final measurement is a von Neumann measurement of σ_z . The last $\pi/2$ pulse ensures that, in the absence of errors, the measurement has a known, deterministic outcome for a given pulse sequence. However, the randomization of the pulse sequence ensures that the outcome is not correlated with any individual pulse or proper subsequence of pulses. For a fixed such subsequence, the randomization of the other pulses ensures that the outcome is uniformly distributed.

The length l of a randomized pulse sequence is its number of $\pi/2$ pulses. The $\pi/2$ pulses are considered to be the ones that advance a computation. The π pulses serve only to randomize the errors. One can view their effect as being no more than a change of the Pauli frame. The Pauli frame consists of a Pauli operator that needs to be applied to obtain the intended computational state in the standard basis [16]. We call the $\pi/2$ and Pauli pulse combinations randomized computational gates. In principle, we can determine a pulse error rate by performing N experiments for each length $l=1, \dots, L$ to estimate the average probability p_l of the incorrect measurement outcome (or “error probability”) for sequences of length l . The relationship between l and p_l can be used to obtain an average probability of error per pulse. Suppose that all errors are independent and depolarizing. Let the depolarization probability of an operation A be d_A and consider a specific pulse sequence consisting of operations $A_0, A_1A_2, \dots, A_{2l+1}A_{2l+2}, A_{2l+3}$, where A_0 is the state preparation, A_1A_2 and the following pairs are the randomized computational gates, and A_{2l+3} the measurement. For the measurement, we can assume that the error immediately precedes a perfect measurement. The state after A_k is a known eigenstate of a Pauli operator or completely depolarized. Depolarization of the state is equivalent to applying a random Pauli or identity operator, each with probability $1/4$. The probability of the state’s not having been depolarized is $\prod_{j=0}^k (1 - d_{A_j})$. In particular, we can express $p_l = E\{[1 - \prod_{j=0}^{2l+3} (1 - d_{A_j})]/2\}$, where the function $E(\dots)$ gives the expectation over the random choices of the A_j . The factor of $1/2$ in the expression for p_l arises because depolarization results in the correct state $1/2$ of the time. The choices of the A_j are independent except for the last $\pi/2$ pulse. Assume that the depolarization probability of the last $\pi/2$ pulse does not depend on the previous pulses. We can then write $p_l = [1 - (1 - d_{if})(1 - d)^l]/2$, where d is the average depolarization probability of a random combination of one $\pi/2$ and one Pauli pulse (a randomized computational gate) and d_{if} combines

the depolarization probabilities of the preparation, initial Pauli pulse, and measurement. Thus p_l decays exponentially to $1/2$, and the decay constant yields d .

A commonly used metric to describe the deviation of an implemented gate from the intended gate is the average fidelity F_a , which is defined as the uniform average over pure input states of the fidelity of the output state with respect to the intended output state. We are interested in the average computationally relevant error per step consisting of a randomized computational gate (“average error” for short). This is given by the expectation over gates of $1 - F_a$ and relates to the depolarization parameter d of the previous paragraph by $1 - F_a = d/2$. In our implementation of the randomized computational gates, the π pulses around the z axis are implemented by changes in rotating frame and do not involve actively applying a pulse. Therefore, on average, the angular distance of the randomized gate’s action is π . As a result, $(1 - d/2)$ represents the average fidelity of pulses with action π .

Although estimates of p_l are sufficient to obtain the average error for a randomized computational gate, it is useful to consider the error behavior of specific randomized computations and even fixed instances of the randomized sequences. For this purpose, the sequences are generated by first producing N_G random sequences consisting of L random computational gates, where the gates are chosen independently without considering the final state. These sequences are considered to be a sample of typical computations. Each sequence is then truncated at different lengths. For each length, a $\pi/2$ pulse is appended to ensure that the final state is an eigenstate of σ_z . The sign of this final pulse is random. The resulting sequences are randomized by inserting the random Pauli pulses. We can then perform experiments to determine the probability of incorrect measurement outcomes for each such sequence and for each truncated computation after randomization by Pauli pulses. To be specific, the procedure is implemented as follows.

Randomized benchmarking for one qubit. This obtains measurement statistics for $N_G N_l N_p N_e$ experiments, where N_G is the number of different computational gate sequences, N_l is the number of lengths to which the sequences are truncated, N_p is the number of Pauli randomizations for each gate sequence, and N_e is the number of experiments for each specific sequence.

(1) Pick a set of lengths $l_1 < l_2 < \dots < l_{N_l}$. The goal is to determine the probability of error of randomized computations of each length.

(2) Do the following for each $j=1, \dots, N_G$.

(2a) Choose a random sequence $\mathcal{G}=\{G_1, \dots\}$ of l_{N_l} computational gates.

(2b) For each $k=1, \dots, N_l$ do the following.

(2b1) Determine the final state ρ_f obtained by applying $G_{l_k} \dots G_1$ to $|0\rangle$, assuming no error.

(2b2) Randomly pick a final computational gate R among the two $\pm x, \pm y, \pm z$ axis $\pi/2$ pulses that result in an eigenstate of σ_z when applied to ρ_f . Record which eigenstate is obtained.

(2b3) Do the following for each $m=1, \dots, N_p$.

(2b3a) Choose a random sequence $\mathcal{P}=\{P_1, \dots\}$ of l_k+2 Pauli pulses.

(2b3b) Experimentally implement the pulse sequence that applies $P_{l_k+2} R P_{l_k+1} G_{l_k} \dots G_1 P_1$ to $|0\rangle$ and measures σ_z , repeating the experiment N_e times.

(2b3c) From the experimental data and the expected outcome of the experiments in the absence of errors [from step (2b2) and the chosen Pauli pulses], obtain an estimate $p_{j,l_k,m}$ of the probability of error. Record the uncertainty of this estimate.

The probabilities of error p_l are obtained from the $p_{j,l_k,m}$ by averaging $p_{l_k} = \sum_{j=1}^{N_G} \sum_{m=1}^{N_p} p_{j,l_k,m} / (N_G N_p)$. We also obtain the probabilities of error for each computational gate sequence $p_{j,l_k} = \sum_{m=1}^{N_p} p_{j,l_k,m} / N_p$. If the errors are independent and depolarizing, the $p_{j,l_k,m}$ and the p_{j,l_k} should not differ significantly from the p_{l_k} . (Significance is determined by the statistical error in the measurement of the $p_{j,l_k,m}$ by the N_e experimental repetitions and the method for inferring expectations of σ_z from the actual measurements performed.) However, if the errors are systematic in the sense that each implemented pulse differs from the ideal pulse by a pulse-dependent unitary operator, this can be observed in the distribution of the $p_{j,l_k,m}$ over m . In this case, the final state of each implemented pulse sequence is pure. The deviation of these pure states from the expected states is distributed over the Bloch sphere as m and j are varied. For example, consider the case where p_{l_k} is close to $1/2$. If the errors are systematic, the $p_{j,l_k,m}$ are distributed as the probability amplitude of $|1\rangle$ for a random pure state. In particular, we are likely to find many instances of j and m where $p_{j,l_k,m}$ is close to 0 or 1, that is, differs significantly from $1/2$. In contrast, if the error is depolarizing, the $p_{j,l_k,m}$ are all close to $1/2$ independent of j and m .

III. TRAPPED-ION-QUBIT IMPLEMENTATION

We determined the computationally relevant error probabilities for computational gates on one qubit in an ion trap. The qubit was represented by two ground-state hyperfine levels of a ${}^9\text{Be}^+$ ion trapped in a linear radio-frequency Paul trap briefly described in Ref. [25]. It is the same trap that has been used in a several quantum information processing experiments [26–30]. The two qubit states are $|\downarrow\rangle$ ($F=2, m_F=-2$) and $|\uparrow\rangle$ ($F=1, m_F=-1$), where for our purposes, we identify $|\downarrow\rangle$ with $|0\rangle$ and $|\uparrow\rangle$ with $|1\rangle$. The state $|\downarrow\rangle$ is prepared by optical pumping, after laser cooling the motional states of the ion. We can distinguish between $|\downarrow\rangle$ and $|\uparrow\rangle$ by means of state-dependent laser fluorescence. Computational gates and Pauli pulses involving x - or y -axis rotations were implemented by means of two-photon stimulated Raman transitions. To ensure that the pulses were not sensitive to the remaining excitations of the motional degrees of freedom, we used copropagating Raman beams. It was therefore not necessary to cool to the motional ground state and only Doppler cooling was used. Pulses involving z -axis rotations were

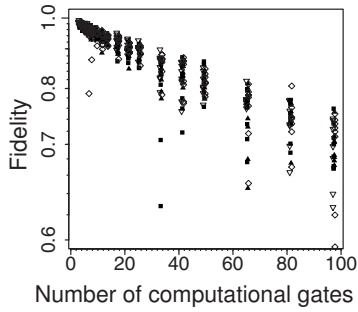


FIG. 1. Fidelity as a function of the number of steps for each randomized sequence. The fidelity (given by 1 minus the probability of error) is plotted on a logarithmic scale. The fidelity for the final state is measured for each randomized sequence. There are 32 points for each number of steps, corresponding to $N_p=8$ randomizations of each of $N_G=4$ different computational sequences. We used $N_e=8160$ for these experiments. Different symbols are used for the data for each computational sequence. The standard error of each point is between 0.001 (near fidelities of 1) and 0.006 (for the smaller fidelities). The scatter greatly exceeds the standard error, suggesting that coherent errors contribute significantly to the loss of fidelity.

implemented by programmed phase changes of one of the Raman beams. This changes the phase of the rotating reference frame and is equivalent to the desired z -axis rotation. The z -axis rotations were accompanied by a delay equivalent to the corresponding x and y pulses.

The Raman beams were switched on and off and shifted in phase and frequency as necessary by means of acousto-optic modulators controlled by a field-programmable gate array (FPGA). The pulse sequences were written in a special-purpose pulse-programming language and precompiled onto the FPGA. The version of the FPGA in use for the experiments was limited to about 100 computational pulses. The longest sequence in our experiments consisted of 96 computational gates. Our initial implementations clearly showed the effects of systematic errors in the distribution of the error probabilities of individual sequences. This proved to be a useful diagnostic and we were able to correct these systematics to some extent. One of the largest contributions to systematic errors was due to Stark shifts. To correct for these shifts, we calibrated them and adjusted phases in the pulse sequences.

IV. EXPERIMENTAL RESULTS

We generated $N_G=4$ random computational sequences and truncated them to the $N_l=17$ lengths $\{2,3,4,5,6,8,10,12,16,20,24,32,40,48,64,80,96\}$. Each truncated sequence was Pauli randomized $N_p=8$ times. Each final pulse sequence was applied to an ion a total of 8160 times in four groups that were interleaved with the other experiments in a randomized order. Pulse durations, qubit-resonant frequencies and Stark shifts were recalibrated automatically at regular intervals. The number of experiments per pulse sequence was sufficient to obtain the probability of incorrect measurement outcome with a statistical error small compared to the

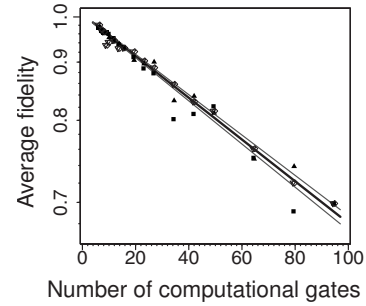


FIG. 2. Average fidelity as a function of the number of steps for each computational sequence. The points show the average randomized fidelity for four different computational gate sequences (indicated by the different symbols) as a function of the length. The average fidelity is plotted on a logarithmic scale. The middle line shows the fitted exponential decay. The upper and lower line show the boundaries of the 68% confidence interval for the fit. The standard deviation of each point due to measurement noise ranges from 0.0004 for values near 1 to 0.002 for the lower values, smaller than the size of the symbols. The empirical standard deviation based on the scatter in the points shown in Fig. 1 ranges from 0.0011 to 0.014. The slope implies an error probability of 0.00482(17) per randomized computational gate. The data is consistent with the gate's errors not depending on position in the sequence ($\chi^2=17.72$, 15 degrees of freedom, significance $p=0.28$).

variation due to randomization and systematic errors. Figure 1 plots the fidelity (one minus the probability of incorrect measurement outcome) of each of the $4 \times 17 \times 8 = 544$ final pulse sequences against the length of the corresponding computational sequence. As explained in the figure caption, the variation in fidelity for each length shows that nondepolarizing errors contribute significantly to error. Figure 2 plots the average fidelity over the eight Pauli randomizations of each computational sequence truncated to the different lengths. Pauli randomization removes coherent errors, significantly reducing the variation in fidelities for different computational sequences. The remaining variation could be due to the small sample of eight Pauli randomizations used to obtain the average. The empirical average probability of error per randomized computational gate can be obtained by fitting the exponential decay and was found to be 0.00482(17). The fit was consistent with a simple exponential decay, which suggests that these gates behave similarly in all computational contexts. The error bars represent standard deviation as determined by nonparameteric bootstrapping [31]. In what follows, if the fits are good, error bars are determined from nonlinear least-squares fits. In the cases where we can obtain a useful estimate of an error per randomized computational gate but the fits are poor, we used nonparameteric bootstrapping.

For our experimental setting, it is possible to perform experiments to quantify the different types of errors as a consistency check. The results of these experiments are in the Appendix and are consistent with the randomized benchmarking data.

V. THEORETICAL CONSIDERATIONS

The average error per randomized computational gate is obtained by fitting an exponential. For general error models,

it is possible that the initial behavior of the measured error probabilities does not represent the average error of interest, and it is the eventual decay behavior that is of interest. In this case, randomized benchmarking determines an asymptotic average error probability (AAEP) per randomized computational gate. It is desirable to relate the empirical AAEP to the average error probability (AEP) of a single randomized computational gate. As discussed above, the AAEP agrees with the AEP if the error of all operations is depolarizing and independent of the gates. It can be seen that for depolarizing errors, this relationship holds even if the error depends on the gates. In general, one can consider error models with the following properties.

Memoryless errors. The errors of each gate are described by a quantum operation. In particular, the “environment” for errors in one gate is independent of that in another.

Independent errors. For gates acting in parallel on disjoint qubits, each gate’s errors are described by a quantum operation acting on only that gate’s qubits.

Stationary errors. The errors depend only on the gate, not on where and when in the process the error occurs.

Subsystem preserving errors. The errors cause no leakage out of the subsystem defining the qubits.

Although the AAEP need not be identical to the AEP, we conjecture that there are useful bounds relating the two error probabilities. In particular, if the AAEP is zero then there is a fixed logical frame in which the AEP is zero. Trivially, if the AEP is zero, then the AAEP is zero.

Randomized benchmarking involves both Pauli randomization and computational gate randomization. The expected effect of Pauli randomization is to ensure that, to first order, errors consist of random (but not necessarily uniformly random) Pauli operators. Computational gate randomization ensures that we average errors over the Clifford group. If, as in our experimental implementation, the computational gates generate only the Clifford group, it takes a few steps for the effect to be close to averaging over the Clifford group. This process is expected to have the effect of making all errors equally visible to our measurement, even though the measurement is fixed in the logical basis and the last step of the randomized computation is picked so that the answer is deterministic in the absence of errors.

VI. BENCHMARKING MULTIPLE QUBITS

Scalable quantum computing requires not only having access to many qubits, but also the ability to apply many low-error quantum gates to these qubits. The error behavior of gates should not become worse as the computation proceeds. Randomized benchmarking can verify the ability to apply many multiqubit gates consistently.

Randomized benchmarking can be applied to two or more qubits by expanding the set of computational gates to include multiqubit gates. The initial state is $|0\dots 0\rangle$. Pauli randomization is performed as before and is expected to convert the error model to probabilistic Pauli errors to first order. Because the size of the Clifford group for two or more qubits is

large, one cannot expect to effect a random Clifford group element at each step. Instead, one has to rely on rapid mixing of random products of generators of the Clifford group to achieve (approximate) multiqubit depolarization. The number of computational steps that is required for approximate depolarization depends on the computational gate set. An example of a useful gate set consists of controlled NOTs (alternatively, controlled sign flips) combined with major-axis $\pi/2$ pulses on individual qubits. By including sufficiently many one-qubit variants of each gate, one can ensure that each step’s computational gates are randomized in the product of the one-qubit Clifford groups. This helps by having the effect of equalizing the probability of Pauli product errors of the same weight (see Ref. [24]).

The one-qubit randomized benchmark has a last step that ensures a deterministic answer for the measurement. For $n > 1$ qubits, one cannot expect deterministic answers for each qubit’s measurement, as this may require too complex a Clifford transformation. Instead, one can choose a random Pauli product that stabilizes the last state and apply a random product of one-qubit $\pi/2$ pulses with the property that this Pauli product is turned into a product of σ_z operators. If there is no error, measuring σ_z for each qubit and then computing the appropriate parity of the measurement outcomes gives a known deterministic answer. With error, the probability of obtaining the wrong parity can be thought of as a one-qubit error probability p for the sequence. If the error is completely depolarizing on all qubits, with depolarization probability d , then $p=d/2$, just as for one qubit. One expects that for sufficiently long sequences, p increases exponentially toward $1/2$ so that the asymptotic average error probability per randomized computational gate can be extracted as for one qubit.

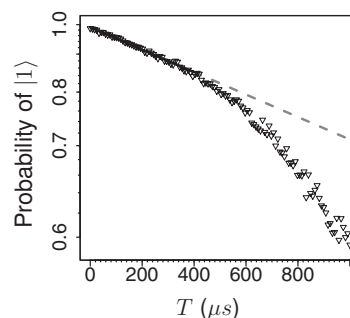


FIG. 3. Measurement of phase decoherence with refocusing. We measured the probability of $|1\rangle$ as a function of time for the standard refocused decoherence measurement. The pulse sequence consisted of a $\pi/2$ pulse at phase 0 followed by a delay of $T/2$, a π pulse at phase π , another delay of $T/2$ and a final $\pi/2$ pulse at phase π . The straight line shows the fit for exponential decay on the interval from 1 to 200 μs . Its extrapolation to larger times is shown dashed. The deviation from an exponential decay at larger times can be attributed to slow phase drifts that are no longer refocused by the single π pulse in the pulse sequence. From the fit, the contribution of unrefocusable phase decoherence to the error probability per step is 0.0037(1). The standard deviation of the plotted points ranges from 0.002 for values near 1 to 0.008 for the smallest values, similar to the apparent scatter of the plotted points.

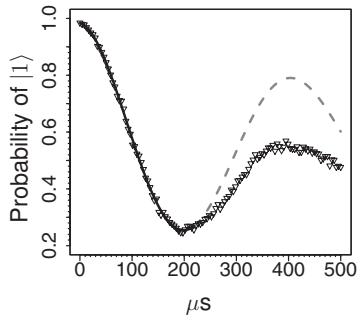


FIG. 4. Measurement of phase decoherence without refocusing. The randomized benchmark does not systematically refocus changes in frequency. To estimate the contribution to error from decoherence including refocusable decoherence, we performed the experiment of Fig. 3 without the refocusing pulse. This is essentially an on-resonance Ramsey experiment. It was not experimentally possible to eliminate the oscillatory shape of the curve by calibrating the frequency indicating that the oscillation was not simply caused by detuning from the resonant frequency. However, the shape is similar to what one would expect from a roughly periodic change in frequency that is not synchronized with the experiment. Such changes could come from magnetic field fluctuations and phase noise due to air currents in the paths of the two Raman beams. To estimate the contribution to the probability of error per step, we fitted an exponentially decaying $\cos(t)$ curve to the points with time coordinates less than $220 \mu\text{s}$. The extrapolation of the fitted curve (dashed) clearly deviates from the data. Note that for sinusoidal phase noise, the curve should be related to a decaying Bessel function. Fits to such a function also deviate from the experimental data, consistent with the phase fluctuations not being sinusoidal. Since the contribution to the probability of error is derived from the short-time behavior, the effect of the different models on the inferred probability of error per step is small. For the fit shown, the inferred contribution to the probability of error per step is $0.0090(7)$, larger than the error per step derived from Fig. 2. This is likely due to the fact that in the randomized sequences, the centering of the explicit π pulses in their intervals reduces this contribution by refocusing.

ACKNOWLEDGMENTS

We thank K. Coakley, S. Glancy, and R. Epstein for their comments on the manuscript. Our work was supported by the Disruptive Technology Office (DTO) under Contract No. 712868 and by the National Institute of Standards and Technology (NIST). It is a contribution of NIST, an agency of the U.S. government, and is not subject to U.S. copyright.

APPENDIX: DIRECT ERROR CHARACTERIZATIONS

We performed experiments to directly quantify the different types of errors in our pulses. These experiments characterize only the initial error (the error of the first gates) and serve as a consistency check for the randomized benchmarking data.

Known sources of errors include (a) phase errors due to fluctuating magnetic fields and changes in path length between the two Raman beams (they are merged on a polariz-

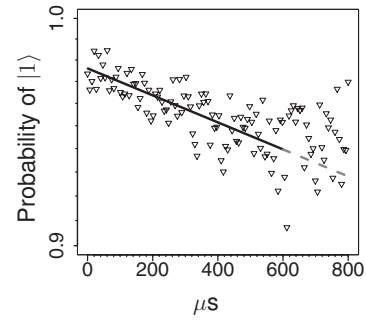


FIG. 5. Contribution of spontaneous emission to phase decoherence. To experimentally determine the contribution of spontaneous emission to decoherence, we applied the two Raman beams separately for half the time of each arm of the refocused decoherence measurement and compared the resulting data to that of Fig. 3 [32]. The points shown here were obtained by dividing the probabilities measured by the corresponding probabilities of Fig. 3, interpolating between the nearest points to match the time coordinates. The straight line shows the fitted exponential decay. The fit was weighted and used linear approximation to determine standard deviations of the points. The standard deviations used range from 0.003 to 0.015, which is substantially less than the apparent scatter of the plotted points. The inferred contribution to the error probability per step is $0.00038(3)$. This contribution can be estimated theoretically [32], which for the relevant configuration gives a value of approximately 0.0003.

ing beam splitter before targeting the ion), (b) amplitude errors due to changes in beam position at the ion and intensity fluctuations not compensated by the “noise eaters” (active-beam intensity stabilization), and (c) spontaneous emission

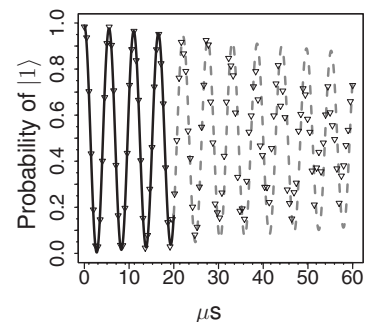


FIG. 6. Rabi flopping experiment. To determine the contribution to the probability of error per step due to pulse area error and associated decoherence, we performed a Rabi flopping experiment. We fitted the points to a decaying cosine curve with a possible phase offset and both linear and quadratic decay. Again, we restricted the fit to an initial segment of the data (black curve). The extrapolation (dashed curve) shows significant deviations. The random uncertainty in the points ranges from 0.002 to 0.007, less than the symbol size of the plotted points. The apparent scatter in the points near the end of the curve is likely due to slow fluctuations in pulse amplitude. The contribution to the probability of error per step as detected in this experiment is $0.006(3)$ if the calibration were based on this experiment. Automatically calibrated pulse times fluctuated by around $0.02 \mu\text{s}$. For pulse times differing by this amount, the contribution to the error per step is $0.007(3)$.

from the upper levels required for the stimulated Raman transition.

Phase decoherence can be measured by observing the decay of signal in a Ramsey spectrometry experiment of the qubit with or without refocusing [32]. Figure 3 shows the probability of observing $|1\rangle$ at the end of a refocused Ramsey experiment as a function of the delay between the first and last $\pi/2$ pulse. By fitting the initial part of the curve to an exponential decay, one can infer the contribution of unrefocusable phase error to each step of the Pauli randomized sequences. We obtained an estimate of 0.0037(1) for this contribution. Figure 4 shows the probability of observing $|1\rangle$ in a similar experiment but with the refocusing pulse omitted. This is an on-resonance Ramsey experiment. The fit suggests a contribution of 0.0090(7) for the error per step. This is larger than the inferred error from the randomized experiments, which can be explained by the refocusing effects of the Pauli randomization. See the caption of Fig. 4 for a discussion of fitting issues. We note that our benchmarking experiments, as well as the error characterizations in this section, were performed without line triggering the experiments, thereby making them sensitive to phase shifts caused by 60

Hz magnetic field fluctuations. Greatly improved decoherence times are typically obtained if such triggering is used.

The contribution of spontaneous emission to phase decoherence can be determined by a refocused Ramsey experiment where the two Raman beams are on separately half the time during the intervals between the pulses [32]. To determine the desired contribution, the probability of $|1\rangle$ as a function of time is compared to the data shown in Fig. 3. The results of the comparisons are in Fig. 5. The inferred contribution to the error probability per step is 0.00038(3), well below the contribution of the other sources of error.

The effect of amplitude fluctuations can be estimated from the loss of visibility of a Rabi flopping experiment. The data are shown in Fig. 6. Modeling the Rabi flopping curve is nontrivial and the fits are not very good. Nevertheless, we can estimate a contribution to the error probability per step from the behavior of the curve during the first few oscillations. This gives a contribution of 0.006(3), consistent with the probability of error per step obtained in the randomized experiments. Note that the contribution measured here also includes errors due to phase fluctuations during the computation pulses.

-
- [1] P. W. Shor, *SIAM J. Comput.* **26**, 1484 (1997).
 [2] R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467 (1982).
 [3] L. K. Grover, in *Proceedings of the 28th Annual ACM Symposium on the Theory of Computation* (ACM, New York, 1996), pp. 212–219.
 [4] D. S. Abrams and C. P. Williams, (1999), e-print arXiv:quant-ph/9908083.
 [5] H. Haffner *et al.*, *Nature (London)* **438**, 643 (2005).
 [6] A. M. Childs, I. L. Chuang, and D. W. Leung, *Phys. Rev. A* **64**, 012314 (2001).
 [7] Y. S. Weinstein, T. F. Havel, J. Emerson, N. Boulant, M. Saraceno, S. Lloyd, and D. G. Cory, *J. Chem. Phys.* **121**, 6117 (2004).
 [8] J. L. O’Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, *Phys. Rev. Lett.* **93**, 080502 (2004).
 [9] Y.-F. Huang and X.-F. Ren, *Phys. Rev. Lett.* **93**, 240501 (2004).
 [10] N. Kiesel, C. Schmid, U. Weber, R. Ursin, and H. Weinfurter, *Phys. Rev. Lett.* **95**, 210505 (2005).
 [11] D. Leibfried *et al.*, *Nature (London)* **422**, 412 (2003).
 [12] M. Riebe, M. Chwalla, J. Benhelm, H. Haeflner, W. Haensel, C. F. Roos, and R. Blatt, *New J. Phys.* **9**, 211 (2007).
 [13] K.-A. Brickman, P. C. Haljan, P. J. Lee, M. Acton, L. Deslauriers, and C. Monroe, *Phys. Rev. A* **72**, 050306(R) (2005).
 [14] J. P. Home, M. J. McDonnell, D. M. Lucas, G. Imreh, B. C. Keitch, D. J. Szwer, N. R. Thomas, S. C. Webster, D. N. Stacey, and A. M. Steane, *New J. Phys.* **8**, 188 (2006).
 [15] M. Steffen, M. Ansmann, R. C. Bialczak, N. Katz, E. Lucero, R. McDermott, M. Neeley, E. M. Weig, A. N. Cleland, and J. M. Martinis, *Science* **313**, 1423 (2006).
 [16] E. Knill, *Nature (London)* **434**, 39 (2005).
 [17] B. W. Reichardt, e-print arXiv:quant-ph/0406025.
 [18] R. Raussendorf and J. Harrington, *Phys. Rev. Lett.* **98**, 190504 (2007).
 [19] J. Emerson, R. Alicki, and K. Życzkowski, *J. Opt. B: Quantum Semiclassical Opt.* **7**, S347 (2005).
 [20] M. Horodecki, P. Horodecki, and R. Horodecki, *Phys. Rev. A* **60**, 1888 (1999).
 [21] F. Magniez, D. Mayers, M. Mosca, and H. Ollivier, e-print arXiv:quant-ph/0512111.
 [22] D. Gottesman, e-print arXiv:quant-ph/9807006.
 [23] S. Bravyi and A. Kitaev, *Phys. Rev. A* **71**, 022316 (2005).
 [24] B. Lévi, C. C. López, J. Emerson, and D. G. Cory, *Phys. Rev. A* **75**, 022314 (2007).
 [25] D. J. Wineland *et al.*, in *International Conference on Laser Spectroscopy*, Avemore, Scotland, edited by H. A. Hinds, A. Ferguson, and E. Riis (World Scientific, Singapore, 2005), pp. 393–402.
 [26] M. D. Barrett *et al.*, *Nature (London)* **429**, 737 (2004).
 [27] J. Chiaverini *et al.*, *Nature (London)* **432**, 602 (2004).
 [28] J. Chiaverini *et al.*, *Science* **308**, 997 (2005).
 [29] D. Leibfried *et al.*, *Nature (London)* **438**, 639 (2005).
 [30] R. Reichle, D. Leibfried, E. Knill, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Nature (London)* **443**, 838 (2006).
 [31] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
 [32] R. Ozeri *et al.*, *Phys. Rev. Lett.* **95**, 030403 (2005).