Randomized Benchmarking of Multiqubit Gates

J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland

Phys. Rev. Lett. **108**, 260503 — Published 26 June 2012

DOI: 10.1103/PhysRevLett.108.260503

Randomized Benchmarking of Multi-Qubit Gates

J. P. Gaebler,* A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke,†
J. D. Jost, J. P. Home,‡ E. Knill, D. Leibfried, and D. J. Wineland
National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA

We describe an extension of single-qubit gate randomized benchmarking that measures the error of multi-qubit gates in a quantum information processor. This platform-independent protocol evaluates the performance of Clifford unitaries, which form a basis of fault-tolerant quantum computing. We implemented the benchmarking protocol with trapped ions and found an error per random two-qubit Clifford unitary of 0.162 ± 0.008, thus setting the first benchmark for such unitaries. By implementing a second set of sequences with an extra two-qubit phase gate inserted at each step, we extracted an error per phase gate of 0.069 ± 0.017. We conducted these experiments with transported, sympathetically cooled ions in a multi-zone Paul trap—a system that can in principle be scaled to larger numbers of ions.

Quantum information processing (QIP) has the potential to solve difficult problems in many-body quantum mechanics and mathematics that lack efficient algorithms on classical computers. Useful QIP will require precise control of many qubits (two-level quantum systems) and implementation of quantum gates (operations that manipulate the quantum states of the qubits) with low error per gate. Here, the error per gate (EPG) is $\varepsilon = 1 - F$, where F is the average gate fidelity defined as the uniform average over pure input states of $\langle\psi|\rho|\psi\rangle$, where ρ is the actual density matrix and $|\psi\rangle$ is the intended output state [1]. Practical fault-tolerant QIP will require EPGs below a threshold of 10^{-4} [2, 3].

Currently, experiments have demonstrated the basic techniques needed for QIP, including the manipulation of small numbers of qubits and the implementation of the required quantum gates [4]. Remaining primary challenges are to scale up to larger numbers of qubits and to decrease the EPG below the fault-tolerant threshold. This requires being able to efficiently characterize or "benchmark" the performance of multi-qubit QIP experiments so as to extract the EPG of specific gates and enable comparison between different quantum computing platforms. With these goals in mind, we developed a benchmarking protocol for arbitrary numbers of qubits and demonstrated an experimental implementation for two qubits. The protocol extends previous work that used randomized sequences of Clifford gates to measure the EPG of one-qubit gates, first implemented in Refs. [5, 6].

Compared to techniques such as process tomography [7, 8], randomized benchmarking offers several key advantages for characterizing EPGs of quantum gates. It can determine EPGs with a number of measurements that scales polynomially with the number of qubits [5, 9], and, because it measures an exponential decay of fidelity as a function of the number of gates in a sequence, imperfections in state preparation and readout do not limit the minimum EPG that one can measure. Also, randomized benchmarking involves gates in the context of long sequences of operations and therefore establishes an EPG within a computational context similar to that expected in the implementation of lengthy QIP algorithms. Therefore, randomized benchmarking [5, 6] has been used to measure one-qubit gate errors in trapped ions [5, 10], superconducting qubits [11, 12], liquid NMR [6], and neutral atoms in an optical lattice [13] reaching $\varepsilon = 2.0(2) \times 10^{-5}$ in [10].

Previous work has characterized two-qubit gates with various techniques. With trapped ions, the fidelity for creating a Bell state has been measured [14–18] and process tomography was used to characterize single and repeated applications of a two-qubit entangling gate [19, 20]. Two-qubit gates have also been studied in superconducting and photonic qubits (see Ref. [4] and citations therein). In a liquid-state NMR system, a randomized benchmarking technique was used to study the errors of sequences of randomized gates on three nuclear spins [6] and found EPGs of 0.0047(3). The gates were randomly chosen in a platform-dependent way from a special-purpose probability distribution where the probability of a two-qubit gate (the CNOT) was 1/3. However, gate sets vary by platform, and other experiments may choose different probability distributions, for example to improve randomization. Therefore, the error probabilities from Ref. [6] may be difficult to compare to those obtained in future experiments.

The multi-qubit protocol we describe establishes a platform-independent error per operation (EPO) for Clifford unitaries by applying random sequences of Clifford unitaries of varying lengths. Clifford unitaries are fundamental to most error-correcting procedures envisioned for quantum computing (e.g., [21]) and thus serve as a foundation on which universal fault-tolerant quantum computing is built. The three main features of the group of Clifford unitaries that make it useful for our purposes are that its members have compact representations that can be efficiently converted to circuits of elementary quantum gates, outcomes of standard measurements of sequences of Clifford unitaries can be efficiently predicted by classi-

cal computation, and the group is sufficiently rich that error operators can be perfectly depolarized. Furthermore, by defining an EPO for Clifford operators, the benchmarking protocol allows for an unbiased comparison between different experimental platforms, even though each experiment may implement the Clifford operators differently. For a system of n qubits, Clifford unitaries can be constructed by combining one-qubit $\pm\frac{\pi}{2}$ rotations, defined as $\hat{R}_u(\pm\pi/2) = e^{\mp i\frac{\pi}{4}\sigma_u}$ with $u = x, y$, about the $\hat{x}$ and $\hat{y}$ axes, and two-qubit CNOT gates. We consider two gates or unitaries that differ only by a global phase to be equivalent. For two qubits, there are 720 non-equivalent Clifford gates (modulo Pauli operators on each qubit) [22].

In addition to determining an EPO for Clifford operations, we determine the EPG of particular gates of our choice by inserting them into the sequences of random Clifford unitaries. Of particular interest are implementations of one of the standard universal two-qubit gates such as the controlled-not (CNOT), phase gate (chosen here) or square-root of swap.

The protocol introduced here improves and extends a protocol based on random Clifford unitaries given in Refs. [9, 23]. The main differences include randomization of the final measurement to detect otherwise-hidden error models, additional Pauli randomization throughout to take advantage of good one-qubit gates, and a method for characterizing individual gates by inserting them into the protocol. This latter method for characterizing gates has now been independently implemented with one superconducting qubit [24]. The emphasis of Ref. [9] is on the relationship between true gate fidelities and the protocol-determined error probabilities, whereas here, we focus on the EPO's as useful system characteristics in their own right.

Here gates are performed in a multi-zone ion trap [25]. The qubit states are the $|F = 1, m_F = 0\rangle \equiv |\uparrow\rangle$ and $|2, 1\rangle \equiv |\downarrow\rangle$ hyperfine states of ^{9}Be$^+$, where F and m_F are the total angular momentum quantum numbers. The energy difference between these states is first-order insensitive to magnetic-field fluctuations at the applied field of 0.011964 T [20, 26]. The phase gate, $\hat{G}$, is implemented via a Mølmer-Sørensen gate (MS) [27], and acts as the diagonal matrix $[1, i, i, 1]$ in the basis $|\uparrow\uparrow\rangle$, $|\uparrow\downarrow\rangle$, $|\downarrow\uparrow\rangle$, and $|\downarrow\downarrow\rangle$. The experiment extends the work of [20] by using longer sequences of gates and a different implementation of the phase gate [28], which acts directly on a magnetic-field-insensitive transition in ^{9}Be$^+$.

We trap four ions in a six-zone linear Paul trap: two ^{9}Be$^+$ ion qubits, and two ^{24}Mg$^+$ ions that are used to sympathetically laser cool the qubit ions during the sequences. The ions form a linear chain along the axis of the trap, which is the axis of weakest confinement. Each experiment begins with the ions in the $|\downarrow\downarrow\rangle$ state and ends with a separate projective measurement of each qubit in the $|\uparrow\rangle$, $|\downarrow\rangle$ basis [20].

Single-qubit rotations $\hat{R}_u$ are implemented using co-propagating laser beams with stimulated-Raman $|\downarrow\rangle \leftrightarrow |\uparrow\rangle$ transitions on the ^{9}Be$^+$ ions after they are separated and held in two different trap zones [20, 26, 29]. The laser beam positions along the trap axis are controlled with an acousto-optic deflector that allows the beams to individually address ions in either trap zone. One-qubit σ_z gates (used for Pauli randomization between Clifford gates [22]) are implemented in software by shifting the phase of all future rotations for that qubit [30]. Identity gates are implemented with a wait time approximately equal to the $\hat{R}_u$ gate durations.

Two-qubit phase gates, $\hat{G}$, are performed with all four ions in the same trap zone, in the order ^{9}Be$^+$–^{24}Mg$^+$–^{24}Mg$^+$–^{9}Be$^+$. In contrast to the one-qubit gates, non-copropagating laser beams are used for the two-qubit gates. The MS gate is performed by the simultaneous application of detuned blue and red sidebands [27]. To implement $\hat{G}$ we surround a MS gate pulse with carrier $\pi/2$-pulses on both ions [22]. The advantage of using $\hat{G}$ as our elementary two-qubit gate rather than the MS gate is that this three-pulse sequence is insensitive to slow changes in the optical path-length difference of the non-copropagating beams [28]. Before performing each $\hat{G}$ gate we sympathetically laser-cool the four-ions, first using Doppler and then Raman sideband cooling of the ^{24}Mg$^+$ ions [20, 26, 29]. This ensures that each time we implement $\hat{G}$, the motional modes along the axial direction are cooled to near the ground state. The cooling light interacts only with ^{24}Mg$^+$ and thus preserves the qubit state coherences. The ability to reinitialize the motional state is the key to performing long sequences with multiple two-qubit gates in the presence of background heating and heating due to ion transport and is likely a necessary ingredient for scalable quantum computing with trapped ions [30, 31].

Sequences of length l for randomized benchmarking are generated as follows. For each l, many sequences of l random Clifford unitaries are produced. At the end of each such sequence, a randomized measurement step is added. This step consists of a Pauli randomization followed by a Clifford unitary that inverts the l preceding Clifford steps. The final Clifford unitary is chosen independently of the Pauli randomization. This ensures that in the absence of errors, the final state is again in the computational basis but randomized. This randomization allows us to detect certain otherwise-hidden error models such as the catastrophic errors described in Ref. [9]. Which basis element it should be in can be computed by use of standard efficient methods for simulating sequences of Clifford unitaries [32]. We compare our measurements of each qubit at the end of each sequence to this expected result to reveal errors.

For the experiment, we must determine an implementation of the Clifford unitaries in terms of the elementary gates available. For n-qubit gates, there exist effi-

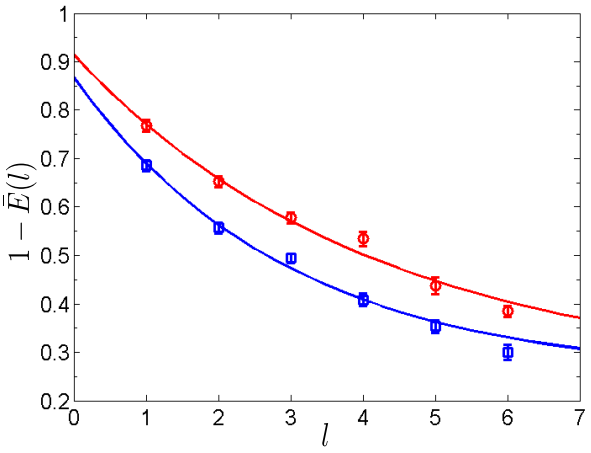

FIG. 1. The red circles show one minus the average probability of measuring an error at the end of sequences of random Clifford unitaries $\bar{E}(l)$ as a function of the sequence length l in the two-qubit benchmarking experiment. By fitting the data to the expression in Eq. (1) (red line), we find an error per random Clifford unitary $\varepsilon_g = 0.162(8)$. The preparation/measurement error, ε_m, is 0.086(22) (recall that measurement error includes the error for an additional inverting gate before detection). Blue squares show the results for running random sequences with an additional $\hat{G}$ inserted after each Clifford unitary. Fitting this data to the same functional form (blue line) and using Eq. 2 yields an error of $\varepsilon_{\hat{G}} = 0.069(17)$ and $\varepsilon_m = 0.132(26)$. The error bars in the plot represent the standard deviation of the mean of the sequences' frequency of correct measurement outcome. Error bars for inferred parameters are based on bootstrap resampling [22, 33].

cient algorithms that translate an arbitrary Clifford unitary into order of $n^2/\log(n)$ elementary one- and two-qubit gates [34, 35], each of which can then be mapped into experimentally available operations. However, for two qubits we used the following optimized strategy: By listing compact circuits of one-qubit rotations and phase gates $\hat{G}$, we determined for each of the 720 Clifford unitaries (modulo Pauli matrices) a circuit with the minimum number of phase gates to implement the corresponding Clifford unitary. On average, 1.5 phase gates and 6.5 effective $\frac{\pi}{2}$ pulses (one times the number of $\frac{\pi}{2}$ pulses plus two times the number of π pulses about the $\pm\hat{x}$ or $\pm\hat{y}$ axes) are required per Clifford unitary including the Pauli randomization.

The process of generating and implementing random sequences at each length is repeated in order to ensure randomization of the unitaries and their associated implementation errors. For our two-qubit benchmarking demonstration, we used the set of lengths $\{1, 2, 3, 4, 5, 6\}$ and generated between 15 and 55 random sequences of each length. We implemented approximately 100 runs for each sequence to determine its probability of error.

The experimental runs yield an average probability of error $E(l)$ for each length l shown in Fig. 1. To analyze $E(l)$ we start by assuming that each step's error behaves as a completely depolarizing channel (see, for example, Ref. [36], pg. 378) characterized by error probability ε_g independent of its gates or position in the sequence. Similarly, we assume an overall error probability ε_m for state preparation, the last inverting gate and its Pauli randomization, and measurement. In this case the mean of $E(l)$ with respect to repetitions of the experiment satisfies

$$\bar{E}(l) = \frac{1}{\alpha_n}\left(1 - (1 - \alpha_n\varepsilon_m)(1 - \alpha_n\varepsilon_g)^l\right) \qquad (1)$$

where $\alpha_n = \frac{2^n}{2^n-1}$ ($\alpha_2 = \frac{4}{3}$ for our two-qubit benchmark). Fitting the average probability of error to the above equation (blue line Fig. 1) we find $\varepsilon_g = 0.162(8)$. Assuming that experimental observations are consistent with the simple exponential behavior suggested by this formula, we use it as the defining formula for the EPO of a random Clifford unitary, regardless of the actual behavior of errors. In particular, we associate the EPO with the decay parameter of the error probabilities $\bar{E}(l)$ rather than a particular exact parameter of the underlying physical errors. If the simple depolarizing assumption does not hold, then $\bar{E}(l)$ may exhibit non-exponential and transient behaviors; however, the randomization is intended to induce behavior that matches the one implied by this assumption. In the experiment we were not able to implement sufficiently long sequences to clearly observe stationary behavior or to determine the extent to which the behavior is nonstationary [22].

To isolate the EPG of the phase gate $\hat{G}$ we generated a second set of sequences by inserting $\hat{G}$ after each random Clifford unitary. The final inverting Clifford unitary is chosen in the same way as before, taking into account the effect of the additional $\hat{G}$ gates to ensure that the final state is a predictable computational basis state in the absence of errors. The average probability of error measured for the implementation of this experiment should also satisfy Eq. 1, but with a different value of ε_g due to the additional operation in each step. In an ideal experiment ε_m should be the same, but the model must take into consideration that it might have changed, for example due to experimental drifts. The EPG is given by

$$\varepsilon_G = \frac{1}{\alpha_n}\left(1 - \frac{1 - \alpha_n\varepsilon_g'}{1 - \alpha_n\varepsilon_g}\right), \qquad (2)$$

where ε_g' is the probability of error of a step consisting of a random Clifford gate with an extra $\hat{G}$ inserted.

The blue data points and curve in Fig. 1 show the results from the $\hat{G}$ benchmark. Curve-fitting and solving the above equations for $\varepsilon_{\hat{G}}$ give an EPG of $\varepsilon_{\hat{G}} = 0.069(17)$. The EPG for our particular choice of $\hat{G}$ in these experiments shows no improvement over the gates

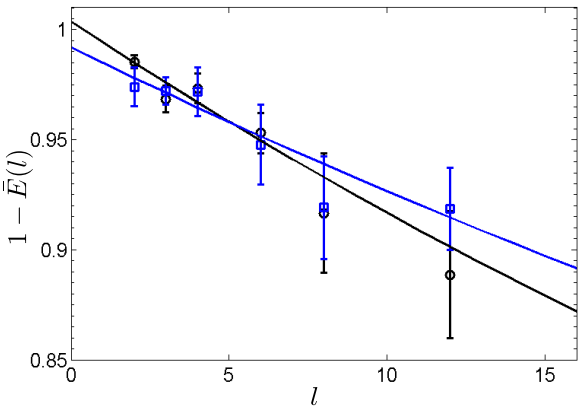

FIG. 2. Black circles and blue squares show one minus the average probability of error for each qubit in the single-qubit benchmarking experiment. The solid lines are the best fits of the data to Eq. 1 with $n = 1$ [5]. The best fit values for ε_g give the error per step (as defined in the text for the single-qubit benchmark), which we find to be 0.010(2) and 0.007(2), respectively.

used in [20], but applies to gates used in computationally relevant contexts in longer sequences.

The main sources of error in the phase gates are due to drifts in the laser beam intensities at the ions' positions, which we estimate to contribute an error of up to 0.03. We estimate spontaneous emission [37, 38] contributes an error probability of 0.013 to the phase gate. Including other known sources of errors predicts an overall error less than what we measure [22]. This suggests that our model of errors for the phase gate is incomplete.

As an independent check on phase-gate fidelity, we measured the state fidelity for a Bell state created by use of the phase gate $\hat{G}$ as in [14, 15, 18]. We determined an error in the Bell state of 0.09(2), which is consistent with the EPG determined by the benchmark. The Bell state error includes additional errors due to three single-qubit rotations on each ion needed to create and analyze the Bell state.

We also performed a benchmark to determine the error in the single-qubit gates. We did not implement the protocol for benchmarking $\hat{G}$ described above, but used the standard one-qubit benchmarking protocol of [5]. In this protocol the length of a sequence is the number of computational gates that consist of a Pauli gate (π-pulse) followed by a Clifford gate ($\frac{\pi}{2}$-pulse) on each qubit. The gate sequence is followed by a Pauli gate and Clifford gate chosen to yield a predictable measurement outcome in the $|\uparrow\rangle$, $|\downarrow\rangle$ basis for each qubit. The Pauli gates are chosen with equal probability to be rotations about the $\hat{x}, \hat{y}$ or $\hat{z}$ axis or the identity. The Clifford gates are chosen with equal probability to be $\frac{\pi}{2}$-pulses about the $\hat{x}$ or $\hat{y}$ axis.

Benchmarks were performed on each qubit in parallel. The results shown in Fig. 2 were executed in one set, after all of the two-qubit benchmarks and following a recalibration of the one-qubit gates. The number of sequences implemented was $15, 13, 6, 13, 12, 14$ for sequence lengths of $2, 3, 4, 6, 8, 12$, respectively. We ran each sequence approximately 100 times. In order to replicate the conditions of the experiment for the two-qubit benchmark the ions were recombined into a single trap zone in each step, recooled and then held for approximately the same duration required to execute $\hat{G}$ before being separated again for application of single-qubit computational gates in the next sequence step. The inferred one-qubit errors per step are 0.010(2) and 0.007(2) for the respective qubits. The EPG for the single-qubit gates combined with the inferred EPG for $\hat{G}$ is consistent with the measured EPO for the two-qubit Clifford operations [22].

Clifford benchmarks as described here can serve as a platform-independent strategy for comparing the quality of quantum operations in a computational context. When benchmarking n qubits, we suggest that the benchmarks are applied to different subsets of the qubits so that comparable EPOs are obtained for $n = 1, 2, 3, \ldots$ qubits. In this way the results can be compared to other experimental platforms that have different numbers of available computational qubits and can also be used for investigating differences in behavior that depend on (for example) geometrical relationships between qubits.

In summary, we have described a protocol for randomized benchmarking of gates in a quantum information processor and implemented the protocol experimentally on two qubits to measure the error per operation of arbitrary two-qubit Clifford unitaries. The protocol we propose is independent of the gate set that is experimentally implemented and so can provide an easily portable method for evaluating the performance of Clifford unitaries on different physical platforms. Furthermore, with this method it is straightforward to isolate the fidelity of a specific two-qubit gate. Looking ahead, this randomized benchmarking protocol should prove useful as different experimental implementations of quantum information processors look to increase the number of qubits and work to decrease the errors towards what is required for fault-tolerance.

This work was supported by NSA, IARPA, ONR and the NIST Quantum Information Program. The authors thank S. Glancy and B. C. Sawyer for careful readings of the manuscript. JPG is supported by NIST through an NRC fellowship. This paper is a contribution by NIST and not subject to U.S. copyright.

* Electronic address: john.gaebler@nist.gov
† Current address: Department of Physics, Amherst College, Amherst, MA, 01002-5000 USA

‡ Current address: Institute for Quantum Electronics, ETH Zurich, 8093 Zurich, CH

[1] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80**, 012304 (2009), arXiv:quant-ph/0606161.

[2] J. Preskill, Proc. R. Soc. Lond. A **454**, 385 (1998).

[3] E. Knill, Nature **463**, 441 (2010).

[4] T. D. Ladd *et al.*, Nature **464**, 45 (2010).

[5] E. Knill *et al.*, Phys. Rev. A **77**, 012307 (2008).

[6] C. A. Ryan, M. Laforest, and R. Laflamme, New Journal of Physics **11**, 013034 (2009).

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

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

[9] E. Magesan, J. M. Gambetta, and J. Emerson, arXiv:1109.6887 (unpublished).

[10] K. R. Brown *et al.*, Phys. Rev. A **84**, 030303(R) (2011), arXiv:1104.2552.

[11] J. M. Chow *et al.*, Phys. Rev. Lett. **102**, 090502 (2009).

[12] J. M. Chow *et al.*, Phys. Rev. A **82**, 040305(R) (2010).

[13] S. Olmschenk, R. Chicireanu, and J. V. Porto, New Journal of Physics **12**, 113007 (2010).

[14] C. A. Sackett *et al.*, Nature **404**, 256 (2000).

[15] D. Leibfried *et al.*, Nature **422**, 412 (2003).

[16] P. C. Haljan *et al.*, Phys. Rev. A **72**, 062316 (2005).

[17] J. P. Home *et al.*, New Journal of Physics **8**, 188 (2006).

[18] J. Benhelm, G. Kirchmair, C. F. Roos, and R. Blatt, Nat Phys **4**, 463 (2008).

[19] M. Riebe *et al.*, Phys. Rev. Lett. **97**, 220407/1 (2006).

[20] J. P. Home *et al.*, Science **325**, 1227 (2009).

[21] D. Gottesman, in *Proceedings of Symposia in Applied Mathematics*, edited by J. Samuel J. Lomonaco (2010), Vol. 68, pp. 13–58, arXiv:0904.2557.

[22] Further details on the multi-qubit benchmarking protocol and the experimental data and analysis are given in J. P. Gaebler *et al.*, arXiv:1203.3733v1 (unpublished).

[23] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504/1 (2011).

[24] E. Magesan *et al.*, (2012), arXiv:1203.4550.

[25] J. D. Jost, Ph.D. thesis, Univ. of Colorado, Boulder, Boulder, CO, 2010.

[26] D. Hanneke *et al.*, Nature Physics **6**, 13 (2010).

[27] A. Sørensen and K. Mølmer, Phys. Rev. Lett. **82**, 1971 (1999).

[28] P. J. Lee *et al.*, J. Opt. B **7**, S371 (2005).

[29] J. D. Jost *et al.*, Nature **459**, 683 (2009).

[30] D. J. Wineland *et al.*, J. Res. Natl. Inst. Stand. Technol. **103**, 259 (1998).

[31] D. Kielpinski, C. Monroe, and D. J. Wineland, Nature **417**, 709 (2002).

[32] D. Gottesman, Ph.D. thesis, Calif. Inst. Tech, Pasadena, California, 1997, quant-ph/9705052.

[33] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).

[34] S. Aaronson and D. Gottesman, Phys. Rev. A **70**, 052328 (2004).

[35] K. N. Patel, I. L. Markov, and J. P. Hayes, Quantum Information and Computation **8**, 0282 (2008).

[36] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2001).

[37] R. Ozeri *et al.*, Phys. Rev. A **75**, 042329 (2007).

[38] H. Uys *et al.*, Phys. Rev. Lett. **105**, 200401 (2010).