Comparing Experiments to the Fault-Tolerance Threshold

Richard Kueng,[1,2,3] David M. Long,[1] Andrew C. Doherty,[1] and Steven T. Flammia[1]
[1]*Centre for Engineered Quantum Systems, School of Physics, University of Sydney, Sydney, 2006 New South Wales, Australia*
[2]*Institute for Theoretical Physics, University of Cologne, D-50937 Cologne, Germany*
[3]*Institute for Physics and FDM, University of Freiburg, D-79104 Freiburg, Germany*
(Received 4 November 2015; published 21 October 2016)

Achieving error rates that meet or exceed the fault-tolerance threshold is a central goal for quantum computing experiments, and measuring these error rates using randomized benchmarking is now routine. However, direct comparison between measured error rates and thresholds is complicated by the fact that benchmarking estimates average error rates while thresholds reflect worst-case behavior when a gate is used as part of a large computation. These two measures of error can differ by orders of magnitude in the regime of interest. Here we facilitate comparison between the experimentally accessible average error rates and the worst-case quantities that arise in current threshold theorems by deriving relations between the two for a variety of physical noise sources. Our results indicate that it is coherent errors that lead to an enormous mismatch between average and worst case, and we quantify how well these errors must be controlled to ensure fair comparison between average error probabilities and fault-tolerance thresholds.

DOI: 10.1103/PhysRevLett.117.170502

The fault-tolerance threshold theorem is a fundamental result that justifies the tremendous interest in building large-scale quantum computers despite the formidable practical difficulties imposed by noise and imperfections. This theorem gives a theoretical guarantee that quantum computers can be built in principle if the noise strength and correlation are below some threshold value [1–3].

To make precise statements of threshold theorems, we must quantify the strength of errors in noisy quantum operations. Ideally we would do this in terms of quantities that can be measured in experiments. A standard measure for quantifying errors in quantum gates is given by the average error rate, which is defined as the infidelity between the output of an ideal unitary gate $\mathcal{U}$ and a noisy version $\mathcal{E}\mathcal{U}$ with noise process $\mathcal{E}$, uniformly averaged over all pure states,

$$r(\mathcal{E}) = 1 - \int d\psi \langle\psi|\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle. \quad (1)$$

This quantity has many virtues: it can be estimated efficiently for any ideal gate $\mathcal{U}$, and in a manner that is independent of state preparation and measurement (SPAM) errors by using the now-standard method of randomized benchmarking [4–7]. Recent experimental implementations include [8–17].

The major drawback of using Eq. (1) to quantify gate errors is that it is only a proxy for the actual quantity of interest, i.e., the fault-tolerance threshold. This is because r captures average-case behavior for a single use of the gate, while fault-tolerance theorems characterize noise in terms of worst-case performance when the gate is used repeatedly in a large computation. The importance of this distinction has recently been emphasized by Sanders *et al.* [18], who gave explicit examples of noise with a large discrepancy between average- and worst-case error and showed that it is possible for the worst-case error to scale like $\sqrt{r}$. For some noise types (such as pure dephasing and depolarizing noise), the worst- and average-case behavior essentially coincide [19]. However, for other classes of errors, notably for experimentally relevant errors in detuning and calibration that lead to over- or under-rotation, the worst-case behavior can be orders of magnitude worse than the average in the relevant regime of $r \ll 1$. Thus, it is not possible to directly compare a measured value of r to a threshold result. Despite this, experimentalists are increasingly wishing to relate the results of benchmarking experiments to fault-tolerance thresholds. There is, thus, a pressing need for techniques that allow for direct comparison between experimentally measurable error rates and fault-tolerance thresholds.

In this Letter, we investigate the relationship between worst-case and average-case error for a wide range of error models that are relevant to experiments. First, we show that while closed -form expressions do not typically exist, well-established theoretical techniques of convex optimization are often sufficient to determine the relationship between average-case and worst-case errors for models of physical interest. The details of these computations are largely relegated to the Supplemental Material [20]. Second, we study a wide range of error models for 1-qubit gates. Our main example is of a 1-qubit gate with combined dephasing and calibration error. This allows us to demonstrate the crossover between a regime dominated by dephasing, where average-case and worst-case errors are not too different, and the limit of a unitary noise, where the worst-case error scales like $\sqrt{r}$. We then turn to general bounds on worst-case error, showing that it scales as $\sqrt{r}$ for all unitary errors and that for a wide class of errors it can be accurately estimated in terms of r and a recently introduced measure of how close an error process is to being unitary. Finally, conventional benchmarking experiments contain a lot more information than is required just to extract r. We find that this information can often be used to show that the worst-case error has an unfavorable scaling. This is an area that we hope will attract much more study in future.

Fault-tolerance thresholds.—A wide range of fault-tolerance thresholds have been reported. The value of the threshold depends greatly on the fault-tolerant procedures that are used, on the noise model that is assumed, and whether the threshold is determined from (possibly conservative) analytic bounds on the error, or from (possibly optimistic) numerical simulations. We emphasize that the errors that are given in theoretical fault-tolerance papers typically refer to some measure of worst-case error. For example, the widely known results of Aliferis and collaborators [26–28] use concatenated error-correcting codes and consider a stochastic adversarial noise model that includes all of the noise processes that we will discuss in this Letter. These papers find that large-scale quantum computation can be performed for errors below a few times 10^{-4}, when that error is quantified by a measure of worst-case error such as the diamond distance that we discuss below. For more optimistic noise models and for fault-tolerant protocols such as the widely known surface code approaches, the threshold is around 10^{-2} based on numerical simulations of Pauli errors [29]. For Pauli noise, however, there is no significant difference between worst-case and average-case errors [19]. The performance of these schemes in the presence of coherent errors is not yet understood.

It is possible to state a version of the threshold theorem directly in terms of r, but given current knowledge the thresholds in these theorems would be roughly the square of current thresholds (around 10^{-8} for [26–28]). It is unclear if this can be significantly improved upon, because it may be that it is the worst-case error that is physically relevant to the success of the computation. However, our results here motivate research into whether current fault-tolerance results could be strengthened to provide significantly improved thresholds when expressed in terms of r for error models sufficiently general to include coherent errors.

Diamond distance.—We will now describe the most commonly used metric of worst-case error for quantum processes. Any candidate measure of distance $\Delta(\mathcal{E},\mathcal{F})$ between noise operations $\mathcal{E}$ and $\mathcal{F}$ should satisfy certain desirable properties [30]. (The operation $\mathcal{F}$ should be thought of as a perfect identity gate for our purposes.) First, like any good distance measure, it should have the structure of a metric, which in particular means it should be symmetric, positive, and obey the triangle inequality. Less obviously, but even more importantly, it should obey two additional properties: chaining and stability. The chaining property,

$$\Delta(\mathcal{E}_2\mathcal{E}_1,\mathcal{F}_2\mathcal{F}_1) \le \Delta(\mathcal{E}_1,\mathcal{F}_1) + \Delta(\mathcal{E}_2,\mathcal{F}_2), \qquad (2)$$

says that composing two noisy operations cannot amplify the error by more than the sum of the two individual errors. Thus, errors can grow at most linearly in the number of operations. The stability property states that the error metric for a single gate should be independent of whether that gate is embedded in a larger computation. Thus, we require

$$\Delta(\mathcal{I}\otimes\mathcal{E},\mathcal{I}\otimes\mathcal{F}) = \Delta(\mathcal{E},\mathcal{F}), \qquad (3)$$

where $\mathcal{I}$ is the identity operation. This ensures that our measure is robust even if the input to the gate is entangled with other qubits in the computation.

The diamond distance, whose formal definition is

$$D(\mathcal{E},\mathcal{F}) = \frac{1}{2}\max_{\rho}\|\mathcal{I}\otimes\mathcal{F}(\rho) - \mathcal{I}\otimes\mathcal{E}(\rho)\|_1, \qquad (4)$$

satisfies each of these physically motivated *desiderata* [1]. It also has an appealing operational interpretation as the maximum probability of distinguishing the output of the noisy gate from the ideal output [1,31]. It is not obvious from the definition how to do practical computations with this quantity, but it can be computed efficiently using the methods of semidefinite programming [32–34]. Because of these properties, the diamond distance is an ideal measure for quantifying noise for the purposes of a fault-tolerance threshold, although in principle other quantities could be employed as well [2].

The only drawback of this quantity is that it is not known how to measure it directly in experiments. It is therefore of interest to have a conversion to, or at least bounds for, diamond distance in terms of the average gate fidelity. To date, the best-known bounds for a d-level quantum gate are [35]

$$\frac{d+1}{d}r \le D \le \sqrt{d(d+1)r},$$

but it is unknown for what conditions these bounds are tight.

Single-qubit calibration and dephasing errors.—In order to discuss the relationship between average-case and worst-case errors in quantum computing demonstration experiments, we will now analyze in detail a simple but physically relevant noise model for a single-qubit gate. Suppose that the gate is implemented by the noisy control Hamiltonian $H_c = J(t)\sigma_z$. Because of experimental imperfections, the control $J(t)$ that is implemented is distinct from the nominal control $J_0(t)$ that would perfectly implement the required gate. Physically, this noise results in two distinct types of errors: dephasing, where $\delta J(t) = J - J_0$ varies stochastically between uses of the gate, and calibration error, where δJ takes the same fixed value each time the gate is used. Where $\delta J(t)$ is stochastically varying, we assume that the noise level does not change with time, and that the noise spectrum for $\delta J(t)$ is mainly confined to frequencies $f > 1/t_g$, where t_g is the time required to perform the gate. When averaged over uses of the gate, the resulting noisy operation is $\mathcal{EU}$, where $\mathcal{U}$ is the desired gate and the noise process amounts to

$$\mathcal{E}(\rho) = p\sigma_z e^{-i\delta\sigma_z}\rho e^{i\delta\sigma_z}\sigma_z + (1-p)e^{-i\delta\sigma_z}\rho e^{i\delta\sigma_z}. \quad (5)$$

In this noise model, the dephasing noise rate p arises from the time-varying noise on the gate, while the unitary over rotation δ results from the fixed miscalibration of the control pulse $J(t)$. [Although we speak here in terms of calibration errors, this also approximately captures the effects of highly non-Markovian errors arising from very low-frequency noise in $J(t)$.]

This noise model roughly captures many experimental gates, but more importantly it will demonstrate the range of behaviors that can be expected in terms of the relationship between average-case and worst-case error. Specifically, when $\delta = 0$ we have a pure dephasing process. For such errors [19] the worst-case error scales like r, so this is the most favorable possible behavior. On the other hand, for $p = 0$ we have a purely unitary rotation error that has the worst possible behavior, where the worst-case error scales like $\sqrt{r}$.

Using well-known techniques [36,37] we find the average error rate for this calibration and dephasing (CD) noise to be $r_{\rm CD} = \frac{2}{3}(p\cos(2\delta) + \sin^2\delta)$. Employing the semidefinite programming approach of Refs. [19,32], we can evaluate the diamond distance for this noise channel and find $D_{\rm CD} = \sqrt{\frac{3}{2}r_{\rm CD} - p(1-p)}$. A logarithmic plot of the ratio $D_{\rm CD}/r_{\rm CD}$ is shown in Fig. 1.

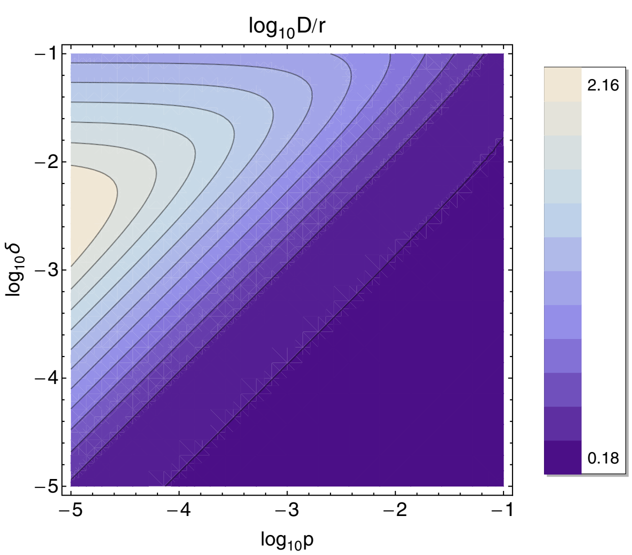

FIG. 1. Average error rate r and worst-case error rate (diamond distance) D for a combination of dephasing and unitary errors. The logarithmic plot is of D/r, which quantifies how much greater the worst-case error is than the average case as a function of a unitary over rotation angle δ and a dephasing probability p, where the exact noise process is given in Eq. (5). When $p \geq \delta$, D and r are comparable to within a small factor, but as soon as $\delta > p$, D rapidly becomes much greater than r.

In the interesting regime of low error we find $r_{\rm CD} \simeq 2(p + \delta^2)/3$, while $D_{\rm CD} \simeq \sqrt{p^2 + \delta^2}$. From this, we can see that when $p \gg |\delta|$ we have $D_{\rm CD} \simeq 3r_{\rm CD}/2$, as for a pure dephasing process, and there is no great difference between worst-case and average-case errors. However, as the calibration error grows, the worst-case error grows significantly. When calibration error dominates, $|\delta| \gg p$, we find $D_{\rm CD} \simeq \sqrt{3r_{\rm CD}/2}$. In this regime, an average error rate $r_{\rm CD}$ of around 10^{-4} corresponds to a worst-case error of more than one percent. Physically, then, we see that as dephasing error is reduced in a particular experimental setting, this places more stringent demands on the calibration required if the average error rate r is to be compared directly to a fault-tolerance threshold.

Single-qubit relaxation errors.—Another natural single-qubit noise process to consider is qubit relaxation or amplitude damping errors [spontaneous emission or a T_1 process in nuclear magnetic resonance (NMR) language], at finite temperature. In this process a qubit with energy splitting E is coupled to a bath at temperature T. Define as in [38] the probability for a decay process during the action of the gate is γp and the probability to go from the ground to the excited state is $\gamma(1-p)$. The ratio of upgoing to downgoing transition rates $p/(1-p) = \exp(-E/k_B T)$ is the Boltzmann factor, which allows us to identify $p = 1/2$ as infinite temperature and $p = 1$ as zero temperature. For this amplitude damping (AD) noise channel we find $r_{\rm AD} = (1 - \sqrt{1-\gamma} + \gamma/2)/3$. Although we have no closed-form expression for the worst-case error for these channels, we have adapted standard techniques in the analysis of semidefinite programs to find the bound $D_{\rm AD} \leq 3r_{\rm AD}\max\{p, 1-p\}$. Therefore, we have a guarantee that the average-case and worst-case errors are not too different. Comparing with a direct evaluation of the semidefinite program, we find $D_{\rm AD} \simeq 3r_{\rm AD}$ for zero temperature ($p = 1$) and low noise $r_{\rm AD} \ll 1$, so this is the tightest bound possible. In the limit of high temperature $p \to 1/2$ we approach a dephasing channel and recover the formula $D_{\rm AD} = 3r_{\rm AD}/2$. This behavior is illustrated in Fig. 2.

Leakage errors.—Another important class of errors encountered in experiment is leakage errors. Modified randomized benchmarking protocols for leakage errors are proposed in [39,40]. In Ref. [39] it was shown that a nearly trivial modification of a standard benchmarking protocol in the presence of leakage errors can still be used to determine the average error rate r, so we again use this figure of merit for comparison. For a leakage model we need to consider a larger space of states, so we add a leakage level $|l\rangle$ to the two 1-qubit states $|0\rangle, |1\rangle$. We follow [40] in distinguishing coherent and incoherent leakage errors and compare the average-case error to the true worst-case error; this will be the diamond distance on the *full* state space including both the leakage and qubit states. Fault-tolerance theorems also exist for leakage error processes [41], and this is the appropriate noise measure to compare with the numerical values found in those papers.

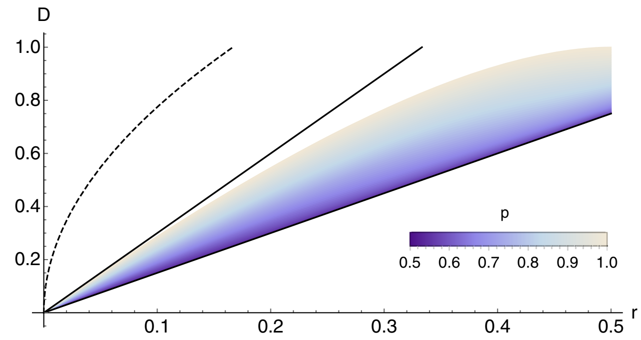

FIG. 2. Tradeoff between average error rate r and the worst-case error rate in terms of the diamond distance D for the thermal amplitude damping channel, where the parameter p controls the temperature with $p = 1$ corresponding to zero temperature and $p = 1/2$ corresponding to infinite temperature. The dashed line is the previous best upper bound [35], while the upper black line is the new bound derived here. The zero-temperature limit ($p = 1$) gives the least-favorable scaling of D with r, but in every case the bound $D \leq 3r$ holds. The infinite-temperature limit ($p = 1/2$) recovers the known value of $D = 1.5r$.

As an example of incoherent leakage (IL), we will consider the case where the qubit state $|1\rangle$ relaxes to $|l\rangle$ with probability p. A benchmarking experiment (following [39]) then obtains the average-case error $r_{\mathrm{IL}} = [1 - \sqrt{1-p} + p]/3$, where this is now the infidelity averaged over states initially in the qubit subspace. Because this process is so similar to the amplitude damping channel, we can use analogous techniques to find the inequality $D_{\mathrm{IL}} \leq 2r_{\mathrm{IL}}$. Thus, for this error process the average-case and worst-case error again almost coincide.

As an example of coherent leakage (CL), consider the unitary noise process $\mathcal{E}_{\mathrm{CL}}(\rho) = U(\delta)\rho U(\delta)^\dagger$ given by $U(\delta) = \exp[-i\delta(|1\rangle\langle l| + |l\rangle\langle 1|)]$. For this noise process one obtains $r_{\mathrm{CL}} = [1 - \cos\delta - \cos^2\delta]/3$. However, as for the unitary errors discussed above, the worst-case error can be much larger than this. We find $\sqrt{3r_{\mathrm{CL}}/2} \leq D_{\mathrm{CL}} = |\sin\delta| \leq \sqrt{2r_{\mathrm{CL}}}$ for all $\delta \in [-\pi/2, \pi/2]$ and, consequently, the worst-case error scales like $\sqrt{r_{\mathrm{CL}}}$. Where leakage errors are possible, it would be important to use the methods of [40], or some other method to bound coherent leakage errors, before comparing the average-case error r to a fault-tolerance threshold.

Unitary errors.—In looking at these examples we have found that unitary or nearly unitary errors appear to result in the largest difference between average-case and worst-case errors. This is true in general. For unitary errors in a d-dimensional space we find the following inequalities:

$$\sqrt{\frac{d+1}{d}}\sqrt{r_U} \leq D_U \leq \sqrt{(d+1)d}\sqrt{r_U}.$$

Thus, any unitary error has a worst-case error scaling like $\sqrt{r_U}$.

A general inequality.—For a large and important class of noise processes, the worst-case error can be directly estimated from benchmarking-type data without side information about the type of error, which generally requires doing full quantum process tomography [42], or one of its SPAM-resistant variants [43,44]. In Ref. [45] a quantity called the unitarity $u(\mathcal{E})$ of a noise process $\mathcal{E}$ was defined (see the Supplemental Material [20] for a precise definition), and it was shown that this can be estimated efficiently and accurately using benchmarking. We find that for all unital noise (i.e., noise where the maximally mixed state is a fixed point) with no leakage, the unitarity and the average error rate together give a characterization of the worst-case error via the inequality [46]

$$c_d\sqrt{u + \frac{2dr}{d-1} - 1} \leq D \leq d^2 c_d\sqrt{u + \frac{2dr}{d-1} - 1}, \quad (6)$$

where $c_d = \frac{1}{2}(1 - 1/d^2)^{1/2}$. Since the unitarity generally obeys the inequality $u \geq [1 - dr/(d-1)]^2$ (see Ref. [45]) we find (for unital noise without leakage) that the worst-case error scaling matches the average case if and only if $u = 1 - 2dr/(d-1) + O(r^2)$.

To illustrate the power of inequality (6), we immediately find that for the single-qubit calibration and dephasing noise model, the condition $1 - u_{\mathrm{CD}} = 4r_{\mathrm{CD}} + O(r_{\mathrm{CD}}^2)$ is both necessary and sufficient to recover the favorable linear scaling between the worst- and average-case errors. In fact, the worst-case error for this channel can be expressed directly in terms of the unitarity as $D_{\mathrm{CD}} = \sqrt{\frac{3}{2}r_{\mathrm{CD}} - \frac{3}{8}(1 - u_{\mathrm{CD}})}$. Also, because the unitarity can be estimated from a benchmarking-type experiment, this gives direct experimental access to worst-case errors for this family of noise models without the need for expensive tomographic methods.

Moreover, inequality (6) allows us to obtain insights into generalizing our conclusions for single-qubit models to few-qubit systems such as those required for entangling quantum gates. A natural generalization of our CD model to 2-qubit calibration and dephasing errors would be an independent dephasing rate p on each qubit and unitary noise given by $e^{iH_{\mathrm{CD2}}}$ where $H_{\mathrm{CD2}} = \delta_1\sigma_z^{(1)} + \delta_2\sigma_z^{(2)} + \epsilon\sigma_z^{(1)}\sigma_z^{(2)}$. The semidefinite programming approach is possible here, but becomes unwieldy because there are so many free parameters. However, both the average error rate and the unitarity are readily computed as in the Supplemental Material [20]. Inequality (6) then allows one to easily and generally explore the tradeoffs in the calibration accuracy of the δ and ϵ parameters such that the overall error remains roughly consistent between the average and worst cases. Furthermore, since u_{CD2} can be measured efficiently in a benchmarking experiment, large values of u can be used to herald that an experiment has left the favorable scaling regime and more characterization and calibration must be done.

Conclusion and outlook.—We have seen that many realistic noise processes admit a linear relation between the average error rate (which is accessible experimentally) and the worst-case error (which is the relevant figure of merit for fault tolerance). The exceptions to this rule are highly coherent errors, where the worst-case error scales proportionally to the square root of the average error rate.

While our methods and results are very general, there are noise sources that we have not tried to fit into our error taxonomy. However, errors such as cross talk [48] and time-dependent or non-Markovian noise [49,50] should be amenable to these methods, and extending our results to cover such noise is an important avenue for future work.

Finally, we reiterate that it is an interesting open question whether it is possible to prove a fault-tolerance threshold result directly in terms of r without the lossy conversion to D. Fault-tolerant circuits are not perfectly coherent because measuring error syndromes necessarily removes certain coherences, and this may provide an avenue to develop stronger theorems.

We thank R. Blume-Kohout, D. Gottesman, J. Gambetta, C. Granade, and J. Wallman for discussions. This work was supported by the Australian Research Council via Excellence for Engineered Quantum Systems Project No. CE11001013, by Intelligence Advanced Research Projects Activity via the Multi-Qubit Coherent Operations program, and by the U.S. Army Research Office Grants No. W911NF-14-1-0098 and No. W911NF-14-1-0103. S. T. F. acknowledges support from an ARC Future Fellowship (Fellowship No. FT130101744).

[1] A. Y. Kitaev, Russ. Math. Surv. **52**, 1191 (1997).
[2] D. Aharonov and M. Ben-Or, in *29th ACM Symposium on Theory of Computing (STOC)* (Associated for Computational Machinery (ACM), New York, 1997), pp. 176–188.
[3] E. Knill, R. Laflamme, and W. H. Zurek, Proc. R. Soc. A **454**, 365 (1998).
[4] J. Emerson, R. Alicki, and K. Życzkowski, J. Opt. B **7**, S347 (2005).
[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, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 080505 (2012).
[7] C. Granade, C. Ferrie, and D. G. Cory, New J. Phys. **17**, 013042 (2015).
[8] 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).
[9] 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).
[10] J. Gaebler, A. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. Jost, J. Home, E. Knill, D. Leibfried *et al.*, Phys. Rev. Lett. **108**, 260503 (2012).
[11] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 060501 (2012).
[12] 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 (2013).
[13] J. M. Chow, J. M. Gambetta, E. Magesan, D. W. Abraham, A. W. Cross, B. Johnson, N. A. Masluk, C. A. Ryan, J. A. Smolin, S. J. Srinivasan *et al.*, Nat. Commun. **5**, 4015 (2014).
[14] R. Barends *et al.*, Nature (London) **508**, 500 (2014).
[15] T. P. Harty, D. T. C. Allcock, C. J. Ballance, L. Guidoni, H. A. Janacek, N. M. Linke, D. N. Stacey, and D. M. Lucas, Phys. Rev. Lett. **113**, 220501 (2014).
[16] J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, J. Phys. Condens. Matter **27**, 154205 (2015).
[17] T. Xia, M. Lichtman, K. Maller, A. W. Carr, M. J. Piotrowicz, L. Isenhower, and M. Saffman, Phys. Rev. Lett. **114**, 100503 (2015).
[18] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, New J. Phys. **18**, 012002 (2015).
[19] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012).
[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.170502, which includes Refs. [21–25], for further technical details and, in particular, Eq. (20) for a precise definition of the unitarity.
[21] J. Watrous, CS 766 Theory of Quantum Information, https://cs.uwaterloo.ca/~watrous/LectureNotes.html, 2011.
[22] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004).
[23] L. Vandenberghe and S. Boyd, SIAM Rev. **38**, 49 (1996).
[24] A. Barvinok, *A Course in Convexity*, Graduate Studies in Mathematics Vol. 54 (American Mathematical Society, Providence, 2002).
[25] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, arXiv:1511.01513.
[26] P. Aliferis, D. Gottesman, and J. Preskill, Quantum Inf. Comput. **6**, 97 (2006).
[27] P. Aliferis and A. W. Cross, Phys. Rev. Lett. **98**, 220502 (2007).
[28] P. Aliferis, F. Brito, D. P. DiVincenzo, J. Preskill, M. Steffen, and B. M. Terhal, New J. Phys. **11**, 013061 (2009).
[29] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, Phys. Rev. A **83**, 020302 (2011).
[30] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Phys. Rev. A **71**, 062310 (2005).
[31] C. W. Helstrom, Inf. Control **10**, 254 (1967).
[32] J. Watrous, Theory Comput. **5**, 217 (2009).
[33] A. Ben-Aroya and A. Ta-Shma, Quantum Inf. Comput. **10**, 77 (2010).
[34] J. Watrous, Chicago J. Theor. Comp. Sci. **2013**, 8 (2013).
[35] J. J. Wallman and S. T. Flammia, New J. Phys. **16**, 103032 (2014).
[36] M. A. Nielsen, Phys. Lett. A **303**, 249 (2002).
[37] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

[38] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2000).
[39] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, Phys. Rev. A **89**, 062321 (2014).
[40] J. J. Wallman, M. Barnhill, and J. Emerson, Phys. Rev. Lett. **115**, 060501 (2015).
[41] P. Aliferis and B. M. Terhal, Quantum Inf. Comput. **7**, 139 (2007).
[42] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).
[43] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Phys. Rev. X **4**, 011050 (2014).
[44] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, arXiv:1310.4492.
[45] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, New J. Phys. **17**, 113020 (2015).
[46] A similar bound was recently derived independently by Wallman [47].
[47] J. J. Wallman, arXiv:1511.00727.
[48] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 240504 (2012).
[49] M. A. Fogarty, M. Veldhorst, R. Harper, C. H. Yang, S. D. Bartlett, S. T. Flammia, and A. S. Dzurak, Phys. Rev. A **92**, 022326 (2015).
[50] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, Phys. Rev. A **93**, 022303 (2016).