Experimental Demonstration of a Cheap and Accurate Phase Estimation

Kenneth Rudinger, Shelby Kimmel, Daniel Lobser, and Peter Maunz

Phys. Rev. Lett. **118**, 190502 — Published 11 May 2017

DOI: 10.1103/PhysRevLett.118.190502

Experimental demonstration of cheap and accurate phase estimation

Kenneth Rudinger,[1] Shelby Kimmel,[2] Daniel Lobser,[3] and Peter Maunz[3]

[1]*Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185,* kmrudin@sandia.gov
[2]*Joint Center for Quantum Information and Computer Science (QuICS), University of Maryland,* shelbyk@umd.edu
[3]*Sandia National Laboratories, Albuquerque, NM 87185*

We demonstrate experimental implementation of robust phase estimation (RPE) to learn the phase of a single-qubit rotation on a trapped Yb^+ ion qubit. We show this phase can be estimated with uncertainty below $4 \cdot 10^{-4}$ radians using as few as 176 total experimental samples, and our estimates exhibit Heisenberg scaling. Unlike standard phase estimation protocols, RPE neither assumes perfect state preparation and measurement, nor requires access to ancillae. We cross-validate the results of RPE with the more resource-intensive protocol of gate set tomography.

INTRODUCTION

As quantum computers grow in size, efficient and accurate methods for calibrating quantum operations are increasingly important [1–4]. Calibration involves estimating the values of experimentally tunable parameters of a quantum operation and, if incorrect, altering the controls to fix the error.

When these tunable parameters are incorrectly set, it causes the system to experience coherent errors. Coherent errors (versus incoherent errors) are more challenging for error correcting codes to correct [5, 6], making it harder to reach fault-tolerant thresholds [7–9]. Hence it is important to correct these types of errors in order to build a scalable quantum computer. While recent techniques using randomized compiling [10] mitigate the effects of coherent errors, removing as much of the coherent errors as possible still gives the best error rates.

Calibration can be challenging without accurate state preparation and measurement (SPAM) estimates [11, 12]. Thus proper calibration of quantum operations will require *robust* protocols, that is, protocols that can accurately characterize gate parameters without highly accurate initial knowledge of SPAM.

A new technique for calibrating the phases of gate operations is robust phase estimation (RPE) [13]. RPE can be used to estimate the rotation axes and angles of single-qubit unitaries. Moreover, it is easy to implement (the sequences required are essentially Rabi/Ramsey experiments), simple and fast to analyze, and obtains accurate estimates with surprisingly small amounts of data.

RPE has advantages over standard robust characterization procedures when it comes to the task of calibration. RPE can estimate specific parameters of coherent errors, whereas randomized benchmarking, while robust, can only estimate the magnitude of errors [14–18]. While compressed sensing approaches can withstand SPAM errors [19, 20], they do not have the Heisenberg scaling RPE achieves. There is a simple analytic bound on the size of SPAM errors that RPE can tolerate (namely less than $1/\sqrt{8}$ in trace distance), unlike the robust Bayesian approach of Wiebe et al., whose error tolerance is less well-understood. [21]. Lastly, RPE is extremely efficient compared to robust protocols that provide complete reconstructions of error maps, like randomized benchmarking tomography [22] and gate set tomography (GST) [23].

Like many other phase estimation procedures, RPE achieves Heisenberg scaling [13]; that is, the estimate error scales inversely with the number of times the quantum operation in question is applied. However, unlike many other protocols, it requires no entanglement such as squeezed states or NOON states [24–30], requires no ancillae [25, 31, 32], and is non-adaptive [33–36].

Finally, compared to many tomography and parameter estimation protocols, the post-experiment analysis of RPE is strikingly simple. There are no Bayesian updates [21, 36, 37], no optimizations [19, 23, 38], and no fits to decaying exponentials [16, 22]. Instead, post-processing can be represented with a dozen lines of pseudo-code, with the most complex operation being an arctangent [39].

Here, we provide the first published experimental demonstration of RPE and investigate its performance. We use RPE to experimentally extract the phase (rotation angle) of a single-qubit unitary. Because we don't know the true values of the parameter, we benchmark the estimate by comparing to GST, which gives a robust, accurate, and reliable estimate, but which requires much more data [23].

We see experimental evidence of Heisenberg scaling in RPE, and we attain an accuracy of $3.9 \cdot 10^{-4}$ radians in our phase estimate using only 176 total samples. Compared to GST, we find that RPE requires orders of magnitude fewer total gates and samples to achieve similar accuracies. However, when experiments involving long sequences are not accessible, we find GST potentially has better performance than RPE. Nonetheless, due to its minimal data requirements, ease of implementation and analysis, and robust estimates of coherent errors, RPE is a powerful tool for efficient calibration of quantum operations.

PRELIMINARIES

We consider estimating the parameter α from the following single-qubit gate (the "X gate"):

$$\hat{X}_{\pi/2+\alpha} = \exp\left[-i\left(\left(\pi/2+\alpha\right)/2\right)\hat{\sigma}_X\right], \tag{1}$$

where $\hat{\sigma}_X$ is the Pauli X operator and α is the rotation error in the X gate. There is no off-axis component to the X gate, as we choose the X axis of the Bloch sphere to be the rotation axis of the X gate [46]. α is a parameter that experimentalists can typically control with ease. (While we do not do so here, RPE may also be used to estimate the rotation angle and and axis misalignment of a rotation gate about an approximately orthogonal axis. [47])

In reality, implemented quantum gates will not be unitary, but instead will be completely positive trace preserving (CPTP) maps. Nonetheless, these CPTP maps will have rotation angles with errors analogous to α, and in the Supplemental Material, we show RPE can extract such angles [39]. For the rest of the paper, with slight abuse of notation, we will use α to refer to this more general CPTP map rotation error.

We use RPE and GST to extract α. Fig. 1 gives a schematic description of GST and RPE circuits. RPE circuits are essentially Rabi/Ramsey sequences; they consist of state preparation ρ, assumed to be not too far in trace distance from $|0\rangle\langle 0|$, followed by repeated applications of the X gate, followed by a measurement operator M, which is assumed to be close in trace distance to $|1\rangle\langle 1|$.

We use "additive error" to denote the maximum bias in the outcome probability of any single RPE experimental sequence. This bias can be due to SPAM errors and incoherent errors in the gates. RPE can tolerate additive error as long as it is less than $1/\sqrt{8}$.

For GST, each sequence consists of a state preparation ρ, followed by a gate sequence F_i to simulate an alternate state preparation. Next a gate sequence g_k is applied repeatedly. Finally, the measurement M is preceded by a gate sequence F_j to simulate an alternative measurement. We refer to F_i and F_j as state and measurement fiducials, respectively, and g_k as a germ [39].

For both RPE and GST, running increasingly longer sequences produces increasingly accurate estimates. We use L to parameterize the length of the sequence, as in Fig. 1. We run sequences with $L \in \{1, 2, 4, 8, \ldots, L_{\max}\}$, where $L_{\max}$ is chosen based on the desired accuracy. In RPE, we repeat the gate of interest either L or $L+1$ times. In GST, we implement all possible combinations of state fiducials, measurement fiducials, and germs, with the germ repeated $\lfloor L/|g_k| \rfloor$ times, where $|g_k|$ is the number of gates in g_k and $\lfloor\cdot\rfloor$ denotes the floor function.

We employ the following notation to keep track of experimental resources. N denotes the number of repetitions (samples taken) of each sequence, and is set to be the same for all sequences in a single RPE or GST experimental run [48]. The total number of unique sequences is a function of the maximum sequence length $L_{\max}$ and is denoted by Q. For RPE, it is given by

$$Q(L_{\max}) = 2\left(1 + \log_2 L_{\max}\right). \tag{2}$$

For GST, $Q(L_{\max})$ is approximately equal to $396\left(1+\log_2 L_{\max}\right)$; see [23] for details. The total number of experimental samples taken is a function of both N and $L_{\max}$ and is denoted by S:

$$S(N, L_{\max}) = N \cdot Q(L_{\max}). \tag{3}$$

(We will sometimes drop the arguments from Q and S if clear from context.)

RPE successively restricts the possible range of the estimated phase using data from sequences with larger and larger L. Inaccuracies result when the procedure restricts to the wrong range. For larger values of $L_{\max}$, there are more rounds of restricting the range, and thus more opportunities for failure. By increasing N when $L_{\max}$ increases, we can limit this probability of failure. Likewise, a large additive error makes it easier to incorrectly restrict the range, but again, taking larger N can increase the probability of success. The interaction between accuracy, N, $L_{\max}$, and additive errors is shown in Fig. 2. This graph shows theoretical upper bounds of the root mean squared error (RMSE) of RPE for fixed sample size N [49]. Fig. 2 shows that, given an additive error δ, there exist good choices for N and $L_{\max}$, provided that $\delta < 1/\sqrt{8}$.

A protocol has Heisenberg scaling when the RMSE of its estimate of a gate parameter scales inversely with the number of applications of a gate. RPE provably has Heisenberg scaling [13], and GST numerically exhibits Heisenberg-like scaling [23]. In this paper, we empirically look for scaling in accuracy and precision that scales as $1/L_{\max}$. This is a good proxy (up to log factors) for Heisenberg scaling.

In practice, experimentalists care less about Heisenberg scaling than about the resources (e.g. N and $L_{\max}$) required to achieve a desired accuracy. Assuming time is the key resource, if experimental reset time is long compared to gate time, S quantifies the experimental cost. On the other hand, if gate time is long compared to experimental reset time, then $L_{\max}$ is the dominant factor.

EXPERIMENTAL RESULTS

We implement GST and RPE on a single $^{171}\mathrm{Yb}^+$ ion in a linear surface ion trap. The qubit levels are the hyperfine clock states of the $^2S_{1/2}$ ground state: $|0\rangle = |F=0, m_F=0\rangle, |1\rangle = |F=1, m_F=0\rangle$. We initialize the qubit close to the $|0\rangle$ state via Doppler cooling and optical pumping; we measure in the computational basis

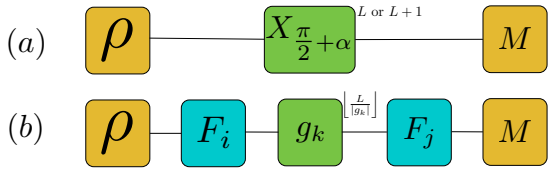

FIG. 1: (Color online) (a) RPE and (b) GST experimental sequences. Each sequence starts with the state ρ and ends with the two-outcome measurement M. (a) An RPE sequence consists of repeating the gate in question either L or $L+1$ times. (b) In GST, a gate sequence F_i is applied to simulate a state preparation potentially different from ρ. This is followed by $\lfloor L/|g_k| \rfloor$ applications of a *germ*—a short gate sequence g_k of length $|g_k|$. Finally, a sequence F_j is applied to simulate a measurement potentially different from M.

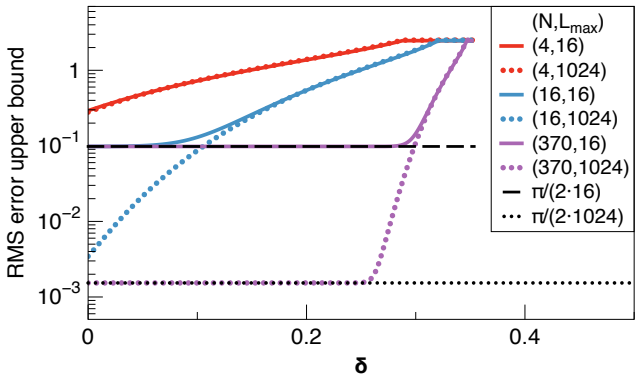

FIG. 2: (Color online) Analytic upper bounds on the RMSE of the RPE phase estimate. Because RPE is potentially biased, the RMSE does not go to zero in the limit of infinite N, but instead, approaches a floor of $\pi/(2L_{\max})$. Larger additive error δ produces a larger bias, and thus require larger N and larger $L_{\max}$ to achieve a small RMSE. For example, $N = 16$ is not large enough to reach the floor for $L_{\max} = 1024$, but increasing N to 370 we easily saturate the bound for most values of δ.

(approximately) via fluorescence state detection [40]. See [23] for experimental details. We used the open-source GST software pyGSTi for numerical analysis, extending its capabilities to include RPE functionality [41].

We take $N = 370$ samples of each GST and RPE sequence, and use $L \in \{1, 2, 4, \ldots, 1024\}$. (For details, see Gate Sequences in [39].) For the GST dataset, this yields $Q = 2,347$ unique sequences and $S = 868,390$ total samples, while the RPE dataset has $Q = 22$ unique sequences and $S = 8,140$ samples.

Looking at Fig. 2, we see $N = 370$ is larger than necessary for RPE with $L_{\max} = 1024$ for additive error less than ~ 0.25. To simulate experiments with small N, we randomly sample (without replacement) from the experimental dataset, so that the new, subsampled dataset has $N << 370$ samples per sequence.

We use several methods to characterize the experimental accuracy of RPE. First, we apply the analytic bounds on RMSE of Fig. 2. We also compare our subsampled RPE estimates to the GST estimate. Unlike RPE, GST is an unbiased estimator [42], so we expect that as we increase N, the RMSE will decrease as $1/\sqrt{N}$. Using the $N = 370$ dataset for GST, we estimate $\alpha = (6.4 \pm 4.9) \cdot 10^{-5}$; the error bars denote a 95% confidence interval derived using a Hessian-based procedure (see [23] for details). On the other hand, using all RPE data we estimate $\alpha = 1.0 \cdot 10^{-4}$, with an RMSE upper bound of $\pi/(2 \cdot L_{\max}) \approx 1.5 \cdot 10^{-3}$ (where this bound comes from Fig 2 with $N = 370$, assuming our additive error is less than 0.25; this assumption is borne out in the next section).

While the RPE estimate is consistent with the GST result, the accuracy is significantly lower, and we thus take α_0, the full data estimate from GST, to be the "true" value of α for the purposes of benchmarking RPE. In particular, throughout this paper, we calculate experimental RMSE by comparing the mean estimate from 100 subsampled datasets to α_0.

Heisenberg Scaling from RPE

To look for Heisenberg scaling in RPE estimates, we perform RPE on 100 subsampled datasets for $L_{\max} \in \{1, 2, 4, \ldots, 1024\}$ with $N = \{8, 32\}$. We see Heisenberg-like scaling in the experimental RMSE in Fig. 3. We also plot $\pi/(2L_{\max})$, which is the analytic upper bound if sufficient samples are taken to compensate for additive error. We see that in practice, the analytic bounds can be pessimistic. Moreover, we see that while the experimental RPE accuracy *is* sensitive to N, increasing N to 32 from 8 does not dramatically improve the RMSE, improving the scaling to $0.200/L_{\max}$ from $0.331/L_{\max}$. Instead, as expected, large increases in accuracy are obtained by moving to larger $L_{\max}$. This Heisenberg-like scaling is especially important for regimes where the time to implement the gate sequence is long relative to SPAM time.

We believe our experimentally derived bounds are significantly better than our analytic bounds in part because our system is well-calibrated. The analytic bounds give a worst-case analysis that accounts for bias caused by adversarial additive error, but RPE is effectively unbiased for our system, up to the accuracy we achieve.

Comparison to GST

Because RPE can be biased, increasing N cannot improve the RMSE below $\pi/(2L_{\max})$ in the worst case (see Fig. 2 and [13]). However since GST is unbiased, it always benefits from increasing N.

We investigate this effect in Fig. 4. We plot the RMSE for experiments with fixed $L_{\max} = 1024$ and varying $N \in \{8, 16, 32\}$. Analytic bounds for RPE are derived using the same method as in Fig 2. Experimental bounds for GST and RPE are derived from comparing the estimates of 100 subsampled datasets to α_0.

While the analytic RPE bounds do not improve with increasing N, the subsampled RPE and GST datasets show standard quantum limit scaling. We expect this for GST, because GST is unbiased. In the case of RPE our experimental system happens to have very small additive error, and so is only very slightly biased. In this case, we expect to see improving estimates with increasing N until our accuracy is about the same size as our bias. Fig. 4 tells us that for systems with relatively large additive error, where large N is feasible but large $L_{\max}$ is not, GST can provide more accurate results.

However, we see in Fig. 4 that GST pays a substantial cost relative to RPE when S is the figure of merit. In Fig. 5, we compare the values of S that RPE and GST each require to achieve a desired accuracy, by analyzing 100 subsampled datasets with fixed $N = 8$ and varying $L_{\max} \in \{1, 2, 4, \ldots, 1024\}$. We see that RPE can achieve similar accuracy to GST while using orders of magnitude fewer total samples. As demonstrated both in Figs. 4 and 5, choosing $N = 8$ and $L_{\max} = 1024$ yields an RPE estimate of α with an RMSE of $3.9 \cdot 10^{-4}$; this costs only $S(8, 1024) = 176$ total experimental samples.

For our system, acquiring the entire RPE and GST datasets took 10.8 minutes and 12.1 hours, respectively, and total experimental time scales linearly with N. Thus we note that had our actual data acquisition rate been $N = 8$, it would have taken 14 s to acquire that RPE dataset and about 15.5 minutes to acquire the GST dataset. As for analysis time, a single RPE dataset can be analyzed in about 0.05 s on a modern laptop. GST analysis takes about 20 s [50]. All datasets and analysis notebooks are available online [43].

CONCLUSIONS

We show that robust phase estimation successfully estimates the phases of single-qubit gates, yielding results that are consistent with the full tomographic reconstruction of gate set tomography, and also exhibits Heisenberg-like scaling in accuracy. In particular, an individual phase may be estimated with a root mean squared error of $3.9 \cdot 10^{-4}$ with as few as 176 total experimental samples.

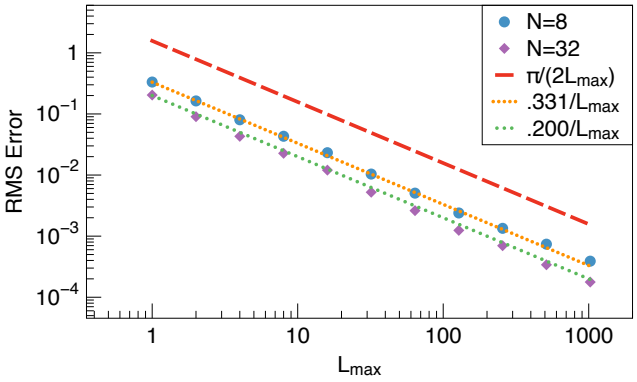

FIG. 3: (Color online) RMSE versus $L_{\max}$ for RPE estimates of α from 100 subsampled datasets of size $N = 8$ and $N = 32$. While analytic bounds are at best $\pi/(2L_{\max})$, we see this can be pessimistic. When the additive errors, which can bias the RPE estimate, are sufficiently small, increasing N improves RMS accuracy.

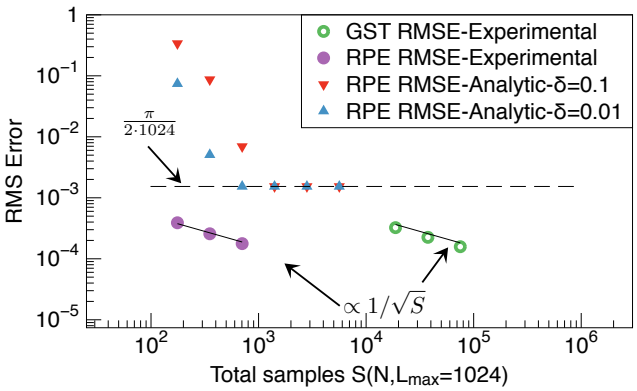

FIG. 4: (Color online) Scaling of RMSE of estimates of α as a function of total samples $S(N, L_{\max})$, with $L_{\max} = 1024$. For the analytic curves, we resample with $N \in \{8, 16, \ldots, 256\}$, while for the experimental curves we restrict ourselves to $N \in \{8, 16, 32\}$ (as higher resampling rates would introduce nontrivial correlations between datasets). For each curve, N increases incrementally from left to right. Analytic bounds are derived using the techniques of Fig. 2. Experimental data points take the RMSE of 100 subsampled datasets for both RPE and GST. While the analytic bounds converge to $\pi/2048$, we see standard quantum limit scaling (i.e., error scaling $\propto 1/\sqrt{S}$) of RPE experimental estimates. As discussed in the text, this is because our experimental device has very low additive error, and thus the RPE estimates are essentially unbiased, and can achieve greater accuracy with increasing number of samples. GST estimates also exhibit standard quantum limit scaling.

Hence, RPE is a strong choice for diagnosing and cal-

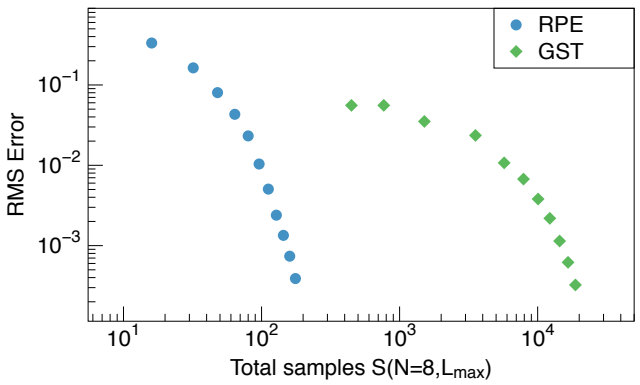

FIG. 5: (Color online) RMSE for RPE and GST estimates of α as function of total number of total samples $S(N, L_{\max})$ using 100 subsampled datasets. As opposed to Fig. 4, here we fix $N = 8$ and vary $L_{\max}$. Each sequential data point corresponds to setting $L_{\max} \in \{1, 2, 3, \ldots, 1024\}$. RPE achieves the same level of accuracy as GST using far fewer resources.

ibrating single-qubit operations. It would be interesting to investigate whether the techniques of RPE can be applied to assessing other errors in single-qubit gate operations in a fast and accurate manner.

ACKNOWLEDGEMENTS

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SK is funded by the Department of Defense. The authors thank Robin Blume-Kohout and Nathan Wiebe for helpful conversations, and Erik Nielsen for extensive software support. This research was funded, in part, by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government.

[1] D. J. Egger and F. K. Wilhelm, Phys. Rev. Lett. **112**, 240503 (2014).
[2] C. Ferrie and O. Moussa, Phys. Rev. A **91**, 052306 (2015).
[3] K. R. Brown, A. C. Wilson, Y. Colombe, C. Ospelkaus, A. M. Meier, E. Knill, D. Leibfried, and D. J. Wineland, Phys. Rev. A **84**, 030303 (2011).
[4] Z. Hou, H.-S. Zhong, Y. Tian, D. Dong, B. Qi, L. Li, Y. Wang, F. Nori, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, New Journal of Physics **18**, 083036 (2016).
[5] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, New Journal of Physics **18**, 012002 (2016).
[6] M. Gutiérrez, C. Smith, L. Lulushi, S. Janardan, and K. R. Brown, Phys. Rev. A **94**, 042338 (2016).
[7] A. Y. Kitaev, Russian Mathematical Surveys **52**, 1191 (1997).
[8] D. Aharonov and M. Ben-Or, in *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97 (ACM, New York, NY, USA, 1997) pp. 176–188.
[9] D. Aharonov and M. Ben-Or, SIAM Journal on Computing **38**, 1207 (2008).
[10] J. J. Wallman and J. Emerson, Phys. Rev. A **94**, 052325 (2016).
[11] C. Stark, Phys. Rev. A **89**, 052109 (2014).
[12] C. Stark, ArXiv e-prints (2015), arXiv:1510.02800 [quant-ph] .
[13] S. Kimmel, G. H. Low, and T. J. Yoder, Phys. Rev. A **92**, 062315 (2015).
[14] 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).
[15] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. A **85**, 042311 (2012).
[16] E. Magesan, J. M. Gambetta, and J. Emerson, Phys. Rev. Lett. **106**, 180504 (2011).
[17] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, New Journal of Physics **17**, 113020 (2015).
[18] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Phys. Rev. A **93**, 012301 (2016).
[19] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, Phys. Rev. Lett. **106**, 100401 (2011).
[20] E. Magesan, A. Cooper, and P. Cappellaro, Phys. Rev. A **88**, 062109 (2013).
[21] N. Wiebe, C. Granade, C. Ferrie, and D. G. Cory, Phys. Rev. Lett. **112**, 190501 (2014).
[22] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Phys. Rev. X **4**, 011050 (2014).
[23] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Nature Communications **8**, EP (2017).
[24] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[25] A. Y. Kitaev, eprint arXiv:quant-ph/9511026 (1995), quant-ph/9511026 .
[26] P. Kok, S. L. Braunstein, and J. P. Dowling, Journal of Optics B: Quantum and Semiclassical Optics **6**, S811 (2004).
[27] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **96**, 010401 (2006).
[28] V. Giovannetti, S. Lloyd, and L. Maccone, **306**, 1330 (2004).
[29] V. Meyer, M. A. Rowe, D. Kielpinski, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, Phys. Rev. Lett. **86**, 5870 (2001).
[30] H. Lee, P. Kok, and J. P. Dowling, Journal of Modern Optics **49**, 2325 (2002).
[31] S. Boixo, S. T. Flammia, C. M. Caves, and J. M. Geremia, Phys. Rev. Lett. **98**, 090401 (2007).
[32] E. M. Kessler, I. Lovchinsky, A. O. Sushkov, and M. D. Lukin, Phys. Rev. Lett. **112**, 150802 (2014).
[33] H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995).

[34] A. Sergeevich, A. Chandran, J. Combes, S. D. Bartlett, and H. M. Wiseman, Phys. Rev. A **84**, 052315 (2011).

[35] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Nature **450**, 393 (2007).

[36] C. Ferrie, C. E. Granade, and D. G. Cory, Quantum Information Processing **12**, 611 (2013).

[37] N. Wiebe and C. Granade, Phys. Rev. Lett. **117**, 010503 (2016).

[38] K. M. Svore, M. B. Hastings, and M. Freedman, Quantum Info. Comput. **14**, 306 (2014).

[39] K. Rudinger, S. Kimmel, D. Lobser, and P. Maunz, "See supplemental material [url], which includes ref. [44]," (2017).

[40] S. Olmschenk, K. C. Younge, D. L. Moehring, D. N. Matsukevich, P. Maunz, and C. Monroe, Phys. Rev. A **76**, 052314 (2007).

[41] E. Nielsen, K. Rudinger, J. K. Gamble IV, and R. Blume-Kohout, "A python implementation of gate set tomography," (2016).

[42] R. Blume-Kohout, J. King Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, ArXiv e-prints (2013), arXiv:1310.4492 [quant-ph] .

[43] K. Rudinger, S. Kimmel, D. Lobser, and P. Maunz, "https://github.com/pygstio/supplemental-info-rpe-1," (2017).

[44] C. King and M. B. Ruskai, IEEE Trans. Inf. Theor. **47**, 192 (2001).

[45] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, New Journal of Physics **11**, 073023 (2009).

[46] This means that state preparation and measurement may not be aligned with Z axis, which contributes to additive error.

[47] More precisely, given such a gate, e.g., $\hat{Y}_{\pi/2+\epsilon}(\theta) = \exp\left[-i\left(\left(\pi/2+\epsilon\right)/2\right)\left(\cos\theta\hat{\sigma}_Y + \sin\theta\hat{\sigma}_X\right)\right]$, RPE can extract both the rotation angle error parameter ϵ and the axis misalignment parameter θ. The gate sequences used to extract ϵ are similar to those used to extract α, while slightly more complicated gate sequences are used to learn θ; see [13].

[48] Although this results in slightly non-ideal scaling in the accuracy of our estimate [45], this is a realistic scenario for experimental implementation.

[49] We use Equations 5.7, 5.8, 5.9, and 5.16 from [13] to create the graph.

[50] All analyses performed on a 2014 MacBook Pro 2.5 GHz Intel Core i7 machine.