Testing the robustness of robust phase estimation

Adam M. Meier, Karl A. Burkhardt, Brian J. McMahon, and Creston D. Herold

Testing the Robustness of Robust Phase Estimation

Adam M. Meier, Karl A. Burkhardt, Brian J. McMahon, and Creston D. Herold*

Georgia Tech Research Institute, Atlanta, GA 30332, USA

(Dated: October 8, 2019)

The Robust Phase Estimation (RPE) protocol was designed to be an efficient and robust way to calibrate quantum operations. The robustness of RPE refers to its ability to estimate a single parameter, usually gate amplitude, even when other parameters are poorly calibrated or when the gate experiences significant errors. Here we demonstrate the robustness of RPE to errors that affect initialization, measurement, and gates. In each case, the error threshold at which RPE begins to fail matches quantitatively with theoretical bounds. We conclude that RPE is an effective and reliable tool for calibration of one-qubit rotations and that it is particularly useful for automated calibration routines and sensor tasks.

I. INTRODUCTION

The standard, circuit model of quantum computation requires discrete, unitary operations (gates). However, physical implementation of these gates is not naturally discrete. For one-qubit gates it is useful to picture a unitary operation as a specific rotation of the Bloch sphere, with an axis and an angle of rotation that can be parametrized by three real numbers. Careful calibration of these parameters is necessary in order to compute accurately using the gate [1–3].

When a single unitary operation is concatenated into a sequence of several identical operations, the resulting angle of rotation is linear in the number gates. Measurements of the quantum state after such a sequence can cause small changes in the angle of the individual gate to appear quadratically as changes in the resulting projection probabilities. As a result, the error of a computation can scale as the square both of the angular error on each operation and of the number of operations. Because of this scaling, small errors in calibrating operation rotation angles using simple calibration experiments can still lead to significant error in the context of longer gate sequences.

Typical experimental methods for estimating gate angles are variations of "Rabi flopping" and Ramsey experiments. In both cases, the qubit is prepared in a known state, an interaction that drives rotation is implemented for a variable duration, and then the system is measured. The projection onto the two qubit states ideally follows a sinusoid as a function of the duration. Fitting the data to this model provides the angular frequency of the sinusoid, which is is called the "Rabi frequency" in the case of Rabi flopping. Multiplying this frequency by the target gate duration provides the gate angle of rotation. These procedures are convenient when their purpose is to pick a gate duration that implements a fixed angle of rotation with little prior knowledge; however, they are not optimal for analyzing a fixed gate. Other common gate characterization methods such as quantum process tomography (QPT) [4], gate set tomography (GST) [5], and their variants are designed to fully characterize one or more fixed gates, and they are not efficient for estimating a single parameter like the gate angle.

In its original form, the Robust Phase Estimation (RPE) protocol was proposed as a practical method for estimating the angles and axes of single qubit rotations. The protocol was initially developed by Kimmel *et al.* [6] and tested in experiments by Rudinger *et al.* [7]. The robustness of RPE refers to its ability to estimate these parameters even when other parameters of the experiment are poorly calibrated or prone to stochastic errors. In this paper we describe three separate experiments that demonstrate the robustness of RPE to injected errors chosen to represent common experimental error sources.

The RPE protocol consists of performing gate sequences of several different lengths. In each sequence the same gate is repeated, so that the final rotation angle compounds linearly. The gate sequences themselves are very similar to the standard Rabi flopping procedure. Two primary differences between Rabi flopping and RPE are: First, RPE repeats an operation with a fixed duration, instead of continuously varying the duration of a single operation. Second, the analysis of RPE iterates a simple estimation formula on data from each sequence length, while Rabi flopping uses curve-fitting of the entire data set. From a practical standpoint, RPE can determine the rotation angle of a fixed gate duration much more efficiently than Rabi flopping, where efficiency is measured by the number of gate applications required to achieve a given precision in the estimation of the angle.

RPE succeeds or fails in a binary fashion; the estimate it produces is either correct (within a pre-determined confidence region) or it is incorrect, sometimes by a large angle. The protocol can fail at each sequence length (independently) due to shot noise or to errors in the implementation of the quantum operations. When each sequence is repeated sufficiently many times to remove shot noise concerns and the operation errors are small enough, the protocol always succeeds in producing an estimate of the rotation angle within its guaranteed confidence region. The robustness of RPE is defined by this feature that the protocol reliably succeeds even when these errors are significant.

* Creston.Herold@gtri.gatech.edu

Rabi flopping is useful for diagnosing experimental problems; in this case protocol efficiency is less important, and an experimenter can visually identify and process unexpected deviations from the sinusoidal model. In contrast, the robustness and efficiency of RPE make it ideal for automated calibration of quantum systems or for automated sensor experiments where there is no human oversight. Ref. [7] has already examined the differences in cost and utility between RPE and GST and finds that RPE is significantly more efficient for single parameter calibrations.

The remainder of this paper is organized as follows. In Section II, we summarize our implementation of the RPE protocol. Our experimental apparatus is described in Section III. In Section IV, we describe the errors against which we test the robustness of RPE. The results of these tests follow in Section V. Finally, we comment in Section VI on the wide range of parameter estimation experiments which could benefit from RPE.

II. RPE PROTOCOL

In this section, we provide a practical description of the RPE protocol as applied in our experiments, where it is used to estimate a single rotation angle. Consider calibrating the rotation angle θ [8] of a qubit operation described by

$$Y_\theta = e^{-i(\theta/2)\sigma_Y} = \cos(\theta/2)\mathbb{I} - i\sin(\theta/2)\sigma_Y \qquad (1)$$

where $\mathbb{I}$ is the identity matrix and σ_Y is the Pauli Y-matrix.

We perform two sets of experiments, where the operation of interest is repeated $\{n\}$ times for various values of n, and the sets are differentiated by the initial qubit state. The experiments all terminate with a single, projective measurement. We repeat each experiment M_n times (subsequently referred to as "samples") and record the number of times we observe the $+1$ eigenstate. The observed number of the $+1$ results for the two sets of experiments are denoted by x_n and y_n, and their values in the limit of large M_n can be predicted using the following formulas:

$$x_n = M_n \left|\langle 1|Y_\theta^n|0\rangle\right|^2 \qquad y_n = M_n \left|\langle 1|Y_\theta^n|+\rangle\right|^2, \quad (2)$$

where the state $|+\rangle$ is equal to $(|0\rangle + |1\rangle)/\sqrt{2}$ and $\{|0\rangle, |1\rangle\}$ are the Z-eigenstates of the qubit. In our experiment we always initially prepare $|0\rangle$. The state $|+\rangle = Y_{\pi/2}|0\rangle$ is generated from the initial state with a $Y_{\pi/2}$ operation.

In order to account for non-idealities in the experiments, we follow the convention of Ref. [6] and write the expected probability of observing the $+1$ eigenstate in the experiments corresponding to x_n and y_n as

$$p_x(n) = \frac{1 - \cos(n\theta)}{2} \pm \delta_x(n)$$

$$p_y(n) = \frac{1 + \sin(n\theta)}{2} \pm \delta_y(n), \qquad (3)$$

where we have introduced additive error terms $\delta_{x,y}$ which may be n-dependent. When the $\delta_{x,y}$ terms are zero, we can estimate the total rotation $n\theta$, modulo 2π, by noting that

$$\tan(n\theta) = \frac{2p_y(n) - 1}{1 - 2p_x(n)}, \qquad (4)$$

and using x_n/M_n and y_n/M_n as estimates of $p_{x,y}(n)$ gives

$$n\theta = \tan^{-1}\left(\frac{y_n - M_n/2}{M_n/2 - x_n}\right), \qquad (5)$$

We note that the inability to discriminate $n\theta = 0$ from 2π can typically be remedied in practice, by repeating the protocol with a gate using the same physics but a smaller intended angle.

When the error terms $\delta_{x,y}$ are non-zero, our estimate of $n\theta$ can be biased. Additionally, if we collect too few samples, shot noise, which is proportional to $1/\sqrt{M_n}$, can lead to erroneous results. Models for the impact of shot noise and additive error on RPE are presented in Ref. [6].

The RPE protocol achieves high efficiency by specifying that only certain values of n (gate repetitions) need to be performed, namely $n = 2^{j-1}$ where $j \in \{1, 2, 3, \ldots, L\}$ and L is chosen as a function of the desired precision of the estimate. The jth measurement restricts the estimate of θ:

$$\theta - \theta_{\text{actual}} \in \left(-\frac{\pi}{2^j}, \frac{\pi}{2^j}\right]. \qquad (6)$$

Each successive doubling of the repetition number cuts the range of the angle estimate bound in half. The supplemental information to [7] provides pseudocode for determining θ including enforcement of the range restriction from preceding steps. Ref. [6] also provides the projected probability of protocol success conditioned only on statistical errors. Here success is defined to be an application of the protocol in which the true value of θ lies within the bounds output by the protocol. With the range restriction provided in Eq. (6), Kimmel et al. proved in [6] that RPE is robust in the presence of errors and will succeed provided the total additive error is bounded by $|\delta_{x,y}(n)| < \delta_{\text{bound}} \equiv 1/\sqrt{8}$ for all repetitions $\{n\}$. When RPE does fail, there is no warning as RPE does not have a built-in failure detection method.

In practice, it is likely to be difficult to estimate the additive error *a priori*. For example, without knowing the angle of the gate under inspection the experimenter will not know the expected final state of the length n sequence and cannot estimate the impact of asymmetric measurement errors. Even more difficult is the case of gate errors associated with colored noise, where no single rotation angle consistently describes the gate for every sequence length. Further study is needed to understand how RPE will perform in this case. Errors that
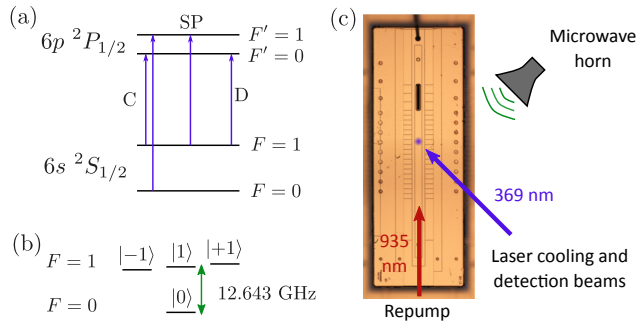
FIG. 1. (a) ^{171}Yb$^+$ energy level diagram. The required 369 nm laser transitions are shown for Doppler cooling (C), initial state preparation (SP), and detection (D). (b) State labels in $^2S_{1/2}$ hyperfine manifold. The qubit levels are $|0\rangle$ and $|1\rangle$. (c) System overview. Laser beam geometry and microwave horn for qubit operations are overlaid on an image of the trap chip.

are mutually uncorrelated, such as measurement errors or errors due to decay events, are simpler to bound as a function of sequence length, and we focus on them in this manuscript.

It is important to note that as additive error increases, more samples are required for RPE to succeed, and the required M_n diverges at the $1/\sqrt{8}$ bound, as described in [6]. A prescription for choosing M_n as a function of the predicted additive error and sequence length is provided therein. Furthermore, because each sequence length constrains the next estimate of θ, the probability that RPE fails is the sum of the probability of failing at each step.

In this paper, we demonstrate experimentally that, for three error sources that are expected to behave additively, the RPE protocol is as robust as promised. This is evaluated by observing the failure rate of RPE for the gate $Y_{\pi/2}$ as a function of injected (artificially introduced) error sources.

III. EXPERIMENTAL SYSTEM

All experiments described herein are performed on a single ^{171}Yb$^+$ ion trapped above a GTRI-Honeywell BGA trap [9]. The relevant portion of the atomic energy level diagram is shown in Figure 1a. Doppler cooling, state preparation, and state detection are all performed with the same 369 nm beam through electro-optic modulation, similar to [10]. This beam is switched on and off with an acousto-optic modulator (AOM). The qubit states are chosen from the ground-state hyperfine manifold (Figure 1b) to be $|0\rangle \equiv |F = 0, m_F = 0\rangle$ and $|1\rangle \equiv |F = 1, m_F = 0\rangle$. We label the auxiliary states $|\pm 1\rangle \equiv |F = 1, m_F = \pm 1\rangle$.

One-qubit rotations are driven with microwaves emitted from a horn outside the vacuum chamber as depicted in Figure 1c. We control the frequency, amplitude, and phase of the microwave drive by mixing an approximately 200 MHz DDS signal with a local oscil-

lator at 12.442 812 GHz. The upper sideband of this mixing is resonant with the transition from $|0\rangle$ to $|1\rangle$. The mixer output is amplified to a maximum of approximately 1.6 W, high enough to achieve a minimum $\pi/2$-rotation time of $T_{\pi/2} \approx 4$ μs. However, for all the reported RPE measurements, we reduce the amplitude and use $T_{\pi/2} = 10$ to 30 μs to reduce amplitude errors caused by the thermal duty cycle of the microwave amplifier chain.

While RPE can be used to calibrate any rotation, we choose $Y_{\pi/2}$ (i.e. $\theta = \pi/2$ in Eq. (1)) for our base gate. This means that we prepare $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ with the same gate we repeat for the RPE sequence. Prior to each measurement set, we use an RPE sequence with no intentionally injected errors to calibrate the $\pi/2$ operation for a fixed $T_{\pi/2}$ by adjusting the microwave amplitude. With a maximum sequence length of 2^7 gates and 128 repetitions of each sequence, RPE bounds the calibration at $\pm\pi/2^8$, or a fractional uncertainty of 0.8%. In practice, however, we observe much lower statistical uncertainty, similar to the behavior observed in [7], and repeated calibration produces the same microwave amplitude to within 0.04%. Following the full set of RPE trials for each type of error, we repeat the amplitude calibration and find that it is consistent to within 0.2%, or one quarter of the RPE estimate bound, at worst.

We initially calibrate the microwave frequency to within 1 Hz of the qubit resonance using Ramsey spectroscopy. Upon recalibrating after each data set, we find that the qubit frequency changed less than 2 Hz. Based on a maximum delay of 100 μs between pulses in our RPE sequences, this contributes at most an inconsequential 2×10^{-4} phase slip between pulses. With these stable, well-calibrated gates, we inject known errors with a significantly larger magnitude as described in the next section.

IV. INJECTING ADDITIVE ERRORS

A. Measurement error

We inject error in the measurement process by intentionally misinterpreting the results of our detection operation. The detection consists of counting photons scattered from the ion over a specified duration. When the ion is measured in the $|1\rangle$ state (bright), we collect on average 19 photons during the 400 us detection interval, and when the ion is measured in the $|0\rangle$ state (dark), we collect 0.1 photons. The experimental dark and bright measurement histograms (Figure 2) are nearly Poissonian, except for mixing of the two Poissonians caused by errors in preparation and by the low end tail observed in bright detection events due to off resonant scattering of the detection light [11]. The measurement error associated with these histograms (before error injection) is 1.2%.
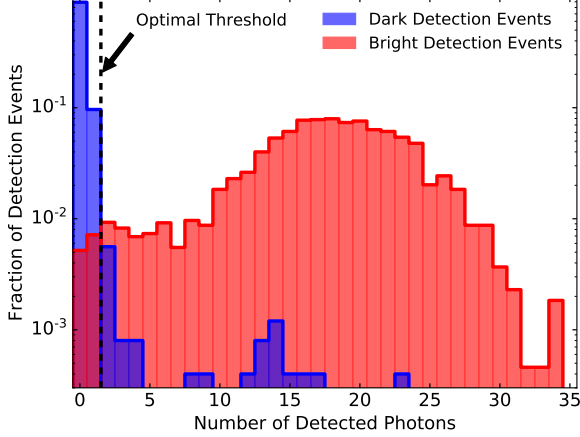
FIG. 2. Relative distribution of observed photon counts for putative $|0\rangle$ (dark) and $|1\rangle$ (bright) states. Each experiment is repeated 2500 times. The dashed line shows the optimal detection threshold of 2 photons. Larger detection errors are introduced by choosing a different threshold. At the optimal threshold, the measurement error is 1.2% (sum of the left-most two red bars).
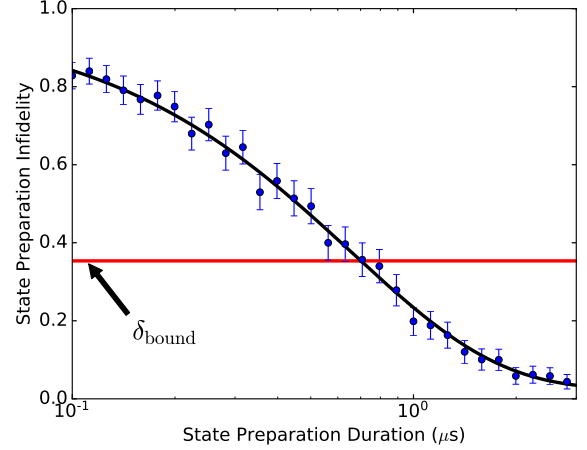

FIG. 3. State preparation error as a function of operation duration. We typically use a 50 μs optical pumping duration to ensure high fidelity initialization in $|0\rangle$; intentionally using a shorter duration injects error. Over the plotted range, the exponential fit to the infidelity E as a function of initialization duration t is $E(t) = 0.95e^{-1.5t} + 0.025$. At short duration, the infidelity limit is due to asymmetric state populations after Doppler cooling. At long duration, the apparent initialization infidelity is limited by measurement error. The $1/\sqrt{8}$ limit for additive error in RPE corresponds to a preparation time of about 700 ns.

To distinguish which state is present after a measurement, we define a threshold for the number of detected photons. If the detected number of photons is less than the chosen threshold, we infer that the ion was in the $|0\rangle$ state; otherwise we infer that it was in the $|1\rangle$ state. The optimal threshold in our experiment was two photons. By intentionally choosing a suboptimal threshold, we introduced an error source in post-processing that is easy to quantify.

In order to convert from a choice of threshold to an additive error estimate δ_{meas}, we must identify the maximum error over all possible sequence outcomes. From the measured reference histograms, we calculate the probability that the chosen threshold mis-identifies an observed photon count as bright (when the ion was dark) or dark (when the ion was bright). For a general sequence, we do not know which of the results to expect, so we take the maximum of these probabilities as our pessimistic estimate of δ_{meas}.

B. Preparation error

The preparation of the $|0\rangle$ state is a stochastic process. The initial state after cooling is a mixture of $^2S_{1/2}$ manifold states. During preparation, the population in the $|0\rangle$ state exponentially approaches some limiting value close to one within a time ≈ 2 μs as depicted in Figure 3. In order to artificially introduce preparation errors, we simply limit the preparation time.

At the laser frequency used for preparation, 369 nm light excites the $^2S_{1/2}, F = 1$ to $^2P_{1/2}, F' = 1$ transitions (see Figure 1a). From the $F' = 1$ states, the ion may decay into any of the $^2S_{1/2}$ manifold states. Repetition of this scattering process depletes population from the $^2S_{1/2}, F = 1$ manifold. With our beam polarization, we expect the rates of scatter from the $|\pm1\rangle$ states to be symmetric and faster than the scatter from the $|1\rangle$ state. Because the $|\pm1\rangle$ states are unaffected by the gates of the RPE sequence, population in these states after initialization is always measured in the bright state and may or may not appear as an error, depending on the ideal result of the sequence. In the worst case, this population sums with the population in the $|1\rangle$ state to constitute the additive error of initialization.

In this worst case, the probability of initialization error after a preparation interval t is equal to the population in the $F = 1$ manifold after that interval. Figure 3 shows this population as a function of the preparation time t in our diagnostic experiment. We use this result to convert between t and δ_{prep}. For example, we observe that a preparation time of about 700 ns corresponds to the theoretical $\delta = 1/\sqrt{8}$ threshold for RPE success. Because this comparison represents a worst case error, the RPE protocol is expected to tolerate even shorter preparation times in some instances.

C. Phase damping error

In order to investigate an additive error that scales with the length of the gate sequence, we illuminate the

ion during the entire sequence with dim 369 nm light tuned to the $^2S_{1/2}, F = 1 \rightarrow \,^2P_{1/2}, F' = 0$ transition. A 369 nm photon scattering off the ion causes it to excite and spontaneously decay if and only if the ion is in the $F = 1$ manifold (including the $|1\rangle$ state). Because we rotated the polarization of the 369 nm light to excite σ^+/σ^- transitions preferentially, any population that decayed to the $|\pm 1\rangle$ states was quickly re-excited and transferred to the $|1\rangle$ state. The effect of this aggregate process, considered as an error, is to dephase the qubit: it does not lead to population mixing of the $|0\rangle$ or $|1\rangle$ states, but it does lead to decay of superpositions of those states. Nielsen and Chuang refer to this as *phase damping* [4]; an example of controllably inducing a dephasing/phase damping channel is given by Urrego *et al.* [12]. The decay events we introduce through this scattering process occur discretely and independently, so we expect the aggregate error probability they induce to approach one as a simple exponential in time.

We control the intensity of the 369 nm light addressing the ion by changing the RF power applied to an AOM. Calibrating the absolute magnitude of the phase damping errors is challenging because we do not have a direct measurement of the rate of photon scatter by the ion. To the extent that the AOM responds proportionally to RF power, however, we can estimate the relative scattering rates at different AOM settings. In the course of performing RPE experiments with a known rotation angle, we can also directly observe the additive error we introduce in each sequence by comparing results with and without the injected errors.

Because the phase damping error only affects superpositions of $|0\rangle$ and $|1\rangle$ states, its effective instantaneous strength changes over the course of each RPE sequence as the superposition state changes. As long as the characteristic time scale of the error is longer than the π-time of the applied rotations and many rotations are applied, it is possible to average this effect out and treat the error as if it has a constant effective strength.

We note as an aside that this approximation breaks down and the error behavior changes notably when the 369 nm intensity is high and the inverse scattering rate is faster than the gate time. In this case, the possibility of excitation and decay constantly project the state and the unitary gates are suppressed. This limit was evident in experiments performed with large injected errors, although it is not relevant for the results presented in the next section.

V. RESULTS

For all experiments testing the robustness of RPE to injected error, we used a maximum sequence length of 2^7 calibrated $Y_{\pi/2}$ gates and a constant number of samples (M_n) independent of the sequence length. For each trial of the RPE protocol, we computed the RPE estimate θ_{est} of the gate angle using the pseudocode in the supplemen-
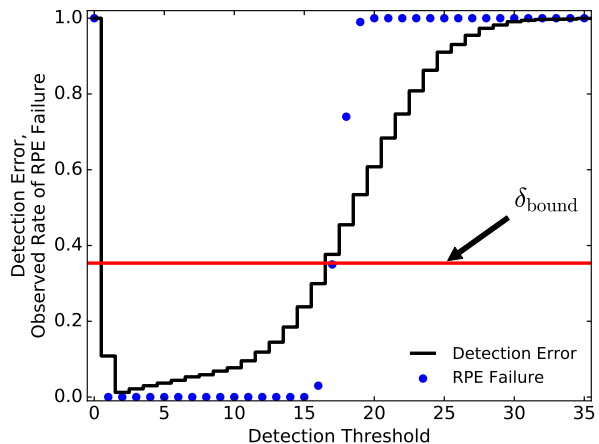

FIG. 4. Observed RPE failure rate as a function of the detection threshold (circles). The solid black curve shows the maximum detection error probability, δ_{meas}, defined in Section IV A and computed from the histograms in Figure 2. We observe a sharp onset of RPE failure when the threshold is set at 17 counts, in good agreement with when $\delta_{meas} > \delta_{bound}$. Similarly, RPE also fails when enough dark detections are mislabeled as bright, a situation which only occurs for a threshold of zero.

tal information to [7]. For each value of injected error, we conducted between 25 and 100 trials of the RPE protocol. The rate of failure, based on Eq. (6), is defined as the fraction of experiments where $|\theta_{est} - \pi/2| > \pi/2^8$.

A. Detection

A common error source in experimental trapped-ion measurements comes from drift in the intensity of the detection light at the ion, due for example to positional drift of the beam. This drift changes the photon scattering rate and leads to a different optimal threshold as described in Section IV A. Measurement degradation is easily observed in, e.g., Rabi flopping, experiments as a decrease in contrast between measurements of $|0\rangle$ and $|1\rangle$ states. It provides an obstacle for automated calibrations because it eliminates the possibility of calibrating against a fixed reference measurement value.

In order to mimic this error and test RPE, we intentionally degrade the measurement operation in RPE by changing the detection threshold used in post-processing. Each RPE sequence used $M_n = 32$ samples, and the observed rate of failure for 100 trials is shown in Figure 4. RPE succeeds in accurately estimating the gate angle despite large measurement error. As expected (see Eq. (V.16) in [6]), we observe a steep transition between success and failure near the theoretical bound of $\delta_{meas} \approx 1/\sqrt{8}$.
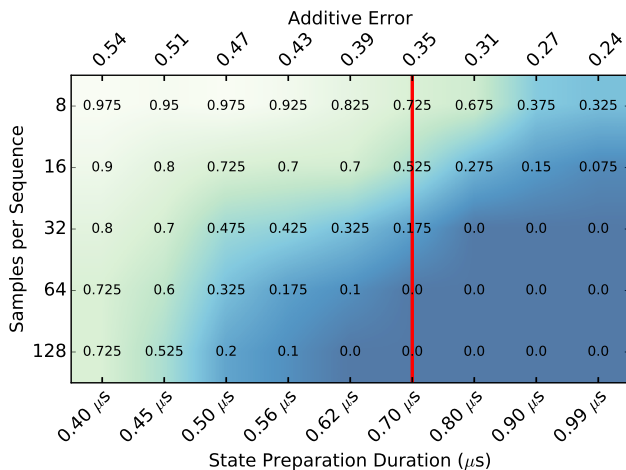
FIG. 5. Observed RPE failure rate as a function of qubit preparation duration (bottom axis) and number of samples (left axis). The background shading also indicates the observed failure rate. The top axis shows the maximum additive error δ_{prep} corresponding to the qubit preparation duration. We estimate that δ_{prep} exceeds $\delta_{\text{bound}} = 1/\sqrt{8}$ beyond 700 ns (vertical red line). The boundary of the onset of failure shifts as a function of the number of samples as expected.

B. Initialization

We intentionally degrade the fidelity of our state preparation by shortening its duration, with consequences described in Section IV B. While preparation is not frequently a large source of error in quantum applications, it can become one when the preparation involves entangled states, such as for recent clock and sensor experiments [13, 14].

We chose to explore the combined effects both of injecting error and of changing the number of samples (M_n) used in the protocol. In the interest of efficiency, especially when measurement is a relatively slow operation, calibration protocols should be optimized to use as few samples as possible. However, when operating near the bounds of the statistics allowed by shot noise, stochastic errors, e.g., due to degraded initialization, can combine with shot noise to lead to calibration failure.

In Figure 5, we show the rate of RPE failure for δ_{prep} between 0.54 (0.40 μs) and 0.24 (0.99 μs). For the center row of Figure 5, we performed $M_n = 32$ samples of each RPE sequence, and we observe agreement between the onset of failure and the $1/\sqrt{8}$ bound based on our estimate of additive error. The number of samples required to reliably achieve a correct estimate increases with the injected error, as indicated by the prescription for M_n given by Eq. (V.17) in [6]. For $M_n < 32$, RPE fails frequently even with injected error below the bound due to increased measurement shot noise.

C. Gate Error: Phase Damping

Spontaneous decay and phase damping bound the coherence time of quantum computations. As with both previous injected errors, phase damping leads to a reduction in contrast in the expected results of calibration experiments. Unlike those errors, the phase damping errors we introduce here accumulate as the experiment duration increases, causing longer sequences to return lower-contrast results. These processes will bound the maximum sequence length that can be used in the RPE protocol before the additive error exceeds its bound. In order to observe this effect, we intentionally induce phase damping by allowing 369 nm (measurement) light to leak into our experiment, with its intensity reduced by $\{-20, -23, -26, -100\}$ dB from our measurement settings. This process is described in greater depth in Section IV C.

In Figure 6 we compare the histograms of RPE estimates for these four levels of laser intensity. Because the intensities are well below saturation, we expect spontaneous scattering to be proportional to intensity. Although assigning an absolute phase damping error rate is difficult in our experiment, we observe the same qualitative behavior for RPE failure as is observed for other error sources: the onset of RPE failure is abrupt, and RPE succeeds over a wide range of injected phase damping errors rates.

In contrast to the detection and initialization errors described previously, the total additive spontaneous emission error for a given gate sequence grows with sequence length. Depending on the type of error, some sequence lengths within the RPE protocol can be more likely to cause the RPE estimate to fail. We observed that for the strongest scattering laser intensity tested (-20 dB), the RPE estimate of the rotation angle only exceeded the bound of Eq. (6) at the longest sequence length leading to a clustering of failure events within twice the claimed range of the true value. In the tests with injected preparation or measurement errors, failures were equally likely for any sequence length, resulting in a flatter error distribution.

VI. CONCLUSIONS

We have shown experimentally that the RPE protocol is robust to errors which occur during state preparation, measurement, and gates, in good agreement with the additive error bound of $1/\sqrt{8}$. Our results are promising for noisy or initially uncalibrated experiments, where the efficiency and robustness of RPE makes it an effective tool for automated gate calibrations.

In order to apply RPE efficiently and successfully, we find that it is important to take into account the dependence of the success probability on the number of samples (M_n). A greater number of samples becomes more important when the additive error is high, so, in order to
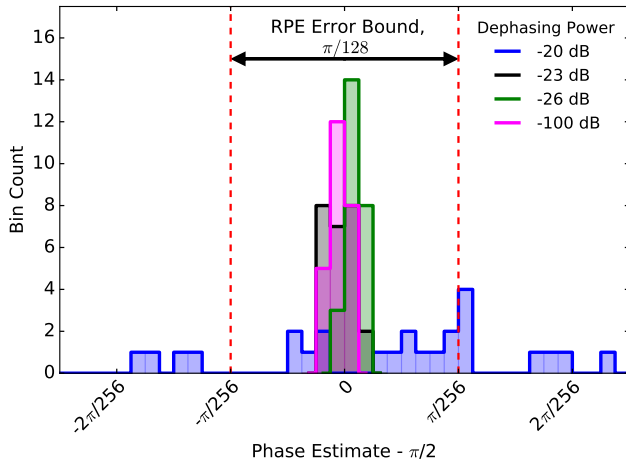
FIG. 6. Histograms of rotation angle estimates from RPE with and without additional phase damping errors. The bounds for protocol success are shown as vertical dashed lines. We injected 369 nm (measurement) light at four intensities to drive phase damping. RPE only fails at the strongest intensity (-20 dB), although even here the majority of trials succeed. At lower intensities, we observe that the distribution of estimates is significantly tighter than the provable bounds reported by the protocol.

maximize efficiency, it is useful to bound the expected error.

Similarly, we observe that the failure of RPE changes in nature depending on whether the error is constant or depends on the sequence length. When initialization and detection errors are larger than gate errors, failures in the RPE protocol can result in estimates that are very far from the true value. In a quantum computing context, this could lead to rare events of mis-calibration causing gates with poor behavior to persist until a new calibration. This catastrophic behavior needs to be treated differently with respect to fault tolerance than typical small-angle coherent errors.

We expect that RPE can also find use within the quantum sensor community. RPE could allow a user to operate a sensor in a poorly controlled environment and with poor initialization and detection operations, as long as the interaction of interest can be repeated reliably. Furthermore, the robustness of RPE to poorly calibrated operations makes it particularly appealing for automated initial calibration of sensors. The main caveat for sensor applications appears to be the need to prepare two orthogonal initial states, although RPE can tolerate significant errors in this operation. In practice, this requirement would be similar to the often-used technique of preparing alternating states for clock measurements [15].

Extensions of RPE to non-static parameters and to two-qubit gates are possible directions for future research. Ion temperature and laser intensity variations are just two examples of time-varying parameters. Depending on the root cause, the rotation angle of a gate may change reliably as a function of sequence length or instead may drift randomly; it should be possible to modify RPE to address or mitigate both of these possibilities.

An extension of RPE that calibrates two-qubit gate angle seems straightforward. One simple and necessary adjustment is to rework RPE to account for measurements from two qubits. However, the full calibration of two-qubit gates typically involves more control parameters, so an important challenge is to describe a routine based on RPE that is capable of calibrating all these parameters together both efficiently and robustly.

[1] K. Wright, K. M. Beck, S. Debnath, J. M. Amini, Y. Nam, N. Grzesiak, J.-S. Chen, N. C. Pisenti, M. Chmielewski, C. Collins, K. M. Hudek, J. Mizrahi, J. D. Wong-Campos, S. Allen, J. Apisdorf, P. Solomon, M. Williams, A. M. Ducore, A. Blinov, S. M. Kreikemeier, V. Chaplin, M. Keesan, C. Monroe, and J. Kim, Benchmarking an 11-qubit quantum computer, arXiv:1903.08181 [quant-ph] (2019).

[2] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, arXiv:1902.08543 [quant-ph] (2019).

[3] J. Kelly, P. O'Malley, M. Neeley, H. Neven, and J. M. Martinis, Physical qubit calibration on a directed acyclic graph, arXiv:1803.03226 [quant-ph] (2018).

[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).

[5] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nat. Commun. **8**, 14485 (2017); J. P. Dehollain, J. T. Muhonen, R. Blume-Kohout, K. M. Rudinger, J. K. Gamble, E. Nielsen, A. Laucht, S. Simmons, R. Kalra, A. S. Dzurak, and A. Morello, Optimization of a solid-state electron spin qubit using gate set tomography, New J.

Phys. **18**, 103018 (2016).

[6] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, Phys. Rev. A **92**, 062315 (2015), arXiv:1502.02677.

[7] K. Rudinger, S. Kimmel, D. Lobser, and P. Maunz, Experimental Demonstration of a Cheap and Accurate Phase Estimation, Phys. Rev. Lett. **118**, 190502 (2017), arXiv:1702.01763.

[8] Note that θ denotes the angle of rotation in this work, and not the rotation axis as in Kimmel *et al.* [6].

[9] N. D. Guise, S. D. Fallek, K. E. Stevens, K. R. Brown, C. Volin, A. W. Harter, J. M. Amini, R. E. Higashi, S. T. Lu, H. M. Chanhvongsak, T. A. Nguyen, M. S. Marcus, T. R. Ohnstein, and D. W. Youngner, Ball-grid array architecture for microfabricated ion traps, J. Appl. Phys. **117**, 174901 (2015).

[10] S. Olmschenk, K. C. Younge, D. L. Moehring, D. N. Matsukevich, P. Maunz, and C. Monroe, Manipulation and detection of a trapped Yb$^+$ hyperfine qubit, Phys. Rev.

A **76**, 052314 (2007), arXiv:0708.0657.

[11] R. Noek, G. Vrijsen, D. Gaultney, E. Mount, T. Kim, P. Maunz, and J. Kim, High speed, high fidelity detection of an atomic hyperfine qubit, Opt. Lett. **38**, 4735 (2013).

[12] D. F. Urrego, J.-R. Álvarez, O. Calderón-Losada, J. Svozilík, M. Nuñez, and A. Valencia, Implementation and characterization of a controllable dephasing channel based on coupling polarization and spatial degrees of freedom of light, Opt. Express **26**, 11940 (2018).

[13] T. Ruster, H. Kaufmann, M. A. Luda, V. Kaushal, C. T. Schmiegelow, F. Schmidt-Kaler, and U. G. Poschinger, Entanglement-Based dc Magnetometry with Separated Ions, Phys. Rev. X **7**, 031050 (2017).

[14] C. F. Ockeloen, R. Schmied, M. F. Riedel, and P. Treutlein, Quantum Metrology with a Scanning Probe Atom Interferometer, Phys. Rev. Lett. **111**, 143001 (2013).

[15] S. M. Brewer, J.-S. Chen, A. M. Hankin, E. R. Clements, C. W. Chou, D. J. Wineland, D. B. Hume, and D. R. Leibrandt, An ^{27}Al$^+$ quantum-logic clock with systematic uncertainty below 10^{-18}, Phys. Rev. Lett. **123**, 033201 (2019), arXiv: 1902.07694.