Robust calibration of a universal single-qubit gate set via robust phase estimation

Citation: Kimmel, Shelby, Guang Hao Low, and Theodore J. Yoder. "Robust calibration of a universal single-qubit gate set via robust phase estimation." Phys. Rev. A 92, 062315 (December 2015). © 2015 American Physical Society

As Published: http://dx.doi.org/10.1103/PhysRevA.92.062315

Publisher: American Physical Society

Persistent URL: http://hdl.handle.net/1721.1/100338

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Robust calibration of a universal single-qubit gate set via robust phase estimation

Shelby Kimmel,[1,2] Guang Hao Low,[3] and Theodore J. Yoder[3]
[1]*Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, Maryland 20742, USA*
[2]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
(Received 9 February 2015; published 8 December 2015)

An important step in building a quantum computer is calibrating experimentally implemented quantum gates to produce operations that are close to ideal unitaries. The calibration step involves estimating the systematic errors in gates and then using controls to correct the implementation. Quantum process tomography is a standard technique for estimating these errors but is both time consuming (when one wants to learn only a few key parameters) and usually inaccurate without resources such as perfect state preparation and measurement, which might not be available. With the goal of efficiently and accurately estimating specific errors using minimal resources, we develop a parameter estimation technique, which can gauge key systematic parameters (specifically, amplitude and off-resonance errors) in a universal single-qubit gate set with provable robustness and efficiency. In particular, our estimates achieve the optimal efficiency, Heisenberg scaling, and do so without entanglement and entirely within a single-qubit Hilbert space. Our main theorem making this possible is a robust version of the phase estimation procedure of Higgins *et al.* [B. L. Higgins *et al.*, New J. Phys. **11**, 073023 (2009)].

DOI: 10.1103/PhysRevA.92.062315 PACS number(s): 03.67.Ac, 03.65.Wj, 03.67.Lx

I. INTRODUCTION

Not all errors in a quantum computation experiment are created equal. There are actually two broad classes of error, unitary errors, also known as systematic errors, and nonunitary errors, also known as decoherence. Both sets of errors need to be corrected below a certain threshold for scalable quantum computation to take place [1,2]. Correcting systematic errors, such as overrotation or off-resonance errors, is typically regarded as the easier task; because these errors are directly related to the controls available to an experimenter, they can be directly corrected by changing those controls. In this respect systematic errors contrast with decoherence, which is typically less affected by an experimenter's control software and more influenced by imperfect or nonideal hardware.

However, even though systematic errors are considered the easier of the two to correct, calibrating gates in a quantum computer to reduce systematic errors can still take hours even for modest system sizes and moreover this calibration may have to be repeated every time the quantum computer is switched on [3]. Not only can this process be inefficient in terms of the precision of the estimates with respect to time, but standard techniques for estimating systematic errors often suffer from measurement bias, leading to inaccurate estimates [4].

To characterize systematic errors, quantum process tomography [5] has long been a valuable tool in the experimental toolkit. However, standard techniques [5] require perfect state preparation, perfect measurement, and at least some perfect gates. Especially during the calibration stage of an experiment, it is unreasonable to assume access to such perfect resources and without them, standard process tomography results in a difficult nonlinear estimation problem [6–8] and hence the estimates obtained using this technique are typically inaccurate. Moreover, systematic errors are controlled by a few key parameters, but unless the measurement basis of the tomography procedure is specialized, as, e.g., in [9], the extraction of those few important parameters can require resources that scale exponentially with the size of the system and can be time consuming even for single-qubit processes.

Recent approaches aim to circumvent the stringent requirements of standard tomography. Randomized benchmarking (RB) [10–12], randomized benchmarking tomography (RBT) [13], and other tools based on randomized benchmarking [14,15] can characterize quantum error processes even when nothing is known about state preparation and measurement. However, these procedures require access to relatively good Clifford operations [16,17]. In addition, other than certain key parameters such as the average fidelity, single parameters cannot be extracted efficiently. While the average fidelity can be learned efficiently using RB, average fidelity gives no information about the nature of the systematic errors on the gates and so is useless for experimentalists who would like to use tomographic data to correct systematic errors.

Another promising approach is gate-set tomography (GST) [18,19]. Gate-set tomography makes no assumptions about state preparation, measurement, or processes, while still obtaining accurate estimates. However, GST is even more inefficient than standard tomography since to learn even a single parameter, one must fully characterize a complete gate set along with state preparation and measurement.

We propose a procedure to estimate simultaneously all the systematic errors in a universal single-qubit gate set. This procedure falls in between existing protocols in terms of required resources and assumptions, but is optimal in terms of asymptotic efficiency. Rather than doing full tomography, we extract only parameters that correspond to systematic errors, precisely the errors that the experimentalists can easily correct. We learn those parameters efficiently and nonadaptively; in fact, we are Heisenberg limited. Like GST, we require no perfect resources and, moreover, we do not require any additional gates besides the ones we are characterizing. We also never require more than a single-qubit Hilbert space. In particular, we never need entangled states, like those often employed in interferometric phase estimation procedures [20,21]. Instead, the source of the quantum advantage in our procedure is the exploitation of long

coherence times of the qubit system and our ability to apply a gate multiple times in series. This allows small variations in gates to coherently accumulate into large observables.

Of course, like other Heisenberg-limited studies [22,23], a finite coherence time ultimately limits the estimation accuracy that we can achieve. However, our procedure does retain Heisenberg scaling against state preparation errors and measurement errors. Thus, while a standard parameter estimation scheme (one that repeatedly prepares a state, applies an operation, and then measures) is limited by uncertainty in the measurement operator, our procedure can obtain Heisenberg-limited, arbitrarily precise parameter estimates even with unknown (but not too large) errors in the measurement operator. In this way, our procedure also has some of the character of randomized benchmarking.

In order to achieve these gains in efficiency and accuracy, we lose some of the flexibility of other procedures. Our procedure will fail if errors are larger than some threshold amount. Also, the procedure is most useful when the experimentalist has precise control over the gates and can undo the systematic errors once they are characterized. We hope that a calibration procedure such as the one we describe could be used to quickly tune up gates before more sophisticated procedures such as RBT or GST are employed to characterize nonsystematic (decoherence) errors.

Our main theorem says that it is possible to perform phase estimation in the presence of errors. In particular, we consider additive errors in the measurement probabilities of experiments. This is a fairly straightforward idea, but it turns out that many different effects can be swept into these additive errors. For example, state preparation and measurement errors can be seen as additive errors. We show how to do phase estimation in the presence of these additive errors and extract two parameters of a process, amplitude and off-resonance errors, instead of only learning the phase of a rotation, as is typical. It turns out that while estimating one of the parameters of interest, the effect of the other parameter can be thought of as another additive error. Moreover, when multiple additive errors occur simultaneously, the result is still an additive error, with (worst-case) magnitude equal to the sum of the magnitudes of the individual additive errors. In particular, we modify and improve a nonadaptive phase estimation technique of Higgins *et al.* [24] to show the following.

Theorem 1. Suppose that we can perform two families of experiments, $|0\rangle$ experiments and $|+\rangle$ experiments, indexed by $k \in \mathbb{Z}^+$, whose probabilities of success are, respectively,

$$p_0(A,k) = \frac{1+\cos(kA)}{2} + \delta_0(k), \tag{1.1}$$

$$p_+(A,k) = \frac{1+\sin(kA)}{2} + \delta_+(k). \tag{1.2}$$

Also assume that performing either of the kth experiments takes time proportional to k and that

$$\sup_k\{|\delta_0(k)|,|\delta_+(k)|\} < 1/\sqrt{8}. \tag{1.3}$$

Then an estimate $\hat{A}$ of $A \in (-\pi,\pi]$ with standard deviation $\sigma(\hat{A})$ can be obtained in time $T = O(1/\sigma(\hat{A}))$ using nonadaptive experiments. On the other hand, if $|\delta_0(k)|$ and $|\delta_+(k)|$ are less than $1/\sqrt{8}$ for all $k < k^*$, then it is possible to obtain an estimate $\hat{A}$ of A with $\sigma(\hat{A}) = O(1/k^*)$ (with no promise on the scaling of the procedure).

We call the terms $\delta_0(k)$ and $\delta_+(k)$ additive errors. While we can only achieve Heisenberg scaling up to arbitrary precision when the additive errors have magnitude less than $1/\sqrt{8}$ for all k, some effects (such as depolarizing errors) cause additive errors that grow with k and so eventually overwhelm the $1/\sqrt{8}$ bound. However, in that case, if k^* is the k at which the errors become too large, our procedure can give an estimate with precision that is $O(1/k^*)$, which is often better than standard procedures, which are limited by uncertainty in state preparation and measurement. More precise bounds on the scaling of standard deviation with time can be found in Sec. V.

The layout of the paper is as follows. In Sec. II we define the notation for single-qubit operations and errors. In Sec. III, we apply Theorem 1 to calibrate systematic errors in a single-qubit gate set and then Sec. IV discusses the robustness of this procedure to sources of error such as imperfect state preparation, measurement noise, and decoherence. In Sec. V we modify and reanalyze the nonadaptive Heisenberg-limited phase estimation procedure of [24] to achieve better scaling and simpler bounds, resulting in the proof of Theorem 1.

II. CHARACTERIZING A UNIVERSAL GATE SET

We consider systematic errors in a universal single-qubit gate set. For the moment, we assume that the implemented gates have systematic errors but no decoherence errors and hence are perfect unitaries. (We relax these assumptions in Sec. IV). Single-qubit unitaries are defined by two parameters: their axis of rotation and their angle of rotation in the Bloch sphere. (See [25] for background on the Bloch sphere).

Two unitary gates are sufficient to create a universal single-qubit gate set. We describe a scheme to characterize a gate set where the two gates are ideally orthogonal. In particular, we consider the case that one gate is a faulty implementation of $Z_{\pi/2}$, a $\pi/2$ rotation about the Z axis of the Bloch sphere, and the other gate is a faulty implementation of $X_{\pi/4}$, a $\pi/4$ rotation about the X axis. We also assume that the experimenter can create an imperfect $|0\rangle$ state, the 1-valued eigenstate of $Z_{\pi/2}$. How good the gates, state preparation, and measurement must be initially for our procedure to work is determined by Theorem 1 and will be made clear in the calibration procedures in Sec. III.

We chose specific rotation angles for our Z and X rotations. This choice is mainly for convenience, since it turns out that access to (i) imperfect versions of the states

$$|0\rangle, \quad |+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}, \quad |\rightarrow\rangle = \frac{|0\rangle+i|1\rangle}{\sqrt{2}} \tag{2.1}$$

and to (ii) a Z_π rotation calibrated to near perfection are sufficient to characterize Z_χ and X_ϕ for any rotations χ and ϕ using our techniques. Only calibration of X_ϕ requires the second condition. These two conditions are satisfied given the gate set in the previous paragraph. Indeed, in an experiment where Z_χ and X_ϕ are available, albeit erroneously, for any χ and ϕ, it would perhaps be best to first calibrate $Z_{\pi/2}$ and $X_{\pi/4}$ rotations so that conditions (i) and (ii) are satisfied before calibrating Z_χ and X_ϕ for arbitrary χ and ϕ.

We now define our universal gates mathematically. Without loss of generality, we can define the Z axis of the Bloch sphere to be aligned with the axis of rotation of our approximate $Z_{\pi/2}$ gate. This means that our initial state preparation may not be aligned with the Z axis, but our scheme is robust against this type of error. Once the axis of our approximate $Z_{\pi/2}$ gate is fixed to the Z axis, the only free parameter is the angle of rotation. Thus, we can write our approximate $Z_{\pi/2}$ gate as

$$Z_{\pi/2}(\alpha) = \cos\left(\frac{\pi}{4}(1+\alpha)\right)\mathbb{I} - i\sin\left(\frac{\pi}{4}(1+\alpha)\right)\mathbb{P}_Z, \quad (2.2)$$

where $\{\mathbb{P}_X,\mathbb{P}_Y,\mathbb{P}_Z\}$ are the Pauli matrices, $\mathbb{I}$ is the 2×2 identity matrix, and α is a parameter that quantifies how far the implemented angle of rotation is from $\pi/2$. When $\alpha = 0$, we have implemented a perfect gate.

Likewise, without loss of generality, we define the X axis of the Bloch sphere so that the axis of rotation of our approximate $X_{\pi/4}$ gate lies along the XZ plane of the Bloch sphere. In this case, the approximate $X_{\pi/4}$ gate has two degrees of freedom: the location of the axis of rotation in the XZ plane of the Bloch sphere and its angle of rotation. More precisely, we can write our approximate $X_{\pi/4}$ gate as

$$X_{\pi/4}(\epsilon,\theta) = \cos\left(\frac{\pi}{8}(1+\epsilon)\right)\mathbb{I} - i\sin\left(\frac{\pi}{8}(1+\epsilon)\right) \times (\cos(\theta)\mathbb{P}_X + \sin(\theta)\mathbb{P}_Z), \quad (2.3)$$

where θ is the angle of the axis of rotation relative to the X axis and ϵ is a parameter that quantifies how far the implemented angle of rotation is from $\pi/4$. When $\epsilon = \theta = 0$ we have implemented a perfect gate.

Our goal is to estimate α, θ, and ϵ, with the expectation that once these systematic errors have been quantified, experimentalists can adjust the controls of the gates to set their values close to 0. If desired, the process can then be repeated: The new values of α, θ, and ϵ can be reestimated and readjusted again.

We will also need notation for a general imperfect X rotation

$$X_\phi(\epsilon,\theta) = \cos\left(\frac{\phi}{2}(1+\epsilon)\right)\mathbb{I} - i\sin\left(\frac{\phi}{2}(1+\epsilon)\right) \times (\cos(\theta)\mathbb{P}_X + \sin(\theta)\mathbb{P}_Z). \quad (2.4)$$

This expression $X_\phi(\epsilon,\theta)$ represents a rotation that is in the XZ plane of the Bloch sphere, which is approximately a rotation by an angle ϕ. In general, the parameters ϵ and θ will depend on ϕ.

In some cases, we will apply the unitary operations $X_{\pi/4}(\epsilon,\theta)$ and $Z_{\pi/2}(\alpha)$ to mixed states instead of pure states. In this case, we will use cursive letters to represent the completely positive and trace-preserving (CPTP) maps corresponding to these unitaries. That is,

$$\mathcal{X}_{\pi/4}(\epsilon,\theta)(\rho) = X_{\pi/4}(\epsilon,\theta)\rho(X_{\pi/4}(\epsilon,\theta))^\dagger, \quad \mathcal{Z}_{\pi/2}(\alpha)(\rho) = Z_{\pi/2}(\alpha)\rho(Z_{\pi/2}(\alpha))^\dagger, \quad (2.5)$$

where the dagger denotes the conjugate transpose.

We use the notation $Z_{\pi/2}(\alpha)^k$ to mean k repeated applications of $Z_{\pi/2}(\alpha)$. Unitaries act right to left, so $Z_{\pi/2}(\alpha)X_{\pi/4}(\epsilon,\theta)$ means apply the X rotation first and then the Z rotation.

III. SEQUENCES FOR ESTIMATING SYSTEMATIC ERRORS

In this section we describe sequences consisting of unitaries $Z_{\pi/2}(\alpha)$ and $X_{\pi/4}(\epsilon,\theta)$, which can be used to estimate the systematic error parameters α,ϵ, and θ. In particular, we would like to obtain observables $p_0(\alpha,k)$, $p_+(\alpha,k)$, $p_0(\theta,k)$, $p_+(\theta,k)$, $p_0(\epsilon,k)$, and $p_+(\epsilon,k)$, as described in Theorem 1. By Theorem 1, such observables will allow us to accurately estimate α, ϵ, and θ as long as the additive errors associated with these observables are not too large. We address the problem of initially bounding additive errors in Appendix C.

In this section we assume that we can prepare the states $|0\rangle$, $|+\rangle$, and $|\rightarrow\rangle$ perfectly and that we can measure (perfectly) the probability of being in the state $|0\rangle$ or the probability of being in the state $|+\rangle$. In Sec. IV we introduce state preparation and measurement errors to our protocols.

A. Estimating α

With the assumption of perfect state preparation and measurement, we can estimate α using standard phase estimation, without having to resort to robust phase estimation. One can verify that

$$|\langle +|Z_{\pi/2}(\alpha)^k|+\rangle|^2 = \frac{1+\cos[-k(\pi/2)(1+\alpha)]}{2}, \quad |\langle +|Z_{\pi/2}(\alpha)^k|\rightarrow\rangle|^2 = \frac{1+\sin[-k(\pi/2)(1+\alpha)]}{2}. \quad (3.1)$$

Comparing with Eqs. (1.1) and (1.2), we see these sequences can be used to estimate α. If N is the number of times we apply $Z_{\pi/2}(\alpha)$, by Theorem 1, we can obtain an estimate of α with standard deviation $O(1/N)$. This is what is meant by Heisenberg scaling or Heisenberg limited. (N is the most natural and unambiguous measure of resource consumption for phase estimation; see the Appendix of [24]).

B. Estimating ϵ

We next describe the sequences used to estimate ϵ. In this section, for ease of explication later in the paper, we will characterize the general gate $X_\phi(\epsilon,\theta)$, where we can always substitute $\pi/4$ for the variable ϕ to obtain the results relevant to $X_{\pi/4}(\epsilon,\theta)$. Let $\phi_\epsilon = \phi(1+\epsilon)$. Again, a simple calculation shows that

$$|\langle 0|X_\phi(\epsilon,\theta)^k|0\rangle|^2 = \frac{1+\cos(k\phi_\epsilon)}{2} + \sin^2\left(\frac{k\phi_\epsilon}{2}\right)\sin^2\theta, \quad |\langle 0|X_\phi(\epsilon,\theta)^k|\rightarrow\rangle|^2 = \frac{1+\sin(k\phi_\epsilon)}{2} - \sin(k\phi_\epsilon)\sin^2\frac{\theta}{2}. \quad (3.2)$$

Comparing with Eq. (1.1), we see this sequence allows us make a measurements with success probabilities $p_{0/+}(\phi_\epsilon,k)$, with $|\delta_0(k)|,|\delta_+(k)| \leqslant \sin^2(\theta)$.

By Theorem 1, as long as $|\theta|$ is less than about 36° (along with our current assumptions of perfect state preparation and measurement), then we can estimate ϕ_ϵ, and hence ϵ (assuming a constant ϕ), with standard deviation $O(1/N)$, where N is the total number of times $X_\phi(\epsilon,\theta)$ is used over the course of the

protocol. In Appendix C we show how to independently bound the size of θ, in order to determine if $|\theta|$ is small enough to apply this protocol.

C. Estimating θ

We now discuss sequences to estimate θ. For the moment, we assume that after estimating α, we are able to set $\alpha = 0$ exactly. In Sec. IV A we will examine what happens to this protocol when α is not zero.

Consider the rotation

$$U = Z_{\pi/2}(0)X_{\pi/4}(\epsilon,\theta)^4 Z_{\pi/2}(0)^2 X_{\pi/4}(\epsilon,\theta)^4 Z_{\pi/2}(0). \quad (3.3)$$

Then, because any single-qubit unitary can be written as a rotation of some angle Φ about an axis $\vec{n}$ in the Bloch sphere, we may write

$$U = \cos\left(\frac{\Phi}{2}\right)\mathbb{I} - i\sin\left(\frac{\Phi}{2}\right)\vec{n}\cdot(\mathbb{P}_X,\mathbb{P}_Y,\mathbb{P}_Z). \quad (3.4)$$

By direct expansion, we find that the Y component of $\vec{n}$ is zero and

$$n_X = -\frac{\cos(\theta)\cos(\pi\epsilon/2)}{\sqrt{1-\sin^2\theta\cos^2(\pi\epsilon/2)}}, \quad (3.5)$$

$$n_Z = \frac{\sin(\pi\epsilon/2)}{\sqrt{1-\sin^2\theta\cos^2(\pi\epsilon/2)}}, \quad (3.6)$$

$$\sin\left(\frac{\Phi}{2}\right) = 2\sin(\theta)\cos\left(\frac{\pi\epsilon}{2}\right)\sqrt{1-\sin^2(\theta)\cos^2\left(\frac{\pi\epsilon}{2}\right)}. \quad (3.7)$$

We define the angle Θ to be such that $\cos(\Theta) = n_X$ and $\sin(\Theta) = n_Z$. Using our notation of Sec. II, we may write $U = X_\Phi(0,\Theta)$ and hence, using the techniques of Sec. III B, we can obtain a Heisenberg-limited estimate of Φ as long as $|\Theta|$ is not too large. All that remains is to show that an estimate of Φ allows us to estimate θ with similar precision and that Θ is not too large.

We have

$$|\Theta| = \arcsin|n_Z| = \arcsin\left|\frac{\sin(\pi\epsilon/2)}{\sqrt{1-\sin^2\theta\cos^2(\pi\epsilon/2)}}\right|, \quad (3.8)$$

which implies that $\sin^2\Theta$ scales as $O(\epsilon^2)$. In particular, if $\sin^2\theta < 1/\sqrt{8}$, as is necessary for estimating ϵ using the methods of Sec. III B, then $|\epsilon| < 0.341$ is sufficient for estimating Φ. We can independently verify whether $|\epsilon|$ is small enough for the protocol to succeed using the techniques of Appendix C.

We now show that estimating Φ is sufficient to estimate θ. We have

$$\sin\frac{\Phi}{2} = 2\sin\theta\cos\frac{\pi\epsilon}{2}\sqrt{1-\sin^2\theta\cos^2\frac{\pi\epsilon}{2}}, \quad (3.9)$$

which can be expanded, assuming small θ, as

$$\sin\frac{\Phi}{2} = 2\theta\cos\frac{\pi\epsilon}{2} + O(\theta^3). \quad (3.10)$$

Since ϵ can be estimated from Sec. III B, we can estimate

$$\theta = \frac{\sin(\Phi/2)}{2\cos(\pi\epsilon/2)}. \quad (3.11)$$

As long as ϵ and θ are not too large, the relationship between Φ and θ is very close to linear, so if we know the standard deviation $\sigma(\hat{\Phi})$ of $\hat{\Phi}$, our estimate of Φ, we can obtain the standard deviation of our estimate of θ, $\sigma(\hat{\theta})$, as

$$\sigma(\hat{\theta}) \leqslant \frac{\sigma(\hat{\Phi})}{4\cos(\pi\epsilon/2)}. \quad (3.12)$$

Since we can estimate Φ with Heisenberg-limited uncertainty, this means we can estimate θ with Heisenberg-limited uncertainty.

In the case that the relationship between θ and Φ is not close to linear (which can be checked using Eq. (3.9)), then while our technique gives a bound on the variance of our estimate of Φ, because we do not know the form of the distribution of this estimate, we cannot easily bound the variance of our estimate of θ. In this case, we recommend using nonparametric bootstrapping [26], which, at the cost of a constant multiplicative overhead, can be used to estimate the variance of the estimate of θ obtained from this procedure, without any assumptions on a linear relationship between Φ and θ. While it is possible that this nonlinearity would keep the estimate of θ from being Heisenberg limited, as long as the variance of our estimate of Φ is small, the relationship between Φ and θ should be approximately linear and so we expect that we will always be Heisenberg limited in our estimate.

In Sec. II we claimed that our techniques can be applied to characterize $Z_\chi(\alpha)$ and $X_\phi(\epsilon,\theta)$ for arbitrary χ and ϕ. Our techniques immediately extend to give estimates of α and ϵ for these rotations, but it may not be immediately clear how to obtain an estimate of θ in this case. The procedure is quite straightforward. First, choose a positive integer q such that $q\phi = t\pi$ for an odd integer t.[1] Construct

$$U_\phi = Z_{\pi/2}(0)X_\phi(\epsilon,\theta)^q Z_{\pi/2}(0)^2 X_\phi(\epsilon,\theta)^q Z_{\pi/2}(0). \quad (3.13)$$

Using the same procedure as before, we can then estimate θ, assuming that $|t\epsilon|$ is not too large [$|t\epsilon| < 0.341$ is sufficient if $\sin^2(\theta) < 1/\sqrt{8}$].

IV. BOUNDING AND QUANTIFYING OTHER ERRORS

In Sec. III we showed how to construct sequences such that, if states are prepared perfectly, measurements are performed perfectly, and the gates are exactly of the form we assume, then one can estimate α, ϵ, and θ at the Heisenberg limit. In this section we show that these assumptions can be relaxed and examine their effect on our protocol.

We will completely restrict ourselves to a Hilbert space of dimension 2. (So we assume that all states and operators exist and act only on this subspace). Let $\mathcal{P}(2)$ be the set of positive

[1] It may happen such a q is impossible to find (e.g., if $\phi = 2\pi/3$). Such cases occur when $\phi = (a/b)\pi$ for a/b a reduced fraction and a even. However, letting $c = a/2^s$ be the odd integer part of a, calibrating a rotation by $\phi' = (c/b)\pi$ is possible and a rotation by ϕ can be obtained by doing 2^s rotations by ϕ'.

semidefinite operators on the Hilbert space of dimension 2. By $A \succeq B$, we mean that $A - B$ is positive semidefinite. Consider a general scenario in which we would like to prepare a state ρ, apply a CPTP map $\mathcal{E}$ (which might be a sequence of gates), and then measure with the positive-operator-valued measure (POVM) $\mathcal{W} = \{W_1, \ldots, W_k\}$. Then the probability of obtaining outcome i is

$$p_i = \mathrm{tr}[W_i \mathcal{E}(\rho)]. \tag{4.1}$$

Suppose, however, that instead of preparing the state ρ perfectly, we prepare the faulty state ρ', apply the faulty CPTP map $\mathcal{E}'$, and measure using the faulty POVM $\mathcal{W}' = \{W_1', \ldots, W_k'\}$. In this case, the probability of obtaining outcome i is

$$p_i' = \mathrm{tr}[W_i' \mathcal{E}'(\rho')]. \tag{4.2}$$

Since we care about additive errors, which are a difference in probability between the desired experiment and the implemented experiment, we would like to bound $|p_i - p_i'|$.

Using the triangle inequality, we have

$$\begin{aligned} |p_i - p_i'| = {} & |\,\mathrm{tr}[W_i \mathcal{E}(\rho)] - \mathrm{tr}[W_i \mathcal{E}'(\rho)]| \\ & + |\,\mathrm{tr}[W_i \mathcal{E}'(\rho)] - \mathrm{tr}[W_i' \mathcal{E}'(\rho)]| \\ & + |\,\mathrm{tr}[W_i' \mathcal{E}'(\rho)] - \mathrm{tr}[W_i' \mathcal{E}'(\rho')]|. \end{aligned} \tag{4.3}$$

Thus the difference in experimental outcome can be split into separate contributions due to gate error, measurement error, and state preparation error. In particular, measurement error is bounded by

$$\delta_{W_i, W_i'} \equiv \max_{\substack{\rho \in \mathcal{P}(2) \\ \mathrm{tr}(\rho) = 1}} |\,\mathrm{tr}[(W_i - W_i')\rho]|, \tag{4.4}$$

state preparation error is bounded by

$$\delta_{\rho,\rho'} \equiv \max_{\substack{W \in \mathcal{P}(2) \\ W \preceq \mathbb{I}}} |\,\mathrm{tr}[W(\rho - \rho')]| = \tfrac{1}{2}\|\rho - \rho'\|_1, \tag{4.5}$$

where $\|\cdot\|_1$ is the l_1 norm or trace distance (see [25]), and the gate error is bounded by[2]

$$\begin{aligned} \delta_{\mathcal{E},\mathcal{E}'} & \equiv \max_{\substack{W,\rho \in \mathcal{P}(2) \\ W \preceq \mathbb{I} \\ \mathrm{tr}(\rho) = 1}} |\,\mathrm{tr}[W\mathcal{E}(\rho)] - \mathrm{tr}[W\mathcal{E}'(\rho)]| \\ & = \tfrac{1}{2} \max_{\substack{\rho \in \mathcal{P}(2) \\ \mathrm{tr}(\rho) = 1}} \|\mathcal{E}(\rho) - \mathcal{E}'(\rho)\|_1. \end{aligned} \tag{4.6}$$

In Sec. IV A we examine the impact of imperfect Z rotations on the gate error contribution to additive errors. In Sec. IV B we analyze the effect of depolarizing errors on the gate error contribution to additive errors. Then in Sec. IV C we look at state preparation and measurement errors and their contributions to additive errors.

[2]We use the bounded rather than completely bounded (diamond) norm here because we are restricting our Hilbert space to be of dimension 2.

A. Errors in Z rotations

In Sec. III C we described a unitary operation U, which involved applying the rotation $Z_{\pi/2}(0)$. Suppose that we cannot implement $Z_{\pi/2}(0)$, but instead can implement $Z_{\pi/2}(\alpha)$. Let U' be the gate that results when $Z_{\pi/2}(0)$ is replaced by $Z_{\pi/2}(\alpha)$ in Eq. (3.3). Let $\mathcal{U}$ and $\mathcal{U}'$ label the corresponding CPTP maps.

Using a similar triangle inequality as in Eq. (4.3), we have that

$$\begin{aligned} & |\,\mathrm{tr}[M_i(\mathcal{U}^k - (\mathcal{U}')^k](\rho)| \\ & \quad \leqslant 2k \max_{\substack{\rho \in Pos(2) \\ \mathrm{tr}(\rho) = 1}} \|[\mathcal{Z}_{\pi/2}(0) - \mathcal{Z}_{\pi/2}(\alpha)](\rho)\|_1 \\ & \quad \leqslant 4k \left|\sin\left(\frac{\pi\alpha}{4}\right)\right|, \end{aligned} \tag{4.7}$$

so a nonzero α contributes at most an amount $k\pi|\alpha|$ to $\delta_{\mathcal{U},\mathcal{U}'}$. For the additive error to be bounded, we require $|\alpha| = O(1/k)$.

In Sec. III A we showed that using $O(N)$ applications of $Z_{\pi/2}(\alpha)$, we could estimate α with standard deviation $O(1/N)$. Assuming that the control of α is precise enough to correct α to within the uncertainty of this estimate, we can obtain a new Z rotation $Z_{\pi/2}(\alpha')$ with $|\alpha'| = O(1/N)$. This improved rotation can them be used to implement the protocol for estimating θ in Sec. III C with standard deviation $O(1/N)$. Notice that both procedures (α and θ estimation) together use $O(N)$ applications of gates, so in the end, we can obtain an estimate of θ, the scales at the Heisenberg limit.

In practice, it is unrealistic to assume that experimentalists have arbitrarily precise controls and so at some point, even if α is estimated very precisely, it cannot be corrected. However, in that case, there is no need to obtain such a precise estimate, for the very reason that it cannot be corrected.

We note that the strategy employed in this section is very general and can be employed for general CPTP errors. However, when the errors have certain structure, we can do better, as in the case of depolarizing errors, which we analyze in the next section.

B. Depolarizing errors

We now consider the effect of depolarizing noise. We look at the case that each applied gate is accompanied by depolarizing noise Λ_γ, where

$$\Lambda_\gamma(\rho) = \gamma\rho + (1 - \gamma)\mathbb{I}/2. \tag{4.8}$$

If we have an experiment that involves a sequence of k gates and the probability of a certain outcome assuming no depolarizing noise is $1/2 + r$ (for $|r| \leqslant 1/2$), then in the presence of depolarizing noise, the probability of that outcome will be $1/2 + \gamma^k r$. This gives a gate error of

$$\delta_{\Lambda_\gamma} = |r|(1 - \gamma^k) \leqslant (1 - \gamma^k)/2. \tag{4.9}$$

For depolarizing errors with $\gamma = 0.99$, which is reasonable for many quantum systems, one could go to sequences of over 100 operations before the depolarizing error would overwhelm the $1/\sqrt{8}$ bound of Theorem 1. Thus, if the depolarizing error is small compared to the uncertainty in state preparation and measurement error, Theorem 1 says that our procedure will give more accurate estimates of the parameters of interest than could be obtained using standard procedures.

In fact, in the case of depolarizing errors, because of their simple form, one can do better than simply incorporating them into additive errors. The procedure of Sec. V can be reanalyzed in the presence of depolarizing errors, allowing for more precise bounds. In the interest of conciseness and clarity, we relegate this analysis to later work.

C. State preparation errors and measurement errors

State preparation and measurement (SPAM) errors are handled very well in general by our procedure. This is because SPAM errors contribute a constant additive error $(\delta_{M_i,M_i'} + \delta_{\rho,\rho'})$ no matter what gates or operations are applied in between state preparation and measurement. As long as these additive errors are not too large, our protocol works. However, there is a challenge in bounding state preparation errors. Up until this point, we have tried to make as few assumptions as possible. However, without good gates or good measurements, it is very difficult to empirically bound the fiducial state preparation error. Therefore, we do have to make an assumption: We assume that the experimenter has an upper bound on the trace distance between their true state preparation $\rho_{|0\rangle\langle 0|}$ and the ideal state preparation $|0\rangle\langle 0|$. (Once gates have been roughly calibrated, better bounds on this distance can then be obtained). In many experimental setups, the prepared state will be extremely close to the ideal [27–29]. We have

$$\delta_{|0\rangle\langle 0|,\rho_{|0\rangle\langle 0|}} \geqslant \tfrac{1}{2}\|\rho_{|0\rangle\langle 0|} - |0\rangle\langle 0|\|_1. \tag{4.10}$$

Now given the initial state $\rho_{|0\rangle\langle 0|}$ and our faulty gates $Z_{\pi/2}(\alpha)$ and $X_{\pi/4}(\epsilon,\theta)$, we would like to create states that are close in trace distance to $|+\rangle$ and $|\rightarrow\rangle$. We will use the states

$$\begin{aligned}\rho_{|+\rangle\langle +|} &= \mathcal{Z}_{\pi/2}(\alpha)\mathcal{X}_{\pi/4}(\epsilon,\theta)^2(\rho_{|0\rangle\langle 0|}),\\ \rho_{|\rightarrow\rangle\langle\rightarrow|} &= \mathcal{X}_{\pi/4}(\epsilon,\theta)^6(\rho_{|0\rangle\langle 0|}).\end{aligned} \tag{4.11}$$

Let $\xi_1 = \max\{\epsilon,\theta,\alpha\}$ and $\xi_2 = \max\{\epsilon,\theta\}$. Then using the triangle inequality, one can calculate that

$$\begin{aligned}\frac{1}{2}\|\rho_{|+\rangle\langle +|} - |+\rangle\langle +|\|_1 &\leqslant \frac{\xi_1}{2}\left(\frac{\pi^4}{8}(12+4\pi+\pi^2)\right)^{1/4}\\ &\quad + \delta_{|0\rangle\langle 0|,\rho_{|0\rangle\langle 0|}} + O\left(\xi_1^{5/4}\right),\\ \frac{1}{2}\|\rho_{|\rightarrow\rangle\langle\rightarrow|} - |\rightarrow\rangle\langle\rightarrow|\|_1 &\leqslant \frac{1}{2}\left(\frac{9\pi^2}{2}\theta^2\epsilon^2\right)^{1/4}\\ &\quad + \delta_{|0\rangle\langle 0|,\rho_{|0\rangle\langle 0|}} + O\left(\xi_2^{5/4}\right).\end{aligned} \tag{4.12}$$

In other words, we can create approximate state preparations, which induce additive errors of the order of the size of the errors in the gates used to create them, plus the base additive error from incorrect preparation of $|0\rangle\langle 0|$.

Let W be a measurement operator that is ideally close to $|0\rangle\langle 0|$. In Appendix C we show how to bound

$$\delta_{|0\rangle\langle 0|,W} = \max_{\rho} |\operatorname{tr}(W\rho) - \operatorname{tr}[|0\rangle\langle 0|\rho)]| \tag{4.13}$$

given access to the state $|0\rangle\langle 0|$ and any other state. As usual, if $\rho_{|0\rangle\langle 0|}$ is used instead of $|0\rangle\langle 0|$, the difference in outcomes will be bounded by $\delta_{|0\rangle\langle 0|,\rho_{|0\rangle\langle 0|}}$. A rotation similar to what is used in state preparation can be applied to W to obtain $W_{|+\rangle\langle +|}$ (an operator close to $|+\rangle\langle +|$) and the additive error for this measurement can be found using the standard triangle inequality strategy we have employed multiple times.

V. NONADAPTIVE HEISENBERG-LIMITED PHASE ESTIMATION

In this section we will prove Theorem 1. First, in Sec. V A, to set up the main ideas, we review, and slightly improve, the proof of Heisenberg scaling without additive errors by Higgins *et al.* [24]. This is sufficient motivation for our proof in Sec. V B.

A. Heisenberg limit without errors

Our proof of Theorem 1 is based on the nonadaptive phase estimation procedure of Higgins *et al.* [24], which states the following.

Theorem 2. Say that we can perform two families of experiments, $|0\rangle$ experiments and $|+\rangle$ experiments, indexed by $k \in \mathbb{Z}$, whose probabilities of success are, respectively,

$$p_0(A,k) = \frac{1+\cos(kA)}{2}, \tag{5.1}$$

$$p_+(A,k) = \frac{1+\sin(kA)}{2}. \tag{5.2}$$

Also assume that performing either of the kth experiments takes time proportional to k. Then, an estimate $\hat{A}$ of $A \in (-\pi,\pi]$ with standard deviation $\sigma(\hat{A})$ can be obtained in time $T = O(1/\sigma(\hat{A}))$ using nonadaptive measurements.

Proof. We reprove Theorem 2 because we use techniques that give improved analytic bounds on the scaling of $T\sigma(\hat{A})$ compared to [24]. These techniques might additionally be of broader use.

For a given k, let $\hat{a}_0$ ($\hat{a}_+$) be the number of successful outcomes of the $|0\rangle$ ($|+\rangle$) experiments, respectively, if M samples are taken of each experiment. Then one can obtain an estimate $\widehat{kA}$ for kA with standard deviation $\sigma(\widehat{kA})$:

$$\begin{aligned}\widehat{kA} &= \operatorname{atan2}(\hat{a}_+ - M/2, \hat{a}_0 - M/2) \in (-\pi,\pi],\\ \sigma(\widehat{kA}) &\propto \frac{1}{\sqrt{M}}.\end{aligned} \tag{5.3}$$

It is tempting to use this to get an estimate $\hat{A} = \widehat{kA}/k$ for A, apparently with standard deviation

$$\sigma(\hat{A}) \propto \frac{1}{k\sqrt{M}} \propto \frac{1}{T}, \tag{5.4}$$

which gives Heisenberg scaling if M is independent of k. Unfortunately, this estimate is deceptive as it is only correct up to factors of $\frac{2n\pi}{k}$, $n \in \mathbb{Z}$, due to the unknown principle range of kA.

To determine the correct range of $\widehat{kA}/k$ while still retaining Heisenberg scaling, Higgins *et al.* instead sample distributions with a range of values of k. In particular, they choose k from $\{k_1,\ldots,k_K\}$, with $k_j = 2^{j-1}$. Let $\hat{A}_j = \widehat{k_jA}/k_j$ be an estimate of A obtained from setting $k = k_j$. Then $\hat{A}_1$ is used to restrict estimates $\hat{A}_j$ for $j > 1$ to the range $(\hat{A}_1 - \pi/2, \hat{A}_1 + \pi/2]$. Continuing in this way, we assume $\hat{A}_{j+1} \in (\hat{A}_j - \pi/2^j, \hat{A}_j + \pi/2^j]$. (This restriction differs slightly from Higgins *et al.*,

in which they assume $\hat{A}_{j+1} \in (\hat{A}_j - \pi/3^j, \hat{A}_j + \pi/3^j]$. This small difference allows us to apply much stronger bounds to the probability of failure at any step).

We immediately see that $\hat{A}_K$ will only be in the correct principle range conditional on all prior estimates $\hat{A}_j$ being within $\pm\frac{\pi}{2k_j}$ of the actual value of A. In other words, the probability

$$p_{\text{error}}(k_j A) \equiv P\left[k_j(\hat{A}_j - A) \geqslant \frac{\pi}{2} \bigvee k_j(\hat{A}_j - A) < -\frac{\pi}{2}\right] \tag{5.5}$$

must be small for all j, where the average is taken over possible estimates $k_j\hat{A}_j$. (We define $p_{\text{error}}(k_j A)$ as stated instead of as $P[|k_j(\hat{A}_j - A)| \geqslant \frac{\pi}{2}]$ in order to obtain slightly better bounds). Any one such error occurring will lead to an incorrect range of $\hat{A}_K$ and thus an incorrect estimate of $\hat{A}$. As the precise value of p_{error} has a significant impact in evaluating the scaling constant of $\sigma(\hat{A}) = O(\frac{1}{T})$, a careful bound on p_{error} is required. In Lemma 1 in Appendix A, we show that if M_j samples are taken of each of the k_jth $|0\rangle$ and $|+\rangle$ experiments,

$$p_{\max}(M_j) \equiv \frac{1}{\sqrt{2\pi M_j} 2^{M_j}} > p_{\text{error}}(k_j A). \tag{5.6}$$

This is a stronger bound than what appears in the work of Higgins *et al.*, which is derived from Hoeffding's bound. This stronger bound in turn allows us to obtain a better analytic bound on the variance of our final estimate.

To calculate the variance of our estimate, we note that if no errors occur in our principal range estimates for all $k_j < k_h$, then the maximum error in our estimate is

$$\xi(h) = \frac{2\pi}{2^h}. \tag{5.7}$$

Furthermore, even if we have no errors in our principal range estimates, our final estimate can still differ from the true value by at most

$$\overline{\xi}(K) = \frac{2\pi}{2^{K+1}}. \tag{5.8}$$

Thus, we can bound the variance of our estimate $\hat{A}$ of A with

$$\begin{aligned}\sigma^2(\hat{A}) &\leqslant [1 - p_{\text{error}}(k_K A)]\overline{\xi}(K)^2 \\ &\quad + \sum_{j=1}^{K} \xi(j)^2 p_{\text{error}}(k_j A) \prod_{i=1}^{j-1} [1 - p_{\text{error}}(k_i A)] \\ &\leqslant [1 - p_{\max}(M_K)]\overline{\xi}(K)^2 + \sum_{j=1}^{K} \xi(j)^2 p_{\max}(M_j).\end{aligned} \tag{5.9}$$

Note that the first term is a variance contribution from the event of *no* errors whereas the second term is the contribution in the event where errors arise.

We assume that running the k_jth $|0\rangle$ or $|+\rangle$ experiment takes time k_j. Then the total time required for our estimate is

$$T = 2\sum_{j=1}^{K} 2^{j-1} M_j. \tag{5.10}$$

As in [24], setting $\delta_{M_j}(\sigma^2(\hat{A})T^2) = 0$, we find that Heisenberg scaling can be attained by setting

$$M_j = \alpha(K - j) + \beta \tag{5.11}$$

for $\alpha, \beta \in \mathbb{Z}^+$. The sum in Eq. (5.9) can be performed by making the replacement $p_{\max}(M_j) \leqslant \frac{1}{\sqrt{2\pi\beta 2^{M_j}}}$. One finds that $\alpha > 2$ is necessary to prevent the sum from growing faster than $\sim 4^{-K}$, which results in poorer-than-Heisenberg scaling. We obtain

$$\begin{aligned}\sigma^2(\hat{A}) &\leqslant \frac{\pi^2}{4^K}\left[1 + p_{\max}(\beta)\left(3 + \frac{16}{2^\alpha - 4}\right)\right], \\ T &< 2^{K+1}(\alpha + \beta), \\ \sigma(\hat{A})T &\leqslant 2\pi(\alpha + \beta)\sqrt{1 + p_{\max}(\beta)\left(3 + \frac{16}{2^\alpha - 4}\right)},\end{aligned} \tag{5.12}$$

which holds for all $K > 0$. ■

Thus Heisenberg scaling can be obtained for any $\alpha > 2$ and $\beta > 0$. Optimizing Eq. (5.12) over the integers gives $\sigma(\hat{A})T < 12.4\pi$ at $\alpha = 3$ and $\beta = 1$. Better bounds of $\sigma(\hat{A})T < 10.7\pi$ can be attained at $\alpha = 5/2$ and $\beta = 1/2$, where fractional values of M_j means one rounds up to the nearest integer value and performs that many experiments. This improved bound also uses a more sophisticated analysis of Eq. (5.9), in which we pull out the last $j = K, K-1, \ldots, K-z$ terms from the sum in Eq. (5.9) and use $p_{\max}(M_j) \leqslant \frac{1}{\sqrt{2\pi M_{K-z}} 2^{M_j}}$ for values of $j < K - z$ to transform the remainder into a geometric sum. These analytic bounds are significant practical improvements over those in [24], where $\sigma(\hat{A})T < 54\pi$ at $\alpha = 8\ln 2$ and $\beta = 23/2$.

We compare our result to the scaling of various other phase estimation procedures (including maximum likelihood and procedures using entanglement) in Appendix B. While the improved analysis of this section gives us better analytic scaling than was previously known for nonadaptive phase estimation, our main motivation is to obtain better results in the presence of additive errors. The present analysis allows us to include much larger additive errors than would have been possible previously.

B. Including additive errors

We now prove Theorem 1.

Proof. We now consider the case that the success probabilities of our experiments differ from the ideal probabilities by constant factors $\delta_0(k_j)$ and $\delta_+(k_j)$ as

$$p_0(A, k_j) = \frac{1 + \cos k_j A}{2} + \delta_0(k_j), \tag{5.13}$$

$$p_+(A, k_j) = \frac{1 + \sin k_j A}{2} + \delta_+(k_j). \tag{5.14}$$

Let

$$\delta_j = \max\{|\delta_0(k_j)|, |\delta_+(k_j)|\}. \tag{5.15}$$

Suppose we use exactly the same procedure to estimate A as in the case of no additive errors. Then in Lemma 2 in Appendix A we show that now

$$\begin{aligned}p_{\max}(M_j, \delta_j) &\equiv \frac{1}{\sqrt{2\pi}(1 - \sqrt{8}\delta_j)} \frac{[1 - (1/2)(1 - \sqrt{8}\delta_j)^2]^{M_j}}{\sqrt{M_j}} \\ &> p_{\text{error}}(k_j A),\end{aligned} \tag{5.16}$$

where $p_{\text{error}}(k_j A)$ is defined in Eq. (5.5).

Now consider replacing M_j by $F(\delta_j,M_j)\times M_j$, where $F(\delta_j,M_j)$ is

$$F(\delta_j,M_j)=\frac{\log[(1/2)(1-\sqrt{8}\delta_j)^{1/M_j}]}{\log[1-(1/2)(1-\sqrt{8}\delta_j)^2]}. \tag{5.17}$$

Then as long as $\delta_j<1/\sqrt{8}\approx 0.354$ we have

$$p_{\max}(F(\delta_j,M_j)M_j,\delta_j)\leqslant\frac{1}{\sqrt{2\pi M_j}2^{M_j}}. \tag{5.18}$$

This bound is the same as Eq. (5.6). This means that by increasing the number of samples of the jth experiment by a factor $F(\delta_j,M_j)$, we can get the same error bounds as if there were no additive errors $\delta_0(k_j)$ and $\delta_+(k_j)$.

Suppose there is some smallest h such that $\delta_h\geqslant 1/\sqrt{8}$. In this case, no matter how many times we repeat the experiments, no matter how many samples we take, $p_{\text{error}}(h)$ will not be bounded. However, we can still use the procedure of the previous section to obtain an estimate of $\hat{A}$ with variance proportional to $4^{-(h-1)}$, by using $F(\delta_j,M_j)M_j$ samples for each $j\leqslant h-1$, proving the second part of Theorem 1.

Furthermore, if

$$\sup_j\delta_j=1/\sqrt{8},\text{ but }\max_j\delta_j\neq 1/\sqrt{8}, \tag{5.19}$$

then we can always increase the number of samples taken of each experiment in order to counteract the effect of additive errors. This means that we can obtain arbitrarily accurate estimates. However, the size of the required $F(\delta_j,M_j)$ blows up, so we will no longer have Heisenberg scaling.

However, if $\sup_j\delta_j<1/\sqrt{8}$, then for all j we have

$$\delta_j<1/\sqrt{8}-e\equiv\delta' \tag{5.20}$$

for some constant e. Then if we take F_jM_j samples of the jth iteration, where

$$F_j=\frac{\log[(1/2)(1-\sqrt{8}\delta')^{1/M_j}]}{\log[1-(1/2)(1-\sqrt{8}\delta')^2]}, \tag{5.21}$$

we can attain the correct bounds on p_{error}. If we set $M_j=\alpha(K-j)+\beta$ as before, M_j is a monotonically decreasing sequence in j, so F_j is a monotonically increasing sequence. Thus, we have $F_j\leqslant F_K$ for all $j=1,2,\ldots,K$.

If for each M_j we replace M_j by F_jM_j, we have increased the total time required by the procedure by at most a constant factor F_K and obtained at least as good a p_{error} at each step as in the case without any errors $\delta_0(k_j)$ or $\delta_+(k_j)$. Thus we can obtain Heisenberg scaling, where $T\sigma_A$ increases by the constant F_K compared to the case without additive errors $\delta_0(k_j)$ or $\delta_+(k_j)$. This completes the proof of Theorem 1. ■

VI. CONCLUSIONS AND OPEN PROBLEMS

There are many ways to extend and refine the ideas of this paper. In particular, while the techniques described here seem to apply broadly for single-qubit operations, it would be both interesting theoretically and of great practical use if these procedures could be extended to multiqubit operations.

Additionally, there is much room for improvement in terms of error analysis. In this work we have suggested treating depolarizing or amplitude damping noise as contributing to additive errors. However, this is essentially a worst-case scenario, in which every process adversarially drives one away from the desired state by as much as possible. In reality, we would expect the repeated applications of the gate to have a twirling effect, thus mitigating, or at least averaging, the effect of noise, as in randomized benchmarking [10]. In addition, it would be of practical relevance to analyze the case where θ_A and ϵ_A are not fixed, but shift over time.

Finally, at least on the surface, our procedure has many similarities to randomized benchmarking: Both procedures are (more or less) robust to SPAM errors and involve applying increasingly lengthy sequences of operations. These similarities draw the following question: Is there an explicit connection between phase estimation and randomized benchmarking?

ACKNOWLEDGMENTS

We are grateful for helpful discussions with Colm Ryan, Blake Johnson, and Marcus da Silva. Funding for S.K. was provided by the Department of Defense. G.H.L. acknowledges support from the ARO Quantum Algorithms program. T.J.Y. acknowledges the support of the NDSEG fellowship program.

APPENDIX A: BOUNDS ON p_{error}

In this section we bound the probability of making an error at any step during our estimation procedure. An error occurs at the jth iteration if

$$\left(k_j(\hat{A}_j-A)\geqslant\frac{\pi}{2}\right)\bigvee\left(k_j(\hat{A}_j-A)<-\frac{\pi}{2}\right). \tag{A1}$$

In the below analysis, we replace $k_j\hat{A}$ with the variable $\hat{\varphi}$ and k_jA with φ. In Lemma 1 we consider the case without additive errors $\delta_0(k_j)$ and $\delta_+(k_j)$ and in Lemma 2 we include these errors.

Lemma 1. For $\varphi\in(-\pi,\pi]$ let

$$p_0=\frac{1+\cos(\varphi)}{2}, \tag{A2}$$

$$p_+=\frac{1+\sin(\varphi)}{2} \tag{A3}$$

and let $\hat{a}_0$ (respectively $\hat{a}_+$) be drawn from the binomial distribution $\mathcal{B}(M,p_0)$ (resp. $\mathcal{B}(M,p_+)$). Let

$$\hat{\varphi}=\operatorname{atan2}\left(\frac{2}{M}\hat{a}_+-1,\frac{2}{M}\hat{a}_0-1\right) \tag{A4}$$

be an estimate for φ (and if $a_0=a_+=M/2$, then $\hat{\varphi}$ is chosen uniformly at random from $(-\pi,\pi]$). Then

$$p_{\text{error}}(\varphi)<\frac{1}{\sqrt{2\pi M}2^M}, \tag{A5}$$

where

$$p_{\text{error}}(\varphi)\equiv P\Big[(\hat{\varphi}-\varphi\geqslant\pi/2)\bigvee(\hat{\varphi}-\varphi<-\pi/2)\Big] \tag{A6}$$

and the probability is taken over the possible outcomes $\hat{a}_0$ and $\hat{a}_+$.

Proof. While Hoeffding's inequality gives a loose bound on $p_{\text{error}}(\varphi)$, we will use a geometric interpretation to obtain a stronger and asymptotically exact result. In particular, we can extract an estimate $\hat{\varphi}$ for φ graphically by plotting the value of $\hat{a}_0$ and $\hat{a}_+$ on orthogonal axes, as shown in Fig. 1(a).

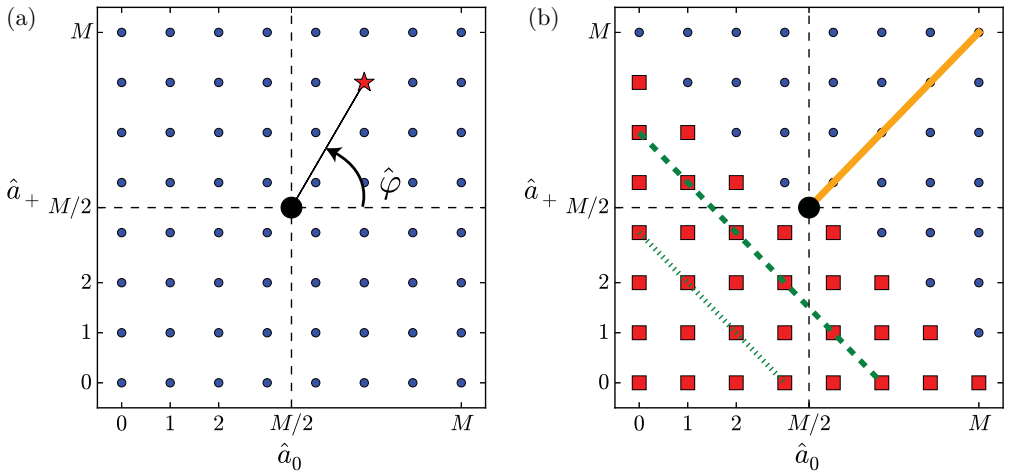

FIG. 1. (Color online) In (a) we show how to calculate $\hat{\varphi}$ given $\hat{a}_0$ and $\hat{a}_+$. Note that $\hat{a}_0$ and $\hat{a}_+$ can take values in $\{0,1,\ldots,M\}$, so the blue circles represent the possible outcomes $(\hat{a}_0,\hat{a}_+)$. In (b) we consider the case that φ lies along the orange line in the upper right quadrant, corresponding to the maximum value of p_{error}. In this case, all of the points with red square markers correspond to errors. We sum the probability of being at one of these points by first calculating the probability of being at one of the points intersected by the green diagonal dashed lines.

Before we take advantage of this geometric interpretation, we first will show that $p_{\text{error}}(\varphi)$ is largest when $\varphi = \pi/4$ and thus we need only analyze $p_{\text{error}}(\pi/4)$. We introduce the substitutions $\hat{y} = \frac{2}{M}\hat{a}_+ - 1$ and $\hat{x} = \frac{2}{M}\hat{a}_0 - 1$ and consider the inner product

$$\hat{r} = (\hat{x},\hat{y}) \cdot (\cos\varphi, \sin\varphi). \tag{A7}$$

Note that $p_{\text{error}}(\varphi)$ corresponds to the probability that $\hat{r}$ is less than 0 (with some small correction because of one-sided error). In the limit of very large M, $\hat{r}$ becomes a weighted sum of two independent normal distributions and is hence a normal distribution itself. As normal distributions are completely characterized by their mean and variance, in this limit, $p_{\text{error}}(\varphi)$ depends only on the mean and variance of $\hat{r}$. In particular, $p_{\text{error}}(\varphi)$ will be largest when the mean of this distribution is smallest and the variance is largest.

Using the well-known properties of binomial distributions and properties of sums of independent distributions, we have

$$E[\hat{r}] = 1, \quad \text{Var}[\hat{r}] = \frac{1}{2M}\sin^2(2\varphi). \tag{A8}$$

Thus the variance of $\hat{r}$ and hence the probability of error is largest when $\varphi = \pi/4 + q\pi/2$ for any integer q. When M is not large, we verify (see Fig. 2) that $p_{\text{error}}(\varphi)$ is indeed largest at $\varphi = \pi/4$.

This leads to a drastic simplification: We need only bound $p_{\text{error}}(\pi/4)$. ($p_{\text{error}}(\varphi)$ for $\varphi = \pi/4 + q\pi/2$ is the same as $\varphi = \pi/4$ by symmetry.) This corresponds to φ lying along the orange line in Fig. 1(b). Then an error occurs when values of $\hat{a}_0$ and $\hat{a}_+$ correspond to the red square markers in Fig. 1(b). Thus, to bound $p_{\text{error}}(\pi/4)$, we calculate the probability of ending up at any one of the red markers. We do this by summing over the cases where $\hat{a}_0 + \hat{a}_+$ is constant and no greater than M, corresponding to the dashed green lines in Fig. 1(b).

For $\varphi = \pi/4$ we have $p_0 = p_+ \equiv p = (2+\sqrt{2})/4$ and the probability of finding $\hat{a}_0 = a_0$ and $\hat{a}_+ = a_+$ is

$$P[a_0,a_+] = \binom{M}{a_0}\binom{M}{a_+}\left(\frac{p}{1-p}\right)^{a_0+a_+}(1-p)^{2M}. \tag{A9}$$

The probability of lying on a line $\hat{a}_0 + \hat{a}_+ = b$ is

$$\begin{aligned} P_{\text{diag}}(b) &= \sum_{a_0=0}^{M} P[a_0, b-a_0] \\ &= \binom{2M}{b}\left(\frac{p}{1-p}\right)^b (1-p)^{2M}. \end{aligned} \tag{A10}$$

Summing over the lines of constant $a_0 + a_+$ up to $M-1$ and including half of the line $a_0 + a_+ = M$, we have

$$\begin{aligned} p_{\text{error}}(\pi/4) &= \sum_{b=0}^{M} P_{\text{diag}}(b) - \frac{1}{2}P_{\text{diag}}(M) \\ &= [p(1-p)]^M \frac{(2M)!}{(M!)^2}\left[H\left(M, \frac{1-p}{p}\right) - \frac{1}{2}\right] \\ &= \frac{(2M)!}{8^M (M!)^2}\left[H\left(M, \frac{2-\sqrt{2}}{2+\sqrt{2}}\right) - \frac{1}{2}\right], \end{aligned} \tag{A11}$$

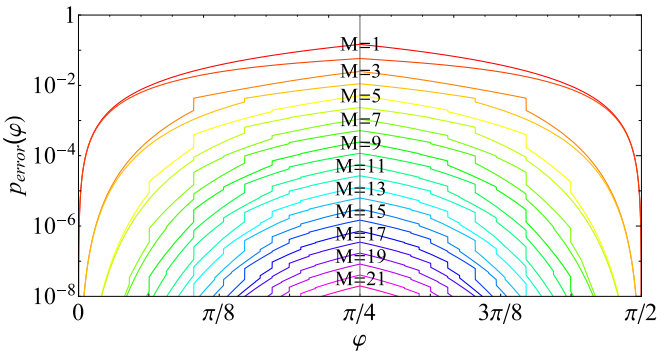

FIG. 2. (Color online) Exact probability of error as a function of φ by enumeration over all possible outcomes $\hat{a}_0$ and $\hat{a}_+$ that lead to errors in $\hat{\varphi}$ defined in Eq. (A6). Different lines correspond to the labeled number of repeats $M = 1,2,\ldots$ from the top. Observe that the maximum occurs at $\varphi = \pi/4$ for all M.

where

$$H(M,z) = \sum_{x=0}^{M} \frac{(M!)^2 z^x}{(M-x)!(M+x)!}. \tag{A12}$$

As the $x = 0$ term is 1 and the ratio of successive terms in $H(M,z)$ is

$$\frac{M-x}{1+M+x} z < z, \tag{A13}$$

we can bound this sum with a geometric series

$$H(M,z) < \sum_{x=0}^{M} z^x < \frac{1}{1-z}. \tag{A14}$$

Using Stirling's approximation $n! \sim \sqrt{2\pi n}(n/e)^n$ and noting that the fractional error of the approximation decreases monotonically with n, we obtain the remarkably simple bound

$$p_{\text{error}}(\varphi) < \frac{1}{\sqrt{2\pi M} 2^M}, \tag{A15}$$

which is tight in the limit $M \to \infty$. ■

We now include additive errors in the analysis.

Lemma 2. For $\varphi \in (-\pi,\pi]$ and δ_0,δ_+ such that $|\delta_0|,|\delta_+| \leqslant \delta < 1/\sqrt{8}$, let

$$p_0 = \frac{1+\cos(\varphi)}{2} + \delta_0, \tag{A16}$$

$$p_+ = \frac{1+\sin(\varphi)}{2} + \delta_+ \tag{A17}$$

and let $\hat{a}_0$ ($\hat{a}_+$) be drawn from the binomial distribution $\mathcal{B}(M,p_0)$ [$\mathcal{B}(M,p_+)$]. Let

$$\hat{\varphi} = \text{atan2}\left(\frac{2}{M}\hat{a}_+ - 1, \frac{2}{M}\hat{a}_0 - 1\right) \tag{A18}$$

be an estimate for φ (and if $a_0 = a_+ = M/2$, then $\hat{\varphi}$ is chosen uniformly at random from $(-\pi,\pi]$). Then

$$p_{\text{error}}(\varphi,\delta_+,\delta_-) < \frac{e}{2\pi} \frac{1}{1-\sqrt{8}\delta} \frac{[1-(1/2)(1-\sqrt{8}\delta)^2]^M}{\sqrt{M}}, \tag{A19}$$

where

$$p_{\text{error}}(\varphi,\delta_+,\delta_-) \equiv P\left[(\hat{\varphi}-\varphi \geqslant \pi/2) \bigvee (\hat{\varphi}-\varphi < -\pi/2)\right] \tag{A20}$$

and the probability is taken over the possible outcomes $\hat{a}_0$ and $\hat{a}_+$.

Proof. This proof will be similar to the proof of Lemma 1, so we will omit some of the details if they parallel the previous result. As done in Lemma 1, we introduce the substitutions $\hat{x} = \frac{2}{M}\hat{a}_0 - 1$ and $\hat{y} = \frac{2}{M}\hat{a}_+ - 1$ and consider

$$\hat{r} = (\hat{x},\hat{y}) \cdot (\cos\varphi, \sin\varphi). \tag{A21}$$

We find in this case that in the limit of large M,

$$E[\hat{r}] = 1 + 2(\delta_0 \cos\varphi + \delta_+ \sin\varphi),$$
$$\text{Var}[\hat{r}] = \frac{1 - [\cos\varphi(2\delta_0 + \cos\varphi)]^2 - [\sin\varphi(2\delta_+ + \sin\varphi)]^2}{M}. \tag{A22}$$

As explained in the proof of Lemma 1, p_{error} is maximized when we simultaneously minimize $\hat{r}$'s expectation and maximize its variance. Using $|\delta_0|,|\delta_+| \leqslant \delta$, we have

$$E[\hat{r}] \geqslant 1 + \sqrt{8}\delta\cos(\varphi - s),$$
$$\text{Var}[\hat{r}] \leqslant \frac{1}{M}\{1 - \cos\varphi^2 \min[1,(2\delta + \sqrt{2}\cos s\cos\varphi)^2] - \sin\varphi^2 \min[1,(2\delta + \sqrt{2}\sin s \sin\varphi)^2]\}, \tag{A23}$$

where $s = \pi(\frac{1}{4} + \frac{j}{2})$, $j = 0,1,2,3$, is used to represent the signs of δ_0 and δ_+. Thus, the worst-case bounds

$$E[\hat{r}] \geqslant 1 - \sqrt{8}\delta,$$
$$\text{Var}[\hat{r}] \leqslant \frac{1}{M}\left[1 - \left(\frac{1}{\sqrt{2}} - 2\delta\right)^2\right] \tag{A24}$$

are obtained when $\delta_0 = \delta_+ = -\delta$ (corresponding to $s = \pi + \pi/4$) and $\varphi = \pi/4$, leading to $p_0 = p_+ \equiv p = (2+\sqrt{2})/4 - \delta$. We thus have $p_{\text{error}}(\varphi,\delta_+,\delta_-) \leqslant p_{\text{error}}(\pi/4, -\delta, -\delta)$.

The bound on $p_{\text{error}}(\pi/4, -\delta, -\delta)$ is then obtained by a calculation identical to the proof of Lemma 1 from Eq. (A9) onward, except with $p = (2+\sqrt{2})/4 - \delta$. We obtain

$$\begin{aligned} &p_{\text{error}}(\pi/4, -\delta, -\delta) \\ &= \frac{(2M)!}{4^M (M!)^2}\left(1 - \frac{1}{2}(1-\sqrt{8}\delta)^2\right)^M \\ &\quad \times \left[H\left(M, \frac{2-\sqrt{2}+4\delta}{2+\sqrt{2}-4\delta}\right) - \frac{1}{2}\right] \\ &< \frac{1}{\sqrt{2\pi}} \frac{1}{1-\sqrt{8}\delta} \frac{[1-(1/2)(1-\sqrt{8}\delta)^2]^M}{\sqrt{M}}. \end{aligned} \tag{A25}$$

Observe that Eq. (A15) is recovered in the absence of additive errors (i.e., when $\delta = 0$). ■

APPENDIX B: SCALING OF PHASE ESTIMATION PROCEDURES

In Sec. V A we gave an analytic bound on the scaling of our Heisenberg-limited phase estimation technique. Optimizing Eq. (5.9) gave $\sigma(\hat{A})T < 10.7\pi$.

This upper bound on the Heisenberg scaling constant should of course be compared to lower bounds. A number of lower bounds are commonly cited in the literature, depending on the specification of allowed resources. The best possible bound is $\sigma(\hat{A})T \geqslant 1$ [30], often used in the atomic clock community [31]. The resources required are similar to those used for our scheme, except that there is no iteration from $j = 1,\ldots,K-1$, so only the largest K experiment is used. However, achieving this bound is only possible when the principle range of A is known, a reasonable assumption when tracking well-characterized frequencies, but not when A is completely unknown.

The next largest bound on the scaling is $\sigma(\hat{A})T \geqslant \pi$ [32], which is achievable using quantum phase estimation. Unlike the above case, A can be completely unknown initially. However, this scheme requires the resource of entanglement between different experimental runs with multiqubit gates,

or nonlocal measurements [33]. Such requirements are technically demanding, which motivates entanglement-free schemes.

Reasonable lower bounds for the entanglement-free scenario can be derived, but proving whether they are achievable remains an open question. For each experiment at some k_j (with k_j as in Sec. V), the amount of information we obtain about A can be quantified by the Fisher information

$$I(A,k_j) = \mathbf{E}\left[\left(\frac{d\ln p(A,k_j)}{dA}\right)^2\right] = k_j^2, \tag{B1}$$

where expectation over success and failure is taken. As the $2M_j$ repeats of the experiment are independent, the total information obtained over all values of k is $I = \sum_{j=1}^{K} I(A,k_j)2M_j$. In the large-$K$ limit, $I = \frac{2}{9}4^K(3\beta+\alpha)$. Using the Cramér-Rao inequality [34] then bounds the variance of $\hat{A}$ obtained via any unbiased estimator, such as maximum-likelihood estimation, by $\sigma^2(\hat{A}) \geqslant F^{-1}$. Thus we obtain

$$\sigma(\hat{A})T \geqslant (\alpha+\beta)\sqrt{\frac{18}{\alpha+3\beta}}. \tag{B2}$$

At the settings of $\alpha = 5/2$ and $\beta = 1/2$, we obtain $\sigma(\hat{A})T \geqslant 2.0\pi$, which is about five times smaller than that obtained through Eq. (5.9).

While maximum likelihood is a reasonable approach for standard phase estimation, once additive errors are included, we no longer have an unbiased estimator, so in this setting it is unfair to compare our bound to that of the Cramér-Rao bound. Once additive errors are included, we do not have an appropriate lower bound on the scaling.

APPENDIX C: INITIAL BOUNDING TECHNIQUES

Our single-qubit calibration procedure works only when the errors are below a certain initial size. Here we show how the initial size of these errors can be bounded by conducting the appropriate experiments.

1. Bounding ϵ and θ

In Sec. III B we showed that we can estimate ϵ and θ at the Heisenberg limit as long as ϵ^2 and θ^2 are not too large. Here we give a procedure to bound the initial size of ϵ and θ.

Let

$$q_0 = |\langle 0|X_{\pi/4}(\epsilon,\theta)^4|0\rangle|^2. \tag{C1}$$

By direct calculation, we have

$$q_0 = \sin^2(\theta) + \cos^2(\theta)\sin^2\left(\frac{t\pi\epsilon}{2}\right). \tag{C2}$$

The maximum value θ can attain is found by setting $\epsilon = 0$. This gives us

$$|\theta| \leqslant \arcsin\sqrt{q_0}. \tag{C3}$$

Likewise, the maximum value ϵ can attain is found by setting $\theta = 0$. This gives us

$$|\epsilon| \leqslant \frac{2\arcsin\sqrt{q_0}}{t\pi}. \tag{C4}$$

Now we just need to bound q_0. Using Hoeffding's bound, if we make V observations of q_0, we can obtain an estimate $\hat{q}_0$ for q_0 such that

$$P(q_0 < \hat{q}_0 + \mu) > 1 - \exp(-2V\mu^2). \tag{C5}$$

Thus we have

$$\begin{aligned} |\theta| &\leqslant \arcsin\sqrt{\hat{q}_0+\mu}, \\ |\epsilon| &\leqslant \frac{2\arcsin\sqrt{\hat{q}_0+\mu}}{t\pi} \end{aligned} \tag{C6}$$

with probability $1 - \exp(-2V\mu^2)$.

2. Bounding measurement error

In this section we show how to bound $\delta_{|0\rangle\langle 0|,W}$ of Eq. (4.13), given access to W, the faulty measurement operator, and the ability to prepare the states $|0\rangle\langle 0|$ and ϱ where ϱ is ideally close to $|1\rangle\langle 1|$. Consider the following measurements:

$$\begin{aligned} G_0 &= \mathrm{tr}(W|0\rangle\langle 0|), \\ G_1 &= \mathrm{tr}(W\varrho). \end{aligned} \tag{C7}$$

Suppose V observations are made of each variable G_0 and G_1, producing estimates $\widehat{G}_0$ and $\widehat{G}_1$ of the respective variables. Then using Hoeffding's bound, we have that

$$\begin{aligned} P(G_0 > \widehat{G}_0 - \mu) &> 1 - \exp(-2V\mu^2), \\ P(G_1 < \widehat{G}_1 + \mu) &> 1 - \exp(-2V\mu^2). \end{aligned} \tag{C8}$$

We will show that if

$$\begin{aligned} G_0 &> \widehat{G}_0 - \mu \equiv \widehat{G}_0^-, \\ G_1 &< \widehat{G}_1 + \mu \equiv \widehat{G}_1^+, \end{aligned} \tag{C9}$$

then

$$\delta_W \leqslant \Delta_1 + \sqrt{\Delta_1^2 + \Delta_2^2/2}, \tag{C10}$$

where

$$\begin{aligned} \Delta_1 &= \frac{(\widehat{G}_0^-)^2 - (\widehat{G}_1^+)^2 - 3\widehat{G}_0^- - 2\widehat{G}_0^-\widehat{G}_1^+ - \widehat{G}_1^+ + 2}{2(\widehat{G}_0^- - \widehat{G}_1^+)}, \\ \Delta_2 &= 2(1 - \widehat{G}_0^-). \end{aligned} \tag{C11}$$

By the union bound, we have

$$P\left(\delta_W \leqslant \Delta_1 + \sqrt{\Delta_1^2 + \Delta_2^2/2}\right) \geqslant 1 - 2\exp(-2V\mu^2). \tag{C12}$$

One can verify that if $\widehat{G}_0 \approx 1$, $\widehat{G}_1 \approx 0$, and $\mu \ll 1$, then Δ_1 and Δ_2 are small and hence δ_W is small.

Since Pauli operators are an orthonormal basis for Hermitian operators, we can write

$$\begin{aligned} W &= \sum_{i=0}^{3} m_i \mathcal{P}_i \\ \varrho &= \frac{1}{2}\left(\mathcal{P}_0 + \sum_{i=1}^{3} r_i \mathcal{P}_i\right), \end{aligned} \tag{C13}$$

where $\mathcal{P}_0 = \mathbb{I}$, $\mathcal{P}_1 = \mathbb{P}_x$, $\mathcal{P}_2 = \mathbb{P}_y$, and $\mathcal{P}_3 = \mathbb{P}_z$. Additionally, W and ϱ must be positive semidefinite and $0 \leqslant \mathrm{tr}(W\rho) \leqslant 1$ for all ρ.

Using Eqs. (C9) and (C13), we have

$$1 \geqslant m_0 + m_3 > \widehat{G}_0^-, \tag{C14}$$

$$0 \leqslant \sum_{i=0}^{3} m_i r_i < \widehat{G}_1^+. \tag{C15}$$

We will use Eq. (C14) to upper bound the size of m_1 and m_2. The eigenvalues of W must lie in the range $[0,1]$. Explicitly evaluating the eigenvalues of W and requiring that they are in this range gives

$$0 \leqslant m_1^2 + m_2^2 \leqslant (1 - m_0)^2 - m_3^2. \tag{C16}$$

Using Eq. (C14), we have

$$1 - m_0 \geqslant m_3 > \widehat{G}_0^- - m_0. \tag{C17}$$

Thus we can write

$$m_3 = f - m_0 \tag{C18}$$

for some $\widehat{G}_0^- < f \leqslant 1$. Plugging Eq. (C18) into Eq. (C16) and taking the derivative with respect to m_0, we find

$$0 \leqslant m_1^2 + m_2^2 \leqslant (1 - f)^2. \tag{C19}$$

Since $\widehat{G}_0^- < f \leqslant 1$, we finally have

$$0 \leqslant m_1^2 + m_2^2 \leqslant (1 - \widehat{G}_0^-)^2, \tag{C20}$$

so

$$|m_1|, |m_2| \leqslant 1 - \widehat{G}_0^-. \tag{C21}$$

Using Eq. (C14) and that $1 - r_3 > 0$, we have

$$\begin{aligned} m_3 &> \frac{\widehat{G}_0^- - m_0 - m_3 r_3}{1 - r_3} \\ &> \frac{\widehat{G}_0^- - \widehat{G}_1^+ - (1 - \widehat{G}_0^-)(|r_1| + |r_2|)}{1 - r_3}, \end{aligned} \tag{C22}$$

where in the second line we have used Eq. (C21). Assuming $\widehat{G}_0^- \approx 1$ and $\widehat{G}_1^+ \approx 0$, the numerator of Eq. (C22) will be positive. Using the positive semidefinite constraint on ϱ, we have $r_3 > -\sqrt{1 - r_1^2 - r_2^2}$, so

$$m_3 > \frac{\widehat{G}_0^- - \widehat{G}_1^+ - (1 - \widehat{G}_0^-)(|r_1| + |r_2|)}{1 + \sqrt{1 - r_1^2 - r_2^2}}.$$

We always want to choose $r_1 = r_2$. If $r_1 \neq r_2$, we can replace r_1 and r_2 by their average, thereby preserving the numerator while increasing the denominator. Thus

$$m_3 > \frac{\widehat{G}_0^- - \widehat{G}_1^+ - 2(1 - \widehat{G}_0^-)|r_1|}{1 + \sqrt{1 - 2r_1^2}}. \tag{C23}$$

We now minimize the right-hand side of Eq. (C23) with respect to r_1 (assuming we are in a regime where $\widehat{G}_0^- \approx 1$ and $\widehat{G}_1^+ \approx 0$), giving

$$m_3 > \frac{2 - (2 - \widehat{G}_0^-)^2 - \widehat{G}_1^+(2\widehat{G}_0^- - \widehat{G}_1^+)}{2(\widehat{G}_0^- - \widehat{G}_1^+)}. \tag{C24}$$

At this point, we can bound the error that results from using W instead of the ideal $|0\rangle\langle 0|$. For an arbitrary state ω such that

$$\omega = \frac{1}{2}\left(\mathcal{P}_0 + \sum_{i=1}^{3} w_i \mathcal{P}_i\right), \tag{C25}$$

we have

$$|\operatorname{tr}(W\omega) - \operatorname{tr}(|0\rangle\langle 0|\omega)| \leqslant \Delta_1(1 + w_3) + \Delta_2\sqrt{\frac{1 - w_3^2}{2}}, \tag{C26}$$

with Δ_1 and Δ_2 given by Eq. (C11), and we have used the trick of replacing w_1 and w_2 by their average. Maximizing (C26) with respect to w_3 we have

$$|\operatorname{tr}(W\omega) - \operatorname{tr}(|0\rangle\langle 0|\omega)| \leqslant \Delta_1 + \sqrt{\Delta_1^2 + \Delta_2^2/2}, \tag{C27}$$

as claimed.

[1] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (ACM, New York, 1997), pp. 176–188.

[2] J. Preskill, Reliable quantum computers, Proc. R. Soc. London Ser. A **454**, 385 (1998).

[3] E. Lucero, http://godwincaruana.me/physicists-find-way-to-boot-up-quantum-computers-72-times-faster-than-previously-possible-phys-org/.

[4] 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, Randomized Benchmarking and Process Tomography for Gate Errors in a Solid-State Qubit, Phys. Rev. Lett. **102**, 090502 (2009).

[5] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).

[6] C. Stark, Global completability with applications to self-consistent quantum tomography, arXiv:1209.6499.

[7] C. Stark, Simultaneous estimation of dimension, states and measurements: Computation of representative density matrices and POVMs, arXiv:1210.1105.

[8] C. Stark, Self-consistent tomography of the state-measurement gram matrix, Phys. Rev. A **89**, 052109 (2014).

[9] A. Bendersky, F. Pastawski, and J. P. Paz, Selective and efficient quantum process tomography, Phys. Rev. A **80**, 032116 (2009).

[10] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Phys. Rev. A **77**, 012307 (2008).

[11] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, Phys. Rev. Lett. **106**, 180504 (2011).

[12] 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, Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).

[13] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Robust Extraction of Tomographic Information via Randomized Benchmarking, Phys. Rev. X **4**, 011050 (2014).

[14] J. J. Wallman, M. Barnhill, and J. Emerson, Characterization of leakage errors via randomized benchmarking, arXiv:1412.4126.

[15] J. J. Wallman, C. Granade, R. Harper, and S. Flammia, Estimating the coherence of noise, arXiv:1503.07865.

[16] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta, Investigating the limits of randomized benchmarking protocols, Phys. Rev. A **89**, 062321 (2014).

[17] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).

[18] R. Blume-Kohout, J. King Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit, arXiv:1310.4492.

[19] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, Phys. Rev. A **87**, 062119 (2013).

[20] M. Kacprowicz, R. Demkowicz-Dobrzański, W. Wasilewski, K. Banaszek, and I. A. Walmsley, Experimental quantum-enhanced estimation of a lossy phase shift, Nat. Photon. **4**, 357 (2010).

[21] M. D. Vidrighin, G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. S. Kim, A. Datta, M. Barbieri, and I. A. Walmsley, Joint estimation of phase and phase diffusion for quantum metrology, Nat. Commun. **5**, 3532 (2014).

[22] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Improvement of Frequency Standards with Quantum Entanglement, Phys. Rev. Lett. **79**, 3865 (1997).

[23] J. Kołodyński and R. Demkowicz-Dobrzański, Phase estimation without *a priori* phase knowledge in the presence of loss, Phys. Rev. A **82**, 053804 (2010).

[24] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements, New J. Phys. **11**, 073023 (2009).

[25] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2004).

[26] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (CRC, Boca Raton, 1994), Vol. 57.

[27] J. E. Johnson, C. Macklin, D. H. Slichter, R. Vijay, E. B. Weingarten, J. Clarke, and I. Siddiqi, Heralded State Preparation in a Superconducting Qubit, Phys. Rev. Lett. **109**, 050506 (2012).

[28] C. Monroe and J. Kim, Scaling the ion trap quantum processor, Science **339**, 1164 (2013).

[29] L. Robledo, L. Childress, H. Bernien, B. Hensen, P. F. A. Alkemade, and R. Hanson, High-fidelity projective read-out of a solid-state spin quantum register, Nature (London) **477**, 574 (2011).

[30] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Optimal frequency measurements with maximally correlated states, Phys. Rev. A **54**, R4649 (1996).

[31] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland, Toward Heisenberg-limited spectroscopy with multiparticle entangled states, Science **304**, 1476 (2004).

[32] D. W. Berry and H. M. Wiseman, Optimal States and Almost Optimal Adaptive Measurements for Quantum Interferometry, Phys. Rev. Lett. **85**, 5098 (2000).

[33] B. C. Sanders and G. J. Milburn, Optimal Quantum Measurements for Phase Estimation, Phys. Rev. Lett. **75**, 2944 (1995).

[34] J. L. Hodges, Jr., and E. L. Lehmann, Some applications of the Cramér-Rao inequality, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1951), pp. 13–22.