This is the accepted manuscript made available via CHORUS. The article has been published as:

Consistency testing for robust phase estimation

Antonio E. Russo, William M. Kirby, Kenneth M. Rudinger, Andrew D. Baczewski, and Shelby Kimmel

Phys. Rev. A **103**, 042609 — Published 15 April 2021

DOI: 10.1103/PhysRevA.103.042609

Consistency testing for robust phase estimation

Antonio E. Russo,[1] William M. Kirby,[2] Kenneth M. Rudinger,[1] Andrew D. Baczewski,[1] and Shelby Kimmel[3]

[1] *Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185*
[2] *Department of Physics and Astronomy, Tufts University, Medford, MA 02155*
[3] *Department of Computer Science, Middlebury College, Middlebury, VT 05753*

We present an extension to the robust phase estimation protocol, which can identify incorrect results that would otherwise lie outside the expected statistical range. Robust phase estimation is increasingly a method of choice for applications such as estimating the effective process parameters of noisy hardware, but its robustness is dependent on the noise satisfying certain threshold assumptions. We provide consistency checks that can indicate when those thresholds have been violated, which can be difficult or impossible to test directly. We test these consistency checks for several common noise models, and identify two possible checks with high accuracy in locating the point in a robust phase estimation run at which further estimates should not be trusted. One of these checks may be chosen based on resource availability, or they can be used together in order to provide additional verification.

I. INTRODUCTION

The phase estimation algorithm is ubiquitous in quantum computing. It is common as an algorithmic primitive [1–6] and is also used for error mitigation [7] and for estimating the parameters of quantum processes [8, 9]. However, error rates in noisy intermediate-scale quantum (NISQ) devices, particularly in state preparation and measurement (SPAM), present a challenge for implementing phase estimation in existing and near-future hardware [10]. This incentivizes the development of intrinsically error-resilient, or robust, protocols for phase estimation [9, 11–13].

Robust phase estimation (RPE) is one such protocol that was originally conceived as a method for characterizing single-qubit gates [14]. Recently, RPE implementations have been experimentally demonstrated on trapped-ion qubits [15, 16] and used to simulate the ground state and low-lying electronic excitations of a hydrogen molecule on a superconducting cloud-based quantum computer [17]. RPE has Heisenberg scaling, so it is optimally fast up to constant factors. It is robust to all errors below a certain threshold. Furthermore, it is easy to implement, as it involves no entangled states, or even any additional registers beyond the register on which the gate of interest acts.

RPE is based on a non-entangled-state version of phase estimation presented in 2009 by Higgins *et al.* [9]. The focal point of an RPE protocol is a particular unitary gate whose rotation angle is to be estimated. The protocol involves multiple generations of experiments where this unitary of interest is repeatedly applied in longer and longer sequences. Roughly, each generation provides an additional bit of precision to the estimate of the phase. The protocol can tolerate a relatively high degree of inaccuracy at any given round, since future generations serve to correct the accuracy. This tolerance is what gives the protocol its robustness to a wide range of errors [18].

The proof of robustness of RPE starts with the assumption that errors do not exceed a certain size, and then shows that under those conditions, by increasing the number of samples by a constant factor we can ensure that the estimates produced will still be accurate, and will still achieve Heisenberg scaling. The problem with this proof is that the errors we would like RPE to be robust to are themselves often expensive to accurately characterize. Thus it is difficult to know whether they violate the threshold required for RPE to work correctly, without resorting to costlier characterization techniques.

We address this difficulty in the present work by describing tests of the self-consistency of RPE, which can herald to the user that errors have exceeded their allowed thresholds. In particular, for several different notions of "consistency" we find sets of underlying angles that can explain the RPE measurement data. Our criteria indicate inconsistency when no such angle exists. Additionally, using realistic error models, we numerically demonstrate that the tests do a good job of flagging when errors start causing inaccuracies in the RPE estimate.

It is important to note that in this paper we are not attempting to tightly characterize resource use in RPE, which has been the primarily focus of prior work [9, 14, 19]. Rather, we test whether an instance of an RPE experiment provides an estimate that is trustworthy, given there are likely aspects of the system (e.g. stochastic error processes) that may not be well-characterized, but nevertheless may impact RPE's success. The tests we develop here are somewhat akin to statistical tests of self-consistency employed in various quantum tomography schemes [20–23]. In those cases, the aim is to perform statistically rigorous testing to see if an estimate appropriately fits the data that generated it. However, as RPE is not a tomographically complete protocol, it does not generate a fully predictive estimate for a gate set (i.e., one that can predict outcomes for any circuit using only the operations in said gate set), so we cannot simply translate the statistical tests used with tomographic protocols, and instead need to develop a different set of tools. Nevertheless, the question our tests aim to answer is quite similar to the tomographic consistency tests: given a dataset and some parameter(s) estimated from it, ought we trust those estimates?

We begin by reviewing the RPE protocol in Section II. We emphasize the multi-generational nature of RPE, which creates the opportunity for self-consistency tests. In Section III, we define various notions of consistency that can be applied to sequences of choices of estimates across generations. We first define four increasingly stringent tests that are related to inter-generational constraints, which we call *plausible consistency*, *consecutive consistency*, *local consistency*, and *uniform-local consistency*. We then define three other types of consistency, which we call *angular-historical consistency*, *probability-historical consistency*, and *intersequence consistency*, which are respectively based on an angular constraint, a statistical constraint, and consistency across different full RPE runs. These definitions ultimately lead to a series of tests that can be applied to data from an RPE experiment in order to determine its failure point, i.e., the generation at which we should cease to trust the estimates. We also provide a full reference implementation of the protocol in the Python package `pyRPE`.

Finally, in Section IV we numerically test the performance of the consistency checks defined in Section III. We apply depolarizing, dephasing, and amplitude damping noise to simulated RPE experiments for a range of error rates and target angles. We find that the angular-historical consistency and intersequence consistency checks outperform the other checks across the board, with both of them on average flagging failure within one generation of the actual failure point for typical rotation angles. These two consistency checks perform similarly, so since the angular-historical consistency check requires only the original set of data, while the intersequence consistency check requires a second RPE run, we suggest using angular-historical consistency as the baseline test, and employing the intersequence consistency test as a double-check if desired. We also find that the probability-historical consistency check flags failure before actual failure occurs with high probability, providing an option for the experimenter who wants a very conservative estimate of when failure occurs.

II. REVIEW OF THE RPE PROTOCOL

RPE is effectively a sequence of Ramsey and Rabi oscillation experiments with logarithmic spacing in the number of repetitions of the unitary under investigation,

$$U_{\mathrm{t}} = \exp[-i\theta\sigma_x/2], \tag{1}$$

where σ_x is the Pauli X matrix, and θ is the parameter we would like to learn. It proceeds across multiple *generations* of experiments indexed by $k = 0, 1, 2, ...$; in the kth generation, U_{t} is applied N_k times. At each generation, RPE produces an estimate $\hat{\theta}_k$ of the rotation angle θ of U_{t}, by combining data from prior generations. N_k is chosen such that it increases with each generation, which refines the estimate, as we will see. In what follows we consider the requirements for implementing RPE for a single-qubit gate, noting that generalization to multi-qubit unitaries is relatively straightforward [17].

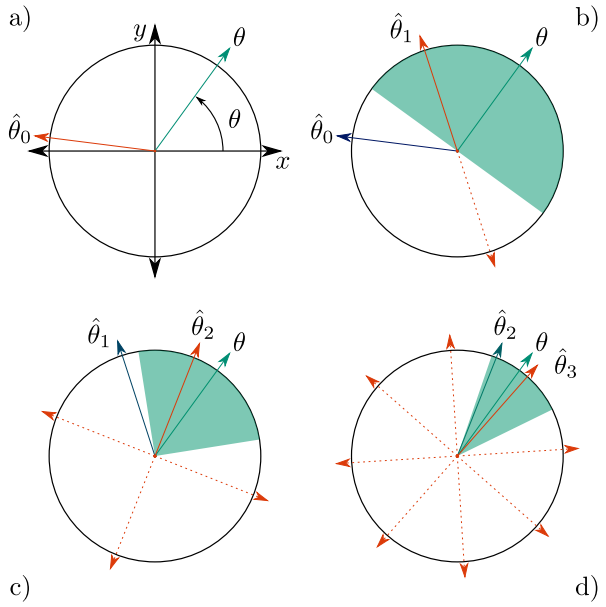

FIG. 1. Illustration of a successful RPE run. Red arrows indicate elements of Θ_k, and the green arrow indicates θ, the correct angle to be estimated. (a) A first measurement $\hat{\theta}_0$ of θ is made. (b) A second measurement of θ is made. Two candidate angles (red arrows) are possible, and RPE selects the one closest to the previous value, $\hat{\theta}_0$ (solid red). The selected angle is within $\pi/2$ of θ, as expected (within the green shaded region). (c) and (d) As before, with four and eight candidates, and shading within $\pi/4$ and $\pi/8$ of θ, respectively.

The RPE protocol requires the ability to (i) apply U_{t} repeatedly, and (ii) prepare the states $|0\rangle$ and

$$|y_+\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle). \tag{2}$$

Using these, we can construct circuits for which the distribution of outcomes encodes θ:

$$P_{\mathrm{c},k} = |\,\langle 0|\,U_{\mathrm{t}}^{N_k}\,|0\rangle\,|^2 = \frac{1}{2}\left(1+\cos(N_k\theta)\right), \tag{3}$$

$$P_{\mathrm{s},k} = |\,\langle y_+|\,U_{\mathrm{t}}^{N_k}\,|0\rangle\,|^2 = \frac{1}{2}\left(1+\sin(N_k\theta)\right). \tag{4}$$

In generation k the circuits represented by Eq. (3) and Eq. (4) are sampled sufficiently many times to generate estimates $\hat{P}_{\mathrm{c},k}$ and $\hat{P}_{\mathrm{s},k}$ of $P_{\mathrm{c},k}$ and $P_{\mathrm{s},k}$, respectively, from the relative frequencies of 0 and 1 measurement outcomes. We do not specify the number of samples that should be taken for each circuit in the protocol, since this has been addressed in previous works [9, 19], and our consistency tests are agnostic to the sampling schedule. The estimates $\hat{P}_{\mathrm{c},k}$ and $\hat{P}_{\mathrm{s},k}$ may be reinserted into Eq. (3) and Eq. (4) to give us a set of candidate estimates

of θ that are compatible with the experimental data,

$$\Theta_k = \left\{ \tilde{\theta} \in [0, 2\pi) \;\middle|\; \exists \hat{\lambda} > 0 \;: \right. \\ \left. \hat{\lambda} \begin{pmatrix} \cos(N_k \tilde{\theta}) \\ \sin(N_k \tilde{\theta}) \end{pmatrix} = \begin{pmatrix} 2\hat{P}_{\mathrm{c},k} - 1 \\ 2\hat{P}_{\mathrm{s},k} - 1 \end{pmatrix} \right\}. \quad (5)$$

The $\hat{\lambda}$ serves to normalize the probabilities, and we will provide a more thorough interpretation of this value in Section IV. Henceforth, all angles are implicitly assumed to be defined modulo 2π. Eq. 5 can be rewritten as

$$\Theta_k = \left\{ \operatorname{atan2}\left(2\hat{P}_{\mathrm{s},k} - 1, 2\hat{P}_{\mathrm{c},k} - 1\right)/N_k + \frac{2\pi n}{N_k} \;\middle|\; n \in \mathbb{Z} \right\}, \quad (6)$$

where n indexes the choice of branch in the branch cut for the arctangent. If $N_k = 1$, this set contains a unique estimate for θ. More generally for $N_k \geq 1$, there are N_k candidate estimates, one in each angular interval $\left((2n-1)\frac{\pi}{N_k}, (2n+1)\frac{\pi}{N_k}\right]$, for $0 \leq n < N_k$. See the dashed red lines in Fig. 1 for a graphical illustration of the angles in this set.

We want to select a single estimate, $\hat{\theta}_k$, from Θ_k at each generation. The criterion used by RPE is to successively choose the $\hat{\theta}_k$ that is closest to the previous estimate $\hat{\theta}_{k-1}$. If we assume that the initial estimate is unique, i.e., that $N_0 = 1$, this is possible with probability 1[24]. To determine which angle is closest to the previous estimate we need a branch cut-independent metric for measuring the distance between angles,

$$|\theta' - \theta''|_{2\pi} = \min\left\{ |\theta' - \theta'' + 2\pi n| \;\middle|\; n \in \mathbb{Z} \right\}. \quad (7)$$

We extend this metric to define the distance from any single angle θ' to a set of angles Θ,

$$d(\theta', \Theta) = \min_{\tilde{\theta} \in \Theta} \left| \tilde{\theta} - \theta' \right|_{2\pi}, \quad (8)$$

the minimizer of which is [25]

$$\mathcal{M}(\theta', \Theta) = \operatorname*{argmin}_{\tilde{\theta} \in \Theta} \left| \tilde{\theta} - \theta' \right|_{2\pi}. \quad (9)$$

Using these definitions, Algorithm 1 states the RPE protocol. In particular, we will henceforth assume that $N_0 = 1$, and $N_{k+1} > N_k$.

While the above description of RPE is made in terms of the selection of the closest angle to the previous selection as in [15], one can consider a procedure where a set of estimate angles is updated at each generation. In Appendix D, we address this formulation of RPE and find that it has equivalent error tolerance to the single-angle approach.

Also, as mentioned above, our goal is not to calculate the resources required from first principles, but to determine if the resources actually used in an experiment are sufficient, given uncharacterized noise in the system.

Algorithm 1 Robust Phase Estimation

Input:
1: $\{\Theta_{k'}\}_{k' \leq k_{\max}}$, the list of candidate estimates for each generation, Eq. (6)
Output:
2: $\hat{\theta}_{k_{\max}}$, the estimate for the underlying angle
Preconditions:
3: $N_{k'} < N_k$ for $k' < k$
4: $N_0 = 1$
Code:
5: **function** RobustPhaseEstimation($\{\Theta_{k'}\}_{k' \leq k_{\max}}$)
6: $k \leftarrow 0$
7: $\hat{\theta}_0 \leftarrow$ the unique element in Θ_0.
8: **while** $k + 1 \leq k_{\max}$ **do**
9: $k \leftarrow k + 1$
10: $\hat{\theta}_k \leftarrow \mathcal{M}(\hat{\theta}_{k-1}, \Theta_k)$
11: **return** $\hat{\theta}_{k_{\max}}$

However, any RPE-like protocol can still achieve Heisenberg scaling in the presence of bounded noise when the number of samples increases by a constant factor, as we show in Appendix B [26]. Thus for optimal efficiency, we suggest starting with a number of samples as prescribed by Refs. [9, 19], which perform a detailed analysis in the error-free case, and then scaling the number of samples by an amount that you believe will overcome your errors according to the scaling of Appendix B. Then use the consistency tests we lay out in the next section to check if the number of samples has been increased sufficiently to overcome the actual errors, or to test at what point in the protocol the noise becomes too large to compensate for.

We suspect that many experimentalists will not actually use the complex sampling schedules suggested in Refs. [9, 19], but rather will take a constant number of samples at each generation, as this schedule achieves near optimal resource efficiency, while being much simpler. In fact, this is what we do in our own numerical simulations. Both the consistency tests presented here, as well as the analysis in Appendix B, can be applied to any sampling schedule.

III. THE SUCCESS AND FAILURE OF RPE

Ideally, the value of $\hat{\theta}_k$ closest to θ would be chosen at each generation. Then $\hat{\theta}_k$ would estimate θ with error at most π/N_k because the N_k elements of Θ_k are equally spaced around the unit circle:

$$\hat{\theta}_k = \mathcal{M}(\theta, \Theta_k) \quad \Leftrightarrow \quad |\hat{\theta}_k - \theta|_{2\pi} < \frac{\pi}{N_k} \quad (10)$$

(see Fig. 1 for an illustration). If, for any reason, the RPE procedure selects a value of $\hat{\theta}_k$ that does not satisfy (10) at some generation k, we say that the procedure failed at generation k.

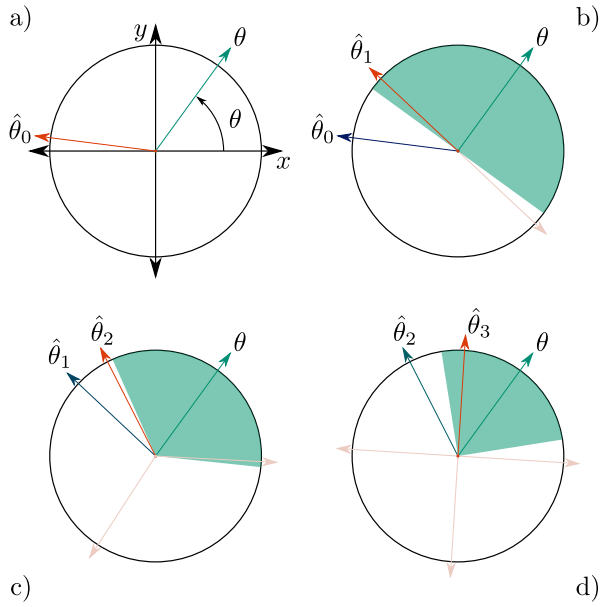

FIG. 2. Illustration of the action of the RPE protocol that fails at intermediate generations, but succeeds at the final generation. Red arrows are elements of Θ_k, and the green arrow is the correct angle be estimated, θ. (a) and (b) As in Fig. 1. (c) An incorrect choice of $\hat{\theta}_2$ occurs—the candidate that is closest to θ was not chosen. (d) Despite this, at generation 3, the correct choice of $\hat{\theta}_3$ is still made.

Generally, we would like to know whether RPE has failed at any given generation. However, to know this with certainty would require knowledge of θ (required to evaluate (10)), which is the parameter being sought. Instead, we develop heuristic consistency checks that evaluate some related conditions but are experimentally accessible, in order to approximate the maximum value of k for which (10) holds.

Note that in order for RPE to succeed at some generation, it is not necessary in general for it to have succeeded at all previous generations. This can happen when RPE fails due to a sufficiently large error in one generation, which is corrected by another sufficiently large error in a subsequent generation. An example illustrating this is shown in Fig. 2. However, such a success mode is not trustworthy, since it depends on the confluence of two errors, each of which would on its own be sufficient to induce failure.

All of the criteria for our checks are based on different notions of *consistency*. These notions are all based on properties that are satisfied in an ideal RPE run, but might fail to be realized in an RPE run in the presence of noise. The first five criteria are based on increasingly stringent constraints on the inter-generational consistency of $\hat{\theta}_k$. We define another criterion based on a condition on the inter-generational probability estimates. Finally, we consider a criterion based on intersequence consistency across different full RPE runs. These definitions provide a series of tests that can be applied to

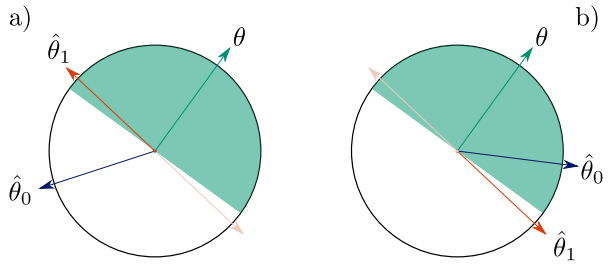

FIG. 3. Pathology in Φ_k: increasing the accuracy of θ_{k-1} can make θ_k *incorrect*. (a) The correct $\hat{\theta}_1$ is chosen, despite the large error in $\hat{\theta}_0$. (b) Even though $\hat{\theta}_0$ is closer to the true angle θ than in (a), it is closer to the incorrect candidate for generation 1, leading to the incorrect choice of $\hat{\theta}_1$. In (a), paradoxically, θ_0 laying further from θ, caused the correct selection of θ_1 for generation 1.

data from an RPE experiment in order to approximately determine its failure point, i.e., the generation at which (10) cease to be satisfied, and we should no longer trust the estimates.

Our first criterion is that there is some "plausible angle" $\tilde{\theta}$ that satisfies Eq. (10) at *every* generation, ruling out the situation in Fig. 2, as well as more typical failure modes like large drift in the estimates.

Criterion 1 (plausible consistency)**.** Consider the set of angles

$$\Phi_k = \left\{ \tilde{\theta} \;\middle|\; \hat{\theta}_k = \mathcal{M}(\tilde{\theta}, \Theta_k) \right\} \tag{11}$$

$$= \left\{ \tilde{\theta} \;\middle|\; |\hat{\theta}_k - \tilde{\theta}|_{2\pi} = d(\tilde{\theta}, \Theta_k) \right\} \tag{12}$$

$$= \left\{ \tilde{\theta} \;\middle|\; |\hat{\theta}_k - \tilde{\theta}|_{2\pi} < \pi/N_k \right\}, \tag{13}$$

i.e., angles for which the choice of $\hat{\theta}_k$ would *not* constitute a failure based on Eq. (10) at generation k [27]. We then check whether such an angle exists in common to all generations, giving us our first criterion:

$$\bigcap_{k=0}^{k_{\max}} \Phi_k \neq \emptyset. \tag{14}$$

Notice that there is no dependence on θ for this criterion, and $\{\Phi_k\}_{k \leq k_{\max}}$ are experimentally derivable quantities.

Because Φ_k are intervals (with size less than π for $k > 0$), their intersection is an interval, which permits efficient classical testing of membership in the intersection. For the later criteria, care will have to be taken to ensure that criterion satisfaction is efficiently testable.

Unfortunately, satisfying Eq. (14) does not provide particularly strong guarantees. The condition is trivially satisfied if $N_k \geq 2N_{k-1}$ for all k (see Remark 4 in Appendix A). Also, the criterion fails to rule out the paradoxical scenario of Fig. 3, in which *more reliable* data can

lead to an incorrect choice of angle. Our next criterion addresses these two concerns by relying on the distance to the *set* Θ_k, rather than the *point* $\hat{\theta}_k$.

Criterion 2 (consecutive consistency)**.** Consider the sets Λ_k, where $\Lambda_0 = [0, 2\pi)$ and

$$\Lambda_k = \left\{ \tilde{\theta} \;\middle|\; d(\tilde{\theta}, \Theta_k) + d(\tilde{\theta}, \Theta_{k-1}) < \frac{\pi}{N_k} \right\} \tag{15}$$

for $k = 1, 2, \ldots$. An angle $\tilde{\theta}$ is in Λ_k if there exist measurements in both Θ_k and Θ_{k-1} that are close to $\tilde{\theta}$. As before, the corresponding consistency test is whether the intersection of the sets Λ_k is nonempty:

$$\bigcap_{k=0}^{k_{\max}} \Lambda_k \neq \emptyset. \tag{16}$$

Criterion 2 is stronger than Criterion 1:

Theorem 1. *The sets* $\{\Phi_k\}_{k \leq k_{max}}$ *and* $\{\Lambda_k\}_{k \leq k_{max}}$ *satisfy:*

$$\bigcap_{k' \leq k} \Lambda_{k'} \subseteq \bigcap_{k' \leq k} \Phi_{k'}. \tag{17}$$

Hence, Eq. (16) *implies Eq.* (14)*.*

The proof may be found in Appendix A. Moreover, for a sequence that satisfies Criterion 2 and Eq. (10) for all $k' < k$, further reduction of $d(\theta, \Theta_{k'})$ will only improve the estimate $\hat{\theta}_k$ (see Corollary 1.1 in Appendix A).

Unlike Φ_k, Λ_k is *not* an interval (for $k > 0$), and testing for membership could introduce exponential classical overhead (since Λ_k is the union of N_k intervals). Fortunately, $\Lambda_k \cap \Phi_k$, and hence $\bigcap_{k' \leq k} \Lambda_{k'}$, *is* an interval:

$$\Lambda_k \cap \Phi_k = \left\{ \theta \;\middle|\; d\left(\theta, (\hat{\theta}_k, \hat{\theta}_{k-1})\right) < D_k \right\}, \tag{18}$$

where $D_k = \frac{\pi}{2N_k} - \frac{1}{2} \left| \hat{\theta}_k - \hat{\theta}_{k-1} \right|_{2\pi}$. The interval is, equivalently, the smallest angular interval containing $\hat{\theta}_k$ and $\hat{\theta}_{k-1}$, expanded by D_k on both sides. Refer to Lemma 1 in Appendix A and the reference Python implementation for further details.

Criterion 2 is based on balancing errors on adjacent generations. While adaptive protocols, such as [28], include detection and adjustment for errors between generations, standard RPE is non-adaptive and makes no such adjustments. Thus a criterion that allows a small error at one generation to compensate for a larger error on another, independent of the pre-determined number of samples to take at each generation, should raise suspicion. This motivates a third, still stronger criterion which forbids balancing errors across subsequent generations.

Criterion 3 (local consistency)**.** Consider a set of angular error bounds on the sequence, $\{\delta\theta_k\}_{k \leq k_{\max}}$, which should satisfy Eq. (21) but may otherwise may be freely chosen. At generation k, the set of angles within those bounds is

$$\Delta_k[\delta\theta_k] = \left\{ \tilde{\theta} \;\middle|\; d(\tilde{\theta}, \Theta_k) < \frac{\delta\theta_k}{N_k} \right\}. \tag{19}$$

We say that a sequence is $(\delta\theta_k)$-locally-consistent if the intersection across all generations is non-empty:

$$\bigcap_{k \leq k_{\max}} \Delta_k[\delta\theta_k] \neq \emptyset. \tag{20}$$

Local consistency is stronger than Criterion 2 if

$$\frac{\delta\theta_k}{N_k} + \frac{\delta\theta_{k-1}}{N_{k-1}} \leq \frac{\pi}{N_k}, \tag{21}$$

since in that case we have

$$\bigcap_{k' \leq k} \Delta_{k'}[\delta\theta_{k'}] \subseteq \bigcap_{k' \leq k} \Lambda_{k'} \subseteq \bigcap_{k' \leq k} \Phi_{k'}. \tag{22}$$

Like Λ_k, Δ_k is not necessarily an interval, but its union is if the set inclusion Eq. (22) holds (see Remark 1), so we demand that Eq. (21) be satisfied. In this case, an interval formulation of this criterion follows directly from Eq. (19):

$$\bigcap_{k' \leq k} \Delta_{k'}[\delta\theta_{k'}] = \bigcap_{k' \leq k} \left\{ \tilde{\theta} \;\middle|\; |\tilde{\theta} - \hat{\theta}_{k'}|_{2\pi} < \frac{\delta\theta_{k'}}{N_{k'}} \right\} \tag{23}$$

Criterion 4 (uniform-local consistency)**.** From Criterion 3, we can obtain a special case for which $\delta\theta_k = \delta\theta$ is independent of k:

$$\delta\theta_k = \frac{\pi}{1 + r}, \tag{24}$$

where $r \geq \frac{N_k}{N_{k-1}}$ for all k; this $\delta\theta_k$ satisfies Eq. (21).

This schedule is intended to be used when N_k grows roughly exponentially in k, i.e., $N_k \approx r^k$. Also, notice that $r \geq 2$, since $N_1 \geq 2$. It balances the error tolerance between generations, so it is a natural choice when the expected error on each generation is the same, which could occur for example when SPAM error is independent of generation and dominates error associated with implementing $U_{\mathrm{t}}^{N_k}$. Note that for the standard RPE case in which $N_k = 2^k$ [14],

$$\delta\theta_k = \frac{\pi}{3}. \tag{25}$$

Criterion 3 (including the special case Criterion 4) leads to a bound on the errors in the estimates of the probabilities of Eq. (3) and Eq. (4). A straightforward geometrical argument — found in, for example, [9] (the "simple geometry" leading to Eq. 2) or [29] (Theorem 4.1) — shows that

$$\max\{|\hat{P}_{\mathrm{c},k} - P_{\mathrm{c},k}|, |\hat{P}_{\mathrm{s},k} - P_{\mathrm{s},k}|\} = \Delta P \leq \frac{\sin(\delta\theta_k)}{2\sqrt{2}}. \tag{26}$$

Criteria of fundamental value

#	Description	Set	Criterion	Interval
1	plausible	$\Phi_k = \left\{\tilde{\theta} \;\middle|\; \hat{\theta}_k = \mathcal{M}(\tilde{\theta}, \Theta_k)\right\}$	$\bigcap_{k\le k_{\max}} \Phi_k \neq \emptyset$	$\bigcap_{k'\le k} \left\{\tilde{\theta} \;\middle|\; \lvert\tilde{\theta} - \hat{\theta}_{k'}\rvert_{2\pi} < \frac{\pi}{N_{k'}}\right\}$
2	consecutive	$\Lambda_k = \left\{\tilde{\theta} \;\middle|\; d(\tilde{\theta}, \Theta_k) + d(\tilde{\theta}, \Theta_{k-1}) < \frac{\pi}{N_k}\right\}$	$\bigcap_{k\le k_{\max}} \Lambda_k \neq \emptyset$	$\bigcap_{k'\le k} \left\{\theta \;\middle|\; d\left(\theta, (\hat{\theta}_k, \hat{\theta}_{k-1})\right) < D_k\right\}$ $D_{k'} = \frac{\pi}{2N_{k'}} - \frac{1}{2}\lvert\hat{\theta}_{k'} - \hat{\theta}_{k'-1}\rvert_{2\pi}$
3	local	$\Delta_k[\delta\theta_k] = \left\{\tilde{\theta} \;\middle|\; d(\tilde{\theta}, \Theta_k) < \frac{\delta\theta_k}{N_k}\right\}$ s.t. $\frac{\delta\theta_k}{N_k} + \frac{\delta\theta_{k-1}}{N_{k-1}} \le \frac{\pi}{N_k}$	$\bigcap_{k\le k_{\max}} \Delta_k[\delta\theta_k] \neq \emptyset$	$\bigcap_{k'\le k} \left\{\tilde{\theta} \;\middle|\; \lvert\tilde{\theta} - \hat{\theta}_{k'}\rvert_{2\pi} < \frac{\delta\theta_{k'}}{N_{k'}}\right\}$

Criteria useful as tests

#	Description		Criterion
4	uniform local		$\bigcap_{k\le k_{\max}} \Delta_k[\pi(1 + N_k/N_{k-1})] \neq \emptyset$
5	angular-historical	$\forall k \le k_{\max}$	$\hat{\theta}_k \in \bigcap_{k'\le k} \Delta_{k'}[\delta\theta_{k'}]$
6	probability-historical	$\forall k' < k \le k_{\max}$	$\max\left\{\lvert\sin N_{k'}\hat{\theta}_{k'} - \sin N_k\hat{\theta}_k\rvert, \lvert\cos N_{k'}\hat{\theta}_{k'} - \cos N_k\hat{\theta}_k\rvert\right\} \le \frac{\sin(\delta\theta_k)}{\sqrt{2}}$
7	intersequence	$\forall k < k_{\max}$	$\lvert\hat{\theta}_k - \hat{\theta}'_k\rvert_{2\pi} \le \frac{2\pi}{N_k}$

TABLE I. Overview of consistency checks of RPE success. The criteria in the first table (of "fundamental value") are used to characterize basic convergence properties of RPE, and include Criteria 1, 2, and (to a lesser extent) 3. Notice they form a logical hierarchy, with later criteria strictly stronger than earlier ones. The second table lists the remainder of the criteria discussed, which we find have value for empirically testing the validity of RPE results. There is no straightforward logical hierarchy among these final 4 criteria.

In particular, to satisfy the hypothesis for uniform-local consistency, with $N_k = 2^k$, we require

$$\Delta P_k \le \sqrt{\frac{3}{32}} \sim 30.6\%, \tag{27}$$

which corrects the numerical value given in [14].

The previous tests are based on the existence of an angle $\tilde{\theta}$ that is consistent with $\{\Theta_k\}_{k\le k_{\max}}$, but do not indicate whether the final output $\hat{\theta}_k$ is such a witness. The following criterion asserts precisely this.

Criterion 5 (angular-historical consistency)**.** For all $k \le k_{\max}$,

$$\hat{\theta}_k \in \bigcap_{k'\le k} \Delta_{k'}[\delta\theta_{k'}]. \tag{28}$$

If this condition holds, the value of $\hat{\theta}_{k_{\max}}$ is not just the terminating measurement in a reasonable RPE measurement sequence, it is also one of the putative underlying angles. In Theorem 3, we show that, if $\delta\theta_{k'-1}/N_{k'-1} > \delta\theta_{k'}/N_{k'}$ for all $k' \le k$ (which holds for $N_k = 2^k$), (28) is equivalent to requiring that the intersection has length greater than $\delta\theta_k/N_k$ for all $k \le k_{\max}$:

$$\left|\bigcap_{k'\le k} \Delta_{k'}[\delta\theta_{k'}]\right| > \frac{L}{N_k}, \tag{29}$$

where $L = \delta\theta_k$. One might imagine fine-tuning the interval length, L, to optimize the performance of the consistency test in identifying the actual failure point. In Section IV, we provide numerical evidence that although such fine-tuning may be possible, it is too sensitive to the error model and the value of the actual angle θ to provide a consistent advantage over using $L = \delta\theta_k$.

The previous five criteria form a hierarchy of consistency checks that are increasingly stringent. We also consider two other criteria that do not strictly fit into this hierarchy. The first is to directly test the probability condition Eq. (26) for each estimate:

Criterion 6 (probability-historical consistency)**.** For all $k' < k \le k_{\max}$,

$$\lvert\sin N_{k'}\hat{\theta}_{k'} - \sin N_k\hat{\theta}_k\rvert \le \frac{\sin(\delta\theta_k)}{\sqrt{2}}, \tag{30}$$

and

$$\lvert\cos N_{k'}\hat{\theta}_{k'} - \cos N_k\hat{\theta}_k\rvert \le \frac{\sin(\delta\theta_k)}{\sqrt{2}}. \tag{31}$$

In other words, probability-historical consistency tests that the probabilities expected from each estimated phase (from Eq. (3) and Eq. (4)) are consistent with the probabilities obtained in previous generations, up to the $\delta\theta_k$ bounds as given in Eq. (26). Because this test is expressed in terms of probabilities rather than angles, it will turn out to be overly pessimistic in the presence of incoherent noise (see Section IV), but it does provide a conservative estimate of the failure point.

Our final test is to compare the results of RPE originating from different sequences of N_k:

Criterion 7 (intersequence consistency)**.** For two sequences of RPE estimates $\hat{\theta}_k$ and $\hat{\theta}'_k$ with sequences $N_k \leq N'_k$, check that

$$|\hat{\theta}_k - \hat{\theta}'_k|_{2\pi} \leq \frac{2\pi}{N_k} \tag{32}$$

for all $k \leq k_{\max}$.

Instead of looking at the single original sequence N_k, and checking for self-consistency, we consider a second sequence N'_k, and check that the resulting estimates are consistent with those of the original sequence. Notice first that if Eq. (32) fails for some generation k, then either

$$|\hat{\theta}_k - \theta|_{2\pi} > \frac{\pi}{N_k} \tag{33}$$

or

$$|\hat{\theta}'_k - \theta|_{2\pi} > \frac{\pi}{N_k} > \frac{\pi}{N'_k}, \tag{34}$$

i.e., at least one of the sequences does not satisfy the true condition for correctness, Eq. (10). Using the notation Φ_k (and Φ'_k, respectively) of Eq. (11), the intersequence consistency condition Eq. (32) is equivalent to

$$\Phi_k \cap \Phi'_k \neq \emptyset. \tag{35}$$

In other words, the intersequence consistency check tests whether there exist any plausible estimates that are consistent with both sequences. In Section IV we find that this approach can provide a very good test of data, but it does of course require an extra sequence's worth of additional experimental data.

We thus have a range of consistency checks that we can use to gain information about the performance of an RPE run. The consistency checks are summarized in Table I. In the next section, we test and compare the performance of these consistency checks by simulating noisy RPE runs.

IV. NUMERICAL PERFORMANCE OF SELF-CONSISTENT CRITERIA

In the previous section, we defined several heuristic consistency tests for an RPE experiment. In this section, we evaluate the performance of these tests by numerically simulating RPE runs with depolarizing, dephasing, and amplitude damping noise.

When one of our consistency tests fails, it flags a generation $k \in \{0, 1, 2, ..., k_{\max}\}$ at which the RPE estimate becomes unreliable. Because we are performing a numerical simulation for a particular target angle θ, the heuristic's failure generation can be compared to the actual failure generation in which RPE is no longer able to correctly estimate θ (i.e., when it fails the condition (10)).

Any single-qubit mixed state ρ may be represented as

$$\rho = \frac{1}{2}\left(\mathbb{1} + x\sigma_x + y\sigma_y + z\sigma_z\right), \tag{36}$$

for $x, y, z \in \mathbb{R}$ satisfying $x^2 + y^2 + z^2 \leq 1$, and Pauli matrices $\sigma_x, \sigma_y, \sigma_z$. For notational and computational convenience we instead represent ρ as a Pauli vector $|\rho\rangle\!\rangle$ (i.e., an element of Hilbert-Schmidt space):

$$|\rho\rangle\!\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\x\\y\\z\end{pmatrix}. \tag{37}$$

Quantum operations are then implemented as superoperators that act on $|\rho\rangle\!\rangle$ by matrix multiplication.

We assume that $U_{\rm t}$ is a rotation about σ_x by some angle θ, the parameter we wish to estimate. We represent $U_{\rm t}$ as a superoperator:

$$U_{\rm t} = R_x(\theta) = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \cos(\theta) & -\sin(\theta)\\ 0 & 0 & \sin(\theta) & \cos(\theta)\end{pmatrix}. \tag{38}$$

Let $|0\rangle$ and $|1\rangle$ denote the σ_z eigenstates with eigenvalues $+1$ and -1, respectively. Ideally, the initial state is $|0\rangle$:

$$\rho_{\rm init} = |0\rangle\langle 0|, \quad |\rho_{\rm init}\rangle\!\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \tag{39}$$

and the measurements of the cosine and sine strings are ideally of the states

$$\begin{aligned}\rho_{\rm c} = |0\rangle\langle 0|, |\rho_{\rm c}\rangle\!\rangle &= \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix},\\ \rho_{\rm s} = |y_+\rangle\langle y_+|, |\rho_{\rm s}\rangle\!\rangle &= \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\\0\end{pmatrix}.\end{aligned} \tag{40}$$

Denoting the realistic noisy (rather than ideal) forms of these quantities with tildes, the probabilities of the cosine and sine measurements are

$$\begin{pmatrix}P_{{\rm c},k}\\ P_{{\rm s},k}\end{pmatrix} = \begin{pmatrix}\langle\!\langle\tilde{\rho}_{\rm c}|\,\tilde{U}_k\,|\tilde{\rho}_{\rm init}\rangle\!\rangle\\ \langle\!\langle\tilde{\rho}_{\rm s}|\,\tilde{U}_k\,|\tilde{\rho}_{\rm init}\rangle\!\rangle\end{pmatrix}. \tag{41}$$

As in Eq. (5), these probabilities are reparameterized as

$$\begin{pmatrix}2P_{{\rm c},k} - 1\\ 2P_{{\rm s},k} - 1\end{pmatrix} = \begin{pmatrix}\lambda\cos\phi\\ \lambda\sin\phi\end{pmatrix}, \tag{42}$$

where ϕ is the maximum likelihood estimate of the angle $N_k\theta$. Notice that $\hat{\lambda}$ of Eq. (5) serves to normalize the

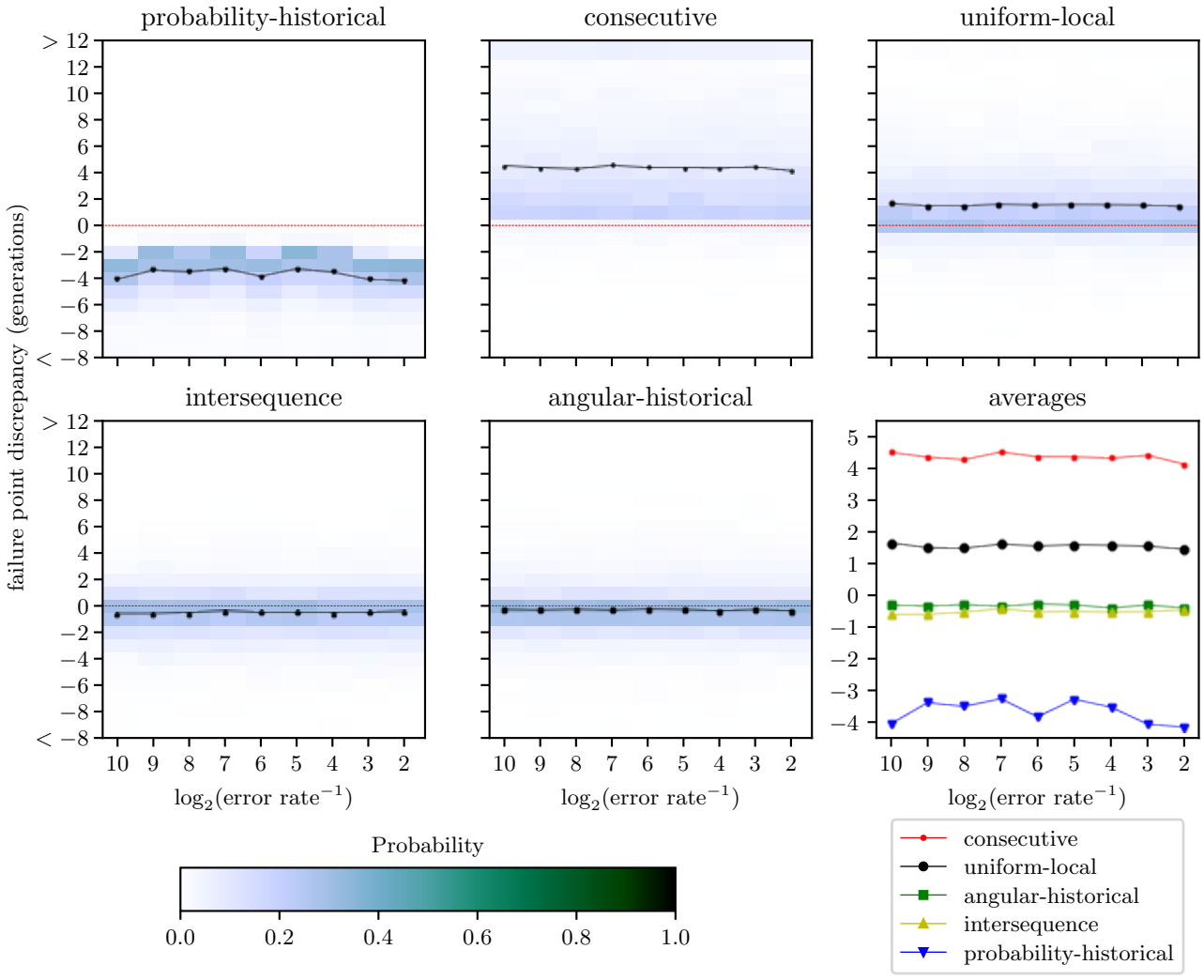

FIG. 4. For varying depolarization error rates and actual angle $\theta = 1.6$, plots the discrepancies between the generations at which each consistency check flagged failure, and the generations at which failure actually occurred. Positive values indicate that the consistency check flagged failure *after* failure actually occurred. 1000 RPE runs were performed per error rate. Bin colors show the proportion of the runs at that error rate they contain. The curves show averages for each error rate. The lower right plot shows the average curves from all of the other plots, for comparison. The dotted red line at discrepancy 0 is placed to help guide the eye.

observed frequencies, while λ above analogously normalizes the expected *probabilities*. In Appendix C, we show that λ renormalizes the measurement counts, $M \mapsto M\lambda^2$. The renormalization occurs when a protocol makes use of only ϕ information, so RPE and the consistency checks—with the exception of the probability-historical consistency check—all have this scaling behavior. In order to simulate depolarization, dephasing, or amplitude damping, after each application of the unitary U_t of Eq. (38) we apply an error operator V defined by the error type and the error rate b, i.e., $\tilde{U}_k = (VU_t)^{N_k}$.

For depolarization, the superoperator is

$$V_{\text{depol.}}(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-b & 0 & 0 \\ 0 & 0 & 1-b & 0 \\ 0 & 0 & 0 & 1-b \end{pmatrix}. \tag{43}$$

Notice that this superoperator commutes with U_t, leading to $\lambda = (1-b)^{N_k}$, while leaving $\phi = N_k\theta$ unaffected.

Dephasing in the $\sigma_x\sigma_y$-plane[30] results in the super-

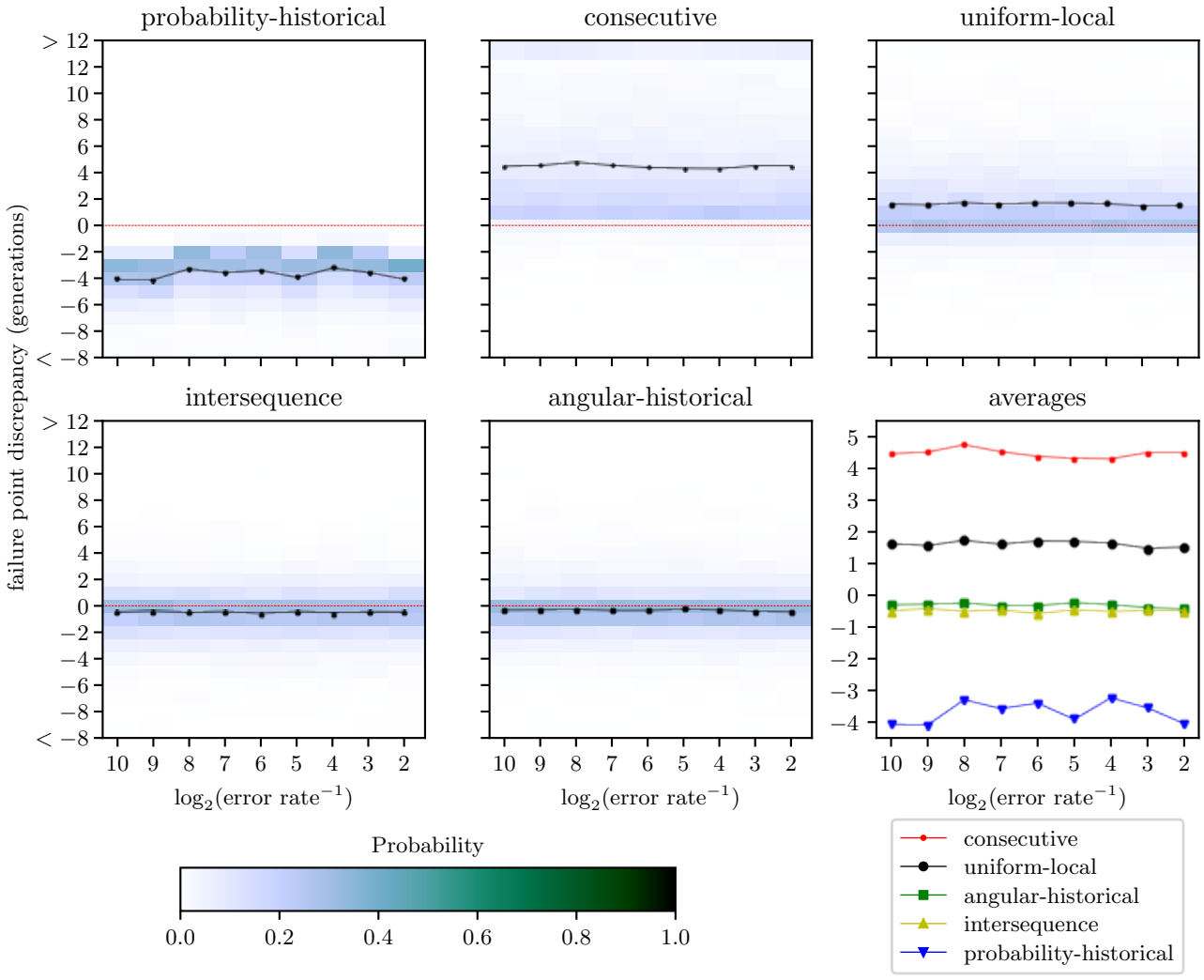

FIG. 5. For varying dephasing error rates and actual angle $\theta = 1.6$, plots the discrepancies between the generations at which each consistency check flagged failure, and the generations at which failure actually occurred. Positive values indicate that the consistency check flagged failure *after* failure actually occurred. 1000 RPE runs were performed per error rate. Bin colors show the proportion of the runs at that error rate they contain. The curves show averages for each error rate. The lower right plot shows the average curves from all of the other plots, for comparison. The dotted red line at discrepancy 0 is placed to help guide the eye.

operator:

$$V_{\text{depha.}}(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-b & 0 & 0 \\ 0 & 0 & 1-b & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{44}$$

In contrast to depolarizing noise, dephasing noise results in a nontrivial $\phi(k)$ and $\lambda(k)$ that we do not attempt to characterize analytically. Nevertheless, the performance of the consistency checks is still qualitatively the same, as seen in Fig. 5.

Finally, we simulate an amplitude damping channel, with decay from $|1\rangle$ to $|0\rangle$. This is described by the superoperator

$$V_{\text{a.d.}}(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-b} & 0 & 0 \\ 0 & 0 & \sqrt{1-b} & 0 \\ b & 0 & 0 & 1-b \end{pmatrix}. \tag{45}$$

Unlike the previous examples, the presence of the b term in the lower-left corner of the matrix—corresponding to the relaxation to the state $|0\rangle$—acts to drive the system to a particular steady state by adding a finite term to the z component of the Bloch vector at every application of

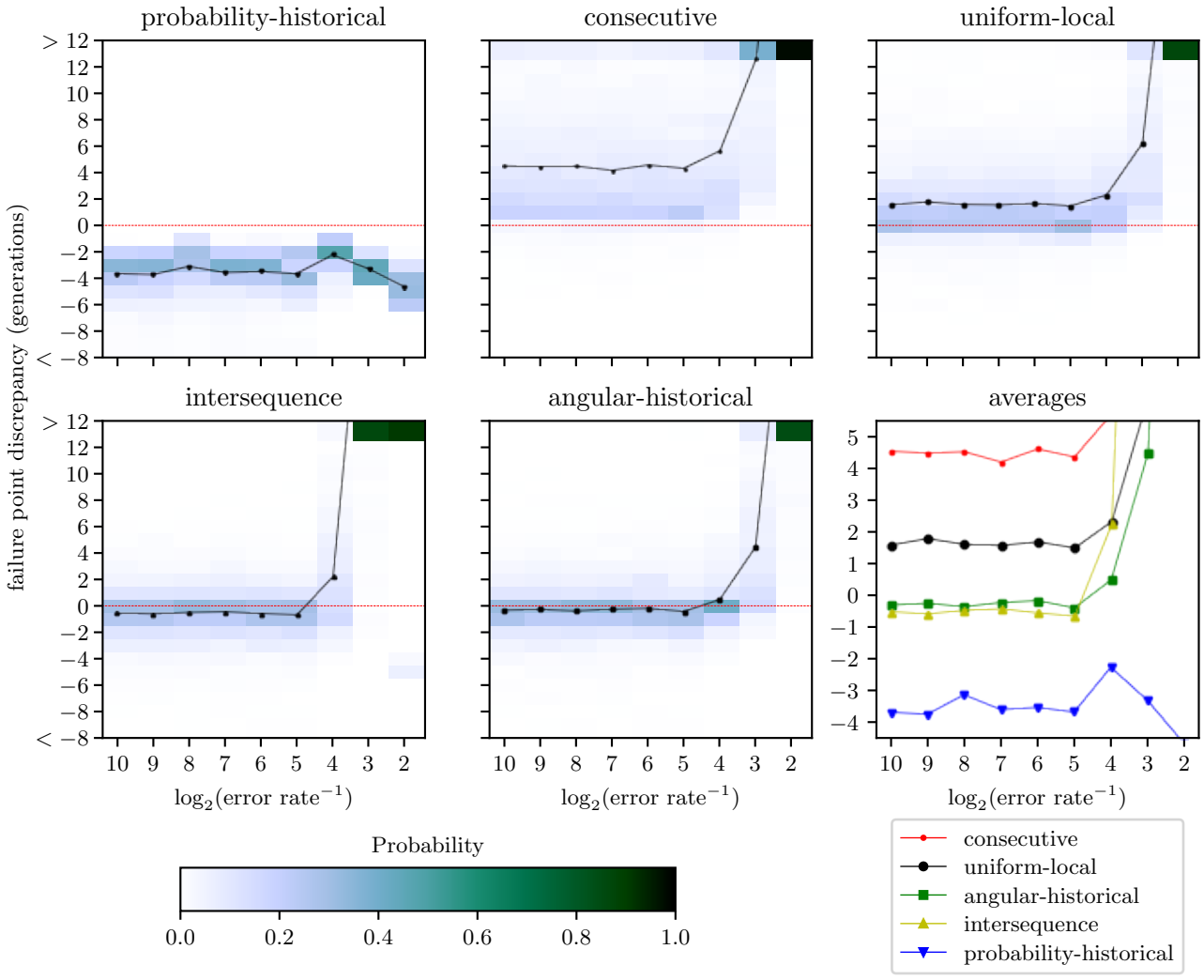

FIG. 6. For varying amplitude damping error rates and actual angle $\theta = 1.6$, plots the discrepancies between the generations at which each consistency check flagged failure, and the generations at which failure actually occurred. Positive values indicate that the consistency check flagged failure *after* failure actually occurred. 1000 RPE runs were performed per error rate. Bin colors show the proportion of the runs at that error rate they contain. The curves show averages for each error rate. The lower right plot shows the average curves from all of the other plots, for comparison. Very strong amplitude damping ($b \geq 1/16$) creates a false signal that RPE will track, as indicated by the residue of failures at generation > 12. This skews the average failure points in those generations towards the maximum probed failure generation of 45, complicating a direct analysis of the average. The dotted red line at discrepancy 0 is placed to help guide the eye.

$\tilde{U}$. The other terms in the noise model act as damping, so for sufficiently large N_k, the system evolves towards a particular fixed $\lambda(\infty)$ and $\phi(\infty)$, which we again do not attempt to determine analytically. Notice that λ increases with b in this scenario, because b serves as the amplitude of the driving term added at every generation. The overall result is that, if the amplitude damping is strong enough, the statistical noise will become irrelevant at high generations, and RPE will begin to track the strong signal for $\phi(\infty)$. Hence any consistency checks that depends on ϕ information only will never flag a failure, because without information about the true angle there is no way to tell that the false signal $\phi(\infty)$ is incorrect. This results in a residue in Fig. 6 for failure discrepancies greater than 12 for strong amplitude damping (i.e., the consistency checks flag failure more than 12 generations after failure actually occurs). Notice that the probability-historical consistency check does detect the error because it is directly sensitive to the length of the Bloch vector λ. For this reason, in general the

probability-historical consistency test is more pessimistic than the other tests (i.e., it flags failure earlier). In the case of strong amplitude damping, this pessimism is justified, but we suggest caution, since other error models may not cause λ to shrink sufficiently to flag a failure. Note that if there is reason to believe that a small λ is an indication of overall infidelity of the system, one might also consider directly checking the magnitude of λ.

In all cases, we simulated RPE runs with each error model for exponentially spaced error rates

$$b = 2^{-i} \quad \text{for} \quad i = 2, \ldots, 10. \tag{46}$$

Additionally, SPAM error was introduced by setting

$$\tilde{\rho}_{\text{init.}} = V_{\text{depol.}}(b_{\text{SPAM}})\rho_{\text{init.}}, \tag{47}$$

and

$$\tilde{\rho}_{\text{c}} = V_{\text{depol.}}(b_{\text{SPAM}})\rho_{\text{c}}. \tag{48}$$

Because many implementations of RPE will use an additional gate to implement ρ_{s}, we injected error due to an imperfect $\pi/2$ rotation:

$$\tilde{\rho}_{\text{s}} = V_{\text{depol.}}(b_{\text{SPAM}})V_{\text{depol.}}(b_{\text{s}})R_x(b_{\text{s}})\rho_{\text{s}}. \tag{49}$$

We choose and present results for a fixed value of $b_{\text{SPAM}} = b_{\text{s}} = 10^{-2}$, but comment that the results are qualitatively the same with both set to zero, $b_{\text{SPAM}} = b_{\text{s}} = 0$. For each error rate and error type, we simulated 1000 runs of the RPE procedure, taking $M = 1000$ samples of the measurement outcomes at each generation, and repeating for a variety of angles θ. The results for

$$\theta = 1.6 \tag{50}$$

are shown in Fig. 4 to 6, with detailed plots of the distributions of discrepancies available in Appendix E. The primary generation sequence in all cases was $N_k = 2^k$, and for the intersequence consistency check the second sequence was $N_0 = 2$, $N_i = 3(2^{i-1})$ for $i \geq 1 = 1, 2, 3, ...$ (with a preliminary generation at $N_{-1} = 1$ that was used to determine a unique $\hat{\theta}_0$, and was not compared to the primary sequence).

Each plot in Fig. 4 to 6 compares the generation at which the heuristic consistency check flagged failure to the actual failure point of the run (as determined by (10)). In particular, we subtract the generation number where failure actually occurred from the generation number that was flagged by the consistency check. Thus positive values in Fig. 4 to 6 indicate that the actual failure occurred *before* failure was flagged by the given consistency check. Data was collected out to 45 generations (and 46 for the intersequence alternative test), and if no failure was detected, a failure point at the following generation was recorded. This choice can only affect the calculation of average failure point. In addition, this only becomes an issue for amplitude damping with strong error rates $b \geq 1/16$.

The results in Fig. 4 to 6 show that the angular-historical consistency check is on average the closest to the actual failure point, in all the cases we studied. (The angular-historical consistency check is found in the center of the bottom row of each figure.) Close behind it is the intersequence consistency check (which appears on the left of the bottom row of each figure). We therefore suggest that if one simply desires to estimate as accurately as possible the actual failure point, one should use the angular-historical consistency check. If one wants an additional verification layer for the resulting failure generations, one could compare these results to those of the intersequence consistency check (which, recall, requires taking a second set of data). Since the angular-historical and intersequence tests perform similarly for most cases of the error models studied in this paper, finding a large difference between them in an experiment would indicate that the underlying error model is outside the regimes investigated here, or is in one of the pathological cases for amplitude damping.

If one instead wants to obtain a conservative estimate of the failure point, i.e., an estimate that precedes the actual failure point with high probability, one should use the check for probability-historical consistency (which appears in the top left corner of each figure). As we can see from Fig. 4 to 6, in every run that we simulated, the check for probability-historical consistency flagged failure before failure actually occurred.

The results in Fig. 4 to 6 are only for $\theta = 1.6$, as noted above. However, their qualitative features appear to hold for almost any θ, the exception being small θ under amplitude damping. In particular, we see in Fig. 7 that at the error rate $b = 2^{-6}$, the angular-historical consistency check and the intersequence consistency check are the closest to correct on average, and the probability-historical consistency check flags failure early.

Testing angular-historical consistency in the case $N_k = 2^k$ amounts to checking that each intersection (29) has size at least $\frac{\pi}{3N_k}$, i.e.,

$$L = \delta\theta_k = \frac{\pi}{3} \tag{51}$$

(by (24)). The fact that (29) is equivalent to (28) for $L = \delta\theta_k$ suggests that this value of L should provide good performance of the consistency check, which is supported by Figs. 4, 5, and 6.

However, as discussed in the paragraph following (29), we could in principle build a heuristic consistency check around any value of L we like, rather than $L = \delta\theta_k$. As for any heuristic consistency check, performance will depend on the specific error model. Consequently, we tested the angular-historical consistency check for a variety of interval widths centered around $L = \delta\theta_k = \pi/3$; an example of the results is shown in Fig. 8. This and the plots for other actual angles show that although the angular-historical consistency check with interval width $L = \pi/3$ (as defined in Criterion 5) is close to optimal, it does flag failure early on average, so it might be pos-

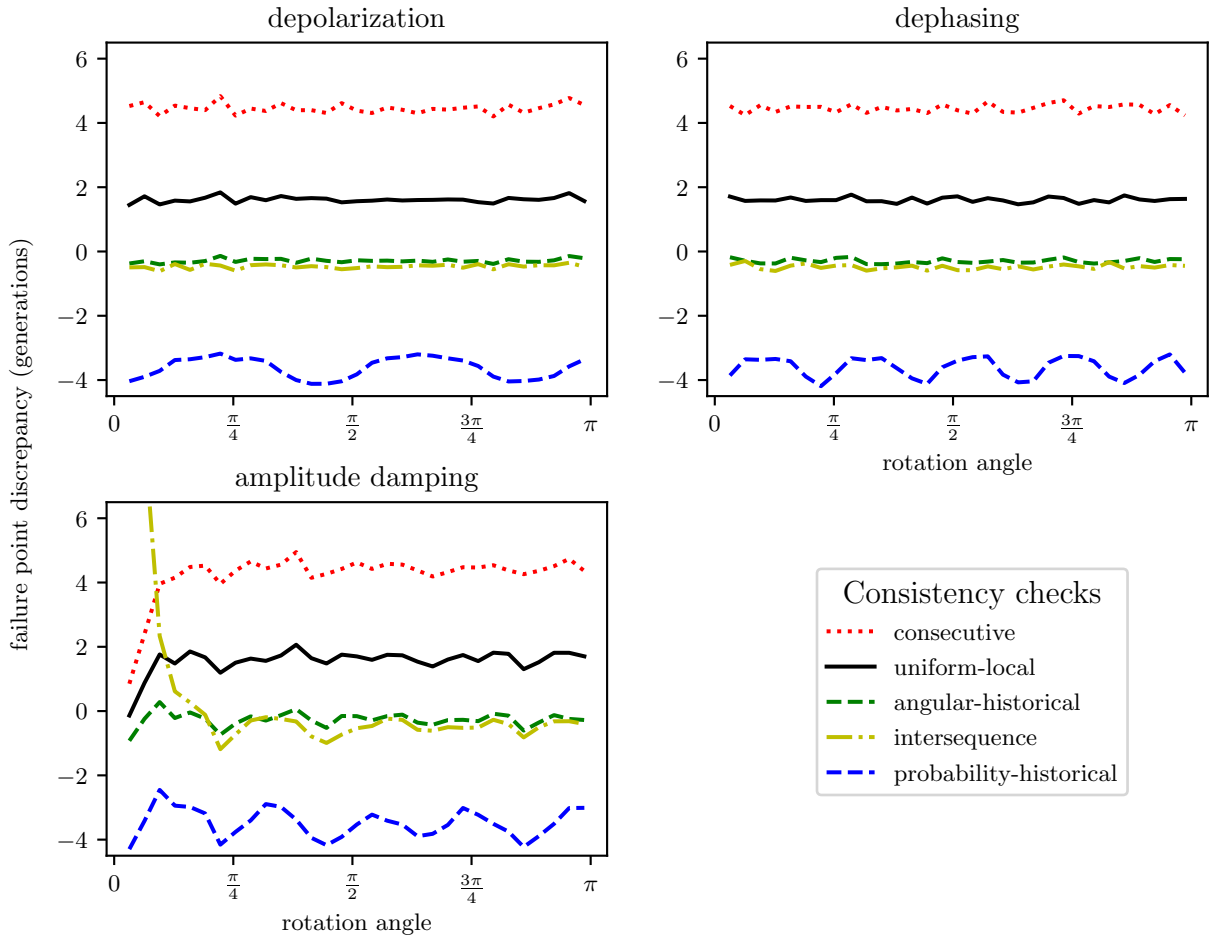

FIG. 7. Dependence of the failure point discrepancies on actual rotation angle for simulated depolarization, dephasing, and amplitude damping at the rate $b = 2^{-6}$. Each data point is an average of 1000 RPE runs. The most notable feature is that all consistency checks have near constant failure point discrepancies for all θ, with the exception of substantial deviations for small θ under amplitude damping.

sible to numerically fine-tune the interval width in order to obtain a more accurate check. However, doing so is sensitive to the specific error model and error rate as well as the actual angle, so the utility of this approach is probably limited, and we instead chose to stick to the theoretically-motivated width of $L = \pi/3$.

V. CONCLUSION

In this work we provided a framework for characterizing the consistency of RPE data based on a variety of efficiently classically verifiable criteria. The implementation of such consistency checks will allow an experimenter to address the worry that, due to systematic errors, their RPE run may have violated the assumptions that guarantee the protocol's performance. Such a violation might result in the protocol returning a dramatically incorrect estimate of the value of the desired parameter. We described seven such checks, and tested them numerically under simulated depolarization, dephasing, and amplitude damping, identifying the angular-interval-consistency check as the most accurate in all cases. This provides a tool that augments the standard RPE protocol by permitting detection of unknown errors with unknown rates, which would otherwise cause hidden failure of the standard RPE protocol.

ACKNOWLEDGMENTS

The authors thank the referees of *Physical Review A* for their thorough and careful read of our paper. W.

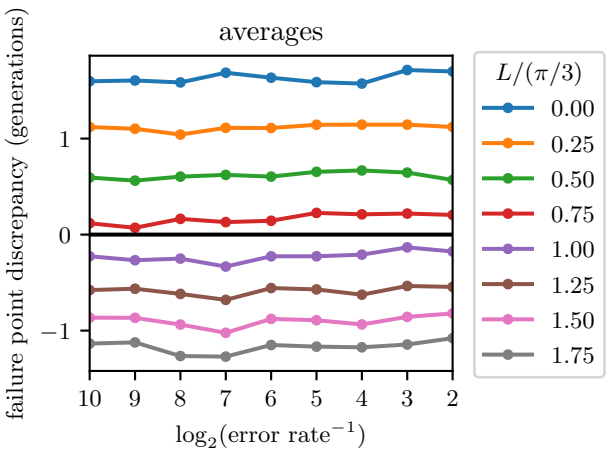

FIG. 8. For varying depolarization error rates and actual angle $\theta = 1.6$, plots the failure discrepancies for the angular-historical consistency check for various minimum interval widths (L, as in (29)). Positive values indicate that the consistency check flagged failure *after* failure actually occurred. Each point is averaged over 1000 runs. The solid black line at discrepancy 0 is placed to help guide the eye.

M. K. acknowledges support from the NSF, Grant No. DGE-1842474, and the NSF STAQ project, Grant No. PHY-1818914. This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Algorithms Team and Quantum Computing Applications Team programs. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for DOE's National Nuclear Security Administration under contract DE-NA0003525.

[1] A. Y. Kitaev, arXiv preprint quant-ph/9511026 (1995).
[2] P. W. Shor, in *Proceedings 35th annual symposium on foundations of computer science* (Ieee, 1994) pp. 124–134.
[3] R. B. Griffiths and C.-S. Niu, Physical Review Letters **76**, 3228 (1996).
[4] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, Phys. Rev. X **6**, 031007 (2016).
[5] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, and M. G. Thompson, Phys. Rev. Lett. **118**, 100503 (2017).
[6] P. M. M. Q. da Cruz, G. Catarina, R. Gautier, and J. Fernández-Rossier, Quantum Science and Technology (2020).
[7] T. E. O'Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean, and R. Babbush, arXiv preprint (2020), arXiv:2010.02538 [quant-ph].
[8] P. Dong and Z.-L. Cao, Journal of Physics: Condensed Matter **19**, 376216 (2007).
[9] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, New Journal of Physics **11**, 073023 (2009).
[10] P. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, *et al.*, Physical Review X **6**, 031007 (2016).
[11] J. Helsen, F. Battistel, and B. M. Terhal, npj Quantum Information **5**, 74 (2019).
[12] T. E. O'Brien, B. Tarasinski, and B. M. Terhal, New Journal of Physics **21**, 023022 (2019).
[13] N. Wiebe and C. Granade, Phys. Rev. Lett. **117**, 010503 (2016).
[14] S. Kimmel, G. H. Low, and T. J. Yoder, Phys. Rev. A **92**, 062315 (2015).
[15] K. Rudinger, S. Kimmel, D. Lobser, and P. Maunz, Phys. Rev. Lett. **118**, 190502 (2017), 1702.01763.
[16] A. M. Meier, K. A. Burkhardt, B. J. McMahon, and C. D. Herold, Physical Review A **100**, 052106 (2019).
[17] A. E. Russo, K. M. Rudinger, B. C. A. Morrison, and A. D. Baczewski, arXiv (2020), 2007.08697.
[18] As the original proof of correctness in [14] contains a flaw (also previously noted in [19]), we reprove the robustness result in a more general form in this paper in Appendix B.
[19] F. Belliardo and V. Giovannetti, arXiv preprint (2020), arXiv:2007.02994 [quant-ph].
[20] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Nature communications **8**, 1 (2017).
[21] N. K. Langford, New Journal of Physics **15**, 035003 (2013).
[22] S. Wölk, T. Sriarunothai, G. S. Giri, and C. Wunderlich, New Journal of Physics **21**, 013015 (2019).
[23] D. Mogilevtsev, Z. Hradil, J. Rehacek, and V. Shchesnovich, Physical review letters **111**, 120403 (2013).
[24] If $\hat{\theta}_{k-1}$ falls exactly half-way between two members of Θ_k, we are faced with an ambiguity. Because this occurs on a set of measure zero and adds distracting complexity to the discussion, a discussion of this contingency is deferred to Appendix A.

[25] When Θ is not a compact set, d is defined using the infimum rather than the minimum. We only define $\mathcal{M}$ in the case that there is a minimizer.
[26] A similar analysis of the robustness of RPE can be found in [14], but as mentioned previously, there is an error in the details of that analysis, so we reprove the result here.
[27] For the $N_0 = 1$ case, the antipodal point $\hat{\theta}_0+\pi$ is excluded from Φ_0 by the third equation, unlike the other two. See Appendix A for details.
[28] D. Grinko, J. Gacon, C. Zoufal, and S. Woerner, arXiv (2019), 1912.05559.
[29] E. van den Berg, arXiv preprint arXiv:1902.11168 (2019).
[30] We choose to simulate dephasing noise along the σ_x- and σ_y-axes because, under the action of U_{t}, the state remains in the $\sigma_y\sigma_z$-plane, so dephasing along σ_y and σ_z would simply look like depolarization.
[31] We have used intervals that are open on both the left and right sides because we have not chosen a convention to use, e.g., the angle that is closer to 0 when restricted to the principal range $[0, 2\pi)$. While this introduces another failure mode to the analysis, albeit a low-probability one, we will see in Theorem 1 that this failure mode is ruled out by consecutive-consistency. Additionally, we will only use Eqs. A5 and A7 as supersets of Φ_0—i.e., not to prove membership in Φ_0.

Appendix A: Derivation of Criteria

This appendix provides proofs for some of the results in Section III, as well as some additional details. We will extensively use the distance d from an angle to a set of angles, as well as the minimizer $\mathcal{M}$ of d, so we repeat their definitions Eq. (7) and Eq. (9) here. First, we restate the point distance on the unit circle,

$$|\theta' - \theta''|_{2\pi} = \min\left\{|\theta' - \theta'' + 2\pi n| \;\middle|\; n \in \mathbb{Z}\right\}. \tag{A1}$$

which induces a set distance,

$$d(\theta', \Theta) = \inf_{\tilde{\theta}\in\Theta} \left|\tilde{\theta} - \theta'\right|_{2\pi}. \tag{A2}$$

Notice this subsumes the definition in Eq. (8). When a unique minimizer exists, we define it as $\mathcal{M}$:

$$\mathcal{M}(\theta', \Theta) = \operatorname*{argmin}_{\tilde{\theta}\in\Theta} \left|\tilde{\theta} - \theta'\right|_{2\pi}. \tag{A3}$$

We will additionally use a version of the triangle inequality appropriate for point-to-set distances.

Theorem (Set Triangle Inequality)**.** $d(\theta, \Theta) \le |\theta - \theta'|_{2\pi} + d(\theta', \Theta)$

Proof.

$$d(\theta, \Theta) = \min_{\tilde{\theta}\in\Theta} \left|\tilde{\theta} - \theta\right|_{2\pi} \le \min_{\tilde{\theta}\in\Theta} \left(\left|\tilde{\theta} - \theta'\right|_{2\pi} + |\theta' - \theta|_{2\pi}\right) = |\theta' - \theta|_{2\pi} + \min_{\tilde{\theta}\in\Theta} \left|\tilde{\theta} - \theta'\right|_{2\pi} = |\theta' - \theta|_{2\pi} + d(\theta', \Theta) \tag{A4}$$

□

1. Plausible Consistency

Recall our definition in Eq. (11) of the plausible consistent angles at each generation k,

$$\Phi_k = \left\{\tilde{\theta} \;\middle|\; \hat{\theta}_k = \mathcal{M}(\tilde{\theta}, \Theta_k)\right\}. \tag{A5}$$

We call Φ_k the plausible angles for generation k because the estimate $\hat{\theta}_k$ chosen at generation k is "correct" (as defined by Eq. (10)) if and only if the actual angle is in Φ_k. Therefore, if the actual angle were any $\tilde{\theta} \in \bigcap_{k\le k_{\max}} \Phi_k$, then the entire sequence of estimates $\{\hat{\theta}_k\}_{k\le k_{\max}}$ would be correct. Hence, the corresponding consistency check is

$$\bigcap_{k=0}^{k_{\max}} \Phi_k \neq \emptyset. \tag{A6}$$

Eq. (A6) is the same as Eq. (14), in the main text.

Remark 1. The plausible angles at generation k are exactly those angles for which distance to the set Θ_k is the same as the distance to the RPE-chosen angle $\hat{\theta}_k \in \Theta_k$,

$$\Phi_k = \left\{\tilde{\theta} \;\middle|\; |\hat{\theta}_k - \tilde{\theta}|_{2\pi} = d(\tilde{\theta}, \Theta_k)\right\}. \tag{A7}$$

Remark 2. The angles within π/N_k of the RPE-chosen angle $\hat{\theta}_k$ are the plausible angles at generation k,

$$\Phi_k = \left\{\tilde{\theta} \;\middle|\; |\tilde{\theta} - \hat{\theta}_k|_{2\pi} < \frac{\pi}{N_{k'}}\right\}. \tag{A8}$$

Note that when $k = 0$, Eqs. A5 and A7 give $\Phi_0 = [0, 2\pi)$, while Eq. (A8) excludes the antipodal point $\hat{\theta}_0 + \pi$. In this case, we define Φ_0 by Eq. (A8) [31].

Remark 3. If the plausible consistency check Eq. (A6) is satisfied as $k_{\max} \to \infty$, we are able to fully resolve the angle. In other words, if there exists a $\tilde{\theta} \in \bigcap_{k<\infty} \Phi_k$, then $\tilde{\theta}$ is unique and $\lim_{k\to\infty} \hat{\theta}_k \to \tilde{\theta}$.

However, as noted in the main text, the implications of Eq. (A6) for finitely many generations of data are limited in some cases.

Remark 4. If $N_k \geq 2N_{k-1}$ for all k, then $\Phi_k \subset \Phi_{k-1}$. Equivalently, $\bigcap_{k\leq k_{\max}} \Phi_k = \Phi_{k_{\max}}$, and the consistency criterion Eq. (A6) is always satisfied.

Proof. Assume that $N_k \geq 2N_{k-1}$, and let $\tilde{\theta} \in \Phi_k$. We prove that $\tilde{\theta} \in \Phi_{k-1}$. First, notice that

$$|\hat{\theta}_{k-1} - \hat{\theta}_k|_{2\pi} \leq \pi/N_k, \tag{A9}$$

since $\hat{\theta}_k$ is chosen to be the element of Θ_k that is closest to $\hat{\theta}_{k-1}$, and Θ_k is made up of N_k equally-spaced angles in $[0, 2\pi)$. Then, using the triangle inequality,

$$\begin{aligned} |\hat{\theta}_{k-1} - \tilde{\theta}|_{2\pi} &\leq |\hat{\theta}_{k-1} - \hat{\theta}_k|_{2\pi} + |\hat{\theta}_k - \tilde{\theta}|_{2\pi} \\ &\leq \frac{\pi}{N_k} + \frac{\pi}{N_k} \\ &= \frac{\pi}{N_k/2} \leq \frac{\pi}{N_{k-1}}, \end{aligned} \tag{A10}$$

where the second line follows from Eq. (A9) and from the definition Eq. (A5) of Φ_k. Hence by Eq. (A5), $\tilde{\theta} \in \Phi_{k-1}$, and thus $\Phi_k \subset \Phi_{k-1}$. □

2. Consecutive Consistency

Recall the definition of consecutive consistency in Criterion 2 (Eq. (16), in the main text),

$$\bigcap_{k=0}^{k_{\max}} \Lambda_k \neq \emptyset, \tag{A11}$$

where

$$\Lambda_k = \left\{ \tilde{\theta} \;\middle|\; d(\tilde{\theta}, \Theta_k) + d(\tilde{\theta}, \Theta_{k-1}) < \frac{\pi}{N_k} \right\} \tag{A12}$$

for $k \in \{1, 2, \ldots, k_{\max}\}$, and $\Lambda_0 = [0, 2\pi)$. Notice first that consecutive consistency does not make reference to the RPE-chosen angles $\hat{\theta}_k$. That Criterion 2 (Eq. (A11)) is stronger than Criterion 1 (Eq. (A6)) is proven in the following theorem:

Theorem 1. *The sets* $\{\Phi_k\}_{k\leq k_{max}}$ *and* $\{\Lambda_k\}_{k\leq k_{max}}$ *satisfy:*

$$\bigcap_{k'\leq k} \Lambda_{k'} \subseteq \bigcap_{k'\leq k} \Phi_{k'}. \tag{A13}$$

Hence, Eq. (16) *implies Eq.* (14).

Proof. We will need to show along the way that each $\hat{\theta}_k = \mathcal{M}(\hat{\theta}_{k-1}, \Theta_k)$ is well-defined—this may fail to be the case if the choice is between two candidate angles that are equidistant to the previous $\hat{\theta}_k$ (i.e., $d(\hat{\theta}_{k-1}, \Theta_k) = \pi/N_k$). We therefore proceed by induction on k for the following statement: there is a unique minimizer $\mathcal{M}(\hat{\theta}_{k-1}, \Theta_k)$ (provided $k > 0$), and

$$\bigcap_{k'\leq k} \Lambda_{k'} \subseteq \bigcap_{k'\leq k} \Phi_{k'}. \tag{A14}$$

The base case is $k = 0$, where Eq. (A14) becomes $\Lambda_0 \subseteq \Phi_0$, which follows directly from the definitions Eq. (A5) and Eq. (A12).

For the induction step, assume that for some k the inductive hypothesis (i.e., Eq. (A14) and uniqueness of the minimizer) holds. Then for any $\tilde{\theta} \in \bigcap_{k'\leq k+1} \Lambda_{k'}$ we have $\tilde{\theta} \in \Phi_k$, and (by Remark 1) $|\hat{\theta}_k - \tilde{\theta}|_{2\pi} = d(\tilde{\theta}, \Theta_k)$. Therefore by the triangle inequality,

$$d(\hat{\theta}_k, \Theta_{k+1}) \leq |\hat{\theta}_k - \tilde{\theta}|_{2\pi} + d(\tilde{\theta}, \Theta_{k+1}) = d(\tilde{\theta}, \Theta_k) + d(\tilde{\theta}, \Theta_{k+1}) < \frac{\pi}{N_{k+1}}, \tag{A15}$$

where the final inequality follows from the definition of Λ_{k+1}. Because the final inequality in Eq. (A15) is strict, the minimizer $\mathcal{M}(\hat{\theta}_k, \Theta_{k+1})$ is unique. Next, set

$$B = \bigcap_{k' \le k} \Lambda_{k'} \subseteq \bigcap_{k' \le k} \Phi_{k'}. \tag{A16}$$

We will show that

$$B \cap \Lambda_{k+1} \subseteq B \cap \Phi_{k+1}. \tag{A17}$$

Let $\tilde{\theta} \in B \cap \Lambda_{k+1}$. Then we have

$$\begin{aligned} \left|\mathcal{M}(\tilde{\theta}, \Theta_{k+1}) - \hat{\theta}_k\right|_{2\pi} &\le \left|\mathcal{M}(\tilde{\theta}, \Theta_{k+1}) - \tilde{\theta}\right|_{2\pi} + \left|\tilde{\theta} - \hat{\theta}_k\right|_{2\pi} \\ &= \left|\mathcal{M}(\tilde{\theta}, \Theta_{k+1}) - \tilde{\theta}\right|_{2\pi} + \left|\tilde{\theta} - \mathcal{M}(\tilde{\theta}, \Theta_k)\right|_{2\pi} \\ &= d(\tilde{\theta}, \Theta_{k+1}) + d(\tilde{\theta}, \Theta_k) \\ &< \frac{\pi}{N_{k+1}}, \end{aligned} \tag{A18}$$

where the first line follows by the triangle inequality, the second line follows because $\tilde{\theta} \in \Phi_k$ and thus $\hat{\theta}_k = \mathcal{M}(\tilde{\theta}, \Theta_k)$, the third line follows by the definition Eq. (A2) of the distance d, and the final inequality follows because $\tilde{\theta} \in \Lambda_{k+1}$. Therefore, since the elements of Θ_{k+1} are separated by $2\pi/N_{k+1}$, by Eq. (A18) $\mathcal{M}(\tilde{\theta}, \Theta_{k+1})$ must be the closest element in Θ_{k+1} to $\hat{\theta}_k$. But that closest element is $\hat{\theta}_{k+1}$, by definition, so $\hat{\theta}_{k+1} = \mathcal{M}(\tilde{\theta}, \Theta_{k+1})$, i.e., $\tilde{\theta} \in \Phi_{k+1}$. □

Corollary 1.1. *If $\tilde{\theta} \in \bigcap_{k' \le k} \Lambda_{k'}$, and $\{\Theta'_{k'}\}_{k' \le k_{max}}$ are another set of measurement data satisfying*

$$d(\tilde{\theta}, \Theta'_{k'}) \le d(\tilde{\theta}, \Theta_{k'}), \tag{A19}$$

then $\tilde{\theta} \in \bigcap_{k' \le k} \Lambda'_{k'} \subseteq \bigcap_{k' \le k} \Phi'_{k'}$, where $\{\Lambda'_{k'}\}_{k' \le k}$ and $\{\Phi'_{k'}\}_{k' \le k}$ are generated by the Θ'.

The corollary shows that any improvement of the measurements that reduces the error to any of the consecutively consistent values $\tilde{\theta} \in \bigcap_{k' \le k_{\max}} \Lambda_{k'}$ will still cause that $\tilde{\theta}$ to be identified as a correct value (c.f. Fig. 3, illustrating that this is *not* the case for $\bigcap_{k' \le k_{\max}} \Phi_{k'}$).

Notice that the Φ_k sets are defined in terms of distance from the $\hat{\theta}_k$, automatically making them intervals. The Λ_k are instead defined in terms of distance to the Θ_k. Theorem 1 and Remark 1 together imply that $\bigcap_{k' \le k} \Lambda_{k'}$ is a simple interval.

Lemma 1. *Let $D = \frac{\pi}{2N_k} - \frac{1}{2}\left|\hat{\theta}_k - \hat{\theta}_{k-1}\right|_{2\pi}$. If $D < 0$, then $\Phi_k \cap \Lambda_k = \emptyset$. Otherwise,*

$$\Lambda_k \cap \Phi_k = \left(\hat{\theta}_{k-1} \mp D, \hat{\theta}_k \pm D\right)_{2\pi}. \tag{A20}$$

The 2π subscript indicates that the interval is circular, and the $\pm$ means that the interval should be interpreted as follows: connect $\hat{\theta}_k$ and $\hat{\theta}_{k-1}$ along the shortest arc, then expand that arc by a circular distance D on both sides. Also, the arc $\Lambda_k \cap \Phi_k$ has length $< \pi$.

Proof. We first prove that the arc $\Lambda_k \cap \Phi_k$ has length $< \pi$. The worst case is when $k = 1$: in this case, assuming that N_k is strictly increasing, $N_1 \ge 2$, so $D \le \pi/4$. Since Θ_1 contains $N_k \ge 2$ angles, $|\hat{\theta}_1 - \hat{\theta}_0|_{2\pi} < \pi/2$. Hence the arc $\left(\hat{\theta}_{k-1} \mp D, \hat{\theta}_k \pm D\right)_{2\pi}$ has length at most π.

For the main proof, first note that by applying Remark 1 to the distances in the definition of Λ_k (Eq. (A12)),

$$\Phi_k \cap \Lambda_k = \Phi_k \cap \left\{\tilde{\theta} \;\middle|\; \left|\tilde{\theta} - \hat{\theta}_k\right|_{2\pi} + \left|\tilde{\theta} - \hat{\theta}_{k-1}\right|_{2\pi} < \frac{\pi}{N_k}\right\}. \tag{A21}$$

Because of the interval representation of Φ_k,

$$\Phi_k \cap \Lambda_k = (\hat{\theta}_k - \pi/N_k, \hat{\theta}_k + \pi/N_k)_{2\pi} \cap \left\{\tilde{\theta} \;\middle|\; \left|\tilde{\theta} - \hat{\theta}_k\right|_{2\pi} + \left|\tilde{\theta} - \hat{\theta}_{k-1}\right|_{2\pi} < \frac{\pi}{N_k}\right\}. \tag{A22}$$

The first interval is a superset of the second set therefore

$$\Phi_k \cap \Lambda_k = \left\{ \tilde{\theta} \;\middle|\; \left|\tilde{\theta} - \hat{\theta}_k\right|_{2\pi} + \left|\tilde{\theta} - \hat{\theta}_{k-1}\right|_{2\pi} < \frac{\pi}{N_k} \right\}. \tag{A23}$$

Because $|\hat{\theta}_{k-1} - \hat{\theta}_k|_{2\pi} < \pi$, for any $\tilde{\theta}$

$$\left|\tilde{\theta} - \hat{\theta}_k\right|_{2\pi} + \left|\tilde{\theta} - \hat{\theta}_{k-1}\right|_{2\pi} = |\hat{\theta}_{k-1} - \hat{\theta}_k|_{2\pi} + \begin{cases} 0, & \tilde{\theta} \in (\hat{\theta}_{k-1}, \hat{\theta}_k)_{2\pi} \ ; \\ 2\min\{|\hat{\theta}_{k-1} - \tilde{\theta}|_{2\pi}, |\hat{\theta}_k - \tilde{\theta}|_{2\pi}\}, & \tilde{\theta} \notin (\hat{\theta}_{k-1}, \hat{\theta}_k)_{2\pi} \ . \end{cases} \tag{A24}$$

Consider the first case in Eq. (A24), where $\tilde{\theta} \in (\hat{\theta}_{k-1}, \hat{\theta}_k)_{2\pi}$: from Eq. (A23) we obtain

$$\tilde{\theta} \in \Phi_k \cap \Lambda_k \quad \Leftrightarrow \quad \left|\hat{\theta}_k - \hat{\theta}_{k-1}\right|_{2\pi} < \frac{\pi}{N_k}. \tag{A25}$$

The right-hand inequality may be rewritten in terms of D as

$$0 < \frac{\pi}{2N_k} - \frac{1}{2}\left|\hat{\theta}_k - \hat{\theta}_{k-1}\right|_{2\pi} = D, \tag{A26}$$

and thus $\tilde{\theta}$ is in the interval in Eq. (A20).

Alternatively, in the second case of Eq. (A24), where $\tilde{\theta} \notin (\hat{\theta}_{k-1}, \hat{\theta}_k)_{2\pi}$, $\tilde{\theta} \in \Phi_k \cap \Lambda_k$ is equivalent to

$$\left|\hat{\theta}_k - \hat{\theta}_{k-1}\right|_{2\pi} + 2\min\{|\hat{\theta}_{k-1} - \tilde{\theta}|_{2\pi}, |\hat{\theta}_k - \tilde{\theta}|_{2\pi}\} < \frac{\pi}{N_k}, \tag{A27}$$

or

$$\min\{|\hat{\theta}_{k-1} - \tilde{\theta}|_{2\pi}, |\hat{\theta}_k - \tilde{\theta}|_{2\pi}\} < \frac{\pi}{2N_k} - \frac{1}{2}\left|\hat{\theta}_k - \hat{\theta}_{k-1}\right|_{2\pi} = D, \tag{A28}$$

which is precisely checking if $\tilde{\theta}$ is within D of the endpoints of the angular interval, also precisely as desired. □

Theorem 2. *Testing for membership in $\bigcap_{k'<k} \Lambda_{k'}$ is equivalent to testing for membership in the intersection of the intervals in Lemma 1.*

Proof. This follows from Lemma 1, noticing that, by Theorem 1, $\bigcap_{k'\le k} \Lambda_{k'} \subseteq \bigcap_{k'\le k} \Phi_{k'}$. □

3. Historical Consistency

Recall the definition of $\Delta_k[\delta\theta_k]$ (Eq. (19) in the main text):

$$\Delta_k[\delta\theta_k] = \left\{ \tilde{\theta} \;\middle|\; d(\tilde{\theta}, \Theta_k) < \frac{\delta\theta_k}{N_k} \right\}, \tag{A29}$$

where $\{\delta\theta_k\}_{k\le k_{\max}}$ is a sequence of positive real numbers. In the following, $|A|$ denotes the length of the interval, A.

Lemma 2. *Suppose $a \in \mathbb{R}$, and $0 \le L' < L$. Then,*

$$b \in (a, a+L) \iff |(b-L', b+L') \cap (a, a+L)| > L'. \tag{A30}$$

This generalizes to angular intervals as long as $L + 2L' < 2\pi$.

Theorem 3. *Consider an RPE schedule N_k with $N_0 = 1$ and $N_k > N_{k-1}$. Assume that $\delta\theta_k$ is a local consistency error bound schedule, i.e., it satisfies $\frac{\delta\theta_k}{N_k} + \frac{\delta\theta_{k-1}}{N_{k-1}} \le \frac{\pi}{N_k}$, and, in addition, $\frac{\delta\theta_{k-1}}{N_{k-1}} > \frac{\delta\theta_k}{N_k}$ for all $0 < k \le k_{max}$. Then the two following statements are equivalent:*

$$\hat{\theta}_k \in \bigcap_{k'\le k} \Delta_{k'} \quad \forall k \le k_{max}, \tag{A31}$$

and

$$\left|\bigcap_{k'\le k} \Delta_{k'}\right| > \frac{\delta\theta_k}{N_k} \quad \forall k \le k_{max}. \tag{A32}$$

Proof. We proceed by induction on $k_{\max}$. For the base case, note that $\hat{\theta}_0 \in \Delta_0[\delta\theta_0]$ (i.e., Eq. (A31) holds), and $N_0 = 1$, so $|\Delta_0[\delta\theta_0]| = 2\delta\theta_0 > \delta\theta_0$ (i.e., Eq. (A32) holds).

Let $k_{\max}$ be a positive integer. The induction hypothesis is that Eq. (A31) and Eq. (A32) are equivalent $\forall k \leq k_{\max} - 1$, i.e., that either both or neither hold. If neither hold, then that remains true even when the $k = k_{\max}$ cases are included. If both hold, then it will be enough to show that under this assumption

$$\hat{\theta}_{k_{\max}} \in \bigcap_{k' \leq k_{\max}} \Delta_{k'} \quad \Leftrightarrow \quad \left| \bigcap_{k' \leq k_{\max}} \Delta_{k'} \right| > \frac{\delta\theta_{k_{\max}}}{N_{k_{\max}}}. \tag{A33}$$

First, notice that by the definitions Eq. (A5) and Eq. (A29),

$$\bigcap_{k' \leq k_{\max}} \Delta_{k'} = (\Phi_{k_{\max}} \cap \Delta_{k_{\max}}) \cap \bigcap_{k' \leq k_{\max}-1} \Delta_{k'}, \tag{A34}$$

and

$$\Phi_{k_{\max}} \cap \Delta_{k_{\max}} = \left(\hat{\theta}_{k_{\max}} - \frac{\delta\theta_{k_{\max}}}{N_{k_{\max}}}, \hat{\theta}_{k_{\max}} + \frac{\delta\theta_{k_{\max}}}{N_{k_{\max}}} \right). \tag{A35}$$

Eq. (A33) is then equivalent to

$$\hat{\theta}_{k_{\max}} \in \bigcap_{k' \leq k_{\max}} \Delta_{k'} \iff \left| (\Phi_{k_{\max}} \cap \Delta_{k_{\max}}) \cap \bigcap_{k' \leq k_{\max}-1} \Delta_{k'} \right| > \frac{\delta\theta_{k_{\max}}}{N_{k_{\max}}}. \tag{A36}$$

Because $\bigcap_{k' \leq k_{\max}-1} \Delta_{k'}$ is an interval of length $L > \frac{\delta\theta_{k_{\max}-1}}{N_{k_{\max}-1}}$, inserting Eq. (A35) and applying Lemma 2 gives the claim. □

Appendix B: Robust Resource Scaling

In Ref [14], RPE was shown to be robust to additive errors, which in turn makes the protocol robust to a range of physical errors including SPAM errors. While there was a mistake in the details of that analysis [19], here we show that the main result still holds, and any protocol that has RPE-like characteristics can be made robust.

Let $X = \{X_1, \ldots, X_{k_{\max}}\}$ be a set of binomial random variables, where it requires cost c_k to obtain a sample of X_k. Suppose there is a protocol that, for each $k \in [k_{\max}]$, takes m_k samples of X_k to create an estimate $\hat{x}_k$ of $\mathbb{E}[X_k]$ (the average value of X_k), where $|\mathbb{E}[X_k] - \hat{x}_k| < \delta$ with probability at least $1 - 2\exp[-2m_k\delta]$. The cost of this protocol is $\sum_{i=1}^{k_{\max}} c_k m_k$. (Here δ is the same constant for all k.) Note that $1 - 2\exp(-2m_k\delta)$ probability of success is natural because many results that involve bounding the success probability of binomial random variables rely on Hoeffding's inequality, which produces this term.

Given such a protocol, we can simulate it using binomial random variables $X' = \{X'_1, \ldots, X'_{k_{\max}}\}$ that approximate X, if we are promised that for all k, $|\mathbb{E}[X_k] - \mathbb{E}[X'_k]| < \epsilon < \delta$ for some constant ϵ. If the cost of sampling X'_k is c_k, then the cost of the new protocol will be only a constant factor more than the original protocol. Consider taking $m'_k = m_k\delta/(\delta - \epsilon)$ samples of X'_k. Then using the Hoeffding inequality for the binomial distribution, we can obtain an estimate $\hat{x}'_k$ of $\mathbb{E}[X'_k]$ to within additive error $\delta - \epsilon$ with probability of error at most

$$2\exp[-2m'_k(\delta - \epsilon)] = 2\exp[-2m_k\delta]. \tag{B1}$$

Because $|\mathbb{E}[X_k] - \mathbb{E}[X'_k]| < \epsilon < \delta$, this estimate $\hat{x}'_k$ is actually within δ of $\mathbb{E}[X_k]$ with probability of error $2\exp(-2m_k\delta)$. Thus we can use our estimates $\hat{x}'_k$ in place of $\hat{x}_k$ in the original protocol and achieve the same result. The cost is $\sum_{i=1}^{k_{\max}} c_k m_k \delta/(\delta - \epsilon)$, as claimed.

The consequence of this analysis is that *any* experiment dealing with binomial random variables that does not require precise estimates of any single variable will still be successful even if those variables become biased, at the cost of a multiplicative, constant overhead. In particular, this means that it is possible to still achieve Heisenberg scaling using the phase estimation protocol outlined here, even in the presence of noise, as long as the noise does not shift the probabilities of the measurement outcomes by more than a constant.

However, this statement is difficult to take advantage of in practice, since knowing how much to increase the sample number requires knowing the size of ϵ. This brings us back to the main purpose of the present work, which is to detect when the noise does not satisfy this property.

Appendix C: Sample Complexity

In this appendix we demonstrate the scaling of sample complexity, as a function of noise in the quantum channel, to achieve a particular target error bound.

1. Preliminaries

For a sufficiently large number of samples M, the binomial distribution is approximately the normal distribution:

$$\mathrm{Binom}_{p,M}(k) \approx \mathrm{Norm}_{Mp,\sqrt{Mp(1-p)}}(k) = \frac{1}{\sqrt{2\pi}\sqrt{Mp(1-p)}} e^{-\frac{1}{2}\frac{(k-Mp)^2}{Mp(1-p)}}. \tag{C1}$$

Also, the product of two normal distributions is a rescaled normal distribution. We prove this here and obtain the rescaling factor. First, observe that

$$(2\pi\sigma_a\sigma_b \,\mathrm{Norm}_{\mu_a,\sigma_a}(x)\,\mathrm{Norm}_{\mu_b,\sigma_b}(x))^{-2} = \exp\left[\left(\frac{x-\mu_a}{\sigma_a}\right)^2 + \left(\frac{x-\mu_b}{\sigma_b}\right)^2\right] \tag{C2}$$

$$= \exp\left[\frac{x^2 - 2\mu_a x + \mu_a^2}{\sigma_a^2} + \frac{x^2 - 2\mu_b x + \mu_b^2}{\sigma_b^2}\right] \tag{C3}$$

$$= \exp\left[\left(\sqrt{\sigma_a^{-2} + \sigma_b^{-2}}\,x - \frac{\mu_a\sigma_a^{-2} + \mu_b\sigma_b^{-2}}{\sqrt{\sigma_a^{-2} + \sigma_b^{-2}}}\right)^2 + \left(\mu_a^2\sigma_a^{-2} + \mu_b^2\sigma_b^{-2} - \frac{(\mu_a\sigma_a^{-2} + \mu_b\sigma_b^{-2})^2}{\sigma_a^{-2} + \sigma_b^{-2}}\right)\right] \tag{C4}$$

$$= \left(\sqrt{2\pi}\sigma\,\mathrm{Norm}_{\mu,\sigma}(x)\right)^{-2} \exp\left[\mu_a^2\sigma_a^{-2} + \mu_b^2\sigma_b^{-2} - \frac{(\mu_a\sigma_a^{-2} + \mu_b\sigma_b^{-2})^2}{\sigma_a^{-2} + \sigma_b^{-2}}\right] \tag{C5}$$

where $\sigma^{-2} = \sigma_a^{-2} + \sigma_b^{-2}$ and $\mu = \sigma^2(\mu_a\sigma_a^{-2} + \mu_b\sigma_b^{-2})$. It follows that

$$\mathrm{Norm}_{\mu_a,\sigma_a}(x)\,\mathrm{Norm}_{\mu_b,\sigma_b}(x) = \mathrm{Norm}_{\mu,\sigma}(x)\frac{\sigma}{\sqrt{2\pi}\sigma_a\sigma_b}\exp\left[-\frac{1}{2}\left(\mu_a^2\sigma_a^{-2} + \mu_b^2\sigma_b^{-2} - \frac{(\mu_a\sigma_a^{-2} + \mu_b\sigma_b^{-2})^2}{\sigma_a^{-2} + \sigma_b^{-2}}\right)\right]. \tag{C6}$$

Better yet, the scale factor can be expressed in terms of another normal distribution:

$$\mathrm{Norm}_{\mu_a,\sigma_a}(x)\,\mathrm{Norm}_{\mu_b,\sigma_b}(x) = \mathrm{Norm}_{\mu,\sigma}(x)\,\mathrm{Norm}_{0,\sigma'}(x'), \tag{C7}$$

where $x' = \mu_a - \mu_b$ and $\sigma'^2 = \sigma_a^2 + \sigma_b^2$.

2. Complexity

We now derive an expression for $p_{\lambda,\phi,M}(\hat{\phi})$, the probability density function for obtaining the estimated angle $\hat{\phi}$ given a distribution by taking M samples, as described by Eq. (42). We suppress the subscripts for typographical clarity. In particular, in the large M limit, λ and M appear only in the combination $\lambda^2 M$, showing that the effect of λ is simply to determine the number M of actual, noisy samples required in order to achieve some desired number M' of effective, noiseless samples, as $M = \lambda^{-2}M'$. Unlike a generic error model, depolarization noise is fully characterized by the λ parameter (i.e., ϕ is unchanged), making this result particularly useful in that case.

The experimental data is used by the RPE algorithm to generate an estimate of the angle of rotation of U_t. As described in Eq. (5), this is the angle $\hat{\phi}$ satisfying

$$\hat{\lambda}\begin{pmatrix}\cos\hat{\phi}\\ \sin\hat{\phi}\end{pmatrix} = \begin{pmatrix}2\frac{\hat{c}}{M} - 1\\ 2\frac{\hat{s}}{M} - 1\end{pmatrix} = \begin{pmatrix}\tau_c\\ \tau_s\end{pmatrix}, \tag{C8}$$

where $\hat{c}$ and $\hat{s}$ are the sample counts from Eq. (3) and Eq. (4), respectively, and M is the total number of measurements. Notice that

$$\frac{\tau_s}{\tau_c} = \tan\hat{\phi} \tag{C9}$$

modulo additive factors of π. If we restrict our attention to the case $\hat{\phi} \in [0, \pi)$, Eq. (C9) uniquely determines τ_s and τ_c (and therefore $\hat{s}$ and $\hat{c}$). It follows that the probability of measuring an angle $\hat{\phi}$ is

$$\sum_{\hat{c},\hat{s}}^{M}{}' \, \mathrm{Binom}_{P_{\mathrm{c},k},M}(\hat{c}) \, \mathrm{Binom}_{P_{\mathrm{s},k},M}(\hat{s}) \,, \tag{C10}$$

where the prime denotes the restriction of the integers $\hat{c}, \hat{s} \in [0, M]$ to those satisfying both $\hat{s} \geq M/2$ for $\hat{\phi} \in [0, \pi)$ and Eq. (C9). If $\hat{\phi} \in [\pi, 2\pi)$, an analogous argument to the following leads to the same conclusions. Using the τ notation, this probability can be expressed as the integral over the ray originating at the origin, making an angle $\hat{\phi}$ with the τ_c-axis:

$$\iint_{\text{ray } \hat{\phi}} d\tau_s \, d\tau_c \left(\sum_{\hat{c},\hat{s}}^{M} \mathrm{Binom}_{P_{\mathrm{c},k},M}(\hat{c}) \, \mathrm{Binom}_{P_{\mathrm{s},k},M}(\hat{s}) \, \delta(\tau_c - (2\hat{c}/M - 1)) \delta(\tau_s - (2\hat{s}/M - 1)) \right). \tag{C11}$$

We identify the integrand as the probability density function, $p_{\tau_c,\tau_s} \, d\tau_c \, d\tau_s$, and notice that, for sufficiently large M, the binomials can be approximated using (C1), yielding

$$p_{\tau_c,\tau_s} = \sum_{\hat{c},\hat{s}}^{M} \mathrm{Norm}_{MP_{\mathrm{c},k},\sqrt{M}\sigma_{\mathrm{c}}}(\hat{c}) \, \mathrm{Norm}_{MP_{\mathrm{s},k},\sqrt{M}\sigma_{\mathrm{s}}}(\hat{s}) \delta\Big(\tau_c - (2\hat{c}/M - 1)\Big) \delta\Big(\tau_s - (2\hat{s}/M - 1)\Big) \tag{C12}$$

$$= \sum_{\hat{c},\hat{s}}^{M} M^2 \, \mathrm{Norm}_{MP_{\mathrm{c},k},\sqrt{M}\sigma_{\mathrm{c}}}(\hat{c}) \, \mathrm{Norm}_{MP_{\mathrm{s},k},\sqrt{M}\sigma_{\mathrm{s}}}(\hat{s}) \delta(M\tau_c - (2\hat{c} - M)) \delta(M\tau_s - (2\hat{s} - M)) \tag{C13}$$

where

$$\sigma_{\mathrm{s}}^2 = P_{\mathrm{s},k}(1 - P_{\mathrm{s},k}), \quad \text{and} \quad \sigma_{\mathrm{c}}^2 = P_{\mathrm{c},k}(1 - P_{\mathrm{c},k}). \tag{C14}$$

If p_{τ_x,τ_y} is to be integrated over a sufficiently large area, the individual privileged points selected by the Dirac-δ's can be neglected (in the large M limit). More precisely, we replace the dirac-δ by a function $D_{i,j}(x, y)$ which is 1 for $i < x < i + 1$ and $j < y < j + 1$, and 0 otherwise:

$$p_{\tau_c,\tau_s} \approx \sum_{\hat{c},\hat{s}}^{M} M^2 \, \mathrm{Norm}_{MP_{\mathrm{c},k},\sqrt{M}\sigma_{\mathrm{c}}}(\hat{c}) \, \mathrm{Norm}_{MP_{\mathrm{s},k},\sqrt{M}\sigma_{\mathrm{s}}}(\hat{s}) D_{2\hat{c}-M,2\hat{s}-M}(M\tau_c, M\tau_s) \tag{C15}$$

$$\approx M^2 \, \mathrm{Norm}_{MP_{\mathrm{c},k},\sqrt{M}\sigma_{\mathrm{c}}}(M(\tau_c + 1)/2) \, \mathrm{Norm}_{MP_{\mathrm{s},k},\sqrt{M}\sigma_{\mathrm{s}}}(M(\tau_s + 1)/2). \tag{C16}$$

Pulling a factor of $1/M$ out of each Norm yields:

$$p_{\tau_c,\tau_s} = \mathrm{Norm}_{P_{\mathrm{c},k},\sigma_{\mathrm{c}}M^{-1/2}}((\tau_c + 1)/2) \, \mathrm{Norm}_{P_{\mathrm{s},k},\sigma_{\mathrm{s}}M^{-1/2}}((\tau_s + 1)/2). \tag{C17}$$

Next, parameterize the probability using Eq. (42),

$$\begin{pmatrix} 2P_{\mathrm{c},k} - 1 \\ 2P_{\mathrm{s},k} - 1 \end{pmatrix} = \begin{pmatrix} \lambda \cos\phi \\ \lambda \sin\phi \end{pmatrix}, \tag{C18}$$

so that

$$\sigma_{\mathrm{s}}^2 = 1 - \lambda^2 \sin^2\phi, \quad \text{and} \quad \sigma_{\mathrm{c}}^2 = 1 - \lambda^2 \cos^2\phi, \tag{C19}$$

and

$$p_{\tau_c,\tau_s} = \mathrm{Norm}_{(\lambda/2)\cos\phi,\sigma_{\mathrm{c}}M^{-1/2}}(\tau_c/2) \, \mathrm{Norm}_{(\lambda/2)\sin\phi,\sigma_{\mathrm{s}}M^{-1/2}}(\tau_s/2). \tag{C20}$$

Because we are interested in the probability density function over $\hat{\phi}$, $p(\hat{\phi}) \, d\hat{\phi}$, we integrate p_{τ_x,τ_y} over a wedge spanning angles from 0 to $\hat{\phi}$,

$$\iint_{\text{wedge}} p_{\tau_c,\tau_s} \, d\tau_x \, d\tau_y = \int_0^{\infty} d\tau \int_0^{\hat{\phi}} \tau \, d\hat{\phi}' \; \mathrm{Norm}_{(\lambda/2)\cos\phi,\sigma_{\mathrm{c}}M^{-1/2}}((\tau/2)\cos\hat{\phi}') \, \mathrm{Norm}_{(\lambda/2)\sin\phi,\sigma_{\mathrm{s}}M^{-1/2}}((\tau/2)\sin\hat{\phi}'), \tag{C21}$$

and then take the derivative with $\hat{\phi}$:

$$p(\hat{\phi}) = \int_0^\infty \tau\, d\tau\ \mathrm{Norm}_{(\lambda/2)\cos\phi,\sigma_c M^{-1/2}}((\tau/2)\cos\hat{\phi})\,\mathrm{Norm}_{(\lambda/2)\sin\phi,\sigma_s M^{-1/2}}((\tau/2)\sin\hat{\phi}) \tag{C22}$$

$$= \frac{2}{\lambda^2\sqrt{\cos\hat{\phi}\sin\hat{\phi}}}\int_0^\infty \tau\, d\tau\ \mathrm{Norm}_{\cos\phi/\cos\hat{\phi},2\sigma_c/((\lambda^2 M)^{1/2}\cos\hat{\phi})}(\tau/\lambda)\,\mathrm{Norm}_{\sin\phi/\sin\hat{\phi},2\sigma_s/((\lambda^2 M)^{1/2}\sin\hat{\phi})}(\tau/\lambda) \tag{C23}$$

$$= \frac{2}{\sqrt{\cos\hat{\phi}\sin\hat{\phi}}}\int_0^\infty \tau\, d\tau\ \mathrm{Norm}_{\cos\phi/\cos\hat{\phi},2\sigma_c/((\lambda^2 M)^{1/2}\cos\hat{\phi})}(\tau)\,\mathrm{Norm}_{\sin\phi/\sin\hat{\phi},2\sigma_s/((\lambda^2 M)^{1/2}\sin\hat{\phi})}(\tau), \tag{C24}$$

where we have absorbed the trigonometric and $\lambda/2$ terms, and changed variable, $\tau/\lambda \to \lambda$. It therefore follows that, in the large M limit, the only effect of changing M on the distribution of observed angles is controlled by the quantity $\lambda^2 M$. In other words, λ renormalizes $M \mapsto \lambda^2 M = M'$, as claimed.

3. Distribution

In this section we further simplify Eq. (C24), our expression for the probability density $p(\hat{\phi})$ of estimated angles $\hat{\phi}$, providing a form that is readily numerically evaluated. This result complements Appendix A of [14], which provides a bound on the integral of the tails of this distribution. Using Eq. (C7), put

$$\sigma''^{-2} = \frac{M'}{4}\left(\sigma_c^{-2}\cos^2\hat{\phi} + \sigma_s^{-2}\sin^2\hat{\phi}\right), \tag{C25}$$

and

$$\mu = \frac{4\left(\sigma_c^{-2}\cos^2\hat{\phi} + \sigma_s^{-2}\sin^2\hat{\phi}\right)}{M'}\left(\frac{\cos\phi/\cos\hat{\phi}}{4\sigma_c^2/(M'\cos^2\hat{\phi})} + \frac{\sin\phi/\sin\hat{\phi}}{4\sigma_s^2/(M'\sin^2\hat{\phi})}\right) \tag{C26}$$

$$= \left(\sigma_s^2\cos^2\hat{\phi} + \sigma_c^2\sin^2\hat{\phi}\right)\left(\sigma_s^2\cos\phi\cos\hat{\phi} + \sigma_c^2\sin\phi\sin\hat{\phi}\right) \tag{C27}$$

$$= \left(1 - \lambda^2(\sin^2\phi\cos^2\hat{\phi} + \cos^2\phi\sin^2\hat{\phi})\right)\left((1-\lambda^2\sin^2\phi)\cos\phi\cos\hat{\phi} + (1-\lambda^2\cos^2\phi)\sin\phi\sin\hat{\phi}\right), \tag{C28}$$

to get

$$p(\hat{\phi}) = \frac{2}{\sqrt{\cos\hat{\phi}\sin\hat{\phi}}}\int_0^\infty \tau\, d\tau\ \mathrm{Norm}_{\mu,\sigma''}(\tau)\,\mathrm{Norm}_{0,\sigma'^2}\left(\frac{\cos\phi}{\cos\hat{\phi}} - \frac{\sin\phi}{\sin\hat{\phi}}\right) \tag{C29}$$

where

$$\sigma'^2 = \frac{4}{M'}\left(\sigma_c^2/\cos^2\hat{\phi} + \sigma_s^2/\sin^2\hat{\phi}\right). \tag{C30}$$

Because we are focusing on the $M \to \infty$ limit, the sign of μ determines whether or not the integral over τ is 0 or 1. That sign is determined by the quantity

$$(1-\lambda^2\sin^2\phi)\cos\phi\cos\hat{\phi} + (1-\lambda^2\cos^2\phi)\sin\phi\sin\hat{\phi} \tag{C31}$$

$$= \cos(\phi-\hat{\phi}) - \lambda^2(\sin^2\phi\cos\phi\cos\hat{\phi} + \cos^2\phi\sin\phi\sin\hat{\phi}) \tag{C32}$$

$$= \cos(\phi-\hat{\phi}) - \lambda^2\sin\phi\cos\phi(\sin\phi\cos\hat{\phi} + \cos\phi\sin\hat{\phi}) \tag{C33}$$

$$= \cos(\phi-\hat{\phi}) - \frac{\lambda^2}{2}\sin(2\phi)\sin(\phi+\hat{\phi}) \tag{C34}$$

$$= \cos\delta - \frac{\lambda^2}{2}\sin(2\phi)\sin(2\phi+\delta) \tag{C35}$$

$$= \cos\delta - \frac{\lambda^2}{2}\sin(2\phi)(\sin(2\phi)\cos\delta + \cos(2\phi)\sin\delta) \tag{C36}$$

$$= \cos\delta\left[1 - \frac{\lambda^2}{2}\sin^2(2\phi)\right] + \frac{\lambda^2}{4}\sin(4\phi)\sin\delta, \tag{C37}$$

where $\delta = \hat{\phi} - \phi$. This quantity is guaranteed to be positive if $\cos\delta > 1/2$, or equivalently $|\delta| < \pi/3$. Then $\mu > 0$ and, because $\sigma'' \to \infty$ in the large-M limit being considered, $\mu/\sigma'' \to \infty$, causing the integral over τ to go to unity:

$$p(\hat{\phi}) = \frac{2\mu}{\sqrt{\cos\hat{\phi}\sin\hat{\phi}}} \operatorname{Norm}_{0,\sigma'}\left(\frac{\cos\phi}{\cos\hat{\phi}} - \frac{\sin\phi}{\sin\hat{\phi}}\right). \tag{C38}$$

On the other hand, if $\mu < 0$, the value of $p(\hat{\phi})$ is zero. Absorbing $\sin\hat{\phi}\cos\hat{\phi}$ from the argument, we get

$$p(\hat{\phi}) = 2\mu \operatorname{Norm}_{0,\sigma_0 M'^{-1/2}}\Big(\sin(\phi - \hat{\phi})\Big). \tag{C39}$$

Appendix D: Limitations of an Alternative, Set Formulation of RPE

In this appendix, we consider only the case that $N_k = 2^k$, and discuss an alternative formulation of RPE. In standard RPE, a single angle $\hat{\theta}_k$ is selected at each generation. Instead, one might imagine identifying sets Ω_k of permitted angles, selected on the basis of their proximity to Θ_k. Contrast this with the sets of angles used in criteria 1, 2, and 3 developed above, which are defined by proximity to a single selected angle $\hat{\theta}_k$. In particular, these Ω_k are not necessarily a single interval, and may be computationally intensive to track. One might therefore expect to obtain additional deductive strength by using these sets, since they require significantly more classical power to manage. However, we show that this alternative formulation cannot tolerate errors greater than $\pi/3$ without failing to exclude infinitely many false candidate values for the angle, and therefore it provides no advantage over standard RPE.

To make this protocol precise, assume that at every generation the measurements suffer an error no greater than some fixed angle $2\pi\alpha$,

$$d(2\pi\tau, \Theta_k) < (2\pi\alpha)2^{-k}, \tag{D1}$$

where $2\pi\tau$ is the "true" angle we are attempting to measure, and we are guaranteeing that the measurements are sufficiently accurate. One may ask the question, "If $2\pi\alpha$ is larger than the $\pi/3$ uniform-approximation limit (the limit for standard RPE), can we still guarantee a valid estimate of the true angle $2\pi\tau$?" Here, we provide a counterexample by showing that, for any $2\pi\alpha > \pi/3$ and integer $j \geq 0$, there is a sequence of measurements satisfying Eq. (D1) that always includes a false angle $2\pi\phi$, defined by

$$2\pi\phi = 2\pi\frac{2^{-j}}{3}. \tag{D2}$$

In other words, for any $2\pi\alpha > \pi/3$, we demonstrate measurements that satisfy the error bound, but converge to any one of infinitely many incorrect $2\pi\phi$.

To construct the counterexample, we first clarify the exact process of this generalized set formulation of RPE, which we parameterize by some $\{\beta_k\}_{k\geq 0}$, with $\beta_k > 0$. As with standard RPE, at every generation, N_k candidate values for θ are provided in Θ_k. We define the permitted subset Ω_k of the angular space as

$$\Omega_k = \Omega_{k-1} \cap \left\{\theta \;\middle|\; \exists\tilde{\theta}_k \in \Theta_k \;:\; \left|\tilde{\theta}_k - \theta\right|_{2\pi} < 2\pi\beta_k\right\} = \bigcap_{k'\leq k}\left\{\theta \;\middle|\; \exists\tilde{\theta}_{k'} \in \Theta_{k'} \;:\; \left|\tilde{\theta}_{k'} - \theta\right|_{2\pi} < 2\pi\beta_{k'}\right\}, \tag{D3}$$

for $k \geq 0$, and letting $\Omega_{-1} = [-\pi, \pi)$. β_k must be at least large enough that Ω_k contains the true angle $2\pi\tau$, which can be as far as $2\pi\alpha$ from $\mathcal{M}(2\pi\tau, \Theta_k)$. Therefore we also require that this generalized set formulation use

$$2\pi\beta_k \geq 2\pi\alpha 2^{-k}. \tag{D4}$$

Without loss of generality, we choose $\beta_k = \alpha 2^{-k}$ to saturate this inequality, because any larger choice of β_k will necessarily create a superset of Ω_k, and therefore also fail to exclude our pathological false angle ϕ.

If $2\pi\phi$ is to serve as the angle in the counterexample, we need to show that $2\pi\phi$ is a member of Ω_k for every $k \geq 0$. This will follow immediately from Eq. (D3) if we show that $2\pi\phi$ is within $2\pi\beta_k$ of Θ_k for each k, i.e.,

$$2\pi\phi \in \left\{\theta \;\middle|\; \exists\tilde{\theta}_k \in \Theta_k \;:\; \left|\tilde{\theta}_k - \theta\right|_{2\pi} < 2\pi\beta_k\right\}. \tag{D5}$$

Assume without loss of generality that $\tau = 0$. Let the error in the measured angle at generation k be $2\pi\epsilon_k$. Then all elements of Θ_k incur an error of $2\pi\frac{\epsilon_k}{N_k}$, so

$$\Theta_k = \left\{\frac{2\pi}{N_k}(q + \epsilon_k) \;\middle|\; q \in \mathbb{Z}_{N_k}\right\}. \tag{D6}$$

If we insert this into Eq. (D5), i.e., replace $\tilde{\theta}_k$ with $\frac{2\pi}{N_k}(q+\epsilon_k)$, then Eq. (D5) is equivalent to

$$\exists q \in \mathbb{Z}_{N_k} \ : \ 2\pi \left| \frac{1}{N_k}(q+\epsilon_k) - \phi \right|_{2\pi} < 2\pi\alpha 2^{-k}, \tag{D7}$$

i.e.,

$$\exists q \in \mathbb{Z}_{N_k} \ : \ |q + \epsilon_k - \phi N_k|_{2\pi} < \alpha. \tag{D8}$$

If we define $F(x) = x - \lfloor x \rfloor$ to be the fractional part of the real number x, then Eq. (D8) becomes

$$F(\epsilon_k - \phi 2^k) < \alpha \quad \text{or} \quad F(\epsilon_k - \phi 2^k) > 1 - \alpha. \tag{D9}$$

Recall that $|\epsilon_k| \leq \alpha$, since α is defined to be the maximum allowed error. Therefore, for any ϕ satisfying

$$F(\phi 2^k) < 2\alpha \quad \text{or} \quad F(\phi 2^k) > 1 - 2\alpha, \tag{D10}$$

there exists ϵ_k such that ϕ satisfies Eq. (D9), i.e., ϕ is a possible angle in the counterexample. Equivalently,

$$\left| F(\phi 2^k) - \frac{1}{2} \right| > \frac{1}{2} - 2\alpha. \tag{D11}$$

Now, suppose $\alpha > \frac{1}{6}$. Then $\frac{1}{6} > \frac{1}{2} - 2\alpha$. It therefore suffices to satisfy

$$\left| F(\phi 2^k) - \frac{1}{2} \right| > \frac{1}{6} \tag{D12}$$

in order to satisfy Eq. (D11). If $\phi = \frac{2^{-j}}{3}$ with $j \leq k$, $F(\phi 2^k)$ is either $\frac{1}{3}$ or $\frac{2}{3}$, both of which satisfy Eq. (D12). For $j < k$, $\phi 2^k < 1/3$, and Eq. (D12) is again also satisfied. This proves the claim.

Appendix E: Numerical Distributions of Predicted Failures

In the numerical surveys of Section IV, 1000 realizations of RPE were simulated, and consistency checks applied on each of these RPE runs. The actual failure points were calculated via Eq. (10), and the discrepancy of the predicted and actual failure points is shown in Figs. 4 to 6. The distributions of these discrepancies are shown in the shading, but one might be additionally interested in each test's ability to respond to varying actual failure points. To address this, we include the multi-plot Fig. 9, Fig. 10, and Fig. 11, displaying the distribution of discrepancies for each actual failure rate, for each test, and for each error rate. Data is only shown for depolarization error rates 1/8, 1/32, 1/128, and 1/512. The distributions are shown as "violin plots," with widths proportional to the number of realizations with a particular generation discrepancy. The red dots are the means of the distribution, and the white dots the mode. Notice the different y-axis scale of the consecutive consistency check, shown on the last page.

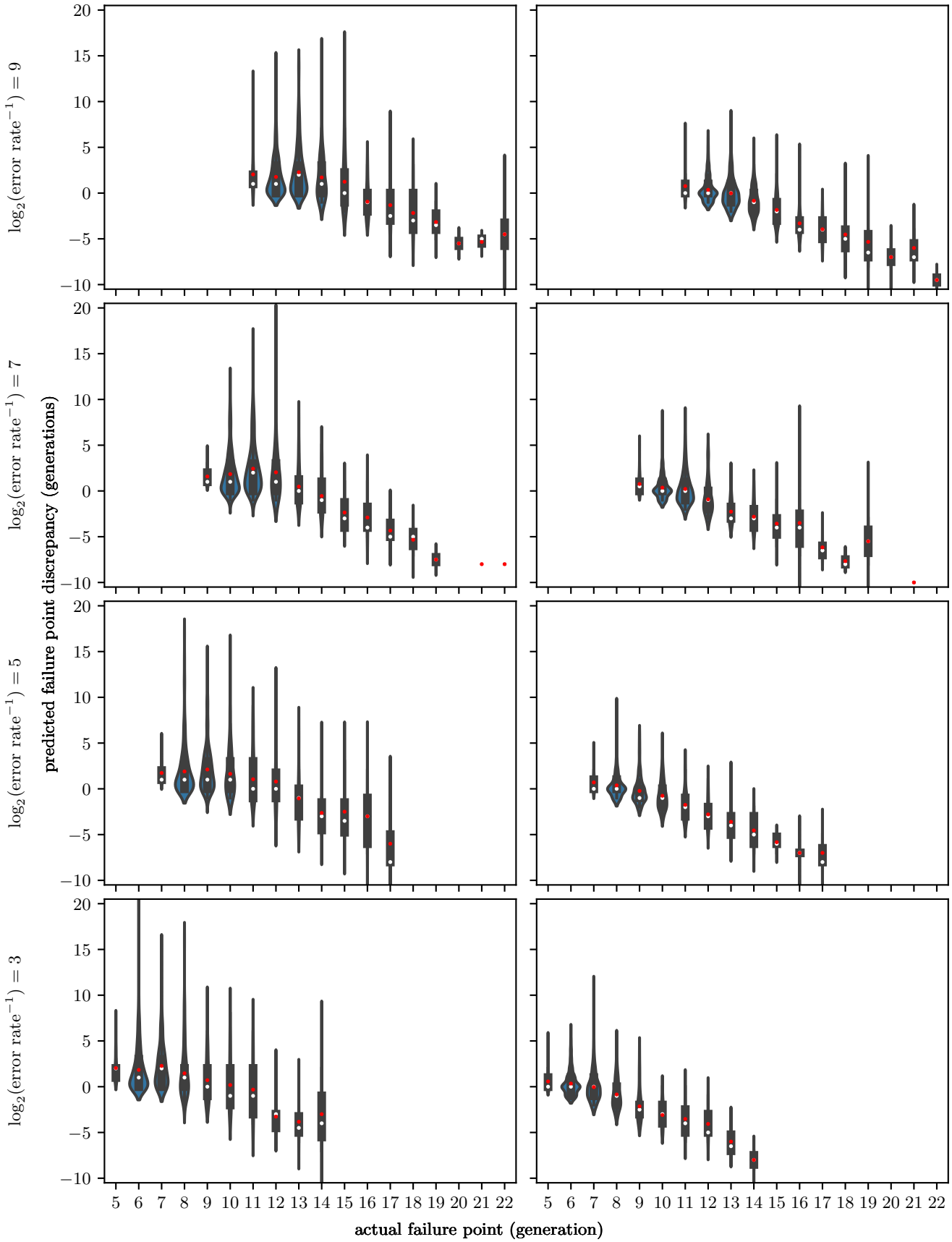

FIG. 9. Uniform-local and Intersequence consistency test failure points as a function of actual failure point.

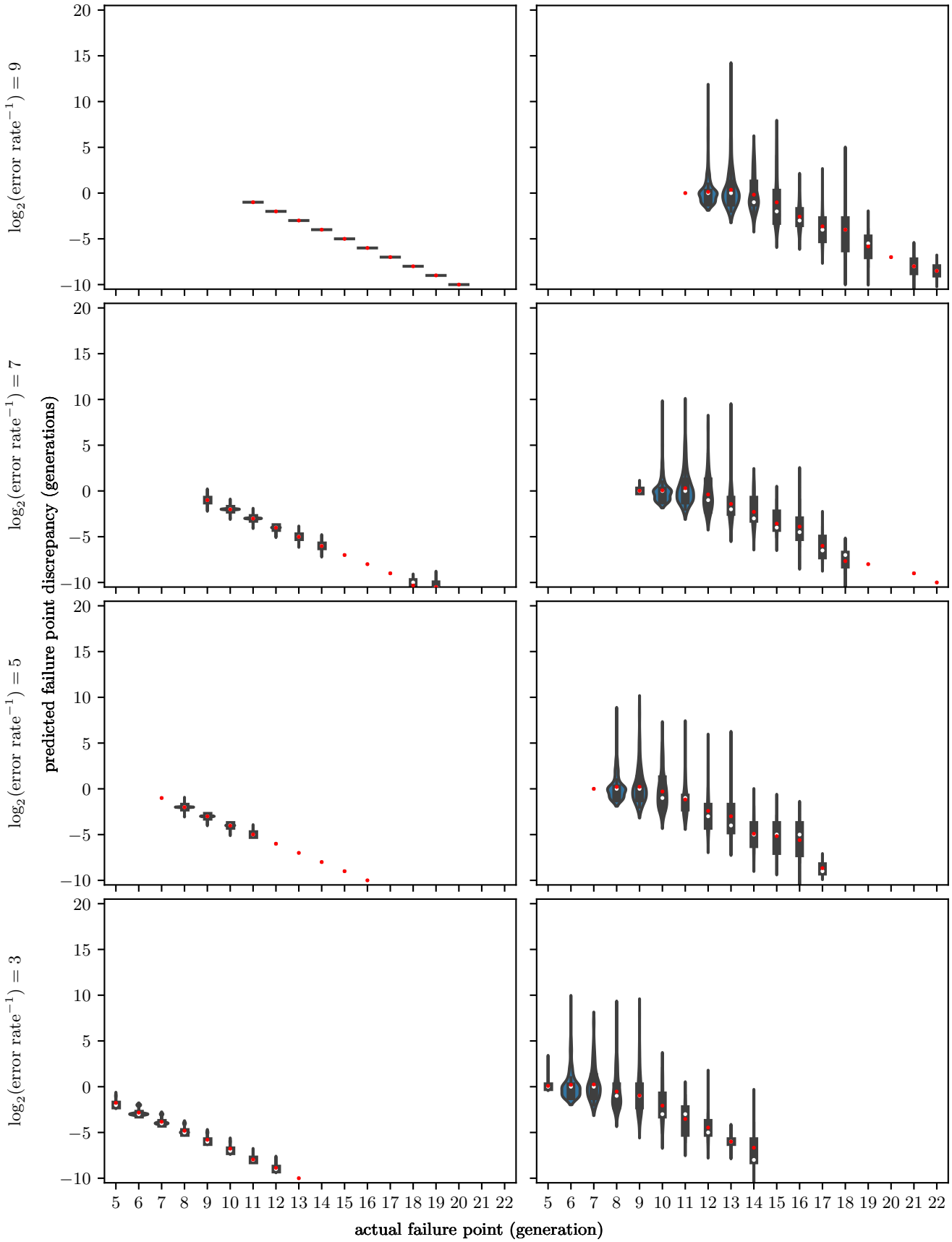

FIG. 10. Probability historical and angular historical consistency test failure points as a function of actual failure point.

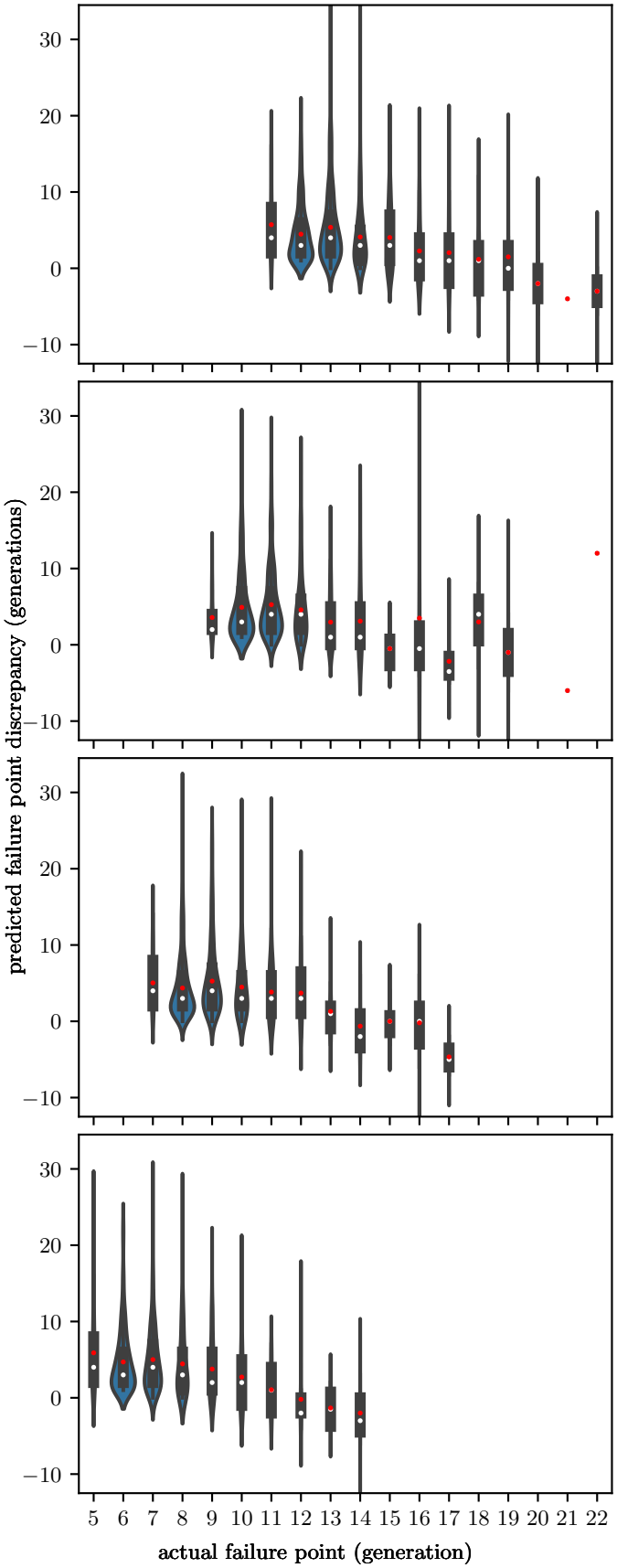

FIG. 11. Consecutive consistency test failure points as a function of actual failure point.