Wildcard error: Quantifying unmodeled errors in quantum processors

Robin Blume-Kohout,1 Kenneth Rudinger,1 Erik Nielsen,1 Timothy Proctor,1 and Kevin Young1
1Quantum Performance Laboratory, Sandia National Laboratories, Albuquerque, NM 87185 and Livermore, CA 94550
(Dated: December 23, 2020)

Error models for quantum computing processors describe their deviation from ideal behavior and predict the consequences in applications. But those processors’ experimental behavior – the observed outcome statistics of quantum circuits – are rarely consistent with error models, even in characterization experiments like randomized benchmarking (RB) or gate set tomography (GST), where the error model was specifically extracted from the data in question. We show how to resolve these inconsistencies, and quantify the rate of unmodeled errors, by augmenting error models with a parameterized wildcard error model. Adding wildcard error to an error model relaxes and weakens its predictions in a controlled way. The amount of wildcard error required to restore consistency with data quantifies how much unmodeled error was observed, in a way that facilitates direct comparison to standard gate error rates. Using both simulated and experimental data, we show how to use wildcard error to reconcile error models derived from RB and GST experiments with inconsistent data, to capture non-Markovianity, and to quantify all of a processor’s observed error.

The performance of a quantum computing processor is constrained and defined by the errors it experiences. Errors cause quantum circuits (programs) run on the processor to yield observed outcomes that deviate from their ideal, intended behavior. Probabilistic error models that describe these deviations are essential tools for assessing performance, predicting results of circuits, and informing error correction [1–11] and error mitigation [12–15] strategies. Error models can be constructed from low-level physics descriptions [16–19], device characterization experiments such as randomized benchmarking (RB) [20–23] or gate set tomography (GST) [24–26], or simple heuristics. But even carefully constructed error models often yield predictions that are visibly inconsistent with experimental data. Inconsistencies are especially problematic when an error model derived from experiments disagrees with the training data used to construct it – which, unfortunately, is common [25–28].

Inconsistency with experimental data makes it hard to take an error model seriously, or to trust its predictions for future experiments. Including more potential error sources can help models fit their training data better, but this increased sophistication incurs increased experimental and analytical costs. And even complex models are unlikely to ever capture all observed errors. The alternative solution we pursue here is to rehabilitate simple error models, by quantifying unmodeled errors and accounting for them in the model’s predictions. This is distinct from using standard statistical techniques based on hypothesis tests [24, 29] to detect statistically significant evidence of unmodeled errors, because those methods do not quantify the size or impact of unmodeled errors. If enough data is taken, even tiny unmodeled effects can generate signatures with overwhelming statistical significance!

In this Letter, we introduce wildcard models, a general technique to quantify unmodeled errors. A wildcard model for a quantum processor augments an error model, weakening its predictions to make them consistent with experimental observations (see Fig. 1). As we demonstrate with simulated and experimental data, a “wildcard error rate” can be assigned to each of a processor’s logic operations, to quantify the rate at which it induces unmodeled errors. Wildcard error rates can be used to quantify many different effects – including non-Markovian errors like leakage [30] – by augmenting an appropriate error model.

**Wildcard models.** For our purposes, an error model $M$ for a quantum processor with logic operations $\{g_i\}$ is simply a map from circuits over $\{g_i\}$ (i.e., circuits composed of the gates in $\{g_i\}$) to probability distributions over their measurement outcomes. A model $M$ predicts that running a circuit $C$ on the modeled quantum processor will result in a sample drawn from the outcome distribution $M(C) = \bar{p}_C$. Many different error models are possible. Process matrices are a common and useful framework that can be learned from experiments [24, 25] or constructed from low-level physical modeling. But to introduce wildcard models we begin with a simpler alternative: the depolarizing model. In this model, each gate acts perfectly except that it induces uniform depolarization with probability $r$. The parameter $r$ can be estimated using

![FIG. 1. Wildcard error models.](image-url)
Survival Probability over dius events. The TVD between two probability distributions $q$ and $R$ makes weaker predictions than qubit Cli model [31]. To illustrate this, we simulated RB with single-

RB data will typically not be consistent with a depolarizing RB. But if the gate errors are not depolarizing, the per-circuit

Red with logic operations \{M \} shows the simulated data and the prediction of the depolariz-

tion (dots) and 1 spec for any $C \in S$. In the RB example of Fig. 2(a), many individual circuits'

output statistics are inconsistent with the depolarizing model (by condition 1), so a wildcard model is required to reconcile them.

Assigning wildcard error to each gate: Structured wildcard models can quantify the “amount” of observed unmodeled error. To do so, we use wildcard models that assign a wildcard error rate $w_g \in [0, 1]$ to each logic operation $g$ and then define $w_C$ as the sum over $C$’s component operations,

$$w_C = \sum_{g \in C} w_g = \sum_{g} n_g(C)w_g = n(C) \cdot \mathbf{w},$$

where $n_g(C)$ is the number of times that $g$ appears in $C$. This family of wildcard models is parameterized by a vector $\mathbf{w} = \{w_g\} \over \{g\}$, so we denote a specific model in the family by $W_{\mathbf{w}}$. This “wildcard error-per-gate” parameterized family is simple, mirrors the general structure of many error models, and is faithful to the basic properties of imperfect quantum operations (e.g., subadditivity of worst-case error rates [36]).

If $n_g(C) = n_w(C)$ for all circuits $C \in S$, then only the sum $w_g + w_\epsilon$ impacts the per-circuit wildcard error $w_C$. In all examples herein, both state preparation and measurement appear once in every circuit, so only the sum of their wildcard error rates matters. To make this explicit, we write $R_C$ is correct. $R_C$ is not a distribution over distributions.

There are other contexts in which point predictions are relaxed to regions. For example, uncertainty in an estimated model’s parameters (“error bars” [32, 33]) can be represented as prediction regions [34]. But that is distinct from what we do here with wildcard error – those regions represent uncertainty about model parameters that could be reduced by taking more data, whereas wildcard models specifically represent out-of-

model effects. These two types of prediction region could be combined, but we do not attempt to do so here.

Statistical consistency: Wildcard models can weaken the predictions of an error model $M$ so that they are consistent with observed data $D = \{\bar{f}_C\}_{C \in S}$. Here, $\bar{f}_C$ is the empirical distribution from $N$ runs of the circuit $C$ and $S$ is some set of circuits over $\{g_i\}$. As $N \rightarrow \infty$, probabilities $\bar{p}$ are consistent with data $\bar{f}$ if and only if $\bar{p} = \bar{f}$, so a wildcard-augmented model $\{M, W\}$ is consistent with data $D = \{\bar{f}_C\}$ if and only if $\delta_{\text{TVD}}(\bar{p}_C, \bar{f}_C) \leq w_C \forall C \in S$.

For finite $N$, statistical consistency has no unique definition. We choose to use a 95% confidence hypothesis test that de-

fines prediction regions $[R_C]$ and data $\{\bar{f}_C\}$ as inconsistent if either [29]: (1) $\bar{f}_C$ rejects $R_C$ by a loglikelihood-ratio (LLR) hypothesis test with $\geq 95.7\%$ familywise confidence for any $C$ [35], or (2) all the data together reject $[R_C]$ by an aggregated LLR test at the 97.5% confidence level. Condition 1 captures effects with a large impact on a few circuits, while condition 2 captures effects with a small impact on many circuits. If a wildcard model satisfies these conditions we call it feasible. In the RB example of Fig. 2(a), many individual circuits’ output statistics are inconsistent with the depolarizing model (by condition 1), so a wildcard model is required to reconcile them.
\( \mathbf{w} = (w_{\text{spam}}, w_{g_1}, w_{g_2}, \ldots, w_{g_n}) \) where \( g_1, g_2, \ldots, g_n \) are unitary logic gates. Sometimes it is useful to consider a restricted model that fixes certain gates’ wildcard error rates to be equal. In our RB example (Fig. 2), we assign each gate the same amount of wildcard error rate. This defines a family of wildcard models parameterized by amount of wildcard error rate. This defines a family of wildcard logic gates. Sometimes it is useful to consider a restricted condition for \( \mathbf{W} \) for all circuits. Equivalently, \( \mathbf{W} \) with wildcard is an admission of that model’s failure. So, we want to add a minimal wildcard model with wildcard is \( \mathbf{W} \) with the data. In this example, wildcard error rates for \( g_i \) and \( g_j \), (a) One instance of simulated data (points, \( w_i \)) error bars, colored according to ideal circuit outcome) versus circuit depth and prediction regions (similarly colored) of a wildcard-augmented model that relaxes the target model’s predictions to make them consistent with the data. In this example, wildcard error rates for \( G_x \) and \( G_y \) were set equal (\( w_i = w_j \)) to facilitate visualization. (b) Results of 1000 independent simulations show that \( w_i \) (blue) and \( w_j \) (green) correlate well with the worst-case error rate (\( \epsilon \)) of the \( G_x \) and \( G_y \) gates. Each gate’s error channel was chosen independently and randomly for each simulation, and \( w_i \) was not set equal to \( w_j \).
circuit depth, from a simulation with $N = 1000$ and where $G_x$ and $G_z$ are subject to the same over-rotation and depolarization rates. It also shows the prediction regions for a minimal wildcard model $W_w$ where $w_x$ and $w_y$ were fixed to be equal (to simplify visualizing these regions). Fig. 3(b) summarizes the results of 1000 independent simulations with randomly chosen rotation and depolarization errors on the gates, comparing each wildcard error assignment to $\epsilon$, for the corresponding gate used in that simulation. There is a strong correlation, although the wildcard error for each gate is consistently slightly less than $\epsilon$, because although GST experiments are sensitive to worst-case error, tomographic analysis is required to infer it exactly.

Quantifying non-Markovian error in GST: GST fits on experimental data often reveal statistically significant model violation [25–27]. Since $n$-qubit GST is designed to capture all $n$-qubit Markovian errors, this indicates non-Markovian errors that cannot be modeled by $n$-qubit process matrices. We could quantify these out-of-model errors by expanding the GST model to include other types of errors by, e.g., using a generic higher-dimensional model [38], or modelling a specific effect like time-dependent gates [27] or leakage (which cannot be modeled as a single-qubit, completely-positive, trace-preserving process matrix). But larger models increase computational and experimental costs, and the errors outside of the GST model might be so small that they are of little practical importance. The easy alternative is to quantify the observed rates of non-Markovian errors in the GST data by augmenting the best-fit GST model with a minimal wildcard model. If the $\epsilon_i$ error rates of the GST gate estimates are much larger than the wildcard errors assigned to the gates, then the GST process matrices represent a good model for the dominant errors.

To demonstrate this analysis we applied it to simulated GST data from a qubit that suffers leakage. We simulated the GST experiments on four gates: $G_i$, $G_x$, $G_z$, and $G_{\text{idle}}$, where $G_{\text{idle}}$ is an idle. All four gates were subject to Markovian errors [39], and the $G_i$ and $G_z$ gates also caused irreversible incoherent leakage at rates of $1 \times 10^{-4}$ and $3 \times 10^{-4}$, respectively, to a state that is detected as $|0\rangle$. Without wildcard, the best-fit GST model is inconsistent with the data (Fig. 4(a)). We constructed the $\|\mathbf{w}\|_1$-minimal wildcard model that reconciles this GST model with the data. It eliminates all of the inconsistency (Fig. 4(b)) using only a tiny amount of wildcard error: $w_i = 1.1 \times 10^{-3}$, $w_x = 1.1 \times 10^{-4}$, $w_y, w_z < 10^{-6}$ and $w_{\text{max}} = 0$ (Fig. 4(d)). The GST models’ predictions are only weakened significantly for circuits that contain many $G_i$ or $G_z$ gates, and the amount of $w$ assigned to them correlates very closely with the TVD between the observed frequencies and the GST model’s predictions (Fig. 4(c)). The wildcard error rates assigned to $G_i$ and $G_z$ are lower than the underlying leakage rates, for two reasons. First, GST found process matrices that managed to explain some of the leakage errors’ effects. Second, while GST circuits are sensitive to leakage, they are not designed for maximal sensitivity to it. Neither these circuits nor the GST model are designed to distinguish leakage from Markovian errors. It is possible that jointly fitting an error model and a wildcard model to data could mitigate the artifacts induced by out-of-model effects, but this remains an open question.

As a final test, we applied wildcard-augmented GST to two sets of experimental GST data from a 2017 trapped-ion experiment on the gates $G_i$, $G_x$, and $G_z$ [25]. The GST fits to these two sets of data exhibit model violation—over $30\sigma$ in the earlier experiment, and about $6\sigma$ in the last experiment. This final experiment was described in Ref. [25] as remarkably Markovian, but even $6\sigma$ is overwhelming statistical evidence [40] — so the errors in that experiment were certainly not perfectly Markovian! To quantify the observed non-Markovianity, we augmented each best-fit GST model with the $\|\mathbf{w}\|_1$-minimal wildcard model. The data from the early experiment could be reconciled by adding $w_{\text{max}} = 0$, $w_i = 5 \times 10^{-3}$, $w_x = 1 \times 10^{-6}$ and $w_y = 1 \times 10^{-4}$ to the GST estimate that had $\epsilon_i$ error rates of $\epsilon_i = 7 \times 10^{-3}$, $\epsilon_x = 8 \times 10^{-5}$ and $\epsilon_y = 9 \times 10^{-6}$. In that experiment $w_i > \epsilon_i$ and $w_y \approx \epsilon_y$, implying that non-Markovian errors are a large proportion of the
total error, and the GST estimate is likely to be unreliable. In contrast, the final experiment required only adding $w_{\text{DAM}} = 0$, $w_1 = 2 \times 10^{-6}$, $w_2 = 7 \times 10^{-8}$, and $w_3 = 2 \times 10^{-6}$ to a GST estimate with $\epsilon_w$ error rates between $7 \times 10^{-5}$ and $1.3 \times 10^{-4}$. In this experiment, the smallest $\epsilon_w$ was around 35× larger than the largest wildcard error rate, so Markovian errors dominate and the GST estimate is reliable. This confirms and quantifies the ad hoc assertion in Ref. [25] that non-Markovian error in the final experiment is negligible compared with the Markovian errors, and it demonstrates that wildcard can be used to show that an error model captures most of the important behavior, despite being technically (and demonstrably) wrong.

**Conclusions:** Wildcard models can rehabilitate error models that are visibly wrong by qualifying their predictions, and wildcard error rates can be useful estimates of how much observed error wasn’t captured by a given model. Although we originally developed wildcard models to quantify non-Markovian errors in GST experiments, they can easily enhance other characterization protocols and other error models. Specific types of error can be quantified by constructing an error model that doesn’t account for the chosen error, adding a wildcard model, and analyzing data designed to reveal the chosen error. Crosstalk is an excellent example, because although crosstalk errors are important for quantum computing [23, 29, 41, 42], general crosstalk cannot be modelled efficiently [43]. The easy alternative is to build a crosstalk-free model [43], fit it to experiments that expose crosstalk, and use wildcard error rates to quantify the rate of crosstalk errors.Wildcard analysis is intrinsically scalable, and because scalable error models must necessarily consider only restricted types of errors, we anticipate that wildcard will be especially valuable for many-qubit characterization protocols [42, 44].

**ACKNOWLEDGMENTS**

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research Quantum Testbed Program, and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA). 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 the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA-0003525. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, the U.S. Department of Energy, or the U.S. Government.

**REFERENCES**

[1] P. D. Johnson, J. Romero, J. Olson, Y. Cao, and A. Aspuru-Guzik, QVECTOR: an algorithm for device-tailed quantum error correction, arXiv (2017), 1711.02249 [quant-ph].

[2] C. J. Trout, M. Li, M. Gutiérrez, Y. Wu, S.-T. Wang, L. Duan, and K. R. Brown, Simulating the performance of a distance-3 surface code in a linear ion trap, New J. Phys. 20, 043038 (2018).

[3] D. K. Tuckett, S. D. Bartlett, and S. T. Flammia, Ultrahigh error threshold for surface codes with biased noise, Phys. Rev. Lett. 120, 050505 (2018).

[4] P. Baireuther, T. E. O’Brien, B. Tarasinski, and C. W. J. Beenakker, Machine-learning-assisted correction of correlated qubit errors in a topological code, Quantum 2, 48 (2018).

[5] A. Piedrafita and J. M. Renes, Reliable Channel-Adapted error correction: Bacon-Shor code recovery from amplitude damping, Phys. Rev. Lett. 119, 250501 (2017).

[6] D. K. Tuckett, A. S. Darmawan, C. T. Chubb, S. Bravyi, S. D. Bartlett, and S. T. Flammia, Tailoring surface codes for highly biased noise, Phys. Rev. X 9, 041031 (2019).

[7] A. Bermudez, X. Xu, M. Gutiérrez, S. C. Benjamin, and M. Müller, Fault-tolerant protection of near-term trapped-ion topological qubits under realistic noise sources, Phys. Rev. A 100, 062307 (2019).

[8] T. E. O’Brien, B. Tarasinski, and L. DiCarlo, Density-matrix simulation of small surface codes under current and projected experimental noise, npj Quantum Information 3, 1 (2017).

[9] M. Gutiérrez, L. Svec, A. Vargo, and K. R. Brown, Approximation of realistic errors by Clifford channels and pauli measurements, Phys. Rev. A 87, 030302 (2013).

[10] E. Magesan, D. Puzzuoli, C. E. Granade, and D. G. Cory, Modeling quantum noise for efficient testing of fault-tolerant circuits, Phys. Rev. A 87, 012324 (2013).

[11] J. Florjanczyk and T. A. Brun, In-situ adaptive encoding for asymmetric quantum error correcting codes, arXiv (2016), 1612.05823 [quant-ph].

[12] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, Quantum computation with universal error mitigation on a superconducting quantum processor, Sci Adv 5, eaaw5686 (2019).

[13] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, Learning-based quantum error mitigation, arXiv (2020), 2005.07601 [quant-ph].

[14] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, arXiv (2020), 2005.10189 [quant-ph].

[15] K. E. Hamilton, T. Kharazi, T. Morris, A. J. McCaskey, R. S. Bennink, and R. C. Pooser, Scalable quantum processor noise characterization, arXiv (2020), 2006.01805 [quant-ph].

[16] D. M. Debroy, M. Li, S. Huang, and K. R. Brown, Logical performance of 9 qubit compass codes in ion traps with crosstalk errors, arXiv (2019), 1910.08495 [quant-ph].

[17] A. Bermudez, X. Xu, M. Gutiérrez, S. C. Benjamin, and M. Müller, Fault-tolerant protection of near-term trapped-ion topological qubits under realistic noise sources, Phys. Rev. A 100, 062307 (2019).

[18] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer’s guide to superconducting qubits, Applied Physics Reviews 6, 021318 (2019).

[19] R. McDermott, Materials origins of decoherence in superconducting qubits, IEEE Trans. Appl. Supercond. 19, 2 (2009).

[20] J. Emerson, R. Alicki, and K. Zyczkowski, Scalable noise estimation with random unitary operators, J. Opt. B Quantum Semiclass. Opt. 7, S347 (2005).

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

[22] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and ro-
bust randomized benchmarking of quantum processes, Phys. Rev. Lett. 106, 180504 (2011).

[23] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Direct randomized benchmarking for multiqubit devices, Phys. Rev. Lett. 123 (2019).

[24] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, arXiv preprint arXiv:2009.07301 (2020).

[25] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nat. Commun. 8, 14485 (2017).

[26] J. P. Dehollain, J. T. Muhonen, R. Blume-Kohout, K. M. Rudinger, J. K. Gamble, E. Nielsen, A. Laucht, S. Simmons, R. Kalra, A. S. Dzurak, et al., Optimization of a solid-state electron spin qubit using gate set tomography, New J. Phys. 18, 103018 (2016).

[27] T. Proctor, M. Revelle, E. Nielsen, K. Rudinger, D. Lobser, P. Maunz, R. Blume-Kohout, and K. Young, Detecting and tracking drift in quantum information processors, Nat. Commun. 11, 5396 (2020).

[28] Y. Wan, D. Kienzler, S. D. Erickson, K. H. Mayer, T. R. Tan, J. J. Wu, H. M. Vasconcelos, S. Glancy, E. Knill, D. J. Wineland, et al., Quantum gate teleportation between separated qubits in a trapped-ion processor, Science 364, 875 (2019).

[29] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, Probing context-dependent errors in quantum processors, Phys. Rev. X 9, 021045 (2019).

[30] C. J. Wood and J. M. Gambetta, Quantification and characterization of leakage errors, Phys. Rev. A 97, 032306 (2018).

[31] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, Effect of noise correlations on randomized benchmarking, Phys. Rev. A 93, 022303 (2016).

[32] R. Blume-Kohout, Robust error bars for quantum tomography, arXiv (2012), arXiv:1202.5270 [quant-ph].

[33] P. Faist and R. Renner, Practical and reliable error bars in quantum tomography, Phys. Rev. Lett. 117, 010404 (2016).

[34] S. Geisser, Predictive Inference (CRC Press, 1993).

[35] When data used to fit the model are recycled for validation (e.g., assigning \( w \)) , a small correction should in principle be made to the LLR test. We ignore this correction, which is safe as long as the number of circuits is much greater than the number of fit parameters.

[36] D. Aharonov, A. Kitaev, and N. Nisan, Quantum circuits with mixed states, in Proceedings of the thirtieth annual ACM symposium on Theory of computing (ACM, 1998) pp. 20–30.

[37] E. Nielsen, K. Rudinger, T. Proctor, A. Russo, K. Young, and R. Blume-Kohout, Probing quantum processor performance with pyGSTi, Quantum Sci. Technol. 5, 044002 (2020).

[38] R. S. Bennink and P. Lougovski, Quantum process identification: A method for characterizing non-markovian quantum dynamics, arXiv preprint arXiv:1803.02438 (2018).

[39] Consisting of depolarization and over-rotation errors with rates of \( 10^{-3} \) and \( 10^{-2} \), respectively.

[40] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. A. Khalek, A. A. Abdelalim, R. Aben, B. Abi, M. Abolins, O. AbouZeid, et al., Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc, Phys. Lett. B 716, 1 (2012).

[41] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Measuring the capabilities of quantum computers, arXiv preprint arXiv:2008.11294 (2020).

[42] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, Nat. Phys. 1 (2020).

[43] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, Detecting crosstalk errors in quantum information processors, Quantum 4, 321 (2020).

[44] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, Nat. Commun. 10, 5347 (2019).