Experimental demonstration of Pauli-frame randomization on a superconducting qubit

Matthew Ware, Guilhem Ribeill, Diego Ristè, Colm A. Ryan, Blake Johnson, and Marcus P. da Silva

1 Raytheon BBN Technologies, 10 Moulton St., Cambridge, MA 02138, USA
2 Rigetti Computing, 775 Heinz Ave., Berkeley, CA 94710, USA
(Dated: March 6, 2018)

The realization of quantum computing’s promise despite noisy imperfect qubits relies, at its core, on the ability to scale cheaply through error correction and fault-tolerance. While fault-tolerance requires relatively mild assumptions about the nature of the errors, the overhead associated with coherent and non-Markovian errors can be orders of magnitude larger than the overhead associated with purely stochastic Markovian errors. One proposal, known as Pauli frame randomization, addresses this challenge by randomizing the circuits so that the errors are rendered incoherent, while the computation remains unaffected. Similarly, randomization can suppress couplings to slow degrees of freedom associated with non-Markovian evolution. Here we demonstrate the implementation of circuit randomization in a superconducting circuit system, exploiting a flexible programming and control infrastructure to achieve this with low effort. We use high-accuracy gate-set tomography to demonstrate that without randomization the natural errors experienced by our experiment have coherent character, and that with randomization these errors are rendered incoherent. We also demonstrate that randomization suppresses signatures of non-Markovianity evolution to statistically insignificant levels. This demonstrates how noise models can be shaped into more benign forms for improved performance.

I. INTRODUCTION

Large-scale quantum computation poses a number of design and control challenges. Significant efforts are in progress [1–3] to meet and overcome challenges associated with initial state preparation, maintaining long coherence times, implementing universal gates, and measuring qubits reliably – all key criteria for building scalable quantum computers [4]. As the coherence timescales of these systems continue to grow, coherent errors can become the dominant source of error. These errors can originate from the miscalibration of qubit rotations, unintentional control frequency detunings, or interactions between systems that are otherwise assumed to be decoupled—all ubiquitous problems for experimental quantum computers. Coherent errors are also particularly difficult to simulate in multiqubit systems, as they can interfere constructively or destructively, making prediction about the performance of quantum error correction codes and fault-tolerant computation quite difficult [5–7]. Moreover, theoretical lower bounds on the tolerable rates for coherent errors indicate they may be much more damaging than stochastic errors [8–11]. One way to address this problem is to transform coherent errors into incoherent stochastic errors, such as random Pauli errors. Here we use a superconducting qubit system to implement an early proposal known as Pauli-Frame Randomization (PFR) [12, 13], and discuss some additional benefits of the randomization process, such as decoupling of slow non-Markovian noise, which we ascribe to randomized decoupling [14].

One significant challenge in determining whether PFR has indeed made coherent errors stochastic is, thanks to the community’s progress towards fault-tolerance, the magnitudes of these errors are on the order of $10^{-3}$ or less in state-of-the-art devices. Which means measuring them reliably runs into limitations of various character-

II. PAULI-FRAME RANDOMIZATION

Pauli-frame randomization is a noise-shaping technique that reduces to applying random Pauli group operations between computational gates [12, 13]. If the computational gates consist of Clifford group operations [23] (a set of operations sufficient for error correction), the effect of these random Pauli operations can be easily tracked [24, 25] so that the computation itself can be unrandomized by reinterpreting measurement results. While
this randomization is designed to have no impact on the ideal computation, it effectively *symmetrizes* the error, much like *twirling* [26–30] and randomized decoupling [14], leading to an effective error operation that corresponds to a mixture of Pauli group operations known as a Pauli channel.

These results can be derived in the limit of perfect randomization operations and gate-independent errors as follows. Throughout this discussion, we represent quantum operations by the corresponding superoperators, denoted with calligraphic upper-case letters (\(\mathcal{A}, \mathcal{B}\), etc.), while the noisy implementations are denoted in the same way but with a tilde. Consider a set of ideal (resp. noisy) Clifford group quantum operations \(\mathcal{C}_i\) (resp. \(\tilde{\mathcal{C}}_i\)). Any sequence of ideal Clifford operations can be randomized by inserting random Pauli group operations between the Clifford group operations. Since Clifford group operations transform Pauli group operation to other Pauli group operations, the overall effect of these random Pauli group operations can be cancelled out by applying a final Pauli group operation at the end of the sequence of gates. Moreover, since the Pauli group is a subgroup of the Clifford group, one may simply combine the \(i\)th random Pauli operation \(P_i\) with the \(i\)th Clifford group operation \(C_i\), to obtain a random Clifford group operations \(D_i\). In other words, a given sequence of Clifford group operations

\[
\mathcal{C}_L C_{L-1} \cdots C_2 C_1
\]

becomes

\[
\mathcal{P}_{L+1} C_L P_L C_{L-1} P_{L-1} \cdots C_2 P_2 C_1 P_1
\]

which results in the randomized sequence of Clifford group operations

\[
\mathcal{D}_L D_{L-1} \cdots D_1,
\]

executed in a randomized experiment—note that many different realizations of (3) should be measured in order to get good statistics about the original Clifford sequence (1).

It is possible to choose all \(\mathcal{P}_i\) independently at random and compensate for their action by flipping observed measurement outcomes in post-processing (as, by construction, we only measure in the computational basis). In order to simplify post-processing, we instead choose \(\mathcal{P}_{L+1}\) to cancel the effect that all other random Pauli group operations would have on measurement results (i.e., \(\mathcal{P}_{L+1}\) is a Pauli frame correction before measurement). In this way the measurement outcome of the randomized and unrandomized experiments can be treated exactly the same, with no additional post-processing for the randomized experiments.

In the presence of gate-independent imperfections, we can analyze the sequences above by replacing each operation with its noisy counterpart. We write the noisy operations \(\tilde{D}_i = \mathcal{E} D_i\) (where \(\mathcal{E}\) is an arbitrary but fixed completely-positive trace-preserving (CPTP) map) to obtain

\[
\tilde{D}_L \tilde{D}_{L-1} \cdots \tilde{D}_2 \tilde{D}_1,
\]

\[
= \mathcal{E} D_L \mathcal{E} D_{L-1} \cdots \mathcal{E} D_2 \mathcal{E} D_1,
\]

\[
= \mathcal{E} \mathcal{P}_{L+1} C_L \mathcal{P}_L \mathcal{E} C_{L-1} \mathcal{P}_{L-1} \cdots \mathcal{E} C_2 \mathcal{P}_2 \mathcal{E} C_1 \mathcal{P}_1.
\]

Defining \(\mathcal{P}^c = \mathcal{C} \mathcal{P} \mathcal{C}^\dagger\), we can write \(\mathcal{C} P = \mathcal{P}^c C\). Similarly, we define \(\mathcal{P}_{n+1} = \mathcal{P}_n \mathcal{P}_{n-1} \cdots \mathcal{P}_1\) (with the base case \(\mathcal{P}_{0,1} = \mathcal{P}_1\)). With these definitions, the entire sequence can then be rewritten as

\[
\mathcal{E} \mathcal{P}_{L+1;1} C_L \mathcal{P}_{L-1;1} \mathcal{E} \mathcal{P}_{L-2;1} C_{L-1} \mathcal{P}_{L-2;1} \cdots \mathcal{E} \mathcal{P}_{1;1} C_1 \mathcal{P}_{1;1} C_1,
\]

where, in the experiments described here, we have chosen \(\mathcal{P}_{L+1;1}\) to be the identity. In other words, we choose \(\mathcal{P}_i\) uniformly at random for \(1 \leq i \leq L\), and choose \(\mathcal{P}_{L+1}\) to get a trivial \(\mathcal{P}_{L+1;1}\). Averaging over uniformly random choices of Pauli operations, we transform each \(\mathcal{E}\) in the sequence into \(\tilde{\mathcal{E}} = \frac{1}{2L} \sum \mathcal{E} \mathcal{P}_i \mathcal{P}_i\), which correspond to twirling \(\mathcal{E}\) over the Pauli group. This, in turn, ensures that the effective error \(\tilde{\mathcal{E}}\) associated with each gate in the sequence corresponds to a statistical mixture of Pauli operations [29], as desired [31].

The calculation outlined above does require rather strong assumptions about the properties of the noise (e.g., that it is gate independent and Markovian), but due to similarities to randomized benchmarking (RB) [16, 32–36], which requires much weaker assumptions, we expect that these strong assumptions are not strictly necessary. Here we experimentally test such a hypothesis with the natural imperfections of a superconducting qubit experiment.

### III. CHARACTERIZATION AND VERIFICATION

The task of checking whether the result of applying PFR to an experiment does indeed result in a Pauli channel is subtle. Modern experiments have very high fidelity to ideal operations so checking that the unrandomized errors are not well described by Pauli channel—i.e., determining that PFR is necessary—is already challenging, since error rates can be on the order of \(10^{-3}\) or less. In both cases, it seems natural to consider long sequences of operations to amplify sensitivity to these small errors.

We choose to use long-sequence GST [19, 20] to observe these small effects, and use a readily available open-source package for experiment design and data analysis [21]. GST is an iterative procedure that refines the tomographic reconstruction of a set of gates by comparing predictions about long gate sequences to experimental observations, and adjusting the reconstruction for better agreement. Since long sequences allow for small perturbations to accumulate, this technique yields unparalleled accuracy [19, 20].

Even with a reconstruction in hand, another subtle question is how to quantify the degree of “non-Pauliness”
of a channel. We use the likelihood ratio test for this purpose [37, 38], adapting the existing functionality within the GST software [21]. The null hypothesis $H_0$ is taken to be that the statistics for each sequence in the GST experiments lead to a separate Binomial distribution of outcomes. More explicitly, for the null hypothesis we only assume the sequences correspond to reproducible experiments with well defined measurement statistics, and ignore the gate structure of the sequences. This corresponds to not making any assumption about Markovianity of the system evolution. We then consider two alternate hypotheses nested within $H_0$: that each gate in the sequence corresponds to a fixed CPTP operation acting on the system (we call this alternate hypothesis $H_1$), and that each gate in the sequence corresponds to a fixed Clifford group operation followed by a fixed Pauli stochastic error operation (we call this alternate hypothesis $H_2$). Note that $H_2$ is contained within $H_1$, since Clifford group operations and random Pauli operations (as well as their composition) are CPTP operations.

We fit data to a model under $H_0$ by maximum-likelihood estimation of the Binomial distribution parameter $p$ associated with each GST sequence. We fit data to a model under $H_1$ using progressive refinement of maximum-likelihood estimation, a heuristic developed for GST [21]. We fit data to a model under $H_2$ by projecting the fit of $H_1$ into a generalized monomial matrix determined by the corresponding noiseless Clifford group operation. The first two fits are part of the standard routines within GST, while the last fit is a small extension to the existing GST routines.

The fitting of data to a model under $H_2$ proceeds as follows. In the Pauli-Liouville representation [39–45], a Clifford group operation is a monomial matrix—each row or column has a single non-zero matrix element, and this matrix element is ±1. In the presence of a Pauli error model, a noisy Clifford group operation will be a generalized monomial matrix, where the ±1 elements of the noiseless matrix are replaced by numbers in the interval $[-1,1]$ (but the 0 matrix elements remain unchanged). Collectively, these matrix elements must live in a simplex equivalent to the probability simplex for the Pauli channel [30]. Thus, the projection of an $H_1$ model onto an $H_2$ model simply corresponds to identifying which matrix elements should be set of zero (i.e., which matrix elements are zero in the ideal gate), and then adjusting the remaining non-zero matrix elements so that the resulting matrix lies in the appropriate simplex.

A. Badness-of-fit

We quantify how well the data is explained by each of the hypotheses discussed above by computing a metric for the quality of the fits obtained, mirroring the calculation performed for generic GST fits [19–21], but extending to the hypotheses we consider here. The basis for this calculation is $L(H_i)$, the likelihood of the observed data given the model fitted under a particular hypothesis $H_i$.

Following Wilk’s theorem [38], we know the log-likelihood ratio $-2 \log \frac{L(H_i)}{L(H_0)}$ has a distribution that asymptotically approaches a $\chi^2$ distribution with degrees of freedom given by the difference in the dimensionality of the two nested hypotheses, under the assumption that the null hypothesis is true. The mean and variance of the asymptotic distribution for the log-likelihood ratio are determined by the number of degrees of freedom. Given the fitted models, the likelihood of the observations under the various hypotheses are computed, and we report the deviation between the observation and the mean predicted, in units of the standard deviation, and call this quantity $N_\sigma$ [46]. Intuitively, if this “badness-of-fit” number is large, we favor the null hypothesis (i.e., the alternate hypothesis fit is bad), but if this number is small, both the alternate and the null hypotheses are valid, and the alternate hypothesis is favored as a parsimonious model.

The log-likelihood ratios allow us to quantify whether (a) the observations are consistent with a Markovian error model (i.e., whether $H_1$ is plausible), and (b) whether the observations are consistent with a Clifford group operation with a Pauli error model (i.e., whether $H_2$ is plausible). In particular, we are interested in testing whether the answer to these questions changes when we apply PFR to our experiments. For this, it is necessary to look at the likelihood of hypotheses in different data sets, which are quantities that cannot be compared meaningfully, as some deviation may arise simply from statistical fluctuations. As will become evident in the experimental results, however, the behavior of the randomized experiments will be qualitatively different (i.e., orders of magnitude more consistent with a Markovian Pauli error model than the unrandomized experiments).

IV. EXPERIMENTS

Device parameters

To test the hypotheses of Section III, we implement the Pauli frame randomization procedure on a superconducting qubit device. The device was fabricated at BBN in collaboration with Raytheon Integrated Defense Systems and consists of four transmon qubits, designed to be similar to those described in [3]. For the experiments described in this Letter, only one qubit (Q1) is measured. Q2–4 are sufficiently far detuned from Q1 to be safely ignored. Q1 is dispersively coupled to a readout resonator with a center frequency of $\omega_r/2\pi = 7.112$ GHz, $\kappa/2\pi = 3.4$ MHz, which is in turn capacitively coupled to a quarter-wave Purcell filter with external $Q = 22$ and a center frequency of $\omega_f = 7.27$ GHz [47] enabling fast qubit readout. Q1 has a fixed 0-1 transition frequency of $\omega_q/2\pi = 4.432$ GHz with an anharmonicity $\alpha/2\pi = 308$ MHz. Coherence times measured for Q1 are $T_1 = 10 \mu s$, $T_2 = 13 \mu s$ and a Hahn echo time of $T_{\text{echo}} = 16 \mu s$. 
FIG. 1. False-color micrograph showing qubit Q₁ (red), resonator R (blue) and Purcell filter F (green). The qubit is dispersively coupled to a λ/4 readout resonator which is capacitively coupled to a Purcell filter with a Q = 22. Qubit control is done through a dedicated drive line (orange) and all qubit readout is done via a central feed line coupled to the Purcell filter.

Electronics and software stack

Making the PFR process experimentally tractable requires leveraging a complex software and hardware control infrastructure. The long-sequence GST (ℓGST) experiments we chose require many runs (1000 shots) of large sets (∼3500) of long sequences (up to ∼1000 gates depending on the germ power). Moreover, under PFR, each shot of each of these experiments must be randomized – we must run both randomized and unrandomized experiments in an interleaved fashion to ensure they experience the same noise (to the extent possible). In contrast, standard process tomography for single qubit experiments requires only 9 different sequences, each made up of 3 gates. Randomized benchmarking, on the other hand, employs long sequences, but usually only a few different sequence lengths, and a few random instances for each length.

The process begins in software with pyGST1 [21] creating standard one-qubit GST sequences. For the data presented in this Letter a germ power of \( L = 1024 \) was used to ensure the GST experiments would have high accuracy (which increases with \( L \)), and be sensitive to non-Markovianity over the long timescale comparable to coherence times. With the ℓGST sequences from pyGSTI in hand, each sequence is then randomized by inserting random Pauli group operations after every gate, as described in Section II. It is worth emphasizing the length of the randomized GST sequence is unchanged as the Pauli group operations are combined with a neighboring Clifford group operation, much like the randomized compiling proposal of Ref. [13]. This randomization occurs in the software preprocessing of the GST gate strings and produces \( N = 1000 \) uniquely randomized GST (rGST) experiments. The original, unrandomized sequences and the 1000 new randomizations respectively define the two experimental cases for comparison.

The gates strings are then passed to our compiler, QGL.jl [49], that translates qubit gate instructions into machine instructions. This involves not only mapping high-level instructions to control pulses but also time-ordering and synchronizing instruction playback between all qubit control and readout channels. This compiler, written in Julia [50] enabled a ∼4× speed-up per sequence over an earlier Python version [51]. Moreover, QGL.jl allows for parallel compilation of experiments using multithreading, increasing the speed-up proportionally to the number of available CPU cores. The compiled instructions are then passed to an arbitrary pulse sequencer (APSII) [52]. The APSII leverages pulse caching and sequencing capabilities which stores a compact representation of pulse sequences and amplitude data [48]. In the case of standard GST, the compiler exploits the repetitive structure of the GST experiments and the the control flow of the APSII for a compact sequence representation that allows uninterrupted playback of sequence instructions. Due to the nature of the randomization process, the rGST experiments lack any kind of repetition or subroutine structure which completely rules out a brute-force approach to writing all \( N = 1000 \) unique
randomizations to hardware. For these values of $L$ and $N$ we produce over 3.5 million single qubit instructions for these 1000 randomizations of $tGST$ the control hardware has to process.

To collect the data for the experiments described here, the randomized experiments were broken into sets of 10 randomizations (10 sets of $\sim$3500 sequences), in order to fit in the hardware DRAM. These 10 single shot runs of rGST are interleaved with 10 shots of unrandomized GST. The process is repeated 100 times until 1000 shots for each GST experiment are acquired. Each data point in Fig. 3 correspond to a single round of the process illustrated in Fig. 2.

Additionally, we use a “diatomic” implementation of single qubit Clifford group operations, meaning each Clifford group operation is performed with two $X_{\pi/2}$ pulses of fixed length (50 ns) and three possible Z-frame updates [45, 53]. This ensures all Clifford operations have equal duration and are composed of a single control pulse.

The room temperature measurement signals were processed with an autodyne technique described in Ref. [54] using the BBN-QDSP digitization architecture [48] for the Innovative Integrations X6-1000M digitizer card. The final state measurement is then fed into the pyGSTi package for gate set reconstruction. pyGSTi also provides the likelihood of $H_0$ and $H_1$, while custom code generates the likelihood of $H_2$.

VI. RESULTS

To test the effectiveness of of PFR we perform the experiment outlined in Sec. III. This entire experiment is repeated seven times, each taking roughly one hour to complete. This repetition allows us to observe how drift affects the results.

The question of the effect on the Pauliessness of errors and the Markovian behavior of the evolution can be seen in Fig. 3. Here the GST reconstruction quality of seven sequential experimental runs are shown with and without randomization.

Several features are immediately apparent in Fig. 3: (1) the Markovian fits to the unrandomized experiments (solid lines) are orders of magnitude worse than the randomized experiments (dashed lines), (2) the Pauli error model fits in the unrandomized experiments (solid red line) are roughly an order or magnitude worse than the randomized experiments (solid blue line), (3) the difference between the quality of the fits for the randomized experiments (blue and red dashed lines). In terms of the hypotheses we have outlined previously, for unrandomized experiments there is a large likelihood discrepancy between $H_0$ and the alternate hypotheses, greatly favoring $H_0$, while for the unrandomized experiments all hypotheses have comparable likelihood, and it is reasonable to take the simplest hypothesis ($H_2$) as the best explanation for the observations.

In other words, these features strongly indicate that the noise in the absence of randomization is not well described by a Markovian error model (in light of comment (1) above), and that even the best Markovian error model does not approximate a Pauli model well (in light of comment (2) above). Conversely, these features indicate that the noise under PFR is very well described by a Markovian Pauli error model. In much simpler terms, the features of non-Pauli error models (i.e., non-trivial off-diagonal matrix elements in the Pauli-Liouville representation [55]) are completely absent in the reconstructions of the randomized experiments, as Fig. 4 illustrates.

VI. SUMMARY

We have demonstrated that Pauli-frame randomization reduces both the non-Markovian features and the non-Pauli model features of errors in single qubit experiments. This demonstration relied on long sequence gate-set tomography, which yields high accuracy reconstructions of all the operations used in the experiments (even with sequences involving as many as 1024 operations), which in turn required a high-degree of automation to make the experiments practical. Without randomization, the experiments were shown to have strong features not present in Markovian time-independent models, and also strong features not present in Pauli error models. With randomization, we were able to show that none of these features were present, and that the experiments were well described.
by Markovian time-independent Pauli error models.

Areas for future work include speeding up the experiments using techniques such as active reset [56, 57], and pushing randomization process onto the hardware FPGA, which would allow for data acquisition of randomized Clifford group circuits without the user having to manually generate random circuits.

VII. ACKNOWLEDGEMENTS

We thank Robin Blume-Kohout, Erik Nielsen, Joel Wallman, and Joseph Emerson for fruitful discussions. We also thank Alan Howsare, Adam Moldawer and Ram Chelakara at Raytheon Integrated Defense Systems for sample fabrication. The qubit design and fabrication were funded by Internal Research and Development at Raytheon BBN Technologies and directed by Thomas Ohki. This work was funded by LPS/ARO grant W911NF-14-C-0048. The content of this paper does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

CONTRIBUTIONS

MW, GR and MPS wrote the manuscript. MW, GR, and DR performed the experiment, while CR and BJ built the control software and electronics infrastructure. GR and MW designed the device and assisted in fabrication. MPS proposed the experiment, and performed experiment design and data analysis.
[17] Cyril Stark, “Self-consistent tomography of the state-measurement gram matrix,” Phys. Rev. A 89, 052109 (2014).
[18] Seth T. Merkel, Jay M. Gambetta, John A. Smolin, Stefano Poletto, Antonio D. Córcoles, Blake R. Johnson, Colm A. Ryan, and Matthias Steffen, “Self-consistent quantum process tomography,” Phys. Rev. A 87, 062119 (2013).
[19] Robin J. Blume-Kohout, John King Gamble, Erik Nielsen, Peter Lukas Wilhelm Maunz, Travis L. Scholten, and Kenneth Michael Rudinger, Turbocharging Quantum Tomography, Tech. Rep. SAND2015-0224 (Sandia National Laboratories, 2015).
[20] Robin Blume-Kohout, John King Gamble, Erik Nielsen, Kenneth Rudinger, Jonathan Mizrahi, Kevin Fortier, and Peter Maunz, “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography,” Nature Communications 8, EP – (2017).
[21] E. Nielsen, L. Saldyt, J. Gross, T. Scholten, K. Rudinger, T. J. Proctor, and D. Nadlinger, “pygsti,” https://github.com/pyGSTio/pyGSTi (2017).
[22] John Watrous, “Semidefinite programs for completely bounded norms,” Theory of Computing 5, 217–238 (2009).
[23] Daniel Gottesman and Isaac L. Chuang, “Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations,” Nature 402, 390–393 (1999).
[24] Daniel Gottesman, “Heisenberg representation of quantum computers,” (1999).
[25] Scott Aaronson and Daniel Gottesman, “Improved simulation of stabilizer circuits,” Phys. Rev. A 70, 052328 (2004).
[26] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters, “Mixed-state entanglement and quantum error correction,” Phys. Rev. A 54, 3824–3851 (1996).
[27] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal, “Quantum data hiding,” IEEE Transactions on Information Theory 48, 580–598 (2002).
[28] Joseph Emerson, Marcus Silva, Osama Moussa, Colm Ryan, Martin Laforest, Jonathan Baugh, David G. Cory, and Raymond Laflamme, “Symmetrized characterization of noisy quantum processes,” Science 317, 1893–1896 (2007).
[29] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine, “Exact and approximate unitary 2-designs and their application to fidelity estimation,” Phys. Rev. A 80, 012304 (2009).
[30] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, “Scalable protocol for identification of correctable codes,” Phys. Rev. A 78, 012347 (2008).
[31] Although the error in the last gate of the sequence is not randomized due to our choice to have the last randomization operation cancel all others, we can treat this as a measurement error. Alternatively, we could choose all randomization operations uniformly at random and changed the measurement outcome to undo the randomization, so that effectively the error in the last gate would also be twirled over the Pauli group.
[32] J. Emerson, R. Alicki, and K. Zyczkowski, “Scalable noise estimation with random unitary operators,” J. Opt. B: Quantum Semiclassical Opt. 7, S347 (2005).
[33] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, “Randomized benchmarking of quantum gates,” Phys. Rev. A 77, 012307 (2008).
[34] Easwar Magesan, J. M. Gambetta, and Joseph Emerson, “Scalable and robust randomized benchmarking of quantum processes,” Phys. Rev. Lett. 106, 180504 (2011).
[35] E. Magesan, J. M. Gambetta, and J. Emerson, “Characterizing quantum gates via randomized benchmarking,” Phys. Rev. A 85, 042311 (2012).
[36] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, “Efficient measurement of quantum gate error by interleaved randomized benchmarking,” Phys. Rev. Lett. 109, 080505 (2012).
[37] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests of statistical hypotheses,” Philos. Trans. Royal Soc. A 231, 289–337 (1933).
[38] S. S. Wilks, “The large-sample distribution of the likelihood ratio for testing composite hypotheses,” Ann. Math. Statist. 9, 60–62 (1938).
[39] K. Blum, Density Matrix Theory and Applications (Plenum Press, 1981).
[40] Debbie W. Leung, Towards Robust Quantum Computation, Ph.D. thesis, Stanford University (2000).
[41] Benjamin Rahn, Andrew C. Doherty, and Hideo Machuki, “Exact performance of concatenated quantum codes,” Phys. Rev. A 66, 032304 (2002).
[42] J. Fern, J. Kempe, S. N. Simic, and S. Sastry, “Generalized performance of concatenated quantum codes—a dynamical systems approach,” IEEE Transactions on Automatic Control 51, 448–459 (2006).
[43] Jerry M. Chow, Jay M. Gambetta, A. D. Córcoles, Seth T. Merkel, John A. Smolin, Chad Rigetti, S. Poletto, George A. Keefe, Mary B. Rothwell, J. R. Rozen, Mark B. Ketchen, and M. Steffen, “Universal quantum gate set approaching fault-tolerant thresholds with superconducting qubits,” Phys. Rev. Lett. 109, 060501 (2012).
[44] Shelby Kimmel, Marcus P. da Silva, Colm A. Ryan, Blake R. Johnson, and Thomas Ohki, “Robust extraction of tomographic information via randomized benchmarking,” Phys. Rev. X 4, 011050 (2014).
[45] Blake R Johnson, Marcus P da Silva, Colm A Ryan, Shelby Kimmel, Jerry M Chow, and Thomas A Ohki, “Demonstration of robust quantum gate tomography via randomized benchmarking,” New Journal of Physics 17, 113019 (2015).
[46] Equivalently, one can compute the $p$-value for the null hypothesis, but we find $N_0$ to be a quantity that is easier to interpret.
[47] Evan Jeffrey, Daniel Sank, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O’Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Ciepl, and John M. Martinis, “Fast accurate state measurement with superconducting qubits,” Phys. Rev. Lett. 112, 190504 (2014).
[48] Colm A. Ryan, Blake R Johnson, Diego Rist, Brian Donovan, and Thomas A. Ohki, “Hardware for dynamic quantum computing,” Review of Scientific Instruments 88, 104703 (2017).
[49] Colm Ryan, Diego Ristè, and Blake Johnson, “QGL.jl master branch (august 2016).” (2016).
[50] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B. Shah, “Julia: A fresh approach to numerical
Appendix A: Error metrics under randomization

As expected the randomization process also suppressed the amount of coherent error present the gate set. In Fig. 5 the effect of PFR on fidelity and diamond-norm error metrics are plotted. In the case of infidelity we see no appreciable difference in the reconstructed infidelity for the gate set, while the diamond-norm changes significantly in the presence of PFR. This is a consequence of the fact that twirling simply averages over different unitaries that individually have roughly the same fidelity to the target gate. Since the overall infidelity is the average of the individual infidelities, we see no change in that quantity. However, the overall diamond distance is not simply the average of the diamond distances, and averaging over different unitaries results in an incoherent process which is harder to distinguish from the target unitary, and so the diamond distance is reduced under PFR. This highlights that the infidelity is insensitive to coherent errors, and that the error rate, as measured by the diamond distance, is greatly improved under PFR.

computing,” SIAM Review 59, 65–98 (2017).
[51] “QGL,” (2018).
[52] “BBN APS II,” (2017).
[53] David C. McKay, Christopher J. Wood, Sarah Sheldon, Jerry M. Chow, and Jay M. Gambetta, “Efficient \( z \) gates for quantum computing,” Phys. Rev. A 96, 022330 (2017).
[54] Colm A Ryan, Blake R Johnson, Jay M Gambetta, Jerry M Chow, Marcus P da Silva, Oliver E Dial, and Thomas A Ohki, “Tomography via correlation of noisy measurement records,” Physical Review A 91, 022118 (2015).
[55] Shelby Kimmel, Marcus P. da Silva, Colm A. Ryan, Blake R. Johnson, and Thomas Ohki, “Robust extraction of tomographic information via randomized benchmarking,” Phys. Rev. X 4, 011050 (2014).
[56] M. A. Rol, C. C. Bultink, T. E. O’Brien, S. R. de Jong, L. S. Theis, X. Fu, F. Luthi, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, D. Deurloo, R. N. Schouten, F. K. Wilhelm, and L. DiCarlo, “Restless tuneup of high-fidelity qubit gates,” Phys. Rev. Applied 7, 041001 (2017).
[57] D. Ristè, C. C. Bultink, K. W. Lehnert, and L. DiCarlo, “Feedback control of a solid-state qubit using high-fidelity projective measurement,” Phys. Rev. Lett. 109, 240502 (2012).
FIG. 5. (Left) Gate set infidelity and (Right) diamond norm distance estimates for all eight experiments. Both the randomized (dashed lines) and unrandomized (solid lines) experiments are plotted but the two cases show no appreciable difference in gate infidelity. All quantities are computed from the reconstructed process matrices. For the randomized case (dashed lines), the infidelity (left) and the diamond norm (right) are comparable. In the case of the unrandomized experiments, there is significant deviation between the diamond norm error rate and the infidelity. This suggests randomization process is substantially suppressing coherent errors which do not effect the infidelity metric. Error bars in both plots are smaller than the data points.