

Fault-Tolerant Logical Gates in the IBM Quantum Experience

Robin Harper^{1,*} and Steven T. Flammia^{1,2}

¹*Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, Australia*

²*Yale Quantum Institute, Yale University, New Haven, Connecticut 06520, USA*



(Received 23 June 2018; published 26 February 2019)

Quantum computers will require encoding of quantum information to protect them from noise. Fault-tolerant quantum computing architectures illustrate how this might be done but have not yet shown a conclusive practical advantage. Here we demonstrate that a small but useful error detecting code improves the fidelity of the fault-tolerant gates implemented in the code space as compared to the fidelity of physically equivalent gates implemented on physical qubits. By running a randomized benchmarking protocol in the logical code space of the $[4,2,2]$ code, we observe an order of magnitude improvement in the infidelity of the gates, with the two-qubit infidelity dropping from 5.8(2)% to 0.60(3)%. Our results are consistent with fault-tolerance theory and conclusively demonstrate the benefit of carrying out computation in a code space that can detect errors. Although the fault-tolerant gates offer an impressive improvement in fidelity, the computation as a whole is not below the fault-tolerance threshold because of noise associated with state preparation and measurement on this device.

DOI: [10.1103/PhysRevLett.122.080504](https://doi.org/10.1103/PhysRevLett.122.080504)

We have entered an exciting stage in the development of quantum computers. Small-scale, prototype quantum devices with a limited number of qubits are beginning to appear, and companies such as IBM and Rigetti are making such devices available in the cloud. Although current quantum devices tend to be too small, have limited interconnectivity between qubits, and are too noisy to allow meaningful quantum computation, they are an important step forward in the aim to build a large-scale universal quantum computer. These small devices are sufficient to act as the test bed for proof of principle concepts such as the implementation of simplified quantum algorithms [1], quantum walks [2], quantum machine learning [3], and testing the ability to detect and correct errors.

Codes that can be implemented on current small noisy quantum devices are perfect test beds for the ideas of fault tolerance. The question we address here is whether such a code can conclusively show a benefit for encoded computation on current small-scale noisy quantum devices. As we discuss later, it is not obvious that the type of noise that exists in current quantum devices will be amenable to such codes. We study the $[4,2,2]$ error detecting code, which is one of the smallest interesting codes. For instance, it can be concatenated with the toric code [4,5], can be viewed as one of the faces of the distance 3 color code [6], or alternatively, as an encoding layer of the C_4/C_6 code of Ref. [7].

Experimental quantum error correction and fault tolerance is still in its infancy, but several impressive results have been achieved. Reference [8] demonstrates the ability to prepare logical qubits of the $[4,2,2]$ code fault tolerantly using trapped atomic ions, and Ref. [9] replicates this on superconducting qubits. References [10,11] implement

repetition codes and observe protection against bit-flip errors. Reference [12] uses quantum error correction to extend the lifetime of quantum information stored in a superconducting resonator, and Refs. [13,14] implement logical gates inside of a quantum code.

Recently, Gottesman [15] suggested that the $[4,2,2]$ code could be implemented fault tolerantly with only five physical qubits. He argued that for a small experiment to conclusively demonstrate fault tolerance, the following must be met. (1) The encoded circuit must have a lower error rate for all circuits in the family of circuits of interest. (2) They must be complete circuits; i.e., they must include the initial state preparation and the final measurements. (3) The original circuit and the encoded circuit must, in the absence of error, produce the same output distribution. (4) It is only meaningful to compare error rates between circuits implemented in the same system. Additionally, Gottesman noted that for the purposes of demonstrating fault tolerance, it is still interesting to consider circuits created from nonuniversal gate sets.

Vuillot [16] attempted to meet Gottesman's criteria using the $[4,2,2]$ code on the IBM Quantum Experience (IQX). Vuillot comprehensively explored the different types of circuits available in the $[4,2,2]$ code, though the methodology used meant it was difficult to extract a clear signal from the data as to the overall benefit (or lack thereof) in utilizing the $[4,2,2]$ code on that noisy device. More recently, Willsch *et al.* [17] simulated the $[4,2,2]$ code numerically and tested it on the IQX. Their analysis suggests that a fault-tolerant protocol could improve performance, provided that errors are due to control errors rather than dominated by decoherence.

While the inclusion of state preparation and measurement (SPAM) errors in Gottesman’s proposed protocol are necessary steps in demonstrating that a computer might meet the fault-tolerance threshold, the inclusion of such errors is not necessary in demonstrating that the use of fault-tolerant gates and appropriate logical encoding improve the fidelity of the logical gates compared to their physical counterparts. In particular, an improvement in the fidelity of such logical gates will address whether the theoretical benefits expected from encoding quantum information can be physically realized or if the noise profile (such as correlated errors) will preclude any such benefit. The answer to this question is not obvious. Consequently, rather than examining the error rate of particular circuits created from fault-tolerant gates, we measure the infidelity of the Clifford group elements created from such gates and the circuits arising from such Clifford elements. Given the proposed uses of the [4,2,2] code discussed above, the ability to accurately create such gates is of high interest. We look at the average infidelity of the gates (defined later) and demonstrate a conclusive advantage to using such fault-tolerant gates to generate these elements and circuits. We do this by utilizing the technique of randomized benchmarking (RB) [18,19] to measure the gate errors in a way that is robust to SPAM errors to show that the logical two-qubit gates outperform their physical counterparts. Using RB in the logical code space [20] of the [4,2,2] code, we demonstrate that the two-qubit average infidelity decreases from 5.8(2)% to 0.60(3)%, an improvement of roughly an order of magnitude.

However, our results do *not* imply that the IBM Q Rueschlikon device is operating below the fault-tolerance threshold. As emphasized by Gottesman [15], the threshold involves improving all aspects of a quantum circuit after the encoding, including the SPAM errors, and our approach using RB is insensitive to SPAM. Thus, while the fault-tolerant encoded gates do improve gate fidelity, the device is not yet convincingly below threshold for complete circuits.

Background.—RB provides an efficient method for the partial characterization of the quality of a gate implementation in a way that is robust to SPAM and to small, arbitrary gate-dependent noise [21,22]. It gives the fidelity between the identity channel and the averaged noise \mathcal{E} over the gate set. The (Haar) average fidelity is

$$F(\mathcal{E}) = \int d\psi \langle \psi | \mathcal{E}(|\psi\rangle\langle\psi|) | \psi \rangle, \quad (1)$$

and the average infidelity is $1 - F$.

Recent work by Carignan-Dugas *et al.* [23] relates average fidelity to the gate set circuit fidelity, a quantity which compares all possible sequences of circuits of m noisy gates implemented from the noisy gate set $\tilde{\mathcal{G}}$ to their ideal analog in \mathcal{G} . A *circuit* is a sequence of m elementary

gates, and the average gate set circuit fidelity is $\mathcal{F}(\tilde{\mathcal{G}}, \mathcal{G}, m) := \mathbb{E}[F(\tilde{\mathcal{G}}_{m:1}, \mathcal{G}_{m:1})]$, where $\tilde{\mathcal{G}}$ and \mathcal{G} are noisy and ideal gates drawn from the respective gate sets. Reference [23] proves for a single qubit (and conjectures, with numerical evidence, for two qubits) that, other than a potential SPAM mismatch, the two fidelities are closely related. This confirms the average infidelity of a gate set as an appropriate metric even when one is considering the fidelity of circuits built from such gates.

The four-qubit code.—The [4,2,2] code is defined by the stabilizer generators $XXXX$ and $ZZZZ$ [24,25]. Reference [15] details how to implement it in a fault-tolerant manner on a system with limited connectivity. To measure the logical qubits in the computational basis, one simply measures all four qubits. Odd parity heralds an error and the run should be discarded.

There are eight logical gates that make up the code gate set. They split into six “active” gates and two “virtual” gates that can be implemented in software by relabeling the physical qubits. These are shown in the table below, where P is the phase gate [$\text{diag}(1, i)$], SWAP_{ij} swaps qubits i and j , $C-Z$ is the controlled- Z gate, and CNOT_{ij} acts from control i to target j .

Physical gates	Logical equivalent
$X^{\otimes 4}, Z^{\otimes 4}$	$I \otimes I$
$X \otimes I \otimes X \otimes I$	$X \otimes I$
$X \otimes X \otimes I \otimes I$	$I \otimes X$
$Z \otimes Z \otimes I \otimes I$	$Z \otimes I$
$Z \otimes I \otimes Z \otimes I$	$I \otimes Z$
$H \otimes H \otimes H \otimes H$	$\text{SWAP}_{12} \circ (H \otimes H)$
$P \otimes P \otimes P \otimes P$	$(Z \otimes Z) \circ (C-Z)$
SWAP_{12}	CNOT_{12}
SWAP_{13}	CNOT_{21}

The final ingredient is the logical $|\overline{00}\rangle$ state preparation. Although Ref. [15] suggests a method for doing this fault tolerantly, the architecture of the IBM-QX5 means implementation is costly in terms of gates required. Since RB is robust to SPAM, it was decided not to prepare the $|\overline{00}\rangle$ state in a fault-tolerant manner (see Ref. [26] for a fuller discussion).

RB with the real-Clifford group.—RB uses long sequences of gates with the aim to amplify small errors in the implementation of these gates. By choosing the sequences of gates from a unitary 2-design [29,30], the average noise channel \mathcal{E} over random sequences of such gates reduces to a depolarizing channel with the same average fidelity $F(\mathcal{E})$.

By sampling over a unitary 2-design, the integral in Eq. (1) is replaced by a sum over the design. Often, the Clifford group is chosen as the unitary 2-design; however in this case the phase gate cannot be implemented in the [4,2,2] code in a fault-tolerant manner. If we limit ourselves to fault-tolerant gates, the lack of a phase gate makes the Clifford group inaccessible.

Our results make essential use of a modified version of RB called real RB [31,32]. In real RB, one can relax the unitary 2-design condition to any *orthogonal* 2-design and still effectively perform RB with the following protocol. (1) Choose a sequence length m and prepare a state (ρ) traditionally in the computational basis (but see later). (2) Apply a chosen number of random gates m independently and uniformly drawn from the orthogonal 2-design followed by a further *inversion gate*, which ideally would result in a net sequence equal to the identity. (3) Measure ρ with E (an effect operator of a positive-operator valued measure) to determine if it has been returned to the starting state.

The above steps are repeated in order to estimate the survival probability \bar{q} over a range of sequence lengths. Reference [32] proves that this can be fit to the model

$$\bar{q}(m, E, \rho) = A + b^m B + c^m C, \quad (2)$$

where, with $\rho_{\pm} = \frac{1}{2}(\rho \pm \rho^T)$, the constants A , B , and C are

$$A = \frac{1}{d} \text{Tr}[E], \quad B = \text{Tr}[E\rho_+] - A, \quad C = \text{Tr}[E\rho_-]. \quad (3)$$

For two qubits, we also have $F(\mathcal{E}) = (9b + 6c + 5)/20$.

The code gate set generates a subgroup of $\mathbf{C}(2)$, the two-qubit Clifford group that we call the realizable group, $\mathbf{R}(2)$. The group $\mathbf{R}(2)$ has only 576 elements, in contrast to the 11 520 elements of $\mathbf{C}(2)$. Moreover, the average number of elementary gates in a group element is reduced from about seven in $\mathbf{C}(2)$ to only about four in $\mathbf{R}(2)$, making it more efficient for RB as well. Finally, $\mathbf{R}(2)$ provably yields an orthogonal 2-design, so the real RB protocol applies [26]. There are subtleties that may impact when fitting the above formulas to an error detecting experiment, although these did not impact our analysis. For a fuller discussion, see Ref. [26].

Implementation details.—We implemented logical RB and physical RB on the device using $\mathbf{R}(2)$ as the twirling group to measure the average infidelities. If there is an advantage to computing in the logical space defined by the [4,2,2] code, then the average infidelity will be lower.

In the Supplemental Material [26], we detail how we perform RB in such a way as to eliminate the fitting parameter A . In addition, for certain obvious choices of ρ , B or C becomes zero [32]. By choosing an appropriate ρ , we can then fit the sequences from those experiments to a simpler fitting sequence. In the computational basis, $\rho = |00\rangle\langle 00|$ and $\rho_- = 0$, thus, $C = 0$. Consequently, if we perform the RB protocol discussed in Ref. [26] with $\rho = |00\rangle\langle 00|$, we can fit to the simplified formula $\bar{q}(m, |00\rangle\langle 00|) = Bb^m + 0.25$ to determine b . This is similar to the ideas discussed in Refs. [33–36] and yields improved parameter estimates for the same sample size.

To determine the c parameter, we need to move outside the computational basis for our choice of ρ . In this case, the optimal method is to rotate ρ by applying a phase gate followed by a Hadamard gate, inverting this after the twirl,

and measuring $\langle 00|$ in the normal way. Since a fault-tolerant phase gate cannot be applied in the code space, to measure c we need to use a non-fault-tolerant phase gate together with the encoded Hadamard gate; see [26] for details. When working outside the code, we can implement the phase and Hadamard gates directly.

In total, this fit procedure requires that we perform four types of runs for the [4,2,2] code: (1) the run in the computational basis (a standard run) and (2) the run initialized with the non-fault-tolerant phase gate (a phased run), and for the physical qubits (3) the standard run in the computational basis and (4) the phased run.

Analysis.—The four averaged fitting charts are shown in Fig. 1(a). This allows us to extract values for b and c and

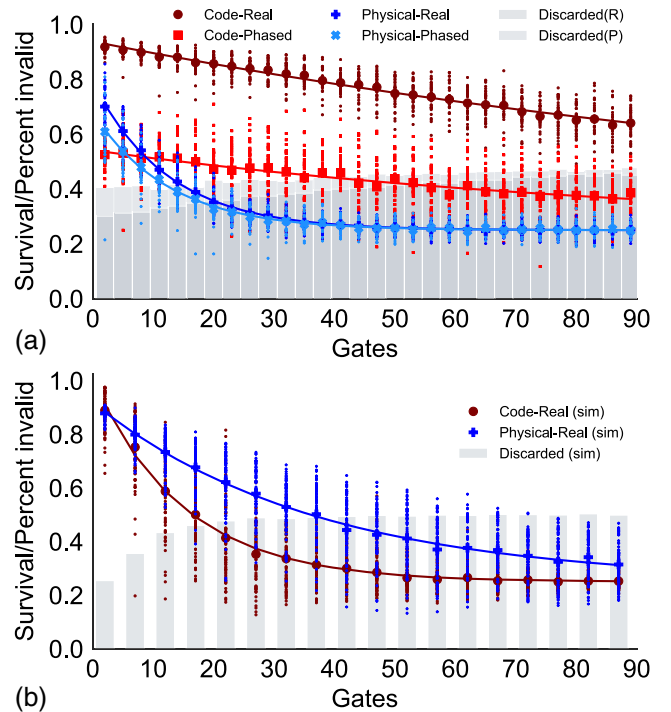


FIG. 1. (a) Fitting charts for the four different types of runs used to extract the b and c parameters in Eq. (2). The data were fit as described in the Supplemental Material [26]. Postselection on invalid code states was used to throw away runs where an error was detected. The bars show the percentage of detected (and discarded) runs for the [4,2,2] code runs. The number of initial errors are greater in the phase run because of the need to perform an initial non-fault-tolerant gate (see text). (b) Simulation showing survival probabilities for logical “real-Clifford” RB using the [4,2,2] code and the equivalent physical two-qubit RB. (Here, we have only shown the “real” runs, not the “phased” runs.) The same high fidelity four-qubit noise map (fidelity ≈ 0.99) is used for the runs, although additional noise is applied to the CNOTs of the run on physical qubits. The noise map contains single-qubit errors as well as ZZ errors between neighboring qubits. Although the combined noise map is high fidelity, the correlated errors are sufficiently strong that the fidelity *decreases* when using encoded gates, meaning the encoded computation performs worse than the bare physical computation in the correlated noise scenario.

calculate the fidelity or infidelity using Eq. (1). The calculations yield an infidelity of 0.60(3)% for the [4,2,2] code and 5.8(2)% for the two physical qubits. The uncertainties were obtained through bootstrapping the experimental data [26].

The question then remains: How much of the benefit is derived from the ability of the code to detect errors, and how much is derived from the virtual CNOTs once computation has been moved into the code space?

We can use the arguments presented in Ref. [20] to determine this. If, instead of using postselection to discard runs that we know are in an erroneous code state, we fit to the complete set of runs obtained for the [4,2,2] code, then this gives us a method of comparing the performance of the code where we are detecting errors and one where we are not. With the caveats that these runs were conducted at different times from the main runs and that a naive RB fit to the data was used (see Ref. [26]), this formulation gave an infidelity that was less than 5.8% but still significantly higher than the infidelity with error detection. This confirms two things. First, there is indeed a benefit in moving computation into a code that allows virtualization of the CNOTs, and second, there is still an additional substantial benefit to using the code to detect errors.

Comparison to correlated noise.—Data on the quantum devices made available by IBM show that there is significant cross talk between connected qubits. One of the interesting questions answered by this Letter is whether such cross talk causes correlations strong enough to defeat any fidelity gains that might occur from using encoded gates operated in a fault-tolerant way. To illustrate why this might be the case, we note that a natural noise model for such correlated noise is to model it as noise of the form $e^{i\theta Z_i Z_{i+1}}$ for small θ . This model accounts for some of the correlated noise observed in recent ion trap experiments [37] and analyzed in Ref. [38].

In Fig. 1(b) we show, by simulation, that even with a high fidelity noise map, the correlated noise can be sufficient to prevent any benefit being observed from computing in the logical code space of the [4,2,2] code.

Discussion.—A useful error detecting code consists of (1) a code space in which logical quantum gates can be performed on the encoded quantum information, (2) an ability to perform certain fault-tolerant gates within the code space, and (3) the ability to detect if a certain number of errors (limited by the distance of the code) have occurred. We have demonstrated how variations of the RB protocol [20,32] can be used to determine, in a well-defined and principled manner, whether the fault-tolerant quantum gates that are supported by such a code space can be performed with lower infidelity than such equivalent gates in the raw physical qubits. These specific RB variations were essential since the [4,2,2] code is not able to perform the full Clifford group fault tolerantly.

The substantial decrease in the infidelity we have observed appears to come from two sources. The first is

that it virtualizes the noisiest gate, the CNOT gate, resulting in a decrease in the infidelity of the averaged gate set noise. The second is that by allowing error detection (and subsequent postselection), we see a further decrease in the infidelity of the averaged gate set, leading to an overall decrease in infidelity of the fault-tolerant gate set by a factor of 10.

We have learned from these experiments that the IBM device does not have noise correlations that are so strong as to preclude an improved fidelity when using an encoded gate. As we have shown via numerical simulations [Fig. 1(a)], a natural noise model [37,38] involving two-qubit correlated errors between neighbors *can* preclude an improved fidelity even when the average error rates are comparable to those we have observed in the IBM device. More experiments are required to fully understand the role that noise correlations play in this device in the context of fault tolerance.

The analysis of Ref. [23] confirms that the average fidelity as measured by RB is closely related (as particularized in their analysis) to the average fidelity for all relevant circuits. Given this, it is clear that the average gate set fidelity will be substantially higher if the fault-tolerant gates are used, and we argue that this suffices for a conclusive demonstration of the benefit of using an error detecting code with fault-tolerant gates. Recent work has also shown [39,40] that encoded noise tends to be less coherent than unencoded noise. Consequently, an improvement in average gate fidelity is also likely to herald an improvement in other metrics relevant for fault tolerance such as the diamond distance, although more experiments would be required to establish this conclusively.

There is at present no complete theory for fitting a RB decay curve with postselection. While this had minimal impact on our present study (for the reasons discussed in Ref. [26]), developing a full theory will be essential for more detailed studies. A further exciting step would be to employ the techniques used here together with those discussed in Ref. [20] to benchmark error-corrected gates. While devices with sufficient connections to support one, or maybe a few, error-corrected qubits are on the verge of becoming available for general experimentation, current implementations do not yet have conditional measurements or reinitialization. In the meantime, error detecting codes on more qubits and implementing different fault-tolerant gate sets can be examined using the techniques discussed above.

We thank D. Gottesman for comments on our draft and T. Yoder for highlighting the impact of postselection on the RB decay curve. This work was supported by the U.S. Army Research Office Grants No. W911NF-14-1-0098 and No. W911NF-14-1-0103, and by the Australian Research Council Centre of Excellence for Engineered Quantum Systems Grant No. CE170100009.

*Corresponding author.

robin.harper@sydney.edu.au

- [1] P. J. Coles, S. Eidenbenz, S. Pakin, A. Adedoyin, J. Ambrosiano, P. Anisimov, W. Casper, G. Chennupati, C. Coffrin, H. Djidjev *et al.*, Quantum algorithm implementations for beginners, [arXiv:1804.03719](https://arxiv.org/abs/1804.03719).
- [2] R. Balu, D. Castillo, and G. Siopsis, Physical realization of topological quantum walks on IBM-Q and beyond, [arXiv:1710.03615](https://arxiv.org/abs/1710.03615).
- [3] D. Ristè, M. P. da Silva, C. A. Ryan, A. W. Cross, A. D. Córcoles, J. A. Smolin, J. M. Gambetta, J. M. Chow, and B. R. Johnson, Demonstration of quantum advantage in machine learning, *npj Quantum Inf.* **3**, 16 (2017).
- [4] C. G. Brell, S. T. Flammia, S. D. Bartlett, and A. C. Doherty, Toric codes and quantum doubles from two-body Hamiltonians, *New J. Phys.* **13**, 053039 (2011), [arXiv:1011.1942](https://arxiv.org/abs/1011.1942).
- [5] B. Criger and B. Terhal, Corrected quantum walk for optimal Hamiltonian simulation, *Quantum Inf. Comput.* **16**, 1261 (2016), [arXiv:1604.04062](https://arxiv.org/abs/1604.04062).
- [6] H. Bombin and M. A. Martin-Delgado, Topological Quantum Distillation, *Phys. Rev. Lett.* **97**, 180501 (2006).
- [7] E. Knill, Quantum computing with realistically noisy devices, *Nature (London)* **434**, 39 (2005).
- [8] N. M. Linke, M. Gutierrez, K. A. Landsman, C. Figgatt, S. Debnath, K. R. Brown, and C. Monroe, Quantum computing with realistically noisy devices, *Sci. Adv.* **3**, e1701074 (2017).
- [9] M. Takita, A. W. Cross, A. D. Córcoles, J. M. Chow, and J. M. Gambetta, Experimental Demonstration of Fault-Tolerant State Preparation with Superconducting Qubits, *Phys. Rev. Lett.* **119**, 180501 (2017), [arXiv:1705.09259](https://arxiv.org/abs/1705.09259).
- [10] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen *et al.*, State preservation by repetitive error detection in a superconducting quantum circuit, *Nature (London)* **519**, 66 (2015), [arXiv:1411.7403](https://arxiv.org/abs/1411.7403).
- [11] J. R. Wootton and D. Loss, Repetition code of 15 qubits, *Phys. Rev. A* **97**, 052313 (2018), [arXiv:1709.00990](https://arxiv.org/abs/1709.00990).
- [12] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang *et al.*, Extending the lifetime of a quantum bit with error correction in superconducting circuits, *Nature (London)* **536**, 441 (2016), [arXiv:1602.04768](https://arxiv.org/abs/1602.04768).
- [13] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell *et al.*, Superconducting quantum circuits at the surface code threshold for fault tolerance, *Nature (London)* **508**, 500 (2014), [arXiv:1402.4848](https://arxiv.org/abs/1402.4848).
- [14] R. W. Heeres, P. Reinhold, N. Ofek, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, Implementing a universal gate set on a logical qubit encoded in an oscillator, *Nat. Commun.* **8**, 94 (2017).
- [15] D. Gottesman, Quantum fault tolerance in small experiments, [arXiv:1610.03507](https://arxiv.org/abs/1610.03507).
- [16] C. Vuillot, Is error detection helpful on IBM 5Q chips?, *Quantum Inf. Comput.* **18**, 0949 (2018), [arXiv:1705.08957](https://arxiv.org/abs/1705.08957).
- [17] D. Willsch, M. Nocon, F. Jin, H. De Raedt, and K. Michielsen, Testing quantum fault tolerance on small systems, *Phys. Rev. A* **98**, 052348 (2018), [arXiv:1805.05227](https://arxiv.org/abs/1805.05227).
- [18] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, *J. Opt. B* **7**, S347 (2005).
- [19] 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).
- [20] J. Combes, C. Granade, C. Ferrie, and S. T. Flammia, Logical randomized benchmarking, [arXiv:1702.03688](https://arxiv.org/abs/1702.03688).
- [21] J. J. Wallman, Randomized benchmarking with gate-dependent noise, *Quantum* **2**, 47 (2018).
- [22] S. T. Merkel, E. J. Pritchett, and B. H. Fong, Randomized benchmarking as convolution: Fourier analysis of gate dependent errors, [arXiv:1804.05951](https://arxiv.org/abs/1804.05951).
- [23] A. Carignan-Dugas, K. Boone, J. J. Wallman, and J. Emerson, From randomized benchmarking experiments to gate-set circuit fidelity: How to interpret randomized benchmarking decay parameters, *New J. Phys.* **20**, 092001 (2018).
- [24] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
- [25] D. Bacon, Operator quantum error-correcting subsystems for self-correcting quantum memories, *Phys. Rev. A* **73**, 012340 (2006).
- [26] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.122.080504> for implementation and data analysis details, which includes Refs. [27,28].
- [27] A. W. Cross, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, Open quantum assembly language, [arXiv:1707.03429](https://arxiv.org/abs/1707.03429).
- [28] R. Cleve, D. Leung, L. Liu, and C. Wang, Near-linear constructions of exact unitary 2-designs, *Quantum Inf. Comput.* **16**, 0721 (2016), [arXiv:1501.04592](https://arxiv.org/abs/1501.04592).
- [29] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
- [30] D. Gross, K. Audenaert, and J. Eisert, Evenly distributed unitaries: On the structure of unitary designs, *J. Math. Phys. (N.Y.)* **48**, 052104 (2007).
- [31] W. G. Brown and B. Eastin, Randomized benchmarking with restricted gate sets, *Phys. Rev. A* **97**, 062323 (2018), [arXiv:1801.04042](https://arxiv.org/abs/1801.04042).
- [32] A. K. Hashagen, S. T. Flammia, D. Gross, and J. J. Wallman, Real randomized benchmarking, *Quantum* **2**, 85 (2018).
- [33] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, Characterizing universal gate sets via dihedral benchmarking, *Phys. Rev. A* **92**, 060302 (2015), .
- [34] J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum *et al.*, Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking., *J. Phys. Condens. Matter* **27**, 154205 (2015).
- [35] M. A. Fogarty, M. Veldhorst, R. Harper, C. H. Yang, S. D. Bartlett, S. T. Flammia, and A. S. Dzurak, Nonexponential fidelity decay in randomized benchmarking with low-frequency noise, *Phys. Rev. A* **92**, 022326 (2015), [arXiv:1502.05119v2](https://arxiv.org/abs/1502.05119v2).
- [36] R. Harper, I. Hincks, C. Ferrie, S. T. Flammia, and J. J. Wallman, Statistical analysis of randomized benchmarking, [arXiv:1901.00535](https://arxiv.org/abs/1901.00535).

- [37] D. Nigg, M. Muller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Delgado, and R. Blatt, Quantum computations on a topologically encoded qubit, *Science* **345**, 302 (2014), [arXiv:1403.5426](#).
- [38] A. Robertson, C. Granade, S. D. Bartlett, and S. T. Flammia, Tailored Codes for Small Quantum Memories, *Phys. Rev. Applied* **8**, 064004 (2017), [arXiv:1703.08179](#).
- [39] E. Huang, A. C. Doherty, and S. Flammia, Performance of quantum error correction with coherent errors, [arXiv:1805.08227](#) [Phys. Rev. A (to be published)].
- [40] S. Beale, J. Wallman, M. Gutiérrez, K.R. Brown, and R. Laflamme, Quantum Error Correction Decoheres Noise, *Phys. Rev. Lett.* **121**, 190501 (2018), [arXiv:1805.08802](#).