Robust Decorrelation of Errors in Quantum Gates by Random Gate Synthesis

Anthony Polloreno*
Rigetti Computing, Berkeley, CA

Kevin C. Young
Sandia National Laboratories, Livermore, CA

(Dated: December 7, 2018)

Thresholds for fault-tolerant quantum computation are often calculated assuming a noise model in which errors are uncorrelated. While convenient for simulation, these error models are often unphysical. Work by Preskill and others has shown that arbitrarily long computations may be performed even in the presence of spatial and temporal correlation, provided the correlation is sufficiently weak and decays sufficiently quickly, but at the cost of a significantly lower threshold. The success of algebraic decorrelation methods, such as dynamical decoupling, demonstrate that quantum control techniques are capable of reducing noise correlations. We propose to introduce similar methods at the physical gate synthesis level to effect the decorrelation of errors in quantum circuits, thereby increasing the threshold for fault-tolerant computation in such systems. We discuss and show numerically that these controls can be made robust to drift, and experimentally demonstrate our approach on a superconducting qubit.

I. INTRODUCTION

Steady progress has been made in the theory of quantum error correction, proving higher thresholds for increasingly general models of noise [1–6]. These results show that quantum computation is feasible, however recent NISQ [7] devices have noise that is not only often above known thresholds, but that also violates fundamental assumptions made by the models used in these results[8–10], such as independence of errors[11] and Markovianity [12].

When noise can be accurately modeled as Markovian, existing work often makes approximations to make problems tractable. For example, Pauli channels are often used to model systems due to their classical simulability[13], even in the absence of physical motivation[14–21]. One approach, then, would be to use these thresholds to give a loose lower bound for thresholds in other systems[20]. Such an approach will produce overly-pessimistic bounds that may not be reflective of a system's actual performance.

When noise is non-Markovian, the situation is much worse. Not only will the Markovian models used in threshold calculations be wrong, but assumptions that underpin characterization routines will be violated. Randomized benchmarking and tomography will report incorrect answers without any syndrome in the case of non-Markovian noise [22] and gate-set tomography will report that a system fails to be Markovian, but without saying in what way.

An early proposal for fixing these problems is randomized decoupling[23, 24] – by introducing classical randomness into quantum computations, both coherent and non-Markovian noise can be transformed to incoherent, independent, and Markovian noise.

One particular way of realizing randomized decoupling is Frame Randomization[25, 26]. Frame Randomization can be used to average over noise by twirling error over an appropriately selected unitary 2-design, by inserting an additional precompilation step in the execution of quantum circuits.[27] This has been demonstrated in existing systems[26], but can require intensive use of classical processors and waveform memory.

Another way of performing randomized deocupling is through mixing unitaries. In [6], Campbell presents an algorithm that, given an oracle for generating unitaries close to a target, produces controls that are quadratically better in diamond norm than any of the individual controls generated by the oracle. In Campbell's work, he discusses such an oracle could be implemented via unitary synthesis approximations algorithms[28], and considers estimating and reducing resources costs that might arise during circuit compilation. In this paper we will we explore a different but related problem: synthesizing physical controls and making them robust to drift. Using optimal control routines as our oracle, we propose to inject additional decorrelating randomness into the system during physical gate synthesis through the use of *robustly balanced controls* (RBCs).

II. ROBUSTLY BALANCED CONTROLS

Consider the problem of trying to approximate a target unitary operation, U_T. Define a control solution $c_i(t)$ by

$$\mathcal{T} e^{-i \int c_i(t) H(t,\vec{\delta})} = U_i \approx U_T \tag{1}$$

where $\vec{\delta}$ are meant to be *quasi-static*[29] parameters of a Hamiltonian, $H(t, \vec{\delta})$. A family of control solutions is called a *balanced control* if there is probability distribu-

* Email: anthony@rigetti.com

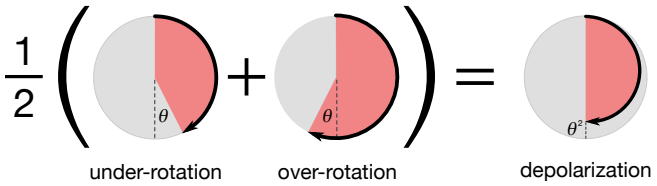

FIG. 1. An example of a balanced control solution. Using optimal control, two implementations of a Z_π gate are designed to have equal and opposite sensitivity to errors (if one implementation over-rotates by angle θ, then the other *under*-rotates by θ). Each time the gate is used, one of these implementations is chosen at random. The resulting quantum channel is equivalent to a perfect implementation of the gate followed by dephasing of $\mathcal{O}\left(\theta^2\right)$.

tion $\vec{\omega}$ that satisfies the following for some small α,

$$\sum_{i=1}^{N} \omega_j U_j \rho U_j^\dagger = DPN[\alpha]\left(U_T \rho U_T^\dagger\right) \tag{2}$$

where $\vec{\omega}$ must be a proper probability distribution and $DPN[\alpha](\rho)$ is a depolarizing noise channel, with strength α. A depolarizing channel is defined as:

$$DPN[\alpha](\rho) \to (1-\alpha)\rho + \alpha \sum p_i \sigma_i \rho \sigma_i \tag{3}$$

with p_i summing to one.

This is equivalent to the condition given by Campbell in [6], that there exist a weight such that

$$\sum \omega_j H_j(t, \vec{\delta})|_{\delta=\vec{0}} = 0 \tag{4}$$

however we will generalize this condition to define an RBC. A balanced control is said to be robust to order ℓ if a generalized form of Equation 4 holds, namely if for all $1 \leq j \leq \ell$:

$$\begin{aligned} \sum_k \omega_k (\sum_{n=0}^{j} D_k^n)^j &= 0 \\ D_k^n = \frac{1}{n!} \frac{\partial^n}{\partial\delta_{i_1} ... \partial\delta i_n} H_k(t, \vec{\delta})|_{\delta=\vec{0}} & \end{aligned} \tag{5}$$

Where the first sum is meant to result from Taylor expansion, and therefore will require matrix reshaping. These conditions imply that an ℓ^{th} order RBC (ℓRBC) is insensitive to the ℓ^{th} order in drift in $\vec{\delta}$. To see this, write the error on each gate in the RBC as $U_j(\vec{\delta}) = exp(-i(\tilde{H}_j(\vec{0}) + \frac{\partial}{\partial\delta_i}\tilde{H}_j(d\delta_i) + \frac{1}{2}\frac{\partial^2}{\partial\delta_i\partial\delta_k}\tilde{H}_j(d\delta_i d\delta_k) + ...)U_T$, where $\tilde{H}(\delta)$ generates the error $U_j(\vec{\delta})U_T^\dagger$ over variations in $\vec{\delta}$. Then, by Taylor expanding $U_j(\vec{\delta})U_T^\dagger$ one finds that the the first ℓ derivatives of an ℓRBC will be zero.

In all of the above, we will be interested in the case where $\alpha \approx \epsilon^2$, with ϵ being diamond distance from each

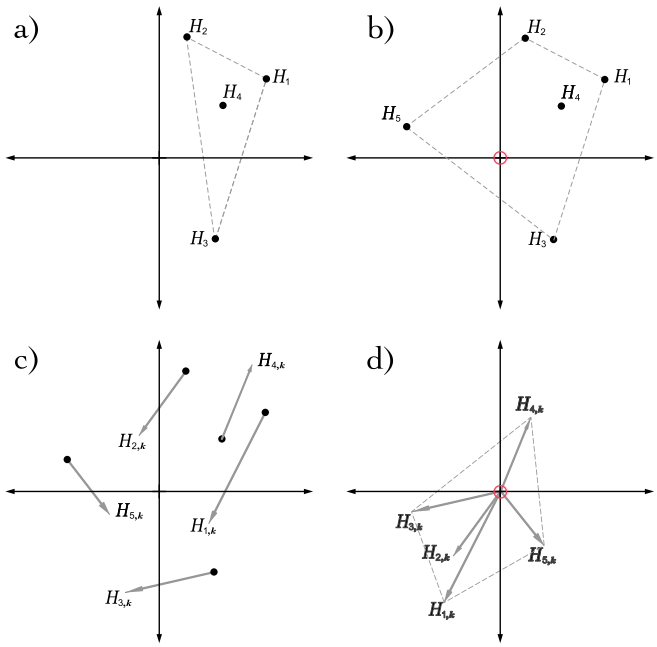

FIG. 2. A target unitary gate can be implemented a number of ways, each with a different effective Hamiltonian error. These error Hamiltonians lie in a vector space. a) Four effective Hamiltonians. The origin is not contained in their convex hull, so there are no balanced control solutions. b) The origin is contained in the covex hull after adding an additional control solution. Because there are more than $n+1$ implementations, there exist an infinite number of balanced control solutions. c) The error Hamiltonians shown with their derivative with respect to a control parameter. As this parameter drifts, a 0^{th}-order balanced control solution may drift, leading to a first-order error. d) The derivatives also lie in a vector space. If the origin lies in their convex hull, then it may be possible to construct a 1^{st}-order robust balanced control solution.

member of the RBC to U_T. Following from Lemma 2 in [6], if $0 \in \text{Conv}[\{\tilde{H}_i(t, \vec{\delta})\}]$, where Conv is the convex hull of its arguments then we know that $\vec{\omega}$ exists that satisfies Equation 2. Then, following from the mixing lemma proven by both Campbell and Hastings[28], we can bound the the diamond distance from a balanced control to U_T by ϵ^2. Moreover, by following a generalized version of the algorithm Campbell presents in [6], we can find families of controls that are ℓRBCs, for any ℓ. In this setting, we can relax Equation 5, to be

$$\sum_k \omega_k D_k^n = 0 \tag{6}$$

In this context, since we are assuming that the derivatives are all of norm ϵ, the above condition guarantees that the error unitary be no worse than $\sum^n \epsilon^2 \delta^n$ near the origin – all orders can be quadratically supressed.

III. A SIMPLE EXAMPLE

As a trivial example, consider a single-qubit rotation-angle error, such as from stochastic laser amplitude fluctuations. An RBC may consist of an X_π pulse, as well as an $X_{-\pi}$ pulse (i.e., a clockwise and counter clockwise rotation of the qubit). In the case of excess amplitude, the X_π pulse will result in an over-rotation error, while the $X_{-\pi}$ pulse will result in an *under*-rotation error. When it comes time to perform the target gate in a quantum circuit, one member of the RBC is chosen uniformly at random. This has the effect of decreasing the norm of the noise channel and decorrelating the over-rotation error (Figure 1). In this simple example, we can analytically find a solution to Equation 2. Specifically, by choosing weights $\omega_j = \frac{1}{2}$, we see:

$$\begin{aligned} &\frac{1}{2}(X^*_{\pi+\epsilon} \otimes X_{\pi+\epsilon} + X^*_{-(\pi+\epsilon)} \otimes X_{-(\pi+\epsilon)})) \\ &= (\sin^2 \frac{\pi+\epsilon}{2} I \otimes I + \cos^2 \frac{\pi+\epsilon}{2} X \otimes X) X \otimes X \\ &\approx DPN[\epsilon^2] X \otimes X \end{aligned} \tag{7}$$

Therefore, for a rotational error of angle $\epsilon > 0$, we see that X_π and $X_{-\pi}$ form a 0RBC, with $\alpha \approx \epsilon^2$.

If we consider the amplitudes to not only be miscalibrated, but also non-Markovian (i.e. such that ϵ is drifting), we find the following channel, for any drift δ:

$$DPN[(\epsilon+\delta)^2] X \otimes X \tag{8}$$

Thus, the noise is always depolarizing, for every $\ell > 0$, this family of controls is an ℓRBC, and by injecting additional randomness the error rate is supressed quadratically.

IV. OPTIMAL CONTROL PROBLEMS

A. Random Gate Synthesis

In this paper, we consider using optimal control to generate RBCs. One could use any of the many available quantum optimal control techniques [30–32], and for our numerics we chose to use the GRAPE algorithm. First described in [30], the GRAPE (GRadient Ascent Pulse Engineering) algorithm is a technique for finding piecewise constant control sequences that approximate a desired unitary, U_T, given a Hamiltonian with controlled and uncontrolled terms. Defining the uncontrolled Hamiltonian as H_0, the control Hamiltonians as $H_{i\neq 0}$, and the *control matrix* c_{ij} as containing the control amplitude associated with the i^{th} time step and the j^{th} Hamiltonian, we can write an the unitary for any timestep of evolution as:

$$U_i = \exp\{-i\Delta t(H_0 + \sum_{j=1}^{n} c_{ij} H_j\} \tag{9}$$

To measure the simularity of our approxiate unitary $U = \prod_i U_{i=1}^n$ to our target unitary U_T, we can define a cost function $J(U) = Tr\{U_T^\dagger U\}$.

To optimize this cost function we can perform the following standard update loop for some threshold value $\varepsilon > 0$ and step size $\delta > 0$:

Gradient Ascent
while $J(U) < (1-\varepsilon)$ **do**
  $c_{ij} \to c_{ij} + \delta \frac{\partial J(U)}{\partial c_{ij}}$
  for $1 \leq i \leq n$ **do**
    $U_i \to \exp\{-i\Delta t(H_0 + \sum_{i=0}^n c_{ij} H_j)\}$
  end for
  $U \to \prod_{i=1}^n U_i$
end while

In general these gradients can be computed by propagating partial derivatives of the cost function with respect to control parameters through each timestep of the via the chain rule. However, in [30] Khaneja et al. derive an update formula that is correct to first order, and efficient to compute. In particular one can show that:

$$\begin{aligned} \frac{\partial J(U)}{\partial u_{ij}} = -2Re\{&\left\langle U_{j+1}^\dagger ... U_N^\dagger U_T | i\Delta t H_j U_j ... U_1 \right\rangle \\ &\left\langle U_j ... U_1 | U_{j+1}^\dagger ... U_N^\dagger U_T \right\rangle\} + \mathcal{O}(\Delta t^2) \end{aligned} \tag{10}$$

Because we our trying to generate high-order RBCs, we want to generate controls that perform well even in the presence of drift. To do this, we modified the gradient in GRAPE to instead be:

$$\frac{\partial \tilde{J}(U)}{\partial u_{ij}} = \int p(\vec{\delta}) \frac{\partial J(U(\vec{\delta}))}{\partial u_{ij}} d\vec{\delta} \tag{11}$$

with $p(\vec{\delta})$ Gaussian distributed, as has been done in previous works [33] to ensure that the optimal control results are robust over a wide range of errors. To make this averaging tractable, we approximate this integral using Gaussian quadrature, approximating the cost function as a low order polynomial[34] (in particular degree 5). Note that despite updating the gradient, we did not change the cost function. The termination condition for the gradient ascent loop is still just that the controls perform well in the absence of drift. This is not a fundamental requirement of the routine, but rather was done to make finding solutions with $J(U) < 1-\epsilon$ less computationally taxing.

B. RBC Approximation

After using GRAPE or another optimal control routine to synthesize a collection of controls, we must find the weights ω_j such that the collection of controls form an ℓRBC as described in Equation 5. Previous authors have considered minimizing the diamond distance to the nearest Pauli or Clifford Channel [35], and while this gives a good theoretical framework it does not constrain the resulting channel to be decomposable into a given family of controls. The following routine, however, will produce a channel defined explicitly in terms of given family of controls. To do this, for each control $U_i(\vec{\delta})$ we find the unitary error channel $\mathcal{E}_i(\vec{\delta})$ such that $\mathcal{E}_i(\vec{\delta})U_i(\vec{\delta}) = U_T$, where U_T is the target gate. By taking the logarithm of these error maps, we then get a family of Hamiltonians $\tilde{H}_j(\delta)$, and consider the following quadratic program whose solutions satisfy the robustness conditions given in Equation 5:

$$\begin{aligned} &\textbf{minimize}: ||D_\ell^T \omega|| \\ &\omega_j \geq 0, |\omega|_1 = 1 \\ &\textbf{subject to: } \forall i < \ell, \sum \omega_j D_j^i = 0 \end{aligned} \tag{12}$$

If the control family is too small a solution won't exist, in particular as in [6] we need $0 \in \text{Conv}[D_j^i]$. This can be accomplished by modifying the oracle in Campbell's algorithm to produce robust controls, in particular by updating the GRAPE cost function to include derivative information, one could produce controls with derivatives of order ϵ. In our case, by using the Gaussian weighted cost function, we ensure that the controls generated should have vanishing first derivative.

To make sure the routine remains practical, we would like to regularize our objective function to enforce sparsity. In this case, lasso regularization [36] is insufficient as we already constrain the one norm of the vector we optimize over to be one. However, the problem of enforcing sparsity in such situations has been considered in [37] and can be expressed via another convex program that extends Equation 12:

$$\begin{aligned} &\textbf{minimize}\{ \underset{\omega_j \geq 0, |\omega|_1 = 1, t \geq 0}{\textbf{minimize}}: ||D_\ell^T \omega|| + t \\ &\textbf{subject to: } \omega_n > \frac{\lambda}{t} \\ &\forall i < \ell, \sum \omega_j D_j^i = 0\} \end{aligned} \tag{13}$$

where λ is a hyperparameter to be optimized over. In practice, for $\lambda = 1E-4$, we have found that this can lead to 1RBCs with fewer than ten members.

V. NUMERICAL RESULTS

In the following two subsections, we present numerical results of our routine, first for a one qubit example, and then for a two-qubit example. The code and RBCs generated for both examples are available online at [38].

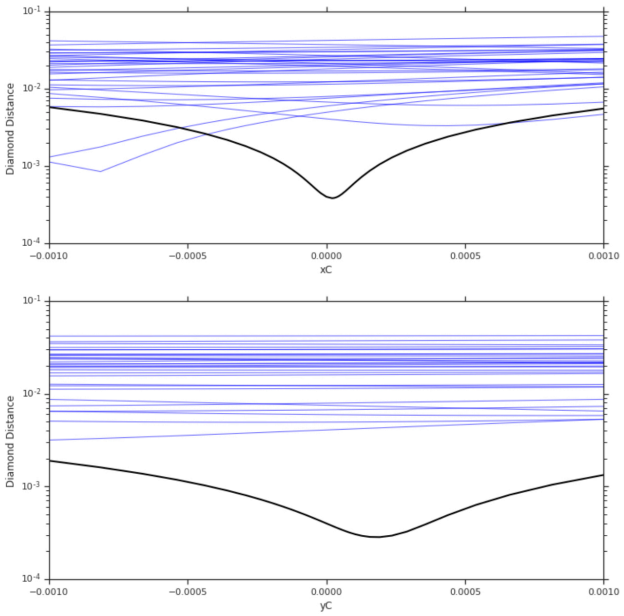

FIG. 3. Placeholder image until we figure out what to plot. This currently shows that for one of the control the diamond norm decreased by over an order of magnitude. The labels need to be bigger.

A. 1Q Gates

For the one-qubit case, we generate a 1RBCs for $RX(\frac{\pi}{2})$. Our control Hamiltonian is given as:

$$H(\delta, \epsilon, t) = \epsilon\sigma_z + (1+\delta)(c_x(t)\sigma_x + c_y(t)\sigma_y) \tag{14}$$

where we take when generating the members of the RBC. We assume that the errors on σ_x and σ_y are perfecly correlated, as mentioned in Section IV. When generating the controls, we chose an total evolution time of $T = FILLIN$, a number of steps $N = FILLIN$, a Gaussian quadrature degree of $d = FILLIN$, and a threshold infidelity of $1E-3$, with $\epsilon, \delta \sim \mathcal{N}(0, .001)$. The results can be seen in V A. In particular, over a wide region about the origin, the diamond norm can be seen to have been reduced by over an order of magnitude from the members of the RBC, and the 1RBC is performs moderately worse at the origin, while outperforming the 0RBC for larger detunings.

B. 2Q Gates

For a two-qubit example, we chose $iSWAP$. Our control Hamiltonian is given as:

$$H(\vec{\epsilon},\vec{\delta},\eta,t) = \sum_{j=1}^{2}(\epsilon_j\sigma_z^j + (1+\delta_j)(c_x^j x(t)\sigma_x^j + c_y^j(t)\sigma_y^j)) + k(XX+YY) \tag{15}$$

We again consider $\epsilon_j, \delta_j \sim .001$, a threshold infidelity of $1E-3$, and a Gaussian quadrature degree of $d = FILLIN$. In the two qubit case, however, the evolution time was constrained by the interaction Hamiltonian. In particular, we required a time of FILLIN so that GRAPE could find controls that performed a high fidelity iSWAP. Due to the naïvety of GRAPE, however, all the controls generated have nearly the same sensitivity to ϵ_j. If we want to generate an ℓRBC for $\ell > 0$, we need to ensure that the derivatives of the controls produced by our routine are both small and various enough that they can be combined in a way that averages to zero. To this end, we manually produced a collection of so-called bang-bang decoupling sequences. An example of such a sequence is show in FILL IN A FIGURE HERE. These controls all are constructed in such a way that one of the two qubits is flipped for some duration of the gate, during which a fraction of accrued RZ errors from frequency detuning errors on either qubit can be echoed away. When the convex solver is given these controls, it finds an optimal weighting that performs comparably with the one qubit example.

VI. EXPERIMENTAL RESULTS

Here we present experimental results from implementing our routine on a fixed-frequency superconducting transmon qubit. In particular we used qubit 8 on the Rigetti 19Q-Acorn chip, whose characterization can be found in [39]. To implement an RBC on this qubit, four incorrectly calibrated approximately Gaussian pulses were produced by scaling the pulseshape for a calibrated 50μs $RX(\frac{\pi}{2})$ pulse by 106.4%, 103.9%, 93.7% and 91.2%.

Using a similar routine to Equation 12, we then generated the optimal weightings ω. Instead of performing the convex optimization routine, we used the COBYLA minimizer in `scipy.optimize` to perform constrained minimization of the off-diagonal elements of the produced channel. This is a form of Sequential Quadratic Programming[40], and is equivalent to the convex optimization problem given for producing a 0RBC. To benchmark the quality of the new balanced channel, we then performed six randomized benchmarking experiments: one for each over- and under-calibrated pulse, one for the calibrated pulse, and one for the balanced channel.

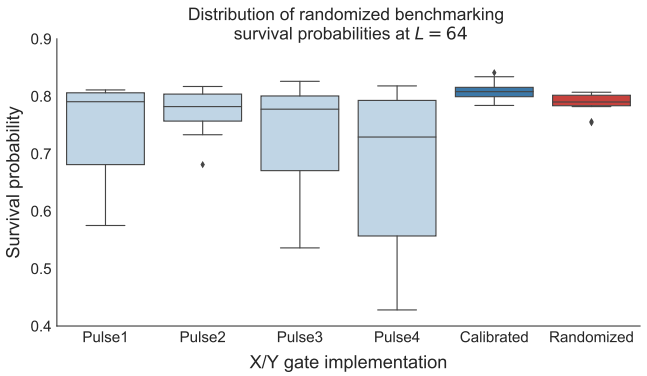

FIG. 4. Randomized benchmarking experiments ran using different pulse definitions. The four plots on the left are from the incorrectly calibrated pulse, while the top right is the calibrated pulse, and the bottom right is the BCS.

We used $N = 1000$ shots per experiment and $K = 10$ sequences per sequence length, for sequence lengths up to $L = 64$[41]. In each case, our Clifford operations were decomposed into RX($\frac{\pi}{2}$)) and RY($\frac{\pi}{2}$)) pulses. In our implementation, these gates are implemented using the same pulse envelope definitions and control electronics, phase shifted by $\frac{\pi}{2}$ radians, and are therefore subject to identical miscalibration errors. The results are shown in Figure 4 for sequence lengths $L = 64$.

In this experiment, fitting the randomized benchmarking data reports one-qubit gate fidelities of 99.3% for the calibrated pulse, 98.9% for Pulse1, 99.1% for Pulse2, 98.9% for Pulse3, 98.5% for Pulse4, and 99.2% for the Randomized pulse. In the cases of each miscalibrated pulse, the error bars extend significantly, which is consistent with the results shown in [29] that for particular non-Markovian error models, noise will manifest as gamma distributed points for each sequence length. On the other hand, Markovian noise, such as depolarizing noise, will result in Gaussian distributed fidelity estimates for each randomized benchmarking sequence length. We see that the coherently miscalibrated controls have long tails, consistent with gamma distributed random variables, while the calibrated and randomized implementations both have much shorter tails, consistent with Gaussian distributed random variables.

VII. CONCLUSION AND FUTURE WORK

We have shown numerically that using RBCs can reduce non-Markovian, coherent error on a quantum channel by more than an order of magnitude in diamond norm, over a wide range of quasi-static values of noise. In addition, we have demonstrated that these approximate controls can be generated through optimal control (GRAPE), and that the minimization problem is tractable.

Future directions for this work include demonstrating

the routine experimentally on a two-qubit gate, moving the random gate selection from a precompilation step to runtime logic onboard the FPGA, investigating other optimization routines such as CRAB [31] and GOAT[32], and using more sophisticated benchmarking routines such as GST[9] to quantitatively investigate the performance of our method.

Another interesting area of research would be using model-free approaches. The numerical work in the paper assumes access to a model of the system, however an experimentalist may not have a model readily available to describe the system, e.g. in the presence of unknown on-chip crosstalk, or an uncalibrated transfer function of the system. Moreover, even if a model is available, it might be computationally inconvenient to simulate, i.e. for more than a few qubits.

In these situations, one approach would be to use *in-situ* optimal control techniques [42–44] to generate candidate controls, and then use an optimizer like Nealder-Mead to perform the minimization. While performing a complete optimization in this way would still be slow, requiring full process tomography, one could intead optimize via partial tomography. By selecting pre– and post –rotations that correspond to measuring pauli moments of interest in the Hamiltonian, such as unwanted $Z \otimes Z$ crosstalk, one could perform optimization over fewer parameters.

VIII. ACKNOWLEDGEMENTS

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-NA0003525.

[1] D. Aharonov, A. Kitaev, and J. Preskill, Physical Review Letters **96** (2006), 10.1103/physrevlett.96.050504.
[2] N. P. Breuckmann, K. Duivenvoorden, D. Michels, and B. M. Terhal, (2016), arXiv:1609.00510.
[3] M. E. Beverland, (2016), 10.7907/z96m34sc.
[4] A. Kubica, M. E. Beverland, F. Brandão, J. Preskill, and K. M. Svore, Physical Review Letters **120** (2018), 10.1103/physrevlett.120.180501.
[5] C. Wang, J. Harrington, and J. Preskill, Annals of Physics **303**, 31 (2003).
[6] E. Campbell, Physical Review A **95** (2017), 10.1103/physreva.95.042306.
[7] J. Preskill, "Quantum computing in the nisq era and beyond," (2018), arXiv:1801.00862.
[8] J. Kelly, P. O'Malley, M. Neeley, H. Neven, and J. M. Martinis, "Physical qubit calibration on a directed acyclic graph," (2018), arXiv:1803.03226.
[9] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Nature Communications **8** (2017), 10.1038/ncomms14485.
[10] P. Klimov, J. Kelly, Z. Chen, M. Neeley, A. Megrant, B. Burkett, R. Barends, K. Arya, B. Chiaro, Y. Chen, A. Dunsworth, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, E. Lucero, J. Mutus, O. Naaman, C. Neill, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. White, S. Boixo, R. Babbush, V. Smelyanskiy, H. Neven, and J. Martinis, Physical Review Letters **121** (2018), 10.1103/physrevlett.121.090502.
[11] E. Knill, R. Laflamme, and W. H. Zurek, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **454**, 365 (1998).
[12] A. Y. Kitaev, Russian Mathematical Surveys **52**, 1191 (1997).
[13] D. Gottesman, (1998), arXiv:quant-ph/9807006.
[14] P. Aliferis and A. W. Cross, Physical Review Letters **98** (2007), 10.1103/physrevlett.98.220502.
[15] E. Knill, Nature **434**, 39 (2005).
[16] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, Physical Review A **83** (2011), 10.1103/physreva.83.020302.
[17] G. Duclos-Cianci and D. Poulin, Physical Review Letters **104** (2010), 10.1103/physrevlett.104.050504.
[18] J. R. Wootton and D. Loss, Physical Review Letters **109** (2012), 10.1103/phalifferis2008accuracyysrevlett.109.160503.
[19] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, Physical Review X **2** (2012), 10.1103/physrevx.2.021004.
[20] D. Puzzuoli, C. Granade, H. Haas, B. Criger, E. Magesan, and D. G. Cory, Physical Review A **89** (2014), 10.1103/physreva.89.022306.
[21] P. Aliferis, D. Gottesman, and J. Preskill, Quantum Information and Computation **8**, 0181 (2008).
[22] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Physical Review A **87** (2013), 10.1103/physreva.87.062119.
[23] L. Viola and E. Knill, Physical Review Letters **94** (2005), 10.1103/physrevlett.94.060502.
[24] L. Viola, in *Proceedings of the 44th IEEE Conference on Decision and Control* (IEEE).
[25] J. J. Wallman and J. Emerson, Physical Review A **94** (2016), 10.1103/physreva.94.052325.
[26] M. Ware, G. Ribeill, D. Riste, C. A. Ryan, B. Johnson, and M. P. da Silva, "Experimental demonstration of pauli-frame randomization on a superconducting qubit," (2018), arXiv:1803.01818.
[27] A. Roy and A. J. Scott, Designs, codes and cryptography **53**, 13 (2009).
[28] M. B. Hastings, "Turning gate synthesis errors into incoherent errors," (2016), arXiv:1612.01011.
[29] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, Physical Review A **93** (2016), 10.1103/physreva.93.022303.
[30] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of Magnetic Resonance **172**,

296 (2005).
[31] T. Caneva, T. Calarco, and S. Montangero, Physical Review A **84** (2011), 10.1103/physreva.84.022326.
[32] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, Physical Review Letters **120** (2018), 10.1103/physrevlett.120.150401.
[33] M. H. Goerz, E. J. Halperin, J. M. Aytac, C. P. Koch, and K. B. Whaley, Physical Review A **90** (2014), 10.1103/physreva.90.032329.
[34] M. Abramowiz and I. Stegun, National Bureau of Standards, US Government Printing Office, Washington, DC (1972).
[35] E. Magesan, D. Puzzuoli, C. E. Granade, and D. G. Cory, Physical Review A **87** (2013), 10.1103/physreva.87.012324.
[36] R. Tibshirani, Journal of the Royal Statistical Society. Series B (Methodological) , 267 (1996).
[37] M. Pilanci, L. E. Ghaoui, and V. Chandrasekaran, in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012) pp. 2420–2428.
[38] A. Polloreno, *Code and data used in this paper.*
[39] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. S. Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papageorge, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, C. A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, R. S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, B. R. Johnson, M. Reagor, M. P. da Silva, and C. Rigetti, "Unsupervised machine learning on a hybrid quantum computer," (2017), arXiv:1712.05771.
[40] S. Wright and J. Nocedal, Springer Science **35**, 7 (1999).
[41] E. Magesan, J. M. Gambetta, and J. Emerson, Physical Review Letters **106** (2011), 10.1103/physrevlett.106.180504.
[42] R.-B. Wu, B. Chu, D. H. Owens, and H. Rabitz, Physical Review A **97** (2018), 10.1103/physreva.97.042122.
[43] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. White, A. Cleland, and J. M. Martinis, Physical Review Letters **112** (2014), 10.1103/physrevlett.112.240504.
[44] C. Ferrie and O. Moussa, Physical Review A **91** (2015), 10.1103/physreva.91.052306.