Advantages of mixed unitary operators for quantum information processing

Anthony M. Polloreno*
Rigetti Computing, Berkeley, CA

Kevin C. Young
Sandia National Laboratories, Livermore, CA

(Dated: January 6, 2019)

Coherent errors in quantum operations are ubiquitous. Whether arising from spurious environmental couplings or errors in control fields, such errors can accumulate rapidly and degrade the performance of a quantum circuit significantly more than an average gate fidelity may indicate. As Hastings and Campbell have recently shown, randomly sampling an ensemble of implementations of a target gate yields an effective quantum channel that well-approximates the target, but with dramatically suppressed coherent error. Our results extend those of Hastings and Campbell to include robustness to drifting external control parameters. We implement these constructions using a superconducting qubit and will discuss randomized benchmarking results consistent with a marked reduction in coherent error.

I. INTRODUCTION

The past decade has seen a dramatic increase in the performance and scale of quantum information processors (QIPs). Gate fidelities are now routinely in the 99% to 99.99% range [1, 2], and dozens of individually-addressable qubits are becoming available on integrated devices. While these advances represent important steps forward on the path towards a computationally useful QIP, the quantum advantage milestone [3] has yet to be definitively reached. The limiting factor, of course, is errors in the quantum gate operations.

The impact of an error in a quantum gate depends strongly on both the magnitude and the nature of the errors. Systematic, or *coherent*, errors can arise from poorly calibrated controls or imperfect gate compilations that induce repeatable, undesired unitary errors on the state of a QIP. Errors of this type are correlated in time and add up coherently. They are are computationally expensive to model and it is difficult to place tight analytic bounds on circuit performance. Contrast this against random, or *stochastic*, errors, which result from high-frequency noise in the controls or the environment. Systems with stochastic errors can be modeled by defining a rate of various discrete errors in the system, such as a bit flip or phase flip. These errors are significantly easier to simulate on a classical computer, and their impact on quantum circuits is much easier to estimate.

Despite the relative ease of modeling stochastic errors, coherent errors are often much more likely to appear in QIPs. While these errors can often be reconstructed using various tomographic techniques, their impact is difficult to predict. The diamond distance can be used to bound the total variation distance (TVD) of a quantum circuit, but it is in general sensitive at first order to repeated application of gate with coherent errors. For long circuits, this can add up extremely quickly. Recent work by Campbell and Hastings[4–6], however, has shown that coherent noise can be strongly suppressed by probabilistically mixing several distinct implementations of the target quantum gates. The resulting effective quantum process has a diamond distance that grows only quadratically in the over/under rotation angle of the component gates.

In this article we discuss various applications of these mixed unitary controls, and show that the advantages of this approach can be made robust to drift in the target gates. We demonstrate that, depending on the objective, different numerical optimzations may be prefered. We present an experimental implementation of single-qubit mixed unitary controls on a superconducting qubit testbed at Rigetti Computing. Using randomized benchmarking, we are able to show a marked improvement in error rates, as well as a reduced variance in circuit outcome probabilities, indicating a reduction in the coherence of the error. We further provide an optimal control approach to the mixed unitary control design problem, and apply our methods in simulation where we construct single- and two-qubit mixed unitary controls which are robust to drift and uncertainty in the control parameters.

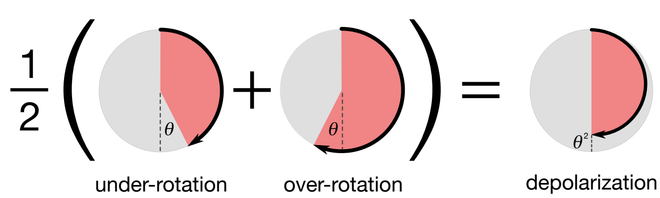

FIG. 1. An example of a mixed unitary process. Using optimal control, two implementations of a Z_π gate are designed to have equal and opposite sensitivity to errors (if one implementation over-rotates by angle θ, then the other *under*-rotates by θ). Each time the gate is used, one of these implementations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

* Email: anthony@rigetti.com

II. MATHEMATICAL PRELIMINARIES

Quantum gate operations are implemented by applying a sequence of classical control fields to some set of qubits. Fluctuations in the environment or imperfections in the controls can cause the state of the qubits to change in a way that is different from what was intended. But if the gates are fairly stable with time and context[7], then we can usually describe their action on the qubit state using *process matrices* – linear, Markovian maps on the state of some qubits. When working with process matrices, it is convenient to write the system density operator using a vectorized representation, and in this article, we'll make use of the generalized Bloch vector,

$$\vec{\rho} = \mathrm{Tr}\,(\rho\vec{\Sigma}), \tag{1}$$

where $\vec{\Sigma}$ is a vector of all 4^n n-qubit Pauli operators. The action of a gate is then given by the usual matrix multiplication:

$$\vec{\rho} \to \mathcal{G}\vec{\rho} = \mathcal{E}\tilde{\mathcal{G}}\vec{\rho}. \tag{2}$$

Here $\mathcal{G}$ is the target operation, $\tilde{\mathcal{G}}$ is the actual gate as implemented, and $\mathcal{E}$ is the effective error channel:

$$\mathcal{E} = \left(\begin{array}{c|c} 1 & \vec{0}^T \\ \hline \vec{m} & R \end{array}\right) \tag{3}$$

The top row of all trace-preserving (TP) maps is fixed to $\{1, 0, 0, 0, \cdots\}$. The rest of the first column, $\vec{m}$, describes any deviations from unitality, as could arise from amplitude damping. If the error channel is unitary, then the error is coherent, and the submatrix R is perfectly antisymmetric, corresponding to a rotation of the generalized Bloch vector. If R is diagonal, then the error channel is Pauli stochastic, with each entry corresponding to the probability that the associated Pauli error occurs in each application of the gate. That is, $\mathcal{E}_{ii}$ gives the probability that the following map is applied:

$$\vec{\rho}_k \to \sum_j \mathrm{Tr}\,(\sigma_j^\dagger \sigma_i \sigma_k \sigma_i^\dagger)\vec{\rho}_k \tag{4}$$

If R is symmetric but not diagonal, then the channel is still stochastic, but the random errors consist of correlated Pauli operators (such as $X + Y$). For a single qubit, this describes everything, but the situation can be slightly more complicated for more qubits.

For a collection of single qubit controls $\{U_i\}$ generated by $\{H_i\}$, the channel that results in general has elements:

$$\mathcal{M}_{jk} = \sum_i p_i \mathrm{Tr}\,(\sigma_j^\dagger U \sigma_k U^\dagger) \tag{5}$$

$$= \sum_i p_i \mathrm{Tr}\,(\sigma_j^\dagger \exp(i\frac{\theta_i}{2}H_i)\sigma_k \exp(-i\frac{\theta_i}{2}H_i)) \tag{6}$$

$$= \mathrm{Tr}\,(\sum_i p_i(\sigma_j^\dagger \sigma_k \cos\theta_i^2 + \sigma_j^\dagger H_i \sigma_k H_i \sin\theta_i^2 \tag{7}$$

$$+ i\sigma_j^\dagger H_i \sigma_k \sin\theta_i \cos\theta_i - i\sigma_j^\dagger \sigma_k H_i \sin\theta_i \cos\theta_i)) \tag{8}$$

The first term corresponds to the identity, and $\sin\theta^2$ is always positive, so we cannot hope to eliminate the second term, but $\sin\theta\cos\theta$ may be positive or negative, and so there is hope that we could possibly combine various implementations to eliminate this term. The result would be a purely stochastic channel.

The problem of minimizing $\sin\theta\cos\theta$ can equivalently be cast as trying to construct a family of vectors, $\sin\theta\cos\theta(\sigma_j^\dagger H_i \sigma_k - \sigma_j^\dagger \sigma_k H_i)$ whose convex hull contains the origin. This can be seen in Figure 2 a) and b).

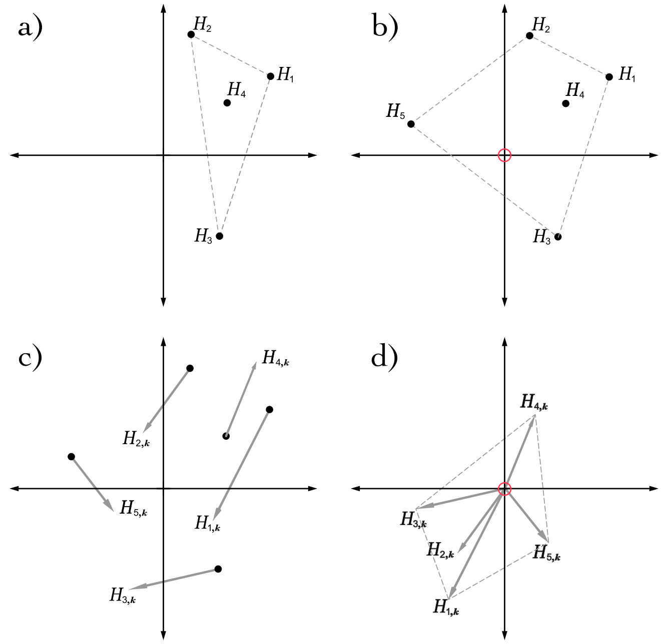

FIG. 2. A target unitary gate can be implemented a number of ways, each with a different effective Hamiltonian error. These error Hamiltonians lie in a vector space. a) Four effective Hamiltonians. The origin is not contained in their convex hull, so there are no 0MUPs. b) The origin is contained in the covex hull after adding an additional control solution. Because there are more than $n + 1$ implementations, there exist an infinite number of 0MUPs. c) The error Hamiltonians shown with their derivative with respect to a control parameter. As this parameter drifts, a 0MUP 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 1MUP.

In addition to the type of errors, we care about the size of an error. The *size* of an error in quantum gates may be quantified in a number of ways. Two of the most common metrics are the average gate fidelity, $\mathcal{F}$, and the diamond norm, $||\cdot||_\diamond$. These may be represented in terms of the error maps as

$$||I-\mathcal{E}||_\diamond = \sup_\rho ||(I\otimes I)(\rho) - (\mathcal{E}\otimes I)(\rho)||_1 \tag{9}$$

$$\mathcal{F}(\mathcal{E}) = \frac{\mathrm{Tr}\,\mathcal{E}+d}{d^2+d} \tag{10}$$

Which metric is relevant depends on the application, and can yield very different numbers. For instance, the diamond norm is generally linear in the over-rotation angle of a quantum operation, while the average gate infidelity (AGI), given by 1-$\mathcal{F}$, is generally quadratic in the over-rotation angle of a quantum operation. Similarly, these two metric perform very differently on stochastic quantum processes, that can be achieved by probabilistically mixing unitary operators.

A *mixed unitary process* (MUP) consists of a set of unitary channels, $\tilde{\mathcal{G}}_j$, and associated weights, $\sum_j \omega_j = 1$. The process matrix for a mixed unitary channel is then the weighted sum of the component channels, $\tilde{\mathcal{G}}_{\mathrm{M}} = \sum_j \omega_j \tilde{\mathcal{G}}_j$, and the associated error channel is simply the weighted sum of the associated error channels, $\mathcal{E}_{\mathrm{M}} = \sum_j \omega_j \mathcal{E}_j$. From this definition, we can compute the AGI of a MUP, by first computing the AGI of an convex combination of channels:

$$\begin{aligned}
\mathcal{F}(\alpha\mathcal{E}_1 + (1-\alpha)\mathcal{E}_2) &= \frac{\mathrm{Tr}\,(\alpha\mathcal{E}_1+(1-\alpha)\mathcal{E}_2)+d}{d^2+d} && (11)\\
&= \alpha\frac{\mathrm{Tr}\,(\mathcal{E}_1)+d}{d^2+d} + (1-\alpha)\frac{\mathrm{Tr}\,(\mathcal{E}_2)+d}{d^2+d} && (12)\\
&= \alpha\mathcal{F}(\mathcal{E}_1) + (1-\alpha)\mathcal{F}(\mathcal{E}_2) && (13)
\end{aligned}$$

where we have used the linearity of the trace.

We therefore see that the AGI of any MUP will be the convex sum of the consituent fidelities, with the same weighting. The diamond norm, however, is non-linear, and can in general be smaller for a MUP than any of the processes being mixed.

Campbell[4] considered the important problem of minimizing the diamond norm of the resulting error channel. Given a collection of component channels with error at most ϵ, he showed that if the Hamiltonians form a convex set containing the origin, then the diamond norm can be quadratically supressed. The diamond norm is a particularly appealing target because the it provides useful error metrics on quantum circuits. However, it is not the only optimization target that can be chosen. If the ultimate goal is to produce a channel whose effect can be Monte Carlo simulated, then it may be useful instead to construct a channel whose errors are Pauli stochastic. Using a constrained numerical optimization routine, such a channel could be produced by minimizing the off-diagonal elements of R in Equation 3. Indeed, this is the approach we consider in Section V. More generally, the are a number of options one may wish to consider when selecting weights to prepare a mixed unitary process.

III. CONSTRUCTING USEFUL UNITARY PROCESSES

A. Diamond Norm Minimization

There are often many possibly ways of implementing any given target quantum gate. Campbell and Hastings, for instance, consider gates compiled using the Solovey-Kitaev algorithm, for which many approximate gate compilations are possible.[4, 5] By selecting from these various implementations at random, they show that the resulting quantum channel can be made to have significantly reduced coherent error. As a simple example of how this occurs, consider a scenario in which we have a single-qubit and four possible implementations of a π-pulse about the σ_x axis. The error channels for these four implementations are themselves unitary rotations about the σ_x axis with rotation angles of $\{-2\epsilon, -\epsilon, \epsilon, 2\epsilon\}$. Such a situation could appear, for instance, if there were amplitude errors on the fields used to affect the gates, and if the control could be implemented by a rotation about the positive or negative σ_x axis.

Gate	H_{eff}	AGI	$\|\|\cdot\|\|_\diamond$
$U_{+2\epsilon}$	$2\epsilon\sigma_x$	$4\epsilon^2$	2ϵ
$U_{+\epsilon}$	$\epsilon\sigma_x$	ϵ^2	ϵ
$U_{-\epsilon}$	$-\epsilon\sigma_x$	ϵ^2	ϵ
$U_{-2\epsilon}$	$-2\epsilon\sigma_x$	$4\epsilon^2$	2ϵ

In this case, the MUP that minimizes the diamond norm is given by:

$$\frac{1}{2}U^*_{+\epsilon}\otimes U_{+\epsilon} + \frac{1}{2}U^*_{-\epsilon}\otimes U_{-\epsilon} \tag{14}$$

More generally, the problem remains: given a collection of channels, how can one efficiently compute a weighting that minimizes a particular metric?

As discussed in [4], a sufficient condition to minimize the diamond norm of a MUP with error generators $\{H_j\}$ to first order is:

$$\sum \omega_j H_j = 0 \tag{15}$$

In [4] Campbell constructs an algorithm that, given an oracle to approximate unitaries, finds a MUP with this property. Alternatively, one can use convex optimization to solve this problem. Consider the matrix whose rows are the vectorized Hamiltonians at our disposal, i.e. for m, $n\times n$ Hamiltonians:

$$H = \begin{pmatrix} H_{1_{11}} & H_{2_{21}} & \dots \\ \vdots & \ddots & \\ H_{m_{11}} & & H_{m_{nn}} \end{pmatrix} \tag{16}$$

If our weighting vector for our MUP is ω, we can rewrite this sum as a matrix product whose two-norm will be zero if and only if the sum is zero. That is, we can instead consider solving the following convex optimization problem:

$$\underset{\omega_j \geq 0, |\omega|_1=1}{\textbf{minimize}} : ||H^T \omega||_2 \tag{17}$$

where the minimization is over all valid probability distributions.

B. Robustly Mixed Unitary Processes

While mixed unitary processes offer significant improvements to gate performance, they fail to take into account the reality that most control electronics experience drift over time scales relevant to QIP performance. Because of this drift, the quality of the MUP will degrade. Thus, we would like to design MUPs that are *robust* to this drift. To enforce robustness, we can consider higher derivatives of the Hamiltonians in 15. Instead of just requiring that the 0^{th} derivative averages to zero, we should impose a similar condition on the derivatives of the Hamiltonians with respect to parameters that may drift:

$$D_j^n = \frac{1}{n!} \frac{\partial^n}{\partial \delta_{i_1} \ldots \partial \delta_{i_n}} H_j(\vec{\delta})|_{\delta=\vec{0}} \tag{18}$$

Then we say that a mixed unitary process is said to be robust to order ℓ (an ℓMUP) if for all $1 \leq j \leq \ell$:

$$\sum_j \omega_j (\sum_{n=0}^{k} D_j^n)^k = 0 \tag{19}$$

In particular, we see that a 0MUP satisfies Equation 15. More generally, these conditions imply that an ℓMUP is insensitive to the ℓ^{th} order in drift in $\vec{\delta}$. To see this, we can rewrite the error on each control in the MUP as:

$$\begin{aligned} \tilde{\mathcal{G}}_j(\vec{\delta}) &= \exp(-i(H_j(\vec{0}) + \frac{\partial}{\partial \delta_i} H_j(d\delta_i) \\ &+ \frac{1}{2} \frac{\partial^2}{\partial \delta_i \partial \delta_k} H_j(d\delta_i d\delta_k) + \ldots))\mathcal{G} \end{aligned} \tag{20}$$

By Taylor expanding Equation 20 in $\vec{\delta}$, one finds that the the first ℓ derivatives of an ℓMUP will be zero. Furthermore, if we are only interested in being first order insensitve to drift and can select controls such that $|D_j^n| \approx \epsilon$, we see that Equation 19 can be approximated as:

$$\sum_k \omega_j D_j^n = 0 \tag{21}$$

This condition guarantees that errors will be supressed quadratically for all derivatives up to order ℓ. A proof is included in the appendix that generalizes the Hastings-Campbell Mixing Lemma in [4]. Namely, Campbell shows that if $0 \in \mathrm{Conv}[\{H_i(\vec{\delta})\}]$, where Conv is the convex hull of its arguments, then an $\vec{\omega}$ exists that quadratically decreases the diamond norm. We prove that $0 \in \mathrm{Conv}[\{D_j^n(\vec{\delta})\}]$ implies there is an $\vec{\omega}$ exists that quadratically decreases the ℓ^{th}-order sensitivity of the diamond norm of an ℓMUP. This condition is illustrated in Figure 2 gives geometric intuition for the conditions required to produce an ℓMUP.

To generate robustly mixed unitary processes, we first define the vectorized derivative matrix D^ℓ in a similar way to 16:

$$D^\ell = \begin{pmatrix} D^\ell_{1_{11}} & D^\ell_{2_{21}} & \cdots \\ \vdots & \ddots & \\ D^\ell_{m_{11}} & & D^\ell_{m_{nn}} \end{pmatrix} \tag{22}$$

Using this, we can then solve the following convex optimization problem, generalizing Equation 17:

$$\begin{aligned} &\underset{\omega_j \geq 0, |\omega|_1=1}{\textbf{minimize}} : ||D^{\ell^T} \omega|| \\ &\textbf{subject to: } \forall n < \ell, \sum \omega_j D_j^n = 0 \end{aligned} \tag{23}$$

with D_k^n defined in Equation 19.

C. Hamiltonian Norm Regularization

While these particular convex optimization problem can be solved through a system of linear equations, casting it as a convex optimization problem allows us to regularize the cost function, and introduce additional constraints. In particular, while this minimization problem is sufficient for quadratically decreasing the diamond norm relative to the *worst* controls in the collection, it does not preferentially select the controls with the least error. That is to say, both $\{U_{+2\epsilon}, U_{-2\epsilon}\}$ and $\{U_\epsilon, U_{-\epsilon}\}$ from Section III A satisfy Equation 17. To encourage the inclusion of controls with smaller error, we may impose an ℓ_2 penalty to our cost function by rewriting it as:

$$\begin{aligned} &\underset{n \in [N]}{\textbf{minimize}} \{ \\ &\quad \underset{\omega_j \geq 0, |\omega|_1=1}{\textbf{minimize}} : ||D^{\ell^T} \omega|| + \eta \sum \omega_j ||D_j^0||_2 \\ &\quad \textbf{subject to: } \forall n < \ell, \sum \omega_j D_j^n = 0 \\ &\} \end{aligned} \tag{24}$$

with $\eta \geq 0$. By making η larger, we can require that the solver decreases the diamond norm of the constituent channel, while using only the best controls.

D. Sparsity Constraints

As a practical consideration, we would also like to regularize our objective function to enforce sparsity. Control electronics often have a limited amount of waveform memory, and thus it is important that MUPs have nontrivial probability support on a small number of controls. As an example of where this would be necessary is Figure 2. In Subfigure b, it's clear that H_4 is unecessary to contain the origin in the convex hull of the error generators. However, if we additionally want our controls to form a 1MUP, we see from Subfigure D that we need H_4 in our control set. Thus we would like the solver to be frugal in which controls it selects. In many machine learning situation, lasso regularization [8] can be used to enforce sparsity, however here it is insufficient as we already constrain the one norm of the vector we optimize over to be one. Conveniently, the problem of enforcing sparsity in such situations has been considered in [9] and can be expressed via another convex program that extends Equation 17:

$$\begin{aligned} &\underset{n\in[N]}{\textbf{minimize}}\{ \\ &\quad \underset{\substack{\omega_j\geq 0, |\omega|_1=1,\\ t\geq 0}}{\textbf{minimize}} : ||D^{\ell^T}\omega|| + t \\ &\quad \textbf{subject to: } \omega_n > \frac{\lambda}{t} \\ &\qquad\qquad \forall n < \ell, \sum \omega_j D_j^n = 0 \\ &\} \end{aligned} \tag{25}$$

where λ is a hyperparameter to be optimized over.

IV. NUMERICAL RESULTS

In the following numerical results, we explore using the methods in Section III to build MUPs. We consider the following model for a single tunable qubit:

$$H(\delta, \epsilon, t) = \epsilon\sigma_z + (1+\delta)(c_x(t)\sigma_x + c_y(t)\sigma_y) \tag{26}$$

We use the GRAPE algorithm as discussed in Section ?? with N=25 steps and total evolution time of π to generate 100 candidate controls, with a standard deviation of $\sigma = .001$ for the distribution in Equation ??. We assume that the errors on σ_x and σ_y are perfecly correlated, as is the case in systems that implement RZ rotations with phase shifts of the control signal. Solving the optimization problem defined in Section III B yields similar MUPs for $RX(\frac{\pi}{2})$ and $RY(\frac{\pi}{2})$, with the results for $RY(\frac{\pi}{2})$ shown in Figure 3. These results demonstrate several properties that make MUPs both useful and tractable.

First, naively generating a 0MUP results in nontrivial support on all the members of the control family. However, by rewriting the minimization to impose the sparsity constraint discussed in Section III B, we can generate

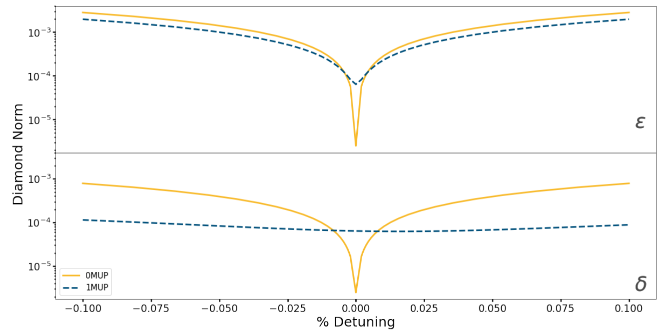

FIG. 3. Numerical results comparing a 0MUP to a 1MUP for a single tunable qubit, for $RY(\frac{\pi}{2})$. The results are qualitatively similar to those for $RX(\frac{\pi}{2})$. In this case the 0MUP outperforms both the 1MUP by two orders of magnitude, and the constituent controls by three orders of magnitude at the origin. However, varying over δ we see that the 1MUP outperforms the 0MUP by up to an order of magnitude when there is .1% drift in the qubit control amplitudes.

a 0MUP that uses just five of the controls. This shows that through adding constraints to our optimization routine, we can make the MUP practically useful. In both cases we impose the same ℓ_2 penalty as described in Section III B, so that the algorithm preferentially selects controls with smaller errors. Imposing this constraint allows us to trade off flatness at the origin for performance.

A. Two Qubit Performance Analysis

In our two-qubit example we consider the following model for two tunable qubits coupled by a resonant exchange interaction, similar to that in [15]:

$$\begin{aligned} H(\vec{\delta}, \vec{\epsilon}, t) = &\sum_{j=1}^{2}(\epsilon_j\sigma_z^j + (1+\delta_j)(c_x^j x(t)\sigma_x^j + c_y^j(t)\sigma_y^j)) \\ &+ \frac{1}{10}(XX + YY) \end{aligned} \tag{27}$$

In this example it was infeasible to use GRAPE to return non-trivial solutions. Instead we manually selected piecewise constant echoing sequences with 500 steps and total evolution time of $\frac{5\pi}{2}$. In particular, we considered $RX(\pi)$, $RX(-\pi)$, $RY(\pi)$ and $RY(-\pi)$ bang-bang sequences [16], consisting of all combinations of simultaneous π pulses activated at multiples of 8 steps from the beginning of the controls, and the same multiple of 8 steps prior to the end of the controls. To give the control family a variety of RF errors, we added on uniformly distributed errors to each π pulse, between $-.25\%$ and $.25\%$.

In this example, we find more modest improvements to performance, as shown in Figure 4. There are now four free parameters to optimize over, and the uncontrolled entangling interaction means that there is little room for

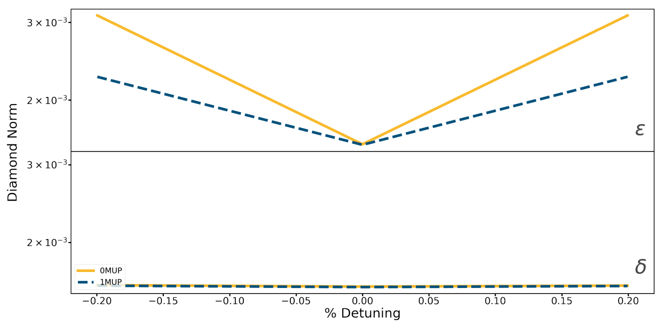

FIG. 4. Numerical results comparing a 0MUP to a 1MUP for a pair of tunable qubits, with a resonant exchange interaction. Shown with lower alpha values are example constituent controls. The 0MUP and 1MUP can be seen to outperform these controls by half of an order of magnitude at the origin. For all detuning values the 1MUP performs as well or better than the 0MUP. When there is .2% drift in the qubit frequency, the 1MUP outperforms members of the control famlies by almost an order of magnitude in diamond norm. Similarly, for .2% drift in the qubit control amplitude, we see that the 1MUP outperforms the the constituent controls by over half an order of magnitude.

variation in the controls. Nonetheless, using a MUP improves performance by half of an order of magnitude at the origin relative to the constituent controls, and up to an order of magnitude away from the origin. For all values of the drifting parameters we see that the 1MUP performs as well or better than 0MUP.

V. 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 [17]. To implement a MUP on this qubit, four incorrectly calibrated Gaussian pulses were produced by scaling the pulseshape amplitude for a calibrated 10 sample 50ns $RX(\frac{\pi}{2})$ pulse by 106.4%, 103.9%, 93.7% and 91.2%.

As discussed in Section III B, we chose here to minimize the off diagonal elements of the process matrix. To benchmark the quality of the MUP, we then performed six randomized benchmarking experiments[18]: one for each over– and under–calibrated pulse, one for the calibrated pulse, and one for the mixed process. We used 1000 shots per experiment, 10 sequences per sequence length, for sequence lengths of $2, 4, 8, 16, 32$ and 64. 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 5 for sequence lengths $L = 64$. Fitting to the randomized

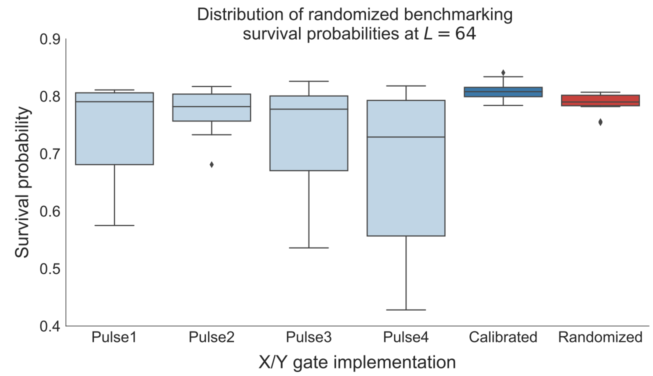

FIG. 5. Randomized benchmarking experiments ran using different pulse definitions. The four plots on the left are from the incorrectly calibrated pulse, while the top right is the calibrated pulse, and the bottom right is the MUP.

benchmarking decay curves, we find 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 MUP, demonstrating that it performs almost as well as the calibrated pulse, and better than the constituent pulses.

Additionally, by minimizing the off-diagonal elements of the process matrix, rather than balancing the Hamiltonian errors, we expect to produce a process with minimal coherent error in the absence of drift. To see that this is the case, we cite the results in [19]. For 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 in our RB experiment 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. Thus, our experiment demonstrates that not only is the performance of the MUP better than the constituent gates, it also has a signicantly less-coherent error channel.

VI. CONCLUSION AND FUTURE WORK

We have shown numerically that using MUPs can reduce 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 control electronics, investigating other optimization routines such as CRAB [11] and GOAT[12], and using more sophisticated benchmarking routines such as GST[20] 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. 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 [21–23] 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 require 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.

VII. 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] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Nature **508**, 500 (2014).

[2] C. Ballance, T. Harty, N. Linke, M. Sepiol, and D. Lucas, Physical Review Letters **117** (2016), 10.1103/physrevlett.117.060504.

[3] J. Preskill, "Quantum computing and the entanglement frontier," (2012), arXiv:1203.5813.

[4] E. Campbell, Physical Review A **95** (2017), 10.1103/physreva.95.042306.

[5] M. B. Hastings, "Turning gate synthesis errors into incoherent errors," (2016), arXiv:1612.01011.

[6] E. Campbell, "A random compiler for fast hamiltonian simulation," (2018), arXiv:1811.08017.

[7] K. Rudinger, T. Proctor, D. Langharst, M. Sarovar, K. Young, and R. Blume-Kohout, "Probing context-dependent errors in quantum processors," (2018), arXiv:1810.05651.

[8] R. Tibshirani, Journal of the Royal Statistical Society. Series B (Methodological) , 267 (1996).

[9] 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.

[10] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of Magnetic Resonance **172**, 296 (2005).

[11] T. Caneva, T. Calarco, and S. Montangero, Physical Review A **84** (2011), 10.1103/physreva.84.022326.

[12] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, Physical Review Letters **120** (2018), 10.1103/physrevlett.120.150401.

[13] 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.

[14] M. Abramowiz and I. Stegun, National Bureau of Standards, US Government Printing Office, Washington, DC (1972).

[15] D. C. McKay, S. Filipp, A. Mezzacapo, E. Magesan, J. M. Chow, and J. M. Gambetta, Physical Review Applied **6** (2016), 10.1103/physrevapplied.6.064007.

[16] L. Viola and S. Lloyd, Physical Review A **58**, 2733 (1998).

[17] 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.

[18] E. Magesan, J. M. Gambetta, and J. Emerson, Physical Review Letters **106** (2011), 10.1103/physrevlett.106.180504.

[19] H. Ball, T. M. Stace, S. T. Flammia, and M. J. Biercuk, Physical Review A **93** (2016), 10.1103/physreva.93.022303.

[20] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Nature Communications **8** (2017), 10.1038/ncomms14485.

[21] R.-B. Wu, B. Chu, D. H. Owens, and H. Rabitz, Physical Review A **97** (2018), 10.1103/physreva.97.042122.

[22] 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.

[23] C. Ferrie and O. Moussa, Physical Review A **91** (2015), 10.1103/physreva.91.052306.

VIII. APPENDIX

A. Robust Mixing Lemma

We begin by generalizing Lemma 2 from [4]. From A3 in [4], we know that $||e^{iM}| - (\mathbb{1} + iM)|| \leq \frac{1}{2}||M||^2$. Thus we can Taylor expand any mixture of unitaries, as:

$$\frac{d}{d\vec{\delta}}\sum_j \omega_j e^{iH_j} = \frac{d}{d\vec{\delta}}(\mathbb{1} + (\sum_j i\omega_j H_j) + \sum \omega_j \sum_{n=2}^{\infty} \frac{(iH_j)^n}{n!})) \tag{28}$$

By assumption, the derivatives of the first order term sum to zero, and so we see by A3 that

$$||\frac{d}{d\vec{\delta}}\sum_j \omega_j e^{iH_j}|| \leq \sum_j \omega_j \frac{c^2}{2} \leq \frac{c^2}{2} \tag{29}$$

Additionally,

$$\begin{aligned}
||\frac{d}{d\vec{\delta}}e^{iH_j}|| &= ||\frac{d}{d\vec{\delta}}(iH_j)e^{iH_j}|| && (30)\\
&\leq ||\frac{d}{d\vec{\delta}}(iH_j)|| + ||\frac{d}{d\vec{\delta}}(iH_j)H_j|| + ||\frac{d}{d\vec{\delta}}(iH_j)\sum_{n=2}^{\infty}\frac{1}{(n)!}(iH_j)^n|| && (31)\\
&\leq ||\frac{d}{d\vec{\delta}}(iH_j)|| + ||\frac{d}{d\vec{\delta}}(iH_j)|| \cdot ||H_j|| + ||\frac{d}{d\vec{\delta}}(iH_j)|| \cdot ||\sum_{n=2}^{\infty}\frac{1}{(n)!}(iH_j)^n|| && (32)\\
&\leq c + c^2 + \frac{1}{2}c^3 && (33)
\end{aligned}$$

And so we see that Lemma 2 generalizes. In all of the above $||\cdot||$ is the Schatten 0-norm, which we will use to upper bound the 1-norm (and hence diamond norm) in the following.

Next we generalize the Mixing Lemma from [4]. We must assume that the 1-norm is uniformly differentiable. In that case, we see that:

$$\begin{aligned}
&= \frac{d}{d\vec{\delta}}\sup|\mathcal{I}\otimes\mathcal{I} - \mathcal{E}(\vec{\delta})\otimes\mathcal{I}|_1 && (34)\\
&= \sup\frac{d}{d\vec{\delta}}|\mathcal{I}\otimes\mathcal{I} - \mathcal{E}(\vec{\delta})\otimes\mathcal{I}|_1 && (35)\\
&\leq \sup|\frac{d}{d\vec{\delta}}(\mathcal{I}\otimes\mathcal{I} - \mathcal{E}(\vec{\delta})\otimes\mathcal{I})|_1 && (36)\\
& && (37)
\end{aligned}$$

And so we see it is sufficient to upperbound:

$$|\frac{d}{d\vec{\delta}}(\mathcal{I}\otimes\mathcal{I} - \mathcal{E}(\vec{\delta})\otimes\mathcal{I})|_1 \tag{38}$$

From Equation 18 in [4]:

$$\begin{aligned}
(\mathcal{V}\circ\mathcal{E} - \mathbb{1})(X) &= \sum_j \omega_j(\tilde{\delta}_j X + X\tilde{\delta}_j^\dagger + \tilde{\delta}_j X\delta_j^\dagger) && (39)\\
\frac{d}{d\vec{\delta}}(\mathcal{V}\circ\mathcal{E} - \mathbb{1})(X) &= \sum_j \omega_j(\frac{d}{d\vec{\delta}}\tilde{\delta}_j X + X\frac{d}{d\vec{\delta}}\tilde{\delta}_j^\dagger + \frac{d}{d\vec{\delta}}\tilde{\delta}_j X\delta_j^\dagger + \tilde{\delta}_j X\frac{d}{d\vec{\delta}}\delta_j^\dagger) && (40)\\
& && (41)
\end{aligned}$$

Thus, by Hölder's inequality we find:

$$||\frac{d}{d\vec{\delta}}(\mathcal{V}\circ\mathcal{E} - \mathbb{1})(X)||_1 \leq ||\sum_j \omega_j\frac{d}{d\vec{\delta}}\tilde{\delta}_j|| + ||\sum_j \omega_j\frac{d}{d\vec{\delta}}\tilde{\delta}_j^\dagger|| + \sum_j \omega_j||\frac{d}{d\vec{\delta}}\tilde{\delta}_j|| \cdot ||\delta_j^\dagger|| + \sum_j \omega_j||\tilde{\delta}_j|| \cdot ||\frac{d}{d\vec{\delta}}\delta_j^\dagger|| \tag{42}$$

By assumption, all of these terms are neglible, and therefore we find:

$$||\frac{d}{d\vec{\delta}}(\mathcal{V} \circ \mathcal{E} - \mathbb{1})(X)||_1 \leq 2a^2 + 2b^2 \tag{43}$$