# Decorrelating Errors in Quantum Gates by Random Gate Synthesis

Kevin C. Young*

*Sandia National Laboratories, Livermore, CA*

Anthony Polloreno

*Rigetti Computing, Berkeley, CA*

(Dated: May 6, 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. Recent work by Preskill and others has shown that arbitrarily long computations may be performed even in the presence of spatial correlation, provided the correlation is sufficiently weak and decays sufficiently quickly with distance, 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 temporal noise correlations. We propose to introduce similar methods to effect the spatial decorrelation of errors in quantum circuits, thereby increasing the threshold for fault-tolerant computation in such systems.

## I. INTRODUCTION

Steady progress has been made in the theory of quantum error correction, proving higher thresholds for increasingly general models of noise [1]. These results show that quantum computation is feasible, however recent NISQ [2] devices have noise that is not only often above thresholds, but that also violates assumptions made by the models used in proofs.

As an example, many models assume that errors are local, however systems can have spatially correlated error on more than two qubits that arise from multi-qubit terms in the Hamiltonian, such as:

$$H(t) = \sum_n \sum_i \sum_{j=x,y,z} a_{\mathbf{j}}^{\mathbf{i}}(t)(\sigma_{j_1}^{i_1} \otimes \sigma_{j_2}^{i_2} \otimes ... \otimes \sigma_{j_n}^{i_n}) \quad (1)$$

where $a_{\mathbf{j}}^{\mathbf{i}}$ is a stochastic process. In a standard nearest-neighbors topology, like those found in super-conducting qubit architectures, these terms might arise from an effective model where high-frequency degrees of freedom are intergrated out to yield high-weight operators. These noise channels are one way in which the assumptions used in proofs of thresholds may be violated.

Another assumption that is instrumental in proving many threshold theorems is Markovianity [3]. There are many ways in which devices can behave in non-Markovian ways, and NISQ devices have already been shown to be subject to drift in controls, [4] and in general non-Markovian noise [5]. Other authors have tried to address these problems [6–8] by using circuit composition to try to remove correlated noise. [WHY IS WHAT WE'RE DOING BETTER?]

To remedy these problems, we take a different approach and propose to inject additional decorrelating randomness into the system through the use of *balanced control solutions* (BCSs). BCSs are families of control solutions that all approximate the same target gate, but with balanced errors for any given instance of the noise Hamiltonian. That is, for a target gate, $U_T$, we seek a family of control solutions, $c_i(t)$, each implementing an approximation $U_i$ to the target gate, such that the family of unitary approximations is *balanced*. A balanced family is one which satisfies, for some small $\alpha$,

$$\frac{1}{N} \sum_{i=1}^{N} \omega_i U_i \rho U_i^\dagger = DPN[\alpha]\left(U_T \rho U_T^\dagger\right) \quad (2)$$

Where $DPN[\alpha](\rho)$ is a *generalized* depolarizing noise channel with strength $\alpha$. (For the rest of this paper, we will refer to them just as depolarizing channels.) Such a channel is defined as:

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

with $p_i$ summing to one. This means that on average, the unitary approximations implement the target unitary followed by a small depolarizing channel. The task of constructing the BCSs will fall to optimal control.

This techique can be used to turn coherent error into incoherent error and correlated error into uncorrelated error. The former gives the benefit that the error incurred from incoherent errors compound linearly, while coherent errors compound quadratically. The latter gives the benefit of reducing non-Markovian effects. This is a particularly useful property of BCSs, as diagnosing non-Markovianity in a gateset is challenging, but essential to perform reliable quantum computation. Many routines exist that can assess the quality of a gateset, however in the presence of non-Markovianity most of them become unreliable.

For example, randomized benchmarking and tomography will report incorrect answers without any syndrome, [CITATION NEEDED], and gateset tomography will report that the gateset failed to be Markovian, but will fail to diagnose in what way it was non-Markovian. Because

* Corresponding author: kyoung@sandia.gov

BCSs change the error to a depolorizing channel, the correlations in the noise can be massively reduced, and the process can be made much closer to Markovian, at no real cost to the fidelity of the implemented gate.

## II.  GRAPE

Generating BCSs can be done in a variety of ways, using any quantum optimal control technique [9, 10] to find a family of controls. For simplicity in this paper we use the GRAPE algorithm to generate candidate pulseshapes to approximate the target gate. First described in [11], the GRAPE (GRadient Ascent Pulse Engineering) algorithm is a technique for finding piecewise constant control sequences that approximate a desired unitary, $U_T$. Defining our uncontrolled Hamiltonian as $H_0$, our control Hamiltonians as $H_{i \neq 0}$, and our *control matrix* $u_{ij}$ as containing control amplitude associated with the $i^{th}$ time step and the $j^{th}$ hamiltonian, we can write our approximate unitary at any timestep as

$$U_i = \exp\{-i\Delta t(H_0 + \sum_{j=1}^{n} u_{ij}H_{ij}\} \quad (4)$$

Then, to measure the simularity of our approximate unitary $U_n$, and our target unitary $U_T$, we can define a cost function $J(U) = Tr\{U_T^\dagger U_n\}$.

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** $J(U_n) < (1 - \varepsilon)$ **do**
    $u_{ij} \to u_{ij} + \delta \frac{\partial J(U)}{\partial u_{ij}}$
    **for** $1 \leq i \leq n$ **do**
      $U_i \to \exp\{-i\Delta t(H_0 + \sum_{i=0}^{n} u_{ij}H_i)\}$
    **end for**
    $U \to \prod_{1}^{n} U_i$
  **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 [11] Khaneja et al. derive a simple update formula that is correct to first order. In particular one can show that:

$$\frac{\partial J(U)}{\partial u_{ij}} = -2Re\{\left\langle U_{j+1}^\dagger...U_N^\dagger U_T | i\Delta t H_j U_j...U_1 \right\rangle \quad (5)$$

$$\left\langle U_j...U_1 | U_{j+1}^\dagger...U_N^\dagger U_T \right\rangle\} + \mathcal{O}(\Delta t^2) \quad (6)$$

In this paper we have modified the update step to use instead use approximately the following gradient:

$$\int p(\vec{\delta}) \frac{\partial J(U(\vec{\delta}))}{\partial u_{ij}} d\vec{\delta} \quad (7)$$

with $p(\vec{\delta})$ Gaussian distributed. This technique has been used in previous works such as [12] to ensure that the optimal control results are robust over a wide range of errors, and we approximate this integral approximately in this paper by using Gaussian quadrature. Doing this ensures that the family of controls produced by our routine perform moderately well over a range of stochastic errors.

## III.  OPTIMAL CONTROL PROBLEM

To explore the utility of BCSs, we consider the following Hamiltonian:

$$H(t) = H_0 + \sum_{i=1}^{n} (c_i(t) + \delta_i(t))H_i \quad (8)$$

for control Hamiltonians $H_{i \neq 0}$, free evolution Hamiltonian $H_0$ and time-varying random variables $\delta_i$ are random variables. Such a model might describe a superconducting qubit quantum processor where control amplitudes for the RF pulses vary over time. Correlations between different $\delta_i$ could arise if two of the control have the same noise source, e.g. $X$ and $Y$ gates in superconducting qubit architectures might use the same AWG and pulse envelope.

After using GRAPE to produce a collection of controls, we must find the weights $w_i$ such that the collection of controls form a BCS as described in (2). To do this, for each control $U_i$ we find the unitary error channel $\mathcal{E}_i$ such that $\mathcal{E}_i U_i = U_T$, where $U_T$ is the target gate. If we consider the Pauli-Liouville representation of this error channel, the diagonal terms are the *stochastic* terms that arise from classical uncertainty, while the off-diagonal terms may more generally arise from *coherent* operations. In particular, we see that we can write a convex sum over these channels as:

$$\frac{1}{N} \sum_{i=1}^{N} w_i \mathcal{E}^\dagger (U_T \rho U_T^\dagger) \mathcal{E} \quad (9)$$

Now, to approximate a depolarizing channel we define our optimal control problem to be the following, which minimizes the off-diagonal terms:

$$\underset{w_0,...,w_N}{\textbf{minimize}}\{\sum_{i \neq j}^{N} |\sigma_i \Lambda(\sigma_j)|^2\}$$

$$\textbf{where } \Lambda(\sigma_j) := \sum_{i=1}^{N} w_i \mathcal{E}_i^\dagger \sigma_j \mathcal{E}_i \quad (10)$$

$$\textbf{subject to } \sum_{i=1}^{N} w_i = 1$$

This can be solved with a constrained minimization algorithm, such as Sequential Least Squares Programming.[CITATION NEEDED]
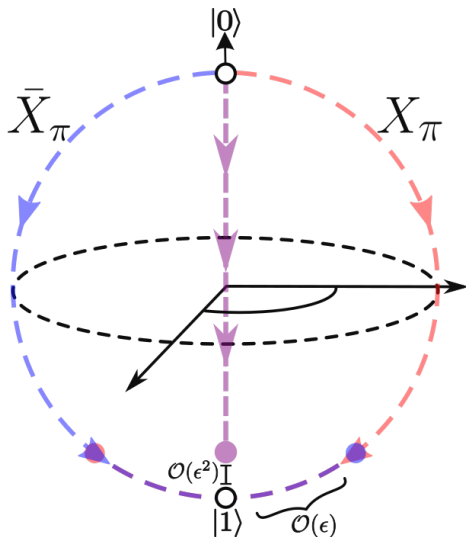
FIG. 1. An example of a BCS. The red path is a $\pi$ pulse over-rotating clockwise, while the blue path is a $\pi$ pulse over-rotating counter-clockwise. The purple path is the average of the two.

Previous authors have considered minimizing the diamond distance to the nearest Pauli or Clifford Channel [13], and while this gives a good theoretical framework, it requires the use of a convex solver, and in particular does not have the restriction that the optimal channel be decomposable into a given family of controls. Therefore we can view our approach as a regularization of this method, where we have imposed the condition that our final channel must arise from a BCS.[IS THIS ACTUALLY A REGULARIZATION? DOES IT IMPOSE SOME KIND OF SPARSITY] In doing so, we move from using a convex optimization routine to compute the diamond norm to needing to computing a sum. In the next section we give a simple example, followed by numerical results for one qubit and two qubits gates in sections V and VI.

## IV. A SIMPLE EXAMPLE

As a somewhat trivial example, consider a single-qubit rotation-angle error, such as result from stochastic laser amplitude fluctuations. A BCS may consist of an $X_\pi$ pulse, as well as an $\bar{X}_\pi$ pulse (i.e., a clockwise and counter clockwise rotation of the qubit). In the case of excess amplitude, the $X_\pi$ pulse will result in an over-rotation error, while the $\bar{X}_\pi$ pulse results in an *under*-rotation error. When it comes time to perform the target gate in a quantum circuit, one member of the BCS 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 solve the minimization problem given by equation 10 analytically. In particular, if we choose the weights in 2 such that $\omega_i = 1$, and we choose to represent ou gates in the vectorized superoperator representation, then:

$$\frac{1}{2}(X_{\pi+\epsilon} + \bar{X}_{\pi+\epsilon})$$
$$= (\sin^2 \frac{\pi+\epsilon}{2}I + \cos^2 \frac{\pi+\epsilon}{2}X)X \qquad (11)$$
$$\approx ((1-\epsilon^2)I + \epsilon^2 X)X$$

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

## V. 1Q GATES

In this section, we present numerical results on generating one-qubit gates that together with $RZ(\theta)$ rotations are universal for one-qubit computation. Our control Hamiltonian is given as:

$$H = \epsilon\sigma_z + (1+\delta)(c_x(t)\sigma_x + c_y(t)\sigma_y) \qquad (12)$$

$\epsilon$ is a normally distributed random variable, centered at 0 with standard deviation .001, corresponding to a time varying Stark shift on the qubit. $\delta$ is also a normally distributed random variable with the same parameters, and is meant to represent random fluctuations in the control electronics, such as diurnal variations resulting from temperature instability. We assume that the errors on $\sigma_x$ and $\sigma_y$ are perfecly correlated, as mentioned in Section III. [ADD ANALYSIS OF PLOTS]

## VI. 2Q GATES

In this section, we present numerical results on a generating one and two-qubit gates that together with $RZ(\theta)$ rotations are universal for two-qubit computation. Our control Hamiltonian is given as:

$$H = \sum_{j=1}^{2} \epsilon_j \sigma_z^j + (1+\delta_j)(c_x^j x(t)\sigma_x^j + c_y^j(t)\sigma_y^j) \qquad (13)$$

We again assume that the standard deviations are .001 on all parameters $\delta_j$ and $c_i^j$ and that errors on the single qubit $\sigma_x$ and $\sigma_y$ rotations are perfectly correlated. [ADD ANALYSIS OF PLOTS]

[1] D. Aharonov, A. Kitaev, and J. Preskill, Physical Review Letters **96** (2006), 10.1103/physrevlett.96.050504.

[2] J. Preskill, "Quantum computing in the nisq era and beyond," (2018), arXiv:1801.00862.
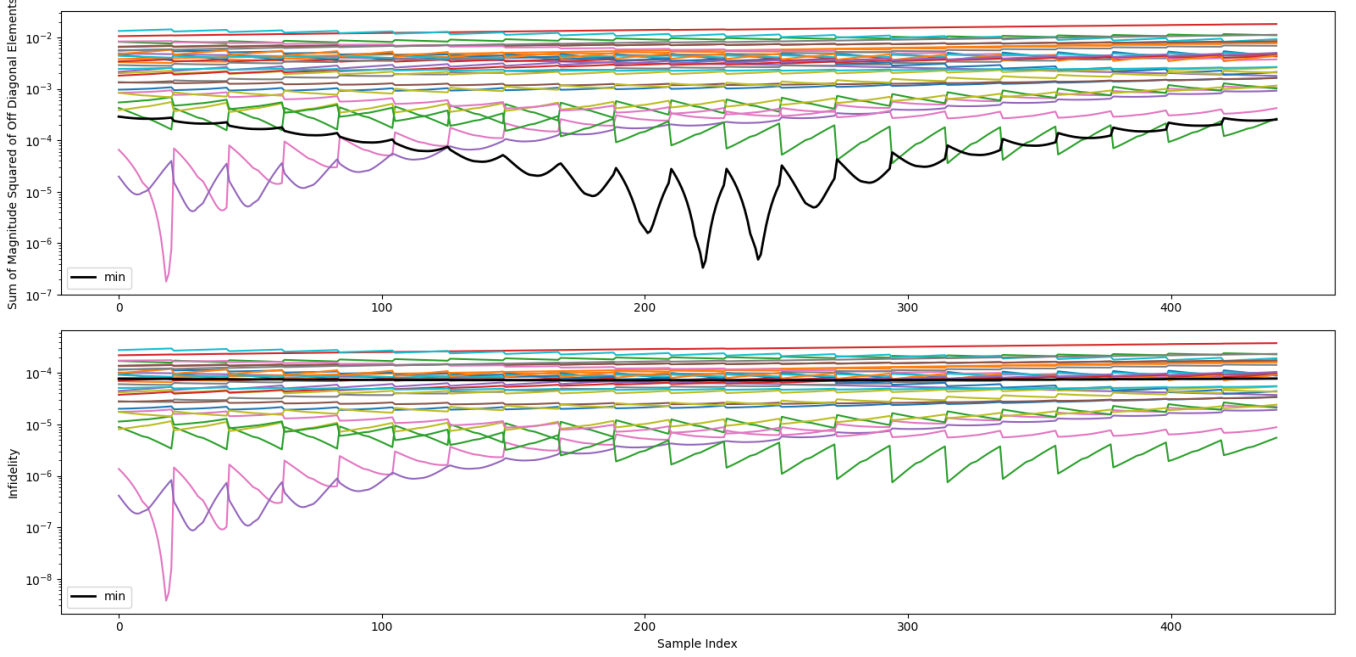
FIG. 2. In the above we have used our method to produce a BCS for a $RX(\frac{\pi}{2})$ pulse, with the Hamiltonian given in 12. A) shows the sum of the squared norms of the off diagonal terms in the Pauli Transfer Matrix of $\mathcal{E}$. B) gives the corresponding infidelity for each member of the BCS, and the optimal choice of weights.
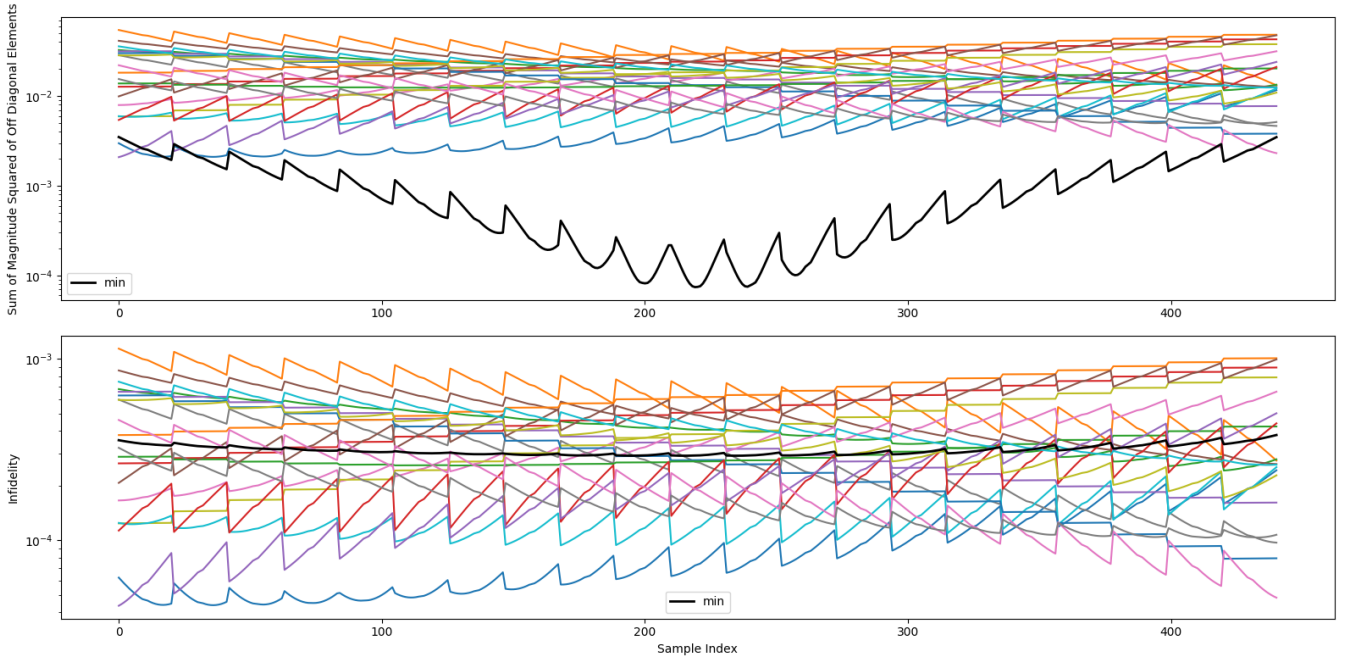


FIG. 3. In the above we have used our method to produce a BCS for a $RY(\frac{\pi}{2})$ pulse, with the Hamiltonian given in 12. A) shows the sum of the squared norms of the off diagonal terms in the Pauli Transfer Matrix of $\mathcal{E}$. B) gives the corresponding infidelity for each member of the BCS, and the optimal choice of weights.
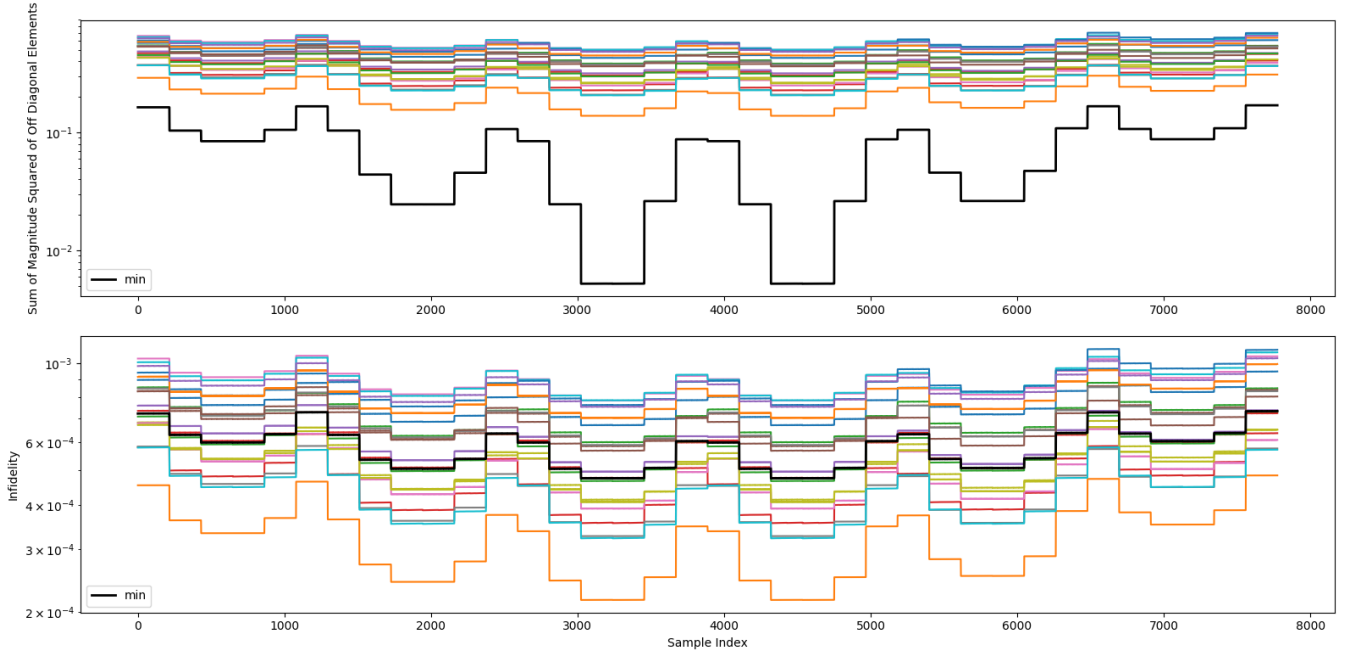
FIG. 4. In the above we have used our method to produce a BCS for a $ZZ(\frac{\pi}{2})$ entangling gate with controls given by 13. A) shows the sum of the squared norms of the off diagonal terms in the Pauli Transfer Matrix of $\mathcal{E}$. B) gives the corresponding infidelity for each member of the BCS, and the optimal choice of weights.

[3] A. Y. Kitaev, Russian Mathematical Surveys **52**, 1191 (1997).

[4] 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.

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

[6] J. J. Wallman and J. Emerson, Physical Review A **94** (2016), 10.1103/physreva.94.052325.

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

[8] B. Heim, K. M. Svore, and M. B. Hastings, "Optimal circuit-level decoding for surface codes," (2016), arXiv:1609.06373.

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

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

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

[12] 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.

[13] E. Magesan, D. Puzzuoli, C. E. Granade, and D. G. Cory, Physical Review A **87** (2013), 10.1103/physreva.87.012324.