Local gradient optimization of modular entangling sequences

A. A. Setser, M. H. Goerz, and J. P. Kestner

Phys. Rev. A **97**, 062339 — Published 26 June 2018

DOI: 10.1103/PhysRevA.97.062339

Local Gradient Optimization of Modular Entangling Sequences

A. A. Setser,[1] M. H. Goerz,[2,3] and J. P. Kestner[1]

[1]*Department of Physics, University of Maryland Baltimore County, Baltimore, Maryland 21250, USA*
[2]*Edward L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA.*
[3]*U.S. Army Research Laboratory, Computational and Information Sciences Directorate, Adelphi, MD 20783, USA.*

Implementation of logical entangling gates is an important step towards realizing a quantum computer. We use a gradient-based optimization approach to find single-qubit rotations which can be interleaved between applications of a noisy nonlocal gate to dramatically suppress arbitrary logical errors, while steering the evolution operator towards the perfectly entangling subset of SU(4) gates. The modularity of the approach allows for application to any two-qubit system, regardless of the Hamiltonian or details of the experimental implementation. This approach is effective for both quasi-static and time-dependent $1/f^{\alpha}$ noise. We also show how the fidelity of the final operation depends on both the fidelity of the local rotations and the noise strength.

I. INTRODUCTION

Quantum computing requires both local (single-qubit) and nonlocal (entangling) gate operations. Fault-tolerant quantum computing requires that all operations on the physical qubits can be performed with an error less than some "quantum error correction threshold", the precise value depending on the particular choice of encoding scheme. Surface codes offer an attractively high threshold of roughly 1% [1], but even then it is desirable to reduce the physical error rate further still in order to reduce the overhead associated with the code.

Errors generally may result from leakage, i.e., loss of population out of the logical subspace of a multilevel system, from *quantum* noise, i.e., dissipation, or from *classical* noise, i.e., fluctuations in the system parameters or control fields. Here, we will focus on (effective) spin-1/2 systems where leakage is not a concern. Dissipation generally depends inherently on the specific qubit implementation and encoding, and can be minimized only by realizing gate operations on a time scale that is short relative to the lifetime of the qubit. Thus, we focus on reducing the sensitivity to classical noise.

In most physical systems, nonlocal gates have significantly higher errors than local gates. For single-qubit gates, noise can be countered using composite pulse sequences, as has been experimentally demonstrated, for example, in NMR [2], trapped ions [3], and spin qubits [4]. For two-qubit entangling gates, however, the situation is more complicated.

One approach to two-qubit dynamical correction is pulse-shaping through optimal control, varying several parameters in the system's Hamiltonian as a function of time with numerically generated shaped pulses in order to minimize a target functional. Optimal control methods include both gradient-free methods, most commonly Nelder-Mead [5], often possible through use of a low-dimensional basis [6, 7]; and gradient based methods, gradient-ascent-pulse-engineering (GRAPE) [8, 9] and Krotov's method [10–13]. These methods have been shown to be a versatile tool for a wide range of tasks in quantum engineering [14, 15]. This includes finding control fields that are robust with respect to noise [16–20]. Although effective, this approach requires a precise and detailed knowledge of both the control Hamiltonian and the form of the noise, including any correlations between the two.

If one does not have such a complete model, an alternative approach is the application of pulse sequences: using the relatively high fidelity of the local gates one can construct a nonlocal gate sequence that cancels certain errors in the elementary nonlocal gates [21–24]. Knowledge of the system Hamiltonian is not required. Here, single-qubit rotations are inserted between applications of an entangling operation in such a way that *any* systematic error present in the entangling operation is canceled. These sequences of single-qubit rotations can generally be applied repeatedly to suppress systematic error to an arbitrary level [25, 26]. However, those analytical results can require a large number of single-qubit rotations, and accumulated imperfections in those rotations quickly diminish the performance of the sequence.

In this paper, we present a new method which seeks to combine the efficiency of numerical pulse-shaping with the agnosticism of composite pulse sequences. We use a gradient-based optimization to numerically find a series of single-qubit rotations to be interleaved between repeated application of a slightly entangling gate such that the entire sequence performs a perfect entangling gate while suppressing all possible logical errors. The modularity of our approach allows for application to any system regardless of the system's Hamiltonian or correlation of the noise with the control. In addition, it performs well not only for systematic noise, but also for time-dependent noise with a 1/f-type spectrum.

II. MODEL AND OPTIMIZATION METHOD

In order to counter the error in the numerical optimization, we sample an ensemble of noisy system evolutions. That is, we consider M separate noise realizations, and require that the optimized single-qubit operations perform well for the average over these realizations [18].

Within each noise realization, indexed by m, we assume that there exists a total evolution operator such that $|\psi(T)\rangle = \hat{\mathsf{U}}^{(m)}\,|\psi(t=0)\rangle$, where T is the total gate duration, and that this operator can be broken into a series of N steps. The time evolution operator for the realization (m) takes the form

$$\hat{\mathsf{U}}^{(m)} = \prod_{n=N}^{1} \underbrace{\exp\left[-\frac{i\pi}{N}\hat{\sigma}_{ZZ}\right]}_{\equiv\hat{\mathsf{Z}}} \underbrace{\exp\left[-\frac{i}{N}\hat{\Delta}_n^{(m)}\right]}_{\equiv\hat{\mathsf{D}}_n^{(m)}} \hat{\mathsf{R}}_n\,. \quad (1)$$

with n running backwards to account for time ordering. Within each step of our evolution operator, $\hat{\mathsf{Z}} \in \mathrm{SU}(4)$ with $\hat{\sigma}_{ZZ} \equiv \hat{\sigma}_Z \otimes \hat{\sigma}_Z$ is a weakly entangling gate. Throughout this work $\hat{\mathsf{Z}}$ is the Nth root of a 2π phase gate. In other words, the sequence consists of slicing an identity operation into N equal pieces and inserting local rotations in between. These local rotations steer the dynamics in arbitrary directions; the resulting total gate $\hat{\mathsf{U}}^{(m)}$ is not restricted to the diagonal gate that a pure $\hat{\sigma}_{ZZ}$ interaction would induce. We do not need to make any assumptions regarding how $\hat{\mathsf{Z}}$ is implemented; the effective Ising operation could itself be composed from any entangling interaction plus local rotations. However, the physical operation in the presence of noise consists of not only the desired entangling gate $\hat{\mathsf{Z}}$ but also a perturbation $\hat{\mathsf{D}}_n^{(m)}$, with

$$\hat{\Delta}_n^{(m)} \equiv \sum_{ij} \delta_{n,ij}^{(m)} \hat{\sigma}_{ij} \quad (2)$$

where the $\delta_{n,ij}^{(m)}$ are the error coefficients and $\hat{\sigma}_{ij} \equiv \hat{\sigma}_i \otimes \hat{\sigma}_j$. The factor of $1/N$ in the exponential reflects our assumption that the error scales with the size of the time slice. Lastly, the $\hat{\mathsf{R}}_n \in \mathrm{SU}(2) \otimes \mathrm{SU}(2)$ are controllable local rotations that are used collectively to counter the effects of the noise operators $\hat{\mathsf{D}}_n^{(m)}$ while steering the total evolution $\hat{\mathsf{U}}^{(m)}$ to a perfect entangling gate.

We choose a parametrization in terms of Euler angles [27],

$$\begin{aligned}\hat{\mathsf{R}}_n = &\exp\left[\frac{i\gamma_{1,n}}{2}\hat{\sigma}_Z\right]\exp\left[\frac{i\beta_{1,n}}{2}\hat{\sigma}_Y\right]\exp\left[\frac{i\alpha_{1,n}}{2}\hat{\sigma}_Z\right]\\ &\otimes\exp\left[\frac{i\gamma_{2,n}}{2}\hat{\sigma}_Z\right]\exp\left[\frac{i\beta_{2,n}}{2}\hat{\sigma}_Y\right]\exp\left[\frac{i\alpha_{2,n}}{2}\hat{\sigma}_Z\right].\end{aligned} \quad (3)$$

The $6N$ parameters $\gamma_{1,1}\ldots\alpha_{2,N}$ are the free control parameters for the optimization.

We now consider the functional to be minimized by our numerical search method. The functional should allow us to optimize towards two goals simultaneously:

1. minimal sensitivity of the sequence to noise, and
2. generation of a perfect entangler (PE), i.e., a gate which can produce a maximally entangled state from an unentangled one [28].

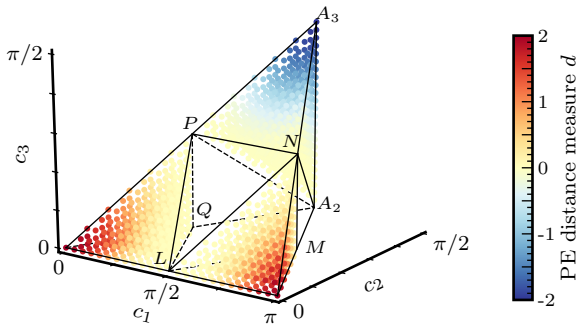

FIG. 1: (Color online.) The Weyl chamber, with the polyhedron of perfect entanglers (edges L–M–A_2–N–P–Q). The colored dots show the value of $d(c_1, c_2, c_3)$, Eq. (7) numerically transformed from the Makhlin invariants (g_1, g_2, g_3) to the Weyl chamber coordinates (c_1, c_2, c_3). For visual clarity, the values for d *inside* the polyhedron are not shown – they may be defined to zero.

Reducing the sensitivity of Eq. (1) to error $\hat{\Delta}^{(m)}$ can be achieved by maximizing the trace overlap of $\hat{\mathsf{U}}^{(m)}$ with the equivalent operator that would have been generated in the absence of noise,

$$F(\hat{\mathsf{U}}^{(m)}) = \frac{1}{16}\left|\mathrm{tr}\left(\hat{\mathsf{O}}^\dagger\hat{\mathsf{U}}^{(m)}\right)\right|^2, \quad \hat{\mathsf{O}} = \prod_{n=N}^{1}\hat{\mathsf{Z}}\hat{\mathsf{R}}_n\,, \quad (4)$$

for every noise realization (m). The gate error for a particular noise realization corresponding to the fidelity $F(\hat{\mathsf{U}}^{(m)})$ is

$$\epsilon(\hat{\mathsf{U}}^{(m)}) = 1 - F(\hat{\mathsf{U}}^{(m)})\,. \quad (5)$$

Note that the target state includes the local rotations, and therefore changes between every iteration of the optimization. Optimization of Eq. (4) serves to produce a known final gate which is robust against noise.

However, even if we obtain the maximum value of 1 for Eq. (4) it does not guarantee that the corresponding noise-free operation is an entangling one, as the interleaved local rotations strongly impact the entanglement dynamics of the final operation. Hence, we must add another term to our functional to ensure the $\hat{\mathsf{U}}$ is a perfect entangler.

For a gradient-based optimization scheme, the figure of merit for the realization of a perfect entangler must analytically connect to the time evolution operator $\hat{\mathsf{U}}$. It is not sufficient to simply calculate the concurrence of $\hat{\mathsf{U}}$ as this is a non-analytic quantity with an undefined gradient. The key to optimizing for maximal entanglement is to note that any SU(4) matrix $\hat{\mathsf{U}}$, can be written in the form of a Cartan decomposition $\hat{\mathsf{U}} = \hat{\mathsf{k}}_1\hat{\mathsf{A}}\,\hat{\mathsf{k}}_2$ [29–31], with the local operations $\hat{\mathsf{k}}_1, \hat{\mathsf{k}}_2 \in \mathrm{SU}(2) \otimes \mathrm{SU}(2)$, and

$$\hat{\mathsf{A}} = \exp\left[-\frac{i}{2}\left(c_1\hat{\sigma}_{XX} + c_2\hat{\sigma}_{YY} + c_3\hat{\sigma}_{ZZ}\right)\right] \quad (6)$$

representing a purely non-local operation $\in$ SU(4)/(SU(2) $\otimes$ SU(2)). The coefficients (c_1, c_2, c_3) that parametrize $\hat{\mathsf{A}}$ may be interpreted as geometric coordinates that identify a two-qubit gate up to single-qubit rotations. The symmetries in Eq. (6) restrict (c_1, c_2, c_3) to form the Weyl chamber [31] shown in Fig. 1. All two-qubit gates which are perfect entanglers lie within a subset of the Weyl chamber [32], the 7-faced polyhedron L–M–A_2–N–P–Q. The idea of optimizing for maximal entanglement, developed in Refs. [33, 34], is to minimize the geometric distance to the closest surface of the polyhedron.

The Weyl chamber coordinates (c_1, c_2, c_3) still cannot be calculated analytically from the gate $\hat{\mathsf{U}}$, but the closely related Makhlin invariants (g_1, g_2, g_3) can [35]. As there is a (non-analytic) mapping $(g_1, g_2, g_3) \to (c_1, c_2, c_3)$ [31], the desired distance can be expressed in terms of the Makhlin invariants as [33]

$$d(g_1, g_2, g_3) = g_3\sqrt{g_1^2 + g_2^2} - g_1\,. \tag{7}$$

Converting Makhlin invariants back to Weyl chamber coordinates, the value of $d(c_1, c_2, c_3)$ is indicated by color in Fig. 1.

The value of d in the top quadrant of the Weyl chamber (the so-called W_1 region, spanned by A_2–A_3–P–N) takes negative values. In order to have a functional that can be minimized to achieve the objective, we must thus change the sign. To identify when a gate is in the W_1 region, we can calculate

$$s = \pi - \cos^{-1} z_1 - \cos^{-1} z_3\,. \tag{8}$$

from the ordered roots $(z_1,\ z_2,\ z_3)$ of the cubic equation [31]

$$z^3 - g_3 z^2 + (4\sqrt{g_1^2 + g_2^2} - 1)z + (g_3 - 4g_1) = 0\,.$$

Also, inside the polyhedron of perfect entanglers, d takes values greater than zero. As we do not wish to bias the optimization towards perfect entanglers on the surface of the polyhedron, we define the value of the functional for any perfect entangler as zero. Thus, in total the functional that should be minimized for the realization of a perfect entangler is

$$\mathcal{D}(\hat{\mathsf{U}}) = \begin{cases} d & d > 0 \text{ and } s > 0 \\ -d & d < 0 \text{ and } s < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $d = d(g_1(\hat{\mathsf{U}}), g_2(\hat{\mathsf{U}}), g_3(\hat{\mathsf{U}}))$. It yields zero if and only if $\hat{\mathsf{U}}$ is a perfect entangler and is positive otherwise [33].

The total optimization function is the sum of the gate error, Eq. (5), and the distance from the closest perfect entangler, Eq. (9), averaged over all M noise realizations,

$$J = \frac{1}{M}\sum_{m=1}^{M}\left(\epsilon(\hat{\mathsf{U}}^{(m)}) + \mathcal{D}(\hat{\mathsf{U}}^{(m)})\right)\,. \tag{10}$$

Minimization of the functional serves to produce a perfect entangler which is robust against noise.

For the minimization of Eq. (10) through variation of the $6N$ free coefficients for the local error gates, we might in principle choose between gradient-free and gradient-based methods. While gradient-free methods are easy to implement, they show slow convergence except for a small number of optimization parameters [36]. Therefore, we look towards a gradient-based search method, which modifies every control parameter $\eta_i \in \{\gamma_{1,1} \ldots \alpha_{2,N}\}$ based on the derivative of the functional, $\frac{\partial J}{\partial \eta_i}$. The optimization is performed through the SciPy optimization package [37], using the L-BFGS-B algorithm [38]. This algorithm offers an increase in convergence and stability through the estimation of the Hessian (the matrix of second derivatives of the optimization functional). In fact, SciPy's implementation of the L-BFGS-B algorithm also allows to estimate the gradient of J numerically. This works for a moderate number of optimization parameters, and is an attractive proposition in our case, as the analytical gradient of Eq. (10) is exceedingly tedious to calculate. However, we note that Eq. (10) was specifically written to guarantee the existence of a well-defined analytical gradient. For larger sequence lengths N than we will consider here, the explicit calculation of the gradient would eventually become necessary, but here we proceed with the numerical approximation.

The convergence and success of the minimization for a single sequence length, N, via the L-BFGS-B algorithm requires a reasonable choice of "guess" parameters as a starting point. In order to maximize efficiency, if the greatest divisor of N is d, we repeat the solution for the length-d sequence $\frac{N}{d}$ times to use as the initial guess for the length-N sequence. This ensures that longer sequences will be constructed from shorter sequences which have already been optimized to be robust against noise. For prime-numbered sequence lengths, the local rotations are initialized as identity operations.

When performing the optimization, we impose termination conditions by specifying tolerances both for the functional and for the gradient of the functional. The optimization terminates when either of the following two conditions is fulfilled.

$$\begin{aligned} \frac{J^k - J^{k+1}}{\max\{|J^k|, |J^{k+1}|, 1\}} &\leq \mathrm{tol}_J, \\ \|\mathrm{proj}(\nabla J)\|_\infty &\leq \mathrm{tol}_{\nabla J}\,. \end{aligned} \tag{11}$$

where $\mathrm{proj}(\nabla J)$ is the projection of the gradient vector onto the space tangent to the active optimization bounds [39]. We find $\mathrm{tol}_J = \mathrm{tol}_{\nabla J} = 2.2 \times 10^{-6}$ to be a sufficient value to identify the minimal error in the optimizations discussed in the following section.

III. OPTIMIZATION RESULTS

For the analysis of the optimized sequences, we will consider the two desired results of the optimization – re-

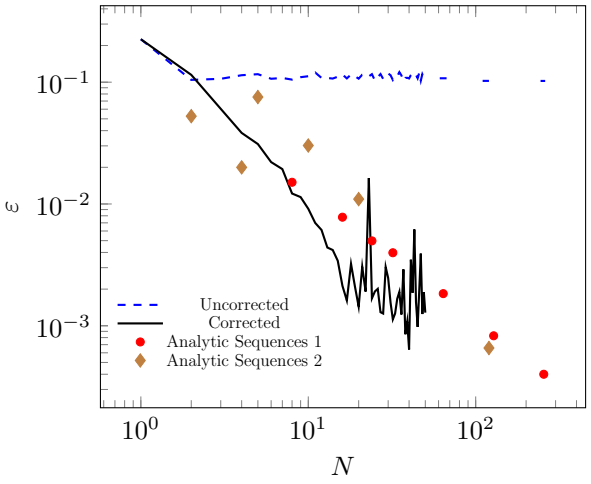

FIG. 2: (Color online.) Gate error in relation to sequence length for quasistatic nonlocal noise, assuming access to perfect single-qubit rotations; plotted on a log-log scale. Both lines shown are to be read at discrete values of N. The first set of analytic sequences are as generated in Ref. [25] and the second set in Ref. [26] (although, strictly speaking, only the last point, at $N = 120$, corresponds to a sequence designed for the generic error being treated here).

alization of a noise-free gate, and steering into the perfect entangling polyhedron – separately. The success of the optimization routine in producing a gate $\hat{\mathrm{U}}$ robust against noise is evaluated through the gate error averaged over noise realizations,

$$\varepsilon = \frac{1}{M}\sum_{m=1}^{M} \epsilon(\hat{\mathrm{U}}^{(m)}) \,. \tag{12}$$

The fidelity of the gate $\hat{\mathrm{U}}^{(m)}$ in a noise realization (m) with respect to the closest perfect entangler is evaluated as $F_{\mathrm{PE}}(\hat{\mathrm{U}}) \in [0,1]$, defined as [33]

$$F_{\mathrm{PE}}(\hat{\mathrm{U}}^{(m)}) = \begin{cases} \cos^2\left(\frac{c_1+c_2-\frac{\pi}{2}}{4}\right) & c_1+c_2 \leq \frac{\pi}{2} \\ \cos^2\left(\frac{c_2+c_3-\frac{\pi}{2}}{4}\right) & c_2+c_3 \geq \frac{\pi}{2} \\ \cos^2\left(\frac{c_1-c_2-\frac{\pi}{2}}{4}\right) & c_1-c_2 \geq \frac{\pi}{2} \\ 1 & \text{otherwise,} \end{cases} \tag{13}$$

where (c_1, c_2, c_3) are the Weyl chamber coordinates of the gate $\hat{\mathrm{U}}^{(m)}$. For evaluation purposes, this fidelity, or equivalently, the error

$$\epsilon_{\mathrm{PE}}(\hat{\mathrm{U}}^{(m)}) = 1 - F_{\mathrm{PE}}(\hat{\mathrm{U}}^{(m)}), \tag{14}$$

is a somewhat more direct measure than the (analytically differentiable) geometric distance to the polyhedron of perfect entanglers in the Weyl chamber that was used to steer the optimization, Eq. (9). Again, the error is averaged over all noise realizations.

$$\varepsilon_{\mathrm{PE}} = \frac{1}{M}\sum_{m=1}^{M} \epsilon_{\mathrm{PE}}(\hat{\mathrm{U}}^{(m)}) \,. \tag{15}$$

All optimized solutions are listed in the Supplemental Material [40].

A. Quasistatic Gate Noise

We first consider quasistatic noise; that is, noise that is constant on the timescale of the operation. Mathematically, this means that $\hat{\Delta}_n^{(m)}$ in Eq. (1) is the same for all time steps (independent of n), i.e., $\delta_{n,ij}^{(m)} \to \delta_{ij}^{(m)}$ in Eq. (2). The error coefficients $\delta_{ij}^{(m)}$ are drawn randomly from a normal distribution with a standard deviation of $\sigma = 0.13$. The standard deviation is chosen such that when no local rotations are made to suppress noise, the initial error, Eq. (12), is approximately 10%, a realistic experimental situation [41]. If we only use one noise realization instead of averaging over many, the optimization runs faster and gives higher fidelities, but clearly the optimal parameters will be tailored to work especially well for that specific noise realization and will not work well for an arbitrary realization. This lack of generality persists even when averaging the optimization over 10 noise realizations. However, we find that averaging over on the order of 100 noise realizations is sufficient to ensure that our results are robust against a general noise realization, i.e., the optimal parameters returned by running the optimization over any random set of 100 noise realizations remain essentially the same.

Figure 2 shows the results of this optimization. The gate error shows a steady decrease with increasing N up to around $N = 16$, after which the returns for increasing N greatly diminish. Note that for the longer sequences there is a lot of variability in the fidelity due to sequences at prime values of N having a significant disadvantage compared to those at non-prime values where the optimization can be initialized to a known good point using the results from shorter sequences. While not shown, the values for $\varepsilon_{\mathrm{PE}}$ are less than or equal to 10^{-8} for all sequences with $N \geq 2$ (including primes); with a majority of sequence lengths resulting in values of $\varepsilon_{\mathrm{PE}} = 0$.

For reference, we also show some points corresponding to analytical results which use local rotations to achieve error suppression [25, 26]. Ref. [25] uses $N = 8k$ applications, with k an integer, of an entangling Hamiltonian with local π pulses about x and y interspersed to isolate a desired σ_{ZZ} term in the Hamiltonian, with residual error terms scaling like $1/N$. In contrast, our present results show scaling closer to $1/N^2$ for small N. Our results saturate around $N = 16$, so eventually the analytical results do better, but one has to go up all the way to $N = 256$ to achieve similar performance with the scheme of Ref. [25]. Ref. [26] uses $N = 120$ entangling operations with local rotations (mostly π rotations) to cancel the leading order error completely generally. For smaller initial noise, this approach is quite superior to suppression as some power of N, but for the large noise values we have taken in this plot it performs similar to the previous analytic scheme.

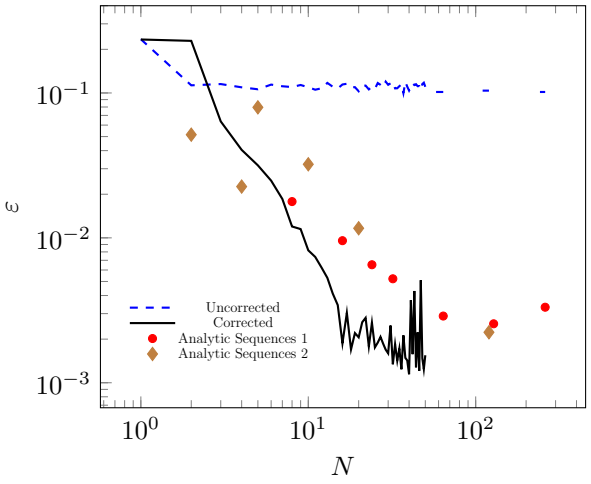

FIG. 3: (Color online.) Gate error in relation to sequence length for the noisy-rotation case, plotted on a log-log scale. All sources of error are quasistatic. The first set of analytic sequences are as generated in Ref. [25] and the second set in Ref. [26]

When some information is known about which type of errors are present, there are shorter variants of with N =2, 4, 5, 10, and 20. We show their performance as well, although they are not designed for the general error case we are considering here. Compared to Refs. [25, 26], our results show an order of magnitude improvement in error suppression for sequences under 50 segments. However, in contrast to our results, both analytical methods allow concatenation to achieve, in principle, arbitrarily small errors at very large N.

So far, we have assumed that the local operations $\hat{\mathsf{R}}_n$ that correct the error can be implemented perfectly. We now consider the effect of noise also in these local gates, i.e., imperfections in the control. To this end, we randomly perturb each of the $6N$ Euler angles $\eta_i \in \{\gamma_{1,1} \ldots \alpha_{2,N}\}$ in Eq. (3) according to

$$\eta_i \to \eta_i' = \eta_i(1+\delta_\eta)\,, \tag{16}$$

where δ_η is an error coefficient drawn randomly from a normal distribution with a standard deviation of $\sigma = 0.01$. This choice corresponds to the local rotations having experimentally realistic fidelities of approximately 99.9% [42], calculated as

$$F_R = \frac{1}{16}\left|\mathrm{tr}\left(\hat{\mathsf{R}}^\dagger\left(\gamma_1',\ldots,\alpha_2'\right)\hat{\mathsf{R}}\left(\gamma_1,\ldots,\alpha_2\right)\right)\right|^2\,, \tag{17}$$

and averaged over both 1000 different sets of error coefficients and 1000 sets of angles drawn randomly from a uniform distribution ranging from -4π to 4π.

The results of this optimization case are shown in Fig. 3. The fidelity again scales favorably with N compared to known analytic sequences [25, 26], as shown in Fig. 3. The scaling does not continue for arbitrarily large sequence length, eventually saturating, but now even the analytical results also saturate due to the accumulation local noise. In contrast, the performance of our sequences optimized in the presence of local noise is only very slightly worse than the performance of the optimal sequences in the absence of local noise. The maximum fidelity achieved with these noisy local rotations is still 99.90%, compared to 99.94% fidelity achieved in the case of perfect local rotations. Thus, the effect of realistic errors in the single-qubit rotations only marginally affects the optimal achievable performance.

Furthermore, the sequence optimized in the presence of local errors does not sacrifice any ability to suppress nonlocal errors; we have checked that using the new set of $6N$ parameters in the case of only nonlocal noise does not have worse performance than the parameters obtained by optimizing specifically for the case of nonlocal noise only. The optimization places no restrictions on the rotation angles. Without local noise, we find that angles remain in $[-\pi,\pi]$. When local noise is introduced, however, the resulting angles are in the range $[-4\pi,4\pi]$, i.e., some rotations must involve multiple cycles around the Bloch sphere in order for local errors to accumulate in an ultimately self-negating way. This is also seen in some analytic pulse sequences [43, 44]. These multiple cycles can be realized physically by, e.g., increasing the rotation time. As with the local-noise-free case, we have $\varepsilon_{\mathrm{PE}} \leq 10^{-8}$ for all sequences with $N \geq 2$.

Recall that the optimization did not target a specific operation a priori, but once the optimization is done once can characterize the specific robust perfect entangler produced. For instance, the operation produced by the $N = 16$ optimization is, in Cartan decomposed form,

$$\hat{\mathsf{U}} = \hat{\mathsf{k}}_1 \exp\{\frac{-i}{2}\,(\,2.250\hat{\sigma}_{XX} + 0.809\hat{\sigma}_{YY} + 0.018\hat{\sigma}_{ZZ}\,)\}\hat{\mathsf{k}}_2\,.$$

With the local operations $\hat{\mathsf{k}}_1$ and $\hat{\mathsf{k}}_2$ parametrized using Pauli vectors,

$$\hat{\mathsf{k}}_1 = \exp\{-i(-0.827\hat{\sigma}_X + 0.527\hat{\sigma}_Y - 0.865\hat{\sigma}_Z)\} \otimes \exp\{-i(-1.292\hat{\sigma}_X + 0.006\hat{\sigma}_Y + 1.589\hat{\sigma}_Z)\}$$
$$\hat{\mathsf{k}}_2 = \exp\{-i(-0.993\hat{\sigma}_X + 0.987\hat{\sigma}_Y - 0.793\hat{\sigma}_Z)\} \otimes \exp\{-i(0.268\hat{\sigma}_X - 0.965\hat{\sigma}_Y - 1.769\hat{\sigma}_Z)\}.$$

Eq. (13) can be used to confirm that the produced operation is indeed a perfect entangler.

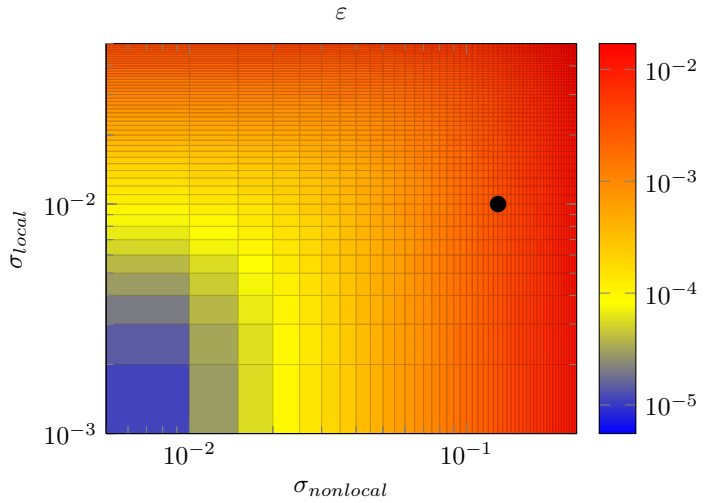

FIG. 4: (Color online.) Gate error of the $N = 16$ solution obtained in the presence of quasistatic noise in both the local and nonlocal operations, in relation to varying standard deviations of noise strengths. Plotted on a log-log-log scale. The solution obtained was optimized at $\sigma_{nonlocal} = 0.13$, $\sigma_{local} = 0.01$ (marked on plot), which corresponds to an uncorrected 90% fidelity in the nonlocal operations and a 99.9% fidelity in the locals.

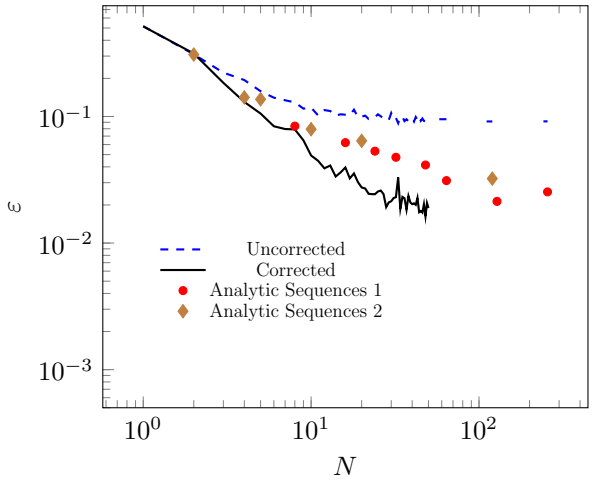

FIG. 5: (Color online.) Gate error in relation to sequence length, plotted on a log-log scale, assuming that both the local and nonlocal noise properties are consistent with those of $1/f$ noise.

While the optimization was performed with an approximate range of values for the noise strength, we wish to see how the solutions hold under varying noise strengths. Figure 4 shows the gate error for the $N = 16$ solution generated under quasistatic noise for a range of noise strengths, both in the local and nonlocal operations. Here we see that the obtained solution works well for all noise values which are less than or equal to the values used in the optimization. The noise strengths used in the optimization can be increased, which causes the obtained solutions to hold for the new range of larger noise values. However, the post-optimization fidelities at each N will be smaller, due to the difficulty of suppressing the now stronger noise. For instance, if we instead optimize sequences with an initial 95% fidelity in the local operations and an 80% fidelity in the nonlocals, the $N = 16$ sequence would now hold for a range of noise values greater than what is shown in Fig. 4. However, when this sequence is now evaluated at the (original) noise values $\sigma_{nonlocal} = 0.13, \sigma_{local} = 0.01$, a fidelity of only 99.5% is obtained rather than the 99.9% obtained under the previous optimization. Thus, by choosing the noise strengths under which the optimization is performed, one can trade between better results for a smaller range of noise values or more modest gains that hold for a larger range of noise.

B. Time-Dependent Gate Noise

We now turn to noise that changes on a faster time scale than the gate operation time. In this case, each $\hat{\Delta}_n^{(m)}$ in Eq. (1) is unique, but taken to be correlated consistent with the properties of $1/f^\alpha$ noise [45]. The same is taken to be true for the noise in the local rotations, δ_n in Eq. (16). For concreteness, in the following we will take $\alpha = 0.7$, consistent with charge noise measurements in semiconductor spin qubits [46]. The time-dependent noise is constructed by first generating a signal that randomly switches between ± 1; also known as a random telegraph process. The switching events themselves have a Poisson distribution. A single process has a characteristic time constant τ which determines the average time spent in one state over the course of the many switches, with a relaxation rate of $\nu = \frac{1}{2\tau}$. The power spectral density of the random telegraph signal is a Lorentzian of width ν. The $1/f^\alpha$ noise is then generated approximately by superimposing a finite number of fluctuators with relaxation rates evenly spaced logarithmically between $\nu_{\min}$ and $\nu_{\max}$.

We assign errors to the various segments of the pulse in the following way. Considering an arbitrary entangling gate time of T, we take $\nu_{\min} = \frac{1}{2(10T)}$ and $\nu_{\max} = \frac{1}{2(T/10)}$ and we use ten random telegraph fluctuators to generate the noise. For a single channel (fixed ij), the error coefficients for separate segments n are drawn by sampling a given time trace of the $1/f$ noise at different times t_n, where t_n is chosen randomly (via a uniform distribution) within the interval $\left[(n-1)\frac{T}{N}, n\frac{T}{N}\right]$ causing the noise within each channel to vary over time. Noise in the local operations is handled in the same way.

As in the quasistatic noise case, we scale the amplitude of the random telegraph signals such that the local rotations possess an approximate 99.9% fidelity on average and the uncorrected entangling operation has a fidelity of roughly 90.0%. In this case, the standard deviations of the sets of discretely sampled $1/f$ noise making up the local and nonlocal errors are $\sigma_{\text{local}} = 0.006$ and $\sigma_{\text{nonlocal}} = 0.2$.

The performance of the optimization routine in the time-dependent noise case is shown in Figure 5. Similar to the quasistatic noise case, the value of $\varepsilon_{\mathrm{PE}}$ is less than or equal to 10^{-8} for all sequences of length $N \geq 2$, with a majority possessing $\varepsilon_{\mathrm{PE}} = 0$. Note that, unlike in the quasistatic case, the uncorrected error diminishes somewhat as N increases. This reflects the fact that the time-dependent noise is less correlated than the quasistatic noise and for long sequences a little of the noise can average out even without any intervention. Comparing gate fidelities, the optimization routine is approximately an order of magnitude worse in the time-dependent case compared to the quasistatic case (99.0% fidelity compared to 99.9%), but still offers an order of magnitude improvement compared to the uncorrected gate. It should also be noted that rotations obtained in the presence of $1/f$ noise are also able to suppress quasistatic errors to the same levels that they were suppressed in the $1/f$ case, while the converse is not true.

In the case where there is $1/f$ nonlocal noise but no local noise (as might be expected if the local gates were practically instantaneous), we achieve a gate error of 5×10^{-3} at $N = 30$, compared to an error of 4×10^{-2} using the sequence of Ref. [26] at $N = 120$ or 2×10^{-2} at length $N = 256$ using the sequence of Ref. [25]. Thus, for $1/f$ noise as in the case of quasistatic noise, this optimization routine has significantly better performance than known analytic pulse sequences.

IV. CONCLUSION

Under reasonable conditions, we have shown that primitive entangling gates of fidelity near 90% can be used to construct composite entangling gates with error rates well below the $\sim 1\%$ fault tolerance threshold of the surface code. A possible difficulty lies in the fact that the interwoven optimized local operations are not simple rational multiples of π, but consist of generic angles. If we restrict the optimization to select the rotations from a small set of fractions of π, the optimization is no longer effective. Thus, experimentally calibrating a set of 10-20 unique local rotations is the cost one has to pay for boosting the fidelity of the nonlocal operation.

We have presented results from numerically optimizing a modular, composite pulse sequence, requiring no knowledge of the underlying Hamiltonian. The high-fidelity local rotations suppress arbitrary logical error in the nonlocal operation, while steering the overall operation into the class of perfect entanglers. Our results show significant improvement over known analytical results, requiring fewer local operations to achieve logical error suppression. The modularity of this approach allows for application to any two-qubit system regardless of the Hamiltonian. The approach has been shown to provide significant error suppression in both quasistatic and $1/f^{0.7}$ noise cases.

Acknowledgments

AAS and JPK acknowledge support by the National Science Foundation under Grant No. 1620740. MHG acknowledges support by ASD(R&E) under their Quantum Science and Engineering Program (QSEP), by the Army High Performance Computing Research Center (AHPCRC) under contract Number W911NF-07-2-0027, and by the U.S. Army Research Laboratory under Cooperative Agreement Number W911NF-16-2-0147.

[1] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, Phys. Rev. A **83**, 020302 (2011).
[2] L. M. K. Vandersypen and I. L. Chuang, Rev. Mod. Phys. **76**, 1037 (2005).
[3] E. Mount, C. Kabytayev, S. Crain, R. Harper, S.-Y. Baek, G. Vrijsen, S. T. Flammia, K. R. Brown, P. Maunz, and J. Kim, Phys. Rev. A **92**, 060301 (2015).
[4] J. Medford, L. Cywiński, C. Barthel, C. M. Marcus, M. P. Hanson, and A. C. Gossard, Phys. Rev. Lett. **108**, 086802 (2012).
[5] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
[6] T. Caneva, T. Calarco, and S. Montangero, Phys. Rev. A **84**, 022326 (2011).
[7] N. Rach, M. M. Müller, T. Calarco, and S. Montangero, Phys. Rev. A **92**, 062343 (2015).
[8] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magnet. Res. **172**, 296 (2005).
[9] P. de Fouquieres, S. Schirmer, S. Glaser, and I. Kuprov, J. Magnet. Res. **212**, 412 (2011).
[10] A. Konnov and V. F. Krotov, Autom. Rem. Contr. **60** (1999).
[11] S. E. Sklarz and D. J. Tannor, Phys. Rev. A **66**, 053619 (2002).
[12] J. P. Palao and R. Kosloff, Phys. Rev. A **68**, 062308 (2003).
[13] D. M. Reich, M. Ndong, and C. P. Koch, J. Chem. Phys. **136**, 104103 (2012).
[14] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, et al., Eur. Phys. J. D **69**, 279 (2015).
[15] C. P. Koch, J. Phys.: Condens. Matter **28**, 213001 (2016).
[16] H. Zhang and H. Rabitz, Phys. Rev. A **49**, 2241 (1994).
[17] R. L. Kosut, M. D. Grace, and C. Brif, Phys. Rev. A **88**, 052326 (2013).
[18] M. H. Goerz, E. J. Halperin, J. M. Aytac, C. P. Koch, and K. B. Whaley, Phys. Rev. A **90**, 032329 (2014).
[19] D. Dong, C. Chen, B. Qi, I. R. Petersen, and F. Nori, Sci. Rep. **5**, 7873 (2015).
[20] C.-H. Huang and H.-S. Goan, Phys. Rev. A **95**, 062325 (2017).
[21] J. A. Jones, Phys. Rev. A **67**, 012317 (2003).
[22] T. Ichikawa, U. Güngördü, M. Bando, Y. Kondo, and

M. Nakahara, Phys. Rev. A **87**, 022323 (2013).
[23] I. Cohen, A. Rotem, and A. Retzker, Phys. Rev. A **93**, 032340 (2016).
[24] Y. Tomita, J. T. Merrill, and K. R. Brown, New Journal of Physics **12**, 015002 (2010).
[25] C. D. Hill, Phys. Rev. Lett. **98**, 180501 (2007).
[26] F. A. Calderon-Vargas and J. P. Kestner, Phys. Rev. Lett. **118**, 150502 (2017).
[27] M. Hamada, Royal Soc. Open Sci. **1**, 140145 (2014).
[28] S. Balakrishnan and R. Sankaranarayanan, Phys. Rev. A **82**, 034301 (2010).
[29] N. Khaneja, R. Brockett, and S. J. Glaser, Phys. Rev. A **63**, 032308 (2001).
[30] B. Kraus and J. I. Cirac, Phys. Rev. A **63**, 062309 (2001).
[31] J. Zhang, J. Vala, S. Sastry, and K. B. Whaley, Phys. Rev. A **67**, 042313 (2003).
[32] P. Watts, M. O'Connor, and J. Vala, Entropy **15**, 1963 (2013).
[33] P. Watts, J. Vala, M. M. Müller, T. Calarco, K. B. Whaley, D. M. Reich, M. H. Goerz, and C. P. Koch, Phys. Rev. A **91**, 062306 (2015).
[34] M. H. Goerz, G. Gualdi, D. M. Reich, C. P. Koch, F. Motzoi, K. B. Whaley, J. Vala, M. M. Müller, S. Montangero, and T. Calarco, Phys. Rev. A **91**, 062307 (2015).
[35] Y. Makhlin, Quantum Inf. Process. **1**, 243 (2002).
[36] M. H. Goerz, K. B. Whaley, and C. P. Koch, EPJ Quantum Technology **2**, 21 (2015).
[37] E. Jones, T. Oliphant, P. Peterson, et al., *SciPy: Open source scientific tools for Python*, `http://www.scipy.org/`.
[38] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, SIAM J. Sci. Comput. **16**, 1190 (1995).
[39] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, ACM Trans. Math. Softw. **23**, 550 (1997).
[40] See Supplemental Material at [insert URL] for a table of solutions containing optimized parameters for each case.
[41] J. M. Nichol, L. A. Orona, S. P. Harvey, S. Fallahi, G. C. Gardner, M. J. Manfra, and A. Yacoby, NPJ Quantum Information **3**, 3 (2017).
[42] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, W. T. C., J. Mutus, A. G. Fowler, B. Campbell, et al., Nature **508**, 500503 (2014).
[43] X. Wang, L. S. Bishop, J. P. Kestner, E. Barnes, K. Sun, and S. Das Sarma, Nature Communications **3**, 997 (2012).
[44] J. P. Kestner, X. Wang, L. S. Bishop, E. Barnes, and S. Das Sarma, Phys. Rev. Lett. **110**, 140502 (2013).
[45] S. Kogan, *Electronic Noise and Fluctuations in Solids* (Cambridge University Press, 1996).
[46] O. E. Dial, M. D. Shulman, S. P. Harvey, H. Bluhm, V. Umansky, and A. Yacoby, Phys. Rev. Lett. **110**, 146804 (2013).