Optimized cross-resonance gate for coupled transmon systems

Susanna Kirchhoff,[1] Torsten Keßler,[1] Per J. Liebermann,[1,*] Elie Assémat,[1] Shai Machnes,[1]
Felix Motzoi,[1,2] and Frank K. Wilhelm[1]

[1]*Theoretical Physics, Saarland University, 66123 Saarbrücken, Germany*
[2]*Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark*

The cross-resonance (CR) gate is an entangling gate for fixed-frequency superconducting qubits. While being simple and extensible, it is comparatively slow, at 160 ns, and thus of limited fidelity due to on-going incoherent processes. Using two different optimal control algorithms, we estimate the quantum speed limit for a controlled-NOT CNOT gate in this system to be 10 ns, indicating a potential for great improvements. We show that the ability to approach this limit depends strongly on the choice of ansatz used to describe optimized control pulses and limitations placed on their complexity. Using a piecewise-constant ansatz, with a single carrier and bandwidth constraints, we identify an experimentally feasible 70-ns pulse shape. Further, an ansatz based on the two dominant frequencies involved in the optimal control problem allows for an optimal solution more than twice as fast again, at under 30 ns, with smooth features and limited complexity. This is twice as fast as gate realizations using tunable-frequency, resonantly coupled qubits. Compared to current CR-gate implementations, we project our scheme will provide a sixfold speed-up and thus a sixfold reduction in fidelity loss due to incoherent effects.

I. INTRODUCTION

Circuit QED is a promising technology for quantum computing. An important requirement for scalability of the architecture is high-accuracy implementation of two-qubit gates. A leading candidate for resonator-mediated interaction between a pair of superconducting qubits is the so-called cross-resonance (CR) gate [1–4] which has been implemented experimentally with over 99% average gate fidelity [5]. The CR gate functions by driving one qubit at the resonant frequency of the other qubit, inducing dynamics in the latter across the connecting resonator, i.e., "cross-resonantly." The design avoids the complexity and noise sources that are present in low-frequency magnetic (flux) tuning of the qubit-qubit interaction [6,7]. It also aims to improve on methods for high-speed addressing of specific two-qubit transitions, by utilizing the spatial addressability that comes from per-qubit control circuitry.

The primary impediments to high-fidelity operation currently come from two sources: incoherent errors, such as T_1 and T_2 processes, and coherent unitary errors, such as crosstalk [8] and frequency crowding [9]. The main method to counteract the former is to shorten gate times as much as possible, but the increased spectral width can drastically increase unitary errors, especially from higher-order corrections to the perturbative model of the gate mechanism.

In this work we systematically optimize CR gate control pulses for best fidelity, by shortening gate times as much as possible to reduce incoherent errors, while avoiding the adverse effects of increased coherent errors. We employ a full Tavis-Cummings model [10,11] which eliminates many of the analytic simplifications that set bounds on analysis of the gate in the deeply nonadiabatic regime. Moreover, it is known that

a careful analysis with regard to parametrization of the control sequence is required to efficiently tailor the control pulses to the constraints of the experimental apparatus [12,13] and to ensure simple pulse shapes which permit experimental calibration of the pulse. By lifting constraints on pulse complexity, we numerically estimate the quantum speed limit (QSL) [14–16] for the gate. However, the QSL is still dependent on other constraints imposed in the optimization problem and thus may also depend on the chosen parametrization of the control pulse. Therefore, we probe the relationship between the QSL and different physically meaningful parametrizations, which leads to a greater understanding of the limitations of the original cross-resonant control scheme and ultimately to improved control strategies.

This manuscript is organized as follows. Section II presents the theoretical model of the system. Section III introduces numerical methods. In Sec. IV we show that the unconstrained quantum speed limit is much shorter than the currently prevailing strategy. Section V describes the first optimization results and point outs the role of higher-frequency components. Section VI studies the QSL with the Fourier parametrization of the drive shape. Section VII explains the influence of the bandwidth constraint and proposes two highly practical solutions. Finally, in Sec. VIII we make concluding remarks.

II. SYSTEM

Let us consider two transmon qubits coupled by a bus resonator. Each transmon is described as an anharmonic oscillator and the coupling to the resonator is described by an appropriately extended Jaynes-Cummings model [17–19]. The qubits consist in the first two levels of these two anharmonic oscillators. As in Ref. [5], we assume that σ_x and σ_y type controls are available. We further assume that the qubit frequencies can be calibrated by a quasistatic flux line. The aim

*perjliebermann@gmail.com

TABLE I. Values of the parameters of the Hamiltonian in Eq. (2), which has an RWA applied. These correspond to typical parameters for the dispersive regime of circuit QED.

Parameter	Value in GHz
$\Delta/(2\pi)$	0.4
$g_1/(2\pi)$	0.1
$g_2/(2\pi)$	0.1
$\alpha_{1,2}/(2\pi)$	-0.32
$\delta_1/(2\pi)$	0.0
$\delta_2/(2\pi)$	-0.67

of the latter is not to provide another time-dependent control function, but to statically shift the qubit frequencies to a value more favorable to the optimization process. With the notations of Ref. [18], the Hamiltonian in the limit of low nonlinearity α [20] takes the form

$$H(t) = \sum_{k=1}^{2} \left[\omega_k b_k^\dagger b_k + \alpha_k b_k^\dagger b_k^\dagger b_k b_k + g_k(ab_k^\dagger + a^\dagger b_k) \right.$$
$$\left. + \sum_{l=1}^{L} \Omega_{k,l}(t) \cos\left(\omega_{k,l}^d t + \phi_{k,l}\right)(b_k + b_k^\dagger) \right] + \omega_r a^\dagger a,$$

(1)

where $a^\dagger, b_1^\dagger$, and $b_2^\dagger$ are the creation operators for the cavity and the two transmons respectively; g_k are the couplings between the resonator and the qubits; ω_r, ω_1, and ω_2 are the respective frequencies of the cavity and the two transmons; $\Omega_{k,l}$ are the low-frequency envelopes of the microwave drive, and $\omega_{k,l}^d$ are the carrier frequencies of the L control drives with phase offset $\phi_{k,l}$. After moving to the rotating frame close to the frequency of the first qubit, and applying a rotating-wave approximation (RWA), the Hamiltonian takes the form

$$H_{\mathrm{RWA}}(t) = \sum_{k=1}^{2} \left[\delta_k b_k^\dagger b_k + \alpha b_k^\dagger b_k^\dagger b_k b_k + g_k(ab_k^\dagger + a^\dagger b_k) \right.$$
$$\left. + \Omega_k^x(t)(b_k + b_k^\dagger) + i\Omega_k^y(t)(b_k^\dagger - b_k) + f_k b_k^\dagger b_k \right]$$
$$+ \Delta a^\dagger a,$$

(2)

where Δ is the detuning of the cavity from the principle (carrier) drive frequency of the controls, and δ_k are the detunings of the transmons. f_k are introduced in the optimization as a mechanism for setting the transmon detunings and/or shifting of the principle drive frequency. The numerical simulations and optimizations presented below include the three first levels of the resonator and the three first levels of each transmon. System parameters are detailed in Table I.

III. NUMERICAL METHODS

The field of quantum optimal control (QOC) provides methodologies by which a quantum system may be driven to a desired state, or undergo a desired evolution, in a fast and efficient manner. With the emergence of quantum technologies [21] the significance and impact of these techniques has grown.

Specifically, the requirement of very-high-fidelity gates for quantum computation and the complexity of the systems involved imply that approximate analytical solutions will not suffice, and numeric optimization theory must be applied to attain the desired process fidelities. Extensive research has gone into the problem of finding optimal driving of quantum systems. The field emerged in the mid-to-late 1980s with the first applications of QOC to chemical reactions and magnetic resonance imaging (MRI) [22–25], with experimental work continuing to this day [26,27]. Since, the scope of QOC has widened considerably, with applications to attosecond physics [28] and high harmonic generation [29], energy flow in biomolecules [30], and quantum computing [31,32], among others. QOC can be applied to both coherent and Markovian dynamics for state generation and other variants (see Refs. [33,34]).

QOC methods can be roughly divided into two categories, gradient-free and gradient-based optimization, where the terms refer to the availability of the gradient of the goal measure to be minimized, with respect to the control parameters. With gradient-free methods, one samples the goal function at one or more points in the control-parameter space and deduces one or more new points for sampling, where the expectation is of an improved goal measure, and then repeats the process. This approach is simple and flexible and is the only possible procedure for closed-loop calibration. However, such methods converge very slowly compared to gradient-driven optimization, particularly when optimizing high-dimensional parameter spaces. The most well-known member of this class of optimizers is the Nelder-Mead algorithm [35], on which the quantum chopped random basis (CRAB) and dressed CRAB (dCRAB) methods [36,37] are based. Other gradient-free algorithms include the covariance matrix adaptation evolution strategy (CMA-ES) [38], the simultaneous perturbation stochastic approximation (SPSA) [39], and genetic algorithms, among others. And while approaches are often better at handling large parameter spaces and the presence of noise, they are still slow to converge compared to gradient-driven methods.

When the gradient of the goal measure with respect to the control variables can be computed quickly (when compared to finite differences), gradient-based methods are preferred. These include the Krotov family of algorithms [40–42] and the gradient ascent pulse engineering (GRAPE) [43] method. Both are derived from the variational formulation of the QOC task [44], where the Schrödinger equation is imposed via a Lagrange multiplier, which turns out to be a conjugate state evolving back in time. The method by which the control fields are updated in both the Krotov and GRAPE methods are defined using time-local expressions, implying a piecewise-constant (PWC) control ansatz, which may be detrimental in cases of bounded bandwidth. Combining Floque theory with the variational approach also offers some advantages [45]. The QOC method gradient optimization of analytic controls (GOAT) (see Ref. [13] and detailed below) utilizes modified Schrödinger equations to compute the gradient and does not resort to variational calculus. We make use of GOAT below due to its simplicity and its flexibility of enforcing control constraints, such as bandwidth or power. For a comprehensive review of QOC, see Refs. [21,46].

Given a system whose dynamics is described by the drift Hamiltonian H_0 and is subject to a set of control Hamiltonians

H_k, the time-dependent Hamiltonian is

$$H(\bar{\alpha},t) = H_0 + \sum_{k=1}^{C} c_k(\bar{\alpha},t)H_k, \tag{3}$$

where c_k are the control functions (ansatz), with details prescribed by the set of parameters $\bar{\alpha}$. One is free to chose any control parametrization, e.g., the Fourier basis,

$$c_k(\bar{\alpha},t) = \sum_{j=1}^{m} A_{k,j} \cos(\omega_{k,j}t + \phi_{k,j}), \tag{4}$$

with

$$\bar{\alpha} = \{A_{k,j}, \omega_{k,j}, \phi_{k,j}\}_{k=1...C, j=1...m}. \tag{5}$$

In the systems investigated in this work we have found the Fourier and Erf parametrizations to produce pulse shapes with low parameter counts. The goal function to minimize is defined as the infidelity (projective SU distance) between the desired gate, U_{goal}, and the implemented gate, $U(T)$,

$$g(\bar{\alpha}) := 1 - \frac{1}{\dim(U)} |\text{Tr}(U_{\text{goal}}^{\dagger} U(T))|, \tag{6}$$

where $U(t)$ is the time-ordered ($\mathbb{T}$) evolution operator

$$U(\bar{\alpha},T) = \mathbb{T} \exp\left(\int_0^T -\frac{i}{\hbar} H(\bar{\alpha},t)dt \right). \tag{7}$$

QOC methods can be roughly divided into two categories: gradient-free and gradient-driven methods. The latter require the gradient of the goal function with respect to the control parameters, $\partial_{\bar{\alpha}} g(\bar{\alpha})$, and are much faster provided this gradient can be computed efficiently. Gradient-free methods are appropriate for closed-loop calibration and when the gradient cannot be determined with ease. Gradient-driven QOC methods require a gradient-driven search method over the parameter space $\bar{\alpha}$. Both GRAPE and GOAT (discussed below) utilize a standard algorithm for that purpose, limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [47].

GRAPE. When $H(\bar{\alpha},t)$ is taken to be a PWC function, the method of choice for QOC is GRAPE [33,43]. We note that when PWC is used as an approximation of a smooth control field, it introduces non-negligible inaccuracies, which are discussed in Ref. [13]. Let us enumerate the time slices $j \in \{1, \ldots, M\}$, each of duration Δt_j, $U_j = \exp(-\frac{i}{\hbar} \Delta t_j H_j) U_{j-1}$, where $H_j := H_0 + \sum_{k=1}^{C} c_{j,k} H_k$ and $U_0 := \mathcal{I}$, $U(T) := U_M$. Here $\bar{\alpha} = \{c_{j,k}\}$. The gradient of the goal function [Eq. (6)], $\partial_{\bar{\alpha}} g(\bar{\alpha})$, can be computed using the chain rule and $\partial_{\bar{\alpha}} U(\bar{\alpha})$. Noting the j,k component of $\bar{\alpha}$ appears only in U_j, the gradient of $U(T)$ is computed by $\partial_{\bar{\alpha}_{j,k}} U(T) = (\Pi_{b=M}^{j+1} U_b)(\partial_{\bar{\alpha}_{j,k}} U_j)(\Pi_{a=j-1}^{1} U_a)$. The expression to compute $\partial_{\bar{\alpha}_{j,k}} U_j = \partial_{c_{j,k}} \exp(-\frac{i}{\hbar} \Delta t_j H_j) = \partial_{c_{j,k}} \exp[-\frac{i}{\hbar} \Delta t_j (H_0 + \sum_{k=1}^{C} c_{j,k} H_k)]$ is rather cumbersome and requires a full eigendecomposition of H_j (see Ref. [33] for details). While computationally expensive, the eigendecomposition can be leveraged to perform exponentiation, propagation, and propagator gradients with little additional numerical effort. Therefore, GRAPE satisfies one of the conditions for a good gradient-based QOC method—the gradient of the propagator can be computed efficiently. The gradient of the propagator is

then used to compute the gradient of the goal function, feeding into the L-BFGS search algorithm, which seeks to minimize the goal function over the $\bar{\alpha}$ parameter space.

GOAT. Consider the gradient of the goal function (6) with respect to $\bar{\alpha}$,

$$\partial_{\bar{\alpha}} g(\bar{\alpha}) = -\text{Re}\left[\frac{g^*}{|g|} \frac{1}{\dim(U)} \text{Tr}(U_{\text{goal}}^{\dagger} \partial_{\bar{\alpha}} U(\bar{\alpha},T)) \right]. \tag{8}$$

Generally, $U(\bar{\alpha},T)$ does not have a closed form [see Eq. (7)], and therefore $\partial_{\bar{\alpha}} U(\bar{\alpha},T)$ cannot be computed directly. As U evolves by the equation of motion (EOM) $\partial_t U(\bar{\alpha},t) = -\frac{i}{\hbar} H(\bar{\alpha},t) U(\bar{\alpha},t)$, we may take the derivative of the U EOM with respect to $\bar{\alpha}$, swapping derivation order, resulting in a system of coupled EOMs:

$$\partial_t \begin{pmatrix} U \\ \partial_{\bar{\alpha}} U \end{pmatrix} = -\frac{i}{\hbar} \begin{pmatrix} H & 0 \\ \partial_{\bar{\alpha}} H & H \end{pmatrix} \begin{pmatrix} U \\ \partial_{\bar{\alpha}} U \end{pmatrix}. \tag{9}$$

$\partial_{\bar{\alpha}} H$ is computed using the chain rule and Eqs. (3) and (4). The coupled time evolution of the propagator (a single equation of motion) and its gradients ($C \times m$ equations—as per the dimension of $\vec{\alpha}$) may be performed by any mechanism for ordinary differential equation (ODE) integration that is accurate and efficient for time-dependent Hamiltonians, such as adaptive Runge-Kutta. A gradient-driven search over the parameter space is performed using L-BFGS.

IV. UNCONSTRAINED QUANTUM SPEED LIMIT

We first look for the QSL for the system described by Eq. (2) and Table I, by using the least-constrained parametrization, i.e., PWC, with no amplitude bounds and high resolution—500 time slices for each gate duration t_g. Assuming the parametrization is sufficiently flexible, the QSL observed results from the physics of the system. One standard method to probe the QSL numerically is to plot a measure of the gate fidelity as a function of the gate duration [19,48]. For different gate durations we perform multiple GRAPE optimizations, starting each with a different random initial guess pulse. The average final gate errors, defined as $g = 1 - \Phi_{\text{goal}}$, are shown in Fig. 1. We observe
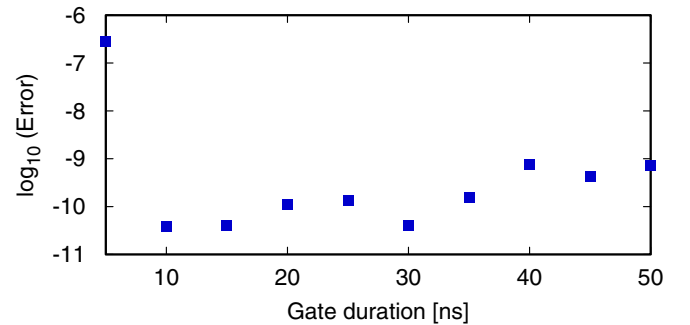

FIG. 1. We observe a 5- to 10-ns quantum speed limit (QSL) for the system described by Eq. (2) and Table I, using arbitrary amplitude piecewise-constant controls with a high sampling rate ($\leqslant 0.1$ ns) and no filter. The unitary error (infidelity; $1 - \Phi_{\text{goal}}$) of the optimized controlled-NOT (CNOT) gates is plotted as a function of gate duration, with each point representing the average error for multiple optimizations starting at random initial pulses. See Sec. IV for a complete discussion.

a clear jump between 5- and 10-ns gate times, which indicates the presence of a QSL. Interestingly, the speed limit is more than an order of magnitude shorter than the gate time presented in Ref. [2], suggesting a potential for reduction of incoherent errors by an order of magnitude. As no constraints are imposed at this stage, these optimal control shapes are beyond the capabilities of most existing experimental setups. However, this is likely to change in the near future, with adoption of AWGs with bandwidths above 10 GHz, such as those demonstrated in the circuit QED setup of Ref. [49], and FPGA control logic. Even without such hardware, by taking into account the constraints of the experiment implementation allows us to implement gates in times which, while higher than the unconstrained QSL, are significantly shorter than the original cross-resonance gate, as discussed in the following sections.

V. SMOOTH CONTROL IN THE TIME DOMAIN

To reduce the complexity of the generated pulse sequence, allowing effective calibration, we optimize control pulses in the PWC time domain, smoothed by a Gaussian filter. Based on the analytical methods presented in Ref. [1], and further refined and experimentally verified in Ref. [2], a cross-resonance gate is generated in a two-transmon system by applying driving on one qubit at the frequency which is resonant with the second qubit. Let us consider the effective coupling J_{eff} between the two qubits that quantifies the effective interaction mediated by the resonator. In the case where J_{eff} is small compared to the detuning between the qubits, δ_2, a drive at frequency $\tilde{\omega}_2 = \omega_2 - (J_{\text{eff}})^2/\delta_2$ generates dynamics that can be described by an effective Hamiltonian of the form

$$H_{\text{eff}} = u_1^{\text{eff}} \sigma_1^z \otimes \sigma_2^x + u_2^{\text{eff}} \sigma_1^x \otimes \mathbb{1}, \tag{10}$$

where σ_i are the Pauli operators and u_i^{eff} denote the relative scaling of the effective interaction. The ZX interaction present in this effective Hamiltonian can generate a CNOT gate directly, up to local rotations [50].

The performance of the initial guess is further improved by limiting the effect of the second term in Eq. (10). Here we assume this spectral constraint (i.e., suppressing the spectral weight of the second term) can be satisfied by using a second off-phase quadrature Ω^y of the control envelope set proportional to the derivative of Ω^x and inversely proportional to the qubit frequency separation δ_2, as in the derivative removal by adiabatic gate method [8,51].

These analytical techniques would not be sufficient to obtain high-fidelity gates at very short times for the model at hand due to the complexity of the level structure. However, the initial guess is found to be relevant enough to be located in the basin of attraction of a higher-fidelity solution, which allows the GRAPE algorithm to converge on a solution with suitably smooth features.

Thus, we choose the following initial control functions:

$$\Omega_1^x(t) = 0, \quad \Omega_1^y(t) = 0,$$

$$\Omega_2^x(t) = 0.4 \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right), \quad \mu = \frac{t_g}{2}, \quad \sigma = \frac{t_g}{4},$$

$$\Omega_2^y(t) = \frac{1}{\delta_2}\dot{\Omega}_1^x(t), \quad f_1 = 0, \quad f_2 = 0.1, \tag{11}$$
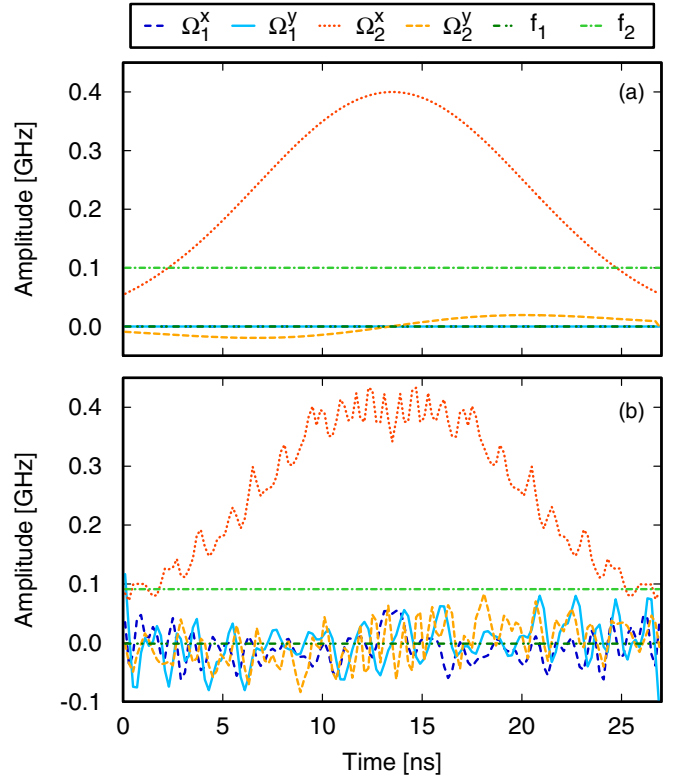

FIG. 2. High-fidelity CR control pulse sequence found with a sampling rate of 5 Gs/s. (a) Initial control guess pulse conforming to the analytical ansatz (11). (b) Optimized controls giving $\Phi_{\text{goal}} = 0.9999$. The pulses are relatively smooth and short (27 ns) but contain a large degree of complexity. See Sec. V for full details.

where the amplitudes are given in gigahertz, and f_1 and f_2 are constant, but tunable, frequency offsets relative to the drive frequency of qubits 1 and 2, respectively. Control amplitudes are bounded, and the sampling rate is lowered to 0.2 ns. When the optimization is run with the parameters in Table I, and a slightly increased gate time of $t_g = 27$ ns is set, we achieve a gate infidelity of 10^{-4}, assuming no incoherent processes. The pre- and postoptimization drive shapes are shown in Fig. 2. We observe that the strong cross-resonant drive remains after the optimization, which indicates that the CR scheme is still the main physical mechanism in play. The pulse induces complicated dynamics to obtain high-fidelity in a very short time, suggesting it may be possible to reduce gate times further, achieving an order of magnitude acceleration compared to Ref. [5], thus reducing the decoherence by a similar fraction and outperforming gates with tunable qubit frequency architectures [52]. The pulse is simplified compared to those used to identify the gate QSL, and the sampling rate is well within the current technological capacities of next generation microwave generators [49]. In Sec. VI we further simplify the parametrization of the pulse, which allows for experimental implementation with a minimal amount of overhead.

VI. QUANTUM SPEED LIMIT WITH THE FOURIER ANSATZ

It is extremely beneficial to arrive at quantum control sequences that are parametrized by a small number of parameters.
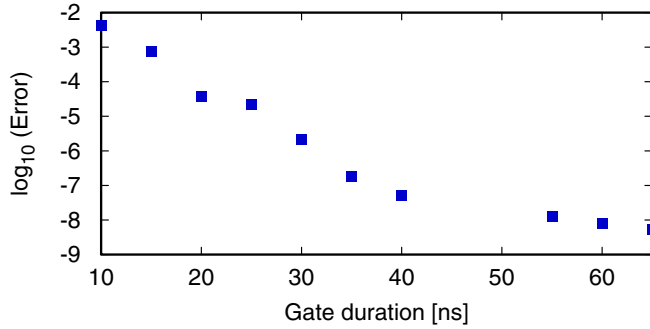
FIG. 3. Determination of the QSL using 167 Fourier components per control. The errors $(1 - \Phi_{\text{goal}})$ averaged over optimized pulses found for different random initial conditions are plotted for different gate durations. Errors are larger and decay more slowly when compared to a time-domain parametrization containing roughly 800 (piecewise-constant) control points, depicted in Fig. 1.

This simplicity can greatly aid in analytical, numerical, and experimental reproducibility of the solution, by allowing easy closed-up *in situ* calibration (or tune-up) of the control sequences. To gauge how the complexity of the pulses increases as we get close to the QSL, we use a parametrization based on the number of Fourier components of the pulse, though any other set of parameters more appropriate for the given control task could be substituted. Each control is parametrized by a truncated Fourier series that is fed to a product of two sigmoids to enforce the global bound and a smooth start and finish. The amplitude coefficients, frequencies, and phases are optimized using the GOAT algorithm [13], a recent gradient base algorithm that is capable of handling any analytic ansatz and can easily accommodate multiple and varied constraints.

We begin with a large number of Fourier components and probe the QSL. The number of components is chosen to offer roughly the same number of parameters to describe the drive shape as we had in the piecewise-constant case of Fig. 1. We observe in Fig. 3 that the QSL is less sharp than what was observed with the piecewise-constant description, which illustrates the clear influence of the choice of the control representation on the control landscape. Then, we iteratively remove the Fourier component with the smallest amplitude and reoptimize. The process terminates when simpler controls are unable to reach the minimum gradient threshold. In this case, component count was lowered down to only nine components. However, this reduction is at the cost of an increased gate time of 70 ns and the appearance of some very high frequencies. This offers a hint to explain why the unconstrained piecewise optimization manages to converge to a gate error of 10^{-10}, whereas the spectral optimization with a smaller frequency range converges only to 10^{-3}. It seems that high-frequency components are necessary for the fine-tuning needed to achieve a high accuracy. Moreover, this could also be a sign that the Fourier ansatz is not the most efficient for this system, and one may wish to try a few other analytic ansatzes. The GOAT package would be well suited for such study.

Nonetheless, limiting the number of Fourier components to 167 we find that the minimum time is around 15 ns to obtain 10^{-3} infidelity, which is consistent with the PWC

case and thus indicates this is likely the QSL for such a regime of control parameters, regardless of the chosen control parametrization. The dependence of QSL on the number of control parameters (15 ns for 167 components vs 70 ns for 9 components) demonstrates that the quantum speed limit is not only a function of time but also a function of pulse complexity. Thus, the Fourier basis is a good choice to minimize pulse complexity, as might be suited to *in situ* tuning of pulse parameters. This representation is also natural for enforcing bandwidth constraints. Lastly, it lends itself to the generation of highly simplified pulses by an iterative process of reducing the number of Fourier components. We do have to keep in mind, however, that microwave-pulse-shaping technology is typically in the time domain.

VII. BANDWIDTH-CONSTRAINED PULSES

The piecewise-constant parametrization, with high-resolution controls, provides a theoretic lower bound to gate times. In practice, time-resolved optimizations are limited by the AWG's bandwidth, as well as other filtering induced by system components. We therefore optimize the pulses for an AWG with a finite time resolution of 1 ns, fine steps of 0.2 ns, buffers of 4 ns in duration at the beginning and the end, and filtering of the signal, consistent with earlier work [12]. Filtering is applied via a Gaussian window function with the standard deviation $\sigma = 0.4$ ns, i.e., a bandwidth of 331 MHz. We consider two control ansatzes: one in which the controls utilize a single carrier frequency (the standard CR scheme), and one in which two carrier frequencies are employed.

When a single carrier frequency is used, a minimum gate time of 70 ns is achieved, with a fidelity of 0.999. GRAPE optimization generated the controls and spectra shown in Fig. 4. Clearly visible is the reduced bandwidth, effected by the control filtering. As a result, the pulse requires significantly more time than the QSL. However, it is still less than half the time required by current implementations of the CR gate [5], implying only half as much fidelity will be lost to incoherent processes.

Moving to two carrier frequencies and retaining the bandwidth limitations on the AWGs, we see a drastic reduction in the gate time. The intuition to this scheme stems from Sec. V, where one can identify two principle components in the frequency spectrum of the controls appearing in Fig. 2. As can be expected, the two principle frequencies are proximate to the resonance frequencies of the two qubits. More precisely, the qubit frequency at 0.67 GHz is shifted by f_2 to 0.57 GHz, which is then bifurcated by the Rabi splitting $2g$ to 0.47 and 0.67 GHz. Our control pulses take the following forms:

$$\Omega_1^x(t) = \Omega_1^{x'}(t) + \cos[(\delta + g)t]\Omega_1^{x''}(t),$$

$$\Omega_1^y(t) = \Omega_1^{y'}(t) + \cos[(\delta + g)t]\Omega_1^{y''}(t),$$

$$\Omega_2^x(t) = \Omega_2^{x'}(t) + \cos[(\delta - g)t]\Omega_2^{x''}(t),$$

$$\Omega_2^y(t) = \Omega_2^{y'}(t) + \cos[(\delta - g)t]\Omega_2^{y''}(t), \qquad (12)$$

where the quotes symbols ($''$) denote the new (AWG) control functions, and $\delta = \delta_2 + f_2$ is the qubit separation. This doubles the number of functions to optimize. A Gaussian filter with a bandwidth of 331 MHz is added to account for the distortion
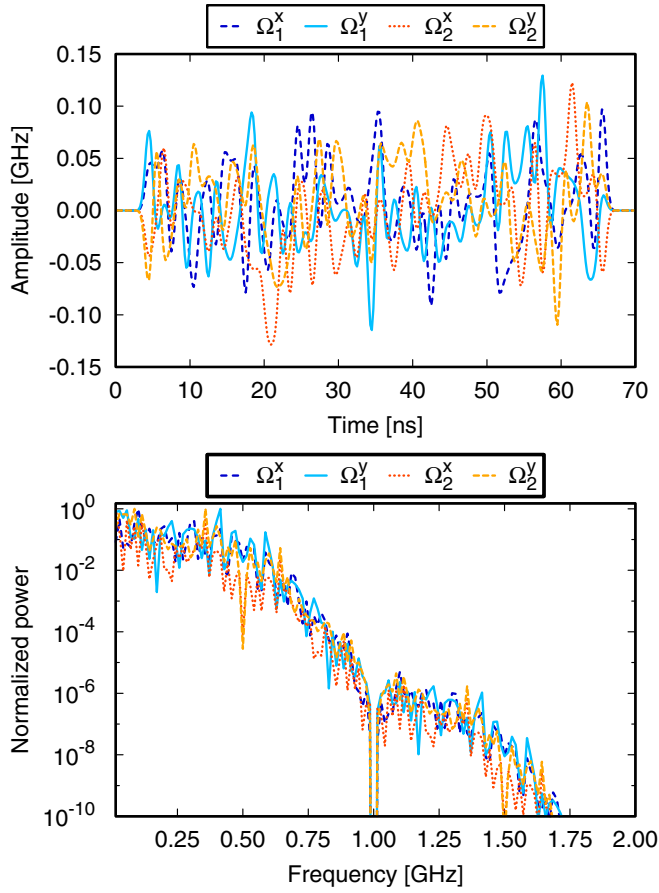
FIG. 4. Pulse shapes and spectra of the optimal drive shape for filtered PWC ansatz with a gate time of 70 ns and a single carrier frequency. A fidelity of 0.999 is achieved using a 331-MHz-bandwidth Gaussian filter.
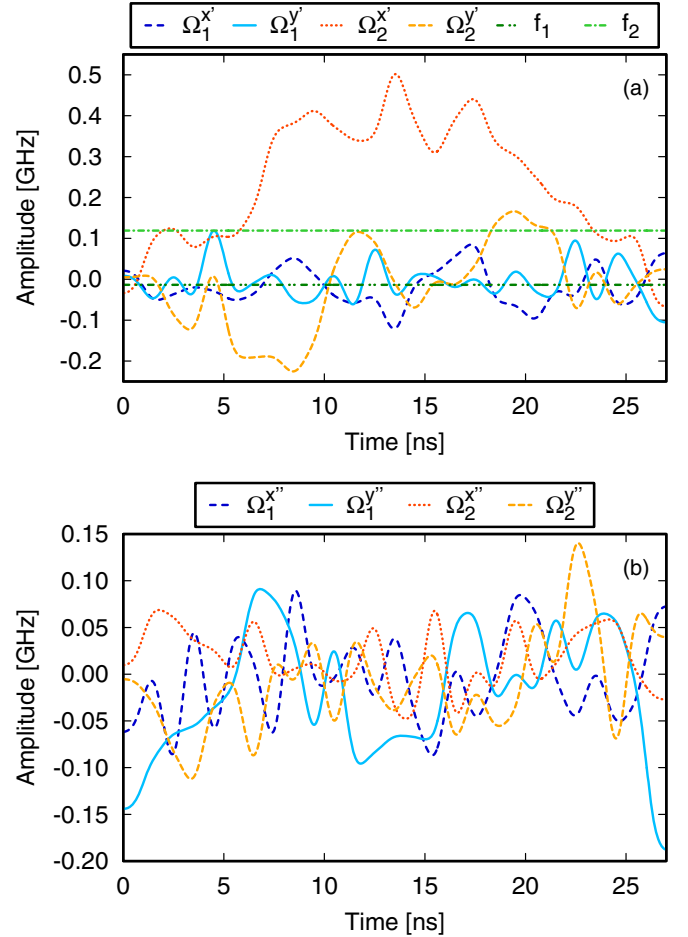

FIG. 5. Optimization results for control pulses using two carrier frequencies and a bandwidth filter, achieving 0.9999 fidelity when incoherent processes are ignored. (a) Controls at qubit 1 frequency. (b) Controls at qubit 2 frequency ($\delta \pm g$).

of the piecewise-constant control functions by wave-form generators [12]. The smooth controls, in addition to being a constraint of the system, improve optimization convergence speed and may help with experimental imperfections and unforeseen low-pass filters in the system. The pulse complexity is reduced not only by decreasing the sampling rate and the bandwidth of the pulses but also by the very short gate time of 27 ns, which indicates that the total number of control points needed is reasonably small, on the order of 100.

The optimization is carried out with coarse pixels of 1 ns and a fine time step of 0.05 ns and reaches a final fidelity of 0.9999. The control functions optimized $\Omega_j^{i'}$, $\Omega_j^{i''}$, and f_j are shown in Fig. 5. The constant value of the frequency detuning is also optimized but its value appears to be stuck in a local minimum and does not evolve significantly during the optimizations, suggesting additional improvement may be achieved by fine-tuning the choice of drive frequencies near the qubit transitions.

In Fig. 6, fidelity dependence on $f_1 \pm \epsilon_1$ and $f_2 \pm \epsilon_2$ is plotted. We notice qubit-driven detuning errors ϵ_i on the order of up to about 2 MHz lead to a drop of fidelity by 2 orders of magnitude. Experimentally, miscalibrations are expected to be much smaller than this amount.

In Fig. 7, we examine the fidelity's dependence on dephasing time, with dephasing and relaxation $T_1 = T_2$ and the cavity decay rate set to $T_P = 100 \, \mu$s. The full evolution is then given
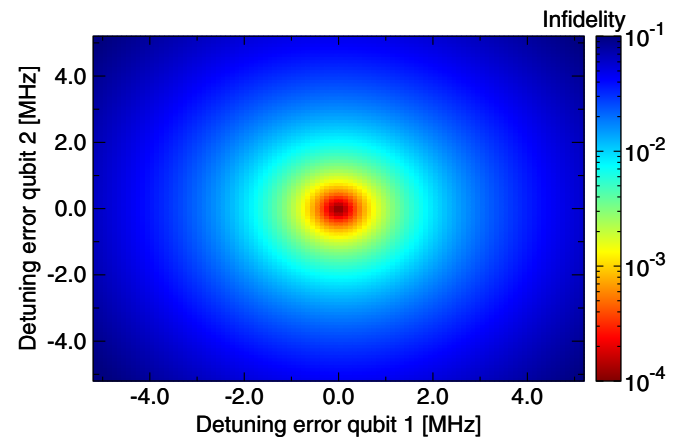

FIG. 6. Infidelity of the CR gate as a function of the miscalibration of qubit 1 and qubit 2 frequencies, with other model parameters specified by Table I. The control pulses are as depicted in Fig. 5.
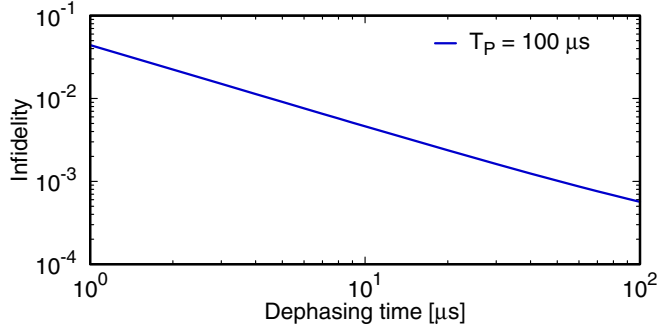
FIG. 7. Infidelity of the CR gate as a function of equal dephasing and decoherence rates. Model parameters follow Table I, and control pulses follow Fig. 5.

by the master equation

$$\dot{\rho} = -i[H,\rho] + \gamma_j \sum_j A_j \rho A_j^\dagger - \frac{1}{2} A_j^\dagger A_j \rho - \frac{1}{2} \rho A_j^\dagger A_j.$$

$$(13)$$

The Lindblad operators are b_i for relaxation, $\sqrt{b_i^\dagger b_i}$ for pure dephasing, and a for the cavity decay. We see that for typical values of experimental dissipation losses ($> 100\ \mu s$), the error is limited only by the precision of our unitary optimization and not by additional nonunitary losses, validating our estimates for the CR quantum speed limit.

Finally, we test the validity of the truncation of the Hilbert space. The optimized control functions were applied to a larger Hilbert space utilizing six levels for the resonator and each transmon. The maximum population reached during the dynamics is noted in Fig. 8. For the transmons, we take the maximum over both qubits $M_r = \max_t[p_r(t)]$ and $M_{\text{Transmons}} = \max_{t; i \in 1,2}[p_{r_i}(t)]$. The population of the resonator's third level, as well as the qubits' fifth level, remain below 10^{-3}. Unsurprisingly, the transmon is leakier than the resonator by just about an order of magnitude: Leakage to level three is approximately 10^{-3} for the resonator and 10^{-2} for the qubits.
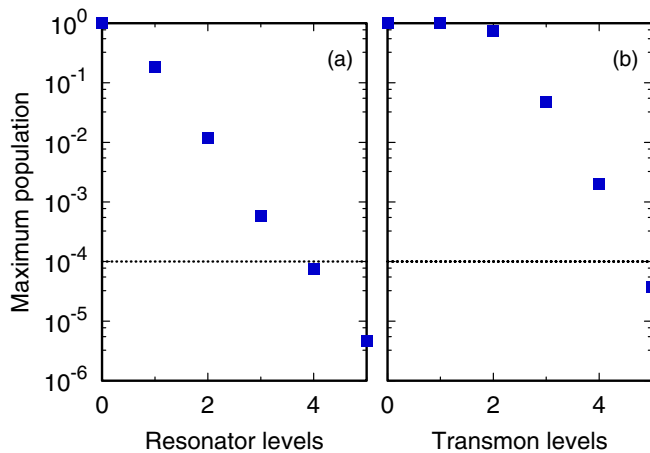

FIG. 8. Maximum transient population leakage to the higher levels of the resonator (a) and the transmons (b) during the dynamics driven by the optimal controls described in Fig. 5 and the model parameters detailed in Table I. Note the logarithmic scaling.

VIII. DISCUSSION

In this work we have demonstrated the very significant benefits which can be derived from applying quantum optimal control to the problem of quantum gate generation and more specifically to superconducting gates. We have chosen to focus on CR gates as one of the leading architectures for superconducting gates, which benefit from not requiring the overhead or noise sources present in tunable-frequency qubits.

The current state of the art for the CR CNOT gate is an infidelity of 0.009 ± 0.002 [5]. One of the primary reasons for the imperfect fidelity is the 160-ns duration of the gate, allowing incoherent processes to induce an infidelity of 0.004 ($T_1 \simeq 40\ \mu s$ and $T_2 \simeq 5\ \mu s$). Therefore, to improve fidelity significantly, one must shorten the gates (in parallel with efforts to increase coherence times and reduce coherent errors).

We note that the model used for the CR circuit is not perfect, as with any experiment: some aspects of the model are unknown (e.g., random defects in the substrate coupling to the circuit), and some characterization gaps and parameter drifts are unavoidable. Therefore, every proposed pulse sequence will have to be calibrated in a closed-loop tune-up process—a gradient-free search over the space of pulse parameters with the goal of minimizing infidelity. Such calibrations are practical only when the pulses are described by a small number of parameters. Therefore, practical pulses must have a simple description.

It is critical to determine the limits of achievable performance with any given circuit design, CR gates included, as this affects the decision of whether new circuits need to be developed to achieve fidelity goals. We have therefore employed several quantum optimal control techniques (GRAPE and GOAT) to determine the quantum speed limit and to design simple bandwidth-constrained control pulses which implement significantly faster CR gates.

Specifically, we have shown that the speed limit of the system is between 5 and 10 ns, implying an incoherence limit of below 5×10^{-4}. Unfortunately, the pulse shapes required to reach such speeds are too complex to reliably implement in the experiment and far too complex to calibrate.

We therefore explored alternate routes: the first follows the standard control scheme, where only the control qubit is driven, and the second requires contemporaneous driving with two carrier frequencies. Both approaches yielded significant improvement over the state of the art.

With a single carrier, we have achieved a CNOT gate in only 70 ns. A further dramatic acceleration can be achieved when a second carrier is introduced: We have identified a pulse sequence which implements a CNOT gate in only 27 ns. For such short durations, incoherent effects induce less than 10^{-3} infidelity. We note that the control fields used are low-pass-filtered to the current circuit's control bandwidth and are therefore directly implementable. Moreover, both pulses are described by less than 100 parameters and are therefore calibratable. To reduce parameter count further, one may employ the GOAT optimal control method, to allow additional control ansatzes to be explored.

Our exploration of the quantum speed limit further shows that, unlike the case where there are no direct qubit drives and the natural (perfect) entangler to use is the iSWAP [53,54],

a CNOT gate is instead a good fit for architectures where such drive lines do in fact exist. Moreover, in comparison to the exhaustive parameter search for a global QSL without a direct line, [54], in our limited local search with a direct line we are able to cut in half the global QSL with our bandwidth-constrained pulses. Nonetheless, one may still be able to use the insights from the former to find an even faster operating regime for the transmons in the latter case, notably by moving the transmon frequencies towards the quasidispersive straddling qutrits (QuaDiSQ) regime [54].

This work also motivates further use of this already prominent gate, being in fact even faster than the entangling gates used in the tunable-frequency implementations, which suffer from extra noises originating at the additional flux-tuning circuitry.

To conclude, we have shown that the application of quantum optimal control to the cross-resonance superconducting CNOT gate can reduce pulse duration from 160 to 27 ns, using control sequences which are well within experimental capabilities, are described by a small number of parameters, and are therefore calibratable. Thus, we demonstrate the potential of reducing incoherent effects fivefold, significantly improving gate fidelity.

[1] G. S. Paraoanu, Phys. Rev. B **74**, 140504 (2006).

[2] J. M. Chow, A. D. Córcoles, J. M. Gambetta, C. Rigetti, B. R. Johnson, J. A. Smolin, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **107**, 080502 (2011).

[3] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 060501 (2012).

[4] J. M. Chow, J. M. Gambetta, E. Magesan, D. W. Abraham, A. W. Cross, B. R. Johnson, N. A. Masluk, C. A. Ryan, J. A. Smolin, S. J. Srinivasan, and M. Steffen, Nat. Commun. **5**, 4015 (2014).

[5] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, Phys. Rev. A **93**, 060302 (2016).

[6] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **76**, 042319 (2007).

[7] P. Groszkowski, A. G. Fowler, F. Motzoi, and F. K. Wilhelm, Phys. Rev. B **84**, 144516 (2011).

[8] F. Motzoi and F. K. Wilhelm, Phys. Rev. A **88**, 062318 (2013).

[9] L. S. Theis, F. Motzoi, and F. K. Wilhelm, Phys. Rev. A **93**, 012324 (2016).

[10] M. Tavis and F. W. Cummings, Phys. Rev. **170**, 379 (1968).

[11] M. Tavis and F. W. Cummings, Phys. Rev. **188**, 692 (1969).

[12] F. Motzoi, J. M. Gambetta, S. T. Merkel, and F. K. Wilhelm, Phys. Rev. A **84**, 022307 (2011).

[13] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, Phys. Rev. Lett. **120**, 150401 (2018).

[14] L. B. Levitin and T. Toffoli, Phys. Rev. Lett. **103**, 160502 (2009).

[15] M. Lapert, Y. Zhang, M. Braun, S. J. Glaser, and D. Sugny, Phys. Rev. Lett. **104**, 083001 (2010).

[16] T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Phys. Rev. Lett. **103**, 240501 (2009).

[17] C. Rigetti, J. M. Gambetta, S. Poletto, B. L. T. Plourde, J. M. Chow, A. D. Córcoles, J. A. Smolin, S. T. Merkel, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, Phys. Rev. B **86**, 100506 (2012).

[18] A. W. Cross and J. M. Gambetta, Phys. Rev. A **91**, 032325 (2015).

[19] D. J. Egger and F. K. Wilhelm, Supercond. Sci. Technol. **27**, 014001 (2014).

[20] B. Khani, J. M. Gambetta, F. Motzoi, and F. K. Wilhelm, Phys. Scr. **2009**, 014021 (2009).

[21] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, Eur. Phys. J. D **69**, 279 (2015).

[22] D. J. Tannor and S. A. Rice, J. Chem. Phys. **83**, 5013 (1985).

[23] S. Conolly, D. Nishimura, and A. Macovski, IEEE Trans. Med. Imaging **5**, 106 (1986).

[24] A. P. Peirce, M. A. Dahleh, and H. Rabitz, Phys. Rev. A **37**, 4950 (1988).

[25] R. Kosloff, S. Rice, P. Gaspard, S. Tersigni, and D. Tannor, Chem. Phys. **139**, 201 (1989).

[26] T. Brixner, N. H. Damrauer, P. Niklaus, and G. Gerber, Nature (London) **414**, 57 (2001).

[27] L. Levin, W. Skomorowski, L. Rybak, R. Kosloff, C. P. Koch, and Z. Amitay, Phys. Rev. Lett. **114**, 233003 (2015).

[28] S. Haessler, T. Balciunas, G. Fan, G. Andriukaitis, A. Pugzlys, A. Baltuska, T. Witting, R. Squibb, A. Zaïr, J. W. G. Tisch, J. P. Marangos, and L. E. Chipperfield, Phys. Rev. X **4**, 021028 (2014).

[29] R. Bartels, S. Backus, E. Zeek, L. Misoguti, G. Vdovin, I. P. Christov, M. M. Murnane, and H. C. Kapteyn, Nature (London) **406**, 164 (2000).

[30] T. Buckup, T. Lebold, A. Weigel, W. Wohlleben, and M. Motzkus, J. Photochem. Photobiol. A **180**, 314 (2006).

[31] E. Zahedinejad, J. Ghosh, and B. C. Sanders, Phys. Rev. Lett. **114**, 200502 (2015).

[32] F. Dolde, V. Bergholm, Y. Wang, I. Jakobi, B. Naydenov, S. Pezzagna, J. Meijer, F. Jelezko, P. Neumann, T. Schulte-Herbrüggen, J. Biamonte, and J. Wrachtrup, Nat. Commun. **5**, 3371 (2014).

[33] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Phys. Rev. A **84**, 022305 (2011), software available at qlib.info.

[34] C. P. Koch, J. Phys.: Condens. Matter **28**, 213001 (2016).

[35] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).

[36] P. Doria, T. Calarco, and S. Montangero, Phys. Rev. Lett. **106**, 190501 (2011).

[37] N. Rach, M. M. Müller, T. Calarco, and S. Montangero, Phys. Rev. A **92**, 062343 (2015).

[38] N. Hansen, S. D. Müller, and P. Koumoutsakos, Evol. Comput. **11**, 1 (2003).

[39] J. C. Spall, IEEE Trans. Autom. Control **37**, 332 (1992).

[40] V. Krotov and I. Feldman, Eng. Cybern. **21**, 123 (1983).

[41] D. J. Tannor, V. Kazakov, and V. Orlov, Control of photochemical branching: Novel procedures for finding optimal pulses and global upper bounds, in *Time-Dependent Quantum Molecular Dynamics*, edited by J. Broeckhove and L. Lathouwers (Springer, Boston, MA, 1992), pp. 347–360.

[42] D. M. Reich, M. Ndong, and C. P. Koch, J. Chem. Phys. **136**, 104103 (2012).

[43] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magn. Reson. **172**, 296 (2005).

[44] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko, *LS Pontryagin: Selected Works*, Vol. 4 (CRC Press, Boca Raton, FL, 1986).

[45] B. Bartels and F. Mintert, Phys. Rev. A **88**, 052315 (2013).

[46] C. Brif, R. Chakrabarti, and H. Rabitz, New J. Phys. **12**, 075008 (2010).

[47] J. Nocedal, Math. Comput. **35**, 773 (1980).

[48] J. J. W. H. Sørensen, M. K. Pedersen, M. Munch, P. Haikka, J. H. Jensen, T. Planke, M. G. Andreasen, M. Gajdacz, K. Mølmer, A. Lieberoth, and J. F. Sherson, Nature (London) **532**, 210 (2016).

[49] J. Raftery, A. Vrajitoarea, G. Zhang, Z. Leng, S. J. Srinivasan, and A. A. Houck, arXiv:1703.00942.

[50] C. Rigetti and M. Devoret, Phys. Rev. B **81**, 134507 (2010).

[51] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Phys. Rev. Lett. **103**, 110501 (2009).

[52] J. Kelly *et al.*, Nature (London) **519**, 66 (2015).

[53] 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).

[54] M. H. Goerz, F. Motzoi, K. B. Whaley, and C. P. Koch, npj Quantum Inf. **3**, 37 (2017).