RESEARCH ARTICLE | DECEMBER 12 2023

# A practical approach to determine minimal quantum gate durations using amplitude-bounded quantum controls

Stefanie Günther ; N. Anders Petersson



*AVS Quantum Sci.* 5, 043802 (2023)
https://doi.org/10.1116/5.0173373



# A practical approach to determine minimal quantum gate durations using amplitude-bounded quantum controls

Cite as: AVS Quantum Sci. **5**, 043802 (2023); doi: 10.1116/5.0173373
Submitted: 21 August 2023 · Accepted: 20 November 2023 ·
Published Online: 12 December 2023



Stefanie Günther[a] and N. Anders Petersson

## AFFILIATIONS

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

[a]Electronic mail: guenther5@llnl.gov



## ABSTRACT

We present an iterative scheme to estimate the minimal duration in which a quantum gate can be realized while satisfying hardware constraints on the control pulse amplitudes. The scheme performs a sequence of unconstrained numerical optimal control cycles that each minimize the gate fidelity for a given gate duration alongside an additional penalty term for the control pulse amplitudes. After each cycle, the gate duration is adjusted based on the inverse of the resulting maximum control pulse amplitudes by re-scaling the dynamics to a new duration where control pulses satisfy the amplitude constraints. Those scaled controls then serve as an initial guess for the next unconstrained optimal control cycle, using the adjusted gate duration. We provide multiple numerical examples that each demonstrate fast convergence of the scheme toward a gate duration that is close to the quantum speed limit, given the control pulse amplitude bound. The proposed technique is agnostic to the underlying system and control Hamiltonian models, as well as the target unitary gate operation, making the time-scaling iteration an easy to implement and practically useful scheme for reducing the durations of quantum gate operations.



## I. INTRODUCTION

Quantum optimal control utilizes numerical optimization tools to design pulses that drive quantum devices. It's potential to improve fundamental operations in quantum information science has by now been demonstrated on various applications, e.g., for quantum state preparation,[1] applications to quantum error correction,[2,3] and the realization of logical quantum gate operations,[4–6] see, for example, Refs. 7 and 8 for an overview. In the current noisy intermediate-scale quantum (NISQ) era, it is desirable to design control pulses with minimal gate duration, such that the quantum operation can be performed before decoherent processes collapse the quantum information into a classical state. However, finding the minimal duration in which such an operation is realizable is challenging. When control pulse amplitudes are unlimited, various lower bounds known as intrinsic *quantum speed limits* (QSL) have been established for the time scale over which a quantum system can evolve, see, e.g., the overview in Ref. 9. Such speed limits can be computed analytically for simple model problems[10–12] and can be estimated numerically for more complex systems by performing multiple optimal control cycles, sweeping over a range of gate durations. Indeed, it has been demonstrated on specific examples that such limits align with the shortest durations for which numerical optimal control techniques converge, when control pulse amplitudes are unlimited.[10,13,14] In practice, however, control pulses are generally subjected to hardware constraints, most commonly given in terms of a maximum drive strength as defined by hardware waveform generating devices. When hardware constraints bound the control pulse amplitude, the time-scale required to realize a quantum operation can be significantly slower than the theoretical (intrinsic) QSL. For unitary operations in the closed system setting, an extrinsic speed limit that takes the maximum control pulse amplitude into account has been derived in Ref. 15. While the lower bound matches the minimum gate duration remarkably well for small single-qubit cases, and the extension in Ref. 16 to general N-level system demonstrates qualitatively reasonable scaling properties for larger numbers of qubits, the lower bound is not sharp in the sense that the actual gate duration is no longer precisely estimated as system sizes increase. Currently, a general technique to find minimal gate durations when control pulse amplitudes are bounded by hardware constraints is not available. Instead, it is common practice to determine the minimum gate duration by trial and error through multiple optimal control cycles, each using a different pulse duration and a different initial guess for the control pulse. Henceforth, the latter approach will be referred to as the brute-force method.

In Ref. 17, a re-seeding technique was proposed that adjusts gate durations based on previous success or failure of the optimal control iterations, combined with a bisection of the resulting gate durations. However, imposing parameter bounds during the optimization updates often leads to non-convex optimization landscapes[18] and slow or stagnant optimization progress, such that optimization cycles need to be performed multiple times for various different initial guesses.

In this paper, we propose an alternative scheme to numerically find minimal time-scales in which a unitary operation can be realized under amplitude-constrained control fields by re-scaling the gate duration based on the optimized control pulse amplitude. Instead of directly imposing box constraints on the control pulse, we include a penalty term in the objective function that minimizes the control pulse energy during the optimization. After solving the unconstrained and penalized optimization problem for a given gate duration, we re-scale the dynamics to a new duration where control pulses satisfy the amplitude constraints. Those stretched (or squeezed) control pulses then serve as an initial guess for the next optimization cycle for the new gate duration. We present numerical evidence for a system of up to three coupled qudits, where this iterative scheme rapidly converges to a final gate duration close to the QSL, yielding control pulses that satisfy hardware-specific amplitude bounds while minimizing the gate time duration.

While the approach presented here is agnostic to the underlying Hamiltonian model and optimization strategy, in Sec. II, we present the system and control Hamiltonian models that are used in our numerical experiments. Section III describes the proposed iterative scheme for finding minimal gate durations and corresponding amplitude-constrained control pulses. Section IV presents the numerical results, and conclusions are drawn in Sec. V.

## II. THE QUANTUM OPTIMAL CONTROL PROBLEM

Given a unitary target gate $V^{\text{target}}$ that represents a logical quantum operation, the goal of quantum optimal control is to design optimal pulses that drive any initial quantum state $\phi(0)$ at time $t=0$ to the unitarily transformed target state $\phi(T) = V^{\text{target}}\phi(0)$ at time $T>0$. To that end, gradient-based optimization methods are typically applied to minimize a cost function that measures the distance between the target and the driven evolution in terms of the gate infidelity,

$$\min J_{cost}(U(T)) = 1 - \left| \frac{1}{N}\mathrm{Tr}\left(U^\dagger(T)V^{\text{target}}\right)\right|^2, \tag{1}$$

where the columns of $U(T)$ describe the evolved quantum state at final time $T$ for a basis of initial states $U(0) = [e_0, \ldots, e_{N-1}]$, where $e_i$ are the canonical basis vectors. We here consider closed quantum systems modeled by Schrodinger's equation, such that the evolved dynamics satisfy

$$\frac{dU(t)}{dt} = -iH(t)U(t), \quad 0 \le t \le T, \quad U(0) = I. \tag{2}$$

The Hamiltonian,

$$H(t) = H_{sys} + H_c(t), \tag{3}$$

decomposes into a time-independent system part, $H_{sys}$, and a time-varying control part, $H_c(t)$. In the laboratory-frame of reference, the control Hamiltonian is given by $\sum_q f_q(t)(a_q + a_q^\dagger)$, where the real-valued control pulse satisfies $f_q(t) = 2\,\mathrm{Re}\{c_q(t)e^{it\omega^d}\}$, where $\omega^d$ is the drive frequency. To slow down the time scales of the state dynamics and the control pulse, we apply the rotating wave approximation (RWA) and select the frequency of rotation to equal the drive frequency, $\omega^{rot} = \omega^d$. In the rotating frame, we consider a general system Hamiltonian for modeling $Q \ge 1$ coupled superconducting qudits,

$$H_{sys} = \sum_{q=1}^{Q}(\omega_q - \omega^{rot})a_q^\dagger a_q - \frac{\xi_q}{2}a_q^\dagger a_q^\dagger a_q a_q + \sum_{p>q} J_{pq}\left(a_p^\dagger a_q + a_p a_q^\dagger\right), \tag{4}$$

where $\omega_q$ denotes the 0–1 transition frequency of qudit $q$, $\xi_q$ is the self-Kerr coefficient, and $J_{pq}$ denotes the dipole–dipole coupling coefficient between qudits $p$ and $q$. Furthermore, $a_q$ ($a_q^\dagger$) denote the lowering (raising) operator for qudit $q$. The action of external control pulses on qudit $q$ in the rotating frame is given by the control Hamiltonian,

$$H_c(t) = \sum_{q=1}^{Q} c_q(t)a_q + c_q^*(t)a_q^\dagger, \tag{5}$$

where $c_q(t) = p_q(t) + iq_q(t)$ denotes the control pulse for qudit $q$ in the rotational frame, and $c_q^*(t)$ denotes its complex conjugate. For notational simplicity, we now drop the index $q$ denoting each subsystem (qudit), noting that it is straightforward to apply the proposed method to systems with multiple qudits.

To parameterize the control pulse $c(t)$ in the rotating frame, we here choose B-spline basis functions,

$$c(t) = \sum_{s=1}^{N_s} \alpha_s B_s(t), \tag{6}$$

where $\alpha_s = \alpha_s^{real} + i\alpha_s^{imag} \in \mathbb{C}$ are the control parameters that are being optimized for; $B_s(t)$ are fixed, quadratic, cardinal B-spline basis functions[19] with local support in time. In the following, we will refer to the set of control parameters $\vec{\alpha} := \{\alpha_s^{real}, \alpha_s^{imag}\}_{s=1}^{N_s}$ as the control vector. It determines the control pulse $c = c(t;\vec{\alpha})$ in the control Hamiltonian and, hence, the solution operator $U = U(t;\vec{\alpha})$ of Schrodinger's equation (2).[20] Parameterizing the control functions using quadratic B-splines provides a compact alternative to discretizing the control functions on the same time step as the underlying dynamical system. The cardinal quadratic B-spline basis functions result in control pulses that are continuously differentiable, see Appendix B. Therefore, they can readily be applied to the control hardware, without the need for smoothing or filtering a stair step control pulse or a bang–bang ansatz. While B-splines have many attractive properties,[21] we note that the time-scaling algorithm presented in Sec. III is not restricted to any particular class of pulse parameterization. The only requirement is that the parameterization must allow the pulses to readily be stretched or compressed to change the pulse duration.

## III. TIME-SCALING ITERATION TO FIND MINIMAL GATE DURATION UNDER PULSE AMPLITUDE CONSTRAINTS

When optimal controls are applied in practice, the pulses are typically subject to amplitude bounds as specified by the pulse-generating hardware, such as an arbitrary waveform generator (AWG). Typically, those constraints are given in terms of a maximum drive strength $b_{max}$ such that optimized pulses need to satisfy

$$\max_{t\in[0,T]} |c(t)| \leq b_{max}. \tag{7}$$

While the amplitude bound varies depending on the specific AWG, we here consider a maximum amplitude bound of $b_{max}/2\pi = 40$ MHz for the numerical results presented below.[22] Standard optimal control approaches incorporate such constraints into the optimization procedure by imposing box constraints on the control parameter vector, e.g., by projecting the gradient-based update steps onto the given bounds, or by using interior point or barrier methods that augment the objective function and enforce the constraint though penalty terms.[23] However, when control parameters hit the bounds, optimization convergence can deteriorate due to non-convex optimization landscapes.[24] As a result, the constrained optimization problem is often much harder to solve than the unconstrained one. In practice, it is often required to repeat multiple optimization cycles, each for a different randomized initial guesses of the control parameters and different pulse durations, to increase the chances of finding a global optima.

In this section, we propose an iterative scheme to automatically adjust the gate duration based on the energy norm of the optimized control pulse. Instead of explicitly imposing box constraints on the control parameter vector during the optimization procedure, we add a penalty term to the objective function that aims to minimize the control pulse energy. The iterative scheme performs a sequence of such unconstrained optimal control iterations, each using an updated gate duration that is selected based on the control pulse energy in the previous iteration.

In each iteration $k$ of the proposed scheme, we solve the following augmented and unconstrained optimal control problem for a given gate duration $T_k$,

$$\min_{\vec{\alpha}} \quad J_{\text{infid}}(U(T_k;\vec{\alpha})) + \gamma J_{\text{energy}}(c(\cdot;\vec{\alpha})). \tag{8}$$

Here, $\gamma > 0$ is a parameter that weights the contribution of the control energy term,

$$J_{\text{energy}}(c(\cdot;\vec{\alpha})) = \frac{1}{T_k}\int_0^{T_k} |c(t;\vec{\alpha})|^2 \, dt, \tag{9}$$

relative to the infidelity, $J_{\text{infid}}$, in the objective function.

Performing an optimization cycle that minimizes the weighted sum of the infidelity at the final time (1) and the control pulse energy (9) yields control pulses with the smallest energy norm possible, while realizing the given target unitary at time $T_k$. We can assume that a gradient-based optimal control technique can solve this unconstrained optimization problem, given that the weight $\gamma$ is chosen appropriately. However, since box constraints are not imposed during the optimization, the resulting control pulses, while being minimal in terms of their energy, might exceed the given hardware bounds $b_{max}$. Figure 1 demonstrates the relation between the pulse duration $T$ and the resulting maximum control amplitude,

$$c_{max} = \max_{t\in[0,T]} |c_*(t)|, \tag{10}$$

for optimized unconstrained control pulses $c_*(t)$ on a four-level qudit Quantum Fourier Transform (QFT) gate (problem specifications are given in Sec. IV). Here, each optimization proceeded until the gate

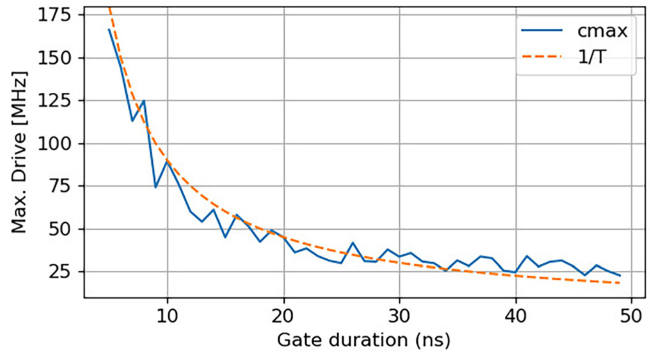


FIG. 1. Unconstrained minimization of the infidelity and the control energy (8), illustrating the relation between the maximum control amplitude and the gate duration $T$, compared with an $1/T$ scaling. Here, the target unitary is a $QFT$ gate on a four-level qudit.

infidelity dropped below the threshold $10^{-4}$. Note that the maximum control amplitude approximately scales as $1/T$.

We make use of the scaling $c_{max} \sim 1/T$ to find a minimal gate duration for the given amplitude constraints by scaling the current gate duration $T_k$ by the ratio between the current maximum control amplitude resulting from solving the unconstrained optimization problem (8) and the given hardware constraint $b_{max}$, i.e., we propose the update scheme as follows:

$$T_{k+1} = sT_k, \quad \text{where} \quad s = \frac{c_{max}}{b_{max}}. \tag{11}$$

Consider the scaled time variable

$$\tau(t) = st \quad \Rightarrow \quad \tau \in [0, T_{k+1}] \quad \text{for} \quad t \in [0, T_k], \tag{12}$$

and the corresponding scaled evolution $\tilde{U}(\tau) := U(\tau/s) = U(t)$. The scaled unitary $\tilde{U}(\tau)$ satisfies the scaled Schrodinger's equation,

$$\dot{\tilde{U}}(\tau) = \frac{dU(\tau/s)}{d\tau} = \frac{1}{s}\dot{U}(\tau/s) = -\frac{i}{s}H(\tau/s)\tilde{U}(\tau) \quad \text{for} \ \tau \in (0, T_{k+1}), \tag{13}$$

and, hence, $\tilde{U}(\tau)$ evolves under the scaled Hamiltonian

$$\tilde{H}(\tau) := \frac{1}{s}H(\tau/s) = \frac{1}{s}H_{sys} + \tilde{c}(\tau)a + \tilde{c}(\tau)^* a^\dagger, \tag{14}$$

with the control pulse

$$\tilde{c}(\tau) := \frac{1}{s}c(\tau/s). \tag{15}$$

The scaled pulse stretches (if $s > 1$) or compresses (if $s < 1$) the original pulse into the time domain $[0, T_{k+1}]$ while preserving its integral,

$$\int_0^{T_k} |c(t)| \, \mathrm{d}t = \int_0^{T_{k+1}} |\tilde{c}(\tau)| \, \mathrm{d}\tau. \tag{16}$$

Preserving the control pulse integral can be further motivated by solving the control problem analytically for a model problem, see Appendix A. By parameterizing the control functions in terms of B-spline basis functions as in (6), the re-scaled controls can be easily

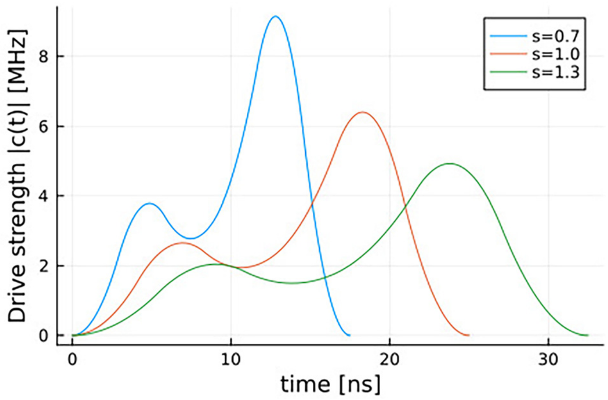


**FIG. 2.** Re-scaling control pulses: by choosing the scaling parameter $s$, the maximum drive strength of the stretched ($s > 1$) or compressed ($s < 1$) control pulse $\tilde{c}(t) = (1/s)c(t/s)$ can be adjusted.

computed by scaling the parameter vector $\tilde{\alpha} = s^{-1}\alpha$ and reevaluating the B-spline basis functions at the scaled time, thereby stretching or compressing their envelops, see Fig. 2. By design [choice of the scaling parameter $s$ in (11)], the scaled control pulses satisfy the hardware bounds exactly with

$$\max_{\tau\in[0,T_{k+1}]} |\tilde{c}(\tau)| = b_{max}. \tag{17}$$

Furthermore, if the dynamics had obeyed the scaled system Hamiltonian $(1/s)H_{sys}$, the scaled controls would have realized the target gate at time $T_{k+1}$. However, because the actual system Hamiltonian is $H_{sys}$, we only use the scaled control pulse as an initial guess to solve the optimization problem (8) for the updated gate duration, $T_{k+1}$. This procedure is repeated until the scaling factor remains constant at $s = 1$, indicating that the minimal gate duration $T_*$ has been found and the corresponding control pulse $c_*(t)$ satisfies the hardware bounds $|c_*(t)| \le b_{max}$. Note that as $s \to 1$, the scaled control pulses provide better and better initial guesses for solving the optimization problem. As a result, fewer and fewer optimization iterations are required.

In the proposed time-scaling scheme, the gate duration increases whenever the control pulse exceeds the amplitude bound and decreases whenever the pulse is below the same bound. However, many iterations of the time-scaling scheme would be needed to precisely determine the minimal gate duration such that the maximum control amplitude equals the prescribed bound. To obtain a practically useful stopping criteria, we therefore introduce a *range* of acceptable maximum control amplitudes $[b_{max} - \delta_b, b_{max}]$, for $\delta_b > 0$. The time-scaling iteration then terminates if $c_{max} \in [b_{max} - \delta_b, b_{max}]$; otherwise, the gate duration and control coefficients are re-scaled as described earlier. The resulting time-scaling iteration is summarized in Algorithm 1.

## IV. NUMERICAL RESULTS

In this section, we verify on multiple test cases that the iterative time-scaling scheme yields gate durations that are close to the smallest possible time in which the target gate can be realized when hardware constraints are in place on the control pulse amplitude. We here choose a maximum allowable amplitude bound of $b_{max}/2\pi = 40$ MHz in the rotating frame and accept the resulting control pulses (and final times) whenever the maximum control pulse amplitudes are within the range of [35, 40] MHz. The test cases are described in Table I, consisting of two single-qudit cases with three and four energy levels, one two-qubit case with two energy levels per qubit, and two three-qubit cases with two energy levels each. The qubits are coupled to their nearest neighbor(s) in a one-dimensional chain with dipole–dipole coupling strength $J = 5$ MHz. System frequencies for each test case are shown in Table II.

**ALGORITHM 1.** Minimizing the gate duration.

**Require:** Specify the range of acceptable maximum control amplitudes: $[b_{max} - \delta_b, b_{max}]$.
**Require:** Pick an initial gate duration $T_0 > 0$ and select an initial guess for the control vector $\vec{\alpha}_0$.
**for** $k = 0, 1, 2, \ldots$ **do**
  Given $\vec{\alpha}_k$, solve the optimization problem
  $\vec{\alpha}_* = \arg\min_{\vec{\alpha}} J_{\text{infid}}(U(T_k;\vec{\alpha})) + \gamma J_{\text{energy}}(c(\cdot;\vec{\alpha}))$
  such that $U(T_k;\vec{\alpha})$ solves $\dot{U}(t;\vec{\alpha}) = -iH(t;\vec{\alpha})U(t;\vec{\alpha})$ with $U(0;\vec{\alpha}) = I$.
  **if** $\max_{t\in[0,T_k]} |c(t,\vec{\alpha}_*)| \in [b_{max} - \delta_b, b_{max}]$ **then**
    Success!
    **return** $T_k, \vec{\alpha}_*$
  **else**
    Update the gate duration and the initial guess for next control vector:

$$T_{k+1} = sT_k, \quad \text{and} \quad \vec{\alpha}_{k+1} = \frac{1}{s}\vec{\alpha}_*,$$
$$\text{where} \quad s = \frac{\max_{t\in[0,T_k]} |c(t;\vec{\alpha}_*)|}{b_{max}}$$

  **end if**
**end for**

We solve the inner optimization problem in Algorithm 1 using L-BFGS as implemented in the quantum control software Quandary.[25] In order to ensure that the inner optimization cycle generates control pulses with minimal energy norm, we choose the penalty coefficient to be $\gamma = 1$. While $\gamma$ can be considered a tunable parameter in the algorithm, we motivate this value of $\gamma$ through a numerical experiment. Here, we study the effect of varying $\gamma$ on the maximum control drive amplitude and infidelity that results from minimizing the objective (8). Because the optimization minimizes the sum of two terms, it is to be expected that increasing $\gamma$ would put more emphasis on minimizing the control energy and less emphasis on the infidelity. Indeed, the results shown in Fig. 3 verify the expected behavior. In the numerical tests presented below, we aim the final infidelity to be less than $10^{-4}$. In the case tested here, $\gamma = 1$ meets the criteria for the infidelity, while also significantly reducing the maximum control drive amplitude.

In order to ensure that an optimal point with minimal energy norm has been reached, we note that it is important to base the optimization stopping criterion on the norm of the gradient, rather than on a small gate infidelity. In the following, we set the gradient threshold to be $10^{-5}$. To stabilize the inner optimization convergence, we further add a Tikhonov regularization term $\gamma_1||\vec{\alpha}||^2$, with parameter $\gamma_1 = 10^{-2}$. The purpose of the Tikhonov term is to regularize the

TABLE I. Test case specifications and target gates.

| Name | System | Target gate |
|---|---|---|
| $QFT_4$ | One qudit, 4 levels | $\frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} \in \mathbb{C}^{4\times 4}$ |
| SWAP02 | One qudit, 3 levels | $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3}$ |
| CNOT | Two qubits<br>coupling $1 \leftrightarrow 2$ at 5 MHz | $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{4\times 4}$ |
| CCNOT | Three-qubit chain<br>coupling $1 \leftrightarrow 2$ at 5 MHz<br>coupling $2 \leftrightarrow 3$ at 5 MHz | $\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & 1 \\ & & & 1 & 0 \end{pmatrix} \in \mathbb{R}^{8\times 8}$ |
| SWAP chain | Three-qubit chain<br>coupling $1 \leftrightarrow 2$ at 5 MHz<br>coupling $2 \leftrightarrow 3$ at 5 MHz | $\begin{pmatrix} 1 & & & & & & & \\ & 0 & & & 1 & & & \\ & & 1 & & & & & \\ & & & 0 & & & 1 & \\ & 1 & & & 0 & & & \\ & & & & & 1 & & \\ & & & 1 & & & 0 & \\ & & & & & & & 1 \end{pmatrix} \in \mathbb{R}^{8\times 8}$ |

optimization problem by adding a convex term that shifts the eigenvalues of the Hessian of the objective function by a small constant factor $\gamma_1$. This factor should be chosen small enough to not alter the optimization outcome.[23] We remark that, when the controls are parameterized by B-spline basis functions, the control energy term (9) can be interpreted as a weighted Tikhonov regularization term, see Appendix C. In addition, to encourage a state evolution where the populations vary slowly in time, we also penalize the second time-derivative of the state populations. To this end, we add a penalty term of the form

$$\frac{\gamma_2}{2T}\int_0^T \sum_{i,j}\left(\frac{\partial^2}{\partial^2 t}|U_{i,j}(t)|^2\right)^2 \mathrm{d}t, \tag{18}$$

where $U_{i,j}$ denotes element $(i, j)$ in the matrix $U$. Again, $\gamma_2 \geq 0$ should be chosen small enough to not dominate the infidelity and the control

TABLE II. System parameters.

| Name | $dim(H)$ | $\omega_q/2\pi$ (GHz) | $\xi_q/2\pi$ (GHz) | $J_{pq}/2\pi$ (GHz) | $\omega^{rot}/2\pi$ (GHz) | $\Delta_B$ (ns) |
|---|---|---|---|---|---|---|
| $QFT_4$ | $d = 4$ | 4.914 | 0.33 | … | 4.584 | 0.3 |
| SWAP02 | $d = 3$ | 5.12 | 0.34 | … | 4.78 | 0.3 |
| CNOT | $d = 2 \times 2$ | 5.12, 5.06 | 0.34, 0.34 | 0.005 | 5.09 | 1.65 |
| CCNOT | $d = 2 \times 2 \times 2$ | 5.18, 5.12, 5.06 | 0.34, 0.34, 0.34 | 0.005, 0.0, 0.005 | 5.12 | 1.65 |
| SWAP chain | $d = 2 \times 2 \times 2$ | 5.18, 5.12, 5.06 | 0.34, 0.34, 0.34 | 0.005, 0.0, 0.005 | 5.12 | 1.65 |

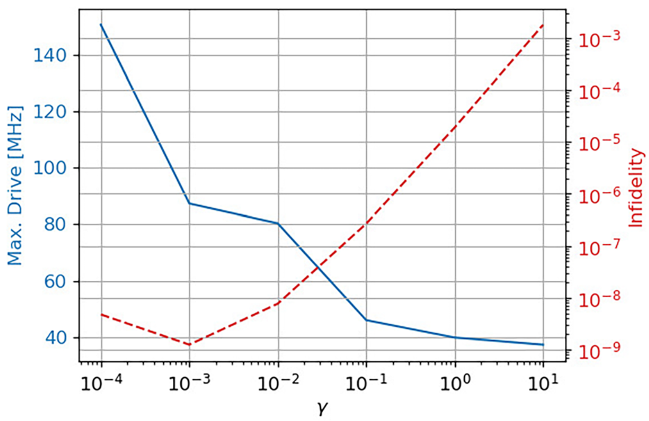


FIG. 3. Maximum control drive amplitude and infidelity when optimizing for the $QFT_4$ gate for various values of the tunable parameter $\gamma$ in the objective function (8). Here, the gate duration is fixed to $T = 20$ ns.

energy terms during the optimization. In the following, we use $\gamma_2 = 10^{-2}$. While the qualitative results of the minimal-gate-duration scheme are not expected to be sensitive to the precise value of $\gamma_2$, a thorough analysis of this term remains to be done.

During the initialization phase of Algorithm 1 ($k = 0$), we assign the elements of the initial control vector from a uniform probability distribution $\vec{\alpha}_0 \sim \mathcal{U}(-0.9b_{max}, 0.9b_{max})$.

For each test case, we deploy Algorithm 1 for a wide range of initial gate durations, $T_0$. Figure 4 shows the convergence history of the intermediate gate durations $T_k$ (left column), as well as the corresponding maximum drive amplitude of the optimized pulses (right column) at each iteration. Across all test cases, we observe that an optimal gate duration can be found within very few iterations (often only two, but never more than eight iterations), meaning that only a small number of optimal control cycles are required to determine the minimal gate duration, for a given bound on the control amplitude. This provides a clear advantage over the standard (brute-force) approach of sweeping over a wide range of gate durations, which can easily require tens to hundreds of optimal control cycles to be completed. The resulting final gate times at the last cycle of the time-scaling scheme are indicated by gray bands in Figure 4, and are also provided in Table III. The width of these bands is determined from the user-prescribed range of allowable maximum amplitudes $[b_{max} - \delta_b, b_{max}]$, which serves as the stopping criterion for the minimal time iteration. Note that we here consider $b_{\max} = 40$ MHz and $\delta_b = 5$ MHz, allowing for a 12.5% deviation from the maximum bound. The variations of final minimal times as presented in Table III are of the same order. Importantly, we observe that the iterations converge to this band for a wide range of initial gate durations, $T_0$.

TABLE III. Minimal gate durations with gate fidelity $\geq 99.9\%$, as achieved from the time-scaling scheme in Algorithm 1 (Fig. 4), and brute-force optimization with box constraints on the control vector (Fig. 5).

| Case | Algorithm 1 (ns) | Brute force (ns) | Smallest rel. diff. (%) |
|---|---|---|---|
| $QFT_4$ | 19 to 23 | 18 to 21 | 5 |
| SWAP02 | 18 to 23 | 18 | 0 |
| CNOT | 68 to 78 | 70 | −2 |
| CCNOT | 200 to 225 | 190 | 5 |
| SWAP chain | 205 to 239 | 190 | 7 |

To evaluate how well the iterative time-scaling scheme recovers the actual minimal gate duration for a given amplitude bound, we compare the resulting gate durations with a brute-force approach that performs a sweep of optimizations for a wide range of (fixed) gate durations, while enforcing box constraints on the control vector using a projected L-BFGS method. Figure 5 shows statistics of optimized gate fidelities, each gathered over ten different optimizations starting from random initial control vectors. Here, variations in the resulting gate fidelities can be attributed to deterioration of the optimization progress, related to the imposed box constraints. Nevertheless, we compare the resulting minimal gate durations as observed in Fig. 5 with those produced by Algorithm 1 as in Fig. 4. Table III summarizes the resulting gate durations, demonstrating that the time-scaling scheme converges to final gate durations that are remarkably close to those obtained with the brute force method. The range of resulting gate durations as produced by Algorithm 1 is a direct consequence of the stopping criteria based on a prescribed range of allowed amplitudes $[b_{max} - \delta_b, b_{max}]$. The range in gate duration hence depends on the user-controlled parameter $\delta_b$. In the last column, we compare the shortest gate durations with the ones obtained from the brute-force optimization approach. We observe a relative difference below 7% for all test cases.

## V. CONCLUSION

This paper presents a practical scheme for approximating the minimal duration for realizing a unitary gate while satisfying given amplitude bounds on the control pulses. The scheme performs a sequence of unconstrained optimal control cycles, each minimizing the gate fidelity alongside an additional penalty term for the control pulse amplitudes. After each cycle, the gate duration is scaled to a new value for which the scaled controls satisfy the amplitude bounds. We provide numerical evidence of convergence to a final gate duration that is close to the shortest achievable duration obtained from sweeping over a large number of gate durations. Updating the gate duration based on the previous control amplitudes converges to the minimal gate duration in a few optimization cycles, drastically reducing the computational costs for practical usage of time-optimal quantum control, which often requires tens to hundreds of optimization cycles to sweep over a range of gate durations in a trial and error fashion. We demonstrate that the proposed scheme converges for a wide range of initial gate durations, indicating that almost no prior knowledge of the minimal gate duration is required. The proposed technique is straightforward to be implemented on top of an existing quantum optimal control code, where we note that it is beneficial to parameterize the control pulses by basis functions such that the pulse amplitudes can easily be adjusted. Furthermore, the proposed method is agnostic to the underlying system and control Hamiltonian models, as well as the target unitary gate operation, making the time-scaling iteration an easy to implement and practically useful scheme for reducing the durations of quantum gate operations.

As illustrated in Fig. 1, the control pulse amplitude can be a non-monotonic function of the gate duration. We conjecture that this

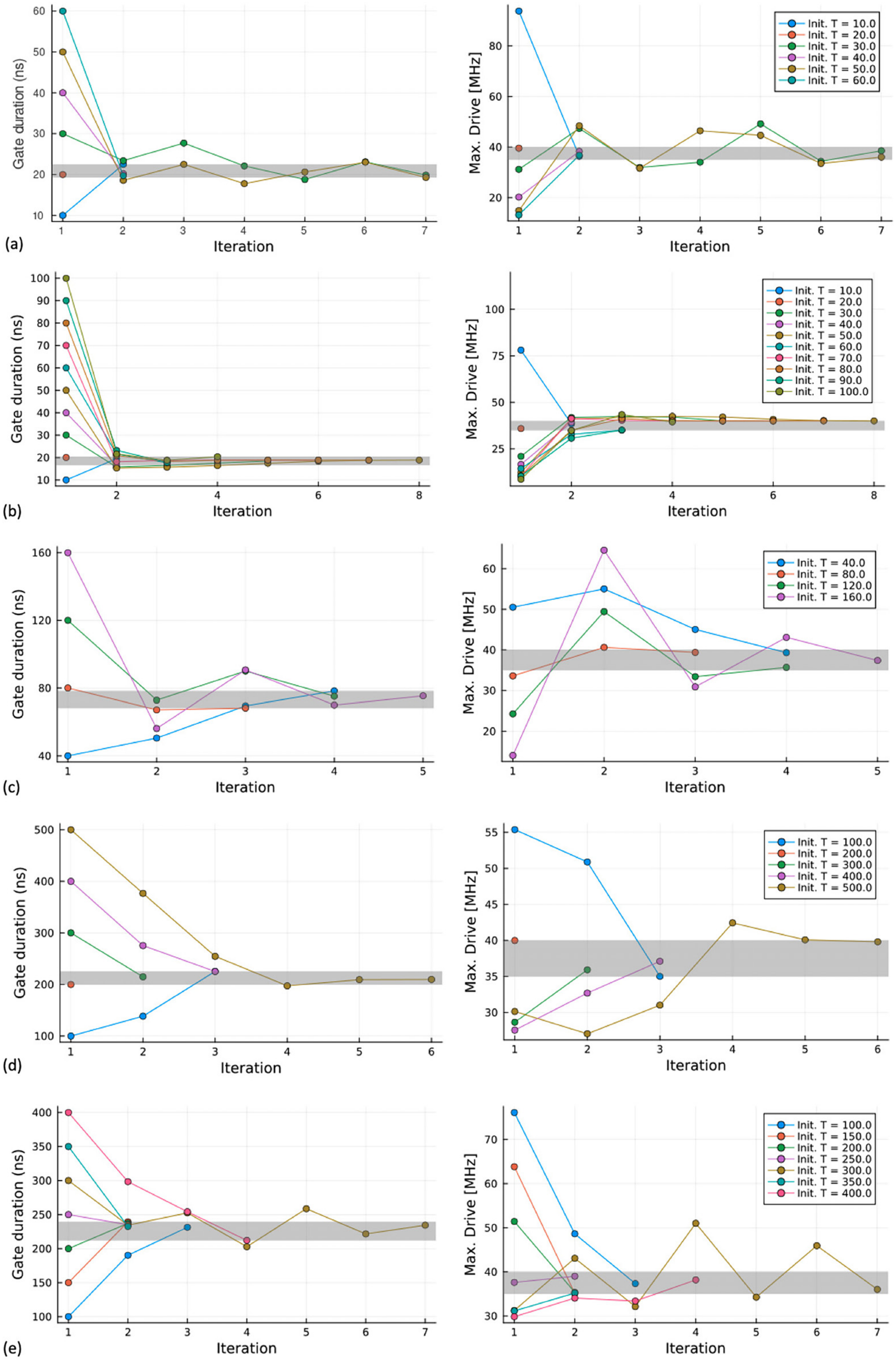


FIG. 4. Optimization history for the time-scaling algorithm (Alg. 1): gate durations (left column) and maximum amplitudes (right column). The gray areas denote the resulting band of final gate times and the prescribed range of acceptable amplitudes, respectively. (a) $QFT_4$ gate on a single qudit. (b) SWAP02 gate on a single qudit. (c) CNOT gate, coupling $J_{12} = 5\,\text{MHz}$. (d) CCNOT (Toffoli) gate on a three-qubit chain. (e) SWAP qubit 1 and 3 on a three-qubit chain.

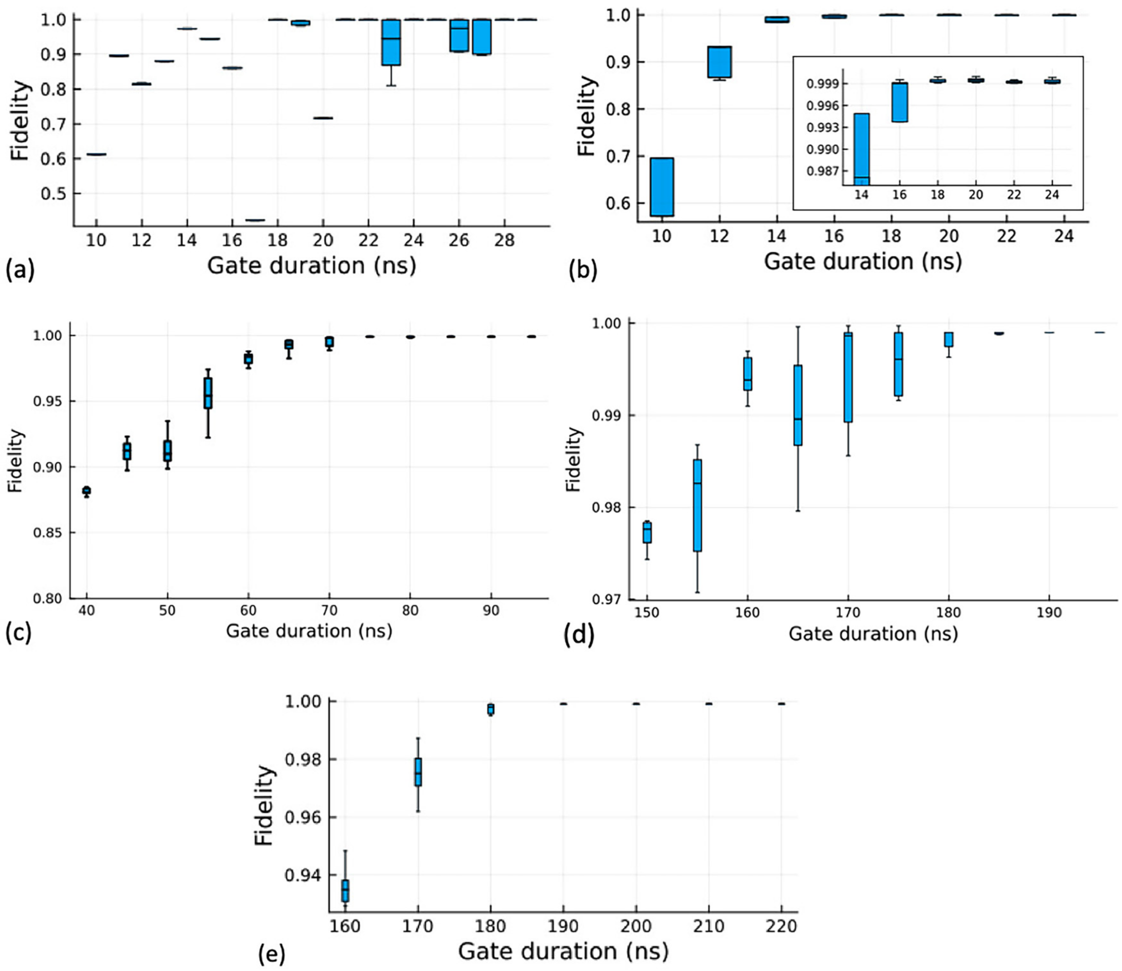


Fig. 5. Estimating the quantum speed limit with the brute-force method: optimized fidelities sweeping over various gate durations, while enforcing hardware constraints directly during the optimization process. For each gate duration, statistics are gathered by performing ten optimization cycles starting from randomly sampled initial control vectors. (a) $QFT_4$ gate on single qudit. (b) SWAP02 gate on single qudit. (c) CNOT gate, coupling $J_{12} = 5$ MHz. (d) Toffoli gate on a three-qubits chain. (e) SWAP qubit 1 and 3 on a three-qubit chain.

phenomena is the underlying cause of the non-monotonic convergence (over- and under-shooting the gate duration) of our algorithm, and we conjecture that convergence may be further improved by damping the updates of the duration. This will be subject of further investigations.

## ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Jonathan Dubois at Lawrence Livermore National Laboratory, CA, who provided expertise and partial funding resources during the preparation of this paper.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344 (LLNL-JRNL-851967).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Stefanie Guenther:** Methodology (equal); Software (lead); Visualization (equal); Writing – original draft (lead); Writing – review & editing (equal). **N. Anders Petersson:** Methodology (equal); Visualization (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## APPENDIX A: THE INFIDELITY IS INVARIANT TO THE CONTROL PULSE INTEGRAL IN A QUBIT MODEL PROBLEM

To gain further insight into the essential properties of a control pulse, we consider the model problem of a driven two-level system in the rotating frame,

$$H(t) = c(t)a^\dagger + c^*(t)a, \quad c(t): \mathbb{R} \mapsto \mathbb{C}. \tag{A1}$$

We consider the displacement transformation,[26]

$$U(t) = D(\beta(t))\tilde{U}(t), \quad \beta(t): \mathbb{R} \mapsto \mathbb{C}, \tag{A2}$$

where the unitary transformation operator is

$$D(\beta) = \exp(\beta a^\dagger - \beta^* a), \quad D(0) = I. \tag{A3}$$

Thus, by taking $\beta(0) = 0$, we have $U(0) = \tilde{U}(0)$, and the transformed unitary solution operator satisfies

$$\dot{\tilde{U}}(t) = -i\tilde{H}(t)\tilde{U}(t), \quad \tilde{U}(0) = I, \tag{A4}$$

$$\tilde{H}(t) = \left(c(t) - i\frac{d\beta(t)}{dt}\right)a^\dagger + \left(c^*(t) + i\frac{d\beta^*(t)}{dt}\right)a, \tag{A5}$$

where we have omitted terms of the form $\delta(t)I$ in $\tilde{H}(t)$ because they only change the global phase in $\tilde{U}$. We conclude that $\tilde{U}(t) = I$ for all $t \geq 0$ if $\beta(t)$ satisfies the ordinary differential equation

$$\frac{d\beta(t)}{dt} = -ic(t), \quad \beta(0) = 0, \iff \beta(t) = -i\int_0^t c(\tau)\, d\tau. \tag{A6}$$

In this case, we have $U(T) = D(\beta(T))\tilde{U}(T) = D(\beta(T))$, and the infidelity at time $t = T$ satisfies

$$J_{\rm infid} = 1 - \left|\frac{1}{N}\mathrm{Tr}\left(V_{\rm target}^\dagger D(\beta(T))\right)\right|^2. \tag{A7}$$

For a given duration $T$, we note that the infidelity only depends on the integral $\beta(T) = -i\int_0^T c(\tau)\, d\tau$. As a result, any control function $\tilde{c}(t)$ with the same integral over time yields the same infidelity.

If we scale time, $t' = st$, and define $\beta'(t') = \beta(t)$, the infidelity with respect to $V_{\rm target}$ becomes invariant to the scaling as long as $\beta'(T') = \beta(T)$. The scaling implies

$$\beta'(T') = -i\int_0^{T'} c'(\tau')\, d\tau' = -i\int_0^T c'(s\tau)s\, d\tau. \tag{A8}$$

Hence, $\beta'(sT) = \beta(T)$ if the scaled control pulse satisfies

$$c'(st) = \frac{c(t)}{s}. \tag{A9}$$

For example, if $s > 1$, the duration of the control pulse is increased, and its amplitude is decreased, such that the time integral of the control pulse remains unchanged.

## APPENDIX B: CONTROL PARAMETERIZATION VIA B-SPLINE BASIS FUNCTIONS

The control function $c(t)$ is parameterized in terms of a weighted sum of B-spline basis functions according to Eq. (6). It is convenient to split the complex-valued function into its real and imaginary parts, $c(t) = p(t) + iq(t)$, where

$$p(t) = \sum_{s=1}^{N_s} \alpha_s^{real} B_s(t), \quad q(t) = \sum_{s=1}^{N_s} \alpha_s^{imag} B_s(t). \tag{B1}$$

Here, $N_s \geq 1$ equals the number of terms in $p(t)$ and $q(t)$, specified by the real-valued design variables $\{\alpha_s^{real}, \alpha_s^{imag}\}_{s=1}^{N_s}$. In this work, we let $B_s(t)$ be quadratic B-spline basis functions centered on a uniform grid in time: $t_s = (s + 0.5)\Delta_B$, with knot spacing $\Delta_B = T/(N_s + 2)$. Each basis function follows by shifting a normalized wavelet function in time, $B_s(t) = \tilde{B}(\tau_s(t))$, where $\tau_s(t) = (t - t_s)/3\Delta_B$ and

$$\tilde{B}(\tau) = \begin{cases} \frac{9}{8} + \frac{9}{2}\tau + \frac{9}{2}\tau^2, & -\frac{1}{2} \leq \tau < -\frac{1}{6}, \\ \frac{3}{4} - 9\tau^2, & -\frac{1}{6} \leq \tau < \frac{1}{6}, \\ \frac{9}{8} - \frac{9}{2}\tau + \frac{9}{2}\tau^2, & \frac{1}{6} \leq \tau < \frac{1}{2}, \\ 0, & \text{otherwise}. \end{cases} \tag{B2}$$

Note that $B_s(t)$ has compact support in $t \in [t_s - 1.5\Delta_B, t_s + 1.5\Delta_B]$. This implies that, at any time $t \in [0, T]$, at most, three terms are non-zero in $p(t)$ and $q(t)$. Since the first and last basis functions are centered at $t_1 = 3\Delta_B/2$ and $t_{N_s} = T - 3\Delta_B/2$, respectively, the control function always starts and ends with zero amplitude.

We note that the quadratic B-spline basis functions used here are continuously differentiable. Control pulses that are only piece-wise continuous, such as those generated by a bang–bang ansatz, can only be *approximated* by a quadratic (second order) B-spline function. If piece-wise constant control pulses are desirable, they could instead be parameterized by zeroth order B-spline functions, where the basis consists of box-car functions. For a general introduction to B-spline basis functions, see Ref. 19. We remark that the minimal time algorithm as presented in Sec. III is agnostic to the underlying parameterization, as long as the generated pulses can be stretched or compressed to a different pulse length.

## APPENDIX C: ENERGY PENALTY TERM AS A REGULARIZER

The integral in the energy penalty term, defined by (9), can be written as

$$\mathcal{J}_{energy} = \frac{1}{T}\int_0^T \left(p(t)^2 + q(t)^2\right) dt. \tag{C1}$$

We here analyze the regularizing effect this term has on the optimization problem, when the control pulses are parameterized by quadratic B-splines, as defined in Appendix B. Because each B-spline basis function is a piece-wise quadratic function of time, it is straightforward to analytically evaluate the integral in the above formula. After some algebra,

$$\mathcal{J}_{energy} = \frac{3\delta}{T}\sum_{j=1}^{N_s}\sum_{k=1}^{N_s}\left(\alpha_j^r W_{j,k}\alpha_k^r + \alpha_j^i W_{j,k}\alpha_k^i\right), \tag{C2}$$

where $W$ is a real, symmetric, penta-diagonal $N_s \times N_s$ matrix,

$$W = \frac{11}{60}\begin{bmatrix} a & b & c & & & & & \\ \ddots & \ddots & \ddots & \ddots & & & & \\ \ddots & \ddots & \ddots & \ddots & \ddots & & & \\ & c & b & a & b & c & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \ddots & \ddots & \\ & & & & c & b & a & \end{bmatrix}, \tag{C3}$$

and the non-zero elements are $a = 1$, $b = (26/66)$, and $c = (1/66)$. The matrix is strictly diagonally dominant for all interior rows, because $a - 2(b + c) = (10/66) > 0$; the diagonal dominance is even greater for the first and last two rows. Thus, the matrix is positive definite, implying that the energy penalty term also serves to regularize the optimization problem, because (C2) represents a weighted norm of the control parameter vector.

## REFERENCES


[1]K. Rojan, D. M. Reich, I. Dotsenko, J.-M. Raimond, C. P. Koch, and G. Morigi, Phys. Rev. A **90**, 023824 (2014).
[2]G. Waldherr, Y. Wang, S. Zaiser, M. Jamali, T. Schulte-Herbrüggen, H. Abe, T. Ohshima, J. Isoya, J. Du *et al.*, Nature **506**, 204 (2014).
[3]F. Gaitan, *Quantum Error Correction and Fault Tolerant Quantum Computing* (CRC Press, 2008).
[4]S. Schirmer, J. Mod. Opt. **56**, 831 (2009).
[5]J. P. Palao and R. Kosloff, Phys. Rev. Lett. **89**, 188301 (2002).
[6]Y. Cho, K. M. Beck, A. R. Castelli, K. A. Wendt, B. Evert, M. J. Reagor, and J. L. DuBois, "Direct pulse-level compilation of arbitrary quantum logic gates on superconducting qutrits," arXiv:2303.04261 (2023).
[7]S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer *et al.*, Eur. Phys. J. D **69**, 279 (2015).
[8]C. P. Koch, J. Phys.: Condens. Matter **28**, 213001 (2016).
[9]S. Deffner and S. Campbell, J. Phys. A **50**, 453001 (2017).
[10]G. C. Hegerfeldt, Phys. Rev. Lett. **111**, 260501 (2013).
[11]N. Khaneja, R. Brockett, and S. J. Glaser, Phys. Rev. A **63**, 032308 (2001).
[12]N. Khaneja, S. J. Glaser, and R. Brockett, Phys. Rev. A **65**, 032301 (2002).
[13]T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Phys. Rev. Lett. **103**, 240501 (2009).
[14]S. Ashhab, P. De Groot, and F. Nori, Phys. Rev. A **85**, 052327 (2012).
[15]C. Arenz, B. Russell, D. Burgarth, and H. Rabitz, New J. Phys. **19**, 103015 (2017).
[16]J. Lee, C. Arenz, H. Rabitz, and B. Russell, New J. Phys. **20**, 063002 (2018).
[17]L. M. Seifert, J. Chadwick, A. Litteken, F. T. Chong, and J. M. Baker, "Time-efficient qudit gates through incremental pulse re-seeding," in *IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2022), pp. 304–313.
[18]X. Ge, R.-B. Wu, and H. Rabitz, Annu. Rev. Control **54**, 314 (2022).
[19]L. Piegl and W. Tiller, *The NURBS Book*, 2nd ed. (Springer-Verlag, New York, 1996).
[20]In the following, the dependence on $\vec{\alpha}$ will be suppressed for improved readability.
[21]M. Unser, "Ten good reasons for using spline wavelets," Proc. SPIE **3169**, 422–431 (1997).
[22]Because we have scaled Schrodinger's equation to make $\hbar = 1$, the unit of energy becomes angular frequency. As a result, the unit of the control pulse amplitude $|c(t)|$ is also angular frequency.
[23]J. Nocedal and S. Wright, *Numerical Optimization* (Springer Science & Business Media, 2006).
[24]B. Russell, H. Rabitz, and R. Wu, "Quantum control landscapes are almost always trap free," arXiv:1608.06198 (2016).
[25]S. Günther, N. A. Petersson, and J. L. DuBois, "Quandary: An open-source C++ package for high-performance optimal control of open quantum systems," in *IEEE/ACM Second International Workshop on Quantum Computing Software (QCS)* (IEEE, 2021), pp. 88–98.
[26]C. Gerry and P. Knight, *Introductory Quantum Optics* (Cambridge University Press, 2004).