Conservation-law-based global bounds to quantum optimal control

Hanwen Zhang,1, 2 Zeyu Kuang, 1, 2 Shruti Puri, 1, 3 and Owen D. Miller1, 2

1Department of Applied Physics, Yale University, New Haven, Connecticut 06511, USA
2Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA
3Yale Quantum Institute, Yale University, New Haven, Connecticut 06511, USA
(Dated: May 14, 2021)

Active control of quantum systems enables diverse applications ranging from quantum computation to manipulation of molecular processes. Maximum speeds and related bounds have been identified from uncertainty principles and related inequalities, but such bounds utilize only coarse system information, and loosen significantly in the presence of constraints and complex interaction dynamics. We show that an integral-equation-based formulation of conservation laws in quantum dynamics leads to a systematic framework for identifying fundamental limits to any quantum control scenario. We demonstrate the utility of our bounds in three scenarios—three-level driving, decoherence suppression, and maximum-fidelity gate implementations—and show that in each case our bounds are tight or nearly so. Global bounds complement local-optimization-based designs, illuminating performance levels that may be possible as well as those that cannot be surpassed.

In this Letter, we develop a framework for computing fundamental limits to what is possible via control of quantum systems. We derive generalized probability conservation laws that must be satisfied by all quantum-mechanical wave functions, adapting a mathematical approach recently developed for light–matter interactions [1, 2], and show that they have unique properties amenable to global bounds via semidefinite programming [3, 4]. We demonstrate the power and utility of our method on three prototypical systems: (1) three-level system driving, where our bounds incorporate sophisticated information about interference between the levels, and can account for constraints on undesirable transitions (as needed in transmons [5], for example), (2) upper bounds to the extent to which decoherence can be suppressed, and (3) the maximum fidelity of a controller-based implementation of a single-qubit Hadamard gate. In each case we supplement our bounds with many local-optimization-based solutions, showing that they come quite close to (and in some cases achieve) our bounds, suggesting that our bounds are tight or nearly so. Our framework applies to open and closed systems, can be extended to related domains in NMR [6–8] and quantum complexity [9–11], and should reveal the limits of what is possible with quantum control.

Quantum control [12–16] refers to the design and synthesis of efficient control sequences that drive a quantum system to maximize a desired objective, such as maximizing overlap with a target state or minimizing error in the implementation of a gate operation. Recent experiments have demonstrated the power of optimal control for wide-ranging applications [17–22]. Because the wave function \( \psi(t) \) that represents a quantum state is nonlinear in the control parameter \( \epsilon(t) \), it is generically difficult to identify globally optimal controls. One strategy is to use local numerical optimization over the control parameters (e.g. GRAPE [6, 23–25], the Krotov method [26–31], and CRAB [32, 33]), optimizing over many initial conditions in the hopes of identifying high-performance local optima. Yet, except in the simplest of systems, one is left uncertain about the best performance possible. Alternatively, there are a variety of global bounds [34–55]; most famously, the Mandelstam–Tamm (MT) bound. The MT bound is a prototype of "quantum speed limits," which more generally have varying levels of complexity but are essentially time-energy uncertainty relations [34–37, 40, 44, 45, 52–55]. The energy measure is typically a matrix norm of the Hamiltonian, but more complex details of the system interactions are not captured. Another class of bounds is obtained by analytically solving Pontryagin’s maximum principle [56], which is only possible in simple cases such as two-level systems [38, 39, 42, 43, 46–48]. Consequently, meaningful, accurate bounds cannot be computed for most quantum control systems of interest.

Formulation—We consider a Hamiltonian of the form \( H_0(t) + H_c(t) = H_0(t) + \epsilon(t)H_c(t) \), where \( H_0 \) is the non-controllable part of the Hamiltonian, \( H_c \) is the controllable part, and \( \epsilon \) is the control parameter to be optimized. We assume the control parameter is bounded between 0 and \( \epsilon_{\max} \) (any other minimum value can be shifted to 0 by replacing \( H_0 \) with \( H_0 + \epsilon_{\min}H_c \)). Our method generalizes to any number of control parameters (cf. SM), but for simplicity we assume one throughout this paper. Any smooth, continuous, bounded control can be approximated with arbitrary accuracy by a “bang–bang” binary control that only takes the values 0 and \( \epsilon_{\max} \) (cf. SM), so we use bang–bang controls in our formulation. Instead of the differential Schrödinger equation for the time-evolution operator \( U(t, t_0) \) (for an initial time \( t_0 \)), we instead start with an integral form (equivalent to the Dyson equation [57, 58] in the interaction picture):

\[
U(t, t_0) = U_0(t, t_0) - \frac{i}{\hbar} \int_{t_0}^{T} G_0^+(t, t')H_c(t')U(t', t_0) \, dt,
\]  

where \( U_0 \) and \( G_0^+ \) are the time-evolution operator and retarded Green's function in the absence of controls (i.e., for \( H_0(t) \)), and \( T \) is the final time. To derive conserva-
tion laws, we start by taking the product of Eq. (1) with $U^{\dagger}(t,t_0)H'_c(t)D_i(t)$ from the left and integrating from an initial time $t_0$ to $T$:

$$\int_{t_0}^{T} U^{\dagger}(t,t_0)H'_c(t)D_i(t)U(t,t_0) \, dt$$

$$+ \frac{i}{\hbar} \int_{t_0}^{T} \int_{t_0}^{T} U^{\dagger}(t,t_0)H'_c(t)D_i(t)G_{0,0}^{\dagger}(t,t')H'_c(t')U(t',t_0) \, dt \, dt'$$

$$= \int_{t_0}^{T} U^{\dagger}(t,t_0)H'_c(t)D_i(t)U_0(t,t_0) \, dt.$$

(2)

The variable $D_i(t)$ can be any time-dependent operator and is an optimization hyperparameter below in Eq. (5); intuitively, allowing different possible choices of $D_i$ enables the isolation of particular times and elements in Hilbert space for which Eq. (2) should be satisfied. The constraint of Eq. (2) depends on both the time-evolution degrees of freedom $U(t)$ and the control variable degrees of freedom $\varepsilon(t)$. However, if we define a new variable $\Phi(t) = \varepsilon(t)H_c(t)U(t,t_0)$, this variable (the analog of a polarization field in electrodynamics [1,59]) can subsume both, leading to the constraints:

$$\int_{t=t_0}^{T} \int_{t'=t_0}^{T} \Phi^{\dagger}(t)D_i(t) \left[ \frac{H^{-1}_c(t)}{\varepsilon_{\max}} \delta(t-t') + i \frac{\hbar}{G_{0,0}^{\dagger}(t,t')} \right] \Phi(t')$$

$$= \int_{t_0}^{T} \Phi^{\dagger}(t)D_i(t)U_0(t,t_0) \, dt,$$

(3)

where $H^{-1}_c$ is interpreted as the pseudo inverse if $H_c$ is not invertible. For any $D_i(t)$, Eq. (3) is a quadratic equation in the variable $\Phi(t)$; the set of all possible $D_i(t)$ imply an infinite number of quadratic constraints, which can ultimately be relaxed to a semidefinite program.

Equation (3) can be interpreted as a generalization of probability conservation. At any time $t_1$, conservation of probability implies unitarity of the time-evolution operator $U(t_1,t_0)$, such that $U^{\dagger}U = I$, where $I$ is the identity operator. From the integral equation for $U$, Eq. (1), the difference $U^{\dagger}U - I$ can be written

$$U^{\dagger}(t_1,t_0)U(t_1,t_0) - I$$

$$= \frac{1}{\hbar^2} \int_{t_0}^{t_1} \int_{t_0}^{t_1} \Phi^{\dagger}(t'',t_0)U_0(t'',t')\Phi(t',t_0) \, dt'' \, dt'$$

$$+ \frac{2}{\hbar} \text{Im} \int_{t_0}^{t_1} U_0(t',t_0)\Phi(t',t_0) \, dt'.$$

(4)

If we take the imaginary part of Eq. (3), and choose $D_i(t)$ to be the identity operator from $t_0$ to $t_1$ (and zero otherwise), the resulting constraint is precisely the one that requires the right-hand side of Eq. (4) to be zero (cf. SM). In other words, a subset of the constraints of Eq. (3) are those which enforce unitary evolution at all times. (In an open system described by a density matrix, unitarity is not preserved and the corresponding constraints instead represent conservation of probability flow, cf. SM.)

We consider all objectives $f$ that are linear or quadratic functions of the time-evolution operator $U$, and therefore linear or quadratic functions of $\Phi = \varepsilon H_c U$. Because the constraints of Eq. (3) are necessary but not necessarily sufficient conditions for describing the quantum evolution, replacing the differential or integral dynamical equations with those constraints leads to a bound on the maximum value (or feasibility) of $f(\Phi)$, as the solution of:

$$\max_{\Phi} f(\Phi)$$

s.t. Equation (3) satisfied for all $D_i(t)$

(5)

where we assume the problem has been discretized in any standard basis [60]. The tradeoff in relaxing the dynamical equation to these constraints, however, is that it is now possible to compute global bounds. Equation (5) is a quadratically constrained quadratic program (QCQP), whose bounds can be computed via standard semidefinite programming (SDP) techniques, such as interior point methods [61, 62]. As the bounds are computed over all possible matrices $D_i$ at all (discrete) times, we label them “D-matrix bounds.” (It is likely possible through clever algorithmic design to solve Eq. (5) over a small subset of the D matrices [1], which is not necessary in the cases below but would generally reduce computational costs, as the total number of possible constraints increases exponentially with the size of the Hilbert space.) This framework applies broadly across quantum control; next, we demonstrate bounds for three prototypical systems.

Applications—First, we compute bounds on driving three-level quantum systems. We consider two three-level systems described by Hamiltonians $H = \hbar \sum_{j=1,2} \omega_j |i\rangle\langle i| - \varepsilon(t) \sum_{i,j=0,1,2} \mu_{ij} |i\rangle\langle j|$: one modeling an asymmetric double-well potential, with exact parameters from Sec. 2.8 of Ref. [13] and given in the SM, and a second modeling a weakly nonlinear harmonic oscillator with nearest-level couplings, as is typically used to model a transmon qubit [5, 63]. (We consider both systems as they have different features: the first, couplings between all levels, and the second, small anharmonicity with hard-to-avoid leakage.) In each case we assume the system starts in the ground state, $|0\rangle$, and that we want to drive it to the first excited state, $|1\rangle$, as rapidly as possible. We denote the probability of occupying state $i$ at time $t$ by $P_i(t) = |\langle i|\psi(t)\rangle|^2$. There are two classes of bounds that we can compute: for a given amount of time $T$, the maximum probability in $|1\rangle$, $P_1(t)$; or, iteratively, the minimum amount of time to achieve near-unity probability in $|1\rangle$.

The black curve of Fig. 1(a) is the computed bound on $P_1(t)$ for the asymmetric-double-well model, for a bounded control field with $|\varepsilon(t)| \leq 0.15$. The shaded region of the figure is impossible to reach: our bounds indicate that any such evolution would necessarily violate at least one of the conservation laws. The grey lines are the results of local computational optimizations; we implemented a gradient-ascent optimization (similar to...
GRAPE) as described in the SM, for many different final times and initial pulse sequences. Also included in the figure are data points corresponding to evaluations of other bounds as applied to this problem: Mandelstam–Tamm (MT), Margolus–Levitin (ML), and Refs. [54, 55]. It takes some effort to map the various bounds to this problem, with varying degrees of looseness, which we discuss in detail in the SM. In particular, however, one can see that each of these bounds predicts minimal times an order of magnitude smaller than our approach. The inset provides a likely explanation: the optimal trajectory (highlighted in red) first populates the second excited state, then transitions to the first excited state through appropriate driving. Such complex dynamics cannot be captured by any previous bound approaches, but can be captured by our approach.

Parts (b–d) of Fig. 1 show results for the transmon-qubit model, with \( \omega_1 = 0.19, \omega_2 = 0.37, \mu_{10} = \mu_{01} = -1, \mu_{21} = \mu_{12} = -\sqrt{2} \) (all other \( \mu_{ij} = 0 \)), and \( |\epsilon(t)| \leq 0.3 \). Fig. 1(b,c) are the transmon analogs of Fig. 1(a). The key novelty that is possible in this case is the addition of a constraint on the excitation probability of the second excited state, [2]. Such “leakage” can be highly detrimental to the practical control of such systems, as they can open up additional decoherence channels [64]. In our approach, we can simply add to Eq. (5) a (quadratic) constraint on the maximum allowed probability in [2]. In Fig. 1(d), we show the bound for the maximum \( P_1(t) \) subject to varying constraints on the maximum allowed \( P_2(t) \), at time \( t = 5 \), which shows the dramatic reduction that is required if state-[2] transitions are to be avoided. Conversely, also in Fig. 1(d), the minimum time for near-unity first-state probability increases dramatically with more stringent constraints (red). Such constraints could not be incorporated into previous bound approaches.

A second example we consider is the extent to which one can prevent decoherence and dissipation due to interactions with the environment. The design of pulses to achieve such a goal has been studied extensively through semi-heuristic “dynamic decoupling” design schemes [65–68], which may not be (and in many cases are not) globally optimal. A typical model of environmental effects is a spin system interacting with a spin bath. We consider a spin-bath system [69] with Hamiltonian \( H_0 = H_S + H_E + H_{\text{int}} \), where \( H_S \) is the system Hamiltonian (two levels split by energy \( \hbar \omega_0 \)), \( H_E \) is the Hamiltonian of the environmental bath \( (H_E = -J \sum_{j=1}^{N} (\sigma_j^x \sigma_{j+1}^x + \lambda \sigma_j^z)) \), and \( H_{\text{int}} \) is the interaction between the system and the bath, \( H_{\text{int}} = -\nu |\downarrow \rangle \langle \downarrow | \otimes \sum_j \sigma_j^z \), with \( \omega_0 = \pi, J = 1, \lambda = 0.5 \), and \( \nu = 2 \) here. The control Hamiltonian here is \( H_c = \epsilon(t) \sigma_z \) on the system only. Rather than use an approximation to the environmental coupling [70], we model the full dynamics of the wave function \( |\psi(t)\rangle \). As a result, we only use a bath of size \( N = 2 \). Despite the bath being unrealistically small, it provides a qualitatively accurate description of the decoherence process [71] and serves as a proof of principle. The system initial state is \( |\uparrow \rangle + |\downarrow \rangle \), while the spin bath is in its ground state. The system density matrix \( \rho^S \) is found by tracing out the bath part of the full density matrix, \( \rho(t) = |\psi(t)\rangle \langle \psi(t)| \). The objective is to maximize \( |\rho^S_{12}| \), the magnitude of the off diagonal elements of \( \rho^S \), which represents the coherence of the system state. Instead of working with the
absolute value (or its square, which is quartic in $|\psi\rangle$), we equivalently maximize $f = \text{Re} \left( \rho_{12}^S e^{i\phi} \right)$ for a given $\phi$, and then iterate over possible values of $\phi$ between 0 and $2\pi$. Fig. 2 shows the bounds on maximal coherence as a function of time for three different bounded controls: $\varepsilon_{\text{max}} = 0.5, 1$ and 2. Also included are actual evolutions for three cases: without control, with a pulse designed by gradient ascent, and pulses designed by a bounded-control version of dynamical decoupling termed “Eulerian Carr-Purcell” [72]. It is possible with strong controls to increase coherence at short times (as is particularly visible in Fig. 2(c)), but that would not be possible over longer time scales. We see that the bounds appear nearly tight, and provide information about what levels of coherence are possible as a function of time.

For the third application, we consider the implementation of a single-qubit Hadamard gate. For a two-level system with Hamiltonian $H = \hbar \omega_0 \sigma_z - \mu \varepsilon(t) \sigma_x$ ($\omega_0 = 0.0784, \mu = 1$) [13], the target time-evolution operator is given by $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. The objective is to compute the maximal fidelity of a quantum gate at time $T$; for computational purposes, it is easier to work with the square of fidelity, $f^2 = \frac{1}{2} \text{Tr} \left\{ U_{12}^\dagger U(T) \right\}^2$. Identifying when the bound approaches 1 then indicates the minimum possible time to perform a gate operation. We consider a bounded control with $\varepsilon_{\text{max}} = 1$. A crucial difference in the gate problem is that multiple inputs map to multiple outputs; the off-diagonal elements of the $D$ matrices in Eq. (3) inherently enforce the corresponding orthogonal-evolution requirement. Fig. 3 shows the fidelity bound as a function of time (solid black), along with time evolutions for locally optimized pulse sequences in the colored lines (optimized for different end times). The bound is tight, or very nearly so, across all times.

Conclusions—Quadratic constraints representing generalized probability-conservation laws offer a framework for quantum control bounds. We have shown that this method can be significantly tighter than previous bounds and more widely applicable. There are further extensions that may be possible as well: in nanophotonic design problems, a hierarchy of bounds with varying analytical and semi-analytical complexity have been discovered as subsets of the $D$-matrix constraints [1, 2, 73–84]; the same may be possible in quantum control. In particular, environment-induced decoherence and dissipation are similar to material-absorption losses in electromagnetism, and may be amenable to general analytical bounds [73, 77]. From an algorithmic perspective, speeding up the bound computations may be possible through leveraging the sparsity of the differential equations [84], probabilistic approaches to the SDP [85], or clever selection of the $D$-matrix constraints [1]. More broadly, our approach and extensions thereof can be applied to problems across the quantum-control landscape, ranging from speed limits and gate fidelity to areas like NMR [6–8] and quantum complexity [9–11].


022326 (2011).


Supplementary Materials:
Conservation-law-based global bounds to quantum optimal control

Hanwen Zhang,1, 2 Zeyu Kuang,1, 2 Shruti Puri,1, 3 and Owen D. Miller1, 2
1 Department of Applied Physics, Yale University, New Haven, Connecticut 06511, USA
2 Energy Sciences Institute, Yale University, New Haven, Connecticut 06511, USA
3 Yale Quantum Institute, Yale University, New Haven, Connecticut 06511, USA
(Dated: May 14, 2021)

CONTENTS

I. Three-level Hamiltonians 1
II. Finding a binary pulse equivalent to a continuous one by local averaging 2
III. Local probability conservation laws as a subset of D-matrix constraints 4
IV. Bound formulation for problems with multiple controllable terms 5
V. Bound formulation for open systems governed by master equations 6
VI. Numerical details 7
   A. Nyström discretization 7
   B. Discretization for quadratic constraints 8
   C. Pulse designs via local optimization 9
VII. Guide to computing prior-literature bounds 10
References 11

I. THREE-LEVEL HAMILTONIANS

The asymmetric double-well case is taken from Sec. 2.8 of [1],
\[
H = \begin{pmatrix}
\omega_0 & 0 & 0 \\
0 & \omega_1 & 0 \\
0 & 0 & \omega_2
\end{pmatrix} - \epsilon(t) \begin{pmatrix}
\mu_{00} & \mu_{01} & \mu_{02} \\
\mu_{10} & \mu_{11} & \mu_{12} \\
\mu_{20} & \mu_{21} & \mu_{22}
\end{pmatrix}
\]
\[
= \begin{pmatrix}
0 & 0 & 0 \\
0 & 0.1568 & 0 \\
0 & 0 & 0.7022
\end{pmatrix} - \epsilon(t) \begin{pmatrix}
-2.5676 & 0.3921 & 0.6382 \\
0.3921 & 2.3242 & -0.7037 \\
0.6382 & -0.7037 & -0.5988
\end{pmatrix}. \tag{1}
\]

We consider a maximum control amplitude of $|\epsilon| \leq 0.15$ in this example.

For the transmon example, the exact Hamiltonian is
\[
H = \begin{pmatrix}
\omega_0 & 0 & 0 \\
0 & \omega_1 & 0 \\
0 & 0 & \omega_2
\end{pmatrix} - \epsilon(t) \begin{pmatrix}
\mu_{00} & 0 & 0 \\
\mu_{10} & 0 & \mu_{12} \\
0 & \mu_{21} & 0
\end{pmatrix}
\]
\[
= \begin{pmatrix}
0 & 0 & 0 \\
0 & 1.9 & 0 \\
0 & 0 & 3.7
\end{pmatrix} + \epsilon(t) \begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & \sqrt{2} \\
0 & \sqrt{2} & 0
\end{pmatrix}, \tag{2}
\]

with a maximum control amplitude $|\epsilon| \leq 0.3$ in this example.
II. FINDING A BINARY PULSE EQUIVALENT TO A CONTINUOUS ONE BY LOCAL AVERAGING

Given a Hamiltonian \( H(t) = H_0(t) + \varepsilon(t)H_c(t) \) with some control \( \varepsilon(t) \), designed controls might be of “bang–bang” type, taking only two discrete values, or they might be smooth, continuous controls, possibly bounded in magnitude. From a theoretical bound perspective, the latter case is subsumed by the former: any smooth, continuous control can be approximated to arbitrarily high accuracy with a particular bang–bang control. Intuitively, one can see that this might be true: if you oscillate a bang–bang control at high enough frequency, oscillating over smaller time scales than any transitions (real or virtual) induced by the Hamiltonian, then the wave function will not respond to the particular high-frequency details of the control but rather to a homogenized average (which will lie between the two extremes). Mathematically, there is a rich literature on the field of “homogenization” theory [2], for applications from material science to optics. Here, we prove that any smooth, continuous control can be approximated by a bang–bang control, which then allows us to use bang–bang controls in the formulation of our bounds. (Note the key point that such bang–bang controls would not ever need to be implemented; they simply inform us of the generality of the bounds.)

Suppose that we are interested in the time evolution of the system from \( t_0 \) to \( t_f \), with continuous control \( \varepsilon_c(t) \) and propagator \( U_c(t_f, t_0) \). We can divide \([t_0, t_f]\) into \( N \) intervals of equal length \( \tau = (t_f - t_0)/N \): \([t_0, t_1], [t_1, t_2], \ldots, [t_{N-1}, t_N]\) \((t_f = t_N)\), such that each interval is shorter than any timescale of the Hamiltonian. We can then write \( U_c(t_f, t_0) = U_c(t_N, t_{N-1}) \cdots U_c(t_2, t_1)U_c(t_1, t_0) \). Since each interval is smaller than any transition time scale, we can apply first-order perturbation theory to each \( U(t_{i+1}, t_i) \) to analyze the accuracy of a bang–bang-control approximation. First, we analyze the accuracy of using a first-order approximation to the time-evolution operator over each interval:

\[
U_c(t_{i+1}, t_i) = U_0(t_{i+1}, t_i) - \frac{i}{\hbar} \int_{t_i}^{t_{i+1}} dt \frac{\partial}{\partial t} U_0(t_{i+1}, t)H(t)U_0(t, t_i)\varepsilon_c(t) + \varepsilon'_i
\]

(Taylor expand \( U_0(t_{i+1}, t)H(t)U_0(t, t_i) \) around \( t = t_i \))

\[
= U_0(t_{i+1}, t_i) - \frac{i}{\hbar} \int_{t_i}^{t_{i+1}} dt \frac{\partial}{\partial t} U_0(t, t_i)H_i \int_{t_i}^{t_{i+1}} dt \varepsilon_c(t) + \varepsilon_i
\]

where \( U_0 \) is the propagator of \( H_0 \) and \( \varepsilon'_i \) is the error term of the perturbation series of \( O(\tau^2) \). Together with the error due to Taylor expansion of \( O(\tau^2) \), the overall local error \( \varepsilon_i \) is of \( O(\tau^2) \). If we stitch these local approximations back to the global propagator, i.e. \( U_c(t_f, t_0) = U_c(t_N, t_{N-1}) \cdots U_c(t_2, t_1)U_c(t_1, t_0) \), we see that the global error is \( \sum_{i=0}^{N-1} U_0(t_{i+1}, t_i)\varepsilon_iU_0(t_i, t_0) \), which is at most \( N \sim 1/\tau \) copies of \( O(\tau^2) \), which is of \( O(\tau) \). Hence the global error goes to zero as \( \tau \to 0 \), and we can approximate \( U_c \) with the local approximations \( \tilde{U}_c \): \( U_c(t_f, t_0) \sim \tilde{U}_c(t_N, t_{N-1}) \cdots \tilde{U}_c(t_2, t_1)\tilde{U}_c(t_1, t_0) \) as \( \tau \to 0 \).

The above perturbation analysis shows that if we can find a bang–bang control \( \varepsilon_b(t) \) with propagator \( U_b(t_f, t_0) \) whose local approximation \( \tilde{U}_b(t_{i+1}, t_i) \) agrees with \( \tilde{U}_c(t_{i+1}, t_i) \) up to the first order in \( \tau \), then we will have

\[
U_c(t_f, t_0) \sim \tilde{U}_c(t_N, t_{N-1}) \cdots \tilde{U}_c(t_2, t_1)\tilde{U}_c(t_1, t_0) \sim \tilde{U}_b(t_N, t_{N-1}) \cdots \tilde{U}_b(t_1, t_0) \sim U_b(t_f, t_0)
\]

as \( \tau \to 0 \). This will show the global equivalence between \( U_c(t_f, t_0) \) and \( U_b(t_f, t_0) \). Such an \( \varepsilon_b(t) \) is easy to find: by the form of the perturbation in Eq. (4), any choice of \( \varepsilon_b \) is valid as long as \( \int_{t_i}^{t_{i+1}} dt \varepsilon_b(t) = \int_{t_i}^{t_{i+1}} dt \varepsilon_c(t) \) for all \( i \), which can be achieved simply by requiring the average of the bang–bang control equaling the average of the continuous control. (This choice of \( \varepsilon_b(t) \) over \([t_i, t_{i+1}]\) has \( \varepsilon_b = \varepsilon_{\text{min}} \) for a duration \( \tau' \) and then switch to \( \varepsilon_b = \varepsilon_{\text{max}} \) for the rest of the time to \( t_{i+1} \), for \( \tau' = \frac{\varepsilon_{\text{max}} - M}{\varepsilon_{\text{max}} - \varepsilon_{\text{min}}} \), where \( M = \int_{t_i}^{t_{i+1}} dt \varepsilon_c(t) \). Certainly, there are other choices as well.)

We illustrate the above proof numerically through the transmon example in the main text. We apply a continuous control \( \varepsilon_c(t) = E \cos(\omega t) \) with \( \omega = \omega_1 \) and \( E = 0.5 \) from \( t_0 = 0 \) to \( T = 12 \). We want to find a binary pulse with \( \varepsilon_{\text{min}} = -\varepsilon_{\text{max}} = E \), such that the dynamics \( \psi_b(t) \) approaches \( \psi_c(t) \) under the continuous pulse when \( \tau \) goes to zeros. We construct such \( \varepsilon_b(t) \) according to the method described in the previous paragraph. In Fig. 1, we compare \( \psi_b \) with \( \psi_c \) for \( \tau = 1.2 \) for (a), 0.3 for (c) and 0.15 for (e) and show their respective pulses in (b), (d) and (f). We see from In Fig. 1 the the evolution under each \( \varepsilon_b(t) \) (red) gets closer to the evolution under \( \varepsilon_c(t) \) (black dash) as \( \tau \) decreases. More explicitly, we plot in Fig. 2 the relative difference between evolution of \( \varepsilon_c \) and \( \varepsilon_b \), measured by \( \frac{(\int_{t_0}^{t_f} |\psi_b(t) - \psi_c(t)|^2 dt)^{1/2}}{(\int_{t_0}^{t_f} |\psi_c(t)|^2 dt)^{1/2}} \), against \( \tau \). We can see that the convergence demonstrates the \( O(\tau) \) behavior proved above. We emphasize again that the purpose of numerical results here is to show the equivalence between continuous and binary pulses when \( \tau \) goes to zero. In the bound computation, we do not need to find such \( \varepsilon_b(t) \) and the bound converges automatically and faster than \( O(\tau) \) in the proof.
To conclude, we have proven that for any continuously valued control $\varepsilon_c(t)$ there is a corresponding binary control $\varepsilon_b(t)$ producing the same time evolution as $\varepsilon_c(t)$, when the local averaging time $\tau$ approaches zero. Hence, a bound derived for bang–bang controls will include all possible bounded continuous controls as well.

![Graph](image_url)

**FIG. 1.** (a) Comparisons between probability time evolution under $\varepsilon_c(t)$ (red) and $\varepsilon_b(t)$ (black dash) for all three levels, from $t_0 = 0$ to $T = 12$. (b) $\varepsilon_c(t) = E \cos(\omega t)$ is plotted in red and the $\varepsilon_b(t)$ in black, obtained through averaging for $\tau = 1.2$. (c) and (d), (e) and (f) are similar to (a) and (b), but the former pair is for $\tau = 0.3$ and the latter is for $\tau = 0.15$. One can see that as $\tau$ gets smaller, the $\varepsilon_b(t)$ oscillates more rapidly and the evolution gets closer to the one produce by $\varepsilon_c(t)$. For $\tau = 0.15$, the effect of $\varepsilon_b(t)$ and $\varepsilon_c(t)$ are almost identical.

![Graph](image_url)

**FIG. 2.** The relative difference between evolution of $\varepsilon_c$ and $\varepsilon_b$, measured by \[ \frac{\int_{t_0}^{t_f} |\psi_b(t) - \psi_c(t)|^2 dt}{\int_{t_0}^{t_f} |\psi_c(t)|^2 dt} \] is plotted against $\frac{\tau}{T}$. It shows the $O(\tau)$ convergence for small $\tau$. 

\[ x = 0 \]
III. LOCAL PROBABILITY CONSERVATION LAWS AS A SUBSET OF D-MATRIX CONSTRAINTS

In this section, we show explicitly that Eq. 4 in the main text, which represents local probability conservation, is a subset of constraints of Eq. 3 in the main text with \( D_i(t) = \eta(t_1 - t)I \) with \( t \geq t_0 \), i.e., identity operator \( I \) supported from \( t_0 \) to \( t_1 \).

Now consider Eq. 2 in the main text,

\[
\int_{t=t_0}^{T} \int_{t'=t_0}^{T} \Phi^\dagger(t) \Phi(t') D_i(t) \left( \frac{H_c^{-1}}{\varepsilon_{\text{max}}} \delta(t - t') + \frac{i}{\hbar} G_0^+(t, t') \right) \Phi(t') dt' dt = \int_{t_0}^{T} \Phi^\dagger(t) D_i(t) U_0(t, t_0) dt,
\]

(7)

where \( G_0^+(t, t') = U_0(t, t') \eta(t - t') \). Now with \( D_i(t) = \eta(t_1 - t)I \), it becomes

\[
\int_{t_0}^{t_1} \Phi^\dagger(t) \frac{H_c^{-1}}{\varepsilon_{\text{max}}} \Phi(t) dt + \frac{i}{\hbar} \int_{t=t_0}^{t_1} \int_{t'=t_0}^{t} \Phi^\dagger(t) U_0(t, t') \Phi(t') dt' dt = \int_{t_0}^{t_1} \Phi^\dagger(t) U_0(t, t_0) dt.
\]

(8)

Taking Hermitian conjugation both side of Eq. (8), and using \( U_0^\dagger(t, t') = U_0(t', t) \) and \( (H_c^{-1})^\dagger = H_c^{-1} \), produces

\[
\int_{t_0}^{t_1} \Phi^\dagger(t) \frac{H_c^{-1}}{\varepsilon_{\text{max}}} \Phi(t) dt - \frac{i}{\hbar} \int_{t=t_0}^{t_1} \int_{t'=t_0}^{t} \Phi^\dagger(t') U_0(t', t) \Phi(t) dt dt' = \int_{t_0}^{t_1} U_0^\dagger(t, t_0) \Phi(t) dt
\]

(9)

Note we can exchange the \( t, t' \) integration in the the following integral

\[
\int_{t=t_0}^{t_1} \int_{t'=t_0}^{t} \ldots dt' dt = \int_{t'=t_0}^{t_1} \int_{t=t_0}^{t} \ldots dt dt'.
\]

As a result, the double integral in Eq. (9) becomes

\[
\int_{t=t_0}^{t_1} \int_{t'=t_0}^{t} \Phi^\dagger(t') U_0(t', t) \Phi(t) dt dt' \quad \text{(exchange} \ t, t' \text{integrations)}
\]

\[
= \int_{t=t_0}^{t_1} \int_{t'=t}^{t_1} \Phi^\dagger(t') U_0(t', t) \Phi(t) dt dt' \quad \text{(relabeling by} \ t \rightarrow t', t' \rightarrow t)
\]

\[
= \int_{t=t_0}^{t_1} \int_{t'=t}^{t_1} \Phi^\dagger(t) U_0(t, t') \Phi(t') dt dt'.
\]

Now we subtract Eq. (8) by Eq. (9) to obtain the imaginary part of Eq. (8), and we get

\[
\frac{1}{\hbar} \int_{t_0}^{t_1} \int_{t'=t_0}^{t} \Phi^\dagger(t) U_0(t, t') \Phi(t') dt dt' = 2 \text{Im} \int_{t_0}^{t_1} \Phi^\dagger(t) U_0(t, t_0) dt,
\]

(10)

which is exactly Eq. 4 in the main text, since \( U^\dagger U = I \) by probability conservation. \( \int_{t_0}^{t_1} \Phi^\dagger(t) \frac{H_c^{-1}}{\varepsilon_{\text{max}}} \Phi(t) dt \) disappears since its imaginary part is zero. This makes sense since \( H_c \) is a Hermitian matrix and no dissipation is present.

Often \( D_i(t) = \delta(t - t')D \) is localized at time \( t' \) (\( D \) is some constant matrix). To connect such \( D_i(t) \) with the above result, we notice that \( \int_{t_0}^{t_1} \delta(t - t') dt' = \eta(t_1 - t) \) for \( t \geq t_0 \). As a result, we can interpret the above conservation law as the imaginary part of a linear combination of constraints with \( D_i(t) = \delta(t - t')I \) for \( t_0 \leq t' \leq t_1 \), a subset of all D-matrix constraints. More generally, \( D \) is not necessarily \( I \) and the real part is also a valid D-matrix constraint. We thus can interpret the entire set of D-matrix constraints as generalized local probability conservation laws.
IV. BOUND FORMULATION FOR PROBLEMS WITH MULTIPLE CONTROLLABLE TERMS

In the main text, we only consider problems with a single control. We show here that our framework extends naturally to problems with multiple controls. Consider a Hamiltonian \( H(t) = H_0(t) + \sum_{j=1}^{N} \epsilon_j(t) H_j^c(t) \), where \( H_0(t) \) is the non-controllable part and \( \sum_{j=1}^{N} \epsilon_j(t) H_j^c(t) \) is the controllable part. To avoid overloading the notation, we here formulate the bound problem for a two-term problem with \( H(t) = H_0(t) + \epsilon_1(t) H_1^c(t) + \epsilon_2(t) H_2^c(t) \), which can be extended to more terms in a straightforward way.

The first step is to write the integral-equation form (analog of Eq. (1) of the main text) for multiple controls. To do this, we will assume the two controls cannot be turned on at the same times. (Note that two controls that can be turned on at the same time is equivalent to three controls that cannot be turned on at the same time, so there is no loss of generality here.) With two independent controls, the integral equation for \( U(t, t_0) \) is:

\[
U(t, t_0) = U_0(t, t_0) - \frac{i}{\hbar} \int_{t_0}^{T} G_0^+(t, t') H_1^c(t) \epsilon_1(t') U(t', t_0) \, dt' - \frac{i}{\hbar} \int_{t_0}^{T} G_0^+(t, t') H_2^c(t) \epsilon_2(t') U(t', t_0) \tag{11}
\]

For times \( \epsilon_2(t) \) is zero, we have

\[
U(t, t_0) = U_0(t, t_0) - \frac{i}{\hbar} \int_{t_0}^{T} G_0^+(t, t') H_1^c(t') \epsilon_1(t') U(t', t_0) \, dt', \tag{12}
\]

and for times \( \epsilon_1(t) \) is zero,

\[
U(t, t_0) = U_0(t, t_0) - \frac{i}{\hbar} \int_{t_0}^{T} G_0^+(t, t') H_2^c(t') \epsilon_2(t') U(t', t_0) \, dt'. \tag{13}
\]

We then left multiply \( U^+(t, t_0) \epsilon_1(t) H_1^c D_1(t) \) to Eq. (12) and \( U^+(t, t_0) \epsilon_2(t) H_2^c D_2(t) \) to Eq. (13), and integrate over \( t \) from \( t_0 \) to \( T \). Then Eq. (12) and Eq. (13) become

\[
\int_{t_0}^{T} dt' U^+(t, t_0) \epsilon_1(t) H_1^c D_1(t) U(t, t_0) = \int_{t_0}^{T} dt' U^+(t, t_0) \epsilon_1(t) H_1^c D_1(t) U_0(t, t_0)
- \frac{i}{\hbar} \int_{t_0}^{T} dt \int_{\{\epsilon_1(t) \neq 0\}} dt' U^+(t, t_0) \epsilon_1(t) H_1^c D_1(t) G_0^+(t, t') H_1^c(t') U(t', t_0), \tag{14}
\]

and

\[
\int_{t_0}^{T} dt' U^+(t, t_0) \epsilon_2(t) H_2^c D_2(t) U(t, t_0) = \int_{t_0}^{T} dt' U^+(t, t_0) \epsilon_2(t) H_2^c D_2(t) U_0(t, t_0)
- \frac{i}{\hbar} \int_{t_0}^{T} dt \int_{\{\epsilon_2(t) \neq 0\}} dt' U^+(t, t_0) \epsilon_2(t) H_2^c D_2(t) G_0^+(t, t') H_2^c(t') U(t', t_0), \tag{15}
\]

We define \( \Phi_1(t) = \epsilon_1(t) H_1^c U_1(t, t_0) \) and \( \Phi_2(t) = \epsilon_2(t) H_2^c U_2(t, t_0) \). The above equations are quadratic in \( \Phi_1 \) and \( \Phi_2 \), thus they hold for \( t' \) from \( t_0 \) to \( T \). Consequently, we can replace the integral range of \( t' \) by \( t_0 \) and \( T \) and obtain constraints

\[
\int_{t_0}^{T} dt \int_{t_0}^{T} dt' \Phi_1^+(t) D_1(t) \left( \delta(t - t') \frac{H_1^c}{\epsilon_{1_{\text{max}}}} + \frac{i}{\hbar} G_0^+(t, t') \right) \Phi_1(t') = \int_{t_0}^{T} dt \Phi_1^+(t) D_1(t) U_0(t, t_0), \tag{16}
\]

and

\[
\int_{t_0}^{T} dt \int_{t_0}^{T} dt' \Phi_2^+(t) D_2(t) \left( \delta(t - t') \frac{H_2^c}{\epsilon_{2_{\text{max}}}} + \frac{i}{\hbar} G_0^+(t, t') \right) \Phi_2(t') = \int_{t_0}^{T} dt \Phi_2^+(t) D_2(t) U_0(t, t_0), \tag{17}
\]

similar to Eq. (3) in the main text, but now with two different components. In addition to these two sets of constraints, we also need to impose that only one of \( \epsilon_1(t) \) and \( \epsilon_2(t) \) can be nonzero at any given time, which is represented in the constraint

\[
\Phi_1^+(t) \Phi_2(t) = 0, \tag{18}
\]
for all times \( t \). Hence bounds for two controls can be formulated as
\[
\max_{\Phi_1, \Phi_2} f(\Phi_1, \Phi_2)
\]
s.t. Equation (16) satisfied for all \( D_i(t) \),
Equation (17) satisfied for all \( D_j(t) \),
\[
\Phi_1^\dagger(t)\Phi_2(t) = 0 \quad \text{for all } t, \text{for all components}.
\]

One can see immediately to generalize to multiterm pulse designs. For \( H(t) = H_0(t) + \sum_{i=1}^N \varepsilon_i(t)H_i^e \), we will have \( N \) different \( \Phi_i \) with its corresponding sets of quadratic constraints. We also need to impose that \( \Phi_1^\dagger(t)\Phi_j(t) = 0 \) for all \( i \neq j \).

Finally, note that if both \( \varepsilon_1(t) \) and \( \varepsilon_2(t) \) can be nonzero at the same time, it is equivalent to considering a problem with three controllable terms: \( \varepsilon_1(t)H_1^e \), \( \varepsilon_2(t)H_2^e \), and \( \varepsilon_1(t)H_1^e + \varepsilon_2(t)H_2^e \), only one of which can be nonzero at any given time.

\[ \text{V. BOUND FORMULATION FOR OPEN SYSTEMS GOVERNED BY MASTER EQUATIONS} \]

In the main text, we only consider problems whose dynamics are governed by the Schrödinger equation. Now we show how to apply our bound framework to open systems described by differential equations for the density matrix \( \rho \). In principle, this framework applies to any equation that is linear in \( \rho \).

\[ \partial_t \rho = -\frac{i}{\hbar}[H, \rho] + \sum_i \gamma_i \left( A_i\rho A_i^\dagger - \frac{1}{2} \{A_i^\dagger A_i, \rho\} \right), \quad (19) \]

where we also assume the Hamiltonian has the form \( H = H_0 + \varepsilon(t)H_e \), and the second term is the dissipative part.

A common trick is to vectorize the equation to form “superoperators” \([3]\), through flattening the \( n^2 \) by \( n^2 \) density matrix \( \rho \) into a column vector \( r \) by ordering the columns of \( \rho \) in a single vector \([4]\). Then all operators are transformed according to this new ordering (basis):

\[ [H, \rho] \rightarrow (I \otimes H - H^T \otimes I) r \]
\[ = (I \otimes H_0 - H_0^T \otimes I) r + (I \otimes H_e - H_e^T \otimes I) r \quad (20) \]
\[ A_i \rho A_i^\dagger \rightarrow \left( (A_i^\dagger)^T \otimes A_i \right) r \quad (21) \]
\[ A_i A_i^\dagger \rho \rightarrow \left( I \otimes (A_i^\dagger A_i) \right) r \quad (22) \]
\[ \rho A_i^\dagger A_i \rightarrow \left( (A_i^\dagger A_i)^T \otimes I \right) r. \quad (23) \]

If we define the vectorized \(-\frac{i}{\hbar}[H_0, \rho]\) by \( L_0 \), the vectorized dissipative term by \( L_D \), and vectorized \(-\frac{i}{\hbar}[H_e, \rho]\) by \( L_e \), we turn the original equation into
\[ \frac{\partial}{\partial t} r = (L_0 + L_D) r + L_e r. \quad (24) \]

This equation again can be turned into the integral form
\[ r(t) = r_0(t) + \int_{t_0}^t dt' G^+(t, t') L_e(t') r(t'), \quad (25) \]

where \( r_0(t) \) is the evolution of \( r \) without the presence of \( L_e \), and \( G^+(t, t') \) is the retarded Green’s function of \( L_e \).

Eq. (25) is the starting point of the bound formulation, just as Eq. (1) in the main text for the Schrödinger equation. Carrying out the same manipulations as in the main text gives the constraints
\[ \int_{t_0}^t dt r^\dagger(t) L_e(t) D_i(t) r(t) + \int_{t_0}^t dt \int_{t_0}^t dt' r^\dagger(t) L_e(t) D_i(t) G^+(t, t') L_e(t') r(t') = \int_{t_0}^t dt r^\dagger(t) L_e(t) D_i(t) r(t), \quad (26) \]

which are analogous to Eq. (2) in the main text. The quadratic constraints can then be formed analogously.
VI. NUMERICAL DETAILS

In this section, we provide details for the numerical scheme used in the main text. Unlike conventional time-stepping methods to solve initial-value differential equations, we directly use the integral equation (Eq. (1) of the main text) to solve for the dynamics. We do this for two reasons: (1) in our bound method, constraints come from the integral equation, so the discretized form of the integral equation is directly available after solving the bound problem, and (2) pulse design is straightforward to perform with methods based on the integral equation. (3) Methods based on the integral equation are equally viable as those based on the differential equation.

A. Nyström discretization

We consider the integral equation in Eq. (1). We consider systems with \( L \) levels, modeling either discrete levels or a subset of continuous states, over times \([t_0, T]\). Here we isolate a single column of the time-evolution operator, essentially selecting an initial state \(|\psi_0\rangle\) and the time-dependent state \(|\psi(t)\rangle = U(t, t_0) |\psi_0\rangle\). Then the integral equation for \(|\psi(t)\rangle\) is

\[
|\psi(t)\rangle = |\psi_0(t)\rangle - \frac{i}{\hbar} \int_{t_0}^{t} U_0(t, t') H_c \varepsilon(t') |\psi(t')\rangle \, dt',
\]

where we have replaced the retarded Green’s function with the time-evolution operator multiplied by a step function (changing the upper limit of the integral to \( t \)). This is in the form of a Volterra integral equation of the second kind.

The Nyström method [5] for discretizing this equation consistent of defining quadrature nodes \( \{t_i\}_{i=0}^{N} \in [t_0, T] \) and corresponding weights, and enforcing Eq. (27) to hold at \( \{t_i\}_{i=0}^{N} \)

\[
|\psi(t_i)\rangle = |\psi_0(t_i)\rangle - \frac{i}{\hbar} \int_{t_0}^{t_i} U_0(t_i, t') H_c \varepsilon(t') |\psi(t')\rangle \, dt', \quad i = 0, 1, 2, \ldots, N,
\]

and then approximating the integral operator by

\[
\frac{i}{\hbar} \int_{t_0}^{t_i} U_0(t_i, t') H_c \varepsilon(t') |\psi(t')\rangle \, dt' \approx \sum_{j=1}^{N} w_{ij} \frac{i}{\hbar} U_0(t_i, t_j) H_c \varepsilon(t_j) |\psi(t_j)\rangle.
\]

for some weights \( w_{ij} \). Then Eq. (27) becomes a linear system

\[
|\psi(t_i)\rangle = |\psi_0(t_i)\rangle - \sum_{j=0}^{i} w_{ij} K_{ij} |\psi(t_j)\rangle, \quad i = 0, 1, 2, \ldots, N,
\]

or more compactly as

\[
\psi = \psi_0 - A\psi,
\]

where \( \psi \) and \( \psi_0 \) are column vector of length \( L(N+1) \) with the \( i \)th column block \( \psi_i = |\psi(t_i)\rangle \) and \( (\psi_0)_i = |\psi_0(t_i)\rangle \) respectively. And \( A \) is a matrix of size \( L(N+1) \times L(N+1) \) whose explicit form we will show later. Then \( \psi \) can be solved as \((I + A)^{-1}\psi_0\).

We note that the convergence rate of \( \psi \) to the exact value \( \psi \) is the same as the order of quadrature used in Eq. (29). The matrix norm of \( A \) is provably small, so \( I + A \) is always well-conditioned. To conclude, the scheme we described above is stable and convergences quickly for good choice of \( \{t_i\}_i^{N} \) and weights.

In this work, we use the trapezoidal rule

\[
\int_{a}^{b} f(x) \, dx \approx \frac{h}{2} (f(a) + 2f(a + h) + \ldots + 2f(b - h) + f(b)) + O(h^2)
\]

where \( h \) is the spacing between the equally spaced nodes from \( a \) to \( b \). As a result, the method we are using has convergence rate equivalent to Runge Kutta of order 2. Without special techniques for the matrix inversion process, this method is not as efficient as Runge Kutta of order 2. Despite of this drawback and the fact that higher order
methods are available, the method is sufficient for the examples we consider. If we apply the trapezoidal rule to Eq. (30), then the nodes \( \{ t_i \}_{i=0}^N \in [0, T] \) are \( t_0, t_1, \ldots, t_N = T \) with \( t_i = t_0 + i h \) and \( h = \frac{T-t_0}{N} \). Eq. (30) then becomes

\[
\begin{align*}
|\psi(t_0)| &= |\psi_0(t_0)| \\
|\psi(t_1)| &= |\psi_0(t_1)| - \frac{h}{2} (K_{10} |\psi(t_0)| + K_{11} |\psi(t_1)|) \\
|\psi(t_2)| &= |\psi_0(t_2)| - \frac{h}{2} (K_{20} |\psi(t_0)| + 2K_{21} |\psi(t_1)| + K_{22} |\psi(t_2)|) \\
&\vdots \\
|\psi(t_N)| &= |\psi_0(t_N)| - \frac{h}{2} (K_{N0} |\psi(t_0)| + 2 \sum_{j=1}^{N-1} K_{Nj} |\psi(t_j)| + K_{NN} |\psi(t_N)|),
\end{align*}
\]

which is Eq. (31) in component form with the trapezoidal rule.

To solve for \( U \), we simply replace the vector \( \psi \) by the matrix \( U = (\psi^1, \psi^2, \ldots, \psi^L) \) and \( \psi_0 \) by the matrix \( U_0 = (\psi^1_0, \psi^2_0, \ldots, \psi^L_0) \), both of size \( L(N+1) \times L \), where \( \psi^i \) represents the discretized version of different states wave function \( |\psi^i(t)| = U(t, t_0) |\psi^i_0(t_0)| \). By comparing Eq. (33) and Eq. (30), we can see that the weight \( w_{ij} = h \) for \( i \neq j \) and \( w_{ij} = h/2 \) for \( i = j \) and \( w_{ii} = 0 \) for \( i < j \). Then the discretized version of Eq. (1) in the main text is given by

\[
U = U_0 - AU,
\]

and \( U = (I + A)^{-1}U_0 \).

**B. Discretization for quadratic constraints**

With the understanding of how the integral equation is discretized, we are ready to give the discretization form of Eq. (3), which constitutes the constraints for computing the bound. Constraints in Eq. (3) are given by

\[
\int_t^T dt \int_t^{t'} dt' \Phi^\dagger(t) D_i(t) \left( \delta(t-t') \frac{H^{-1}}{\varepsilon_{\text{max}}} + \frac{i}{\hbar} G_i^0(t, t') \right) \Phi(t')
= \int_t^T dt \Phi^\dagger(t) D_i(t) U_0(t, t_0),
\]

or

\[
\int_t^T dt \int_t^{t'} dt' \Phi^\dagger(t) D_i(t) \frac{i}{\hbar} U_0(t, t') \Phi(t') + \int_t^T dt \Phi^\dagger(t) D_i(t) \frac{H^{-1}}{\varepsilon_{\text{max}}} \Phi(t)
= \int_t^T dt \Phi^\dagger(t) D_i(t) U_0(t, t_0),
\]

We have seen in the previous section on how to discretize the \( t' \) integral. The \( t \) integral is in fact much easier with fixed integration limit, and we can apply the trapezoidal rule directly to the integral \( \int_t^T dt f(t) \approx \sum_{i=0}^N w_i f(t_i) \) with the same equally spaced nodes \( \{ t_i \}_{i=0}^N \) and weights \( w_i \) in Eq. (32). Then Eq. (36) becomes

\[
\Phi^\dagger D_i W G \Phi + \Phi^\dagger D_i W \frac{H^{-1}}{\varepsilon_{\text{max}}} \Phi = \Phi^\dagger D_i W U_0,
\]

where \( \Phi \) and \( U_0 \) are matrix of size \( L(N+1) \times L \). \( G \) is the discretized version of the integral operator \( \int_t^T dt' U(t, t') \).

Similar to the way we discretized Eq. (29), we have the \( ij \) block of \( G \)

\[
G_{ij} = w_{ij} \frac{i}{\hbar} U_0(t_i, t_j)
\]
with $w_{ij}$ the weight in Eq. (33). $W$ is the matrix contains the effect of weights $w_i$ from the $t$ integral, and it is of the form

$$W = \begin{pmatrix} w_0 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} \otimes I_{L \times L},$$

(39)

where $\otimes$ is the kronecker product for matrices and $I_{L \times L}$ represents an identity matrix of dimension $L$. Here $H_c^{-1}$ is

$$H_c^{-1} = I_{(N+1) \times (N+1)} \otimes H_c^{-1}$$

(40)

Also, since $D_i = \delta(t - t_i)D$ and $D$ is some constant matrix of dimension $L$ by $L$, we have

$$D_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \otimes D,$$

(41)

where the first diagonal matrix is the discretized $\delta(t - t_i)$ and is nonzero only in the $i$th diagonal entry. $D$ of dimension $L$ by $L$ consists of $L^2$ unit basis matrices. For a two level system, all possible independent $D$ consists of $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Higher dimension systems can be generalized accordingly.

By varying $D_i$ over all possible $t_i$ and $D$, we obtain all possible constraints in our problem in the form of a $L$ by $L$ matrix equality, which must hold componentwise. $\Phi$ can be flattened column by column into a vector, and the equality constraints for each components can then be transformed into a quadratic constraint of the flattened $\Phi$. Then semidefinite relaxation can be carried out.

C. Pulse designs via local optimization

In this section, we describe how to efficiently compute the gradient $\frac{\partial f(U)}{\partial p}$ of the objective $f(U)$ for all $t$ from $[t_0, T]$ by solving for the dynamics twice only. Now, we return to Eq. (34), the discretized version of Eq. (1) in the main text. To make the variables $\varepsilon(t_i)$ here more prominent, we separate $\varepsilon(t_i)$ from $A$ in Eq. (34) and write it as

$$U = U_0 - GH_c \varepsilon U,$$

(42)

where

$$H_c = I_{(N+1) \times (N+1)} \otimes H_c,$$

(43)

$$\varepsilon = \begin{pmatrix} \varepsilon(t_0) \\ \varepsilon(t_1) \\ \vdots \\ \varepsilon(t_N) \end{pmatrix} \otimes I_{L \times L},$$

(44)

and $G$ is given in Eq. (38). We also denote the column vector $(\varepsilon(t_0) \; \varepsilon(t_1) \; \ldots \; \varepsilon(t_N))$ by $p$.

The goal here is to compute the discretized gradient $\frac{\partial f(U)}{\partial p}$ in a way that the work required is independent of the number of variable $\varepsilon(t_i)$. To do so, we first use the CR calculus to formally treat $U$ and $U^*$ as independent variables [6]; the chain rule yields

$$\frac{\partial f(U)}{\partial p} = \frac{\partial f(U)}{\partial U} \frac{\partial U}{\partial p} + \frac{\partial f(U)}{\partial U^*} \frac{\partial U^*}{\partial p} = 2 \Re \frac{\partial f(U)}{\partial U} \frac{\partial U}{\partial p}. $$

(45)
Then we differentiate Eq. (42) with respect to \( p \) and obtain

\[ G H_c \frac{\partial \varepsilon}{\partial p} U = -(I + A) \frac{\partial U}{\partial p}, \quad (46) \]

and we can solve for \( \frac{\partial U}{\partial p} = -(I + A)^{-1} G H_c \frac{\partial \varepsilon}{\partial p} U \). Substituting this into Eq. (45) gives

\[ \frac{\partial f(U)}{\partial p} = -2 \text{Re} \left( \frac{\partial f(U)}{\partial U} \right) (I + A)^{-1} G H_c \frac{\partial \varepsilon}{\partial p} U, \quad (47) \]

Since \( G, H_c \) are known \( \frac{\partial \varepsilon}{\partial p} \) can be calculated analytically, then if we define the so-called “adjoint” solution [7, 8]

\[ (I + A)^T U_{\text{adj}} = \left( \frac{\partial f(U)}{\partial U} \right)^T, \quad (48) \]

we can obtain the gradient

\[ \frac{\partial f(U)}{\partial p} = -2 \text{Re} U_{\text{adj}}^T G H_c \frac{\partial \varepsilon}{\partial p} U, \quad (49) \]

by solving for \( U_{\text{adj}} = (I + A^T)^{-1} \left( \frac{\partial f(U)}{\partial p} \right)^T \) from Eq. (48) and \( U = (I + A)^{-1} U_0 \) from Eq. (42) only. This allows rapid computation of \( \frac{\partial f(U)}{\partial p} \). Then standard optimization methods, such as gradient descent, can be applied to maximize or minimize \( f(U) \). All designed pulses in the main text are obtained by this approach described here.

VII. GUIDE TO COMPUTING PRIOR-LITERATURE BOUNDS

In this section, we describe how we computed the bounds from prior literature in Fig. 1 of the main text: Mandelstam–Tamm (MT), Margolus–Levitin (ML), and the bounds from Refs. [9, 10]. We start with the MT bound: the minimum time, which we denote \( \tau_{\text{MT}} \), for a given evolution is given by

\[ \tau_{\text{MT}} = \frac{\pi}{2 \Delta H}, \quad (50) \]

where the standard deviation \( \Delta H \) is given by \( \Delta H = \sqrt{\langle H^2 \rangle - \langle H \rangle^2} \). We do not know the mean of the Hamiltonian over the optimal trajectory, but can bound its square below by zero, which gives \( \Delta H \leq \sqrt{\langle H^2 \rangle} \). We don’t know the value of \( \langle H^2 \rangle \) over the optimal trajectory either, but it can be bounded above by the square of the largest possible eigenvalue of the Hamiltonian, which we denote \( \lambda_{\text{max}} \). This eigenvalue can be found by a grid-search-based optimization over all possible instantaneous Hamiltonians (assuming bounded controls), and we are left with the bound:

\[ \tau_{\text{MT}} \geq \frac{\pi}{2 \lambda_{\text{max}}}, \quad (51) \]

The ML bound, which we denote \( \tau_{\text{ML}} \), follows similar reasoning. Now given by \( \tau_{\text{ML}} = \pi/2 \langle H \rangle \), we again do not know the mean energy, but can now bound it above by \( \lambda_{\text{max}} \), which implies that

\[ \tau_{\text{ML}} \geq \frac{\pi}{2 \lambda_{\text{max}}}, \quad (52) \]

By this reasoning, the MT and ML bounds coincides when presented with uncertainty about the optimal trajectory and bounded in this way.

Next, we consider the bound of Ref. [9]. In this case, one defines two functions incorporating information about the controllable and non-controllable Hamiltonians, \( H_c \) and \( H_0 \), respectively:

\[ C(U_g, H_c) = \frac{\sqrt{2} \left( d - \sum_{j=1}^{d} \left| \langle \phi_j^{(c)} | U_g | \phi_j^{(c)} \rangle \right| \right)}{2}, \quad (53) \]
\[ C(U_g, H_0) = \frac{\sqrt{2} \left( d - \sum_{j=1}^{d} \left| \langle \phi_j^{(0)} | U_g | \phi_j^{(0)} \rangle \right| \right)}{2}, \quad (54) \]
where \( U_g \) is a target evolution operator, \( d \) is the dimensionality, and \( \phi_j^{(c)} \) and \( \phi_j^{(0)} \) are the eigenfunctions of the controllable and non-controllable Hamiltonians, respectively. From these functions, the time bounds of Ref. [9] can be written:

\[
T \geq \max \left\{ \frac{2C(U_g, H_c)}{\|H_0\|}, \frac{2C(U_g, H_0)}{|f_{\text{max}}|\|H_c\|} \right\},
\]

(55)

where \( |f_{\text{max}}| \) is the maximum amplitude of the bounded control parameter. Equation (55) is straightforward to compute, though hard to apply to the maximum transition probability in a multilevel system, as there is no specific target unitary evolution. In this case, to compute a meaningful bound at all (for the sake of comparison), we instead consider the best numerically-optimized control sequence, and use its evolution operator evaluated at the final time as the target unitary. In this way, we take a known achievable propagator and can generate a meaningful bound to compare it to. It should be noted that if one were to want to derive a strict bound in this way, it would be required to find the smallest bound over all possible unitaries that correspond to unity probability in the desired state; hence, our computation of a bound is actually an over-estimate of the “true” bound using this approach, which is necessarily smaller (and thereby looser).

The bound of Ref. [10] is similarly for a target unitary, but now uses a different expression. Now, the minimum time is given by:

\[
T \geq \max_{V \in \text{Stab}(H_c)} \frac{\|[U_g, V]\|}{\|[H_0, V]\|}.
\]

(56)

This bound arises from maximizing the ratio of norms on the right-hand side of Eq. (56) over all matrices that commute with \( H_c \). Generically, there is no simple way to optimize over all such matrices. To compute a reasonable bound, then, we choose \( H_c \) for \( V \), as it is guaranteed to commute with itself. Then we evaluate the ratio in Eq. (56), again using a single unitary as described above (instead of minimizing over all possible unitaries). In both Eq. (55) and Eq. (56), the Frobenius norm is the norm that is used.