Universal gates for protected superconducting qubits using optimal control

Mohamed Abdelhafez,[1] Brian Baker,[2] András Gyenis,[3] Pranav Mundada,[3] Andrew A. Houck,[3] David Schuster,[1] and Jens Koch[2]

[1]*The James Franck Institute and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA*
[2]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*
[3]*Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA*

(Received 28 August 2019; accepted 30 January 2020; published 18 February 2020)

We employ quantum optimal control theory to realize quantum gates for two protected superconducting circuits: the heavy-fluxonium qubit and the 0-π qubit. Utilizing automatic differentiation facilitates the simultaneous inclusion of multiple optimization targets, allowing one to obtain high-fidelity gates with realistic pulse shapes. For both qubits, disjoint support of low-lying wave functions prevents direct population transfer between the computational-basis states. Instead, optimal control favors dynamics involving higher-lying levels, effectively lifting the protection for a fraction of the gate duration. For the 0-π qubit, offset-charge dependence of matrix elements among higher levels poses an additional challenge for gate protocols. To mitigate this issue, we randomize the offset charge during the optimization process, steering the system towards pulse shapes insensitive to charge variations. Closed-system fidelities obtained are 99% or higher and show slight reductions in open-system simulations.

DOI: 10.1103/PhysRevA.101.022321

I. INTRODUCTION

The ability to perform fast, high-fidelity gate operations on qubits is critical for quantum information processing. A host of research over the past decades has pursued optimal strategies to realize qubit gates [1–12]. Realizing universal sets of quantum gates has been achieved through a variety of techniques for different qubit implementations [13–19]. One standard example of a minimal universal set of gates consists of the single-qubit Hadamard gate H, the single-qubit T gate, and one two-qubit entangling gate such as the controlled-Z gate. Realizing these building blocks with high fidelities is crucial in order to meet the gate-fidelity threshold required for error correction codes [20,21].

With superconducting qubits, optimized gates with fidelities exceeding 99% have been proposed [20,22–25], for example, for the transmon qubit [26,27]. Driving transitions in the computational subspace of the transmon qubit is facilitated by direct matrix elements between the states $|0\rangle$ and $|1\rangle$ whose wave functions reside in the same cosine-potential well. Recently, a new generation of superconducting qubits has been introduced which feature disjoint support: low-energy wave functions are localized in different potential wells so that matrix elements of local operators are exponentially suppressed [28–31]. Therefore, these qubits are intrinsically protected from spontaneous transitions between the computational-basis states. Two of the most promising protected superconducting qubits are the heavy-fluxonium qubit [30,32,33] and the 0-π circuit [29,34–36]. Performing high-fidelity gates on these qubits is challenging precisely because of the lack of direct transition matrix elements. For the heavy-fluxonium qubit, these forbidden transitions have been successfully accessed by stimulated Raman processes [30,31]. For the 0-π circuit, a recent study has proposed dc-voltage signals for realizing either an X gate or Hadamard gate [36]. Here, we argue that optimal-control theory is a promising route to explore the options for high-fidelity gates in protected qubits such as heavy fluxonium and 0-π.

Optimal-control theory, applied to quantum systems, achieves a set of optimization targets, the primary target usually consisting of a maximized gate or state-transfer fidelity. Additional constraints associated with specific experimental systems may be added, and include smoothing of control pulses and limiting their amplitudes [37–39], as well as accounting for the limited time resolution of arbitrary waveform generators [40]. There are many different implementations of optimal-control algorithms. Examples of such algorithms include implementations for closed [1,3,5,6,24] and open [1,25,41] quantum systems; most of them are gradient based. Some of these algorithms are available as open-source packages [42–44], and we here utilize the automatic-differentiation [45] based quantum optimizer we previously introduced in Ref. [24]. Automatic differentiation allows for the flexibility of adding optimization targets without calculating their analytical gradients. We utilize and further develop this optimal-control implementation to obtain a universal set of gates for the protected heavy-fluxonium and 0-π qubits. For the latter, we find that optimal-control pulses strikingly succeed in overcoming the obstacle of offset-charge dependent matrix elements.

II. OPTIMAL CONTROL THEORY

Quantum optimal control helps steer the time evolution of quantum systems to realize a desired state transfer, unitary operation, or readout protocol [37,38,46]. This is accomplished by optimizing a set of external control pulses $\{u_1(t), \ldots, u_M(t)\}$ which couple to the quantum system via

TABLE I. Relevant contributions to the cost functional. Indices k, j label the kth control pulse and the jth discretized time step.

Cost functional contribution	Explanation
$C_1 = 1 - \frac{1}{n^2}\lvert\mathrm{Tr}(U_t^\dagger U_f)\rvert^2$	Infidelity of realized unitary U_f relative to target gate U_t
$C_2 = \sum_{k,j} \lvert u_{kj} - u_{kj-1}\rvert^2$	First derivatives of the control parameters
$C_3 = \sum_{k,j} \lvert u_{kj}\rvert^2$	Control pulse energy

control operators $\{\mathcal{H}_1, \ldots, \mathcal{H}_M\}$ and, thus, change the system dynamics. The resulting time-dependent Hamiltonian has the general form

$$H(t) = \mathcal{H}_0 + \mathcal{H}_c(t), \tag{1}$$

where $\mathcal{H}_0$ is the intrinsic system Hamiltonian, also known as drift Hamiltonian, and $\mathcal{H}_c(t) = \sum_{k=1}^{M} u_k(t)\mathcal{H}_k$. The task of optimization is to determine a set of control pulses which minimize a cost functional $C[\{u_k(t)\}]$. This functional encodes the infidelity of the target process, and may include additional optimization constraints crucial for achieving realistic pulses.

We briefly review the pertinent contributions to the cost functional employed in our work. In the case of a target unitary operation U_t, acting on a closed system, the primary cost to be minimized is the gate infidelity

$$C_1 = 1 - F_c = 1 - \frac{1}{n^2}\lvert\mathrm{Tr}(U_t^\dagger U_f)\rvert^2. \tag{2}$$

Here, U_f is the unitary realized by a given set of control pulses and n denotes the dimension of the Hilbert space. Secondary optimization targets are utilized to smooth control pulses and limit their signal power so as to enable their implementation in the laboratory setting. In addition, cost penalties for occupation of certain higher-lying states help avoid leakage and ensure the validity of the inevitable Hilbert-space truncation. The individual contributions to the cost functional are summarized in Table I.

A common technique for cost minimization consists of gradient ascent pulse engineering (GRAPE) [1] based on explicit analytical expressions for the gradients of C with respect to each of the control pulses $u_k(t)$. Here, we utilize an automatic-differentiation optimizer [24] built on TensorFlow [47], which avoids the need for hard-coded analytical gradients of each new contribution to the cost functional. Appendix A covers the optimization process in greater detail.

To assess gate fidelities in the presence of dissipation and dephasing, we employ a Lindblad master equation description [48] of the quantum system weakly interacting with its environment,

$$\frac{d\rho}{dt} = -i[H(t), \rho] + \sum_l \gamma_l \left[c_l \rho c_l^\dagger - \frac{1}{2}\{c_l^\dagger c_l, \rho\}\right]. \tag{3}$$

Here, ρ is the reduced density matrix of the system and $\{c_l\}$ a set of jump operators capturing relaxation and dephasing processes with associated rates $\{\gamma_l\}$. The metric we use for open-system gate fidelity is given by

$$F_o = \frac{1}{n^2}\mathrm{Tr}(L_t^\dagger L_f), \tag{4}$$

where $L_t = U_t \otimes U_t^*$ is the target superoperator and L_f is the final superoperator defined by $L_f\rho(0) = \rho(t)$, i.e., it propagates a vectorized version of the system density matrix (see Appendix B for more details).

III. OPTIMIZED UNIVERSAL-GATE SET FOR THE HEAVY FLUXONIUM QUBIT

A. Single-qubit gates

The fluxonium qubit [32] is a promising superconducting circuit that may, in its most recent variants as "heavy fluxonium" [30,32], outperform the widely used transmon qubit [26,27]. In contrast to the transmon, heavy fluxonium combines strong Josephson nonlinearity with T_1 protection due to disjoint support of its lowest-lying localized wave functions. Heavy fluxonium devices utilize a decreased capacitive energy E_C, which emphasizes the localization of states [30,32]. Moreover, fluxonium eigenenergies are intrinsically insensitive to slow offset charge variations [49]. The protection granted by disjoint state support, however, also complicates the realization of universal gate operations by means of external microwave pulses: matrix elements for direct transitions between disjoint-support states remain exponentially suppressed. In this section, we show that optimal control algorithms can nevertheless yield efficient protocols for a universal gate set. Such protocols necessitate involvement of higher qubit levels, and we carefully evaluate fidelity limitations arising from temporary occupation of these states.

Experimentally, gates for heavy fluxonium have been realized by driving Raman transitions [30,31], which utilize intermediary higher-energy states to assist indirect transitions between the protected states. We will demonstrate a similar approach, exploiting the availability of intermediary state transitions using optimal control theory. The optimal-control formalism offers greater flexibility in terms of pulse shape, and yields fast, high-fidelity single-qubit gates with gate times below 100 ns and fidelities exceeding 99.9%. We obtain optimized pulse shapes for X, H, and T gates, thereby establishing a blueprint for realizing arbitrary single-qubit gate operations.

As typical in circuit QED [50,51], each gate is realized by a microwave pulse applied to a transmission-line resonator which, in turn, is coupled to the qubit. The corresponding Hamiltonian for this driven, generalized Jaynes-Cummings model is

$$H_{JC} = \sum_l \epsilon_l \lvert l\rangle\langle l\rvert + \omega_r a^\dagger a + \sum_{l,l'} g_{ll'} \lvert l\rangle\langle l'\rvert (a^\dagger + a) + u(t)(a^\dagger + a), \tag{5}$$

where ϵ_l, $\lvert l\rangle$ are the fluxonium eigenenergies eigenstates labeled by index l and ω_r is the resonator frequency. The relative coupling strengths are given by $g_{ll'} = g\langle l\rvert n_\phi \lvert l'\rangle$, where n_ϕ is the fluxonium charge operator. Fluxonium eigenenergies and

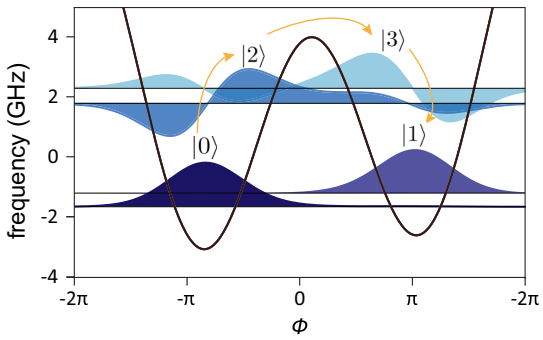

FIG. 1. First four fluxonium wave functions, slightly away from the flux sweet spot ($\Phi_{\text{ext}} = 0.45\Phi_0$). The lowest-lying states $|0\rangle$ and $|1\rangle$ are localized and have practically disjoint support. The auxiliary states $|2\rangle$ and $|3\rangle$ delocalize over both potential wells and serve as intermediate states for quantum gates. Gates involving population transfer between $|0\rangle$ and $|1\rangle$ such as X or H gates utilize the delocalized states for transfer across the potential barrier.

eigenstates are governed by the Hamiltonian [32,49]

$$\mathcal{H}_0 = 4E_C n_\phi^2 + \tfrac{1}{2}E_L\phi^2 - E_J\cos(\phi + 2\pi\,\Phi_{\text{ext}}/\Phi_0), \qquad (6)$$

in which E_C, E_L, and E_J denote the capacitive, inductive, and Josephson energies, respectively. Φ_{ext} is the external magnetic flux threading the loop formed by the junction and inductor. For the heavy-fluxonium qubit, we choose realistic device parameters $E_C/h = 0.5$ GHz, $E_L/h = 0.25$ GHz, and $E_J/h = 4.0$ GHz, and a flux working point slightly away from half-integer flux, $\Phi_{\text{ext}} = 0.45\Phi_0$. This places the system in the protected regime of nearly degenerate states $|0\rangle$ and $|1\rangle$ with disjoint support; see Fig. 1. (Operating the qubit away from the half-integer flux sweet spot increases sensitivity to dephasing from $1/f$ flux noise, which we monitor closely in our analysis).

Throughout this work, we focus on dispersive control of the qubit, in which the drive tone $u(t)$ steers dynamics within the qubit subsystem, but leaves the resonator state essentially unchanged. This allows us to exclude the resonator subspace from explicit simulation within the optimal-control algorithm. We verify in a separate simulation that the resonator state is unaffected by the drive tone, i.e., the average photon number obeys $\langle a^\dagger a\rangle \ll 1$ throughout the evolution. (It is interesting to note that incorporating resonator degrees of freedom and abandoning the dispersive regime offers additional ways for optimal control [52] which are beyond the scope of this paper). In the resulting driven-fluxonium Hamiltonian

$$H(t) = \mathcal{H}_0 + \mathcal{H}_c(t), \qquad (7)$$

we properly account for the fact that the drive on the qubit is filtered through the resonator. The dispersive coupling between qubit and resonator produces an effective drive on the qubit of the form

$$\mathcal{H}_c(t) = u(t)\sum_{l,l'}\frac{2g\omega_r\,\langle l|\,n_\phi\,|l'\rangle}{(\epsilon_l - \epsilon_{l'})^2 - \omega_r^2}\,|l\rangle\langle l'|\,. \qquad (8)$$

See Appendix C for details [53]. For our simulation, we consider a coupling strength and resonator frequency of $g/2\pi = 300$ MHz and $\omega_r/2\pi = 7.5$ GHz, respectively.

Using closed-system optimal control [54], we optimize the control pulse $u(t)$ to realize three different single-qubit gates: the Pauli-X gate, Hadamard gate, and the T gate,

$$X = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad H = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},\quad T = \begin{pmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{pmatrix}.$$

The latter two gates are known to form a universal set of single-qubit gates [55]. Optimization must balance two conflicting requirements: gate times t_g should be as short as possible to minimize the influence of dissipation and dephasing; at the same time, the maximum pulse amplitude $\max|u(t)|$ must remain small enough to avoid population of the resonator with unwanted photons. We find that pulses with t_g on the order of a few tens of nanoseconds satisfy these conditions while also producing gates with high fidelities. In addition to the cost-functional contribution C_1, quantifying the target-gate infidelity, we employ additional cost contributions C_2 and C_3 to limit the time derivatives and maximum amplitude of the pulse $u(t)$. Suppressing the maximum amplitude ensures that occupation of the resonator with spurious photons is minimized. The cost on pulse derivatives helps eliminate unnecessary high-frequency components of $u(t)$ and render the pulses as smooth as possible, which is important for experimental applications, since instruments generating these control fields have a finite impulse response.

The pulses we obtain have a gate duration of $t_g = 60$ ns and closed-system fidelities >99.9%. This choice of a relatively long gate duration is driven by limiting the overall pulse power. The panels of Figs. 2(a)–2(c) show the pulse $u(t)$ in the time domain and its discrete Fourier transform $\bar{u}(f)$ in the frequency domain. While interpreting optimized pulses is notoriously difficult, we note that general features of the three pulses and their frequency components can be given physical meaning. The Pauli-X and Hadamard gates both exhibit relatively well defined peaks in their Fourier spectra $\bar{u}(f)$, which coincide with the relevant transition frequencies among the lowest four levels primarily involved in the performance of the gate operation; see inset in Fig. 2(a). Visual inspection of $u(t)$ further reveals the staggered application of different frequency components. The initial and final ∼5 ns time windows are dominated by high-frequency components related to transferring the system from the $|0\rangle$, $|1\rangle$ subspace to the delocalized states $|2\rangle$, $|3\rangle$ (and back). The central time window between $t = 5$ ns and 55 ns shows involvement of the low-frequency components associated with the transfer between the intermediary states $|2\rangle$ and $|3\rangle$. The T gate, by contrast, exhibits a Fourier spectrum $\bar{u}(f)$ with only a single dominant frequency component corresponding to the $|1\rangle \leftrightarrow |3\rangle$ transition. This is plausible, since the T gate does not necessitate population transfer across the potential well. The transition peak for $|1\rangle \leftrightarrow |3\rangle$ facilitates the needed $e^{i\pi/4}$ phase accumulation for the $|1\rangle$ state.

Further evidence for this interpretation is given by Fig. 3, showing the probabilities for occupying the various fluxonium eigenstates as a function of time. For the Pauli-X gate and the Hadamard gate, occupation probabilities $p_l(t) = |\langle l|\psi(t)\rangle|^2$ are obtained for the example of initial qubit state $|\psi(0)\rangle = |0\rangle$. As expected, the X gate transfers population into the final state $|1\rangle$, while the H gate takes $|0\rangle$ into an equal

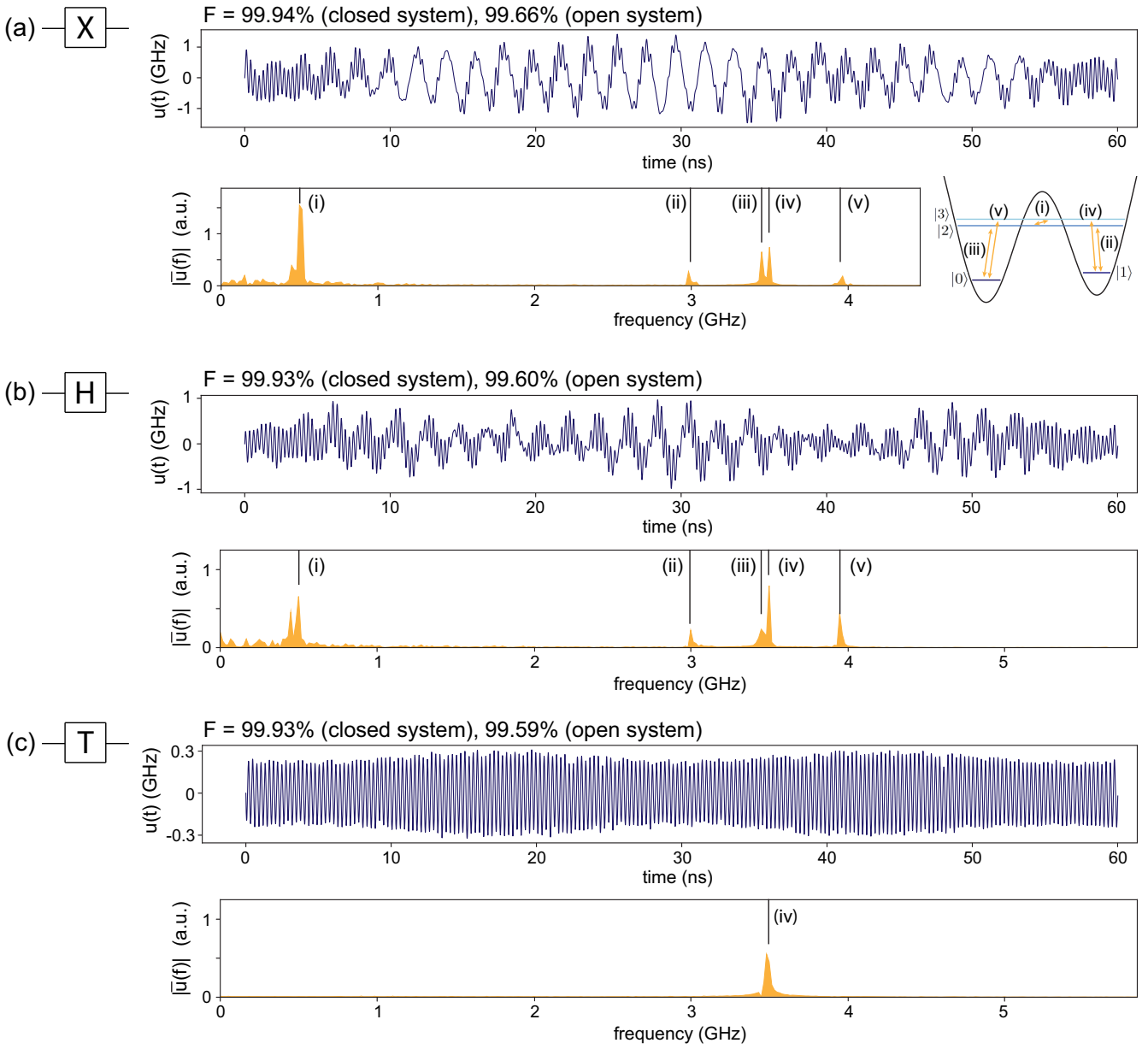

FIG. 2. High-fidelity single-qubit gates for heavy fluxonium. (a) Optimized pulse shape $u(t)$ and its discrete Fourier transform $\bar{u}(f)$ for the Pauli-X gate, achieving a gate fidelity of 99.94%. The Fourier transform exhibits distinct peaks that align with the transition frequencies among the involved levels (see inset). (b) Corresponding pulse data for the Hadamard gate with a fidelity of 99.933%. (c) Optimized pulse for the T gate with 99.933% gate fidelity. The Fourier transform shows a single peak centered at the $|1\rangle \leftrightarrow |3\rangle$ transition, serving to induce the required phase shift of $\pi/4$ for state $|1\rangle$.

superposition of $|0\rangle$ and $|1\rangle$. Both of these gates rely on the auxiliary states $|2\rangle$ and $|3\rangle$ to transfer population between the qubit computational states. By contrast, the T gate exhibits qualitatively dissimilar behavior since there is no need for state transfer across the fluxonium potential barrier. Instead, the much weaker control field only causes a small amount of intermediate population transfer from $|1\rangle$ to $|3\rangle$ for phase accumulation.

Operating the fluxonium qubit away from its half-integer flux sweet spot makes the gate fidelity more vulnerable to the detrimental effects of $1/f$ flux noise. At an external flux of $0.45\Phi_0$, we expect flux noise to limit the dephasing time T_φ and affect gate fidelities. To assess this issue, we follow a hybrid approach in which we evaluate dephasing rates due to classical $1/f$ noise, and then incorporate these rates in the Lindblad master equation. It must be emphasized that the latter step is a compromise we accept to avoid the heavier framework of non-Markovian master equations strictly appropriate for the inclusion of $1/f$ noise. This compromise is justified for a bound on the fidelity loss δF in the present context, as gate durations are small compared to relevant dephasing times, $t_g \ll T_\varphi$. (Note that the exponential decay modeled by the Lindblad treatment is more rapid than the actual Gaussian decay at short times). Following Refs. [35,56], we consider the Gaussian decay (up to logarithmic corrections) of the off-diagonal elements of the density matrix, and assign

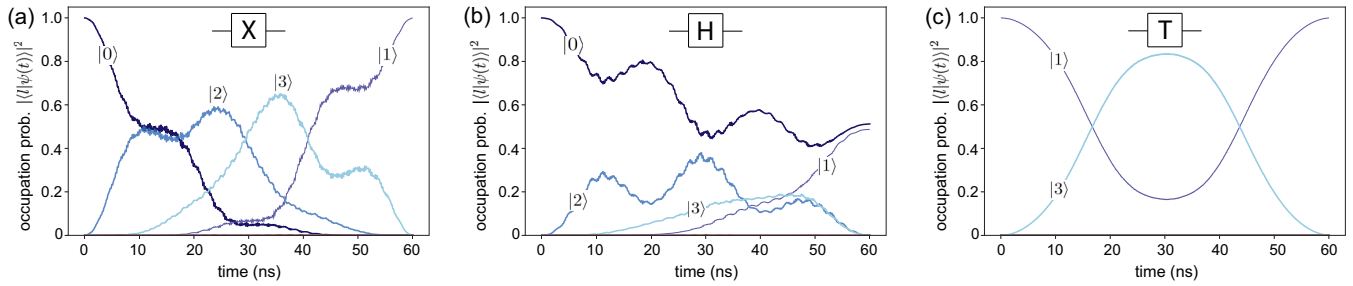

FIG. 3. Time evolution of state populations for 60 ns high-fidelity single-qubit gates. (a) The time evolution of states involved in the X gate shows state transfer between the qubit computational states via the delocalized $|2\rangle$ and $|3\rangle$ states. (b) For the H gate, states are transferred into an (approximately) equal superposition of $|0\rangle$ and $|1\rangle$. (c) In the T gate, the $|1\rangle$ state acquires an additional phase due to temporary state transfer into the $|3\rangle$ state.

the standard deviation as the effective dephasing time. The leading-order result from inspection of the density-matrix element $\rho_{ll'}$ is given by

$$1/\left(T_\varphi^{\Phi_{\rm ext}}\right)_{ll'} = A_{\Phi_{\rm ext}} \left|\partial_{\Phi_{\rm ext}} \omega_{ll'}\right| \sqrt{2|\ln \omega_{\rm ir} t|}. \tag{9}$$

Here, $A_{\Phi_{\rm ext}} = 1\mu\Phi_0$ is the flux-noise amplitude [57] and $\omega_{ll'}$ the frequency difference between fluxonium states $|l\rangle$ and $|l'\rangle$. In our calculations, we use $\omega_{\rm ir}/2\pi = 1$ Hz as the low-frequency cutoff and $t = 10$ μs as the measurement time scale [35]. For the heavy-fluxonium parameters stated above, the extracted dephasing rates are of the order of ~ 1 μs, and specifically $(T_\varphi^{\Phi_{\rm ext}})_{10} = 3.1$ μs for the two computational states.

Gate fidelities are also negatively affected by depolarization processes. Here, we consider dielectric surface loss as a likely candidate for limiting the T_1 time. While direct transitions among the computational states $|0\rangle$ and $|1\rangle$ are exponentially suppressed due to their disjoint support, transitions involving the delocalized levels $|2\rangle$ and $|3\rangle$ can occur. The corresponding transition rates are given by

$$\gamma_{ll'} = \Gamma |\langle l|n_\phi|l'\rangle|^2, \tag{10}$$

where we fix the rate constant Γ by using the estimate $1/\gamma_{02} = 50$ μs, a realistic intrawell decay time observed in experiments using similar device parameters [58] and further supported by dielectric loss theory [33].

Since the most relevant noise channels give rise to decoherence times that are about 10^2 times larger than t_g, we expect that open-system simulation using the optimized pulses should only lead to small changes in gate fidelities. In our calculation, we use a master equation of the form

$$\frac{d\rho(t)}{dt} = -i[H(t), \rho(t)] + \left(\mathbb{D}[c_0] + \sum_{l<l'} \mathbb{D}[c_{ll'}]\right)\rho(t), \tag{11}$$

where dephasing due to flux noise is captured by the diagonal jump operator $c_0 = \sum_l \sqrt{(\gamma_\varphi)_{l0}}\,|l\rangle\langle l|$ and depolarization due to dielectric loss by $c_{ll'} = \sqrt{\gamma_{ll'}}\,|l\rangle\langle l'|$. The Lindblad damping superoperator has the usual form $\mathbb{D}[c]\rho = c\rho c^\dagger - \frac{1}{2}\{c^\dagger c, \rho\}$. We calculate the resulting open-system gate fidelities by means of the expressions detailed in Appendix B. The resulting open-system fidelities for the X, H, and T gates are 99.66%, 99.60%, and 99.59%, respectively. This should be compared to the corresponding closed-system fidelities of 99.94%, 99.93%, and 99.93%.

B. Controlled-Z gate

To obtain a set of gates universal for multiqubit unitaries, we demonstrate an optimized controlled-Z (CZ) gate,

$$\mathrm{CZ} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

with a gate time of $t_g = 60$ ns. We consider two heavy-fluxonium qubits with identical circuit parameters of the same values as in Sec. III A, but biased by different magnetic fluxes $\Phi_{\rm ext,1} = 0.45\Phi_0$ (target qubit) and $\Phi_{\rm ext,2} = 0.455\Phi_0$ (control qubit). Biasing the qubits by this small flux offset results in coupling-induced energy shifts of $\sim$10 MHz for higher level, noncomputational states, which can be used for state entanglement. The interaction leaves computational states essentially unshifted due to suppression of the $\langle 0|n_\phi|1\rangle$ matrix element.

We couple each qubit to a shared resonator through a small coupling capacitor. The resulting Hamiltonian

$$H(t) = \mathcal{H}_0 + u_1(t)\mathcal{H}_1 + u_2(t)\mathcal{H}_2 \tag{12}$$

generalizes Eq. (7). Here, $\mathcal{H}_0 = \mathcal{H}_0^{(1)} + \mathcal{H}_0^{(2)} + \mathcal{H}^{(1,2)}$, where $\mathcal{H}_0^{(i)}$ are the Hamiltonians for the two fluxonia ($i = 1, 2$) and $\mathcal{H}_i$ are the dispersively filtered drives acting on each qubit, in the form given in Eq. (8). The qubits are driven by two separate pulses $u_1(t)$ and $u_2(t)$ for target and control, respectively. Due to the coupling to a shared resonator, there is an effective mutual coupling $\mathcal{H}^{(1,2)}$ between the two fluxonium qubits that allows for entanglement generation; see Appendix C for details.

As shown in Fig. 4(a), optimal-control theory yields pulses $u_1(t)$ and $u_2(t)$ that activate a two-fluxonium CZ gate. As in the case of a single fluxonium qubit, we employ cost contributions C_1, C_2, and C_3 in the optimization. The bottom panel of Fig. 4(a) monitors the system time evolution in terms of the occupation probabilities of participating states. We have confirmed that the case with $|11\rangle$ acting as initial state shows the largest amount of intermediate population transfer. This is consistent with the fact that this state must acquire a phase factor of $e^{i\pi}$, accomplished by the observed excursion into states $|13\rangle$, $|31\rangle$, and $|33\rangle$. Like the T gate, only transitions between states $|1\rangle$ and $|3\rangle$ are necessary to accumulate phase factors. Figure 4(b) depicts the gate unitary achieved by optimization. U_f is represented in the product basis $|ml\rangle$, with

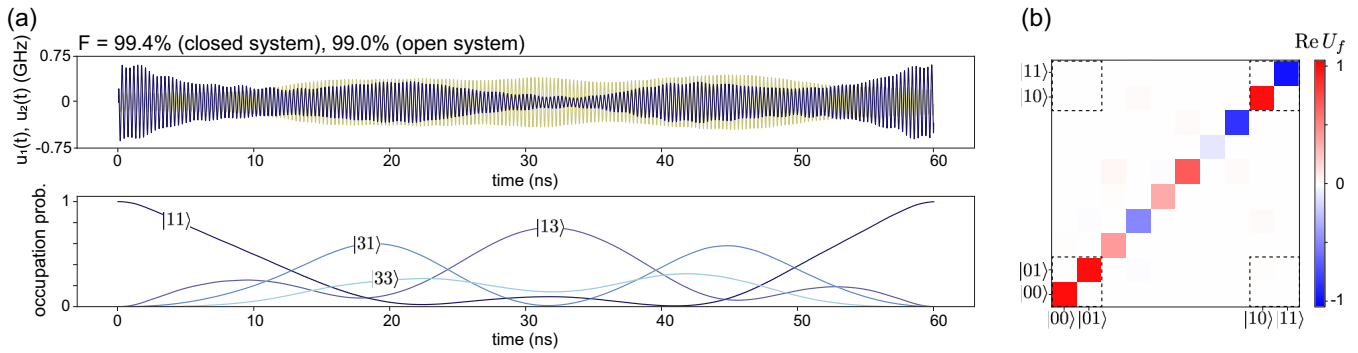

FIG. 4. Controlled-Z gate for two heavy-fluxonium qubits with a gate time of 60 ns [$\Phi_{\text{ext},1} = 0.45\Phi_0$ (target qubit) and $\Phi_{\text{ext},2} = 0.455\Phi_0$ (control qubit)]. (a) The top panel shows optimized pulses acting on the target and control qubit, $u_1(t)$ and $u_2(t)$, respectively, achieving a closed-system fidelity of 99.4% and open-system fidelity of 99.0%. The cost functionals used are C_1, C_2, and C_3. The bottom panel shows occupation probabilities of system eigenstates $|ml\rangle$, with $|11\rangle$ chosen as the initial state. $|11\rangle$ undergoes the significant population to intermediate states so as to induce a phase of $e^{i\pi}$, as required for the CZ gate. (b) Real part of the resulting unitary U_f achieved by optimization, showing levels $|0l\rangle$ with $0 \leqslant l \leqslant 7$, $|10\rangle$, and $|11\rangle$. Matrix elements between states in the computational subspace are marked by dashed squares.

m and l labeling control and target qubit levels, respectively. The relevant elements in the 4×4 computational subspace are marked by dashed rectangles and have entries which closely match the controlled-Z target unitary. Overall, the optimized pulse trains realize the CZ gate with a closed-system fidelity of 99.4% using the same gate time $t_g = 60$ ns. Open-system simulations including noise contributions from $1/f$ flux noise and dielectric loss result in an open-system fidelity of 99.0%.

Our optimal-control results for the CZ gate may be compared to the recent work by Nesterov *et al.* [59]. The setup in that work differs in the utilization of *direct* capacitive or inductive coupling between the fluxonia, which are then driven without shared resonator by a microwave tone with a Gaussian envelope. The pulse is optimized over the amplitude and drive frequency, rather than using a general-purpose optimal-control package. For the same gate time of 60 ns and direct capacitive coupling, they report a similar closed-system fidelity of 99.3% employing a single pulse on only one qubit.

IV. OPTIMIZED SINGLE-QUBIT GATE SET FOR THE 0-π QUBIT

The 0-π circuit [29,34–36] further extends the protection afforded by the heavy-fluxonium qubit by combining exponential suppression of *both* relaxation and dephasing due to disjoint wave-function support and robust ground-state degeneracy. We briefly review the physics of the 0-π circuit and the parameters required for its protected regime. The circuit consists of two superinductors of inductance E_L, two Josephson junctions with Josephson energy E_J and junction capacitance C_J, and two large shunt capacitors C. The ideal 0-π Hamiltonian reads

$$\mathcal{H}_0 = \frac{(q_\theta - n_g)^2}{2C_\theta} + \frac{q_\phi^2}{2C_\phi} + E_L\phi^2 - 2E_J\cos(\theta)\cos(\phi - \pi\,\Phi_{\text{ext}}/\Phi_0), \qquad (13)$$

where n_g is the offset charge and $q_\theta = 2en_\theta$, $q_\phi = 2en_\phi$ are the charge operators canonically conjugate to the two degrees of freedom θ and ϕ. The effective capacitances associated with these two variables are $C_\theta = 2(C + C_J) + C_g$ and $C_\phi = 2C_J + C_g$, where C_g is a small capacitance due to coupling to ground and external voltage lines [36]. The external magnetic flux threading the circuit loop is denoted Φ_{ext}.

For the 0-π qubit to realize the desired intrinsic protection, circuit parameters must satisfy several conditions. To achieve localization along the θ axis, the effective mass in θ needs to be heavy compared to that in ϕ direction, $C_\theta \ll C_\phi$, and the local potential wells deep enough to hold localized states, $e^2/2C_\theta \ll E_J$. The latter condition also renders the qubit charge-noise insensitive. To suppress sensitivity to flux noise, wave functions should be delocalized along the ϕ axis, obtained when $E_L \ll E_J$, $e^2/2C_\phi$. Achieving this parameter regime remains experimentally challenging. Here, we choose an "optimistic" parameter set previously considered in Groszkowski *et al.* [35], namely $E_L/h = E_C/h = 40$ MHz, $E_J/h = 10$ GHz, and $E_{CJ}/h = 20$ GHz. This choice provides an appropriate amount of qubit protection. The eigenspectrum as a function of external flux Φ_{ext} is shown in Fig. 5(a) along with several eigenfunctions. The lowest two, $|0\rangle$ and $|1\rangle$, span the computational subspace and are localized along $\theta = 0$ and $\theta = \pi$, respectively. States higher up in the spectrum, such as $|13\rangle$ and $|14\rangle$, are delocalized in the θ direction and will play an important role in the gate protocols.

Similar to the situation with the heavy-fluxonium qubit, disjoint support of the computational basis states in the 0-π qubit provides intrinsic protection from decoherence, but inevitably also prevents one from driving direct transitions between the two qubit states. Di Paolo *et al.* [36] achieved gate operations between states indirectly via a square voltage pulse that drives transitions via intermediate higher excited levels. Depending on device parameters, this strategy results either in an X gate or a Hadamard gate, but does not readily yield a gate set universal for single-qubit operations. For our optimal-control search, we consider the more conventional method of dispersively coupling the 0-π qubit to a resonator via n_θ, and driving the qubit via this resonator with a microwave pulse. Together, the drift and control Hamiltonian for 0-π acquire a form analogous to that encountered for heavy fluxonium in

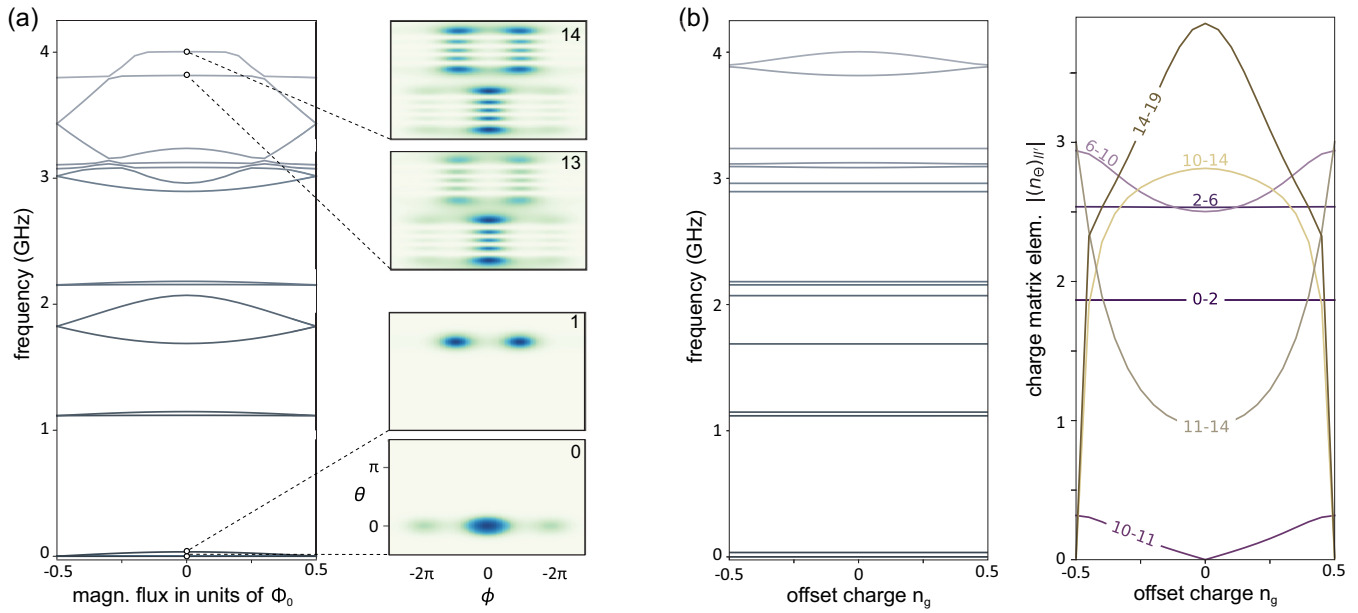

FIG. 5. Spectrum and matrix elements of the 0-π qubit. (a) Lowest 14 eigenenergies vs magnetic flux, and select 0-π eigenfunctions at $\Phi_{\text{ext}} = 0$ and offset charge $n_g = 0.25$. The lowest two eigenfunctions, forming the computational subspace, are localized around $\theta = 0$ and $\theta = \pi$, and are nearly degenerate. The top two eigenfunctions shown are delocalized states which occupy both potential wells, and serve as auxiliary states in gate operations. (b) Eigenenergies and n_θ charge matrix elements vs offset charge n_g. As eigenstates start delocalizing along θ, eigenenergies become weakly dependent on offset charge. Charge matrix elements show significant dependence on n_g for transitions between delocalized high-energy states, e.g., $|11\rangle \rightarrow |14\rangle$. (Parameters: $E_L/h = E_C/h = 40$ MHz, $E_J/h = 10$ GHz, and $E_{CJ}/h = 20$ GHz).

the previous section,

$$H(t) = \mathcal{H}_0(n_g) + \mathcal{H}_c(t). \tag{14}$$

The control Hamiltonian $\mathcal{H}_c(t)$ takes the form of Eq. (8), in which we take the filtered drive to couple to the θ degree of freedom (i.e., n_ϕ is replaced with n_θ). Employing optimal control to find the appropriate pulse shapes $u(t)$ gives sufficient flexibility for realizing a variety of single-qubit gates.

However, one challenge concerning 0-π gates which has not previously been discussed is revealed by Fig. 5(b), showing the dependence of charge matrix elements $(n_\theta)_{jj'} = \langle j|n_\theta|j'\rangle$ on the offset charge n_g. Among low-lying, θ-localized states, these matrix elements are practically n_g insensitive as expected. By contrast, as higher-energy states start delocalizing in the θ direction, offset-charge dependence of matrix elements becomes significant. This offset-charge sensitivity may affect gate operations which utilize higher-energy states as a means to transfer probability amplitude between the $\theta = 0$ and $\theta = \pi$ wells. The problem is exacerbated by the fact that offset charge is not controlled in experiments and is subject to significant fluctuations due to $1/f$ charge noise [60,61]. Our strategy is thus to steer the optimizer towards control solutions that are maximally insensitive to offset-charge fluctuations. Employing a scheme analogous to that previously used by Khani *et al.* [62], we have enhanced the optimal-control code to allow for drift and control Hamiltonians to vary from iteration to iteration, allowing us to choose random values of offset charge (using a uniform distribution over $0 \leqslant n_g < 1$) for each individual iteration of the optimizer. Directly applying the gradients from each iteration results in a stochastic-gradient-descent [63,64] process. With careful tuning of cost-function weights, this process converges to an average solution balancing all possible values of n_g.

A second challenge concerns the inevitable presence of disorder in circuit components which can lead to spurious coupling to a harmonic, low-energy degree of freedom, the ζ mode [34–36]. To avoid the overhead of a significant increase in Hilbert space dimension, we apply the optimal-control formalism to the ideal 0-π system, and verify subsequently that weak coupling to the ζ mode does not significantly reduce gate fidelities.

Figure 6 presents the results from optimal-control theory for the three single-qubit gates X, H, and T. For all 0-π gates, we again choose a gate duration of $t_g = 60$ ns, which aims to balance, on one hand, gate fidelity benefiting from short gate times and, on the other hand, overall pulse power which decreases as gate duration is increased. Both the X and Hadamard gate [Figs. 6(a) and 6(b)] require probability-amplitude transfer between computational-basis states, and are seen to result in similar level-population dynamics accessing excited states $|\geqslant 10\rangle$. Delocalization of these states in the θ variable enables the population transfer between the computational states, but also temporarily lifts the protection granted by disjoint-support wave functions.

For both X and H gates, the plots of occupation probability amplitude $p_j = |\langle j|\psi(t)\rangle|^2$ show initial and final phases of state transfer out of and into the computational-basis subspace. The sequences of either even- or odd-numbered levels reflect transitions among states centered at $\theta = 0$ and π, respectively. (The states $|4\rangle$ and $|5\rangle$ do not contribute due to the lack of connecting matrix elements). Beyond these initial and final phases, the dynamics is dominated by an extended intermediate phase during which higher-excited states participate.

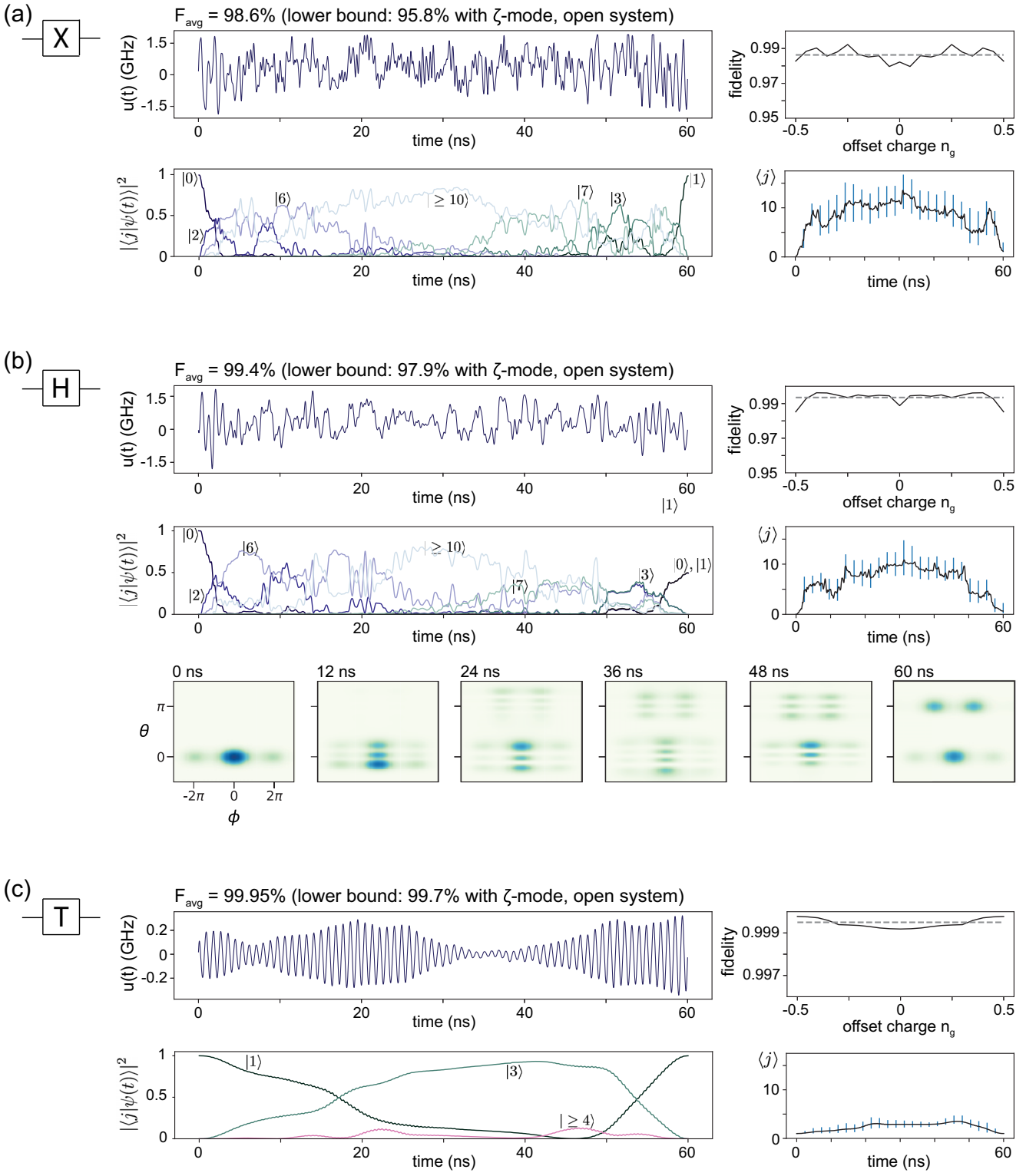

FIG. 6. Optimized pulses for 0-π single-qubit gates at $\Phi_{\text{ext}} = 0$. For X gate, Hadamard gate, and T gate, (a)–(c) show panels with pulse trains $u(t)$, fidelities vs offset charge, occupation probabilities, and mean and standard deviation of occupied levels (using $n_g = 0.25$). Fidelity variations vs offset charge are observed to be small compared to the average fidelity. The additional bottom panel in (b) depicts time evolution snapshots of 0-π wave functions for the optimized Hadamard gate.

As expected, these states are delocalized in θ as illustrated by the intermediate-time evolution snapshots for the Hadamard gate [Fig. 6(b)]. As depicted in the plots of the average level occupied and the standard deviation (calculated separately for levels above and below the average), the intermediate-phase dynamics is not readily interpreted as a sequence of transitions among higher levels, but rather involves evolution among superpositions of such states.

Pulse shape and population dynamics are qualitatively different for the T gate, which does not require transfer of amplitude from the $\theta = 0$ to the $\theta = \pi$ well; see Fig. 6(c). Instead, the $e^{i\pi/4}$ phase accumulation required for the $|1\rangle$ state is obtained by temporary occupation of state $|3\rangle$ and subsequent backtransfer into $|1\rangle$.

The closed-system fidelities, averaged over offset charge n_g, are 98.6% for the X gate, 99.4% for the H gate, and 99.95% for the T gate. The sub-99.9% fidelities for the X and H gates can be attributed to the significant offset-charge dependence of the transition matrix elements $(n_\theta)_{jj'}$ between high-energy delocalized states; see Fig. 5(b). Since pulses for X and H gates require occupation of these high-energy auxiliary states intermediately, optimization must compromise between higher-fidelity solutions for different fixed offset charges n_g. The trade-off is most significant between the offset charge values $n_g = 0$ and 0.5. As shown in Fig. 5(b), certain transitions between delocalized states are symmetry forbidden at $n_g = 0$, such as the $|10\rangle \rightarrow |11\rangle$ transition. This forces the system to take a detour into higher-energy states such as $|13\rangle$ or $|14\rangle$. The situation is reversed for $n_g = 0.5$, where $|10\rangle \rightarrow |11\rangle$ is allowed, but $|10\rangle \rightarrow |14\rangle$ is strongly suppressed. The fidelity plots for the X and H gates in Figs. 6(a) and 6(b) are consistent with this trade-off, showing decreases in fidelity at $n_g = 0$ and 0.5. The resulting pulse shapes appear more complicated than their heavy-fluxonium counterparts and their Fourier transforms (not shown) do not exhibit the well-resolved peaks observed in the fluxonium case.

We next investigate the performance of the optimized pulses in the presence of circuit disorder—in particular, in C and L. Such disorder leads to spurious coupling to the harmonic ζ mode with mode frequency $\Omega_\zeta = \sqrt{8E_L E_{C_\zeta}}/\hbar$, where $E_{C_\zeta} = e^2/2C_\zeta$ and $C_\zeta = 2C + C_g \approx 2C$ [35,36]. The resulting Hamiltonian reads

$$H = H_{0-\pi} + \Omega_\zeta a^\dagger a + \sum_{j,j'} (\kappa_{jj'} |j\rangle\langle j'| a + \text{H.c.}), \tag{15}$$

where $|j\rangle$ are 0-π eigenstates and $\kappa_{jj'} = \kappa^\phi_{jj'} + i\,\kappa^\theta_{jj'}$ are coupling strengths defined by

$$\kappa^\theta_{jj'} = \frac{1}{2} E_{C\Sigma}\, dC \left(\frac{32E_L}{E_{C_\zeta}}\right)^{1/4} \langle j|n_\theta|j'\rangle\,, \tag{16}$$

$$\kappa^\phi_{jj'} = \frac{1}{2} E_L dE_L \left(\frac{8E_{C_\zeta}}{E_L}\right)^{1/4} \langle j|\phi|j'\rangle\,. \tag{17}$$

We assume the relative disorder in inductance and capacitance to be at the level of $dL = dC = 5\%$. Disorder also leads to an additional component to the drive that couples to the ζ mode,

$$n_\theta \rightarrow n_\theta - \beta n_\zeta\,, \tag{18}$$

where n_ζ is the charge-number operator for the ζ mode and $\beta = C\,dC/C_\zeta$ (see Appendix A of Ref. [36]).

As noted in Ref. [35], the coupling to the ζ mode opens up an unwanted shot-noise dephasing channel that is absent in the symmetric 0-π device, and can become a dominant source of dephasing. In addition to shot noise, we consider $1/f$ charge noise. Since the induced dephasing rates are offset-charge dependent, we consider a worst-case scenario by maximizing dephasing rates over n_g. Charge noise impacts the system at intermediate times of the gate protocol, when the system occupies unprotected high-energy states that are delocalized in the θ variable. We further take into account the effect of dissipation due to dielectric surface loss. Like in heavy-fluxonium, while direct transitions between $|1\rangle$ and $|0\rangle$ are suppressed due to disjoint support, transitions between excited states can still occur. Using realistic parameters and operating at the $\Phi_{\text{ext}} = 0$ sweet spot, we have confirmed that other noise channels such as critical-current fluctuations and flux noise are subdominant and do not lead to significant reductions of the fidelity.

For the open-system simulation, we employ the obtained pulses and evolve the system composed of 0-π and ζ mode under the Lindblad master equation [see Eq. (11)]. Shot noise is incorporated using absorption and relaxation rates $\kappa_\zeta n_{\text{th}}(\Omega_\zeta)$ and $\kappa_\zeta n_{\text{th}}(\Omega_\zeta) + 1$, respectively, where the ζ-mode thermal occupation $n_{\text{th}}(\Omega_\zeta)$ is 2.29 at a temperature of 15 mK and $1/\kappa_\zeta = 100\ \mu$s. Charge-noise dephasing is incorporated using dephasing rates $(\gamma_\varphi)_{j0}$, evaluated using the analog of Eq. (9), here with derivative evaluated with respect to n_g. Finally, we model dielectric decay rates in the same way as for fluxonium, i.e., we take $\gamma_{jj'} = \Gamma\,|\langle j|n_\theta|j'\rangle|^2$ and fix the rate constant Γ by taking $1/\gamma_{02} = 50\ \mu$s.

Based on this master-equation simulation, we obtain conservative lower bounds on average gate fidelities for the three single-qubit gates: 95.8% for X, 97.9% for H, and 99.7% for T. Evidently, inclusion of noise and coupling to the ζ mode mainly affects the X and H gates. This fidelity loss is primarily due to two factors: shot-noise dephasing induced by the ζ mode, as well as occupation of higher delocalized states with enhanced sensitivity to charge noise. (Note that lower bounds report the worst fidelity, reached for a particular n_g). We thus find that the main barrier to achieving high-fidelity single-qubit gates for 0-π is coupling to the ζ mode, offset-charge fluctuations, and dielectric surface loss.

V. CONCLUSION

In summary, we have used optimal control theory for implementing quantum gates for protected superconducting qubits whose computational states have practically disjoint support. We have presented optimal control pulses yielding a fully universal set of gates for the heavy-fluxonium qubit and a set universal for single-qubit gates for the 0-π qubit. Specifically, we considered the Pauli-X gate, Hadamard gate, and T gate with closed-system fidelities of >99.9% for heavy-fluxonium; likewise a controlled-Z gate with a closed-system fidelity of 99.4%. For the 0-π qubit, we implemented an enhanced optimal-control method by allowing offset charge to vary for each optimizer iteration. Applying gradients from each iteration results in a stochastic-gradient-descent process. This process converges to an average solution yielding a fidelity averaged over a range of n_g. We presented pulses with closed-system average fidelities of 98.6%, 99.4%, and 99.95% for X, H, and T, respectively. Remarkably, this method thus provides a way to find control pulses with good fidelities which are roughly insensitive to random offset charge changes.

All constructed gates represent a compromise between limiting drive powers to realistic values and mitigating the effects of noise. To assess the fidelity losses in the open system, we incorporated optimized pulses into a master equation treatment. For heavy fluxonium, $1/f$ flux dephasing and dielectric surface losses were the primary sources of decoherence. The resulting open-system fidelities obtained were >99% for single-qubit gates and 99.0% for the controlled-Z gate. For 0-π, shot-noise dephasing in the ζ mode, $1/f$ offset-charge dephasing, and dielectric surface loss were the most relevant noise sources. The resulting conservative lower bounds on average fidelities for 0-π + ζ mode were 95.8%, 97.9%, and 99.7% for X, H, and T.

Future work should consider combining drive pulses with the active cooling scheme proposed in Ref. [36], and constructing the more challenging 0-π multiqubit gates. Further extensions may also entail using open-system optimization algorithms [25,41] to partially mitigate gate-fidelity losses due to noise. These will involve significantly more computational overhead. We believe that optimal control provides a promising avenue toward universal gates on today's protected superconducting qubits.

ACKNOWLEDGMENTS

We thank Peter Groszkowski, Anjali Premkumar, and Agustin Di Paolo for valuable discussions. We gratefully acknowledge support from the Army Research Office through Grants No. W911NF-15-1-0421 and No. W911NF-19-1-0016.

M.A. and B.B. contributed equally to this work.

APPENDIX A: GRAPE OPTIMIZATION

In this Appendix we describe the GRAPE optimization scheme summarized in Sec. II and utilized throughout the paper. The optimizer iteratively adjusts the set of control fields $\{u_1(t), \ldots, u_M(t)\}$ such that the deviation of the realized unitary U_f from the target unitary U_t is minimized. For the numerical treatment, but also motivated by the finite resolution of arbitrary waveform generators, the gate time t_g is discretized into a large number N of small time intervals δt and corresponding discrete time instances $t_j = j\delta t$. The control amplitudes at these times, $u_{kj} = u_k(t_j)$, can be grouped into a vector $\mathbf{u} = (u_{kj}) \in \mathbb{R}^{NM}$, and form the set of optimization parameters.

The primary quantity to be optimized is the gate fidelity

$$F_c(\mathbf{u}) = \frac{1}{n^2}|\mathrm{Tr}(U_t^\dagger U_f(\mathbf{u})|^2, \tag{A1}$$

where U_t is the desired target unitary, $U_f(\mathbf{u})$ the unitary realized by means of the control field, and n the relevant Hilbert space dimension. The realized unitary is computed by a decomposition into short-time propagators,

$$U_f = U_N U_{N-1} \cdots U_1 U_0, \tag{A2}$$

where $U_j = \exp(-iH_j\delta t)$ describes the unitary evolution of the system from time t_j to $t_j + \delta t$, and

$$H_j = \mathcal{H}_0 + \sum_{k=1}^{M} u_{kj}\mathcal{H}_k \tag{A3}$$

is the Hamiltonian at time t_j.

In addition to the infidelity cost function $C_1(\mathbf{u}) = 1 - F_c(\mathbf{u})$, we also want to include additional costs to ensure that the total control-pulse energy is not too large and that the signals $u_j(t)$ do not vary too rapidly. We thus define a composite cost function

$$C(\mathbf{u}) = C_1(\mathbf{u}) + \alpha_2 C_2(\mathbf{u}) + \alpha_3 C_3(\mathbf{u}), \tag{A4}$$

in which C_2 and C_3 are the contributions quantifying pulse derivatives and energy, as shown in Table I. The coefficients α_m are the corresponding cost-function weights. (These weights may need to be adjusted intermittently throughout the course of optimization in order to help navigate toward higher fidelities and more desirable pulse shapes).

The most basic way of minimizing the cost functional $C(\mathbf{u})$ is to apply updates of $\mathbf{u}$ by following the opposite direction of the gradient,

$$\mathbf{u}_{p+1} = \mathbf{u}_p - \eta_p \frac{\partial C(\mathbf{u}_p)}{\partial \mathbf{u}_p}. \tag{A5}$$

Here, p denotes the pth iteration and η_p denotes the step size on the pth iteration. This gradient descent can be improved, e.g., by using ADAM [65], a "momentum-accelerated" gradient descent method that iteratively updates $\mathbf{u}$ using moving averages of both the gradient direction and magnitude. We implement ADAM with an exponentially decaying step size ("learning rate") given by $\eta_p = \eta_0 e^{-\beta p}$, in which the parameters η_0 and β, like the regularization coefficients α_2 and α_3, may need to be intermittently adjusted.

Gradients $\partial C(\mathbf{u}_p)/\partial \mathbf{u}_p$ are computed using automatic differentiation, a central tool in machine learning [45]. Rather than hard-coding analytical gradients, this algorithm decomposes $C(\mathbf{u}_p)$ into its computational graph of elementary operations, each having known derivatives. Starting at the outermost function C in the graph, these derivatives are evaluated iteratively in a back-propagation process. The gradient is obtained from these elementary derivatives according to the chain rule.

Throughout this paper, the initial pulses $\mathbf{u}_0$ are chosen as Gaussian white noise, thus suppressing any bias towards any particular pulse shape. The truncated Hilbert-space dimensions for the fluxonium and 0-π qubits are $n = 10$ and $n = 30$, respectively. Computations are carried out on a CPU with a step count of $N = 5000$ for the chosen 60 ns gate duration, amounting to a time step $\delta t = 12$ ps. The total number of iterations typically ranged from 1000–2500, resulting in runtimes of the order of a few hours.

APPENDIX B: OPEN-SYSTEM GATE FIDELITY

To assess the effects of dissipation on optimized gates, we employ a definition of open-system fidelity that correctly reduces to the closed-system fidelity when the coupling to the environment is eliminated. This allows for consistent

comparison between open and closed-system fidelities. The definition makes use of the density matrix in vectorized (coherence vector) form, stacking the rows of the $n\times n$ density matrix ρ into a $n^2\times 1$ vector $|\rho\rangle\!\rangle$,

$$\rho = \sum_{ij} \rho_{ij} |i\rangle \langle j| \longrightarrow |\rho\rangle\!\rangle = \sum_{ij} \rho_{ij} |i\rangle |j\rangle . \tag{B1}$$

The evolution of the density matrix is then written with the help of the $n^2\times n^2$ superoperator L as $|\rho(t)\rangle\!\rangle = L|\rho(0)\rangle\!\rangle$.

A consistent measure for the open-system gate fidelity is then given by

$$F_o = \frac{1}{n^2}\mathrm{Tr}(L_t^\dagger L_f), \tag{B2}$$

where L_t is the target superoperator and L_f is the final achieved superoperator. This metric correctly reduces to the expression of the closed-system trace fidelity $F_c = |\mathrm{Tr}(U_t^\dagger U_f/n)|^2$ when dissipation is switched off. To prove this, we first describe the action on vectorized states equivalent to matrix multiplication from the right and left on ρ,

$$A\rho = \sum_{ij} \rho_{ij} A |i\rangle \langle j| \longrightarrow \sum_{ij} \rho_{ij} A |i\rangle |j\rangle = A\otimes I|\rho\rangle\!\rangle,$$

$$\rho B = \sum_{ij} \rho_{ij} |i\rangle \langle j| B \longrightarrow \sum_{ij} \rho_{ij} |i\rangle B^T |j\rangle = I\otimes B^T|\rho\rangle\!\rangle.$$

In the case of closed-system dynamics, the density matrix is a pure state $|\psi\rangle\langle\psi|$, and the evolution reduces to

$$L_c\rho = L_c |\psi\rangle \langle\psi| = U |\psi\rangle \langle\psi| U^\dagger = U\rho U^\dagger, \tag{B3}$$

where U is the closed-system propagator. In the vectorized picture, this reads

$$L_c\rho = U\rho U^\dagger \longrightarrow U\otimes U^*|\rho\rangle\!\rangle. \tag{B4}$$

Therefore, computation of F_o yields

$$F_o = \frac{1}{n^2}\mathrm{Tr}\big(U_t^\dagger U_f \otimes U_t^T U_f^*\big) = \frac{1}{n^2}|\mathrm{Tr}(U_t^\dagger U_f)|^2 = F_c, \tag{B5}$$

consistent with the closed-system gate fidelity.

APPENDIX C: DISPERSIVELY FILTERED DRIVE AND QUDIT COUPLING

This Appendix briefly summarizes the derivation of the dispersive drive term [Eq. (8)] and the effective qudit-qudit coupling mediated by a resonator. The derivation follows as a slight generalization from Ref. [66] and is based on the dispersive Schrieffer-Wolff transformation.

Consider a generalized Jaynes-Cummings Hamiltonian describing two qudits coupled to a resonator, the latter driven by a microwave tone $u(t)$:

$$H = H_0 + V + u(t)(a^\dagger + a). \tag{C1}$$

Here, $H_0 = \omega_r a^\dagger a + H_0^{(1)} + H_0^{(2)}$ comprises the resonator and bare qudit Hamiltonians. $V = V^{(1)} + V^{(2)}$ describes the qudit-resonator coupling which is of the form

$$V^{(k)} = \sum_{ll'} g_{ll'}^{(k)} |l_k\rangle\langle l'_k| (a^\dagger + a). \tag{C2}$$

$|l_k\rangle$ are the eigenstates of qudit k and $g_{ll'}^{(k)} = g\langle l_k|n_\phi|l'_k\rangle$ the coupling matrix elements with overall strength g. In the dispersive regime, detunings are large compared to the coupling, i.e., $\lambda_{ll'}^{(k)} = g_{ll'}^{(k)}/\Delta_{ll'}^{(k)} \ll 1$ with $\Delta_{ll'}^{(k)} = \epsilon_l^{(k)} - \epsilon_{l'}^{(k)} - \omega_r$ denoting the detuning between qudit-k transition $l\to l'$ and the resonator and $\epsilon_l^{(k)}$ the lth eigenenergy of qudit k. For our second-order treatment, we only require the leading order of the Schrieffer-Wolff transformation generator,

$$S \equiv -i\sum_k \sum_{ll'} \big(\lambda_{ll'}^{(k)} a - \lambda_{l'l}^{(k)} a^\dagger\big) |l_k\rangle\langle l'_k| , \tag{C3}$$

with the second-order result

$$H' = H_0 + u(t)(a^\dagger + a) + V + [iS, H_0 + u(t)(a^\dagger + a)] + [iS, V] + \tfrac{1}{2}[S,[S, H_0 + u(t)(a^\dagger + a)]]. \tag{C4}$$

First, focus on the first-order term $[iS^{(k)}, u(t)(a^\dagger + a)]$. This term is the leading drive contribution on qudit k. Evaluating the commutator yields

$$[iS^{(k)}, u(t)(a^\dagger + a)] = 2g\omega_r u(t)\sum_{ll'} \frac{|l_k\rangle\langle l_k| n_\phi |l'_k\rangle\langle l'_k|}{\big(\epsilon_l^{(k)} - \epsilon_{l'}^{(k)}\big)^2 - \omega_r^2},$$

which captures the filtered drive on qudit k.

Second-order terms in Eq. (C4) become essential for calculation of the effective qudit-qudit coupling. The terms that lead to coupling are the "cross" commutators, $[iS^{(k)}, V^{(j)}]$ and $\frac{1}{2}[S^{(k)},[S^{(j)}, H_0]]$ $(k\neq j)$. Together, these generate the coupling term

$$\mathcal{H}^{(1,2)} = \frac{g^2}{2}\big(n_\phi^{(1)}\otimes \tilde{n}_\phi^{(2)} + \tilde{n}_\phi^{(1)}\otimes n_\phi^{(2)}\big), \tag{C5}$$

where

$$\tilde{n}_\phi^{(k)} \equiv 2\omega_r \sum_{ll'} \frac{\langle l_k| n_\phi |l'_k\rangle}{\big(\epsilon_l^{(k)} - \epsilon_{l'}^{(k)}\big)^2 - \omega_r^2} |l_k\rangle\langle l'_k| . \tag{C6}$$

The above coupling term is the natural generalization of the two-qubit coupling discussed in Ref. [67].

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

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

[3] P. de Fouquieres, S. G. Schirmer, S. J. Glaser, and I. Kuprov, J. Magn. Reson. **212**, 412 (2011).

[4] S. E. Sklarz and D. J. Tannor, Phys. Rev. A **66**, 053619 (2002).

[5] R. Eitan, M. Mundt, and D. J. Tannor, Phys. Rev. A **83**, 053426 (2011).

[6] C. Gollub, M. Kowalewski, and R. de Vivie-Riedle, Phys. Rev. Lett. **101**, 073002 (2008).

[7] R. Nigmatullin and S. G. Schirmer, New J. Phys. **11**, 105032 (2009).

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

[9] J. P. Palao and R. Kosloff, Phys. Rev. A **68**, 062308 (2003).

[10] Y. Maday and G. Turinici, J. Chem. Phys. **118**, 8191 (2003).

[11] A. Borzì, J. Salomon, and S. Volkwein, J. Comput. Appl. Math. **216**, 170 (2008).

[12] P. Ditz and A. Borzì, Comput. Phys. Commun. **178**, 393 (2008).

[13] T. Sleator and H. Weinfurter, Phys. Rev. Lett. **74**, 4087 (1995).

[14] S.-L. Zhu and Z. D. Wang, Phys. Rev. Lett. **89**, 097902 (2002).

[15] C. Zu, W.-B. Wang, L. He, W.-G. Zhang, C.-Y. Dai, F. Wang, and L.-M. Duan, Nature (London) **514**, 72 (2014).

[16] F. Schmidt-Kaler, H. Häffner, S. Gulde, M. Riebe, G. Lancaster, T. Deuschle, C. Becher, W. Hänsel, J. Eschner, C. Roos *et al.*, Appl. Phys. B **77**, 789 (2003).

[17] T. Monz, K. Kim, A. S. Villar, P. Schindler, M. Chwalla, M. Riebe, C. F. Roos, H. Häffner, W. Hänsel, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **103**, 200503 (2009).

[18] L. K. Castelano, E. F. de Lima, J. R. Madureira, M. H. Degani, and M. Z. Maialle, Phys. Rev. B **97**, 235301 (2018).

[19] R. W. Heeres, P. Reinhold, N. Ofek, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, Nat. Commun **8**, 94 (2017).

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

[21] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).

[22] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **102**, 090502 (2009).

[23] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **82**, 040305(R) (2010).

[24] N. Leung, M. Abdelhafez, J. Koch, and D. Schuster, Phys. Rev. A **95**, 042318 (2017).

[25] M. Abdelhafez, D. I. Schuster, and J. Koch, Phys. Rev. A **99**, 052327 (2019).

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

[27] J. A. Schreier, A. A. Houck, J. Koch, D. I. Schuster, B. R. Johnson, J. M. Chow, J. M. Gambetta, J. Majer, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. B **77**, 180502(R) (2008).

[28] A. Kitaev, arXiv:cond-mat/0609441.

[29] P. Brooks, A. Kitaev, and J. Preskill, Phys. Rev. A **87**, 052306 (2013).

[30] N. Earnest, S. Chakram, Y. Lu, N. Irons, R. K. Naik, N. Leung, L. Ocola, D. A. Czaplewski, B. Baker, J. Lawrence, J. Koch, and D. I. Schuster, Phys. Rev. Lett. **120**, 150504 (2018).

[31] U. Vool, A. Kou, W. C. Smith, N. E. Frattini, K. Serniak, P. Reinhold, I. M. Pop, S. Shankar, L. Frunzio, S. M. Girvin, and M. H. Devoret, Phys. Rev. Appl. **9**, 054046 (2018).

[32] V. E. Manucharyan, J. Koch, L. I. Glazman, and M. H. Devoret, Science **326**, 113 (2009).

[33] L. B. Nguyen, Y.-H. Lin, A. Somoroff, R. Mencia, N. Grabon, and V. E. Manucharyan, Phys. Rev. X **9**, 041041 (2019).

[34] J. M. Dempster, B. Fu, D. G. Ferguson, D. I. Schuster, and J. Koch, Phys. Rev. B **90**, 094518 (2014).

[35] P. Groszkowski, A. D. Paolo, A. L. Grimsmo, A. Blais, D. I. Schuster, A. A. Houck, and J. Koch, New J. Phys. **20**, 043053 (2018).

[36] A. Di Paolo, A. L. Grimsmo, P. Groszkowski, J. Koch, and A. Blais, New J. Phys. **21**, 043002 (2019).

[37] T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, and S. J. Glaser, J. Magn. Reson. **167**, 68 (2004).

[38] K. Kobzar, T. E. Skinner, N. Khaneja, S. J. Glaser, and B. Luy, J. Magn. Reson. **170**, 236 (2004).

[39] K. Kobzar, T. E. Skinner, N. Khaneja, S. J. Glaser, and B. Luy, J. Magn. Reson. **194**, 58 (2008).

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

[41] S. Boutin, C. K. Andersen, J. Venkatraman, A. J. Ferris, and A. Blais, Phys. Rev. A **96**, 042315 (2017).

[42] J. R. Johansson, P. D. Nation, and F. Nori, Comput. Phys. Commun. **184**, 1234 (2013).

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

[44] H. J. Hogben, M. Krzystyniak, G. T. P. Charnock, P. J. Hore, and I. Kuprov, J. Magn. Reson. **208**, 179 (2011).

[45] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, J. Mach. Learn. Res. **18**, 1 (2018).

[46] Q. M. Chen, R. B. Wu, T. M. Zhang, and H. Rabitz, Phys. Rev. A **92**, 063415 (2015).

[47] M. Abadi *et al.*, arXiv:1603.04467.

[48] H. P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).

[49] J. Koch, V. Manucharyan, M. H. Devoret, and L. I. Glazman, Phys. Rev. Lett. **103**, 217004 (2009).

[50] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. Schoelkopf, Nature (London) **431**, 162 (2004).

[51] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **69**, 062320 (2004).

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

[53] We intentionally avoid a direct drive on the qubit which would require Purcell filters and complicate experimental scalability.

[54] Here, "closed system" refers to a system isolated from the environment and should not be confused with closed-loop optimization which we do not consider.

[55] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).

[56] G. Ithier, E. Collin, P. Joyez, P. J. Meeson, D. Vion, D. Esteve, F. Chiarello, A. Shnirman, Y. Makhlin, J. Schriefl, and G. Schön, Phys. Rev. B **72**, 134519 (2005).

[57] M. D. Hutchings, J. B. Hertzberg, Y. Liu, N. T. Bronn, G. A. Keefe, M. Brink, J. M. Chow, and B. L. T. Plourde, Phys. Rev. Appl. **8**, 044003 (2017).

[58] A. Premkumar (private communication).

[59] K. N. Nesterov, I. V. Pechenezhskiy, C. Wang, V. E. Manucharyan, and M. G. Vavilov, Phys. Rev. A **98**, 030301(R) (2018).

[60] A. B. Zorin, F.-J. Ahlers, J. Niemeyer, T. Weimann, H. Wolf, V. A. Krupenin, and S. V. Lotkhov, Phys. Rev. B **53**, 13682 (1996).

[61] A. Pourkabirian, M. V. Gustafsson, G. Johansson, J. Clarke, and P. Delsing, Phys. Rev. Lett. **113**, 256801 (2014).

[62] B. Khani, S. T. Merkel, F. Motzoi, J. M. Gambetta, and F. K. Wilhelm, Phys. Rev. A **85**, 022306 (2012).

[63] N. Ketkar, Stochastic gradient descent, in *Deep Learning with Python: A Hands-on Introduction*, edited by N. Ketkar (Apress, New York, 2017), pp. 113–132.

[64] M. Zinkevich, M. Weimer, L. Li, and A. J. Smola, Parallelized stochastic gradient descent, in *Advances in Neural Information Processing Systems 23*, edited by J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (Curran Associates, Inc., Red Hook, NY, 2010), p. 2595.

[65] D. P. Kingma and J. Ba, arXiv:1412.6980.

[66] G. Zhu, D. G. Ferguson, V. E. Manucharyan, and J. Koch, Phys. Rev. B **87**, 024510 (2013).

[67] J. Majer, J. M. Chow, J. M. Gambetta, J. Koch, B. R. Johnson, J. A. Schreier, L. Frunzio, D. I. Schuster, A. A. Houck, A. Wallraff, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Nature (London) **449**, 443 (2007).