Supplementary Information for Universal Quantum Control through Deep Reinforcement Learning

Murphy Yuezhen Niu[1*,2], Sergio Boixo[2*], Vadim N. Smelyanskiy[2*], Hartmut Neven[2*]

[1]Department of Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA
[2]Google, 340 Main Street, Venice Beach, California 90291, USA

*To whom correspondence should be addressed;

A Gmon Hamiltonian and Control Noise

In this section, we derive the two-qubit gmon Hamiltonian in the perturbative regime. It is obtained by quantizing a phenomenological classical Hamiltonian for nonlinear LC circuits containing Josephson junctions, as explained below.

The supercurrent across a Josephson junction between two superconducting media, in the absence of any external field, persists and undergoes quantum oscillations once connected to a capacitor. Intuitively, it is the quantum tunneling across the junction that drives the supercurrent. Such a macroscopic quantum effect is associated with a phenomenological classical circuit model, in which the effective voltage across the Josephson junction comes from the changing phase difference ϕ across the junction, i.e.,

$$V = \frac{\phi_0}{2\pi}\frac{d\phi}{dt}\,, \tag{1}$$

where $\phi_0 = h/2e$ and e is the electron charge. Moreover, the supercurrent depends on the phase

difference as

$$I = I_c \sin(\varphi) \ , \tag{2}$$

where the critical current I_c is a parameter characterizing the junction. Once we connect the junction to a capacitor of capacitance C, we obtain a nonlinear quantum oscillation. Writing the potential energy of the inductor according to

$$U = \int IV dt = \frac{\phi_0 I_c}{2\pi} \int \sin(\phi) \frac{d\phi}{dt} dt = -E_J \cos\phi, \tag{3}$$

the energy contribution from the capacitor

$$T = \frac{1}{2} C V^2 = \frac{1}{2} C (\frac{\phi_0}{2\pi} \dot{\phi})^2 \tag{4}$$

can be regarded as the oscillator's kinetic energy term. The Hamiltonian for the Josephson junction oscillator, to lowest order in ϕ, is therefore

$$H_0 = U_0 + T = \alpha \phi^2 + \beta \dot{\phi}^2 \ , \tag{5}$$

with α and β determined by the effective capacitance and inductance of the superconducting circuit. For example, we can add an additional inductor to the Josephson LC circuit to change the circuit's potential energy.

To analyze the quantum dynamics of this phenomenological model, we perform second quantization of ϕ by treating it as one of the two quadratures of a bosonic field:

$$\hat{\phi} = \kappa \frac{\hat{a}^\dagger + \hat{a}}{\sqrt{2}} \ , \tag{6}$$

where $\kappa = (\beta/\alpha)^{1/4}$. The conjugate basis $\dot{\hat{\phi}}$ takes the form

$$\dot{\hat{\phi}} = \kappa^{-1} \frac{\hat{a}^\dagger - \hat{a}}{\sqrt{2} i} \ . \tag{7}$$

Under this quantization, the harmonic part of the Josephson LC circuit's Hamiltonian is

$$\hat{H}_0 = \omega_0 \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) \ , \tag{8}$$

where we have set $\hbar = 1$ and defined $\omega_0 = \sqrt{\kappa}$.

Now let us include the lowest order anharmonic term from Eq. (3) to the Josephson LC circuit's Hamiltonian:

$$\hat{H}_1 = \gamma \hat{\phi}^4 = \frac{\gamma \kappa^4}{4} (\hat{a} + \hat{a}^\dagger)^4 \tag{9}$$

$$= \frac{\eta}{8} \left[\left(\hat{a}^2 + (\hat{a}^\dagger)^2 \right)^2 + \left(\hat{a}^2 + (\hat{a}^\dagger)^2 \right) (2\hat{n} + 1) + (2\hat{n} + 1) \left(\hat{a}^2 + (\hat{a}^\dagger)^2 \right) + (2\hat{n} + 1)^2 \right] , \tag{10}$$

with $\hat{n} = \hat{a}^\dagger \hat{a}$, γ determined by the configurations of additional inductor or capacitor elements added to the Josephson LC circuit, and η is the anharmonicity parameter.

Two superconducting qubits can interact by capacitive coupling between two Josephson LC circuits. To the lowest order, such capacitive coupling energy depends on $\dot{\phi}_1$ and $\dot{\phi}_2$ as:

$$\hat{H}_2 = -\alpha'(t) \dot{\phi}_1 \dot{\phi}_2 = -\frac{\alpha'(t)}{2\kappa_1 \kappa_2} (\hat{a}_1^\dagger - \hat{a}_1)(\hat{a}_2^\dagger - \hat{a}_2) \ , \tag{11}$$

where the coupling strength $\alpha'(t)$ is tunable.

Additionally, we also have capacitive coupling between a strong non-depleted microwave pulse and our quantized bosonic modes that is given by

$$T_{\text{micro}} = \frac{1}{2} C V_{\text{drive}} V_{\text{qubit}} \ . \tag{12}$$

The microwave drive-induced voltage is proportional to the field amplitude,

$$V_{\text{drive}} = f(t) e^{-i(\omega t - \varphi)} + f(t) e^{i(\omega t + \varphi)} \tag{13}$$

with c-number $f(t)$ determined by the amplitude of the microwave drive, φ being the phase of the drive and ω being the frequency of the drive. The quantized voltage of the bosonic field, on

the other hand, can be represented by the conjugate phase operator defined in Eq. (7)

$$V_{\text{qubit}} = i(\hat{a}_j - \hat{a}_j^\dagger) \tag{14}$$

up to a real-valued constant factor. Putting together the voltage dependence of the control field and the quantized bosonic modes, we obtain the microwave pulse controlled capacitive energy:

$$\hat{H}_3 = \sum_{j=1,2} i\left(\hat{a}_j - \hat{a}_j^\dagger\right) f_j(t)\cos(\omega_j t + \varphi_j)\ , \tag{15}$$

where ω_j, φ_j, and $f_j(t)$ represent, respectively, the frequency, phase, and amplitude of each microwave pulse. The overall system Hamiltonian thus consists of four parts:

$$\hat{H} = \hat{H}_0 + \hat{H}_1 + \hat{H}_2 + \hat{H}_3. \tag{16}$$

We further simplify this Hamiltonian by switching to the interaction picture

$$\hat{H}_I = U\left(\hat{H} - \hat{\tilde{H}}_0\right)U^\dagger \tag{17}$$

with $U = e^{-i\hat{\tilde{H}}_0 t}$ and

$$\hat{\tilde{H}}_0 = (\bar{\omega}_0 + \eta)\hat{n} = \tilde{\omega}_0\hat{n}\ . \tag{18}$$

The Hamiltonian for two bosonic modes hosted by two coupled gmon circuits in the interaction picture thus takes the form

$$\begin{aligned}\hat{H}_I = \sum_{j=1}^{2}\delta_j\hat{n}_j + \frac{\eta}{8}\sum_{j=1}^{2}[&\left(\hat{a}_j^2 e^{-2i\tilde{\omega}_0 t} + (\hat{a}_j^\dagger)^2 e^{2i\tilde{\omega}_0 t}\right)^2 + \left(\hat{a}_j^2 e^{-2i\tilde{\omega}_0 t} + (\hat{a}_j^\dagger)^2 e^{-2i\tilde{\omega}_0 t}\right)(2\hat{n}_j + 1)\\ &+ (2\hat{n}_j + 1)\left(\hat{a}_j^2 e^{-2i\tilde{\omega}_0 t} + (\hat{a}_j^\dagger)^2 e^{2i\tilde{\omega}_0 t}\right) + 4\hat{n}_j^2 - 4\hat{n}_j + 1]\\ &- \frac{\alpha'(t)}{2\kappa_1\kappa_2}(\hat{a}_1^\dagger\hat{a}_2^\dagger e^{2i\tilde{\omega}_0 t} + \hat{a}_1\hat{a}_2 e^{-2i\tilde{\omega}_0 t} - \hat{a}_2^\dagger\hat{a}_1 - \hat{a}_1^\dagger\hat{a}_2)\\ + \sum_{j=1,2} i\Big(\hat{a}_j(e^{-i(\tilde{\omega}_0+\omega_j)t-i\varphi} &+ e^{-i(\tilde{\omega}_0-\omega_j)t+i\varphi}) - \hat{a}_j^\dagger(e^{i(\tilde{\omega}_0+\omega_j)t+i\varphi} + e^{i(\tilde{\omega}_0-\omega_j)t-i\varphi})\Big) f_j(t)\ ,\end{aligned} \tag{19}$$

where $\delta_j = \bar{\omega}_0 - \omega_0$ is the detuning of each qubit from the initial harmonic frequency. The gate time of the gmon system t is usually set to be much longer than $1/\eta$ satisfying $t\tilde{\omega}_0 \gg 1$. We can therefore apply the rotating-wave-approximatoin (RWA) to omit the highly-oscillating components whose phase varies at a rate equal or or faster than $\tilde{\omega}_0$. As a result, we obtain the following Hamiltonian for the gmon circuit bosonic modes in RWA basis:

$$\hat{H}_{\text{RWA}} = \frac{\eta}{2}\sum_{j=1}^{2}\hat{n}_j(\hat{n}_j - 1) + g(t)(\hat{a}_2^\dagger \hat{a}_1 + \hat{a}_1^\dagger \hat{a}_2) + \sum_{j=1}^{2}\delta_j(t)\hat{n}_j + \sum_{j=1}^{2} i f_j(t)\left(\hat{a}_j e^{i\varphi_j(t)} - \hat{a}_j^\dagger e^{-i\varphi_j(t)}\right) , \quad (20)$$

where we assumed that the microwave frequencies ω_j are not too far from the qubit frequency $\tilde{\omega}_0$ such that $\Delta\omega = \tilde{\omega}_0 - \omega_j \ll \tilde{\omega}_0$, and define $g(t) = \alpha'(t)\frac{\kappa_1\kappa_2}{2}$.

	η	$g(t)$	$\delta_j(t)$	$f_j(t)$	$\varphi_j(t)$
amplitude	200 MHz	[-20, 20] MHz	[-20, 20] MHz	[-20, 20] MHz	[0, 2π]
error amplitude	$\pm$ 1 MHz	$\pm$ 1 MHz	$\pm$1 MHz	$\pm$ 1 MHz	

Table 1: Hamiltonian control parameter range.

The above Hamiltonian describes the two coupled Josephson LC circuits in the perturbative regime of its anharmonicity. The parameter range of its coupling terms are listed in Table. 1. We require that the magnitudes of our two-qubit coupling g-pulse $g(t)$, detuning $\delta_j(t)$ and microwave pulse amplitude $f_j(t)$ are at least one magnitude lower than η to meet our leakage minimization condition. Moreover, we make the modulation rate of all parameters $g(t), \delta_j(t), f_j(t)$ to be within the frequency range $[-50, 50]$ MHz due to experimental limitations. Projected onto the qubit basis, this Hamiltonian takes the form:

$$\hat{H}_{\text{RWA}} = \frac{g(t)}{2}(\sigma_1^x\sigma_2^x + \sigma_1^y\sigma_2^y) + \sum_{j=1}^{2}\left[\frac{\delta_j(t)}{2}\sigma_j^z - f_j(t)\left(\sin\varphi_j(t)\sigma_j^x + \cos\phi_j(t)\sigma_j^y\right)\right]. \quad (21)$$

The stochastic training environment is implemented by adding amplitude fluctuation sampled from a zero mean Gaussian distribution of 1 MHz standard deviation to the six control

amplitudes $\eta \to \eta + \delta\eta$, $g(t_k) \to g(t_k) + \delta g(t_k)$, $\delta_j(t_k) \to \delta_j(t_k) + \delta\delta_j(t_k)$ and $f_j(t_k) \to \delta f_j(t_k)$ in every discretized time step $t_k \in [0, \Delta t, 2\Delta t, \ldots, N\Delta t]$.

B Complete Leakage Bound

In the following sections, we derive the leakage bound L_{tot}, that accounts for different sources of leakage errors during a time-dependent Hamiltonian evolution. We start with the formulation of TSWT for defining a rotated computational basis, in which direct coupling induced leakage errors are suppressed to higher order. Next, we prove that the non-adiabatic leakage from our generalization of the adiabatic theorem under TSWT is sub-dominant to obtain the final leakage bound L_{tot} defined in (4) of the main text.

B.1 Time-dependent Schrieffer-Wolff transformation

The success of the Schrieffer-Wolff transformation [1] rests upon the difference in energy scales between the energy gap separating different subspaces and the coupling terms within and between different subspaces. To generalize the previous formulation, we decompose the system Hamiltonian into three parts

$$\hat{H}(t) = \hat{H}_0 + \hat{H}_1(t) + \hat{H}_2(t), \tag{22}$$

$$\hat{H}_0 = \sum_{\alpha} \sum_{m \in \Omega_\alpha} E_1 |m\rangle\langle m|, \tag{23}$$

$$\hat{H}_1(t) = \sum_{\alpha} \sum_{m \in \Omega_\alpha} \langle m|\hat{H}_1^{\alpha}(t)|m'\rangle |m\rangle\langle m'|$$

$$\hat{H}_2(t) = \sum_{\alpha \neq \alpha'} \sum_{m \in \Omega_\alpha, , m' \in \Omega_{\alpha'}} \langle m|\hat{H}_2^{\alpha,\alpha'}(t)|m'\rangle |\alpha, m\rangle\langle \alpha', m'| \ ,$$

where Ω_α with $\alpha \in \{0, 1, 2, \ldots\}$ represents different subspaces and $\{|m\rangle\}$ denotes the orthogonal basis states spanning each subspace. The first term, $\hat{H}_0$, accounts for the time-independent part of the system Hamiltonian. The second term, $\hat{H}_1(t)$, accounts for the time-dependent

coupling within each subspace Ω_α, which we call the block-diagonal term. The third part, $\hat{H}_2(t)$, accounts for the coupling between different subspaces which we call the block-off-diagonal term. In order to apply TSWT's perturbative expansion, we assume the magnitudes of three parts of the system Hamiltonian obey: $\Delta = \min_{\alpha\neq 0}|E_\alpha - E_0| \gg |\langle m|\hat{H}_1^\alpha(t)|m'\rangle| \sim |\langle m|\hat{H}_2^{\alpha,\alpha'}(t)|l\rangle|\ \epsilon$, for $\vee\, m, m' \in \Omega_\alpha, l \in \Omega_{\alpha'}, \alpha \neq \alpha'$ throughout the time-dependent Hamiltonian evolution, where we use Δ and ϵ to represent the different energy scales satisfying $\epsilon/\Delta << 1$. Under this assumption, each subspace Ω_α is separated from the other by an energy gap much larger than the coupling within Ω_α or between Ω_α and Ω_β for $\beta \neq \alpha$. Now we perform the Schrieffer-Wollf transformation represented by the anti-Hermitian operator $\hat{S}(t)$ which rotates the original basis state $|\psi\rangle$ to $|\tilde{\psi}\rangle = e^{-\hat{S}(t)}|\psi\rangle$. In this rotated basis, the effective Hamiltonian $\hat{\mathbb{H}}$ can be found by

$$i\frac{d}{dt}|\tilde{\psi}\rangle = \hat{\mathbb{H}}|\tilde{\psi}\rangle = i\frac{de^{-\hat{S}}}{dt}e^{\hat{S}}|\tilde{\psi}\rangle + ie^{-\hat{S}}\frac{d}{dt}|\psi\rangle, \tag{24}$$

$$\hat{\mathbb{H}} = -i\sum_{j=0}^{\infty}\frac{1}{(j+1)!}[\dot{\hat{S}},\hat{S}]_j + e^{-\hat{S}}\hat{H}e^{\hat{S}}\ , \tag{25}$$

where $\hat{S}(t)$ contains non-zero terms only between different subspaces, and is thus block-off-diagonal. The goal of TSWT is to find a perturbative solution of the rotation $\hat{S}(t) = \epsilon\hat{S}_1(t) + \epsilon^2\hat{S}_2(t) + \ldots + \epsilon^n\hat{S}_n(t)$ that block-diagonalizes the system Hamiltonian [1] such that the effective Hamiltonian $\hat{\mathbb{H}}$'s block-off-diagonal terms are suppressed to an order of $O\left(\frac{\epsilon^{n+1}}{\Delta^n}\right)$ for the nth-order perturbative solution of $\hat{S}(t)$. Below, we provide the derivation of $\hat{S}(t)$ up to the second order as a function of $\hat{H}(t)$.

We start by expanding the effective Hamiltonian in Eq. (25) into

$$\hat{\mathbb{H}} = -i\sum_{j=0}^{\infty}\frac{1}{(j+1)!}[\dot{\hat{S}},\hat{S}]_j + \sum_{j=0}\frac{1}{j!}[\hat{H}(t),\hat{S}(t)]^j \tag{26}$$

using the algebraic relation

$$e^{-\hat{S}(t)}\hat{H}(t)e^{\hat{S}(t)} = \sum_{j=0} \frac{1}{j!}[\hat{H}(t), \hat{S}(t)]^j, \tag{27}$$

$$[\hat{H}(t), \hat{S}(t)]^j = [\ldots[[\hat{H}(t), \underbrace{\hat{S}(t)], \hat{S}(t)]\ldots, \hat{S}(t)}_{\text{j many}}]. \tag{28}$$

We can now separate Eq. (26) into its block-diagonal and block-off-diagonal components and choose the correct form of $\hat{S}(t)$ to cancel the block-off-diagonal component to a chosen order. For simplicity, henceforth we do not write down the operator's time-dependence.

As defined before, $\hat{H}_0$ and $\hat{H}_1$ are block-diagonal, while $\hat{S}$ and $\hat{H}_2$ are block-off-diagonal. Therefore, even-order commutators in Eq. (28) between $\hat{H}_0$ or $\hat{H}_1$ with $\hat{S}$ or $\dot{\hat{S}}$ are block-diagonal, while their odd-order counterparts are block-off-diagonal. Also, the odd-order commutators between $\hat{H}_2$ or $\dot{\hat{S}}$ and $\hat{S}$ are block-diagonal and their odd counterparts are block-off-diagonal. We thus obtain block-diagonal part of Eq. (27):

$$\hat{\mathbb{H}}_{\mathrm{d}} = -i\sum_{j=0}^{\infty} \frac{1}{(2j+2)!}[\dot{\hat{S}}, \hat{S}]^{2j+1} + \sum_{j=0}^{\infty} \frac{1}{2j!}[\hat{H}_0 + \hat{H}_1, \hat{S}]^{2j} + \sum_{j=0}^{\infty} \frac{1}{(2j+1)!}[\hat{H}_2, \hat{S}]^{2j+1}. \tag{29}$$

Similarly we obtain the block-off-diagonal part of the effective Hamiltonian:

$$\hat{\mathbb{H}}_{\mathrm{od}} = -i\sum_{j=0}^{\infty} \frac{1}{(2j+1)!}[\dot{\hat{S}}, \hat{S}]^{2j} + \sum_{j=0}^{\infty} \frac{1}{(2j+1)!}[\hat{H}_0 + \hat{H}_1, \hat{S}]^{2j+1} + \sum_{j=0}^{\infty} \frac{1}{2j!}[\hat{H}_2, \hat{S}]^{2j}. \tag{30}$$

Now we would like to set the block-off-diagonal component to zero up to the order $O(\epsilon^3/\Delta^2)$. With a slight abuse of notation, we use ϵ below as shorthand for the unitless value ϵ/Δ. We solve such diagonalization through perturbative expansion of the TSWT rotation, $\hat{S} = \epsilon\hat{S}_1 + \epsilon^2\hat{S}_2 + \ldots + \epsilon^n\hat{S}_n$ and $\dot{\hat{S}} = \epsilon^2\dot{\hat{S}}_1 + \epsilon^3\dot{\hat{S}}_2 + \ldots + \epsilon^{n+1}\dot{\hat{S}}_n$, such that we solve order by order for the rotation $\hat{S}_n(t)$ that cancels the block-off-diagonal parts of the Hamiltonian of order $O(\epsilon^n)$. Here, we assume that the TSWT rotation changes at a rate that is of the same order as ϵ: $|\dot{\hat{S}}_n| \sim \epsilon|\hat{S}|$.

Following the same convention, we rewrite the original Hamiltonian as $\hat{H} = \hat{H}_0 + \epsilon(\hat{H}_1 + \hat{H}_2)$ according to relative amplitudes of different components, and insert it together with expansion of $\hat{S}(t)$ to obtain the order by order perturbative expansion of the block-diagonal and block-off-diagonal parts of the effective Hamiltonian:

$$\hat{\mathbb{H}}_{\mathrm{d}} = \hat{H}_0 + \epsilon\hat{H}_1 + \frac{1}{2}\epsilon^2[\hat{H}_2, \hat{S}_1] + \frac{1}{2}\epsilon^3[\hat{H}_2, \hat{S}_2] - i\frac{1}{2}\epsilon^3[\dot{\hat{S}}_1, \hat{S}_1] + O\left(\epsilon^4\right), \tag{31}$$

$$\begin{aligned}\hat{\mathbb{H}}_{\mathrm{od}} =& \epsilon[\hat{H}_0, \hat{S}_1] + \epsilon\hat{H}_2 + \epsilon^2[\hat{H}_0, \hat{S}_2] + \epsilon^2[\hat{H}_1, \hat{S}_1] - i\epsilon^2\dot{\hat{S}}_1 \\ &+ \epsilon^3[\hat{H}_1, \hat{S}_2] + \frac{\epsilon^3}{3}[[\hat{H}_2, \hat{S}_1], \hat{S}_1] - i\epsilon^3\dot{\hat{S}}_2 + O\left(\epsilon^4\right).\end{aligned} \tag{32}$$

Perturbatively diagonalizing the Hamiltonian to the first order in ϵ, we have

$$[\hat{H}_0, \hat{S}_1] + \hat{H}_2 = 0\ , \tag{33}$$

which gives us the matrix expression for the first order SW rotation

$$\hat{S}_1^{\alpha,\alpha'} = \frac{\hat{H}_2^{\alpha,\alpha'}}{E_{\alpha'} - E_\alpha}, \tag{34}$$

where $\hat{H}_2^{\alpha,\alpha'}$ is the Hamiltonian between subspace Ω_α and $\Omega_{\alpha'}$ and is itself a matrix. To second order in diagonalization, we have

$$[\hat{H}_0, \hat{S}_2] + [\hat{H}_1, \hat{S}_1] - i\dot{\hat{S}}_1 = 0\ , \tag{35}$$

which immediately yields the matrix representation of the second order SW rotation between subspace Ω_α and Ω'_α:

$$\hat{S}_2^{\alpha,\alpha'} = \frac{\hat{H}_1^\alpha \hat{H}_2^{\alpha,\alpha'} - \hat{H}_2^{\alpha,\alpha'}\hat{H}_1^{\alpha'}}{(E_{\alpha'} - E_\alpha)^2} - \frac{i\dot{\hat{H}}_2^{\alpha,\alpha'}}{(E_{\alpha'} - E_\alpha)^2} \tag{36}$$

where we use $\hat{H}_1^\alpha$ as the sub-dominant Hamiltonian terms within the subspace Ω_α. Inserting Eq. (34) and (36) into Eq. (31) and (32) gives us the block-diagonal and block-off-diagonal

parts of the effective Hamiltonian after second order TSWT:

$$
\begin{aligned}
(\hat{\mathbb{H}}_{\rm d})^{\alpha} =& \hat{H}_0^{\alpha} + \hat{H}_1^{\alpha} - \sum_{\alpha' \neq \alpha} \frac{\hat{H}_2^{\alpha,\alpha'} \hat{H}_2^{\alpha',\alpha}}{(E_{\alpha'} - E_{\alpha})} \\
&+ \frac{1}{2} \sum_{\alpha' \neq \alpha} \frac{\hat{H}_2^{\alpha,\alpha'} \hat{H}_2^{\alpha',\alpha} \hat{H}_1^{\alpha} - \hat{H}_1^{\alpha} \hat{H}_2^{\alpha,\alpha'} \hat{H}_2^{\alpha',\alpha}}{(E_{\alpha'} - E_{\alpha})^2} \\
&+ \sum_{\gamma \neq \alpha} \frac{i \left[\dot{\hat{H}}_2^{\alpha,\gamma} \hat{H}_2^{\gamma,\alpha} - \hat{H}_2^{\alpha,\gamma} \dot{\hat{H}}_2^{\gamma,\alpha} \right]}{(E_{\alpha} - E_{\gamma})^2} + O(\epsilon^4),
\end{aligned} \tag{37}
$$

$$
\begin{aligned}
(\hat{\mathbb{H}}_{\rm od})^{\alpha,\alpha'} &= \hat{H}_{\rm od}^{\alpha,\alpha'} + \Delta \hat{H}_{\rm od}^{\alpha,\alpha'}, \\
&= \frac{(\hat{H}_1^{\alpha})^2 \hat{H}_2^{\alpha,\alpha'} - 2 \hat{H}_1^{\alpha} \hat{H}_2^{\alpha,\alpha'} \hat{H}_1^{\alpha'} + \hat{H}_2^{\alpha,\alpha'} (\hat{H}_1^{\alpha'})^2}{(E_{\alpha'} - E_{\alpha})^2} \\
&\quad + \frac{2}{3(E_{\alpha'} - E_{\alpha})} \sum_{\gamma} \left[\frac{\hat{H}_2^{\alpha,\alpha'} \hat{H}_2^{\alpha',\gamma} \hat{H}_2^{\gamma,\alpha'}}{E_{\alpha'} - E_{\gamma}} - \frac{\hat{H}_2^{\alpha,\gamma} \hat{H}_2^{\gamma,\alpha} \hat{H}_2^{\alpha,\alpha'}}{E_{\alpha} - E_{\gamma}} \right] \\
&\quad - i \left[\frac{\dot{\hat{H}}_1^{\alpha} \hat{H}_2^{\alpha,\alpha'} + 2 \hat{H}_1^{\alpha} \dot{\hat{H}}_2^{\alpha,\alpha'} - 2 \dot{\hat{H}}_2^{\alpha,\alpha'} \hat{H}_1^{\alpha'} - \hat{H}_2^{\alpha,\alpha'} \dot{\hat{H}}_1^{\alpha'}}{(E_{\alpha'} - E_{\alpha})^2} \right] - \frac{\ddot{\hat{H}}_2^{\alpha,\alpha'}}{(E_{\alpha'} - E_{\alpha})^2} + O(\epsilon^4).
\end{aligned} \tag{38}
$$

With this block-off-diagonal Hamiltonian in hand, we continue to evaluate the population leakage out of the qubit subspace caused by this direct coupling in both off-resonant and resonant regimes in the following section.

B.2 Leakage Bound

Direct Coupling Leakage Bound:

The non-zero block-off-diagonal part of the Hamiltonian $\hat{\mathbb{H}}_{\rm od}$ directly couples the qubit subspace to the higher energy subspace. To evaluate the population that transition out of the qubit subspace due to this direct coupling, we adopt the interaction picture with state basis $|\psi(t)\rangle_I = U_{\rm d}^{-1}(t)|\psi(t)\rangle$ that relates to the Schrödinger picture basis' initial state $|\psi(0)\rangle$ by a block-diagonal Hamiltonian evolution $U_{\rm d}(t) = \mathbb{T}[e^{-i\int_0^t \hat{\mathbb{H}}_{\rm d}(\tau) d\tau}]$, where $\mathbb{T}$ represents the time-ordering. The Schrödinger equation in the interaction picture is $i\frac{d}{dt}|\psi(t)\rangle_I = U_{\rm d}^{-1}(t)\hat{\mathbb{H}}_{\rm od}(t)U_{\rm d}(t)|\psi(0)\rangle_I$

in which the lowest-order solution given $|\hat{\mathbb{H}}_{\rm od}(t)|$ is of order $O(\epsilon^3/\Delta^2)$:

$$|\psi(t)\rangle_I \approx \left[I - i\int_0^t U_{\rm d}^{-1}(\tau)\hat{\mathbb{H}}_{\rm od}(\tau)U_{\rm d}(\tau)d\tau\right]|\psi(0)\rangle \tag{39}$$

where we have inserted the initial condition $|\psi(0)\rangle_I = |\psi(0)\rangle$. Because $U_{\rm d}(t)$ preserves the computational subspace, the leakage is thus evaluated by the sum of the amplitudes of all excited states outside the qubit subspace arising from the non-zero block-off-diagonal Hamiltonian after TSWT:

$$\begin{aligned}
L_{\text{direct}}(t) &= \sum_{\alpha\neq 0, m\in\Omega_\alpha} |\langle m(t)|\left(U_{\rm d}(0)|\psi(0)\rangle - U_{\rm d}(t)|\psi(t)\rangle_I\right)| \\
&= \sum_{\alpha\neq 0, m\in\Omega_\alpha} \left|\langle m(t)|\int_0^t U_{\rm d}(t,\tau)\hat{\mathbb{H}}_{\rm od}(\tau)U_{\rm d}(\tau,t)d\tau|\psi(t)\rangle\right| \\
&\overset{1}{\approx} \sum_{\alpha\neq 0, m\in\Omega_\alpha} \left|\langle m(t)|\left(U_{\rm d}(t,\tau)\frac{1}{\Delta_\alpha}\hat{\mathbb{H}}_{\rm od}(\tau)U_{\rm d}(\tau,t)d\tau\right)\right|_{\tau=0}^{t} \\
&- \int_0^t U_{\rm d}(t,\tau)\frac{1}{\Delta_\alpha}\frac{d}{d\tau}\left(\hat{\mathbb{H}}_{\rm od}(\tau)\hat{\mathbb{H}}_{\rm d}^{-1}\frac{dU_{\rm d}(\tau,t)}{d\tau}\right)d\tau|\psi(t)\rangle\Big| \\
&\overset{2}{=} \sum_{\alpha\neq 0, m\in\Omega_\alpha} \left|\langle m(t)|\left(U_{\rm d}(t,\tau)\frac{1}{\Delta_\alpha}\hat{\mathbb{H}}_{\rm od}(\tau)U_{\rm d}(\tau,t)d\tau\right)\right|_{\tau=0}^{t} \\
&- \left(2U_{\rm d}(t,\tau)\frac{1}{\Delta_\alpha^2}\frac{d\hat{\mathbb{H}}_{\rm od}(\tau)}{d\tau}\frac{dU_{\rm d}(\tau,t)}{d\tau}\right)\Bigg|_{\tau=0}^{t} + \int_0^t U_{\rm d}(t,\tau)\frac{1}{\Delta_\alpha^2}\frac{d^2\hat{\mathbb{H}}_{\rm od}(\tau)}{d\tau^2}U_{\rm d}(\tau,t)|\psi(t)\rangle\Bigg| \\
&\overset{3}{\leq} \frac{\|\hat{\mathbb{H}}_{\rm od}(0)\|}{\Delta(0)} + \frac{\|\hat{\mathbb{H}}_{\rm od}(t)\|}{\Delta(t)} + 2\frac{\|\hat{\mathbb{H}}_{\rm od}(0)\|}{\Delta^2(0)} + 2\frac{\|\dot{\hat{\mathbb{H}}}_{\rm od}(t)\|}{\Delta^2(t)} + \int_0^t \frac{1}{\Delta^2(t)}\left\|\frac{d^2\hat{\mathbb{H}}_{\rm od}(t)}{dt^2}\right\|dt
\end{aligned} \tag{40}$$

where: approximation 1 is obtained from integration by parts , replacing $U_{\rm d}(\tau,t)$ with $\frac{dU_{\rm d}(\tau,t)}{d\tau}\hat{\mathbb{H}}_{\rm d}^{-1}$ and using $\langle m_\alpha|\hat{\mathbb{H}}_{\rm d}^{-1}|\psi_0\rangle \sim 1/\Delta_\alpha$; equality 2 is obtained by two more integrations by parts, and inequality 3 follows from the triangle inequality. The middle two terms of (40) are at least one magnitude smaller than the remaining terms in either the resonant or off-resonant regimes. Hence they can be omited in our final leakage bound:

$$L_{tot} = \frac{\|\hat{\mathbb{H}}_{\rm od}(0)\|}{\Delta(0)} + \frac{\|\hat{\mathbb{H}}_{\rm od}(t)\|}{\Delta(t)} + \int_0^t \frac{1}{\Delta^2(t)}\left\|\frac{d^2\hat{\mathbb{H}}_{\rm od}(t)}{dt^2}\right\|dt \tag{41}$$

We now compare different terms of the leakage bound (4) of the main text in the resonant and off-resonant frequency regimes separately. First, consider the off-resonant frequency regime such that the Fourier components of the Hamiltonian's modulation have a frequency range bounded above by a value much smaller than Δ. This condition implies that for any nth order time-derivative we have $\frac{d^n||\hat{\mathbb{H}}_{\rm od}||/dt^n}{||\hat{\mathbb{H}}_{\rm od}||} \sim \epsilon^n \ll \Delta^n$, where the modulation rate of the system Hamiltonian is much smaller than the minimum energy gap separating different energy subspaces. The TSWT assumption in the off-resonant regimes implies: $\|\hat{\mathbb{H}}_{\rm d}\| \sim \epsilon$ and $\|\hat{\mathbb{H}}_{\rm od}\| \sim \epsilon^3/\Delta^2$. Moreover, in the off-resonant regime, the total runtime T obeys $1/T \ll \Delta$ to guarantee that the modulation frequency is smaller than the energy gap. The first two terms in Eq. (40), $\frac{\|\hat{\mathbb{H}}_{\rm od}(0)\|}{\Delta(0)} + \frac{\|\hat{\mathbb{H}}_{\rm od}(t)\|}{\Delta(t)} \sim O(\frac{\epsilon^3}{\Delta^3})$, dominate in this regime, because they are separated from the remaining terms by a factor of $1/\Delta$ with $\Delta = \min_t \Delta(t)$.

Now consider the resonant regime in which the frequency components of the Hamiltonian modulation have magnitudes that are similar to the energy gap so that $\frac{d^n||\hat{\mathbb{H}}_{\rm od}||/dt^n}{||\hat{\mathbb{H}}_{\rm od}||} \sim \Delta^n$, where unwanted couplings between the qubit subspace and a higher energy subspace is of approximately the same frequency as the energy gap. The TSWT assumption still holds with the amplitudes of the time-dependent Hamiltonians obeying $\|\hat{H}_1\| \sim \|\hat{H}_2\| \sim \epsilon$. The last term in Eq. (40), $\int_0^t \frac{1}{\Delta^2(t)} \left|\left|\frac{d^2\hat{\mathbb{H}}_{\rm od}(t)}{dt^2}\right|\right| dt \sim O(\frac{\epsilon^3}{\Delta^3})$, is dominant in this regime because it is separated from the remaining terms by a factor of $1/\Delta$.

Non-adiabatic Leakage Bound:

A generalized adiabatic theorem given below bounds the leakage amplitude into higher-energy subspaces.

Theorem C.1. Let $\hat{H}(s)$ be a twice differentiable Hamiltonian parametrized by a unitless re-scaled time $s \in [0,1]$ comprising three parts: $\hat{H}(s) = H_0 + \hat{\mathbb{H}}_{\rm d}(s) + \hat{\mathbb{H}}_{\rm od}(s)$. The time-invariant term $\hat{H}_0 = \sum_{\alpha=0}^{\infty} \sum_{m\in\Omega_\alpha} E_\alpha |m\rangle\langle m|$ ensures there is a large constant energy gap between the lowest energy subspace Ω_0 and other higher energy subspaces. The time-varying

term $\hat{\mathbb{H}}_{\rm d}(s)$ accounts for couplings within each non-degenerate subspace Ω_α, and $\hat{\mathbb{H}}_{\rm od}(s)$ accounts for the coupling between different subspaces. There is a separation between the energy gap and inter/intra-subspace coupling: $\Delta = \min_\alpha |E_\alpha - E_0| \gg |\hat{\mathbb{H}}_{\rm d}(s)| \sim |\hat{\mathbb{H}}_{\rm od}(s)| \sim \delta$. Let $|\phi_0(s)\rangle = \sum_{m\in\Omega_0} a_m(s)|m\rangle$ be an instantaneous eigenstate in the lowest energy subspace Ω_0 at physical time sT. Let $|\psi(s)\rangle$ be the state evolved from the same initial state $|\phi_0(0)\rangle$ at time $s = 0$ under the total Hamiltonian $\hat{H}(s)$ to time s. We have the following inequality that bounds the difference between these two states at the final time T by:

$$
\begin{aligned}
L_{\text{non-adiabatic}} \leq \frac{1}{T}\Bigg[& \frac{1}{\Delta^2(s)}(||\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}|| + T||[\hat{\mathbb{H}}_{\rm d}(s), \hat{\mathbb{H}}_{\rm od}(s)]||)_{s=1} + \frac{1}{\Delta^2(s)}(||\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}|| \\
& + T||[\hat{\mathbb{H}}_{\rm d}(s), \hat{\mathbb{H}}_{\rm od}(s)]||)_{s=0}\Bigg] + \int_0^1 \frac{5}{\Delta^3(s)}\left(||\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}|| + T||[\hat{\mathbb{H}}_{\rm d}(s), \hat{\mathbb{H}}_{\rm od}(s)]||\right)^2 ds \\
& + \int_0^1 \frac{1}{\Delta^2(s)}\Bigg(T||[\hat{\mathbb{H}}_{\rm d}(s), [\hat{\mathbb{H}}_{\rm d}(s), \hat{\mathbb{H}}_{\rm od}(s)]]|| + 2||[\hat{\mathbb{H}}_{\rm d}(s), \frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}]|| + 2||[\frac{d\hat{\mathbb{H}}_{\rm d}(s)}{ds}, \hat{\mathbb{H}}_{\rm od}(s)]|| \\
& + \frac{1}{T}||\frac{d^2\hat{\mathbb{H}}_{\rm od}(s)}{ds^2}||\Bigg) ds \ , \qquad (42)
\end{aligned}
$$

where we choose an appropriate global phase for the initial state following the convention in [2].

Proof: The proof is a generalization of Goldstone's original proof in Ref. [2] that accounts for the inter-subspace dynamics induced by $\hat{\mathbb{H}}_{\rm d}$ and intra-subspace coupling induced by $\hat{\mathbb{H}}_{\rm od}$. To begin, we change from the Schrödinger picture $|\psi(s)\rangle$ to the interaction picture $|\tilde{\psi}(s)\rangle$. We define the unitary evolution under the time-dependent diagonal Hamiltonian $\hat{\mathbb{H}}_{\rm d}(s)$ as

$$
U_{\rm d}(s,0) = \mathbb{T}[e^{-iT\int_0^1 \hat{\mathbb{H}}_{\rm d}(s)ds}] \qquad (43)
$$

which includes quantum dynamics within each subspace Ω_α and satisfies the Schrödinger's equation:

$$
\frac{d}{ds}U_{\rm d}(s,0) = -iT(\hat{\mathbb{H}}_{\rm d}(s))U_{\rm d}(s,0). \qquad (44)
$$

In the rotated basis $|\tilde{\psi}(s)\rangle = U_{\rm d}(s,0)|\psi(s)\rangle$, the quantum dynamics induced by $\hat{\mathbb{H}}_{\rm d}(s)$ within each subspace is absent because

$$\frac{i}{T}\frac{d}{ds}|\psi(s)\rangle = (\hat{H}_0 + \hat{\mathbb{H}}_{\rm d}(s) + \hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0)|\tilde{\psi}(s)\rangle, \tag{45}$$

$$\frac{i}{T}\frac{d}{ds}\left(U_{\rm d}(s,0)\right)|\tilde{\psi}(s)\rangle + U_{\rm d}(s,0)\frac{d}{ds}|\tilde{\psi}(s)\rangle = (\hat{H}_0 + \hat{\mathbb{H}}_{\rm d}(s) + \hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0)|\tilde{\psi}(s)\rangle, \tag{46}$$

$$\rightarrow \frac{d}{ds}|\tilde{\psi}(s)\rangle = U_{\rm d}(0,s)(\hat{H}_0 + \hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0)|\tilde{\psi}(s)\rangle = \hat{\tilde{H}}(s)|\tilde{\psi}(s)\rangle, \tag{47}$$

$$\hat{\tilde{H}}(s) = U_{\rm d}(0,s)(\hat{H}_0 + \hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0) = \hat{H}_0 + U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0) \tag{48}$$

where we use $\hat{\tilde{H}}(s)$ to represent the effective Hamiltonian in the rotated basis.

Now we continue to establish the adiabatic theorem in the rotated basis. If the system Hamiltonian changes slowly compared to the energy gap Δ, then the transition from the ground state living in the lowest-energy subspace Ω_0 to an excited state in a higher-energy subspace is negligible. Let $|\tilde{\phi}_n(s)\rangle$ be the instantaneous energy eigenstate of the effective Hamiltonian in the interaction picture :

$$\hat{\tilde{H}}(s)|\tilde{\phi}_n(s)\rangle = \tilde{E}_n(s)|\tilde{\phi}_n\rangle, \tag{49}$$

where the eigenvalues $\{\tilde{E}_n\}$ are that of the effective Hamiltonian $\hat{H}_0 + U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0)$, which differs from the original energies $\{E_n\}$ of the static Hamiltonian $\hat{H}_0 = \sum_{n=0}^{\infty} E_n|\phi_n\rangle\langle\phi_n|$ by a factor of $O(\frac{\epsilon^3}{\Delta^3})$ under our assumptions.

By appropriately choosing the initial states global phase $e^{i\int_0^T E_0(t)dt}$, we can ensure that the lowest energy state in the subspace Ω_0 has eigenvalue $E_0 = 0$ and will thus remain zero throughout the evolution:

$$\hat{\tilde{H}}(s)|\tilde{\phi}_0(s)\rangle = 0. \tag{50}$$

The deviation from the adiabatic evolution between the actual state and the instantaneous eigenstate is measured by

$$\delta(T) = |\tilde{\psi}(T)\rangle - |\tilde{\phi}_0(T)\rangle = U_T(1,0)|\tilde{\phi}_0(0)\rangle - |\tilde{\phi}_0(T)\rangle\ , \tag{51}$$

where we use $U_T(1,0)$ as the unitary transformation induced by the effective Hamiltonian in the rotated basis $\hat{\tilde{H}}(s)$ for time T:

$$U_T(1,s) = \mathbb{T}[e^{-iT\int_s^1 \hat{\tilde{H}}(s)ds}] \,, \tag{52}$$

which obeys the Schrödinger equation:

$$\frac{d}{ds}U_T(1,s) = iTU_T(1,s)\hat{\tilde{H}}(s) = iTU_T(1,s)(\hat{H}_0 + U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0)). \tag{53}$$

Similar to Goldstone's proof [2], we rewrite the adiabatic deviation in Eq. (51) as

$$\delta(T) = \int_0^1 \frac{d}{ds}\left[U_T(1,s)|\tilde{\phi}_0(s)\rangle\right] ds \tag{54}$$

After differentiating by parts, we have

$$\delta(T) = \int_0^1 ds\frac{d}{ds}\left[U_T(1,s)\right]|\tilde{\phi}_0(s)\rangle + \int_0^1 dsU_T(1,s)\,\frac{d}{ds}|\tilde{\phi}_0(s)\rangle \tag{55}$$

Inserting Eq. (53) into Eq. (55) we get

$$\begin{aligned}\delta(T) =& iT\int_0^1 dsU_T(1,s)\hat{\tilde{H}}(s)|\tilde{\phi}_0(s)\rangle + T\int_0^1 dsU_T(1,s)\,\frac{d}{ds}|\tilde{\phi}_0(s)\rangle \\ =& T\int_0^1 dsU_T(1,s)\,\frac{d}{ds}|\tilde{\phi}_0(s)\rangle \,,\end{aligned} \tag{56}$$

where we simplified the expression using the fact that $|\tilde{\phi}_0(s)\rangle$ is the zero eigenvalue instantaneous eigenstate of the rotated Hamiltonian. We then differentiate Eq. (50) with respect to s on both sides to obtain:

$$\frac{d\hat{\tilde{H}}(s)}{ds}|\tilde{\phi}_0(s)\rangle + \hat{\tilde{H}}(s)\frac{d}{ds}|\tilde{\phi}_0(s)\rangle = 0, \tag{57}$$

$$\rightarrow \hat{\tilde{H}}(s)\frac{d}{ds}|\tilde{\phi}_0(s)\rangle = -\frac{d}{ds}\left[\hat{H}_0 + U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0)\right]|\tilde{\phi}_0(s)\rangle = -\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}|\tilde{\phi}_0(s)\rangle, \tag{58}$$

$$\rightarrow \frac{d}{ds}|\tilde{\phi}_0(s)\rangle = -\hat{G}'\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}|\tilde{\phi}_0(s)\rangle, \tag{59}$$

where we used the fact that $\hat{H}_0$ is a constant, employed the shorthand notation $\hat{\tilde{H}}_{\rm od}(s) = U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s))U_{\rm d}(s,0)$, and defined $\hat{G}' = \hat{\tilde{H}}^{-1}$. This inverse of effective Hamiltonian satisfies

$$\hat{G}' = \sum_{n\neq 0}\frac{1}{\tilde{E}_n}|\phi_n(s)\rangle\langle\phi_n(s)| \approx \sum_{n\neq 0}\frac{1}{E_n}|\phi_n(s)\rangle\langle\phi_n(s)| + O(\frac{\epsilon^4}{\Delta^3}) \approx \hat{G}\,, \tag{60}$$

where we used the same order-of-magnitude analysis as was employed Eq. (49): the rotated basis energy gap between different energy subspaces is the same as the original energy gap up to the order $O(\frac{\epsilon^4}{\Delta^3})$ with $\epsilon \ll \Delta$. Inserting Eqs. (53) and (59) into Eq. (56) and integrating by parts, we obtain:

$$\delta(T) = \int_0^1 \frac{1}{iT}\frac{dU_T(1,s)}{ds}\hat{\tilde{H}}^{-1}(-\hat{G}'\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds})|\tilde{\phi}_0(s)\rangle, \tag{61}$$

$$= \frac{i}{T}U_T(1,s)(\hat{G}')^2\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}|_{s=0}^{s=1} - \frac{i}{T}\int_0^1 dsU_T(1,s)\frac{d}{ds}\left[(\hat{G}')^2\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}|\tilde{\phi}_0(s)\rangle\right]. \tag{62}$$

Now, using the results from Ref. [2], it is straightforward to bound the magnitude of the adiabatic deviation given in Eq. (62) by

$$\begin{aligned}||\delta(T)|| \leq & \frac{1}{T}\left[||(\hat{G}')^2\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}||_{s=0} + ||(\hat{G}')^2\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}||_{s=1} + \int_0^1 ||\frac{d}{ds}(\hat{G}')^2\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}|\tilde{\phi}_0(s)\rangle||ds\right] \\ & \leq \frac{1}{T}\left[||(\hat{G}')^2\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}||_{s=0} + ||(\hat{G}')^2\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}||_{s=1}\right] \\ & + \frac{1}{T}\left[\int_0^1 ds\left(5||(\hat{G}')^3||\,||\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}||^2 + ||(\hat{G}')^2||\,||\frac{d^2\hat{\tilde{H}}_{\rm od}(s)}{ds^2}||\right)\right].\end{aligned} \tag{63}$$

Expanding the derivative of the off-diagonal Hamiltonian in the rotated basis we have

$$\begin{aligned}\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds} &= \frac{d}{ds}\left(U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s)U_{\rm d}(s,0)\right), \\ &= iTU_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm d}(s)\hat{\mathbb{H}}_{\rm od}(s)U_{\rm d}(s,0) + U_{\rm d}(0,s)\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}U_{\rm d}(s,0) \\ &\quad - iTU_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s)\hat{\mathbb{H}}_{\rm d}(s)U_{\rm d}(s,0), \\ &= iTU_{\rm d}(0,s)[\hat{\mathbb{H}}_{\rm d}(s),\hat{\mathbb{H}}_{\rm od}(s)]U_{\rm d}(s,0) + U_{\rm d}(0,s)\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}U_{\rm d}(s,0).\end{aligned} \tag{64}$$

This result and the triangle inequality give us the upper bound:

$$||\frac{d\hat{\tilde{H}}_{\rm od}(s)}{ds}|| \leq ||\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}|| + T||[\hat{\mathbb{H}}_{\rm d}(s), \hat{\mathbb{H}}_{\rm od}(s)]||. \tag{65}$$

Differentiating Eq. (64) with respect to s we obtain:

$$\begin{aligned}
&\frac{d^2\hat{\tilde{H}}_{\rm od}(s)}{ds^2} = -T^2U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm d}^2(s)\hat{\mathbb{H}}_{\rm od}(s)U_{\rm d}(s,0) + iTU_{\rm d}(0,s)\frac{d\hat{\mathbb{H}}_{\rm d}(s)}{ds}\hat{\mathbb{H}}_{\rm od}(s)U_{\rm d}(s,0) \\
&+ iTU_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm d}(s)\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}U_{\rm d}(s,0) + T^2U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm d}(s)\hat{\mathbb{H}}_{\rm od}(s)\hat{\mathbb{H}}_{\rm d}(s)U_{\rm d}(s,0) \\
&+ iTU_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm d}(s)\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}U_{\rm d}(s,0) + U_{\rm d}(0,s)\frac{d^2\hat{\mathbb{H}}_{\rm od}(s)}{ds^2}U_{\rm d}(s,0) - iTU_{\rm d}(0,s)\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}\hat{\mathbb{H}}_{\rm d}(s)U_{\rm d}(s,0) \\
&+ T^2U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm d}(s)\hat{\mathbb{H}}_{\rm od}(s)\hat{\mathbb{H}}_{\rm d}(s)U_{\rm d}(s,0) - iTU_{\rm d}(0,s)\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}\hat{\mathbb{H}}_{\rm d}(s)U_{\rm d}(s,0) \\
&- iTU_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s)\frac{d\hat{\mathbb{H}}_{\rm d}(s)}{ds}U_{\rm d}(s,0) - T^2U_{\rm d}(0,s)\hat{\mathbb{H}}_{\rm od}(s)\hat{H}_{\rm d}^2(s)U_{\rm d}(s,0) \\
&= T^2U_{\rm d}(0,s)[\hat{\mathbb{H}}_{\rm d}(s),[\hat{\mathbb{H}}_{\rm d}(s),\hat{\mathbb{H}}_{\rm od}(s)]]U_{\rm d}(s,0) + i2TU_{\rm d}(0,s)[\hat{\mathbb{H}}_{\rm d}(s),\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}]U_{\rm d}(s,0) \\
&+ U_{\rm d}(0,s)\frac{d^2\hat{\mathbb{H}}_{\rm od}(s)}{ds^2}U_{\rm d}(s,0) + i2TU_{\rm d}(0,s)[\frac{d\hat{\mathbb{H}}_{\rm d}(s)}{ds},\hat{\mathbb{H}}_{\rm od}(s)]U_{\rm d}(s,0)
\end{aligned} \tag{66}$$

This result gives us the following upper bound on the second derivative in the non-adiabatic deviation:

$$\begin{aligned}
||\frac{d^2\hat{\tilde{H}}_{\rm od}(s)}{ds^2}|| \leq& T^2||[\hat{\mathbb{H}}_{\rm d}(s),[\hat{\mathbb{H}}_{\rm d}(s),\hat{\mathbb{H}}_{\rm od}(s)]]|| + 2T||[\hat{\mathbb{H}}_{\rm d}(s),\frac{d\hat{\mathbb{H}}_{\rm od}(s)}{ds}]|| \\
&+ 2T||[\frac{d\hat{\mathbb{H}}_{\rm d}(s)}{ds},\hat{\mathbb{H}}_{\rm od}(s)]|| + ||\frac{d^2\hat{\mathbb{H}}_{\rm od}(s)}{ds^2}||.
\end{aligned} \tag{67}$$

Inserting (65) and (67) into (63) completes the proof.

We now evaluate the magnitudes of each term in our non-adiabatic leakage bound in the off-resonant regime. Under the same assumption as our analysis for coherent leakage errors, we can bound the order of magnitudes of non-adiabatic leakage errors from (42) by $\|\hat{\mathbb{H}}_{\rm d}\hat{\mathbb{H}}_{\rm od}\|/\Delta^2 \sim O(\epsilon^4/\Delta^4)$. This is one-order-of-magnitude lower than that from the direct coupling leakage

bound, which is of order $\|\hat{\mathbb{H}}_{\rm od}\|/\Delta \sim O(\epsilon^3/\Delta^3)$. As the result, off-resonant leakage errors are dominated by the direct coupling leakage errors.

C Control Filter Design

In this section, we describe in greater detail the design of the control filters used to suppress the first term of the leakage error that was bounded by (4) in the main text. Each time-dependent control trajectory is discretized into N steps in a piecewise constant representation, with neighboring steps separated by a short time Δt. The modulation rate of the control signal is determined by this smallest time step according to $F_{\text{sample}} = 1/\Delta t$. We take $\hbar = 1$ throughout this appendix. Because the first term of the leakage bound from (4) in the main text

$$\int_0^1 \frac{1}{\Delta^2(s)} \frac{1}{T} \left\| \frac{d^2 \hat{\mathbb{H}}_{\rm od}(s)}{ds^2} \right\| ds, \tag{68}$$

is proportional to the second time-derivative of the block-off-diagonal Hamiltonian after TSWT, to suppress it to a small value we must ensure that the frequency component of any control action proposed by the RL agent is sufficiently small compared to the energy gap Δ.

Without affecting the Markov nature of the control problem, while limiting the frequency components of the control, we chose to apply a two-pole normalized double-exponential smoothing filter [3] to the proposed control action. At each time step $1 \leq n \leq N$, the new control Hamiltonian depends both on the proposed control action $\hat{H}_n^{RL}$ set by the RL agent, and on the control taken in the previous two steps, $\hat{H}_{n-1}$ and $\hat{H}_{n-2}$, according to:

$$\hat{H}_n = a_1 \hat{H}_n^{RL} - b_1 \hat{H}_{n-1} - b_2 \hat{H}_{n-2}. \tag{69}$$

Here the filter coefficients are chosen according to the bandwidth B_w of the actuated control

and the modulation rate of the control signals as follows [3]:

$$\alpha = \exp[-\frac{\pi B_w}{F_{\text{sample}}}], \tag{70}$$

$$a_1 = (1-\alpha)^2,\ b_1 = -2\alpha,\ b_2 = \alpha^2. \tag{71}$$

D Evaluation of the Average Fidelity

In this section, we present our method for evaluating average fidelity, which is based on previous results of Nielsen [4]. The average fidelity,

$$\bar{F}(\mathcal{E}, U) = \int d\psi \langle\psi|U^\dagger \mathcal{E}(|\psi\rangle\langle\psi|)U|\psi\rangle, \tag{72}$$

measures the average performance of a quantum gate over a uniform distribution of the input quantum state under the noisy quantum channel $\mathcal{E}$ that describes the actual control realizations. It can be measured in experiments by a randomized benchmarking process. Theoretically, we evaluate the average fidelity by summing over all possible overlaps with the Pauli operators, i.e.,

$$\bar{F}(\mathcal{E}, U) = \frac{\sum_j \text{Tr}[U_{\text{target}} U_j^\dagger U_{\text{target}}^\dagger \mathcal{E}(U_j)] + d^2}{d^2(d+1)}, \tag{73}$$

where: the two-qubit Pauli operator $U_j = (\sigma_1^x)^{h_1}(\sigma_1^z)^{q_1}(\sigma_2^x)^{h_2}(\sigma_2^z)^{h_q}$ with $h_j, q_j \in \{0, 1\}$ satisfies $\text{Tr}[U_j U_k^\dagger] = \delta_{j,k} d$; $d = 4$ is the dimension of the two-qubit subspace dimension; $\mathcal{E}$ is the trace-preserving quantum operation that represents noisy realization of an ideal quantum gate control; U_{target} represents the target quantum gate. It is not hard to see that the necessary and sufficient condition for average fidelity to be one is to have $\mathcal{E}(\rho) = U_{\text{target}} \rho U_{\text{target}}^\dagger$. We evaluate $\mathcal{E}(\rho)$ of a zero mean Gaussian control-noise model with standard deviation σ_{noise}, by sampling the full control trajectory $\hat{H} + \delta\hat{H}$ under this noise model and averaging the evolved quantum state over different instances:

$$\mathcal{E}(\rho) = \mathbb{E}_{\delta\vec{c} \sim N(0, \sigma_{\text{noise}})} U(\vec{c} + \delta\vec{c})\rho U^\dagger(\vec{c} + \delta\vec{c}). \tag{74}$$

Such averaging can change a pure state into a mixed state and thus accounts for the magnitude of decoherence induced by stochastic control errors. Figure 4 of the main text is evaluated under 60 samples per noise model parameter.

References and Notes

[1] Y. Goldin, Y. Avishai, *Phys. Rev. B* **61**, 16750 (2000).

[2] S. P. Jordan, *arXiv:0809.2307* (2008).

[3] J. O. Smith, *Introduction to digital filters: with audio applications*, vol. 2 (Julius Smith, 2007).

[4] M. A. Nielsen, *Physics Letters A* **303**, 249 (2002).