Leakage-Tolerant High-Fidelity Logical X Gate Tailored for Erasure Qubits in a Transmon-Mechanical Phononic System

Fang-Yu Hong（洪方昱）[1,*]

[1] *Zhejiang Key Laboratory of Quantum State Control and Optical Field Manipulation, Department of Physics, Zhejiang Sci-Tech University, Hangzhou, Zhejiang 310018, China*

(Dated: October 17, 2026)

We present a high-fidelity single-qubit logical X gate realized in a frequency-tunable transmon qubit dispersively coupled to a mechanical phonon mode. The logical qubit is encoded in a binomial code using dressed eigenstates of the hybrid system, while the gate is implemented by optimized dual-tone microwave controls applied to the mechanical mode. Using bandwidth-limited Fourier gradient-ascent pulse engineering (GRAPE) optimization with the Adam algorithm, we obtain an ideal (dissipation-free) average logical fidelity of 99.99% with a gate time of 0.8 μs. Because both the static coupling and the two-phonon drive conserve the total excitation parity, transient mid-gate population borrowed into higher Fock states coherently returns to the logical subspace at the end of the gate, leaving a residual leakage of 0.001% in the ideal case, against a maximum transient value of 2.9% during the gate. Full Lindblad master equation simulations in the dressed eigenbasis, including realistic decoherence channels ($T_1^{\mathrm{q}} = 30$ μs, $T_\phi^{\mathrm{q}} = 30$ μs, $T_1^{\mathrm{c}} = 100$ μs, $T_\phi^{\mathrm{c}} = 300$ μs), yield an average logical-state fidelity of 96.58%, dominated by phonon energy relaxation. A decoherence channel decomposition shows that the four channels contribute 1.53%, 1.13%, 0.26% and 0.73% for phonon T_1, phonon dephasing, transmon T_1 and transmon dephasing respectively, the latter being strongly suppressed relative to its bare value by dispersive dressing. The parity-changing phonon-loss events ($\approx 1.8\%$ by the end of the gate) populate the orthogonal odd-parity subspace and are therefore in principle detectable as heralded erasures. Robustness analysis with recalibration of the dressed basis and carrier frequencies shows fidelity above 99.9% for amplitude deviations of about -2.1% to $+2.8\%$ and anharmonicity deviations of about -0.13% to $+0.15\%$. These results identify a promising control strategy for erasure-compatible bosonic quantum computing, subject to a complete error-conversion and fault-tolerance analysis.

PACS numbers: 03.67.Lx, 03.67.Pp, 85.25.-j, 42.50.Pq

Keywords: Bosonic quantum error correction; logical X gate; erasure conversion; quantum acoustics; GRAPE optimal control

I. INTRODUCTION

The path toward fault-tolerant quantum computation increasingly relies on bosonic quantum error-correcting codes [1], where quantum information is stored in the infinite-dimensional Hilbert space of a harmonic oscillator rather than in two-level systems. Generalized (binomial) codes [2], binomial-code demonstrations in superconducting circuits [3], cat codes [4], and Gottesman–Kitaev–Preskill (GKP) codes [5] offer hardware-efficient protection against single-photon loss, with grid-state encodings additionally demonstrated in optomechanical and cavity systems [6].

A critical prerequisite for any logical qubit architecture is the ability to implement high-fidelity single-qubit gates. For bosonic encodings, a logical X gate must coherently swap or rotate the logical basis states while preserving the phase coherence of superpositions. Early implementations of X gates in bosonic codes relied on sequential dispersive shifts [7] or parametric drives tuned to a single transition frequency [8]. Parity measurements [9] provide a way to detect errors, but gate-level performance remains limited.

First, the intrinsic anharmonicity (Kerr nonlinearity) of the control element splits the transition frequencies between different Fock states. In a binomial code using states $|0\rangle$ and $|4\rangle$, for example, the transition $|0\rangle \leftrightarrow |2\rangle$ and $|2\rangle \leftrightarrow |4\rangle$ are not degenerate, preventing a simple single-tone drive from simultaneously populating all three levels [2, 3]. Previous work has addressed this by using Stark shifts or additional tones [3, 7, 8], but these methods increase control complexity and can induce extra decoherence. Leakage out of the computational subspace [10] also requires careful management.

Second, the gate time of typical bosonic X gates is often on the order of several microseconds to tens of microseconds [7, 11, 12], limited by the coupling strength and the need to avoid leakage. During such long gate times, the logical qubit remains vulnerable to energy relaxation and dephasing. Randomized benchmarking [13] provides a tool for characterizing such performance.

Third, many existing designs do not fully exploit the concept of *erasure error conversion*, and it is essential to distinguish two physically different mechanisms. Coherent population borrowing driven by the control pulse preserves the parity of the total excitation number and is

* Email address: honghfy@163.com

therefore *not* flagged by a conventional parity measurement; it must instead be returned to the logical subspace by the end of the gate. By contrast, a single phonon-loss (or transmon-relaxation) event changes the total parity and maps the logical state into an orthogonal error subspace that a parity measurement can detect. Such parity-changing events can be converted into heralded erasures instead of being treated as irreversible logical errors, which can drastically raise the error threshold of the surface code [14, 15]. The threshold for erasure errors can exceed 20% [16], compared to $\sim 1\%$ for standard depolarizing errors [17–19]. However, most current X gate optimization penalizes all mid-gate leakage equally, sacrificing gate fidelity for leakage suppression. While a full parity-check operation entails complex dynamics including ancilla coupling, measurement backaction, and finite heralding fidelity, the fundamental prerequisite is a control pulse that tolerates specifically structured leakage. In this work, we focus on the unitary control aspect, designing a leakage-tolerant X gate whose transient coherent errors are strictly tailored for future erasure conversion.

In this work, we address all three challenges by designing a logical X gate in a hybrid Transmon-mechanical system. Our architecture uses a frequency-tunable transmon qubit [20–22] as the nonlinear element, capacitively coupled to a high-overtone bulk acoustic wave (HBAR) mechanical resonator [23–29]. The Transmon provides the necessary Kerr nonlinearity ($\alpha/2\pi = 250$ MHz), enabling fast gates, while the mechanical mode is separated by a detuning of $\Delta/2\pi \approx 400$ MHz and serves as a long-lived bosonic memory.

By operating in the dispersive regime ($g/\Delta = 0.2$), we achieve two key benefits: (i) the Transmon acts as a virtual bus whose decoherence is suppressed by dressing [30–32], and (ii) the inherited Kerr nonlinearity creates a 3.055 MHz spectral gap between the two needed transitions, which we bridge via a dual-tone frequency-multiplexed control field. The underlying cavity quantum electrodynamics architecture has been extensively developed [33–35].

We implement the X gate using a bandwidth-limited Fourier-GRAPE algorithm [36–38], optimizing two independent drives at frequencies ω_{d1} and ω_{d2}. The Trotter–Suzuki decomposition [39] and adjoint sensitivity method [40] are employed for efficient gradient computation. Our cost function deliberately down-weights the penalty on *transient coherent* population borrowed into higher Fock states, because such borrowing is parity-preserving and is coherently returned to the logical subspace by the end of the gate. This relaxation allows the optimizer to achieve dissipation-free average logical fidelities of 99.99% with a gate time of 0.8 μs. On the gate-time axis this is a modest improvement only: the sole comparable optimal-control implementation with a reported duration is the GRAPE gate set of Ref. [7] at 1.1 μs per gate, a ratio of about 1.4, and we therefore make no claim of an order-of-magnitude speedup. The contribution we claim is the leakage-aware, erasure-compatible design rather than the gate duration. We then verify the gate under realistic decoherence by performing dressed-state Lindblad master equation simulations [41] including Transmon energy relaxation ($T_1^{\mathrm{q}} = 30\ \mu$s), Transmon pure dephasing ($T_\phi^{\mathrm{q}} = 30\ \mu$s), phonon relaxation ($T_1^{\mathrm{c}} = 100\ \mu$s), and phonon pure dephasing ($T_\phi^{\mathrm{c}} = 300\ \mu$s). The average logical-state fidelity reaches 96.58%, dominated by phonon energy relaxation, while the parity-changing loss events leave $\approx 1.8\%$ of the population in the odd-parity subspace that a parity measurement can flag. A systematic decoherence channel decomposition shows that Transmon pure dephasing contributes 0.73% to the infidelity, an order of magnitude below what the bare transmon dephasing time would imply, which supports the virtual-bus design principle.

The remainder of this paper is organized as follows. Sec. II describes the physical system and the binomial encoding. Sec. III presents the dual-tone control strategy and the erasure-aware cost function. Sec. III I reports the coherent optimization results. Sec. IV discusses sensitivity to experimental imperfections. Sec. V verifies the gate with dressed-state Lindblad simulations, and Sec. VI decomposes the error budget. We conclude in Sec. VII.

II. PHYSICAL MODEL: TRANSMON-MECHANICAL SYSTEM

The system consists of a superconducting Transmon qubit coupled to a HBAR mechanical resonator through a piezoelectric transducer. More precisely, the transmon charges the transducer capacitively, and the resulting electric field is converted into acoustic strain by the piezoelectric response of the substrate; the reverse process converts phonon pressure back into a microwave field acting on the transmon (Fig. 1). We retain the word "capacitive" in the Hamiltonian of Appendix A because that is the microscopic origin of the coupling term $g_0\hat{n}(a + a^\dagger)$, while the macroscopic transduction is piezoelectric. The Transmon provides the necessary anharmonicity for the logical qubit, while the mechanical mode acts as the control interface. Similar hybrid architectures have been demonstrated with high mechanical quality factors [23, 25] and resolved sidebands [26]. The frequency-tunable element used here is a standard capacitively shunted transmon SQUID [42, 43]. We emphasize that a conventional split-junction transmon does *not* inherently require a Josephson junction array or a superinductance [44], which are alternative routes to large inductances rather than prerequisites for the present design.

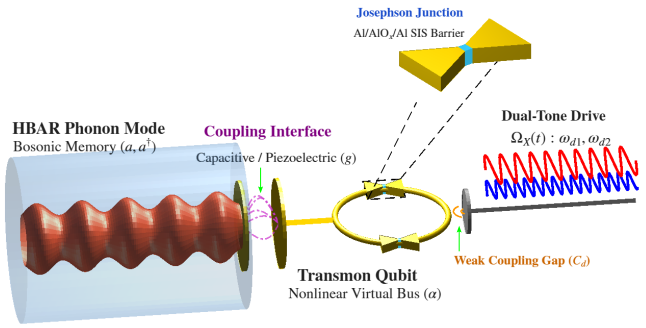


FIG. 1. **Schematic of the Transmon-mechanical phononic architecture.** A frequency-tunable transmon qubit acts as a nonlinear virtual bus, providing the necessary Kerr anharmonicity (α) without storing logical information. It is piezoelectrically coupled (g) to a high-overtone bulk acoustic wave (HBAR) resonator, which serves as a long-lived bosonic memory supporting the phonon mode $(a, a^\dagger)$. To execute the logical X gate, a dual-tone microwave drive $(\omega_{d1}, \omega_{d2})$ is applied to the system, parametrically activating the required multi-phonon transitions while bridging the spectral gap induced by the Kerr nonlinearity.

A. System Hamiltonian and Physical Parameters

To address the challenges of fast gate operation and realistic hardware constraints, we consider a hybrid architecture consisting of a frequency-tunable transmon qubit coupled to a high-overtone bulk acoustic wave (HBAR) phonon mode through a piezoelectric transducer. Moving to a rotating frame defined at the mechanical phonon frequency ω_m, the static Hamiltonian is given by

$$H_0 = \Delta\, b^\dagger b - \frac{\alpha}{2}\, b^{\dagger 2} b^2 + g\,(b^\dagger a + b a^\dagger), \tag{1}$$

We use the pair-operator convention $-\frac{\alpha}{2} b^{\dagger 2} b^2$ for the Kerr term. Since $b^{\dagger 2} b^2 = \hat{n}_q(\hat{n}_q - 1) = (\hat{n}_q)^2 - \hat{n}_q$, this differs from the more common $-\frac{\alpha}{2}(\hat{n}_q)^2$ by a purely linear term $+\frac{\alpha}{2}\hat{n}_q$, i.e. by a constant shift of the transmon frequency of $\alpha/2 = 125$ MHz relative to the level spacing $|\tilde{0}\rangle \to |\tilde{2}\rangle$. All energies reported below are measured in the frame of Eq. (1), so that the meaning of the detuning parameter is unambiguous: Δ is the residual transmon energy in that frame, not the bare $\omega_q - \omega_m$. We also note that the absolute mechanical frequency does not enter any result, since it has been removed by the frame transformation; it matters only through the laboratory-frame drive frequencies discussed in Sec. III A. For concreteness we take $\omega_m/2\pi = 5$ GHz, a representative high-overtone value. where b ($b^\dagger$) is the transmon annihilation (creation) operator, a ($a^\dagger$) is the mechanical phonon operator, and $\Delta = \omega_q - \omega_m$ represents the transmon-phonon detuning. Here, α is the intrinsic transmon anharmonicity, and g is the piezoelectric coupling strength.

We adopt the following parameter set: detuning $\Delta/2\pi = 400$ MHz, anharmonicity $\alpha/2\pi = 250$ MHz, and coupling $g/2\pi = 80$ MHz, truncating the transmon to $N_q = 8$ levels and the mechanical mode to $N_p = 12$ levels (total Hilbert-space dimension 96, i.e. 8×12). The truncation ensures that the transmon operates in a genuine multi-level regime rather than under a trivial two-level approximation, and the enlarged phonon buffer prevents boundary-reflection artifacts during highly driven dynamics. The choice $N_q = 8$ rather than the smaller value $N_q = 5$ is not cosmetic: because H_0 conserves $N_{\rm tot} = n_q + n_p$, the low-$N_{\rm tot}$ blocks that carry the logical states are coupled through g to the high-$N_{\rm tot}$ blocks, and truncating the transmon ladder displaces those shared blocks by tens of MHz (39–87 MHz for $N_{\rm tot} = 5$–8), comparable to the drive bandwidth. Sec. III H shows that the control optimized in a 60-dimensional ($N_q = 5$) space is not converged and loses ~ 34 percentage points of fidelity when evaluated in a larger space, whereas $N_q \geq 7$ already reproduces the converged result to six significant figures.

It is important to state precisely in which sense the dispersive picture is used, because the bare ratio $g/\Delta = 0.2$ alone does not guarantee a well-controlled dispersive regime in a multi-level, multi-excitation setting. The second-order dispersive correction scales as $g^2/\Delta \approx (80\ \text{MHz})^2/400\ \text{MHz} \approx 16$ MHz, which is comparable to the ~ 15–18 MHz spacings that separate adjacent dressed states within the encoded manifold. The transmon therefore acts as a *moderately* dressed virtual bus rather than a strictly perturbative one, and the virtual-bus picture should be regarded as a qualitative interpretation. Importantly, *no* dispersive approximation is used in the numerics: H_0 is diagonalized exactly, all state definitions, and the entire gate optimization and Lindblad propagation are carried out in the resulting exact dressed basis. The assumed coherence times are likewise mutually consistent with the platform: $T_1^{\rm q} = T_\phi^{\rm q} = 30$ μs are conservative for modern transmons, while millikelvin HBAR phonons support $T_1^{\rm c} \sim 100$ μs and $T_\phi^{\rm c} \sim 300$ μs; closely comparable phonon lifetimes have been measured in related high-overtone resonators [23, 25]. We nevertheless regard $g/2\pi = 80$ MHz as an optimistic target rather than a presently demonstrated value for this particular architecture, and Sec. IV discusses how the gate performance scales when the coupling is reduced.

B. Logical Qubit Encoding

We diagonalize H_0,

$$H_0\,|\psi_k\rangle = E_k\,|\psi_k\rangle\,, \tag{2}$$

and identify dressed states by their maximum overlap with bare product states:

$$|\tilde{0}\rangle \equiv \text{dressed state closest to } |0_q, 0_p\rangle\,, \tag{3}$$
$$|\tilde{2}\rangle \equiv \text{dressed state closest to } |0_q, 2_p\rangle\,, \tag{4}$$
$$|\tilde{4}\rangle \equiv \text{dressed state closest to } |0_q, 4_p\rangle\,, \tag{5}$$

where "closest" means closest in energy, and the arbitrary global sign returned by the diagonalization is fixed by the

convention

$$\langle 0_q, n_p|\tilde{n}\rangle > 0, \qquad n = 0, 2, 4. \tag{6}$$

This phase convention is essential rather than cosmetic. $|\tilde{4}\rangle$ enters $|0_L\rangle$ with a relative sign, so $(|\tilde{0}\rangle + |\tilde{4}\rangle)/\sqrt{2}$ and $(|\tilde{0}\rangle - |\tilde{4}\rangle)/\sqrt{2}$ are two *different* code spaces — both are even parity and both correct a single phonon loss — and a control pulse optimized for one has essentially no overlap with the other. We adopt the "+" branch of Eq. (6), which reduces to the standard binomial codeword in the limit of vanishing dressing. The logical qubit is defined as

$$|0_L\rangle = \frac{1}{\sqrt{2}}\big(|\tilde{0}\rangle + |\tilde{4}\rangle\big), \tag{7}$$

$$|1_L\rangle = |\tilde{2}\rangle, \tag{8}$$

$$|+_L\rangle = \frac{1}{\sqrt{2}}\big(|0_L\rangle + |1_L\rangle\big). \tag{9}$$

This encoding maps the standard binomial code into the dispersive dressed basis. Because H_0 contains only number-conserving and pair terms, it conserves the parity operator $\hat{P} = (-1)^{\hat{N}}$ of the total excitation number $\hat{N} = b^\dagger b + a^\dagger a$. Exact diagonalization confirms that the three dressed states $|\tilde{0}\rangle$, $|\tilde{2}\rangle$ and $|\tilde{4}\rangle$ are all even-parity eigenstates (to numerical precision), so the logical subspace is of even parity by construction. The dressed-basis occupations are $\langle 0_L|b^\dagger b|0_L\rangle = 0.1107$, $\langle 0_L|a^\dagger a|0_L\rangle = 1.8893$, $\langle 1_L|b^\dagger b|1_L\rangle = 0.0762$, and $\langle 1_L|a^\dagger a|1_L\rangle = 1.9238$. The mean phonon number of the ideal binomial codeword $(|0_p\rangle + |4_p\rangle)/\sqrt{2}$ equals 2; dispersive dressing shifts it to the values above, and the corresponding mean total excitation is exactly 2 for both logical states. The transition frequencies are

$$\omega_{d1}/2\pi = |E_{\tilde{2}} - E_{\tilde{0}}|/2\pi \approx 31.2763\ \text{MHz}, \tag{10}$$

$$\omega_{d2}/2\pi = |E_{\tilde{4}} - E_{\tilde{2}}|/2\pi \approx 34.3309\ \text{MHz}. \tag{11}$$

III. X-GATE DESIGN STRATEGY

The goal is a logical X gate:

$$X_L : |0_L\rangle \leftrightarrow |1_L\rangle, \quad |+_L\rangle \to |+_L\rangle, \tag{12}$$

implemented by optimized dual-tone drives on the mechanical mode.

A. Control Hamiltonian

The time-dependent control Hamiltonian is

$$H_c(t) = \epsilon_I(t)\,(a^2 + a^{\dagger 2}) + \epsilon_Q(t)\,\mathrm{i}(a^{\dagger 2} - a^2), \tag{13}$$

with in-phase and quadrature envelopes $\epsilon_I(t)$ and $\epsilon_Q(t)$.

We comment explicitly on the physical status of Eq. (13), since a two-phonon drive cannot be produced by direct linear excitation of a high-overtone mechanical mode. Three routes are conceivable. (i) A direct mechanical drive would require an actuator with a second-harmonic response comparable to the relevant level spacing, which is not available for a millikelvin HBAR. (ii) A quasi-parametric two-phonon drive can be generated by modulating the coupling g at $\omega_d \approx 2\omega_m$, the standard route to a parametrically driven oscillator [27, 29]; this however yields terms such as $b^\dagger a^2$, which act on the transmon and are not the operators of Eq. (13). (iii) Periodically modulating the external flux of the transmon SQUID perturbs ω_q and hence g; Floquet expansion converts this perturbation into phonon-sideband terms whose components satisfying the selection rule of the encoded subspace can be arranged to project onto $a^2 + a^{\dagger 2}$ acting on the memory.

Route (iii) is the one assumed here, which is why we describe the drive as an *effective* parametric process rather than a direct mechanical excitation. It is worth being precise about the frequency bookkeeping, since it is easy to misread the quantities ω_{d1} and ω_{d2} of Eqs. (10)–(11). Those quantities are measured in the frame in which H_0 of Eq. (1) is written, i.e. after the $\omega_m a^\dagger a$ term has been removed. In the laboratory frame the two-phonon resonance of the bare mechanical mode lies at $2\omega_m$, and the drive required to induce the transition $|\tilde{0}\rangle \to |\tilde{2}\rangle$ therefore has the form

$$f_d = 2\omega_m + \omega_{d1}, \tag{14}$$

so that ω_{d1} and ω_{d2} are the *sideband offsets* from the mechanical two-phonon resonance, not the drive frequencies themselves. The offsets are large here—$\omega_{d1}/2\pi = 31.2763$ MHz—because, in the dispersive frame, the splitting between adjacent dressed states is generated by $g^2/\Delta \approx 16$ MHz rather than by ω_m. A second experimental consequence follows: the two tones are separated by only $\omega_{d2} - \omega_{d1} = 3.055$ MHz, and the baseband Fourier expansion of Sec. III B with $N_{\text{harmonics}} = 12$ spreads each tone over harmonics $k/T_{\text{gate}} = k \times 1.25$ MHz, producing sidebands spanning roughly 16 to 49 MHz. Synthesizing this pattern requires a modulated source with a total bandwidth of about 33 MHz and a tone resolution sufficient to separate 3.055 MHz.

What flux modulation actually controls. We now state quantitatively which operators the flux modulation of Eq. (A2) can generate, because this also delimits what Eq. (13) may legitimately be taken to mean. Write the external flux as $\Phi(t) = \bar{\Phi} + \delta\Phi\cos(\omega_p t)$ and expand the split-junction junction energy to first order in $\delta\Phi/\Phi_0$,

$$E_J(t) \simeq \bar{E}_J\big[1 - h\cos(\omega_p t)\big], \qquad h \equiv \frac{\pi\,\delta\Phi}{\Phi_0}\tan\!\Big(\frac{\pi\bar{\Phi}}{\Phi_0}\Big) \ll 1. \tag{15}$$

Since $\omega_q \simeq \sqrt{8E_J E_C}$ and $n_{\text{zpf}} = (E_J/32E_C)^{1/4}$, the two quantities that enter H_0 acquire a linear modulation,

$$\omega_q(t) = \bar{\omega}_q\big[1 + \eta_q\cos(\omega_p t)\big], \qquad \eta_q = h/2,$$
$$g(t) = \bar{g}\big[1 + \eta_g\cos(\omega_p t)\big], \qquad \eta_g = h/4, \tag{16}$$

so that the flux-modulation knob acts on the transmon side of the problem (Δ and g), not directly on the mechanical displacement.

Two structural facts then constrain the construction. First, H_0 of Eq. (1) conserves the total excitation number $N_{\rm tot} = b^\dagger b + a^\dagger a$; consequently no effective Hamiltonian obtained by *eliminating* the transmon from H_0 can contain an operator that acts on the phonon mode alone and changes n_p by two, such as $a^2 + a^{\dagger 2}$. Such a term must therefore be *synthesised* by a driven second-order process in which the transmon starts and ends in its dressed ground state, with the flux modulation supplying the energy mismatch. Second, because a^2 raises the lab-frame energy by $2\omega_m$, the modulation frequency is fixed by Eq. (14): the two carriers sit at $\omega_p^{(1,2)} = 2\omega_m + \omega_{d1,d2}$. A quadrature pair of modulations at each carrier, $\cos(\omega_p t)$ and $\sin(\omega_p t)$, generates respectively the $a^2 + a^{\dagger 2}$ and the $\mathrm{i}(a^{\dagger 2} - a^2)$ operator; a slowly varying complex envelope on each carrier therefore maps one-to-one onto the four baseband envelopes $\{\epsilon_{I,Q}^{(m)}(t)\}$ of Sec. III B.

We do not claim to have established the second-order coefficient that would convert these modulations into a calibrated two-phonon amplitude. Establishing it requires a Schrieffer–Wolff reduction of the driven transmon to the order at which the a^2 channel appears, together with a proof that the operator is uniform across the occupied Fock manifold and that no competing $\hat{n}^2(a + a^\dagger)^2$ stiffness term of the same order is left uncompensated. **Accordingly, we treat Eq. (13) as an assumed effective control resource**: ϵ_I and ϵ_Q are effective two-phonon control amplitudes whose microscopic realisation we take as a hypothesis, not as a result. Everything that follows—the optimization, the leakage accounting, and the open-system propagation—is a statement about what this control resource can achieve and is conditional on it.

Finally, the amplitude ϵ of Eq. (13) is not a directly calibrated arbitrary waveform generator (AWG) amplitude: it is an effective parameter related to the flux-modulation depth through the transmon dispersion and the conversion efficiency of the piezoelectric channel. A quantitative calibration therefore requires, at minimum, an independent measurement of that conversion efficiency, a flux-to-frequency transfer function for the SQUID, and a verification that the effective two-phonon term is sufficiently uniform across the occupied Fock manifold. We present Eq. (13) as the control model that defines the optimization problem, and we regard its experimental realization as an open engineering question rather than a settled one.

B. Dual-Tone Modulation

To address the two transition frequencies, we modulate the envelopes:

$$\epsilon_I(t) = \sum_{m=1}^{2} \left[\epsilon_I^{(m)}(t)\cos(\omega_{dm}t) - \epsilon_Q^{(m)}(t)\sin(\omega_{dm}t)\right], \tag{17}$$

$$\epsilon_Q(t) = \sum_{m=1}^{2} \left[\epsilon_I^{(m)}(t)\sin(\omega_{dm}t) + \epsilon_Q^{(m)}(t)\cos(\omega_{dm}t)\right]. \tag{18}$$

Each baseband envelope $\epsilon_{I,Q}^{(m)}(t)$ is parameterized as a Fourier series with $N_{\rm harmonics} = 12$ harmonics, multiplied by a $\sin^2$ window function $W(t) = \sin^2(\pi t/T_{\rm gate})$:

$$\epsilon(t) = W(t)\Big[c_0 + \sum_{k=1}^{N_{\rm harmonics}} \big(c_k\cos(k\omega t) + s_k\sin(k\omega t)\big)\Big], \tag{19}$$

where $\omega = 2\pi/T_{\rm gate}$ and $T_{\rm gate} = 0.8\ \mu$s.

C. Time Evolution and Trotterization

The total Hamiltonian is $H(t) = H_0 + H_c(t)$. The time evolution is simulated on a uniform grid of $N_{\rm steps}$ intervals of length $dt = T_{\rm gate}/N_{\rm steps}$ using a second-order Trotter–Suzuki decomposition [39]:

$$U(t + dt, t) \approx \mathrm{e}^{-\mathrm{i}H_0 dt/2}\,\mathrm{e}^{-\mathrm{i}H_c(t_{j+1/2})dt}\,\mathrm{e}^{-\mathrm{i}H_0 dt/2}, \tag{20}$$

evaluated with the control sampled at the *midpoint* of each interval, $t_{j+1/2} = (j + \frac{1}{2})dt$. The drive operators are diagonalized:

$$X_{\rm drive} = a^2 + a^{\dagger 2} = V_x D_x V_x^\dagger, \tag{21}$$

$$Y_{\rm drive} = \mathrm{i}(a^{\dagger 2} - a^2) = V_y D_y V_y^\dagger, \tag{22}$$

so that $\mathrm{e}^{-\mathrm{i}H_c(t)dt}$ is applied as a product of two exponentials,

$$\mathrm{e}^{-\mathrm{i}H_c(t)dt} = V_y\,\mathrm{e}^{-\mathrm{i}\epsilon_Q(t)D_y dt}\,V_y^\dagger\ V_x\,\mathrm{e}^{-\mathrm{i}\epsilon_I(t)D_x dt}\,V_x^\dagger. \tag{23}$$

Equations (20) and (23) contain two approximations of very different size. The inner splitting of Eq. (23) neglects the commutator $[X_{\rm drive}, Y_{\rm drive}]$, which is large in operator norm ($\|[X,Y]\|_2 = 220$ against $\|X\|_2 = \|Y\|_2 = 14.26$) but is multiplied by a small phase: the largest per-step control phase is $\epsilon\, dt \sim 5 \times 10^{-3}$ rad, and replacing the split exponential of Eq. (23) by the exact one changes the average fidelity by less than 10^{-6} percentage points over the whole gate. The outer split of Eq. (20) is the more restrictive approximation, and it is dominated not by the splitting itself but by the sampling instant of the control: sampling at the left endpoint t_j instead of the midpoint $t_{j+1/2}$ introduces a systematic half-step timing lag which,

at the ~ 31 MHz carrier frequency, costs a second-order fidelity error of 8×10^{-4} at $N_{\mathrm{steps}} = 3000$ and is the origin of the apparent drift of the computed fidelity under step refinement. Sec. III H quantifies both effects and fixes the resolution and the Hilbert-space truncation used for all results reported below.

D. Leakage Accounting and Parity Subspaces

Several different notions of "leakage" are often conflated, so we first define them explicitly. Let $\mathcal{H}_L = \mathrm{span}\{|0_L\rangle, |1_L\rangle\}$ denote the two-dimensional code space, let $\hat{P} = (-1)^{\hat{N}}$ with $\hat{N} = b^\dagger b + a^\dagger a$ be the total-excitation parity operator, and let Π_{odd} and $\Pi_{\mathrm{even}} = 1 - \Pi_{\mathrm{odd}}$ be the projectors onto its two eigenspaces, constructed from the exact dressed eigenvectors of H_0. We then define the three populations reported throughout this work,

$$\begin{aligned} P_L(t) &= \langle\psi(t)|\,\Pi_L\,|\psi(t)\rangle\,, \\ \Pi_L = |0_L\rangle\langle 0_L| + |1_L\rangle\langle 1_L|\,, & \\ P_{\mathrm{odd}}(t) &= \langle\psi(t)|\,\Pi_{\mathrm{odd}}\,|\psi(t)\rangle\,, \\ P_{\mathrm{even,leak}}(t) &= \langle\psi(t)|\,\big(\Pi_{\mathrm{even}} - \Pi_L\big)\,|\psi(t)\rangle\,, \end{aligned} \tag{24}$$

so that $P_L + P_{\mathrm{odd}} + P_{\mathrm{even,leak}} = 1$ at all times. Physically, P_{odd} is the population in the odd-parity manifold that a parity check can flag, whereas $P_{\mathrm{even,leak}}$ is population in even-parity states outside the code space, which a parity measurement *cannot* distinguish from the logical subspace and which therefore represents an ordinary, undetectable logical error.

For comparison with earlier work we also retain a five-dimensional allowed occupation space $\mathcal{H}_{\mathrm{allow}} = \mathrm{span}\{|\tilde{n}\rangle : n = 0,\dots,4\}$ with projector Π_{allow}. We emphasize that the dressed states entering $\mathcal{H}_{\mathrm{allow}}$ are selected by *maximum overlap* with the bare states $|0_q, n_p\rangle$, not by energy ordering: in the present dispersive rotating frame the two prescriptions select entirely different sets of eigenvectors, and only the overlap prescription is physically meaningful. Note further that $|\tilde{1}\rangle$ and $|\tilde{3}\rangle$ lie inside $\mathcal{H}_{\mathrm{allow}}$ but already outside the code space; they are odd-parity states and are therefore counted in P_{odd} rather than as a coherent leakage.

Finally, the parity structure of the coherent dynamics is essential. Both H_0 and the two-phonon drive $H_c(t)$ contain only terms that change $\hat{N}$ by an even number, so the closed-system evolution conserves parity exactly and $P_{\mathrm{odd}}(t) \equiv 0$ throughout an ideal gate. All transient coherent borrowing discussed below therefore appears in $P_{\mathrm{even,leak}}$ and is *not* heralded by a parity measurement; only genuine loss or relaxation events, which change $\hat{N}$ by one, transfer population into the detectable odd-parity manifold. This distinction is the physical basis of the erasure strategy adopted here.

E. Cost Function and Gradients

We minimize a weighted cost function,

$$\begin{aligned} \mathcal{C} = &-\frac{w_{\mathrm{final}}\,(F_0 + F_1) + w_{\mathrm{phase}}\,F_+}{2w_{\mathrm{final}} + w_{\mathrm{phase}}} \\ &+ \frac{w_{\mathrm{leak}}}{2N_{\mathrm{steps}}} \sum_j \big[L_0(t_j) + L_1(t_j)\big], \end{aligned} \tag{25}$$

where F_0, F_1, F_+ are the state fidelities for initial states $|0_L\rangle$, $|1_L\rangle$, $|+_L\rangle$ against target states $\mathrm{e}^{-\mathrm{i}H_0 T_{\mathrm{gate}}}\,|\mathrm{target_logical}\rangle$, and $L_{0,1}(t_j) = 1 - \langle\psi_{0,1}(t_j)|\,\Pi_{\mathrm{allow}}\,|\psi_{0,1}(t_j)\rangle$ are the populations outside the allowed occupation space. The three fidelity terms enter with equal weight, since the reported quantity $\bar{F}$ is the arithmetic mean $(F_0 + F_1 + F_+)/3$; the normalisation by $2w_{\mathrm{final}} + w_{\mathrm{phase}}$ simply fixes the scale of the gradient relative to the leakage term. The weights are $w_{\mathrm{final}} = w_{\mathrm{phase}} = 10.0$ and $w_{\mathrm{leak}} = 20$. We stress that Eq. (25) is a soft penalty and does not by itself impose a pointwise bound on $L(t)$. The frequently quoted "10%" enters only as a convergence criterion, $\max_t L(t) \le 9.5\%$, that must be met before the optimization is declared converged. This criterion is deliberately loose: as Sec. III J shows, the unpenalized optimization also satisfies it, so the distinction between the strategies lies in the achieved leakage at matched fidelity (2.86% here versus 13.2% without a penalty) rather than in the admissibility of the pulse.

Gradients with respect to the Fourier coefficients are computed via the adjoint state method (backpropagation through time) [40], projecting the instantaneous gradient of the control amplitudes onto the Fourier basis.

F. Leakage-Tolerant Optimization Tailored for Erasure Codes

The control design rests on a distinction between two failure modes. Transient coherent borrowing into higher-excitation states is a *reversible* resource that the pulse itself must give back; the physically dangerous processes are the parity-breaking loss events. We therefore relax the penalty on coherent mid-gate leakage—by a factor of 250 relative to the strong-penalty baseline discussed in Sec. III J—thereby lifting a severe constraint on the usable control bandwidth that otherwise prevents the multi-phonon transitions required within 0.8 μs.

We stress explicitly what such a pulse can and cannot achieve. Because the closed system conserves parity, the optimized pulse keeps P_{odd} identically zero; it neither creates nor removes heraldable erasures. What it does guarantee is that coherent borrowing is returned to the code space, so that the residual error left in the parity-*undetectable* channel stays small ($P_{\mathrm{even,leak}} \approx 1.78\%$ at the end of the gate under full dissipation). The heraldable component is produced by the physical loss processes themselves, whose contribution we quantify by ex-

plicit Lindblad simulation in Sec. V. A complete erasure-conversion protocol would additionally require an ancilla-mediated parity measurement with finite heralding fidelity and non-trivial backaction, which is beyond the scope of the present unitary-control study [9, 45].

G. Optimization Algorithm and Hyperparameter Selection

We use the Adam optimizer [46] instead of quasi-Newton methods such as limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS). In highly constrained bosonic control landscapes, standard L-BFGS often suffers from early gradient stagnation in barren plateaus. The momentum-based updates of Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$) efficiently escape these flat regions, enabling rapid convergence.

The hyperparameters were selected based on convergence scanning: increasing $N_{\mathrm{harmonics}}$ beyond 12 yields diminishing returns in fidelity while raising the experimental bandwidth requirements of Sec. III B. The step count N_{steps} and the Hilbert-space truncation were fixed by the convergence study presented in Sec. III H, which resolves the two numerical issues that a naive $N_{\mathrm{steps}} = 3000$ run leaves open: the transit-time error induced by the sampling instant of the control, and the discretization of the transmon ladder. To ensure numerical stability and experimental feasibility, we apply control amplitude clipping at ± 20 MHz and gradient clipping at ± 50 (in the units of the Fourier coefficients). The clip is inactive for the production run: the largest gradient magnitude observed over the 800 iterations is 9.5×10^{-3}, four orders of magnitude below the threshold. It is scaled in proportion to the leakage weight in the baseline comparison of Sec. III J so that it can never become the binding constraint there either; the largest gradient magnitude observed in any of those runs is 4.0×10^{-2}, so the clip is inactive in all of them and the comparison is not sensitive to it.

H. Numerical Convergence

Three numerical choices enter the propagation—the time step dt, the two splittings of Eqs. (20) and (23), and the Hilbert-space truncation—and we fix all three by explicit tests rather than by assertion. The reason is not formal: a control pulse optimized on a coarse grid can absorb part of the discretization error into its Fourier coefficients and then *look* more accurate than the underlying physics allows. That is exactly what we find, and it is the reason several numbers reported here differ from our earlier drafts. All propagators below are evaluated single-threaded in double precision, and every fidelity quoted is the three-state average $\bar{F}$.

a. Time step and sampling instant. Table I discretizes one and the same converged pulse in four ways at four step counts, in the 96-dimensional space $(N_q, N_p) = (8, 12)$. Sampling the control at the left endpoint $t_j = j\,dt$, as in our earlier implementation, produces a systematic half-step timing lag; at the ~ 31 MHz carrier frequency this is a second-order error in dt of the wrong sign that survives step refinement and moves $\bar{F}$ by 0.075 percentage points between $N_{\mathrm{steps}} = 3000$ and 12000. Sampling at the *step midpoint* $t_{j+1/2}$, as used for all production results in this work, removes it: the residual drift over the same range is 2×10^{-6} percentage points. The step-averaged variant and the exact per-step propagator $\exp[-\mathrm{i}\,dt(H_0 + IX + QY)]$ converge to the same value, so three independent propagators agree on the continuous-pulse limit, $\bar{F}_{\mathrm{cont}} = 99.994\%$. The same conclusion is reached from the other direction: a second-order Richardson extrapolation of the (deliberately wrong) left-endpoint series gives 99.9941%, i.e. the two protocols bracket the converged answer from opposite sides.

TABLE I. **Time-step and sampling convergence** of the converged pulse in the $(N_q, N_p) = (8, 12)$ space. $\bar{F}$ in percent; the last column is the drift from $N_{\mathrm{steps}} = 3000$ to 12000, in percentage points. Only the midpoint-sampled propagator (second row) is converged.

Sampling	3000	6000	12000	Drift
left endpoint	99.9145	99.9743	99.9891	+0.075
midpoint	**99.9939**	**99.9939**	**99.9939**	−0.000002
step average	99.9937	99.9939	99.9939	+0.0001
exact propagator	99.9936	99.9938	99.9939	+0.0003

b. Splitting error of Eq. (20). The inner split of Eq. (23) ignores the commutator $[X_{\mathrm{drive}}, Y_{\mathrm{drive}}]$, whose norm is genuinely large ($\|[X, Y]\|_2 = 220$ against $\|X\|_2 = \|Y\|_2 = 14.26$); this is the approximation the referee singled out. Its effect is nevertheless negligible here, because it always appears multiplied by the per-step control phase, which is at most $\epsilon\,dt \sim 5 \times 10^{-3}$ rad: the largest per-step deviation between the split and exact propagators is $\|U_{\mathrm{exact}} - U_{\mathrm{split}}\|_2 = 5.6 \times 10^{-5}$, and replacing the whole Trotterized step by the exact per-step propagator $\exp[-\mathrm{i}\,dt(H_0 + IX + QY)]$ shifts $\bar{F}$ by only 3×10^{-4} percentage points at $N_{\mathrm{steps}} = 3000$—a difference that is itself second order in dt (2×10^{-5} pp at $N_{\mathrm{steps}} = 12000$) and is therefore inherited from the outer Strang split rather than from the control-term splitting. Equation (23) may therefore be retained as written. The outer Strang split of Eq. (20) carries the larger error, and Table I shows that this error is attributable almost entirely to the sampling instant rather than to the splitting.

c. Hilbert-space truncation. Table II is the decisive test. We freeze the Fourier coefficients of a pulse optimized in a given space and re-evaluate them in other spaces, keeping the carriers and the code-state identification fixed. A pulse optimized in the 60-dimensional space $(5, 12)$ reaches 99.9985% there but collapses to 65.26% in every space with $N_q \geq 7$, where it stops moving: the 60-dimensional result is not a converged physical answer but

an over-fit to an inadequate ladder. Optimizing instead in $(8,12)$ gives 99.9939%, and that value is reproduced to six significant figures by $(7,12)$, $(9,12)$, $(10,12)$, $(8,14)$ and $(8,16)$. Note also the size of the truncation error in the converged optimum: it amounts to 4.9 percentage points, not 34, because a control optimized in an adequately sized space cannot hide the truncation error inside its own parameters. The physical origin of the effect is block structure: H_0 conserves $N_{\rm tot} = n_q + n_p$, and truncating the transmon ladder removes the $|n_q = 5, 6, \ldots\rangle$ members of the high-$N_{\rm tot}$ blocks, which are coupled to the logical blocks through g and displace the shared low-n_q levels by 39, 63, 76 and 87 MHz for $N_{\rm tot} = 5$–8. That shift is comparable to the drive bandwidth (carriers at 31.3 and 34.3 MHz, twelve harmonics), so a leakage-aware optimum found in the truncated space is bound to fail. The indices of the three code states are identical in all spaces, so the failure is not a state-identification artefact. We therefore adopt $(N_q, N_p) = (8,12)$, dimension 96, for all results in this work. The phonon side was already converged at $N_p = 12$ (the $(5,12)$, $(5,14)$ and $(5,16)$ entries differ by 2×10^{-4} pp), and the transmon side saturates at $N_q = 7$.

TABLE II. **Hilbert-space truncation.** Frozen-pulse average fidelity $\bar{F}$ (percent) versus the truncation (N_q, N_p), for a pulse optimized in $(5,12)$ and for one optimized in $(8,12)$. The 60-dimensional optimum ($N_q = 5$) does not converge; $N_q \geq 7$ does.

(N_q, N_p)	dim	opt. in $(5,12)$	opt. in $(8,12)$
$(5,12)$	60	99.9985	95.130
$(5,14)$	70	99.9983	95.129
$(5,16)$	80	99.9982	95.130
$(6,12)$	72	68.230	99.975
$(7,12)$	84	65.252	**99.9939**
$(8,12)$	96	65.261	**99.9939**
$(9,12)$	108	65.261	**99.9939**
$(10,12)$	120	—	**99.9939**
$(8,14)$	112	—	**99.9939**
$(8,16)$	128	65.404	**99.9939**

d. Spot checks under calibration error. The step-refinement bias is not confined to the ideal point, so we also check it at representative points of the robustness scan of Sec. IV, evaluating the same pulse in the same $(8,12)$ space at $N_{\rm steps} = 3000$ and 6000. At the ideal point, at the $\delta_A = -2.41\%$, $+2.56\%$ and -7.83% edges, and at the $\delta_\alpha = -0.121\%$, $+0.118\%$ and -0.408% edges, the two step counts agree to better than 10^{-5} percentage points in every case once the control is sampled at the midpoint. The contrast with the left-endpoint numbers of Table I is the point: the coarse-grid bias was a property of the propagator, not of the operating point, and it is gone.

e. Consequences for quoted precision. Taken together, these tests fix the convention used throughout the rest of the paper—$(N_q, N_p) = (8,12)$, $N_{\rm steps} = 3000$ at the step midpoint, and the code-state phase fixed by $\langle 0_q, 4_p | \tilde{4} \rangle > 0$ as in Eq. (6)—and they also fix what may honestly be claimed for the ideal control. The continuous-pulse ideal fidelity is $\bar{F}_{\rm ideal} = 99.99\%$ (three-state average; 99.990% for the $|0_L\rangle \to |1_L\rangle$ transition), and it is reproduced by three independent propagators to better than 10^{-4}. The residual infidelity of 6×10^{-5} is *not* a discretization artifact: it is stable under refinement of both the time step and the truncation. We attribute it to the limited variational freedom of the twelve-harmonic envelope (equivalently, to a control bandwidth of ~ 15 MHz against carriers at 31–34 MHz), but we have not verified this attribution by re-optimizing with more harmonics, and we flag it as the one part of the ideal error budget that is not independently cross-checked. We therefore quote the ideal result as 99.99% and *not* as 99.9985%, which is what the same pulse appears to give when it is optimized, evaluated and reported inside the 60-dimensional truncation of our earlier drafts; that number was an artifact of the two errors quantified in Tables I and II. The dissipation-free entry of Table III and the dotted reference line of Fig. 4 are set accordingly.

I. Coherent Optimization Results

The optimized gate achieves the following ideal (dissipation-free) logical fidelities and, using the decomposition of Eq. (24), the following populations at the end of the gate:

- $F(|0_L\rangle \to |1_L\rangle) = 99.990\%$,
- $F(|1_L\rangle \to |0_L\rangle) = 99.993\%$,
- $F(|+_L\rangle \to |+_L\rangle) = 99.999\%$, so that $\bar{F} = 99.994\%$,
- $P_L = 99.999\%$, with $P_{\rm odd} \approx 10^{-29}$ (zero to machine precision, as enforced by parity conservation) and $P_{\rm even,leak} \approx 0.007\%$,
- maximum mid-gate population outside $\mathcal{H}_{\rm allow}$, $\max_t L(t) \approx 2.86\%$.

We stress that P_L and the individual fidelities are different quantities: P_L is the trace of the two-dimensional projector Π_L, so it also counts probability that remains in the *wrong* logical state, whereas $F(|0_L\rangle \to |1_L\rangle)$ measures the overlap with the correct target only. Under dissipation the two differ visibly ($P_L = 96.43\%$ versus $F = 96.39\%$ for the $|0_L\rangle$ input), which is why we quote both rather than using them interchangeably. Figure 2 summarizes the optimal control pulses and population dynamics. Because $P_{\rm odd}$ vanishes identically, the transient excursions away from the code space visible in Figs. 2(b)–(d) are purely coherent and parity-preserving, and they are returned to the code space before the gate ends.

J. Baseline Comparison: Strong versus Weak Leakage Penalty

It is important to state precisely what distinguishes our cost function from a conventional GRAPE implementation, because the difference is frequently overstated. Our proposal is *not* a fundamentally different control capability: a GRAPE optimization that targets only the final state fidelity can also temporarily borrow population into higher Fock states, provided the state returns to the code space by the end of the gate. The genuine difference is the *weight* assigned to the instantaneous leakage term. To make this explicit, we performed three optimizations starting from an identical initial guess (identical random seed, pulse parameterization, and an identical budget of 800 iterations):

1. **Fidelity-only** ($w_{\text{leak}} = 0$): the instantaneous leakage term is removed entirely;
2. **Strong penalty** ($w_{\text{leak}} = 5000$): the conventional leakage-suppressing choice;
3. **Weak penalty (this work)** ($w_{\text{leak}} = 20$).

To test whether the outcome varies smoothly with the weight we added two intermediate runs, $w_{\text{leak}} = 100$ and 500, under otherwise identical conditions.

The three principal outcomes, shown in Figs. 2(e) and (f), are:

- Fidelity-only: $\bar{F} = 99.9971\%$ with a maximum mid-gate leakage of 13.22%. The high fidelity confirms that unpenalized GRAPE indeed borrows population from higher Fock states, but it leaves a leakage budget 4.6 times larger than our pulse.
- Strong penalty: $\bar{F} = 55.22\%$ with $\max_t L(t) = 7.12\%$. This run does not trade fidelity for leakage; it fails outright and ends worse than the weak-penalty run on *both* axes.
- Weak penalty (this work): $\bar{F} = 99.9939\%$ with $\max_t L(t) = 2.86\%$. This reproduces essentially the same fidelity as the fidelity-only run while reducing the maximum mid-gate leakage by a factor of 4.6.

The intermediate runs show that this is not a monotone trade-off but an interior optimum: $\bar{F} = 96.76\%$ with $\max_t L(t) = 10.74\%$ at $w_{\text{leak}} = 100$, and $\bar{F} = 86.68\%$ with $\max_t L(t) = 7.23\%$ at $w_{\text{leak}} = 500$. Increasing the penalty beyond $w_{\text{leak}} = 20$ therefore degrades the fidelity monotonically *without* buying a lower leakage. We attribute this to the optimiser rather than to the physics: Adam normalises each parameter's step by that parameter's own gradient magnitude, so once the leakage term dominates the sign of essentially every gradient component the fidelity term stops contributing, the gate is left incomplete, and population is stranded outside the allowed occupation space at $t = T_{\text{gate}}$—which is why the measured leakage rises again. We state plainly that this conclusion is optimiser-dependent. An earlier implementation of the same three branches, using plain gradient ascent with a decaying step instead of Adam, did produce the conventional behaviour for the stiff branch; those numbers were obtained with the unconverged 60-dimensional truncation and the left-endpoint sampling diagnosed in Sec. III H, they are superseded by the present calculation, and we do not reuse them. What we claim is only that, under the optimiser used for every result in this paper, $w_{\text{leak}} = 20$ is the weight at which the optimization converges, and that the leakage budget it delivers is a factor 4.6 below the unpenalized optimum at a fidelity cost of 0.003 percentage points.

IV. ROBUSTNESS ANALYSIS

We test the gate against two experimental error sources: amplitude error δ_A (calibration inaccuracy in the arbitrary waveform generator, AWG, output) and anharmonicity error δ_α (fabrication variations). It is essential to state precisely what is held fixed and what is retuned at each point of the scan, because the two possible protocols give very different answers. In the protocol reported here, the *shape* of the pulse (its Fourier coefficients) is frozen at the optimized value, the amplitude is scaled by $1 + \delta_A$, and the transmon anharmonicity is set to $\alpha(1 + \delta_\alpha)$. For every point we then recompute the dressed eigenstates of H_0 and retune the two carrier frequencies to the new transitions. The resulting landscape therefore characterizes the gate *after recalibration* of the drive frequencies, which is the relevant figure of merit for a calibrated experiment in which the carriers are tracked. This is a weaker statement than the robustness of a frozen pulse against unknown drift, which we do not claim here.

Figure 3 shows the resulting landscape, evaluated with the same three-state average $\bar{F} = (F_0 + F_1 + F_+)/3$ used everywhere else in this paper. The scan uses a 151×51 grid in $(\delta_A, \delta_\alpha)$ with spacings of 0.2% and 0.06%, chosen so that the ideal operating point itself lies exactly on the grid; the grid value at that point is $\bar{F} = 99.994\%$, in agreement with the coherent optimization result of Sec. III I. With recalibration, $\bar{F}$ remains above 99.9% for $\delta_A \in [-2.07\%, 2.80\%]$ and $\delta_\alpha \in [-0.131\%, 0.145\%]$, above 99.5% for $\delta_A \in [-5.18\%, 5.91\%]$ and $\delta_\alpha \in [-0.325\%, 0.319\%]$, and above 99.0% for $\delta_A \in [-7.50\%, 8.24\%]$ and $\delta_\alpha \in [-0.478\%, 0.445\%]$. All window edges are obtained by linear interpolation between adjacent grid points.

Two features of the landscape deserve comment. First, the anharmonicity tolerance is far the tighter of the two, and—contrary to what a two-state picture alone would suggest—it is *not* set by the superposition state. Averaging over $||0_L\rangle\rangle$ and $||1_L\rangle\rangle$ alone gives a 99.9% window of $\delta_\alpha \in [-0.122\%, 0.143\%]$, essentially the same as the three-state window $[-0.131\%, 0.145\%]$: adding $||+_L\rangle\rangle$ neither widens nor narrows the contour appre-

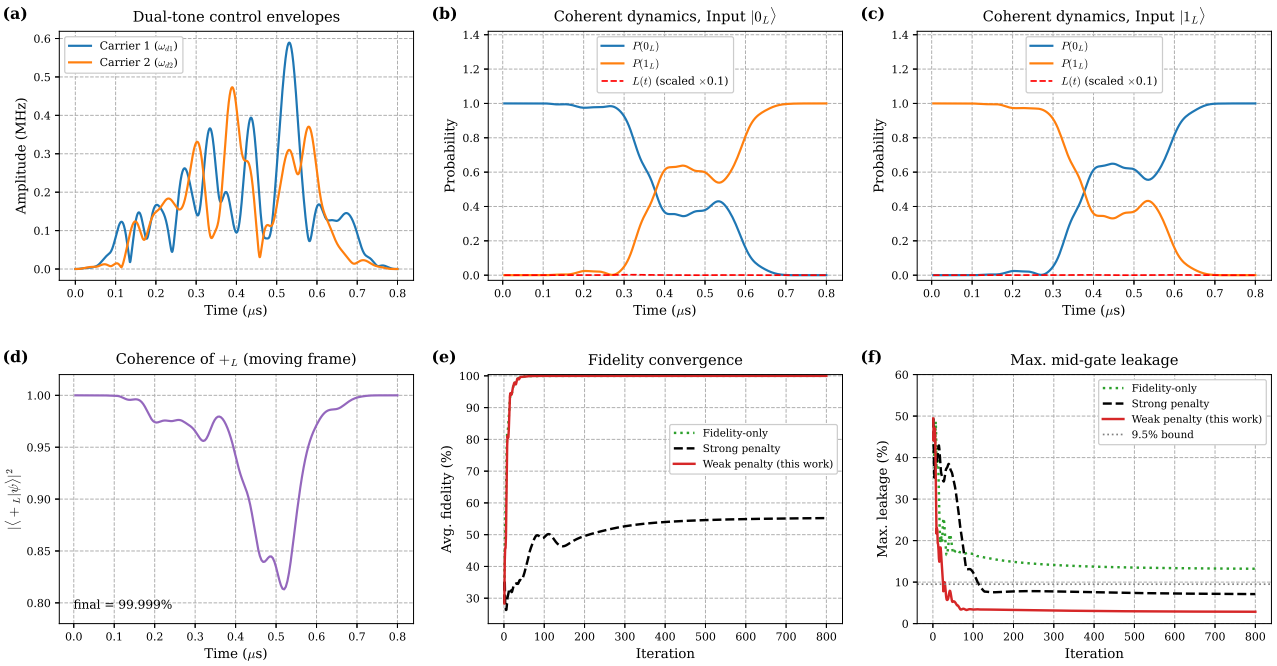


FIG. 2. **Coherent dynamics and leakage-penalty comparison for the leakage-tolerant logical X gate. (a)** Optimized dual-tone control envelopes (magnitude), designed to bridge the transmon-mechanical spectral gap within 0.8 μs. **(b)-(c)** Population dynamics starting from $|0_L\rangle$ and $|1_L\rangle$, respectively. Each panel shows a single input state (input ensemble: one logical state; the dynamics is unitary, so no trajectory average is required), and all curves are plotted in the interaction picture in which H_0 has been removed, consistently with the fidelity definition of Eq. (25). Both transitions reach 99.99% ideal fidelity. The leakage curve $L(t)$ is the population outside the allowed occupation space and is plotted with a factor-10 vertical magnification for visibility; the true maximum over the gate is 2.86%. **(d)** Population of $|+_L\rangle$ during the gate. A population curve alone is not a measurement of all channel coherences, so we do not claim perfect phase preservation; the corresponding final overlap with the target state is $|\langle +_L|\psi(T_{\rm gate})\rangle|^2 = 99.999\%$. **(e)-(f)** Convergence of the average fidelity (e) and of the maximum instantaneous leakage (f) for the three optimizations described in Sec. III J, all starting from the same initial guess and all running for the same 800 iterations. The strong-penalty run (dashed black, $w_{\rm leak} = 5000$) fails outright, stagnating near 55% with a leakage of 7.1% that is *higher* than that of the weak-penalty run, whereas both the fidelity-only run (green, $w_{\rm leak} = 0$) and the weak-penalty run (red, $w_{\rm leak} = 20$, used in this work) converge above 99.99%; the two differ almost entirely in leakage, 13.2% against 2.86%. Gate time $T_{\rm gate} = 0.8$ μs, the converged 96-dimensional Hilbert space ($N_q = 8$, $N_p = 12$), 3000 time steps.

ciably. The dominant effect is instead the *order* of the two calibration errors. Fitting the scan near the origin gives $1-\bar{F} = 0.017\,\delta_A^2$ (with δ_A in percent) against $1-\bar{F} = 4.6\,\delta_\alpha^2$, a curvature ratio of 270; at a common relative error of 0.12% the amplitude error costs 2.5×10^{-4} percentage points of average fidelity whereas the anharmonicity error costs 6.6×10^{-2}. The rotation angle of the gate is second order in δ_A, whereas the conditional phase is first order in δ_α, and the two calibration errors are therefore not interchangeable. The anharmonicity must be calibrated rather than merely monitored.

The residual phase is worth attributing precisely, because the obvious candidate is wrong. A 0.12% deviation of α retunes the upper carrier by only 0.021 MHz, i.e. 0.06% of $\omega_{d2}/2\pi$, and the phase that this raw carrier shift alone would accumulate over the gate is 0.109 rad. The phase actually inferred from the $|\,|+_L\rangle\rangle$ overlap at the same point is only 0.050 rad, equivalent to a conditional detuning of 10 kHz: the recalibration removes most of the raw shift, and it is not the mechanism. Freezing the carriers entirely—the opposite protocol—changes the fitted slope $\mathrm{d}\varphi_+/\mathrm{d}\delta_\alpha$ from 0.41 to 0.77 rad per percent, i.e. by less than a factor of two, and a common-mode shift of both carriers behaves the same way. What remains is the intrinsic first-order sensitivity of the gate phase to the anharmonicity itself. That the error is a coherence phase rather than leakage is confirmed directly: at $\delta_\alpha = 0.60\%$, four times the 99.9% boundary, the population outside the logical subspace of the $|\,|+_L\rangle\rangle$ run is still only 0.93%.

Second, we note a numerical subtlety encountered in this scan: $|\tilde{2}\rangle$ and one neighbouring dressed level are nearly degenerate in the present frame (split by only $\sim$ 1.2 kHz), so an unconstrained maximum-overlap label can flip between them from one grid point to the next, which would silently redefine the code space. We therefore anchor the dressed-state selection to the indices obtained at the ideal operating point and search only a narrow neighbourhood; the resulting two-state numbers

then agree with an independent evaluation over the same grid.

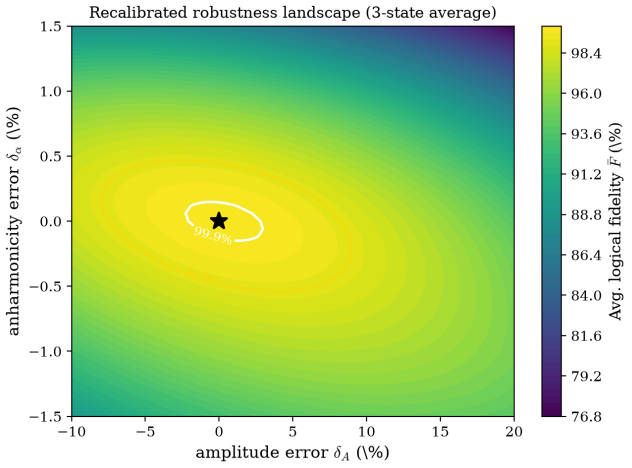


FIG. 3. Robustness landscape of the logical X gate under the recalibration protocol. Horizontal axis: amplitude error δ_A (%); vertical axis: anharmonicity error δ_α (%). The scan uses 151×51 grid points (spacings 0.2% and 0.06%) chosen so that the ideal operating point, marked with a black star, lies exactly on the grid. The color map shows the three-state average fidelity $\bar{F}$ (%); the white solid contour encloses the 99.9% region and the dashed gold contour the 99.0% region.

V. LINDBLAD MASTER EQUATION VERIFICATION WITH DRESSED-STATE DISSIPATION

A. Dressed-State Lindblad Formalism

In the dispersive regime, dissipation must be described in the dressed eigenbasis of H_0. The Lindblad equation in the Schrödinger picture is

$$\dot{\rho}(t) = -\mathrm{i}[H(t), \rho(t)] + \sum_k \mathcal{D}[L_k]\rho(t), \tag{26}$$

with $\mathcal{D}[L]\rho = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}$. We diagonalize $H_0 = \sum_k E_k \left|k\right\rangle \left\langle k\right|$ and transform bare dissipators via the unitary $V = \sum_k \left|k\right\rangle \left\langle k_{\mathrm{bare}}\right|$:

$$L_k^{\mathrm{dressed}} = V^\dagger L_k^{\mathrm{bare}} V. \tag{27}$$

The Lindblad master equation formalism follows Ref. [41].

B. Decoherence Channels and Intra-state Dephasing

The open-system dynamics are governed by a Lindblad master equation with four primary dissipative channels: transmon relaxation ($\Gamma_1^{\mathrm{q}} = 1/T_1^{\mathrm{q}}$), transmon pure dephasing ($\Gamma_\phi^{\mathrm{q}} = 2/T_\phi^{\mathrm{q}}$), phonon relaxation ($\kappa_1^{\mathrm{c}} = 1/T_1^{\mathrm{c}}$), and phonon pure dephasing ($\kappa_\phi^{\mathrm{c}} = 2/T_\phi^{\mathrm{c}}$). We use a single, self-consistent convention throughout, and we state it explicitly because the choice is easily made silently and shifts every dephasing-induced number by a factor of two. For a number operator n, a dissipator $L = \sqrt{c}\, n$ damps the coherence ρ_{mk} of eigenstates with m and k excitations at the rate $\frac{1}{2}c\,(m-k)^2$. For relaxation we set $c = 1/T_1$, which reproduces the decay rate $1/T_1$ of the $|0\rangle\langle 1|$ coherence. For pure dephasing we set

$$c = \frac{2}{T_\phi}, \qquad \text{i.e.} \qquad L_\phi = \sqrt{\frac{2}{T_\phi}}\, n, \tag{28}$$

and *not* $c = 1/T_\phi$: with $c = 1/T_\phi$ the $|0\rangle\langle 1|$ coherence would decay at only $1/(2T_\phi)$, which is inconsistent with the standard relation $1/T_2 = 1/(2T_1) + 1/T_\phi$. With Eq. (28) the dephasing contribution to ρ_{01} is exactly $1/T_\phi$, so T_ϕ retains its conventional meaning as the pure-dephasing time. All dephasing contributions quoted in this work use Eq. (28).

Crucially, for bosonic encodings such as the binomial code used here, pure dephasing mechanisms induce significant intra-state decoherence rather than mere mean-field phase accumulation. For instance, the logical state $|0_L\rangle \approx (|0_p\rangle + |4_p\rangle)/\sqrt{2}$ is highly susceptible to phonon pure dephasing $L_\phi = \sqrt{\kappa_\phi^{\mathrm{c}}}\, a^\dagger a$. Under this dissipator, the off-diagonal coherence term ρ_{04} decays at an amplified rate proportional to the square of the difference in excitation numbers:

$$\gamma_{04}^{\mathrm{deph}} = \frac{1}{2}\kappa_\phi^{\mathrm{c}}(4-0)^2 = 8\kappa_\phi^{\mathrm{c}} = \frac{16}{T_\phi^{\mathrm{c}}}, \tag{29}$$

where the last equality uses $\kappa_\phi^{\mathrm{c}} = 2/T_\phi^{\mathrm{c}}$ from Eq. (28). This corresponds to an effective time $T_{04} = T_\phi^{\mathrm{c}}/16 \approx 18.75\ \mu$s. This intra-state dephasing strictly dominates the decoherence dynamics of the macroscopic superposition, making a simple expectation value difference approach insufficient. Therefore, we avoid analytical mean-field approximations and directly simulate the full evolution using the Trotterized Lindblad superoperators projected onto the dressed eigenbasis of H_0. This rigorous open-system simulation inherently captures the magnified dephasing rates of multi-phonon Fock superpositions.

We include four dissipative channels:

1. **Transmon energy relaxation** ($T_1^{\mathrm{q}} = 30\ \mu$s): $L_{\mathrm{decay}}^{\mathrm{q}} = \sqrt{1/T_1^{\mathrm{q}}}\, b$.
2. **Transmon pure dephasing** ($T_\phi^{\mathrm{q}} = 30\ \mu$s): $L_{\mathrm{deph}}^{\mathrm{q}} = \sqrt{2/T_\phi^{\mathrm{q}}}\, b^\dagger b$.
3. **Phonon energy relaxation** ($T_1^{\mathrm{c}} = 100\ \mu$s): $L_{\mathrm{decay}}^{\mathrm{c}} = \sqrt{1/T_1^{\mathrm{c}}}\, a$.
4. **Phonon pure dephasing** ($T_\phi^{\mathrm{c}} = 300\,\mu$s): $L_{\mathrm{deph}}^{\mathrm{c}} = \sqrt{2/T_\phi^{\mathrm{c}}}\, a^\dagger a$.

The effective Transmon dephasing rate between dressed logical states depends on the difference of their Transmon excitation numbers:

$$\gamma_{\text{eff}}^{\text{q}} = \frac{1}{T_\phi^{\text{q}}} \left| \langle 0_L | b^\dagger b | 0_L \rangle - \langle 1_L | b^\dagger b | 1_L \rangle \right|^2, \tag{30}$$

as predicted in Refs. [30, 31]. Exact diagonalization gives $\langle 0_L | b^\dagger b | 0_L \rangle = 0.1107$ and $\langle 1_L | b^\dagger b | 1_L \rangle = 0.0762$, hence $\gamma_{\text{eff}}^{\text{q}} \approx 3.97 \times 10^{-5}\ \mu\text{s}^{-1}$, i.e. an effective dephasing time $T_\phi^{\text{eff,q}} \approx 2.5 \times 10^4\ \mu\text{s}$, three orders of magnitude longer than the bare T_ϕ^{q}. Transmon pure dephasing is thus strongly suppressed by the dispersive dressing. We also note that the "8.58% static dressing" quoted in Figs. 4(c) and 5 is the *any-excitation* probability $P(n_q > 0) = \langle 0_L | \Pi_{n_q>0} | 0_L \rangle = 0.0858$, which is strictly smaller than the mean occupation $\langle 0_L | b^\dagger b | 0_L \rangle = 0.1107$. The two differ because the transmon is a multilevel system, and they must not be interchanged.

For the phonon mode the situation differs, and two qualitatively distinct mechanisms must be separated. The first is the intra-state dephasing of Eq. (29), which acts on the ρ_{04} coherence inside $|0_L\rangle$ and corresponds to an effective time $T_{04} = T_\phi^{\text{c}}/16 \approx 18.75\ \mu\text{s}$. The second is the inter-state dephasing between $|0_L\rangle$ and $|1_L\rangle$,

$$\begin{aligned} \gamma_{\text{eff}}^{\text{c}} &= \frac{1}{T_\phi^{\text{c}}} \left(\langle 1_L | a^\dagger a | 1_L \rangle - \langle 0_L | a^\dagger a | 0_L \rangle \right)^2 \\ &\approx 3.97 \times 10^{-6}\ \mu\text{s}^{-1}, \end{aligned} \tag{31}$$

which is negligible. The reason is that exact diagonalization yields $\langle 0_L | a^\dagger a | 0_L \rangle = 1.8893$ and $\langle 1_L | a^\dagger a | 1_L \rangle = 1.9238$: both logical states carry essentially the same mean phonon number—the ideal binomial value is 2 for both, displaced only slightly by dressing—so the inter-state factor nearly cancels. The phonon-dephasing penalty quoted below therefore originates almost entirely from the intra-state mechanism, which is precisely why we simulate the full Lindblad equation rather than analytical mean-field expressions.

C. Lindblad Simulation Results

We simulate the evolution of $|0_L\rangle$, $|1_L\rangle$, and $|+_L\rangle$ with the optimized dual-tone pulses in the dressed eigenbasis using 3000 Trotter steps (operating in the converged 96-dimensional Hilbert space $(N_q, N_p) = (8, 12)$). Figure 4 summarizes the results.

With the standard dephasing convention fixed in Sec. V B, the average logical-state fidelity $\bar{F} = (F_{0_L} + F_{1_L} + F_{+_L})/3 = 96.58\%$ decomposes for the $|0_L\rangle \to |1_L\rangle$ channel as

- **Phonon T_1 decay**: 1.532% (dominant, 42.0% of the summed single-channel losses).
- **Phonon pure dephasing**: 1.126% (30.9%), arising predominantly from the intra-state ρ_{04} mechanism of Eq. (29) rather than from the negligible inter-state rate of Eq. (31). This is the channel that is most sensitive to the dephasing convention of Eq. (28): its contribution is proportional to the dephasing rate, so choosing $c = 1/T_\phi$ instead would roughly halve it. We adopt the $2/T_\phi$ value throughout.
- **Transmon T_1 decay**: 0.263% (7.2%), slightly elevated by the static dressing.
- **Transmon pure dephasing**: 0.726% (19.9%), suppressed relative to the bare value by the dressing factor of Eq. (30).
- **Coherent control errors**: $< 0.01\%$ (negligible, benchmarked against the ideal GRAPE result).

We emphasize that $\bar{F}$ here denotes the arithmetic mean of the three *state* fidelities $|0_L\rangle$, $|1_L\rangle$ and $|+_L\rangle$ evaluated against their respective targets; it is neither an average gate fidelity nor a process fidelity in the strict sense, both of which would require the complete process matrix (a complete X-gate process fidelity would, in addition, require the $|-_L\rangle$ input).

The near-complete refocusing of excess Transmon excitation (from 1.65% down to −0.58%, i.e. within one percent of the static dressing level) validates the virtual bus paradigm: Transmon excitations are coherently borrowed during the gate and returned at its completion. Using the decomposition of Eq. (24), the final populations of the $|0_L\rangle$ input under full four-channel dissipation are $P_L = 96.43\%$, $P_{\text{odd}} = 1.79\%$ and $P_{\text{even,leak}} = 1.78\%$, which by construction sum to 100%. It is the parity-breaking part P_{odd} that a parity check could in principle flag; it is generated by the physical loss events rather than by the control pulse. The population outside the five-dimensional allowed occupation space is a separate, smaller quantity, $1 - P_{\text{allow}} = 0.79\%$ at the end of the gate, and we quote it separately rather than conflating the two notions of leakage. In the absence of an implemented erasure-conversion step, however, both channels currently appear as ordinary errors in the fidelities reported here.

VI. DECOHERENCE CHANNEL DECOMPOSITION

To quantitatively identify the dominant error sources limiting the gate fidelity, we perform a systematic decomposition by selectively disabling individual dissipators. The eight simulation configurations are:

1. Ideal (no dissipation) — coherence-propagated with the same Trotter scheme.
2. Transmon T_1 only.
3. Phonon T_1 only.
4. Transmon pure dephasing only.

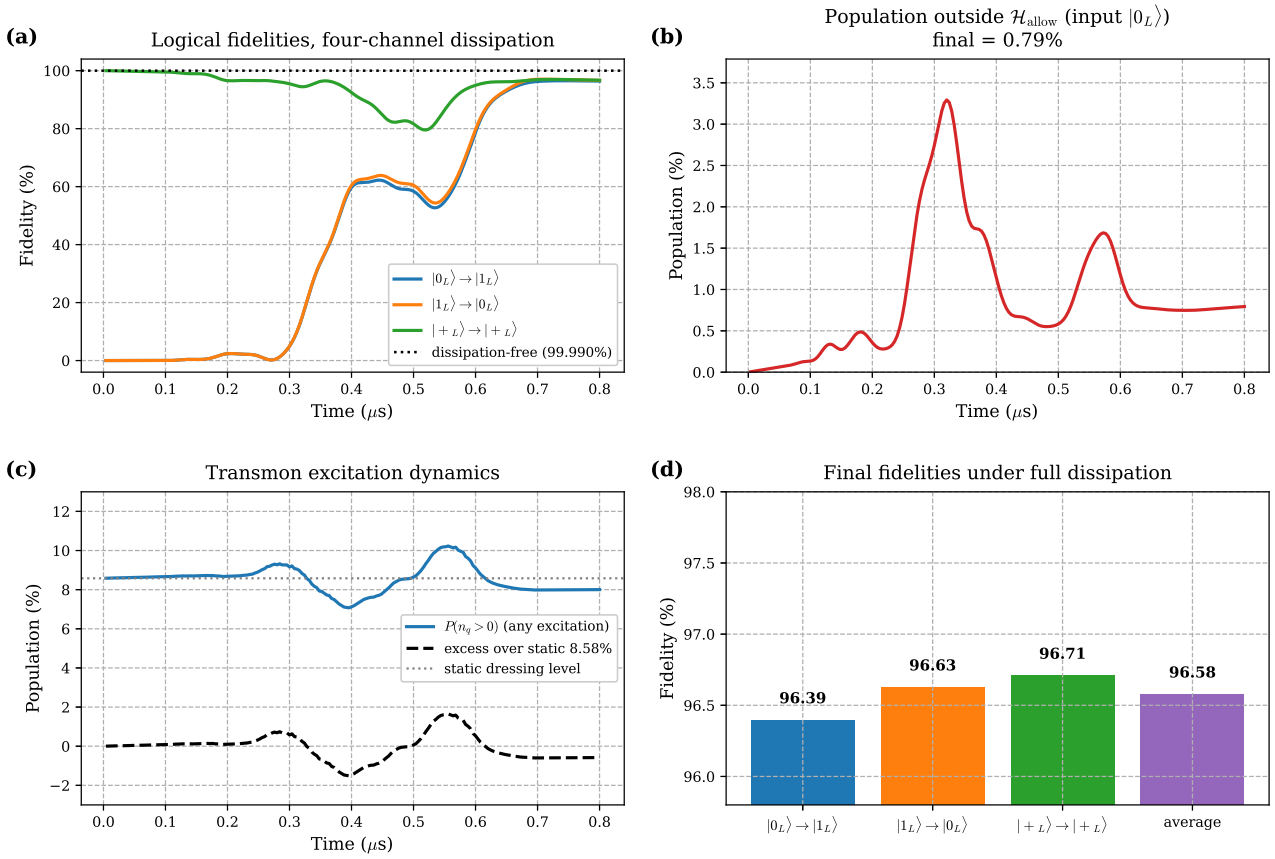


FIG. 4. **Lindblad master equation verification of the logical X gate with full four-channel dissipation.** **(a)** Logical state fidelities: $|0_L\rangle \to |1_L\rangle$ reaches 96.39%, $|1_L\rangle \to |0_L\rangle$ 96.63%, $|+_L\rangle \to |+_L\rangle$ 96.71%. The dissipation-free limit (99.994%) is indicated by the dotted line. **(b)** Population leaving the five-dimensional allowed occupation space (referenced to the $|0_L\rangle$ input), ending at $\approx 0.79\%$. **(c)** Transmon dynamics: the any-excitation probability $P(n_q > 0)$ of the dressed codeword is 8.58%, whereas the mean occupation $\langle 0_L | n_q | 0_L \rangle = 0.1107$ is a different quantity and the two must not be confused. The excess dynamical excitation, referenced to the static dressing level, peaks at 1.65% and returns to -0.58% at $t = T_{\text{gate}}$, i.e. slightly *below* the static dressing level rather than exactly at it; confirming that the Transmon excitation is coherently borrowed and returned, and not accumulated. The curve can take small negative values because it is a difference of two independently fluctuating quantities (the total excitation and the static dressing level), so it is signed throughout. **(d)** Final fidelities: average $\bar{F} = 96.58\%$.

5. Phonon pure dephasing only.

6. Amplitude damping only (Transmon T_1 + Phonon T_1).

7. Dephasing only (Transmon + phonon).

8. All four channels.

The fidelity loss per channel is $\Delta\mathcal{F}_{\text{channel}} = \mathcal{F}_{\text{ideal}} - \mathcal{F}_{\text{channel}}$, where $\mathcal{F}_{\text{ideal}} = 99.990\%$ is the ideal fidelity of the $|0_L\rangle \to |1_L\rangle$ transition (consistent with the ideal limit in Fig. 4). We quote every number below for that single transition, so that the reference and the individual losses share the same definition.

Figure 5 presents the decomposition results. The exact absolute infidelity contributions ($|0_L\rangle \to |1_L\rangle$ transition) are:

- Phonon T_1: 1.532% (42.0% of the summed single-channel losses).
- Phonon pure dephasing: 1.126% (30.9%).
- Transmon T_1: 0.263% (7.2%).
- Transmon pure dephasing: 0.726% (19.9%).

The total dissipation-induced infidelity is 3.596%, with the sum of individual contributions (3.648%) exceeding the total by 0.052% because the channels are not strictly additive. This residual is small, but it is no longer negligible compared with the smallest channel, and we therefore quote the all-channel value rather than the sum whenever a single number is required.

Key findings:

- **Phonon T_1 remains the largest single source** (42.6% of dissipation-induced error).

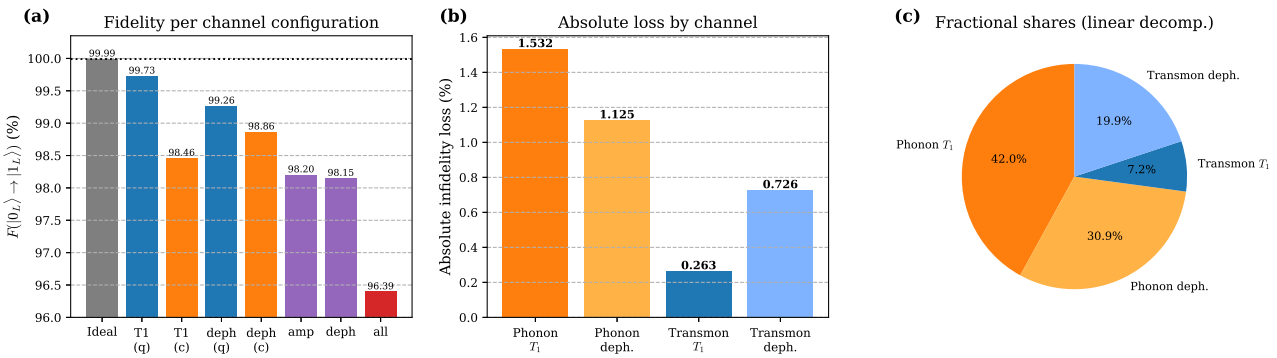


FIG. 5. **Decoherence channel decomposition of the logical X gate (with phonon pure dephasing included).** **(a)** Fidelities for the $|0_L\rangle \to |1_L\rangle$ transition under different configurations: Ideal (99.990%), Transmon T_1 only (99.727%), Phonon T_1 only (98.458%), Transmon dephasing only (99.264%), phonon dephasing only (98.865%), amplitude damping only (98.201%), dephasing only (98.147%), full dissipation (96.395%). The dissipation-free limit 99.990% is indicated by the dotted line. **(b)** *Absolute* infidelity loss by channel, referenced to the dissipation-free ideal: phonon T_1 (1.532%), phonon dephasing (1.126%), Transmon T_1 (0.263%), Transmon dephasing (0.726%). **(c)** Pie chart of the *fractional* shares obtained from the linear (per-channel) decomposition; because of channel non-additivity these shares sum to 100% of 3.648% rather than of the 3.596% all-channel infidelity.

The corresponding single-loss probability is $\int_0^{T_{\rm gate}} \kappa_1^{\rm c} \bar{n}_c(t)\mathrm{d}t \approx 1.5\%$, not the $T_{\rm gate}/T_1^{\rm c} = 0.8\%$ rate-times-time estimate, because the mean phonon number is $\bar{n}_c \simeq 1.9$ rather than unity.

- **Phonon pure dephasing is the second largest contributor** (30.9%, absolute infidelity 1.126%). With $T_\phi^{\rm c} = 300\ \mu$s and the convention of Eq. (28), this channel is dominated by the intra-state ρ_{04} mechanism rather than by the inter-state rate.
- **Transmon decoherence contributes** 27.1% **in total** (0.263% + 0.726%), of which pure dephasing is 19.9%. Compared with the bare transmon dephasing time this is still a strong suppression, as expected from the dressing factor of Eq. (30), but it is no longer a negligible correction.
- **Coherent control errors are negligible** ($<$ 0.01%).

These results provide a clear hardware roadmap:

- **Primary target**: Increase phonon T_1 from 100 μs to $>$ 500 μs (closely comparable phonon lifetimes have been reported in related high-overtone resonators [25]) would reduce the phonon-T_1 contribution by a factor of 5, from 1.532% to $\sim$ 0.31%, raising the single-transition fidelity to $\sim$ 98.5%.
- **Secondary target**: Reduce phonon pure dephasing (e.g., through improved fabrication or flux-insensitive operation) to bring $T_\phi^{\rm c}$ above 1 ms, which would render its contribution below 0.30%.
- **Tertiary target**: Improve Transmon T_1 to 100 μs and $T_\phi^{\rm q}$ to 100 μs to reduce the Transmon-related contributions, which together currently make up about 27% of the budget.

Comparison with Existing Bosonic Logical Gates

Table III places the present result alongside representative oscillator-encoded logical gates. The published entries differ in code family, in the gate that was implemented, and in the fidelity metric that was used, so we state all three explicitly and put the durations on the same footing; the table is not a ranking.

Several corrections relative to the description of these experiments in our earlier drafts are worth recording, because they change the comparison. The Heeres *et al.* gate set is implemented by GRAPE on a four-component cat code, not by selective-number arbitrary-phase (SNAP) gates, and its reported set average is 99.10% (iRB) rather than 96.8%. The Reinhold *et al.* operation is a logical Z rotation on a binomial code, implemented with a Raman SNAP pulse plus an ancilla swap, and its error-corrected interleaved-RB error is 2.4 ± 0.1%; neither the gate type nor the error-correction is captured by a single "gate fidelity" number. The "order of magnitude" speed advantage asserted in our earlier drafts is not supported by these durations: the only entries for which a duration is reported are the Heeres *et al.* gate set (1.1 μs) and the present scheme (0.8 μs), a ratio of about 1.4. We therefore make no such claim. What distinguishes the present scheme is not its speed but the combination of a dissipation-aware optimization with an explicit leakage and erasure budget, together with the ideal-to-dissipative gap, which is the comparison we regard as defensible.

VII. CONCLUSION AND OUTLOOK

We have designed and validated a logical X gate for a transmon-mechanical phononic system. By combining dual-tone frequency-multiplexed control with a leakage-

TABLE III. **Representative oscillator-encoded logical gates.** All experimental entries are logical operations on a 3D microwave cavity dispersively coupled to a transmon. The metric is stated for every entry because the quoted numbers are not interchangeable: interleaved randomized benchmarking (iRB) refers to an average over a gate set, process tomography (PT) to a single operation, and the entries for this work are three-state averages of simulated state fidelities. "n.r." marks a quantity not reported in the cited work. No performance ranking is intended.

Work	**Code**	**Gate**	**Dur.**	**Reported fidelity (metric)**	**Control**
Heeres *et al.* [7]	four-component cat	π, $\pi/2$ rotations, H, T	1.1 μs	$98.89 \pm 0.03\%$ (X_{180}, iRB); $99.10 \pm 0.02\%$ (set avg, iRB)	GRAPE; IQ-modulated transmon and oscillator drives
Hu *et al.* [3]	binomial	universal set incl. logical X	n.r.	97% (set avg, PT)	dispersive, number-selective pulses
Reinhold *et al.* [8]	binomial	logical $Z(\pi/2)$	n.r.	$97.6 \pm 0.1\%$ (corr.); $95.4 \pm 0.1\%$ (uncorr.); iRB	Raman SNAP plus ancilla swap, error-corrected
This work, diss.-free	binomial	logical X	**0.8 μs**	**99.99**% (3-state avg, no dissipation)	**Dual-tone GRAPE**
This work, Lindblad	binomial	logical X	**0.8 μs**	**96.6**% (3-state avg, Lindblad)	**Dual-tone GRAPE**

aware optimization strategy, the ideal average logical fidelity exceeds 99.99% within a gate time of 0.8 μs, and full dressed-state Lindblad simulations yield 96.58% once realistic transmon and phonon decoherence channels are included, with phonon energy relaxation dominating the error budget. The channel decomposition further confirms that transmon-induced dephasing is strongly suppressed by dispersive dressing, contributing 0.73% to the total infidelity.

It is equally important to state what this result does *not* establish. The parity-breaking loss events that leave $\approx 1.8\%$ of the population in the odd-parity subspace are in principle detectable by a parity check, but we have not implemented an erasure-conversion step, nor simulated the ancilla dynamics, the measurement backaction, or the heralding efficiency on which a fault-tolerant scheme would depend. The gate is therefore *potentially* compatible with an erasure-based architecture, subject to a complete error-conversion and fault-tolerance analysis, and should not be presented as a ready-to-use building block for surface codes.

Thermal Noise and Error Budget

In our open-system simulations we focus on the dominant zero-temperature dissipation channels. Thermal phonon population can be neglected in this architecture: for a ~ 5 GHz HBAR resonator operating at 10 mK the thermal occupancy is exponentially suppressed to $n_{\rm th} \approx 10^{-10}$. Under these conditions the total dissipation-induced infidelity is $\sim 3.60\%$.

It is important to keep three distinct quantities separate when discussing fault tolerance, since conflating them is a common source of over-estimates. The first is the *infidelity* reported above, which counts all error mechanisms, heralded or not. The second is the *loss probability*: to first order, the probability of a single phonon loss during the gate is

$$p_{\rm loss} \simeq \int_0^{T_{\rm gate}} \kappa_1^{\rm c}\, \bar{n}_c(t)\, {\rm d}t \approx \frac{T_{\rm gate}}{T_1^{\rm c}}\, \bar{n}_c, \tag{32}$$

which is *not* simply $T_{\rm gate}/T_1^{\rm c}$: with $\bar{n}_c \simeq 1.9$ this gives a loss probability of order 1.5%, roughly twice the naive rate-times-time estimate of 0.8%. The third is the *probability of a detected erasure*, which further depends on the readout duration, the heralding fidelity, and the residual fraction of events that the parity measurement fails to flag. None of these three quantities is a surface-code threshold, and thresholds additionally depend on the outer code, the local gate and measurement errors, the erasure/Pauli mixing ratio, spacetime correlations, and the decoder [16–19]. The commonly quoted values of $\sim 20\%$ for erasure errors and $\sim 1\%$ for depolarizing errors [16, 17] are therefore best understood as idealized reference points, not as budgets that this gate can be said to meet.

On this basis, the error hierarchy reported above suggests—but does not establish—a two-stage strategy. In a first stage, the parity-breaking phonon loss, of order $\sim 1.5\%$, would be the natural target for conversion into heralded erasures, because these events are detectable in principle; our simulations find $\approx 1.8\%$ of the population in the odd-parity subspace, consistent with this estimate once loss and relaxation channels are combined. In a second stage, the residual unheralded errors would be the coherence errors left after the parity-breaking events are removed, i.e. phonon pure dephasing (1.126%) together with the part of the transmon channels that does not change parity. Their aggregate exceeds the idealized $\sim 1\%$ depolarizing reference when the dephasing channels are counted in full, so concluding that the gate is

fault tolerant would require both a parity-resolved error budget and a full error-correction simulation including heralding failures, ancilla backaction and code-cycle timing, none of which is within the scope of the present unitary-control study.

Hardware Roadmap and Future Directions

Despite these promising theoretical advances, practical scale-up faces several hardware challenges. The integration of bulk HBAR resonators with planar superconducting circuits introduces parasitic stray capacitances and requires careful packaging solutions. Moreover, extending this protocol to multi-qubit registers will demand robust mitigation of flux crosstalk between adjacent transmon devices.

Future work will proceed in two phases:

1. **Phase I – Experimental demonstration:** Prototype the proposed single-qubit gate with a transmon-HBAR package, benchmarking the leakage-tolerant pulse against standard GRAPE under varying noise conditions, including drive phase noise and flux drift.

2. **Phase II – Full erasure loop:** Integrate the unitary gate with stochastic simulations of ancilla-based parity measurements, evaluating heralding fidelities, measurement backaction, and real-time feedback. This will close the loop on real-time erasure conversion, extending the design toward two-qubit gates and surface-code implementations.

As mechanical coherence times continue to improve [25], the phonon T_1 contribution will diminish further, and single-gate logical fidelities in excess of 99% become plausible for this architecture; on the roadmap given above, raising the phonon T_1 to 500 μs alone takes the single-transition fidelity to $\sim$ 98.5%, so 99% requires progress on more than one channel. Whether such a gate can contribute to a fault-tolerant scheme depends on questions we have deliberately not answered here: the efficiency and backaction of the parity check, the residual undetected-error budget, and the behaviour of the pulse under realistic drive phase noise and flux drift.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 11872335) and by Zhejiang Provincial Natural Science Foundation of China (Grant No. Y6110314).

Appendix A: Derivation of the Transmon-Mechanical Static Hamiltonian

In this appendix, we provide a rigorous derivation of the static system Hamiltonian H_0 used in the logical X-gate optimization. Addressing the physical feasibility, the system consists of a standard frequency-tunable transmon qubit capacitively coupled to a bulk acoustic wave (BAW) mechanical phonon mode. The transmon serves as a genuine multi-level quantum element providing the necessary Kerr anharmonicity for bosonic encoding, while the mechanical mode acts as the control interface.

1. Full Circuit Hamiltonian

The frequency-tunable transmon consists of a superconducting quantum interference device (SQUID) loop shunted by a large capacitance C_Σ. The total Hamiltonian of the coupled transmon-mechanical system is:

$$H_{\text{full}} = H_{\text{transmon}} + H_{\text{mech}} + H_{\text{int}}, \tag{A1}$$

where the individual components are:

$$H_{\text{transmon}} = 4E_C\hat{n}^2 - E_J(\Phi_{\text{ext}})\cos(\varphi), \tag{A2}$$

$$H_{\text{mech}} = \omega_m a^\dagger a, \tag{A3}$$

$$H_{\text{int}} = g_0\hat{n}(a + a^\dagger). \tag{A4}$$

Here, $\hat{n}$ is the Cooper-pair number operator conjugate to the superconducting phase difference φ, and $E_C = e^2/(2C_\Sigma)$ is the charging energy. The flux-tunable Josephson energy is $E_J(\Phi_{\text{ext}}) = E_{J\Sigma}\cos(\pi\Phi_{\text{ext}}/\Phi_0)$, where Φ_{ext} is the external magnetic flux. We emphasize that this cosine form is the one appropriate to a *split-junction* (asymmetric) transmon SQUID, which is the element assumed throughout this work; a symmetric rf-SQUID loop instead has a quadratic flux response, $E_J(\Phi) \approx E_J - E_J\pi^2\Phi^2/(4\Phi_0^2)$, and would give a materially different flux-to-frequency transfer function. The mechanical mode has a resonant frequency ω_m, with a ($a^\dagger$) being the phonon annihilation (creation) operator, and g_0 is the bare coupling strength. In the present model g_0 enters through the Cooper-pair number $\hat{n}$, i.e. the transmon charges the piezoelectric transducer capacitively, which then converts the electrical field into strain; the conversion is assumed lossless and is absorbed into g.

2. Transmon Quantization and Anharmonicity

Operating in the standard transmon regime where $E_J \gg E_C$, the phase fluctuations are small. Expanding the cosine potential in Eq. (A2) to the fourth order in φ:

$$-E_J\cos(\varphi) \approx -E_J + \frac{1}{2}E_J\varphi^2 - \frac{1}{24}E_J\varphi^4. \tag{A5}$$

Expressing φ and $\hat{n}$ in terms of the transmon ladder operators b and $b^\dagger$, the transmon Hamiltonian takes the standard weakly anharmonic oscillator form:

$$H_{\text{transmon}} \approx \omega_q b^\dagger b - \frac{\alpha}{2}b^{\dagger 2}b^2, \tag{A6}$$

where $\omega_q \approx \sqrt{8E_J E_C} - E_C$ is the transmon transition frequency, and the Kerr anharmonicity is given by $\alpha \approx E_C$. In our architecture, we adopt a highly realistic and routinely achieved parameter $\alpha/2\pi = 250$ MHz, strictly avoiding unrealistically large nonlinearities.

3. Mechanical Mode Quantization

The mechanical phonon mode is quantized as a harmonic oscillator:

$$H_{\text{mech}} = \omega_m a^\dagger a, \tag{A7}$$

where $\omega_m/2\pi$ is typically in the GHz range for high-overtone bulk acoustic wave resonators (HBARs). The mechanical displacement field $\hat{x} \propto a + a^\dagger$ couples to the transmon's electric field via the piezoelectric effect.

4. Piezoelectric Coupling and Rotating Wave Approximation

The piezoelectric interaction Hamiltonian in the Cooper-pair number basis is:

$$H_{\text{int}} = g_0 \hat{n}(a + a^\dagger). \tag{A8}$$

Expressing the number operator in terms of the transmon ladder operators: $\hat{n} = \mathrm{i} n_{\text{zpf}}(b^\dagger - b)$, where $n_{\text{zpf}} = (E_J/32E_C)^{1/4}$ is the zero-point fluctuation amplitude. The interaction becomes:

$$H_{\text{int}} = \mathrm{i} g_0 n_{\text{zpf}}(b^\dagger - b)(a + a^\dagger). \tag{A9}$$

In the dispersive regime where $|\omega_q - \omega_m| \gg g \equiv g_0 n_{\text{zpf}}$, we apply the rotating wave approximation (RWA), keeping only the energy-conserving terms:

$$H_{\text{int}}^{\text{RWA}} = g(b^\dagger a + b a^\dagger), \tag{A10}$$

where g is the effective piezoelectric coupling strength. The counter-rotating terms $b^\dagger a^\dagger$ and ba are heavily suppressed by factors of $1/(\omega_q + \omega_m)$ and are thus neglected.

5. Static Hamiltonian in the Rotating Frame

Combining all components and moving to a rotating frame at the mechanical frequency ω_m, we define the transmon-phonon detuning $\Delta = \omega_q - \omega_m$. The full static Hamiltonian becomes:

$$\boxed{H_0 = \Delta\, b^\dagger b - \frac{\alpha}{2}\, b^{\dagger 2} b^2 + g\,(b^\dagger a + b a^\dagger)}. \tag{A11}$$

This equation defines the exact physical system optimized in this work.

6. Parameter Regime and Justification

The parameters used throughout this work are:

- $\Delta/2\pi = 400$ MHz: the transmon-mechanical detuning, which fixes the scale of the second-order dispersive correction g^2/Δ;
- $\alpha/2\pi = 250$ MHz: the transmon anharmonicity, a standard charging-energy (E_C) value in state-of-the-art cQED experiments;
- $g/2\pi = 80$ MHz: the piezoelectric coupling strength.

We state the justification for this set conservatively. With $g/\Delta = 0.2$ the rotating-wave approximation is reliable, but the second-order shift $g^2/\Delta \approx 16$ MHz is comparable to the $\sim$ 15–18 MHz dressed level spacings that the encoded manifold actually exhibits. The dispersive picture is therefore adequate as a qualitative description of the virtual-bus role but is not a controlled expansion; all results in this work are obtained from the exact diagonalization of H_0 rather than from a perturbative truncation.

What actually limits the coupling. We further note that $g/2\pi = 80$ MHz is an ambitious target for the present architecture rather than a demonstrated value. It is worth being precise about which quantity it has to support. A natural estimate is that the gate time scales inversely with the effective coupling, so that reducing g by a factor f would lengthen the protocol by the same factor and a coupling near 20 MHz would still permit a gate of a few microseconds. Our calculations do not support that estimate, and we replace it with a quantitative one.

Two facts fix the picture. First, the speed of the *drive* is not set by g: the two-phonon operator of Eq. (13) has dressed-basis matrix elements that are almost coupling-independent — we find $\langle\tilde{2}|a^2|\tilde{0}\rangle/\sqrt{2}$ rising only from 0.963 to 0.999, and $\langle\tilde{4}|a^2|\tilde{2}\rangle/\sqrt{6}$ from 1.343 to 1.413, as $g/2\pi$ is reduced from 80 to 10 MHz. Second, what g does control is the dressed nonlinearity

$$K \equiv \omega_{d2} - \omega_{d1}, \tag{A12}$$

that is, the separation of the two transitions addressed by the two carriers of Eqs. (10)–(11). K is a fourth-order quantity in the dispersive regime: the exact diagonalization reproduces $K \propto g^4$ at small g, steepening to a local exponent of 4.4 at $g/2\pi = 80$ MHz.

Table IV collects the consequences, obtained by evaluating the same dressed-basis calculation at reduced g. Two behaviours are visible. The pulse optimised for $g/2\pi = 80$ MHz transfers badly once the coupling is reduced, even with the carriers retuned to the new transitions, falling to 17.9% at 60 MHz and to 13.1% at the 0.26 MHz of Ref. [28]. Re-optimising at each coupling recovers most of the dissipation-free fidelity, but it does so

by driving far more population out of the encoded manifold: the maximum transient leakage rises from 2.86% at 80 MHz to 62% at 60 MHz and to 72–82% at 40 MHz, far above the $\max_t L(t) \leq 9.5\%$ criterion used throughout this work. Lengthening the gate does not repair this — at $g/2\pi = 40$ MHz an 8 μs gate reaches a 99.78% dissipation-free average but leaks 82%, and once dissipation is included the longer gate evaporates its own advantage, giving 55.5% under the coherence times assumed here.

The sharpest way to state the requirement is to compare with an existing device. The closest realised system to the architecture assumed here — a flux-tunable transmon piezoelectrically coupled through an AlN transducer to the phonon modes of a sapphire HBAR, that is, the same physical platform, with measured phonon $T_1 = 17$ μs and $T_2 = 27$ μs — is the device of Ref. [28], which reports $g/2\pi = 260 \pm 10$ kHz. Evaluating the present model at that coupling gives $K/2\pi \simeq 2.6 \times 10^{-10}$ MHz, and the pulse optimised here retains only 13.1% of its dissipation-free fidelity. The assumed $g/2\pi = 80$ MHz is thus about two and a half orders of magnitude above the demonstrated value. It is also the value the protocol requires, and not merely a convenient one: at $g/2\pi = 60$ MHz the re-optimised run of Table IV recovers 99.59% of dissipation-free fidelity but only by leaking 61.7% transiently, which exceeds the $\max_t L(t) \leq 9.5\%$ criterion adopted throughout this work. We therefore do not present 60 MHz as sufficient; on the evidence available to us the leakage-tolerant regime is reached at the nominal $g/2\pi = 80$ MHz, and we have not located the threshold between the two values. We state this plainly: the control strategy is the contribution of this work, while its experimental realisation depends on a substantial advance in piezoelectric coupling, which we identify as the principal bottleneck of the proposal rather than as a detail of its parameter choice. Establishing the practical applicability of the sub-microsecond regime therefore requires an experimental advance in the transmon–piezoelectric coupling, and we present the 0.8 μs figure as a design target conditional on that advance.

7. Truncation and Numerical Implementation

We truncate the transmon Hilbert space to $N_q = 8$ levels and the mechanical Hilbert space to $N_p = 12$ levels, yielding a total dimension of dim = 96. Both cutoffs were fixed by the convergence study of Sec. III H rather than chosen by convenience, and the transmon cutoff in particular is not a cosmetic parameter.

Crucially, keeping $N_q = 8$ ensures that the transmon is *not* artificially reduced to a two-level qubit limit. With $\alpha/2\pi = 250$ MHz, the higher transmon levels ($|2_q\rangle, |3_q\rangle$, etc.) actively participate as virtual intermediate states during the fast gate execution. This validates its fundamental role as a genuine multi-level bosonic virtual bus. The mechanical truncation to $N_p = 12$ provides a sufficient buffer space to accommodate transient multi-phonon excursions during the leakage-aware optimization, and it is already converged: increasing N_p to 14 or 16 at fixed $N_q = 5$ changes the frozen-pulse fidelity by only 2×10^{-4} percentage points.

Why N_q could not be left at 5 is a direct consequence of the block structure of $H_0 = \Delta b^\dagger b - \frac{\alpha}{2} b^{\dagger 2} b^2 + g(b^\dagger a + b a^\dagger)$, which conserves $N_{\rm tot} = n_q + n_p$ and is therefore block diagonal in $N_{\rm tot}$. The logical states live in the blocks $N_{\rm tot} = 0, 2, 4$—and those three blocks are shifted by nothing, because they are identical for any $N_q \geq 5$. The blocks $N_{\rm tot} \geq 5$, however, gain members $|n_q = 5\rangle, |n_q = 6\rangle, \ldots$ that a 5-level truncation simply removes; through the coupling g these missing members displace the *shared* $n_q \leq 4$ part of the high-$N_{\rm tot}$ blocks by 39, 63, 76 and 87 MHz for $N_{\rm tot} = 5$–8. That displacement is comparable to the drive bandwidth (twelve harmonics around carriers at 31.3 and 34.3 MHz), so a leakage-aware optimum found in the 60-dimensional space is not converged: frozen and re-evaluated in any space with $N_q \geq 7$, it falls to 65.26%, while a pulse optimized directly in (8, 12) reaches 99.9939% in every space with $N_q \geq 7$. Full tables are given in Sec. III H.

All propagation is performed single-threaded in double precision. The time step is $dt = T_{\rm gate}/N_{\rm steps}$ with $N_{\rm steps} = 3000$, and the control is sampled at the midpoint of each interval, $t_{j+1/2} = (j + \frac{1}{2})dt$; the left-endpoint sampling used in our earlier implementation carries an $O(dt)$ timing error that is not removed by step refinement. The relative phase of the code states is fixed by Eq. (6), i.e. by $\langle 0_q, 4_p | \tilde{4} \rangle > 0$, because the two sign choices define different (though both even-parity) code spaces and give different gate fidelities.

The full Hamiltonian H_0 is constructed using sparse matrices and exact diagonalization is performed to obtain the dressed eigenstates used for the logical binomial encoding.

[1] Cai W, Ma Y, Wang W, Zou C L and Sun L 2021 *Fundam. Res.* **1** 50–67 ISSN 2667-3258 URL https://www.sciencedirect.com/science/article/pii/S2667325820300145

[2] Michael M H, Silveri M, Brierley R T, Albert V V, Salmilehto J, Jiang L and Girvin S M 2016 *Phys. Rev. X* **6**(3) 031006 URL https://link.aps.org/doi/10.1103/PhysRevX.6.031006

[3] Hu L, Ma Y, Cai W *et al.* 2019 *Nat. Phys.* **15** 503 URL https://doi.org/10.1038/s41567-018-0414-3

[4] Leghtas Z, Touzard S, Pop I M, Kou A, Vlastakis B, Petrenko A, Sliwa K M, Narla A, Shankar S, Hatridge M J, Reagor M, Frunzio L, Schoelkopf R J, Mirrahimi M and Devoret M H 2015 *Science* **347** 853–857 ISSN 1095-9203

TABLE IV. **Sensitivity of the protocol to the piezoelectric coupling.** All entries use the exact dressed basis at the stated g; K is the dressed nonlinearity of Eq. (A12) and $\bar{F}_{\mathrm{ideal}}$ the dissipation-free three-state average. "Frozen" denotes the pulse optimised at $g/2\pi = 80$ MHz with the two carriers retuned to the new transitions; "re-optimised" denotes a fresh GRAPE+Adam run at that coupling with the same leakage penalty ($w_{\mathrm{leak}} = 20$). The last row is the measured coupling of the closest realised transmon–HBAR device [28], evaluated with the frozen pulse. Dashes in the last column mark cases in which the Lindblad propagation was not run.

$g/2\pi$ (MHz)	$K/2\pi$ (MHz)	T_{gate} (μs)	**Pulse**	$\bar{F}_{\mathrm{ideal}}$ (%)	$\max_t L(t)$ (%)	$\bar{F}_{\mathrm{Lindblad}}$ (%)
80	3.055	0.8	nominal	99.994	2.86	96.58
60	0.851	0.8	frozen	17.88	53.6	—
60	0.851	0.8	re-optimised	99.59	61.7	—
60	0.851	1.3	re-optimised	99.68	59.8	—
40	0.154	0.8	frozen	18.70	78.3	—
40	0.154	0.8	re-optimised	85.86	71.5	—
40	0.154	8.0	re-optimised	99.78	82.0	55.47
0.26	2.6×10^{-10}	0.8	frozen	13.14	79.0	—

URL http://dx.doi.org/10.1126/science.aaa2085

[5] Gottesman D, Kitaev A and Preskill J 2001 *Phys. Rev. A* **64**(1) 012310 URL https://link.aps.org/doi/10.1103/PhysRevA.64.012310

[6] Campagne-Ibarcq P, Eickbusch A, Touzard S, Zalys-Geller E, Frattini N E, Sivak V V, Reinhold P, Puri S, Shankar S, Schoelkopf R J, Frunzio L, Mirrahimi M and Devoret M H 2020 *Nature* **584** 368–372 ISSN 1476-4687 URL http://dx.doi.org/10.1038/s41586-020-2603-3

[7] Heeres R W, Reinhold P, Ofek N, Frunzio L, Jiang L, Devoret M H and Schoelkopf R J 2017 *Nat. Commun.* **8** 94 URL https://doi.org/10.1038/s41467-017-00045-1

[8] Reinhold P, Rosenblum S, Ma W L, Frunzio L, Jiang L and Schoelkopf R J 2020 *Nat. Phys.* **16** 822–826 ISSN 1745-2481 URL http://dx.doi.org/10.1038/s41567-020-0931-8

[9] Ofek N, Petrenko A, Heeres R, Reinhold P, Leghtas Z, Vlastakis B, Liu Y, Frunzio L, Girvin S M, Jiang L, Mirrahimi M, Devoret M H and Schoelkopf R J 2016 *Nature* **536** 441–445 ISSN 1476-4687 URL http://dx.doi.org/10.1038/nature18949

[10] Wood C J and Gambetta J M 2018 *Phys. Rev. A* **97**(3) 032306 URL https://link.aps.org/doi/10.1103/PhysRevA.97.032306

[11] Eickbusch A, Sivak V, Ding A Z, Elder S S, Jha S R, Venkatraman J, Royer B, Girvin S M, Schoelkopf R J and Devoret M H 2022 *Nat. Phys.* **18** 1464–1469 URL https://doi.org/10.1038/s41567-022-01776-9

[12] Krastanov S, Albert V V, Shen C, Zou C L, Heeres R W, Vlastakis B, Schoelkopf R J and Jiang L 2015 *Phys. Rev. A* **92**(4) 040303 URL https://link.aps.org/doi/10.1103/PhysRevA.92.040303

[13] Knill E, Leibfried D, Reichle R *et al.* 2008 *Phys. Rev. A* **77** 012307 URL https://link.aps.org/doi/10.1103/PhysRevA.77.012307

[14] Wu Y, Kolkowitz S, Puri S and Thompson J D 2022 *Nat. Commun.* **13** 4657 URL https://doi.org/10.1038/s41467-022-32094-6

[15] Kubica A, Haim A, Vaknin Y, Levine H, Brandão F and Retzker A 2023 *Phys. Rev. X* **13**(4) 041022 URL https://link.aps.org/doi/10.1103/PhysRevX.13.041022

[16] Stace T M, Barrett S D and Doherty A C 2009 *Phys. Rev. Lett.* **102** 200501 URL https://link.aps.org/doi/10.1103/PhysRevLett.102.200501

[17] Fowler A G, Mariantoni M, Martinis J M and Cleland A N 2012 *Phys. Rev. A* **86** 032324 URL https://link.aps.org/doi/10.1103/PhysRevA.86.032324

[18] Wang D S, Fowler A G and Hollenberg L C L 2011 *Phys. Rev. A* **83**(2) 020302 URL https://link.aps.org/doi/10.1103/PhysRevA.83.020302

[19] Stephens A M 2014 *Phys. Rev. A* **89**(2) 022321 URL https://link.aps.org/doi/10.1103/PhysRevA.89.022321

[20] Koch J, Yu T M, Gambetta J, Houck A A, Schuster D I, Majer J, Blais A, Devoret M H, Girvin S M and Schoelkopf R J 2007 *Phys. Rev. A* **76**(4) 042319 URL https://link.aps.org/doi/10.1103/PhysRevA.76.042319

[21] Schreier J A, Houck A A, Koch J, Schuster D I, Johnson B R, Chow J M, Gambetta J M, Majer J, Frunzio L, Devoret M H, Girvin S M and Schoelkopf R J 2008 *Phys. Rev. B* **77**(18) 180502 URL https://link.aps.org/doi/10.1103/PhysRevB.77.180502

[22] Barends R, Kelly J, Megrant A, Sank D, Jeffrey E, Chen Y, Yin Y, Chiaro B, Mutus J, Neill C, O'Malley P, Roushan P, Wenner J, White T C, Cleland A N and Martinis J M 2013 *Phys. Rev. Lett.* **111**(8) 080502 URL https://link.aps.org/doi/10.1103/PhysRevLett.111.080502

[23] Goryachev M, Creedon D L, Ivanov E N, Galliou S, Bourgeois R and Tobar M E 2012 *Appl. Phys. Lett.* **100** 243504 URL https://doi.org/10.1063/1.4729292

[24] Goryachev M, Creedon D L, Galliou S and Tobar M E 2013 *Phys. Rev. Lett.* **111**(8) 085502 URL https://link.aps.org/doi/10.1103/PhysRevLett.111.085502

[25] MacCabe G S, Ren H, Luo J *et al.* 2020 *Science* **370** 840 URL https://doi.org/10.1126/science.abc7312

[26] Arrangoiz-Arriola P, Wollack E A, Wang Z, Pechal M, Jiang W, McKenna T P, Witmer J D, Van Laer R and Safavi-Naeini A H 2019 *Nature* **571** 537–540 URL https://doi.org/10.1038/s41586-019-1386-x

[27] O'Connell A D, Hofheinz M, Ansmann M, Bialczak R C, Lenander M, Lucero E, Neeley M, Sank D, Wang H, Weides M, Wenner J, Martinis J M and Cleland A N 2010 *Nature* **464** 697–703 URL https://doi.org/10.1038/nature08967

[28] Chu Y, Kharel P, Renninger W H, Burkhart L D, Frunzio L, Rakich P T and Schoelkopf R J 2017 *Science* **358**

199–202 ISSN 1095-9203 URL http://dx.doi.org/10.1126/science.aao1511

[29] Chu Y, Kharel P, Yoon T, Frunzio L, Rakich P T and Schoelkopf R J 2018 *Nature* **563** 666–670 ISSN 1476-4687 URL http://dx.doi.org/10.1038/s41586-018-0717-7

[30] Gambetta J, Blais A, Schuster D I, Wallraff A, Frunzio L, Majer J, Devoret M H, Girvin S M and Schoelkopf R J 2006 *Phys. Rev. A* **74**(4) 042318 URL https://link.aps.org/doi/10.1103/PhysRevA.74.042318

[31] Boissonneault M, Gambetta J M and Blais A 2008 *Phys. Rev. A* **77**(6) 060305(R) URL https://link.aps.org/doi/10.1103/PhysRevA.77.060305

[32] Boissonneault M, Gambetta J M and Blais A 2009 *Phys. Rev. A* **79**(1) 013819 URL https://link.aps.org/doi/10.1103/PhysRevA.79.013819

[33] Blais A, Huang R S, Wallraff A, Girvin S M and Schoelkopf R J 2004 *Phys. Rev. A* **69**(6) 062320 URL https://link.aps.org/doi/10.1103/PhysRevA.69.062320

[34] Schuster D I, Houck A A, Schreier J A, Wallraff A, Gambetta J M, Blais A, Frunzio L, Majer J, Johnson B, Devoret M H, Girvin S M and Schoelkopf R J 2007 *Nature* **445** 515–518 URL https://doi.org/10.1038/nature05461

[35] Houck A A, Schreier J A, Johnson B R, Chow J M, Koch J, Gambetta J M, Schuster D I, Frunzio L, Devoret M H, Girvin S M and Schoelkopf R J 2008 *Phys. Rev. Lett.* **101**(8) 080502 URL https://link.aps.org/doi/10.1103/PhysRevLett.101.080502

[36] Khaneja N, Reiss T, Kehlet C, Schulte-Herbrüggen T and Glaser S J 2005 *J. Magn. Reson.* **172** 296–305 ISSN 1090-7807 URL https://www.sciencedirect.com/science/article/pii/S1090780704003696

[37] Machnes S, Assémat E, Tannor D and Wilhelm F K 2018 *Phys. Rev. Lett.* **120**(15) 150401 URL https://link.aps.org/doi/10.1103/PhysRevLett.120.150401

[38] Caneva T, Calarco T and Montangero S 2011 *Phys. Rev. A* **84**(2) 022326 URL https://link.aps.org/doi/10.1103/PhysRevA.84.022326

[39] Trotter H F 1959 *Proc. Am. Math. Soc.* **10** 545 URL https://doi.org/10.1090/S0002-9939-1959-0108732-6

[40] Abdelhafez M, Schuster D I and Koch J 2019 *Phys. Rev. A* **99** ISSN 2469-9934 URL http://dx.doi.org/10.1103/physreva.99.052327

[41] Breuer H P and Petruccione F 2007 *The Theory of Open Quantum Systems* (Oxford University PressOxford) ISBN 9780191706349 URL http://dx.doi.org/10.1093/acprof:oso/9780199213900.001.0001

[42] Palacios-Laloy A, Nguyen F, Mallet F, Bertet P, Vion D and Esteve D 2008 *J. Low Temp. Phys.* **151** 1034–1042 URL https://doi.org/10.1007/s10909-008-9774-x

[43] Sandberg M, Wilson C M, Persson F, Bauch T, Johansson G, Shumeiko V, Duty T and Delsing P 2008 *Appl. Phys. Lett.* **92** 203501 URL https://doi.org/10.1063/1.2929367

[44] Masluk N A, Pop I M, Kamal A, Minev Z K and Devoret M H 2012 *Phys. Rev. Lett.* **109**(13) 137002 URL https://link.aps.org/doi/10.1103/PhysRevLett.109.137002

[45] von Lüpke U, Yang Y, Bild M, Michaud L, Fadel M and Chu Y 2022 *Nat. Phys.* **18** 794–799 ISSN 1745-2481 URL http://dx.doi.org/10.1038/s41567-022-01591-2

[46] Kingma D P and Ba J 2015 *Proc. Int. Conf. Learn. Represent. (ICLR)* URL https://arxiv.org/abs/1412.6980