

Fermilab

Fast Sideband Control of a Weakly Coupled Multimode Bosonic Memory

FERMILAB-PUB-25-0171-SQMS

arXiv:2503.10623

This manuscript has been authored by Fermi Forward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Fast Sideband Control of a Weakly Coupled Multimode Bosonic Memory

Jordan Huang,^{1,*} Thomas J. DiNapoli,¹ Gavin Rockwood,¹ Ming Yuan,² Prathyankara Narasimhan,¹ Eesh Gupta,³ Mustafa Bal,⁴ Francesco Crisa,⁴ Sabrina Garattoni,⁴ Yao Lu,⁴ Liang Jiang,² and Srivatsan Chakram^{1,†}

¹*Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA*

²*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*

³*Department of Applied Physics, Stanford University, Stanford, California 94305, USA*

⁴*Superconducting Quantum Materials and Systems Division,*

Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA

(Dated: March 14, 2025)

Circuit quantum electrodynamics (cQED) with superconducting cavities coupled to nonlinear circuits like transmons offers a promising platform for hardware-efficient quantum information processing. A critical challenge is developing control protocols that mitigate cavity errors arising from ancilla-induced decoherence, nonlinearities, and crosstalk. We address this by weakening the dispersive coupling while also demonstrating fast, high-fidelity multimode control by dynamically amplifying gate speeds through transmon-mediated sideband interactions. This approach enables transmon-cavity SWAP gates, for which we achieve speeds up to 30 times larger than the bare dispersive coupling. Combined with transmon rotations, this allows for efficient, universal state preparation in a single cavity mode, though achieving unitary gates and extending control to multiple modes remains a challenge. In this work, we overcome this by introducing two sideband control strategies: (1) a shelving technique that prevents unwanted transitions by temporarily storing populations in sideband-transparent transmon states and (2) a method that exploits the dispersive shift to synchronize sideband transition rates across chosen photon-number pairs to implement transmon-cavity SWAP gates that are selective on photon number. We leverage these protocols to prepare Fock and binomial code states across any of ten modes of a multimode cavity with millisecond cavity coherence times. We demonstrate the encoding of a qubit from a transmon into arbitrary vacuum and Fock state superpositions, as well as entangled NOON states of cavity mode pairs—a scheme extendable to arbitrary multimode Fock encodings. Furthermore, we implement a new binomial encoding gate that converts arbitrary transmon superpositions into binomial code states in 4 μ s (less than $1/\chi$), achieving an average post-selected final state fidelity of 96.3% across different fiducial input states. By using precalibrated transmon and sideband pulses, our work demonstrates multimode control with significantly reduced calibration requirements, enabling efficient unitary operations using sideband interactions in multimode cQED systems.

I. INTRODUCTION

Superconducting cavities coupled to nonlinear, ancillary circuits form a promising platform for quantum computing based on circuit quantum electrodynamics (cQED) [1]. This architecture leverages the superconducting circuit's nonlinearity to manipulate quantum information stored in high-coherence cavity modes [2]. When coupled to a multimode cavity, this architecture also allows multiplexed control of many cavity modes with only a few control lines [3, 4]. Universal control of a harmonic oscillator has been realized through a variety of control schemes [5, 6]. These control schemes have been extended to multiple oscillators to demonstrate entanglement [7, 8] and gate operations between multiple cavity modes, using both fixed-frequency transmons [9, 10] and tunable coupler circuits [11, 12]. The harmonic oscillator's expanded Hilbert space and straightforward decoherence mechanisms—primarily photon loss—have also enabled demonstrations of hardware-efficient quantum error correction (QEC) using bosonic codes [13–17].

The dispersive coupling to a nonlinear circuit, while crucial for implementing universal cavity control, introduces ancilla-mediated errors that are frequently the dominant error channels. In particular, the cavity modes inherit decay from the lossy, nonlinear circuit through the inverse-Purcell effect, setting a fundamental upper bound on cavity coherence—an effect that becomes increasingly significant as cavity lifetimes are extended [18–21]. In multimode architectures, the always-on dispersive interaction and the nonlinearities inherited by the cavity modes introduce coherent crosstalk between modes [22], further contributing to gate infidelity and placing constraints on the scalability of these systems.

One approach to mitigating these challenges is to weaken the dispersive interaction, thereby reducing the magnitude of ancilla-mediated errors. However, this reduction in interaction strength also slows gate speeds in traditional control schemes that rely on photon-number-selective ancilla rotations [5, 8, 23, 24]. To address this trade-off, control strategies utilizing conditional cavity displacements have been developed. In these schemes, interaction rates are enhanced beyond the bare dispersive coupling by using large cavity displacements as an interaction switch [6, 25]. Another method for increasing gate speeds beyond the dispersive coupling strength

* jah499@scarletmail.rutgers.edu

† schakram@physics.rutgers.edu

is to use ancilla displacements to activate charge-driven sideband interactions [19]. The transmon's quartic non-linearity can activate various parametric processes [26], such as cavity beamsplitters [27, 28] and transmon-mode SWAP gates using two drive photons [29]. The fastest transmon-mediated sideband interaction for a given drive strength uses a single drive photon to convert two transmon excitations into a single photon in a coupled cavity mode [9, 30].

We use such charge-driven sidebands to realize a tunable Jaynes-Cummings interaction (within the transmon's $|g\rangle$ - $|f\rangle$ manifold) between the transmon and any target mode of a multimode cavity. Here, the coupling is modulated by the strength of the transmon drive. When combined with transmon rotations, this enables efficient, universal state preparation using the Law-Eberly protocol [31]—originally developed to control the motional states of trapped ions [32] and later demonstrated in superconducting circuits [33]. This protocol reverses the sequence that brings the target state to the vacuum one photon at a time and has a sequence length sub-linear in the maximum photon number of the target state. However, incommensurate sideband interaction rates across different photon-number states from the $\sqrt{n}$ bosonic enhancement pose significant challenges in implementing arbitrary gate operations [34]. Furthermore, calibrating the arbitrary transmon and sideband rotation angles required for a given unitary can be challenging, especially in the presence of large Stark shifts. This difficulty is analogous to that encountered when realizing target unitaries via optimal control pulses generated through algorithms such as GRAPE [5, 35]. In traditional dispersive control [23] or with conditional displacement gates [6], employing discrete families of pulses that can be independently calibrated and optimized offers advantages in both ease of implementation and robustness against classical control errors and quantum state-dependent shifts [36, 37].

In this work, we overcome these challenges by introducing two sideband control strategies: the first is a shelving technique that temporarily stores populations in the transmon $|e\rangle$ state so that they remain unaffected by sideband drives. The second method implements photon-number-selective sideband operations by matching sideband rates across chosen photon numbers, using the sideband drive detuning as a control knob. Using these strategies, we introduce a method of state preparation and unitary synthesis that is implemented through a discrete set of precalibrated transmon and sideband pulses. Using these techniques, we prepare binomial code states [38] across ten cavity modes in approximately 1 μ s. We also implement unitary operations that encode a qubit from the transmon into vacuum and Fock state superpositions, as well as into entangled NOON states between chosen cavity mode pairs. Finally, we implement a new encoding unitary that transfers a transmon superposition into a binomial logical qubit in any target mode at a rate exceeding the dispersive shift. We demon-

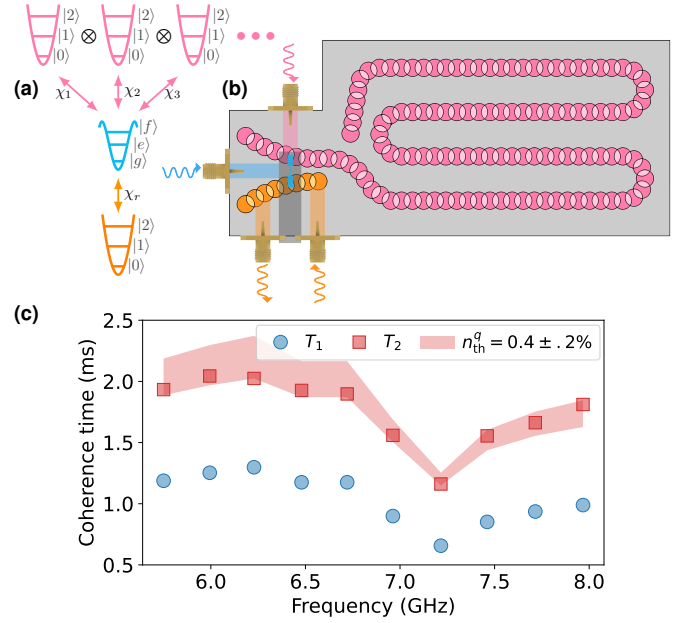


FIG. 1. **Multimode bosonic memory device.** (a) Device architecture and (b) schematic depicting the multimode cavity-based quantum processor [4, 8] comprised of a multimode cavity (pink) with cavity modes coupled to and controlled by a single transmon circuit (blue), which is measured using a readout cavity (orange). The cavities are fabricated using the flute method [4, 8], by drilling offset, overlapping, evanescent holes in a monolithic block of high-purity (5N5) aluminum. The hole locations are indicated in the schematic. (c) The measured coherence times of the storage cavity modes. The shaded regions indicate the inferred qubit thermal population that would give rise to the measured cavity T_2 .

strate these multimode operations on a cavity memory with ten active modes with a weak dispersive coupling to a transmon ancilla. In this regime, we achieve SWAP operations that are 30 times faster than the bare dispersive coupling.

II. MULTIMODE BOSONIC MEMORY

Our system architecture and device schematic are shown in Fig. 1(a) and (b). The device features a multimode rectangular waveguide cavity supporting several high-coherence modes, each with lifetimes ranging from 0.6–1.3 ms, as shown in Fig. 1(c). The cavity is fabricated from a monolithic block of high-purity (5N5) aluminum using the flute method, where the cavity volume is formed by drilling offset, overlapping, evanescent holes [4]. The mode spectrum is set by the cavity dimensions, with the lowest mode around 5.75 GHz. We realize a nearly evenly spaced mode spectrum, with separations of approximately 250 MHz, by tapering the cavity height. The transmon is inserted at the taller end of the cavity

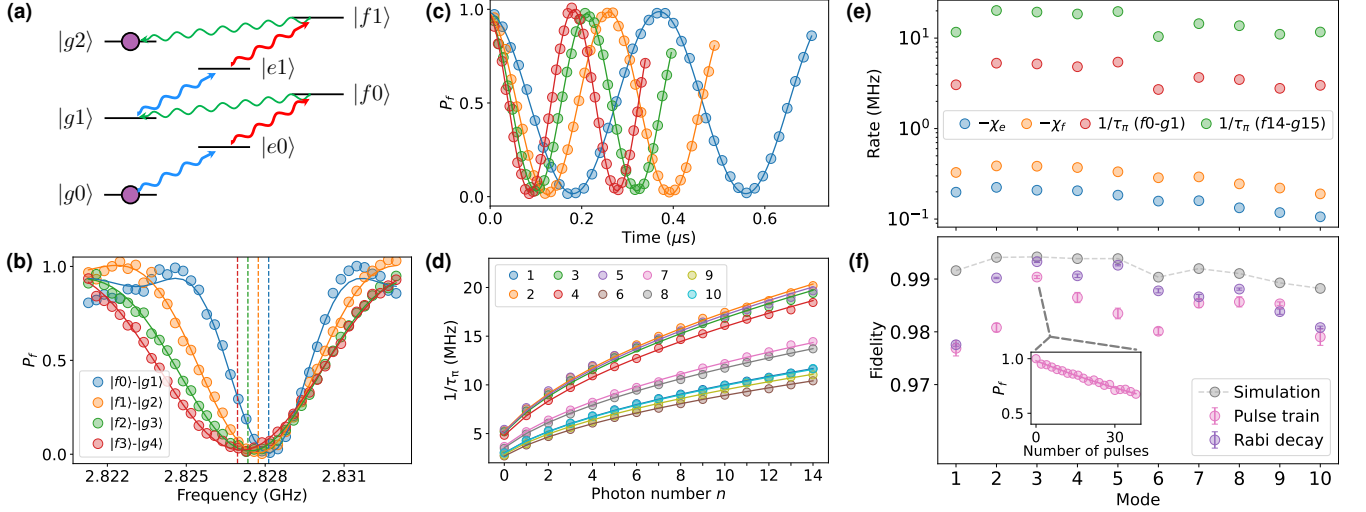


FIG. 2. **Charge-driven sideband interactions in the multimode memory.** (a) Energy level diagram for a transmon and a single cavity mode illustrating how to climb the levels with transmon rotations (blue and red) and $|f, n\rangle$ - $|g, n+1\rangle$ sideband transitions (green). (b) Spectroscopy and (c) and stimulated vacuum Rabi oscillations of $|f, n\rangle$ - $|g, n+1\rangle$ transitions for mode 3 of the storage cavity. (d) Sideband Rabi oscillation rate as a function of photon number for the first ten modes of the multimode memory, up to a photon number of $n = 14$. (e) Comparison of the maximum achieved SWAP rate for the $|f0\rangle$ - $|g1\rangle$ and $|f13\rangle$ - $|g14\rangle$ transitions against the bare dispersive interaction strengths for the $|e\rangle$ and $|f\rangle$ states. (f) The pink markers show the $|f0\rangle$ - $|g1\rangle$ SWAP fidelities inferred from a pulse train experiment. The inset shows the pulse train experiment performed for mode three. The data is fit to an exponential decay as a function of pulse number to extract the fidelity. The purple markers show the $|f0\rangle$ - $|g1\rangle$ SWAP fidelities inferred from fitting the $|f0\rangle$ - $|g1\rangle$ oscillations to an effective Rabi model as in [12]. The grey markers show the sideband interaction fidelities extracted from Lindblad master equation simulations.

with transmon-mode couplings chosen to realize dispersive shifts of $\chi \approx 100$ – 200 kHz. With the transmon's $T_1 \approx 56$ μ s, this leads to an inverse-Purcell limit of about 60 ms for the coherence times of the cavity modes due to their coupling to the lossier transmon. The transmon frequency is $f_q = 4.606$ GHz, its dephasing time is $T_2 \approx 66$ μ s, and the thermal population is 0.3 ± 0.1 %. The transmon is also coupled to a separate flute-style cavity whose fundamental mode at $f_r = 8.067$ GHz is used for transmon readout. In addition to readout input and output drives, the system has direct transmon and multimode storage cavity drive lines.

III. FAST CHARGE-DRIVEN SIDEBAND INTERACTIONS

The quartic nonlinearity of the transmon mediates a sideband interaction between the states $|f, n\rangle$ and $|g, n+1\rangle$, where n is the number of photons in the target cavity mode [39]. This interaction is initiated by applying a charge drive on the transmon at the frequency difference between its $|f\rangle$ state and the cavity mode. While these states do not directly couple, an interaction is facilitated through virtual states which are coupled to $|f, n\rangle$ and $|g, n+1\rangle$ through the dressing from the transmon drive and the Jaynes–Cummings coupling with the target

cavity mode (the virtual states being $\{|e, n\rangle, |e, n+1\rangle\}$ for small drive amplitudes). The resulting sideband interaction can be described as $g_{sb,i} (|g\rangle \langle f| b_i^\dagger + |f\rangle \langle g| b_i)$, where b_i is the lowering operator for photons in cavity mode i . Its interaction rate, computed to lowest order in the nonlinearity, without making a rotating wave approximation for the transmon displacement, is given by

$$g_{sb,i} = \frac{\sqrt{2}(\omega_q + K + \Delta_i)}{2\omega_q + K + \Delta_i} \frac{\epsilon}{g_i} \chi_i \approx \frac{\epsilon}{\sqrt{2}g_i} \chi_i. \quad (1)$$

Here, ω_q is the transmon's $|g\rangle$ - $|e\rangle$ frequency, K is its anharmonicity, $\Delta_i = \omega_q - \omega_{c,i}$ is the transmon-mode detuning, g_i is the bare transmon-mode coupling, and χ_i is the dispersive shift of the transmon $|e\rangle$ state from addition of a photon in mode i . In this form, it is clear that the sideband interaction can be enhanced beyond χ with a strong drive ϵ . Higher-order corrections to the sideband rate are discussed in Section S3 B of the Supplementary Information.

In conjunction with transmon rotations, the sideband interaction allows us to climb the ladder of Fock levels of any cavity mode, as schematically illustrated in the level diagram shown in Figure 2(a). We measure the resonance frequency of the $|f, n\rangle$ - $|g, n+1\rangle$ transition by performing sideband spectroscopy measurements after preparing the transmon in the $|f\rangle$ state. We start with the cavity in the

ground state and locate the $|f, 0\rangle$ - $|g, 1\rangle$ resonance. Next, we sweep the pulse duration to identify the π -pulse time corresponding to adding a single photon in the target cavity mode. This implements a transmon-cavity SWAP gate. The process is iteratively repeated for higher photon numbers, identifying the $|f, n\rangle$ - $|g, n+1\rangle$ transitions and extracting the corresponding π pulse times (shown in Figure 2(b) and (c)). We observe that the resonance frequency of the n -photon sideband is offset by $n\chi_f$ due to the dispersive shift of the $|f\rangle$ state. Additionally, the interaction rates exhibit a $\sqrt{n}$ bosonic enhancement with photon number as shown in Figure 2(d). Figure 2(e) compares the rate of the $|f, 0\rangle$ - $|g, 1\rangle$ SWAP gate to the dispersive interaction (χ_e, χ_f) of the modes. We achieve SWAP gate times nearly 30 times faster than χ_e by using strong transmon drive strengths ranging from 0.1–1.3 GHz. At these drive strengths, the transition frequency has a significant drive-amplitude-dependent Stark shift (approximately 10–50 MHz). Here, the Floquet modes adiabatically connected to the undriven transmon states have significant support with higher transmon levels. To maximize the interaction fidelity, we use a smooth ramp (bump function) that is slow enough to allow adiabatic evolution of the undriven states to their corresponding Floquet modes and fast enough to induce high-fidelity population exchange. This corresponds to the ramp initializing the initial state into an equal superposition of Floquet modes at the final drive amplitude through a Landau-Zener-like transition (see Section S7 of the Supplementary Information).

We characterize the fidelity of the sideband SWAP gate using an error amplification experiment where we measure the $|f\rangle$ state population following a train of $|f0\rangle$ - $|g1\rangle$ π -pulses. An optimally calibrated sideband pulse results in a nearly exponential decay of the population, as shown in the inset of Fig. 2(f), whereas calibration errors lead to non-exponential behavior due to coherent over- or under-rotations. We fit the exponential decay to extract a pulse fidelity, which includes errors from decoherence and leakage to higher transmon levels. To isolate the contribution of decoherence alone, we fit the decay of a long-time sideband Rabi oscillation to an effective Rabi decay model [12]. The fidelities extracted from both pulse train and Rabi decay measurements are shown in Fig. 2(f), along with a comparison with master equation simulations. The pulse-train fidelities range from approximately 97.5–99% and are found to be marginally lower than the Rabi decay fidelity, attributed to residual leakage errors caused by imperfections in the pulse ramps. The fidelities from Rabi decay are found to be consistent with master equation simulations that account for transmon and cavity decay and dephasing. In addition to decoherence and leakage errors, the pulses used in the gate protocols are susceptible to coherent errors arising from calibration drifts during the measurements. These drifts are attributed to drive-amplitude fluctuations, which lead to frequency errors due to the large Stark shifts. We mitigate this by performing iterative cal-

ibrations of the sideband frequencies and π -pulse times for all target photon numbers before each experiment. We detail the procedure followed to calibrate and characterize the sideband pulses in Section S4 of the Supplementary Information. At sufficiently high sideband drive strengths, which vary between modes, fidelities degrade due to transmon ionization [40], imposing a fundamental limit on gate speed. To avoid this, we constrain the drive strength below this threshold.

We also measure the thermal population of the cavity modes by using a new sideband-based protocol that extends the standard protocol used for transmon thermal population measurements [41]. This is estimated by comparing the contrast of $|f0\rangle$ - $|g1\rangle$ sideband rotations with and without initially preparing the transmon in $|f\rangle$, yielding cavity-mode thermal populations in the range of 0.3–2%. To further reduce thermal excitations, we implement transmon reset using the $|f0\rangle$ - $|g1\rangle$ sideband with our readout resonator, which has a short lifetime of approximately 600 ns. Combined with sidebands to the storage modes, we implement a sequential sideband reset protocol applicable to any storage mode. In this protocol, detailed in Section S6 of the Supplementary Information, the cavity state is reset one photon at a time by shuttling photons to the transmon $|f\rangle$ state and removing it through the readout mode. This increases our experimental repetition rate by two orders of magnitude (7 ms to 50 μ s) and cools the cavity modes to a thermal population of 0.2–0.7%, consistent with the expected thermal population of our readout resonator, as inferred from transmon dephasing.

IV. SINGLE-MODE SIDEBAND CONTROL

We demonstrate single-mode control over any cavity mode by preparing both Fock states $|n\rangle$ and vacuum-Fock superpositions $\frac{|0\rangle+|n\rangle}{\sqrt{2}}$ by climbing the Jaynes-Cummings ladder with transmon rotations and sideband pulses. Fock states are prepared using a sequence of broadband π_{ge} and π_{ef} pulses followed by $|f, n\rangle$ - $|g, n+1\rangle$ sideband π pulses. To achieve superposition states, we employ our *shelving* method—temporarily storing part of the state in the $|e\rangle$ manifold, which is transparent to the $|f, n\rangle$ - $|g, n+1\rangle$ sidebands. We track the evolution of each component throughout the sequence, arranging it so that only one component is selectively transferred via sideband transitions while the others remain unaffected. Without shelving, the components would evolve at different rates due to the incommensurate sideband interaction rates across different photon numbers. Figure 3(a) shows the pulse sequence used to prepare the superposition state $\frac{|0\rangle+|2\rangle}{\sqrt{2}}$, with shelving pulses indicated by blue underlines. Figure 3(b) depicts the corresponding state evolution, with the numbers corresponding to the pulse number and the arrows indicating the superposition state at the end of

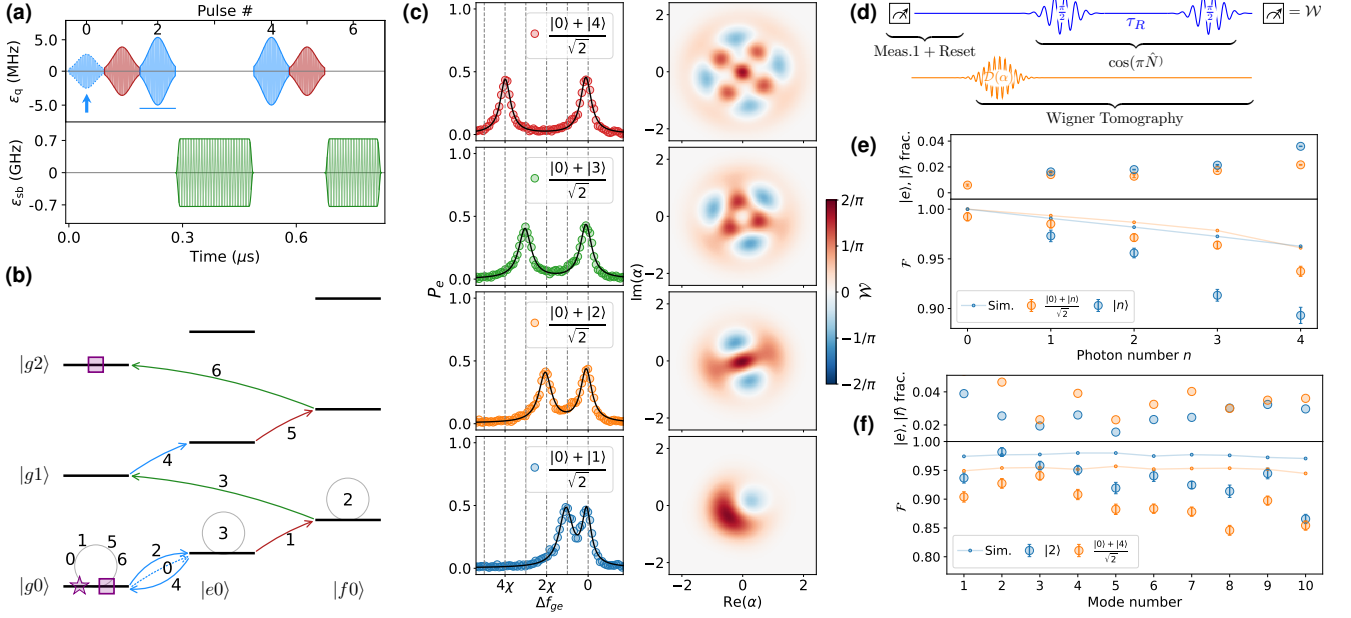


FIG. 3. **Single-mode control using transmon rotations and sideband transitions.** (a) Pulse sequence for preparing $(|0\rangle + |2\rangle)/\sqrt{2}$ in a target mode in ~ 800 ns, corresponding to $\approx \frac{1}{6\chi}$ for this mode (3). The arrow indicates the initial pulse used to prepare a transmon superposition, while the blue underline denotes a shelving pulse. An arbitrary superposition of $\cos\theta|0\rangle + e^{i\varphi}\sin\theta|n\rangle$ can be encoded by adjusting the initial transmon pulse rotation. (b) Energy level diagram representing the combined Hilbert space of the transmon and one of the cavity modes, illustrating the state transitions induced by the pulse sequence in (a). The purple star indicates the initial state, and the purple squares indicate the final superposition state. The numbers indicate the states occupied after each pulse in the sequence. The circles denote the state remaining unchanged by the corresponding pulse. (c) (Left) Photon number-resolved spectroscopy and (right) Wigner tomography of the cavity state after preparing $(|0\rangle + |n\rangle)/\sqrt{2}$ for $n = 1-4$, post-selected on the transmon being in $|g\rangle$. For the preparation of $|n\rangle$ in any cavity mode, the shelving pulses are omitted and the transmon is initialized in $|e\rangle$ with a π_{ge} pulse. (d) The pulse sequence used for Wigner tomography with post-selection, comprising an initial transmon measurement followed by a standard Wigner tomography sequence that includes a cavity displacement and subsequent parity measurement. (e) (Top) Fraction of shots not in $|g\rangle$ in the first measurement as a function of n following preparation of $|n\rangle$ and $(|0\rangle + |n\rangle)/\sqrt{2}$. (Bottom) Corresponding fidelities obtained from Wigner tomography conditioned on the transmon being in $|g\rangle$. (f) (Top) Fraction of population not in $|g\rangle$ and (bottom) Wigner tomography fidelities post-selected on the transmon being in $|g\rangle$ after preparing binomial code states in any of the cavity modes.

each pulse. By adjusting the initial transmon pulse (indicated by the blue arrow), we note that the sequence implements a unitary operation that encodes transmon superpositions into arbitrary vacuum-Fock state superpositions, $\cos\theta|0\rangle + e^{i\varphi}\sin\theta|n\rangle$.

We characterize the prepared states by performing photon-number-resolved transmon spectroscopy to obtain the Fock-state populations of the cavity state. We also use Wigner tomography to fully characterize the cavity state density matrix. These protocols assume that the transmon is in the $|g\rangle$ state and initially disentangled from the cavity. However, decoherence and coherent calibration errors can lead to residual entanglement between the transmon and cavity modes at the end of gate operations. We mitigate the resulting uncontrolled state-characterization errors by measuring the transmon before state characterization and post-selecting on outcomes where the transmon is in $|g\rangle$. To minimize the ef-

fect of cavity decoherence during the initial post-selection measurement, we implement a fast integrated readout and reset pulse [42], detailed in Section S8 C of the Supplementary Information. Figure 3(c) (left) shows the photon-number-resolved spectroscopy of the transmon, while Figure 3(c) (right) shows the corresponding reconstructed Wigner functions of the prepared superposition states up to $n = 4$, both after post-selection. The pulse sequence used for Wigner tomography with post-selection is shown in Figure 3(d).

The fidelities for Fock ($|n\rangle$) and vacuum-Fock superposition states ($(|0\rangle + |n\rangle)/\sqrt{2}$) as a function of n for mode three are shown in Figure 3(e) (bottom), with the fraction of shots not in $|g\rangle$ shown on top. Similarly, the fidelities for preparing binomial code states in any of the ten modes of the cavity, post-selected on the transmon in $|g\rangle$, are shown in Figure 3(f). The experimental results are compared with master simulations incorporating transmon

and cavity decoherence. The discrepancy is attributed to a combination of imperfect calibration of the sideband pulses, and underestimating the fidelity of the state due to noise in the Wigner tomography measurements. The error bars for the fidelities are extracted by considering both the noise in the data and the uncertainty in the parity measurement calibration. The noise in the data leads to both a spread in the reconstructed fidelity as well as a systematic lowering. The calibration and error analysis for Wigner tomography are detailed in Section S8 D of the Supplementary Information. We note that the weak dispersive coupling results in a long duration for the parity measurement ($1/(2\chi) \sim 2.5 \mu\text{s}$), and the cavity-state measurement is affected by ancilla decoherence. This can potentially be avoided in the weak coupling limit by instead measuring the characteristic function of the cavity state through conditional displacements [6, 15]. In our case, to obtain accurate state fidelities, we employ an error mitigation technique by rescaling the parity measurement to account for transmon decoherence during the parity measurement (see Section S8 A in the Supplementary Information). We also account for the reduction in contrast due to the finite bandwidth of the parity measurement.

V. MULTIMODE ENTANGLEMENT

We extend the protocol used to encode vacuum-Fock superpositions to multiple modes and prepare entangled NOON states of the form $\cos \theta |N0\rangle + e^{i\varphi} \sin \theta |0N\rangle$. The pulse sequence to implement this is shown in Figure 4(a). Shelving pulses to the $|e\rangle$ manifold (underlined in blue) ensures that sideband pulses only act on one state of the superposition at a time. The angles θ, φ are controlled by the initial transmon $|g\rangle$ - $|e\rangle$ pulse (indicated by an arrow), and the sequence following it realizes a unitary that encodes the qubit into a NOON state superposition for a cavity state starting off in the vacuum.

We first prepare symmetric maximally entangled NOON states ($\theta = \pi/2$) and characterize them using photon-number-resolved spectroscopy of the transmon, shown in Figure 4(b) (performed here without post-selection). The dispersive interaction leads to the cavity states $|N0\rangle$ and $|0N\rangle$ having peaks in the transmon spectrum that differ in frequency by $N\Delta\chi$, where $\Delta\chi$ is the difference in dispersive shifts between the two modes. The similar dispersive shifts for the modes of our system ($\Delta\chi \approx 30 \text{ kHz}$) allow us to only resolve the peaks for higher photon number NOON states ($N > 2$). However, the experiment demonstrates that the state does not have populations in other photon numbers to under 2%. To confirm the coherence of our state preparation, we perform a Ramsey-like experiment where we decode the NOON states $|N0\rangle$ and $|0N\rangle$ back to the transmon $|g\rangle$ and $|e\rangle$ states. We then apply a $\frac{\pi}{2}$ -pulse with varying phase and observe oscillations that verify the coherence of the prepared NOON states. By varying the delay time

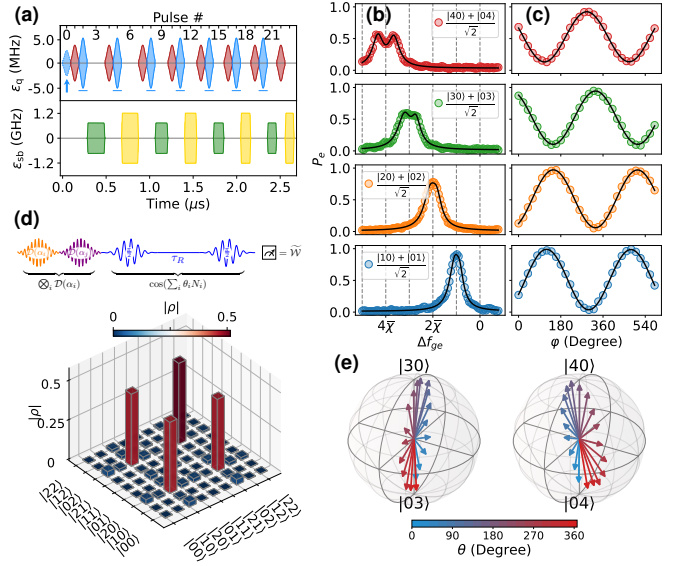


FIG. 4. Preparation and characterization of two-mode entangled states. (a) Pulse sequence for preparing a NOON state with $N = 4$. The initial pulse preparing an arbitrary transmon $|g\rangle$ - $|e\rangle$ superposition is indicated by an arrow while the blue underlines indicate the shelving pulses. (b) Photon-number-resolved spectroscopy of NOON states (between modes three and five) up to $N = 4$. $\bar{\chi}$ is the average dispersive shift of the two modes. (c) Coherence of the NOON state measured using a $\pi/2$ pulse with varying phase after decoding the NOON state back to the transmon $|g\rangle$ - $|e\rangle$ subspace. (d) The density matrix of a NOON state ($N = 2$) extracted using multimode Wigner tomography (e) Tomography of $|\psi\rangle = \cos \theta |N0\rangle + e^{i\varphi} \sin \theta |0N\rangle$ versus θ within the NOON state subspace, for $N = 3, 4$. This was measured by decoding the NOON state back into the transmon $|g\rangle$ - $|e\rangle$ subspace and performing qubit tomography.

between the encode and the decode operations, we can measure the decay time of the NOON state, finding it to be $1.27 \pm 0.05 \text{ ms}$ for $N = 1$ (see Section S10 of the Supplementary Information).

Additionally, we characterize NOON states up to $N = 2$ using multimode Wigner tomography [8, 43]. This is implemented through the pulse sequence shown in Figure 4(d) (top) by displacing both cavities to a set of chosen displacements and measuring a generalized parity-like operator $\tilde{\Pi} = \cos(\theta_1 \hat{n}_1 + \theta_2 \hat{n}_2)$, where $\hat{n}_i$ is the photon number in mode i . The measurement of $\tilde{\Pi}$ is performed through a transmon Ramsey experiment with delay time τ_R , so that $\theta_i = \chi_i \tau_R \approx \pi$. The density matrix is then obtained by inverting the measurements while enforcing physicality constraints. The result for $N = 2$ is shown in Figure 4(d) (bottom). We extract post-selected fidelities of $\mathcal{F} = 0.95 \pm 0.012$ and $\mathcal{F} = 0.905 \pm 0.012$ for NOON states with $N = 1$ and 2, respectively. Lastly, we verify that we can accurately prepare any state within the $\{|N0\rangle, |0N\rangle\}$ subspace by varying the amplitude and

phase of the initial transmon pulse, decoding the state, and subsequently performing qubit tomography. We visualize the resulting states on the $\{|N0\rangle, |0N\rangle\}$ Bloch sphere, as shown in Figure 4(e) for $N = 3, 4$. This confirms the expected trajectory and the ability to prepare the state at an arbitrary point in the encoded NOON state subspace.

The space of two-mode superposition states that can be realized with sideband transitions and transmon rotations, by shelving in the $|e\rangle$ manifold, is restricted to basis states of the form $\{|n0\rangle, |0p\rangle\}$. We introduce a second control technique that allows for the implementation of unitary operations that encode arbitrary multimode Fock states $\{|n_1 n_2 \dots\rangle, |p_1 p_2 \dots\rangle\}$, as shown in Section S4 of the Supplementary Information. This method addresses the challenge of realizing gates amid incommensurate sideband transition rates across different photon numbers by leveraging the residual dispersive interaction, which is not available on other hardware platforms, such as trapped ions and mechanical resonators. As shown in Figure 2(b), the dispersive interaction causes sideband transitions corresponding to different photon numbers ($|f, n\rangle \rightarrow |g, n+1\rangle$) to vary in frequency by $n\chi_f$. These dispersive shifts allow us to use the detuning of the sideband drive as a control parameter to synchronize transitions for different photon numbers. This enables *photon-number-selective* (PNS) sideband transitions, where we perform a π pulse on a specific photon number while effectively performing no transition (e.g., 2π rotation) on another state. This technique is similar to those applied in other superconducting gate protocols [11, 44].

VI. SIDEBAND-BASED ENCODING UNITARY FOR THE BINOMIAL CODE

We demonstrate our new method of unitary synthesis by implementing an encoding gate for the binomial code—an operation that maps an arbitrary qubit in the transmon $|g\rangle$ - $|e\rangle$ subspace, $u|g\rangle + v|e\rangle$, onto a corresponding superposition of binomial code states, $u|2\rangle + v\frac{|0\rangle+|4\rangle}{\sqrt{2}}$. While this unitary has previously been demonstrated in the strong-dispersive regime [14, 16] (with $\chi \sim 2$ MHz), we demonstrate a fast encoding gate in the weak-dispersive regime with nearly an order of magnitude smaller χ than in previous demonstrations. We implement our sideband-based binomial encoding unitary using the pulse sequence shown in Figure 5(a). The sequence is again composed entirely of sideband and transmon rotations, with the key addition of two PNS sidebands, marked by the black arrows. The state transitions effected by the PNS sidebands (following pulse 13 of the sequence) are illustrated in the energy level diagram in Figure 5(b). The complete sequence of state transitions for both $|g0\rangle$ and $|e0\rangle$ input states is provided in Fig. S15 of the Supplementary Information.

In our scheme, the PNS sidebands are crucial for separating different cavity Fock states (e.g., $|g2\rangle$ and $|g4\rangle$)

into distinct transmon manifolds, enabling these states to be addressed either simultaneously with transmon pulses or independently with fast, unselective sideband pulses through the use of shelving. Specifically, the two PNS sidebands perform a π -pulse for the $|f1\rangle \leftrightarrow |g2\rangle$ and $|f0\rangle \leftrightarrow |g1\rangle$ transitions, while implementing 2π and 4π rotations, respectively, for the $|f3\rangle \leftrightarrow |g4\rangle$ transition. These selective sidebands are the rate-limiting steps of the operation, scaling with the dispersive interaction but benefiting from favorable scaling factors due to their dependence on the dispersive shift of the $|f\rangle$ state and the involvement of higher photon numbers in the binomial code (see Section S9 of the Supplementary Information). With this, we implement the binomial encoding gate in a time shorter than the timescale set by the $|e\rangle$ dispersive shift, $1/\chi_e$, making the gate fast in an intermediate- χ regime where Purcell-limited cavity lifetimes can still be long. This gate speed can be further enhanced by using the transmon straddling regime [45] to enhance χ_f without a commensurate increase in cavity Purcell decay. In our work, the inverse-Purcell-limited cavity lifetime is approximately 60 ms, which is comparable to the best reported lifetimes in niobium-based cavities [18–20].

We characterize the unitary gate by performing Wigner tomography of the cavity states following the encoding operation, after initializing the transmon at six cardinal points on the transmon $|g\rangle$ - $|e\rangle$ Bloch sphere, as shown in Figure 5(e). The Wigner tomography is implemented with post-selection to characterize the cavity state accurately. The resulting states are consistent with expectations and are further verified by photon-number-resolved spectroscopy of the transmon, shown in Figure 5(f), which displays peaks corresponding to the expected even Fock states, with populations consistent with those extracted from Wigner tomography to $\sim 2\%$. We illustrate the encoded unitary by projecting the density matrices extracted from Wigner tomography into the subspace of binomial code states and reconstructing the states in the Bloch sphere of logical qubit, as shown in Figure 5(g). The fraction of the shots with the transmon not in $|g\rangle$ range from 8–9% as shown in Figure 5(h) (top). The post-selected fidelities (on $|g\rangle$) for the different fiducial states are shown in Figure 5(h) (bottom) and range from 95.6–96.9%. These results are consistent with master equation simulations (see Section S9 in the Supplementary Information) and are near the coherence limit. The detailed error budget for the Binomial encoding gate is shown in Fig. S17 of the Supplementary Information. We find the gate infidelity to be dominated by ancilla decoherence, even for the post-selected fidelities. With state-of-the-art transmon coherence times ($T_1, T_\varphi = 500, 200 \mu\text{s}$), we expect post-selected $|g\rangle$ fractions of 0.3% and binomial encode fidelities of $\geq 99\%$ without post selection. The fidelities can also be improved by using mid-circuit measurements and feed-forward. These can be selectively implemented only for the slower PNS sidebands to significantly reduce infidelity.

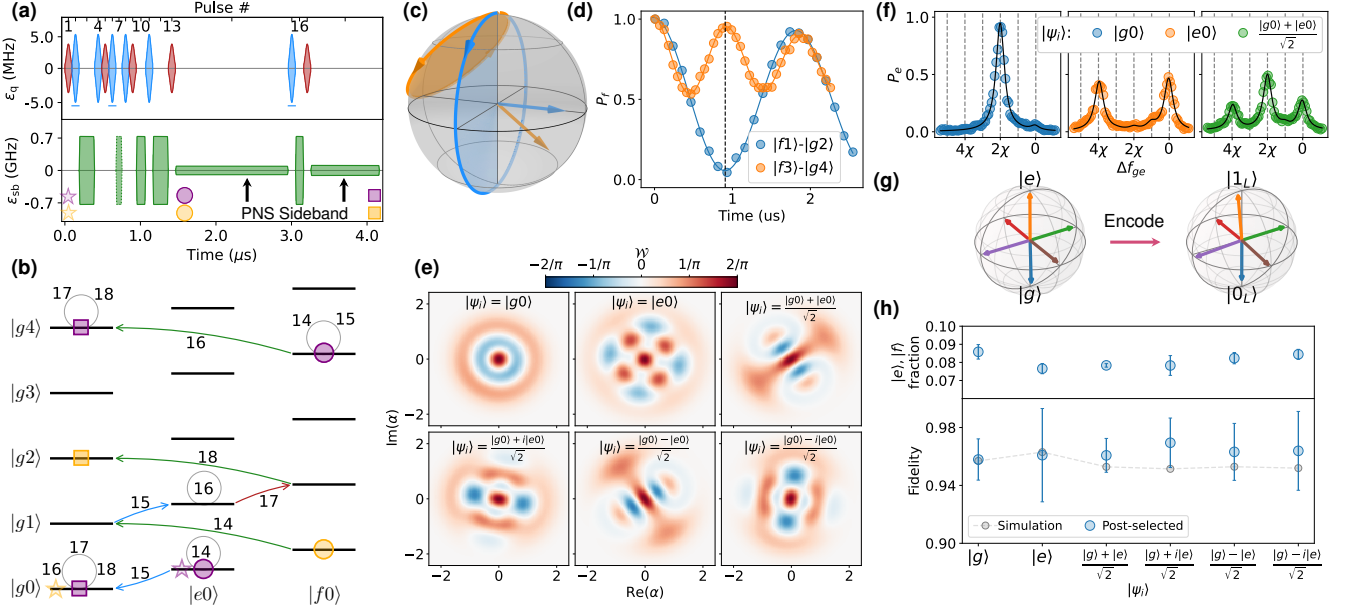


FIG. 5. **Unitary encoding operation for the binomial code.** (a) Pulse sequence for an encoding operation which maps transmon states to binomial code states ($|g\rangle \rightarrow |2\rangle$ and $|e\rangle \rightarrow (|0\rangle + |4\rangle)/\sqrt{2}$). The rate-limiting photon-number-selective (PNS) sidebands are indicated by the arrows and the shelving pulses are indicated by underlines. (b) Energy level diagram of the combined Hilbert space of the transmon and a cavity mode. The gold circles represent the state after pulse 13 of the encoding sequence for the initial state of $|g0\rangle$ (gold star), and the purple circle corresponds to the initial state of $|e0\rangle$ (purple star). The subsequent pulses use PNS sidebands to bring the state back onto the $|g\rangle$ manifold of the transmon, completing the encoding gate. State transitions are tracked by arrows, while the gray circles denote no state change for that particular superposition branch during a given pulse. The final binomial code states are indicated by purple and orange squares. Numbers indicate the states occupied at the end of each corresponding pulse in the sequence. (c) Bloch sphere illustration showing the principle of PNS sidebands, with a π rotation applied to the target transition and cycling of the second transition. (d) Rabi oscillations for the $|f1\rangle \leftrightarrow |g2\rangle$ and $|f3\rangle \leftrightarrow |g4\rangle$ transitions, showing the timing for a π and 2π rotation, respectively, corresponding to pulse 18 of the encoding sequence. (e) Reconstructed Wigner functions of the cavity states after the encoding operation, demonstrating the preparation of corresponding binomial code states when the transmon is initialized in six cardinal states. (f) Photon-number-resolved spectroscopy of the prepared Binomial code states. (g) Bloch sphere representation of the encoding operation, illustrating the transition from the transmon Bloch sphere to the logical binomial code Bloch sphere. (h) (Top) The fraction of the population not in $|g\rangle$ after the encoding operation for six cardinal transmon states. (Bottom) Wigner tomography fidelities of the binomial code states following the encoding operation (blue markers) compared against the expected fidelities from Lindblad master equation simulations (gray markers).

VII. CONCLUSION

In summary, we have used transmon-mediated charge-driven sideband interactions to demonstrate fast sideband control of a weakly coupled multimode bosonic memory. We achieve transmon-cavity SWAP gates in 10 modes in 150–350 ns, which is 15–30 times faster than the bare dispersive coupling. Using this, we implement a tunable Jaynes–Cummings interaction within the transmon $\{|g\rangle, |f\rangle\}$ subspace and introduce two strategies to address challenges associated with implementing unitary gates: a shelving technique that prevents unwanted transitions by temporarily storing populations in the sideband-transparent $|e\rangle$ manifold of the transmon, and exploiting the dispersive shift to synchronize sideband transition to implement photon-number-selective sideband gates. Using this, we demonstrated

few-microsecond gates that encode a qubit into vacuum-Fock superposition in arbitrary modes and NOON states in arbitrary pairs of modes. Our encoding protocol is straightforwardly generalizable to arbitrary multimode Fock encodings. In addition, we have also prepared binomial code states across ten cavity modes. Finally, we implemented a new binomial encoding gate with average fidelities of 96.3% for the encoded binomial states after post-selection. This gate is executed in approximately 4 μ s, comparable with previous demonstrations, but with a bare dispersive coupling that is an order of magnitude weaker.

The shelving techniques introduced in this work can be extended to implement a larger space of gates by using higher transmon states [46, 47]. For example, access to the $|h\rangle$ state allows for autonomous error correction following a controlled-parity gate for the binomial code.

The techniques presented in this work are highly relevant for implementing fast sideband-based gate operations in quantum processors and memories comprised of state-of-the-art ultra-low-loss niobium cavities without spoiling cavity coherence [48]. While building a multimode random-access memory requires fixing many-body coherent errors with new random-access architectures [49], the large contrast between the speed of sideband gates and the bare dispersive shift that we have demonstrated can potentially allow for their mitigation through frequency-robust sideband and transmon pulses [22] and dynamical decoupling. Our control techniques allow for efficient control of multiple logical qubits in our multimode memory and tests of gate operations, code concatenation, and multimode bosonic codes [50–52]. Our results provide a promising path toward fast bosonic unitary operations in multimode circuit QED systems, addressing key challenges in bosonic quantum error correction and quantum computing.

ACKNOWLEDGEMENTS

We are grateful for insightful discussions with David Schuster, Michael Gershenson, Jens Koch, Tanay Roy, Xinyuan You, Taeyoon Kim, Shyam Shankar, Wolfgang Pfaff, Angela Kou, and Yuan Liu. We acknowledge Cole Woloszyn and Paul Pickard of the Rutgers Physics and Astronomy Machine Shop for the cavity fabrication. We acknowledge David Schuster, Connie Miao, and Sara Sussman for their support in setting up the QICK measurement hardware; Taekwan Yoon and Zhixin Wang for the setup of Zurich instruments hardware; and David Van Zaanten, Alexander Romanenko, and Anna Grassellino for their support in transmon fabrication. We sincerely thank Michael Gershenson for providing access to his laboratory space and appreciate the support of Michael Gershenson, Plamen Kamenov, Ethan Kasaba, Andre Barbosa, and Raymond Niu in setting up the dilution refrigerator and microwave hardware used in these measurements.

This work is supported by the U.S. Department of Energy, Office of Science and National Quantum Information Science Research Centers, Superconducting Quantum Materials and Systems Center (SQMS) under contract number DE-AC02-07CH11359, and by the Army Research Office under Grant Number W911NF-23-1-0096 and W911NF-23-1-0251.

-
- [1] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf, *Nature* **431**, 162 (2004).
 - [2] M. Reagor, W. Pfaff, C. Axline, R. W. Heeres, N. Ofek, K. Sliwa, E. Holland, C. Wang, J. Blumoff, K. Chou, *et al.*, *Physical Review B* **94**, 014506 (2016).
 - [3] R. Naik, N. Leung, S. Chakram, P. Groszkowski, Y. Lu, N. Earnest, D. McKay, J. Koch, and D. Schuster, *Nature Communications* **8**, 1904 (2017).
 - [4] S. Chakram, A. E. Oriani, R. K. Naik, A. V. Dixit, K. He, A. Agrawal, H. Kwon, and D. I. Schuster, *Physical Review Letters* **127**, 107701 (2021).
 - [5] R. W. Heeres, P. Reinhold, N. Ofek, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, *Nature Communications* **8**, 94 (2017).
 - [6] A. Eickbusch, V. Sivak, A. Z. Ding, S. S. Elder, S. R. Jha, J. Venkatraman, B. Royer, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, *Nature Physics* **18**, 1464 (2022).
 - [7] C. Wang, Y. Y. Gao, P. Reinhold, R. W. Heeres, N. Ofek, K. Chou, C. Axline, M. Reagor, J. Blumoff, K. Sliwa, *et al.*, *Science* **352**, 1087 (2016).
 - [8] S. Chakram, K. He, A. V. Dixit, A. E. Oriani, R. K. Naik, N. Leung, H. Kwon, W.-L. Ma, L. Jiang, and D. I. Schuster, *Nature Physics* **18**, 879 (2022).
 - [9] S. Rosenblum, Y. Y. Gao, P. Reinhold, C. Wang, C. J. Axline, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, *et al.*, *Nature communications* **9**, 652 (2018).
 - [10] Y. Y. Gao, B. J. Lester, K. S. Chou, L. Frunzio, M. H. Devoret, L. Jiang, S. Girvin, and R. J. Schoelkopf, *Nature* **566**, 509 (2019).
 - [11] B. J. Chapman, S. J. de Graaf, S. H. Xue, Y. Zhang, J. Teoh, J. C. Curtis, T. Tsunoda, A. Eickbusch, A. P. Read, A. Koottandavida, *et al.*, *PRX Quantum* **4**, 020355 (2023).
 - [12] Y. Lu, A. Maiti, J. W. Garmon, S. Ganjam, Y. Zhang, J. Claes, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *nature communications* **14**, 5767 (2023).
 - [13] N. Ofek, A. Petrenko, R. Heeres, P. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. Girvin, L. Jiang, *et al.*, *Nature* **536**, 441 (2016).
 - [14] L. Hu, Y. Ma, W. Cai, X. Mu, Y. Xu, W. Wang, Y. Wu, H. Wang, Y. Song, C.-L. Zou, *et al.*, *Nature Physics* **15**, 503 (2019).
 - [15] P. Campagne-Ibarcq, A. Eickbusch, S. Touzard, E. Zalys-Geller, N. E. Frattini, V. V. Sivak, P. Reinhold, S. Puri, S. Shankar, R. J. Schoelkopf, *et al.*, *Nature* **584**, 368 (2020).
 - [16] Z. Ni, S. Li, X. Deng, Y. Cai, L. Zhang, W. Wang, Z.-B. Yang, H. Yu, F. Yan, S. Liu, *et al.*, *Nature* **616**, 56 (2023).
 - [17] V. V. Sivak, A. Eickbusch, B. Royer, S. Singh, I. Tsioutsios, S. Ganjam, A. Miano, B. Brock, A. Ding, L. Frunzio, *et al.*, *Nature* **616**, 50 (2023).
 - [18] A. Romanenko, R. Pilipenko, S. Zorzettil, D. Frolov, M. Awida, S. Belomestnykh, S. Posen, and A. Grassellino, *Physical Review Applied* **13**, 034032 (2020).
 - [19] O. Mihil, B. Guttel, U. Goldblatt, S. Hazanov, L. M. Joshi, D. Chausovsky, N. Kahn, E. Çiftiyerek, F. Lafont, and S. Rosenblum, *PRX Quantum* **4**, 030336 (2023).

- [20] A. E. Oriani, F. Zhao, T. Roy, A. Anferov, K. He, A. Agrawal, R. Banerjee, S. Chakram, and D. I. Schuster, arXiv preprint arXiv:2403.00286 <https://doi.org/10.48550/arXiv.2403.00286> (2024).
- [21] I. Pietikäinen, O. Černotík, A. Eickbusch, A. Maiti, J. W. Garmon, R. Filip, and S. M. Girvin, PRX Quantum **5**, 040307 (2024).
- [22] X. You, Y. Lu, T. Kim, D. M. Kurkcuolu, S. Zhu, D. Van Zanten, T. Roy, Y. Lu, S. Chakram, A. Grassellino, *et al.*, Physical Review Applied **22**, 044072 (2024).
- [23] R. W. Heeres, B. Vlastakis, E. Holland, S. Krastanov, V. V. Albert, L. Frunzio, L. Jiang, and R. J. Schoelkopf, Physical review letters **115**, 137002 (2015).
- [24] L. Bretheau, P. Campagne-Ibarcq, E. Flurin, F. Mallet, and B. Huard, Science **348**, 776 (2015).
- [25] A. A. Diringer, E. Blumenthal, A. Grinberg, L. Jiang, and S. Hacoen-Gourgy, Phys. Rev. X **14**, 011055 (2024).
- [26] A. Wallraff, D. Schuster, A. Blais, J. Gambetta, J. Schreier, L. Frunzio, M. Devoret, . f. S. Girvin, and R. Schoelkopf, Physical Review Letters **99**, 050501 (2007).
- [27] W. Pfaff, C. J. Axline, L. D. Burkhardt, U. Vool, P. Reinhold, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, Nature Physics **13**, 882 (2017).
- [28] Y. Y. Gao, B. J. Lester, Y. Zhang, C. Wang, S. Rosenblum, L. Frunzio, L. Jiang, S. Girvin, and R. J. Schoelkopf, Physical Review X **8**, 021073 (2018).
- [29] M. Mollenhauer, A. Irfan, X. Cao, S. Mandal, and W. Pfaff, arXiv preprint arXiv:2407.16743 <https://doi.org/10.48550/arXiv.2407.16743> (2024).
- [30] M. Pechal, L. Huthmacher, C. Eichler, S. Zeytinoğlu, A. Abdumalikov Jr, S. Berger, A. Wallraff, and S. Filipp, Physical Review X **4**, 041010 (2014).
- [31] C. K. Law and J. H. Eberly, Phys. Rev. Lett. **76**, 1055 (1996).
- [32] A. Ben-Kish, B. DeMarco, V. Meyer, M. Rowe, J. Britton, W. M. Itano, B. Jelenković, C. Langer, D. Leibfried, T. Rosenband, *et al.*, Physical review letters **90**, 037902 (2003).
- [33] M. Hofheinz, H. Wang, M. Ansmann, R. C. Bialczak, E. Lucero, M. Neeley, A. O’connell, D. Sank, J. Wenner, J. M. Martinis, *et al.*, Nature **459**, 546 (2009).
- [34] Y. Liu, J. Sinanan-Singh, M. T. Kearney, G. Mintzer, and I. L. Chuang, Physical Review A **104**, 032605 (2021).
- [35] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of Magnetic Resonance **172**, 296 (2005).
- [36] T. Fösel, S. Krastanov, F. Marquardt, and L. Jiang, arXiv preprint arXiv:2004.14256 <https://doi.org/10.48550/arXiv.2004.14256> (2020).
- [37] M. Kudra, M. Kervinen, I. Strandberg, S. Ahmed, M. Scigliuzzo, A. Osman, D. P. Lozano, M. O. Tholén, R. Borgani, D. B. Haviland, *et al.*, PRX Quantum **3**, 030301 (2022).
- [38] M. H. Michael, M. Silveri, R. Brierley, V. V. Albert, J. Salmilehto, L. Jiang, and S. M. Girvin, Physical Review X **6**, 031006 (2016).
- [39] S. Zeytinoğlu, M. Pechal, S. Berger, A. A. Abdumalikov, A. Wallraff, and S. Filipp, Phys. Rev. A **91**, 043846 (2015).
- [40] R. Shillito, A. Petrescu, J. Cohen, J. Beall, M. Hauru, M. Ganahl, A. G. Lewis, G. Vidal, and A. Blais, Phys. Rev. Appl. **18**, 034031 (2022).
- [41] X. Jin, A. Kamal, A. Sears, T. Gudmundsen, D. Hover, J. Miloshi, R. Slattey, F. Yan, J. Yoder, T. Orlando, *et al.*, Physical Review Letters **114**, 240501 (2015).
- [42] D. T. McClure, H. Paik, L. S. Bishop, M. Steffen, J. M. Chow, and J. M. Gambetta, Physical Review Applied **5**, 011001 (2016).
- [43] K. He, M. Yuan, Y. Wong, S. Chakram, A. Seif, L. Jiang, and D. I. Schuster, Nature communications **15**, 4138 (2024).
- [44] T. Tsunoda, J. D. Teoh, W. D. Kalfus, S. J. de Graaf, B. J. Chapman, J. C. Curtis, N. Thakur, S. M. Girvin, and R. J. Schoelkopf, PRX Quantum **4**, 020354 (2023).
- [45] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Physical Review A—Atomic, Molecular, and Optical Physics **76**, 042319 (2007).
- [46] Z. Wang, R. W. Parker, E. Champion, and M. S. Blok, arXiv preprint arXiv:2407.17407 <https://doi.org/10.48550/arXiv.2407.17407> (2024).
- [47] S. Rosenblum, P. Reinhold, M. Mirrahimi, L. Jiang, L. Frunzio, and R. Schoelkopf, Science **361**, 266 (2018).
- [48] Y. Lu, T. Kim, T. Roy, X. You, O. V. Pronitchev, M. Bal, S. Garattoni, F. Crisa, D. Bafia, R. M. Pilipenko, P. Heidler, A. C. Y. Li, Y. Lu, D. van Zanten, S. Zorzetti, R. Harnik, A. Murthy, S. Zhu, A. Vallières, Z. Huang, J. Koch, A. Grassellino, S. Chakram, and A. Romanenko, Bull. Am. Phys. Soc. (2025).
- [49] R. Banerjee, S. Chakram, A. Oriani, F. Zhao, K. He, A. Agrawal, and D. Schuster, Bull. Am. Phys. Soc. (2022).
- [50] I. L. Chuang, D. W. Leung, and Y. Yamamoto, Physical Review A **56**, 1114 (1997).
- [51] V. V. Albert, S. O. Mundhada, A. Grimm, S. Touzard, M. H. Devoret, and L. Jiang, Quantum Science and Technology **4**, 035007 (2019).
- [52] B. Royer, S. Singh, and S. Girvin, PRX Quantum **3**, 010335 (2022).

Supplementary Information: Fast Sideband Control of a Weakly Coupled Multimode Bosonic Memory

Jordan Huang,^{1,*} Thomas J. DiNapoli,¹ Gavin Rockwood,¹ Ming Yuan,² Prathyankara Narasimhan,¹ Eesh Gupta,³ Mustafa Bal,⁴ Francesco Crisa,⁴ Sabrina Garattoni,⁴ Yao Lu,⁴ Liang Jiang,² and Srivatsan Chakram^{1,†}

¹*Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA*

²*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*

³*Department of Applied Physics, Stanford University, Stanford, California 94305, USA*

⁴*Superconducting Quantum Materials and Systems Division,*

Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA

(Dated: March 14, 2025)

CONTENTS

S1. Multimode processor	1	A. Photon-number-selective sideband calibration	16
A. Device design and fabrication	1	B. Gate characterization and error budget	18
S2. Quantum, microwave, and cryogenic hardware	2	S10. NOON state encoding gate	19
S3. Multimode system Hamiltonian and Lindbladian	2	A. Multimode Wigner tomography	19
A. Multimode Jaynes-Cummings model with a transmon	2	B. NOON state coherence using encode and decode	21
B. Charge-driven sideband interactions	2	S11. Data normalization	21
C. Calibrating the dispersive and Kerr shifts	4	References	21
D. Long-lived mode excited by readout pulse	4		
E. Cavity mode coherence times	4		
F. Cavity mode thermal occupations	6		
S4. Multimode cavity control with sidebands	6		
A. Universal encoding of multimode Fock states	7		
B. Comparison with Law-Eberly state preparation	7		
S5. Transmon and cavity reset with sequential sidebands	8		
S6. Calibrating and optimizing sideband pulses	8		
A. Pulse train calibration	8		
S7. Floquet analysis of sideband interactions	10		
A. Overview of Floquet theory	10		
B. Unitary evolution of drives with smooth ramping	11		
S8. Wigner tomography	12		
A. Error mitigation of parity measurement	12		
B. Wigner tomography with post-selection	13		
C. Fast reset of the readout cavity	14		
D. Wigner tomography error analysis	16		
S9. Binomial code encoding gate	16		

S1. MULTIMODE PROCESSOR

A. Device design and fabrication

The device consists of a multimode storage cavity and a readout cavity, bridged by a transmon. Both are seamless rectangular waveguide cavities fabricated using the flute method [1], created by drilling offset, overlapping, evanescent holes into a monolithic block of 5N5 aluminum. Microwave signals are applied through four SMA ports: a direct storage cavity drive, a direct transmon drive, and two readout cavity drives (input and output). Measurement information is extracted through the single output line connected to the readout cavity. The transmon is positioned to achieve relatively weak coupling to the storage modes, resulting in a dispersive shift of approximately 100-200 kHz.

After machining, the cavity surfaces undergo an etching process similar to that described in [2]. The cavity is submerged in Transene Aluminum Etchant Type A, heated at 30 °C to 50 °C, with constant agitation provided by a magnetic PTFE stirrer at 800 RPM. The process is carried out in two intervals: the first lasting two hours and the second lasting one hour. Between intervals, the used etchant is replaced to prevent saturation, and the cavity is immersed in a bath of deionized (DI) water. Following etching, the cavity is rinsed thoroughly in DI water and dried with nitrogen gas. This procedure removes approximately 45 μm of material, as verified using a test block of 5N5 aluminum treated concurrently.

* jah499@scarletmail.rutgers.edu

† schakram@physics.rutgers.edu

The transmon was fabricated by the Superconducting Quantum Materials and Systems (SQMS) center at Fermi National Laboratory at the Pritzker Nanofabrication Facility (PNF) at the University of Chicago, using the recipe outlined in [3]. The sputtered base layer consists of a Niobium underlayer capped with Tantalum. The Al-AlO_x-Al Josephson junctions were fabricated by double-angle shadow evaporation using the Dolan bridge method.

S2. QUANTUM, MICROWAVE, AND CRYOGENIC HARDWARE

For the bulk of the measurements presented in the paper, the control and readout pulses were digitally synthesized using a Xilinx ZCU216 RFSoc FPGA running the Quantum Instrumentation Control Kit (QICK) [4] firmware and software. For the readout pulse, a signal from the RFSoc ZCU216 is mixed with an LO signal from a microwave function generator (Keysight N5183B). Figure S1 displays a wiring diagram of the experimental setup. Initial characterization of the cavity device was performed using the Zurich Instruments SHFQC Qubit Controller. The required phase stability between qubit, sideband, and cavity drive tones was achieved by taking the frequencies of the local oscillators associated with each drive to be integer multiples of the clock frequency.

The cavity is bolted to a copper plate that is attached to the base stage (8 mK) of a Bluefors LD-400 dilution refrigerator. The cavity is enclosed inside a mu-metal can that is wrapped outside with copper tape. Surrounding the cavity inside the mu-metal can is eccosorb foam (Fig. S1).

S3. MULTIMODE SYSTEM HAMILTONIAN AND LINDBLADIAN

A. Multimode Jaynes-Cummings model with a transmon

The Hamiltonian of our system is described by $\mathcal{H} = \mathcal{H}_{\text{sys}} + \mathcal{H}_d$ where

$$\mathcal{H}_{\text{sys}} = \underbrace{\tilde{\omega}_q c^\dagger c - E_J \left(\cos(\varphi) + \frac{\varphi^2}{2} \right)}_{\text{Transmon}} + \sum_i \underbrace{\left(\tilde{\omega}_{c,i} b_i^\dagger b_i + g_i (c - c^\dagger)(b_i - b_i^\dagger) \right)}_{\text{Coupling}}. \quad (\text{S1})$$

Here, we have multiple cavity modes (with lowering operator b_i , where i is the mode index) electric-dipole coupled to the same transmon and have written the transmon in the linearized oscillator basis, where c is its lowering operator. $\tilde{\omega}_q$ and $\tilde{\omega}_{c,i}$ are the bare frequencies of the lin-

ear part of the transmon and cavity modes, respectively. $\varphi = \varphi_{\text{zpt}}(c + c^\dagger)$ is the superconducting phase across the transmon's Josephson junction. Additionally, we have direct transmon and cavity drives given by

$$\mathcal{H}_d = -2i\epsilon_q \cos(\omega_d t + \phi_d)(c - c^\dagger) - 2i \sum_i \epsilon_{c,i} \cos(\omega_{d,c,i} t + \phi_{d,c,i})(b_i - b_i^\dagger). \quad (\text{S2})$$

B. Charge-driven sideband interactions

We derive the sideband interaction using a four-wave mixing picture. We begin by considering the transmon coupled to one of the cavity modes, along with a direct transmon drive to activate the interaction. In order to prepare our Hamiltonian in an interaction frame, we first diagonalize the coupling between the linear parts of the system. Henceforth, ω_q and ω_r will refer to the dressed frequencies of the linear part of the transmon and cavity, respectively. Next, we go into the displaced frame of the drive with $U = \exp(\alpha^*(t)c - \alpha(t)c^\dagger)$ where $\alpha(t) = e^{-i\omega_d t} \epsilon_q / (\omega_d - \omega_q) - e^{i\omega_d t} \epsilon_q / (\omega_d + \omega_q)$. Finally, we go into the rotating frame of the transmon and the cavity with $U = \exp(i\omega_q c^\dagger c t) \exp(i\omega_c b^\dagger b t)$. The phase operator transforms as

$$\varphi = \varphi_{\text{zpt}} \left(\beta_q e^{-i\omega_q t} c + \beta_c e^{-i\omega_c t} b + \beta_q i e^{-i\phi_d} e^{-i\omega_d t} \xi + \text{h.c.} \right) \quad (\text{S3})$$

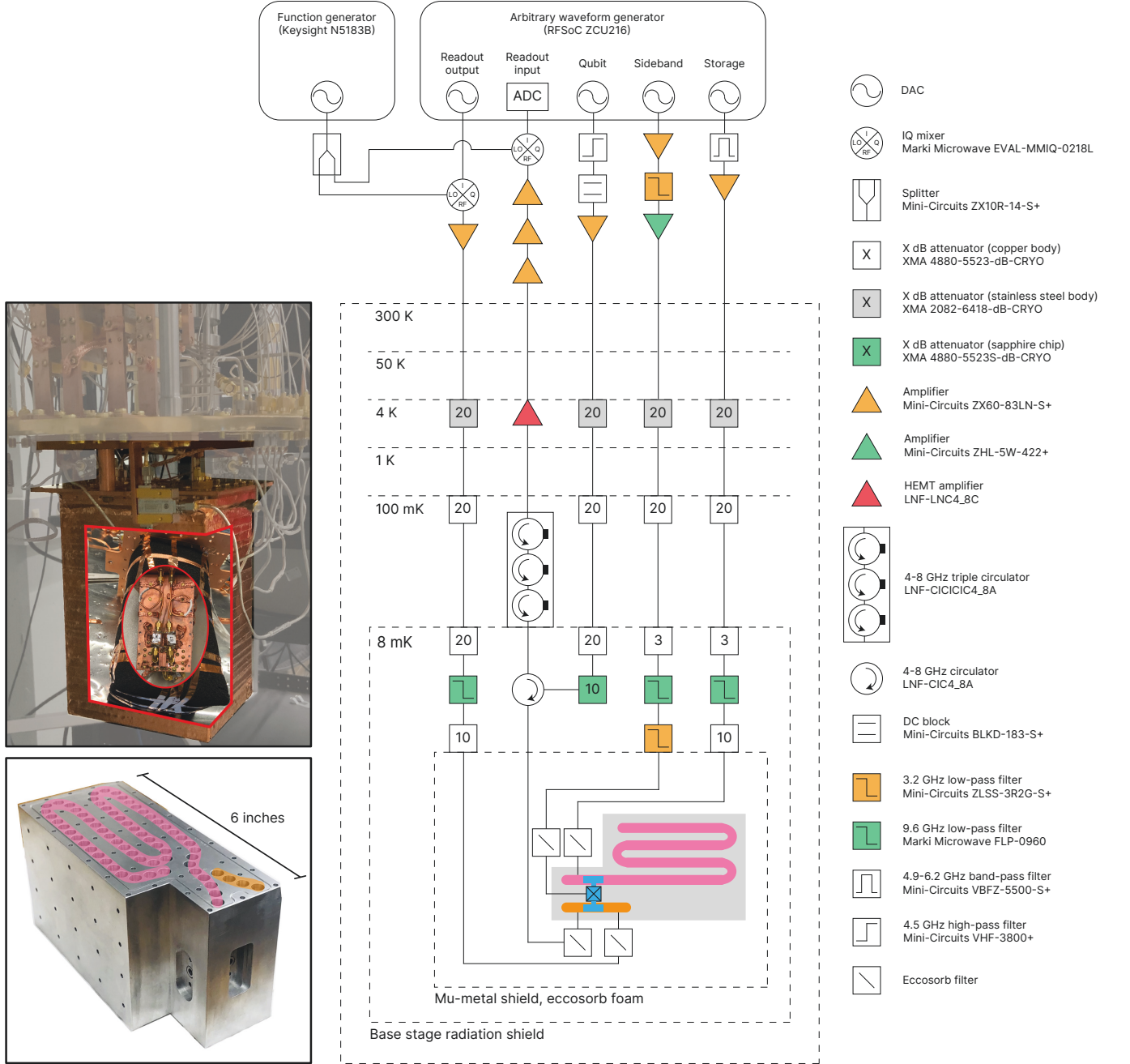
where β_q and β_c are the participations of the transmon and the cavity in the phase across the junction, respectively. In the dispersive regime where $g \ll |\tilde{\omega}_q - \tilde{\omega}_r|$, we have $\beta_q \approx 1$ and $\beta_c \approx -g/\tilde{\Delta}$ where $\tilde{\Delta} = \tilde{\omega}_q - \tilde{\omega}_c$. $\varphi_{\text{zpt}} = (2E_C/E_J)^{1/4}$ is the zero-point fluctuation of the phase and $\xi = 2\omega_d \epsilon_q / (\omega_d^2 - \omega_q^2)$ is the effective transmon displacement. When $|\varphi| \ll 1$, we may expand the nonlinearity of our system as

$$\mathcal{H}_{\text{nl}} = -E_J \left(\frac{\varphi^4}{4!} - \frac{\varphi^6}{6!} + \dots \right). \quad (\text{S4})$$

The nonlinearity in conjunction with the transmon drive can activate parametric interactions which can be made resonant through a judicious choice of the drive frequency ω_d . The fastest interactions involving the cavity mode use the fourth-order nonlinearity φ^4 to convert two transmon photons into a photon in any of the cavity modes. This gives rise to an interaction of the form $\tilde{g}_{\text{sb}}(ie^{-i\phi_d} c^\dagger c^\dagger b - ie^{i\phi_d} c c b^\dagger)$, where

$$\tilde{g}_{\text{sb}} = \frac{gK\xi}{\Delta}. \quad (\text{S5})$$

Here, $K \approx -E_C$ is the transmon anharmonicity and $\Delta = \omega_q - \omega_c$, where we have made the approximation $\Delta \approx \tilde{\Delta}$. This interaction can connect any two transmon levels



that differ by two-photons, such as $|f, n\rangle$ and $|g, n+1\rangle$ or $|h, n\rangle$ and $|e, n+1\rangle$. The $|f, n\rangle$ - $|g, n+1\rangle$ sideband interaction is made resonant by choosing $\omega_d = 2\omega_q + K - \omega_c$, the difference frequency between the transmon's $|f\rangle$ state and the cavity. A SWAP operation in the $|g\rangle$ - $|f\rangle$ mani-

fold is given by $g_{sb}(ie^{-i\phi_d}|f\rangle\langle g|b - ie^{i\phi_d}|g\rangle\langle f|b^\dagger)$, where

$$g_{sb} = \frac{\sqrt{2}gK\xi}{\Delta} = \frac{\sqrt{2}(\omega_q + K + \Delta)}{2\omega_q + K + \Delta} \frac{\epsilon}{g}\chi. \quad (\text{S6})$$

Here, $\chi = 2g^2K/(\Delta(\Delta + K))$ is the dispersive shift of the $|e\rangle$ state. In this form, we see that ϵ can be increased such that this interaction strength is larger than χ . In-

cluding the contribution from higher-order nonlinearities (beyond φ^4), while making an RWA to drop counter-rotating interaction terms, modifies the sideband rate to

$$\tilde{g}_{sb} = \frac{gK\xi}{\Delta} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(\varphi_{zpt}\xi)^{2n-2}}{(n-1)!(n-1)!}. \quad (S7)$$

We see that the sideband rate becomes nonlinear in drive strength ϵ . Because of the opposite signs of successive nonlinear terms, this formula predicts a downturn in the rate at sufficiently large ϵ .

C. Calibrating the dispersive and Kerr shifts

In the dispersive regime where $g \ll |\omega_q - \omega_{c,i}|$, the system can be described by a simplified Hamiltonian

$$\mathcal{H} = \omega_q c^\dagger c + \frac{K}{2} c^\dagger c^\dagger c c + \sum_i \left(\omega_{c,i} b_i^\dagger b_i + \frac{K_i}{2} b_i^\dagger b_i^\dagger b_i b_i + \chi_i c^\dagger c b_i^\dagger b_i \right), \quad (S8)$$

which can be derived from Eq. (S1) using second-order perturbation theory. We measure ω_q , the transmon's anharmonicity K using standard $|g\rangle$ - $|e\rangle$ and $|e\rangle$ - $|f\rangle$ two-tone spectroscopy measurements that are refined with transmon Ramsey measurements. We also measure the transmon life times using standard T_1 measurements, and extract decoherence times T_2 and T_2^* using Ramsey experiments with and without a spin-echo pulse. We measure the transmon's thermal population ($|e\rangle$ population) by measuring the $|f\rangle$ -state population after driving the $|e\rangle$ - $|f\rangle$ transition, with and without an initial π_{ge} pulse.

We measure the cavity mode frequencies $\omega_{c,i}$ in the single-photon regime using the transmon via their cross-Kerr interaction (the dispersive shift χ_i). This interaction changes the transmon frequency based on the photon number in the cavity. We perform cavity spectroscopy using a cavity drive pulse sent through the direct storage drive port while sweeping its frequency. The cavity state is subsequently probed by a narrow-band photon number-resolved transmon pulse driven on the zero-photon peak, which affects a π pulse conditioned on the cavity being in vacuum, thereby measuring the zero-photon population. When on resonance, the cavity pulse excites the mode and reduces the zero-photon population, leading to a dip in the spectrum. The cavity mode frequency is further refined by a cavity-Ramsey experiment [5], where we excite the cavity with two displacement pulse, while varying the time τ between them and advancing the phase of the second displacement pulse by $\omega_R \tau$. This is followed by a measurement of the zero-photon population through a resolved transmon π -pulse.

We measure the dispersive shift χ_i of the cavity modes by performing a transmon $|g\rangle$ - $|e\rangle$ Ramsey experiment after adding a photon in the target mode using charge-sidebands, using the sequence $\pi_{ge} \rightarrow \pi_{ef} \rightarrow \pi_{f0g1}$. We

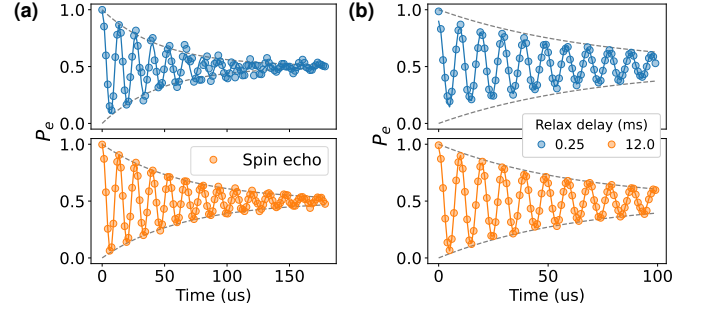


FIG. S2. **Spurious mode excited by readout pulse.** (a) Ramsey experiment with (top) and without (bottom) a spin echo. (b) Ramsey experiment with a 0.25 ms delay time (top) and a 12 ms delay time (bottom) between experiments for the same readout pulse settings.

also measure $\chi_{f,i}$ using the same protocol, with the Ramsey experiment between $|g\rangle$ and $|f\rangle$ transmon levels. With calibrated transmon pulses, the difference in the oscillation frequency from the Ramsey frequency programmed in software gives a precise measurement of χ_i limited only by linewidth of the transmon.

D. Long-lived mode excited by readout pulse

We observe spurious beating and reduced contrast in transmon Ramsey measurements, which depend on the magnitude and duration of the readout pulse at 8.017 GHz, as well as the experiment duty cycle. These effects are eliminated by a spin-echo sequence, as shown in Fig.S2(a), and are also mitigated in a standard Ramsey experiment by reducing the duty cycle, as shown in Fig.S2(b). This suggests that the readout tone excites a long-lived mode in the system that is dispersively coupled to the transmon. Qubit spectroscopy indicates that this is a long-lived mode rather than a two-level system (TLS) [6], though its dispersive shift does not correspond to the storage cavity mode closest to the readout mode at 7.967 GHz. We estimate the lifetime of the spurious mode to be on the order of milliseconds. To mitigate this issue, we minimize the readout drive power while maintaining readout fidelity and increase the time delay between experiments.

E. Cavity mode coherence times

We measure the single-photon lifetime (T_1) of a cavity mode by first adding a photon to the mode via the sequence $\pi_{ge} \rightarrow \pi_{ef} \rightarrow \pi_{f0g1}$. After waiting for a variable delay, we transfer the remaining one-photon population back to the transmon using a π_{f0g1} pulse and measure the transmon's $|f\rangle$ population as a function of wait time to extract the cavity's T_1 . To measure the single-photon dephasing time, we perform a cavity Ramsey experiment

Parameter	Hamiltonian Term ($\hbar = 1$)	Value
Transmon frequency ν_{ge}	$\pi\nu_{ge}(e\rangle\langle e - g\rangle\langle g)$	4.606 GHz
Transmon frequency ν_{ef}	$\pi\nu_{ef}(f\rangle\langle f - e\rangle\langle e)$	4.492 GHz
Transmon frequency ν_{fh}	$\pi\nu_{fh}(h\rangle\langle h - f\rangle\langle f)$	4.371 GHz
Transmon $ e\rangle \rightarrow g\rangle$ decay time T_1		55.83 μ s
Transmon $ f\rangle \rightarrow e\rangle$ decay time T_1		28.85 μ s
Transmon $ g\rangle - e\rangle$ dephasing time T_2^*, T_2		47.20, 65.65 μ s
Transmon $ g\rangle - f\rangle$ dephasing time T_2^*, T_2		36.58, 41.06 μ s
Transmon thermal population $\bar{n}$		0.47 %
		0.32 % (with transmon reset)
Storage mode frequency ν_c	$2\pi\nu_c b^\dagger b$	{5.750, 5.994, 6.228, 6.479, 6.720, 6.962, 7.216, 7.461, 7.715, 7.967} GHz
Storage mode dispersive shift χ	$2\pi\chi_e b^\dagger b e\rangle\langle e $ $2\pi\chi_f b^\dagger b f\rangle\langle f $	{-197, -217, -202, -208, -171, -150, -165, -133, -117, -106} kHz {-356, -383, -394, -391, -363, -299, -304, -245, -220, -190} kHz
Storage mode single-photon decay time T_1		{1.187, 1.253, 1.298, 1.175, 1.175, 0.899, 0.656, 0.851, 0.936, 0.989} ms
Storage mode single-photon dephasing time T_2^*		{1.932, 2.044, 2.0243, 1.926, 1.897, 1.559, 1.1613, 1.555, 1.662, 1.811} ms
Storage mode thermal population $\bar{n}$		{0.82, 0.85, 0.29, 0.54, 0.38, 1.30, 1.93} % {0.26, 0.55, 0.48, 0.58, 0.49, 0.39, 0.59, 0.62, 0.73, 0.70} % (with cavity reset)
Readout mode frequency ν_r	$2\pi\nu_r b_r^\dagger b_r$	8.0174 GHz
Readout mode dispersive shift χ_e	$2\pi\chi_e b_r^\dagger b_r e\rangle\langle e $	-845 kHz
Readout mode single-photon decay time T_1		0.61 μ s
Readout mode thermal population $\bar{n}$		0.54 %

TABLE S1. Measured system parameters.

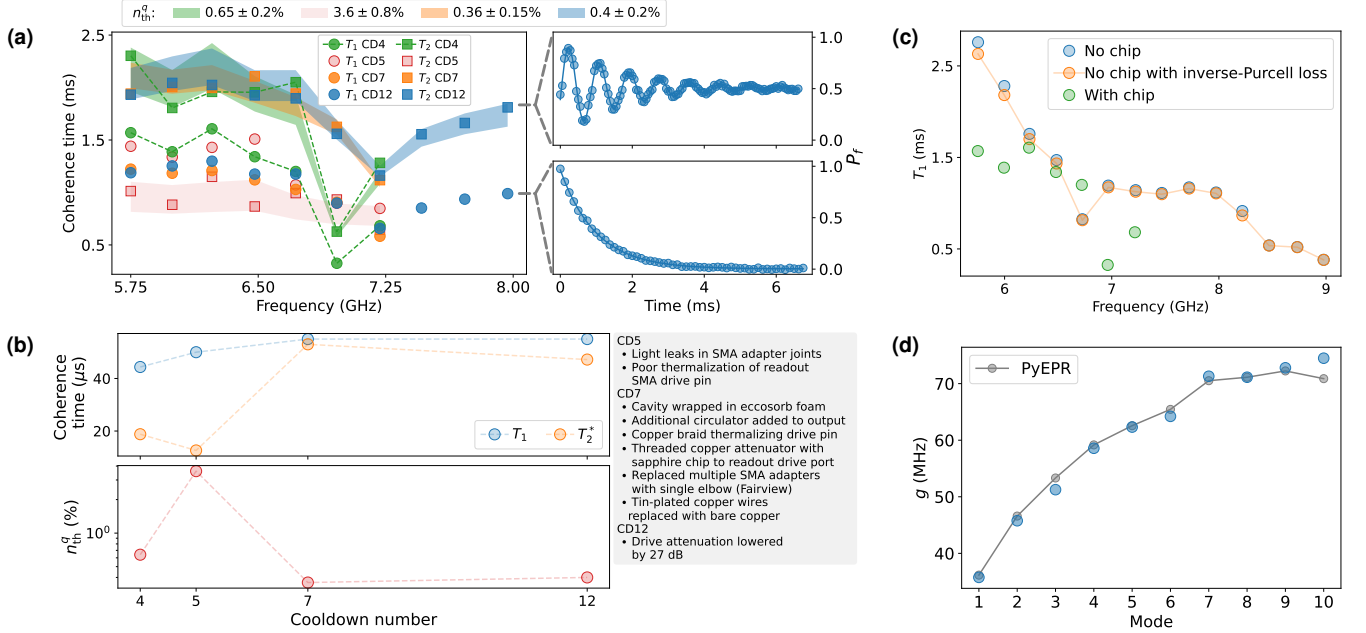


FIG. S3. **Transmon-cavity characterization.** (a) Coherence times of the cavity modes across different cooldowns. The shaded regions indicate the inferred qubit thermal population that would give rise to the measured cavity T_2 . The right inset shows the cavity T_2 Ramsey (top) and the T_1 experiment (bottom) for a given mode. (b) Transmon coherence times (top) and inferred transmon thermal population from cavity coherence (bottom) across different cooldowns. The grey box highlights important changes made between cooldowns. The major change that lowered the transmon thermal population was wrapping the cavity in eccosorb foam. The transmon T_2^* was increased by a factor of 4 through changes made to the output lines. (c) Cavity T_1 with and without the transmon and sapphire substrate. (d) The blue circles denote the bare coupling (g) derived from experimentally measured χ 's for the modes. The grey circles are predictions from pyEPR [7] simulations.

by preparing the superposition state $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ in the target mode. After allowing the state to idle in the cavity for a variable delay, we map the cavity states $\{|0\rangle, |1\rangle\}$ onto the transmon states $\{|g\rangle, |e\rangle\}$ using the pulse sequence $\pi_{f0g1} \rightarrow \pi_{ef}$, where we sweep the phase of the π_{ef} pulse as a function of the delay time. Finally, we apply a transmon $\frac{\pi}{2}_{ge}$ pulse to convert the phase information into population, thereby performing a Ramsey experiment on the $\{|0\rangle, |1\rangle\}$ subspace.

The resulting T_1, T_2 times for the cavity modes are shown in Fig. 1 of the main text and in Table S1. We note that the cavity mode T_1 times were found to be lower than for the bare cavity without the transmon chip (Fig. S3(c)). This additional decay was beyond that expected from the inverse-Purcell loss from transmon decay, estimated from the measured transmon-mode couplings. The extra loss is attributed to the sapphire substrate on which the transmon is patterned. We also note that the mode T_1 's degraded from the initial cooldowns (Fig. S3(a)) despite no change made to the packaging and filtering. No change was observed in the transmon lifetimes across cooldowns.

A summary of the changes made between cooldowns and their effect on transmon coherences times and thermal populations is shown in Fig. S3. The thermal population of the transmon was initially high, which we attributed to inadequate shielding from high-frequency infrared radiation above the superconducting gap, known to create quasiparticles. This was mitigated by wrapping the cavity in eccosorb foam, which reduced the thermal population from 3% $\rightarrow$ 0.5%. The readout mode's thermal population was also reduced from 1.32% to 0.54% during the same cooldown. While there were several changes, we particularly note the additional circulator added to the readout output line, which likely reduced thermal photons leaking in through it.

F. Cavity mode thermal occupations

We measure the cavity thermal population by applying an $|f0\rangle\text{--}|g1\rangle$ -pulse of varying duration on the target cavity mode with the transmon initialized in the ground state. Given the negligible thermal population of the transmon $|f\rangle$ state, the small residual oscillation that is observed is due to the thermal occupation of the one-photon state of the mode. We measure this population precisely by comparing the oscillation with that observed with the transmon prepared in $|f\rangle$. The resulting thermal occupations are shown in Fig. S6(d) ranging from 0.2–2% across the modes. We note that the modes that have higher thermal occupation also have lower quality factor from coupling to the direct storage drive line. In order to mitigate this state preparation error, we perform a sequential sideband reset protocol prior to the start of every experiment, where we transfer the cavity population to the $|f\rangle$ and cool the state by driving the $|f0_r\rangle\text{--}|g1_r\rangle$ transition with the readout cavity, which has

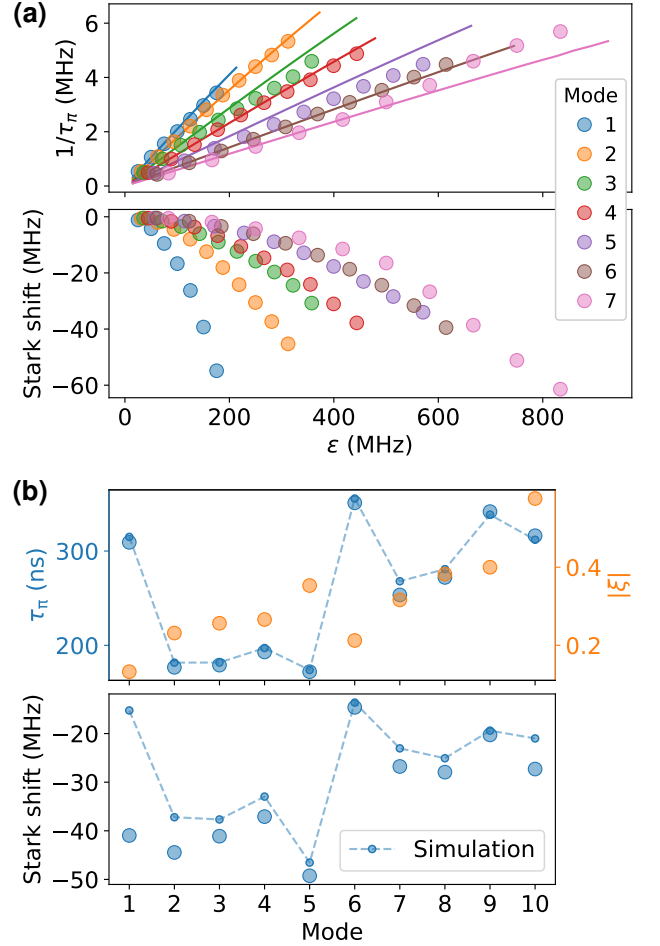


FIG. S4. **Sideband characterization.** (a) (Top panel) Measured sideband rates (circles) as a function of transmon drive strength for the first seven modes of the storage cavity. Solid lines represent simulated rates obtained from Floquet numerics. (Bottom panel) Drive-induced Stark shift of the sideband resonance as a function of drive strength. (b) (Top panel) Measured π -pulse times (blue circles) for the $|f0\rangle\text{--}|g1\rangle$ sideband across modes at a fixed drive strength, corresponding to the one used for binomial state preparation experiments in Fig. 3(f) of the main text. Simulated π -pulse times (blue dots) and corresponding transmon displacement required to achieve the same speed (orange circles, right axis) are also shown. (Bottom panel) Measured drive-induced Stark shift of the sideband resonance (blue circles) at the corresponding $|f0\rangle\text{--}|g1\rangle$ drive strengths. Numerical predictions are shown as blue dots.

a T_1 of 0.61 μs (see Section S5).

S4. MULTIMODE CAVITY CONTROL WITH SIDEBANDS

The sideband interaction allows us to engineer an effective multimode Jaynes-Cummings-like Hamiltonian

where we have control of the coupling strength, given by

$$\mathcal{H} = \sum_i \left(\underbrace{\chi_{e,i} b_i^\dagger b_i |e\rangle \langle e| + \chi_{f,i} b_i^\dagger b_i |f\rangle \langle f|}_{\text{Dispersive shift}} + \underbrace{g_{\text{sb},i}(t)(b_i |f\rangle \langle g| + b_i^\dagger |g\rangle \langle f|)}_{\text{Tunable coupling}} \right) \quad (\text{S9})$$

The tunable Jaynes-Cummings coupling, combined with transmon drives on the $|g\rangle$ - $|e\rangle$, $|e\rangle$ - $|f\rangle$, and direct $|g\rangle$ - $|f\rangle$ two-photon Raman transition, provides a versatile set of interactions that enable the intuitive derivation of analytic pulse sequences for gate operations using a Jaynes-Cummings ladder picture.

A key challenge when relying solely on sideband interactions ($g_{\text{sb}} \gg |\chi|$) is their lack of photon-number selectivity and the incommensurate sideband Rabi rotation rates across different photon-number states. As detailed in the main text, we address this issue through two complementary strategies. First, we implement a shelving technique, where we temporarily store selected states in the transmon's $|e\rangle$ manifold. Since the sideband interaction occurs only in the $|g\rangle$ - $|f\rangle$ manifold and is detuned from the $|e\rangle$ manifold by twice the transmon anharmonicity, this suppresses transitions of the shelved states by a factor of $\sim 10^{-4}$. Second, we leverage the dispersive interaction available in our system to introduce a photon-number-selective sideband pulse (see Section S9 A), which implements selective sideband operations on a given photon number state while leaving another non-shelved state unchanged (up to a phase).

A. Universal encoding of multimode Fock states

We introduce a protocol to implement a unitary gate U that encodes the transmon ground and excited states into two arbitrary multimode Fock states:

$$U (\cos \theta |g\rangle + e^{i\varphi} \sin \theta |e\rangle) \otimes |0 \dots 0\rangle = |g\rangle \otimes (\cos \theta |n_1 \dots n_M\rangle + e^{i\varphi} \sin \theta |p_1 \dots p_M\rangle). \quad (\text{S10})$$

We prove that we can construct this unitary using the following protocol. We first use sideband and qubit rotations to take $|g\rangle$ to a desired multimode Fock state $(|g\rangle \otimes |n_1 \dots n_M\rangle)$ while $|e\rangle$ is left shelved $(|e\rangle \otimes |0 \dots 0\rangle)$. Next, we periodically shelve the state $|g\rangle \otimes |n_1 \dots n_M\rangle$ while preparing the second multimode Fock state one photon at a time using transmon rotations and sideband transitions, up to a state one-photon away from the target $(|e\rangle \otimes |0 \dots 0\rangle \rightarrow |f\rangle \otimes |p_1 - 1 \dots p_M\rangle)$. During the preparation of the second state, periodic shelving can be skipped if a direct $|g\rangle$ - $|f\rangle$ transition is used. Before the final sideband pulse, we unshelve the first state into the $|g\rangle$ -manifold $(|g\rangle \otimes |n_1 \dots n_M\rangle)$. We then bring the second

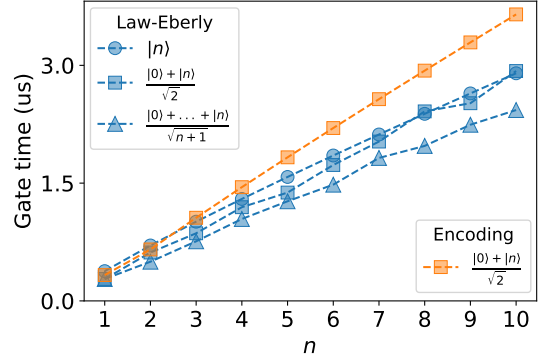


FIG. S5. **Comparison of the scaling with photon number for Law-Eberly state preparation and our encoding protocol.** The blue markers show state preparation times using the Law-Eberly protocol for various cavity states as a function of n . The orange markers show the gate times for encoding vacuum-Fock state superpositions as a function of n using the scheme detailed in Section S4 A.

state to the $|g\rangle$ manifold using a photon-number-selective sideband pulse. This construction is not unique, as we can permute the order in which we climb the multimode Jaynes-Cummings ladder.

With this protocol, the number of sideband and qubit pulses required for each is $N_{\text{tot}} = \sum_{i=1}^M (n_i + p_i)$, in addition to shelving pulses. Crucially, the speed of the photon number-selective pulse is proportional to the detuning of the two encoding states with respect to their dispersive shifts (see Section S9 A). In the special condition where at least one of n_i or p_i are equal 0, a photon number-selective sideband is not required at the last step. The NOON states $(|N0\rangle + |0N\rangle)/\sqrt{2}$ are an example of such a state, for which we implement an encoding gate with just fast sideband transitions (see Section S10).

B. Comparison with Law-Eberly state preparation

The control Hamiltonian in Eq. S9 can be used to prepare arbitrary cavity states in a single mode using the Law-Eberly protocol [8]. This protocol prepares an arbitrary quantum state by finding the sequence of sideband and qubit rotations that removes one photon at a time from the cavity mode to transition from the initial state to $|g0\rangle$. To prepare the target state, we simply apply the reverse of this gate sequence.

In contrast to [8], the Hamiltonian in Eq.S9 also includes photon-number-dependent dispersive shifts in the qubit and cavity-mode frequencies. In the single-mode case, the original Law-Eberly protocol can be extended to incorporate the dispersive Hamiltonian in Eq.S9 by introducing photon-number-dependent shifts in the sideband and qubit pulse frequencies.

With this protocol, the number of sideband and qubit pulses required each to implement the unitary U that

prepares an arbitrary state,

$$U|0\rangle = \sum_{n=0}^N c_n |n\rangle, \quad (\text{S11})$$

is $N_{\text{max}} = N$, the highest number of photons in the state. Additionally, it is faster to prepare states whose populations are more distributed in the photon number basis (Fig. S5). This speedup is because the qubit and sideband pulses that recombine superpositions in the photon number basis do not necessarily need to complete full π rotations.

As shown in Fig. S5, the number of gates required for the encoding protocol scales linearly with the total number of photons in the superposition state. In contrast, the Law-Eberly state preparation protocol scales linearly with only the highest photon number in the target state. Importantly, we emphasize that the encoding protocol is not merely a state preparation method but a full gate operation that encodes transmon states into arbitrary multimode Fock states.

S5. TRANSMON AND CAVITY RESET WITH SEQUENTIAL SIDEBANDS

We increase the experiment's duty cycle by implementing sequential transmon and cavity reset, cooling both the transmon and cavity mode one photon at a time using sideband pulses between the transmon and the lossy readout cavity. Cooling the state $|g, N\rangle$ by one photon is achieved via a $\pi_{f, N-1, g, N}$ pulse, followed by driving the $|f, 0_r, N-1\rangle$ - $|g, 1_r, N-1\rangle$ transition through the readout mode to dissipate the transmon excitation. This process is repeated N times until the cavity reaches the ground state. Similarly, the $|e, N\rangle$ state is reset by first applying a π_{ef} pulse, followed by driving the $|f, 0_r, N\rangle$ - $|g, 1_r, N\rangle$ transition via the readout mode before proceeding with the same protocol used to cool $|g, N\rangle$.

These protocols and their corresponding pulse sequences are illustrated in Fig. S6(a) and (b). A full reset cycle sequentially cools all n -photon states up to N to the ground state for both $|g\rangle$ and $|e\rangle$ transmon states. Multiple reset cycles are applied to mitigate population redistribution caused by incommensurate rotations from fast sideband pulses and to remove population from cavity states where the transmon remains in $|f\rangle$. Spectroscopy and Rabi oscillations of the $|f, 0_r\rangle$ - $|g, 1_r\rangle$ sideband with the readout mode is shown in Fig. S6(c) indicating a readout mode decay time of 0.61 μs . The improvement in the transmon and cavity mode thermal populations with reset are shown in Fig. S6(d). The final thermal populations following reset are comparable to the thermal population of the readout cavity, inferred independently from shot-noise dephasing of the transmon [9].

S6. CALIBRATING AND OPTIMIZING SIDEBAND PULSES

Maximizing the fidelity of strongly driven interactions requires smooth pulses that allow adiabatic evolution of the undriven eigenstates to their corresponding Floquet modes (See Section S7B). As a result, we cannot simply use a rectangular pulse and must implement smooth ramps. Our sideband pulse shape is a flat top, ramped up and down using a bump function, with the pulse envelope given by,

$$\epsilon(t) = \begin{cases} \epsilon_{\text{max}} \exp\left(2 + \frac{2}{\left(\frac{t-\tau}{\tau}\right)^2 - 1}\right), & t \leq \tau \\ \epsilon_{\text{max}}, & \tau < t < T - \tau \\ \epsilon_{\text{max}} \exp\left(2 + \frac{2}{\left(\frac{t-T+\tau}{\tau}\right)^2 - 1}\right), & t \geq T - \tau \end{cases} \quad (\text{S12})$$

where τ is the ramp time and T is the total length of the pulse. This function has the feature of being infinitely differentiable, with the ability to do fine adjustments of the effective pulse time. We note that we observed similar performance with a simpler $\sin^2$ ramp function. The envelope is digitally combined in the FPGA with a carrier at the sideband frequency ω_{sb} .

The sideband interaction is induced by applying charge drives to the transmon, which leads to a drive-amplitude-dependent Stark shift of the transmon energies, which shifts the sideband resonance frequency. We prepare the transmon in $|f\rangle$ and find the resonance frequency through sideband spectroscopy. We sweep ω_{sb} and measure the $|f\rangle$ population, fitting the resulting spectrum to a sinc function to extract the resonance frequency. We subsequently drive the sideband on resonance and measure the $|f\rangle$ population as a function of pulse time to extract the π -time. We perform this calibration iteratively for different ϵ , operating at an ϵ_{max} where there is no drop in sideband Rabi oscillation contrast.

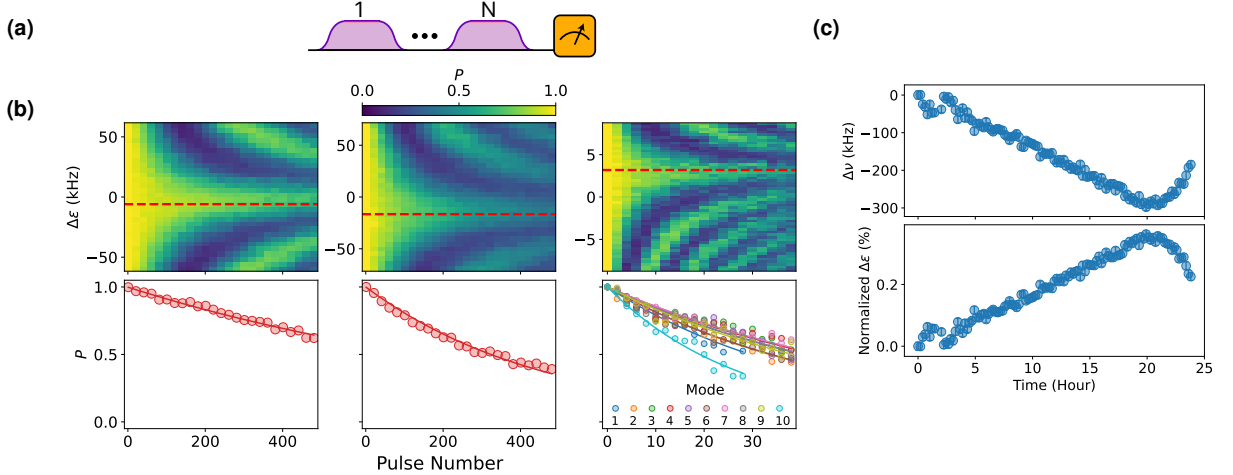
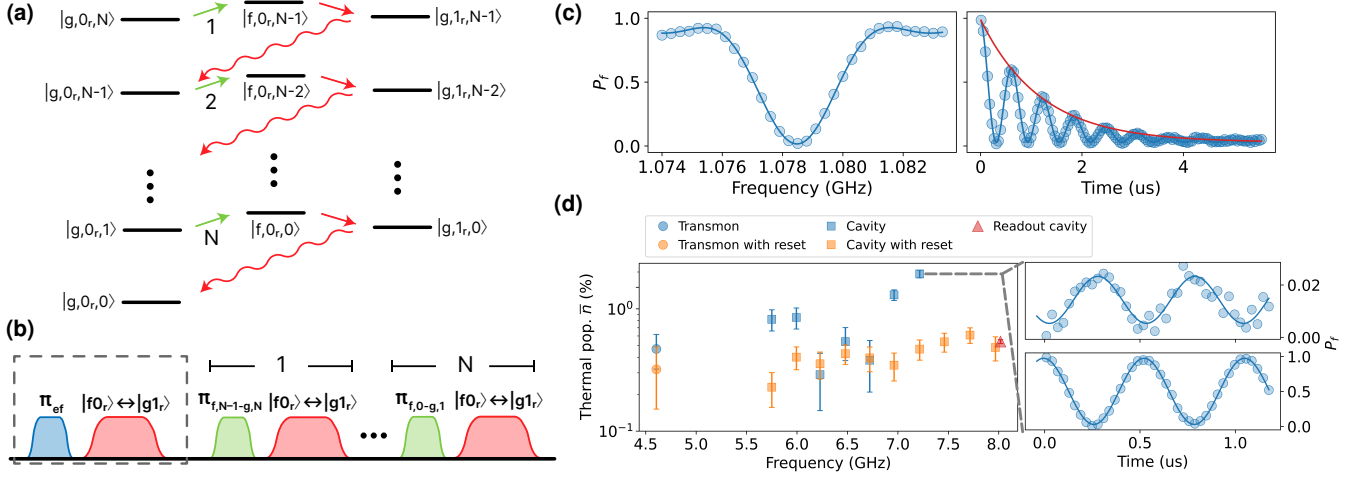
We estimate the sideband interaction fidelity by fitting the Rabi oscillations to

$$P_f = \frac{1}{2} e^{-\kappa t} (1 + e^{-\kappa_{\varphi} t} \cos(2g_{\text{sb}} t)). \quad (\text{S13})$$

Following [10], the fidelity of a π -pulse is given by $F \approx 1 - \frac{\pi}{2} \frac{\kappa + \kappa_{\varphi}/2}{g_{\text{sb}}}$, where κ is the effective decay and κ_{φ} the effective dephasing of the interaction. This inferred fidelity is plotted in Fig. 2(f) of the main text for all modes. Fig. S8 displays the Rabi oscillations used to extract the fidelity for mode three.

A. Pulse train calibration

Optimization of the sideband pulse involves tuning the pulse duration, amplitude (ϵ_{max}), and drive frequency. This is performed using a pulse train to amplify rotation errors, where a sequence of π -pulses is applied, and the $|f\rangle$ state population is measured as a function of the sequence length. We select a target π -pulse duration that



is an integer multiple of the envelope fabric of the Xilinx DAC and perform a fine sweep of the drive amplitude $\epsilon_{\max}$ near the expected π -pulse amplitude while applying sequences of even (or odd) numbers of π -pulses, as shown in Fig. S7(b). Given the 16-bit amplitude resolution, this enables a fine sweep of the pulse area. Similar

control can also be achieved by adjusting the ramp time; however, this also affects the adiabaticity condition for the pulse.

Since the pulse amplitude also influences the optimal detuning due to the Stark shift, determining the true optimal pulse requires tuning both the amplitude and de-

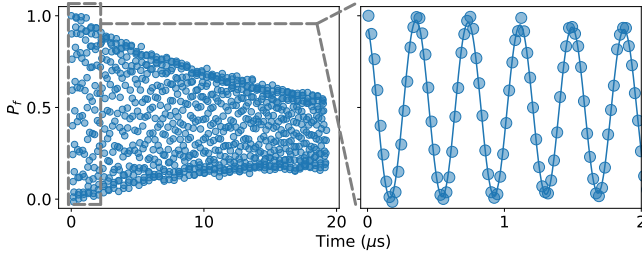


FIG. S8. **Sideband interaction fidelity.** The sideband Rabi oscillations (third mode shown) are fit to Eq. S13 to extract the decay and dephasing times. The evolution for the first 2 μ s are plotted on the right.

tuning. Our current experiments were affected by $\sim 0.1\%$ per hour drifts in the pulse amplitude, which we attribute to ambient temperature fluctuations impacting the gain of the 5 Watt amplifier used on the sideband drive line. The resulting fluctuations in the sideband resonance frequency over time are shown in Fig. S7(c). These drifts made the two-dimensional sweep of both ϵ_{max} and ω_{sb} nonviable. To mitigate this, we instead relied on a one-dimensional sweep of the pulse amplitude near the optimal detuning, which makes the pulse error quadratic in the detuning error. The optimal calibration is determined by selecting ϵ_{max} that minimizes the residual error to an exponential fit.

A comparison of the fidelities extracted from the pulse train calibration and those inferred from Rabi decay are shown in Fig. 2(f) of the main text. The pulse train fidelities are slightly below those inferred from Rabi decay because the pulse train captures leakage effects associated with the pulse profile, such as the ramp. As a result, the Rabi decay fidelities measure the interaction fidelity while the pulse train fidelities measure the pulse fidelity and includes leakage errors. Our pulse profiles can be further optimized to mitigate this.

S7. FLOQUET ANALYSIS OF SIDEBAND INTERACTIONS

In this section, we explain the numerical procedure used to simulate the sideband interaction. In our simulations, we model the drive signal as a classical monochromatic drive on the transmon, as in Eq. S2. To compute the resonance frequency and rate, we first compute the Floquet quasienergy spectrum. The sideband interaction resonance manifests as an avoided crossing in the spectrum, which we fit to extract the frequency and rate. This can be understood in analogy with a driven two-level system, which we detail below. Finally, we use the Floquet formalism to understand the role of the ramp in reducing leakage out of the interaction subspace. This is especially relevant for strongly-driven transmon interactions that lead to significant hybridization between the transmon levels. We derive bounds on the ramp timescale

that are required to achieve an optimal fidelity interaction.

A. Overview of Floquet theory

The Floquet formalism provides solutions for the time dynamics of systems whose Hamiltonian $\mathcal{H}(t)$ is periodic with period T , i.e., $\mathcal{H}(t) = \mathcal{H}(t + T)$. It is particularly useful for analyzing periodically driven systems, as it captures the effects of strong drives beyond the perturbative regime and the rotating wave approximation (RWA).

The unitary time-evolution operator is given by $U(t_2, t_1) = \mathcal{T} \exp \left(-i \int_{t_1}^{t_2} \mathcal{H}(t) dt \right)$, where $\mathcal{T}$ is the time-ordering operator. When $\mathcal{H}(t)$ is periodic, it can be written in the form

$$U(t_2, t_1) = e^{-iK[t_0](t_2)} e^{-i\mathcal{H}_F[t_0] \times (t_2 - t_1)} e^{iK[t_0](t_1)}. \quad (\text{S14})$$

Here, we have decomposed the dynamics within a rotating frame given by $U(t) = e^{iK[t_0](t)}$, where $K[t_0](t)$ is the Floquet kick operator. In this frame, the dynamics are generated by a time-independent Floquet Hamiltonian $\mathcal{H}_F[t_0]$. The Floquet Hamiltonian and the kick operator are defined as

$$e^{-i\mathcal{H}_F[t_0]T} = U(t_0 + T, t_0) \quad (\text{S15})$$

$$e^{-iK[t_0](t)} = U(t, t_0) e^{i\mathcal{H}_F[t_0] \times (t - t_0)}, \quad (\text{S16})$$

where $U(t_0 + T, t_0)$ is the unitary over one period. The brackets are used to indicate that the terms are gauge-dependent on the reference stroboscopic time t_0 . The kick operator has the same periodicity as the Hamiltonian ($K[t_0](t + T) = K[t_0](t)$) and $e^{-iK[t_0](t)} = I$ at stroboscopic times $t = t_0 + nT$, where n is an integer. The Floquet modes are defined as the eigenstates of the unitary over one period:

$$U(t_0 + T, t_0) |\Phi_m[t_0]\rangle = e^{-i\epsilon_m T} |\Phi_m[t_0]\rangle. \quad (\text{S17})$$

Here, $|\Phi_m[t_0]\rangle$ is the Floquet mode indexed by m and ϵ_m is its corresponding quasienergy. We compute the Floquet modes and quasienergies by numerically diagonalizing the unitary over one period.

Sideband interactions are analogous to Rabi oscillations within a two-state subspace. In this subspace, the Hamiltonian is described by a driven two-level system:

$$\mathcal{H} = \frac{\omega_q}{2} \sigma_z + \Omega (e^{-i\omega_d t} \sigma_+ + e^{i\omega_d t} \sigma_-). \quad (\text{S18})$$

The Floquet Hamiltonian can be derived by transforming into the rotating frame given by $U(t) = e^{-i\frac{\omega_d}{2}(t - t_0)(\sigma_z - I)}$. This yields

$$\mathcal{H}_F[t_0] = \frac{\Delta}{2} \sigma_z + \frac{\omega_d}{2} + \Omega (e^{i\omega_d t_0} \sigma_+ + e^{-i\omega_d t_0} \sigma_-), \quad (\text{S19})$$

where $\Delta = \omega_q - \omega_d$. The eigenfrequencies of this Hamiltonian now correspond to the Floquet quasienergies:

$$\epsilon_{\pm} = \frac{1}{2} \left(\omega_d \pm \sqrt{4\Omega^2 + (\omega_q - \omega_d)^2} \right) \mod \omega_d. \quad (\text{S20})$$

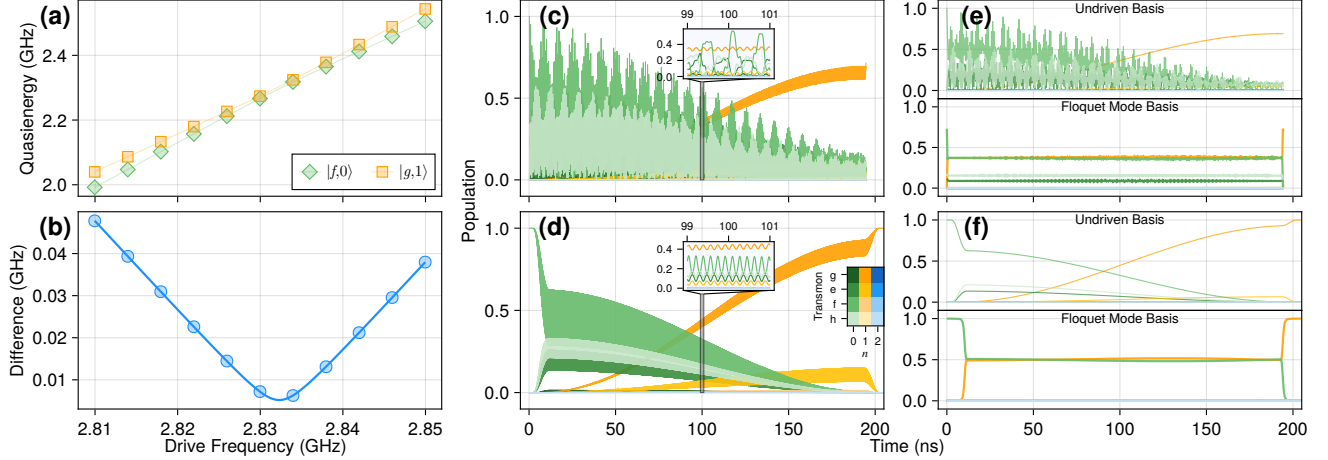


FIG. S9. **Numerical characterization of the sideband interaction.** (a) Quasienergies of the Floquet modes $|f,0\rangle_F$ and $|g,1\rangle_F$ for mode three as a function of drive frequency. (b) The fit of the difference in quasienergies is used to extract the resonance frequency and interaction rate. (c) Population evolution in the undriven basis for an $|f0\rangle$ - $|g1\rangle$ sideband with a ramp time of 0.5 ns, showing significant leakage. The inset shows the micro-oscillations in the populations. (d) Corresponding evolution for an $|f0\rangle$ - $|g1\rangle$ sideband with a ramp time of 11.6 ns. (e) Projection of the populations at stroboscopic times into the undriven Hamiltonian basis (top) and the Floquet mode basis (bottom) for a 0.5 ns ramp time. With such a short ramp, the initial state starts in a superposition of many Floquet modes, including those outside of the interaction subspace. (f) Corresponding evolution for a ramp time of 11.6 ns. Here, the ramp initializes the state into an equal superposition of $|f,0\rangle_F$ and $|g,1\rangle_F$ (see Section S7B), minimizing leakage and off-resonant rotation.

The strength and resonance frequency of the sideband interaction are determined by identifying the two relevant interacting Floquet modes in the quasienergy spectrum. As in the two-level case, the size of the avoided crossing (2Ω) gives the interaction strength, while the resonance frequency corresponds to the drive frequency at which the gap is minimized (Fig. S9(a) and (b)).

B. Unitary evolution of drives with smooth ramping

Smoothly ramping up (and down) the drive strength mitigates leakage out of the interaction subspace when compared to abruptly turning on the drive, such as when using a square pulse. Eq. S14 provides a way to understand the first of two roles played by the ramp. The first, when its timescale is much longer than the Hamiltonian period ($\tau_{\text{ramp}} \gg T$), is to adiabatically evolve the undriven eigenstates of the system into their corresponding Floquet modes (and back, on the ramp down). By noting that the Floquet modes are eigenstates of $\mathcal{H}_F[t_0]$, this adiabatic transitioning of the undriven interaction subspace to its corresponding subspace in the Floquet mode basis prevents leakage.

The second role of the ramp is to initialize the starting state into an equal superposition of Floquet modes within the interaction subspace, which ensures that the Rabi oscillations are full contrast. This is performed by diabatically traversing a Landau-Zener transition among

the Floquet modes.

For the sideband interaction, the interaction subspace is $\{|f,n\rangle_F, |g,n+1\rangle_F\}$, where the subscript is used to indicate the corresponding Floquet mode. Here, the Landau-Zener velocity is $v_{\text{LZ}} \sim |\omega_{\text{SS}}|/(2\pi\tau_{\text{ramp}})$, where $\omega_{\text{SS}} = \omega_{\text{sb}}(\epsilon = \epsilon_{\text{max}}) - \omega_{\text{sb}}(\epsilon = 0)$ is the Stark shift of the sideband resonance at maximum drive amplitude. The avoided crossing energy is $2g_{\text{sb}}$, where g_{sb} is the sideband interaction rate. Hence, the Landau-Zener diabaticity can be defined as

$$\Gamma_{\text{LZ}} = \frac{2\pi|\omega_{\text{SS}}|}{\tau_{\text{ramp}}g_{\text{sb}}^2} = \frac{4\pi\Delta^2}{\tau_{\text{ramp}}|K|g^2}, \quad (\text{S21})$$

where we have used Eq. S6 for g_{sb} and $\omega_{\text{SS}} \approx -2K\xi^2$, derived from perturbation theory. Since the avoided crossing must be traversed diabatically, Γ_{LZ} must be large. Hence, the time of the ramp τ_{ramp} must obey the conditions

$$\frac{2\pi}{\omega_{\text{sb}}} \ll \tau_{\text{ramp}} \ll \frac{4\pi\Delta^2}{|K|g^2}, \quad (\text{S22})$$

where the lower bound prevents leakage out of the interaction subspace and the upper bound ensures full contrast oscillations. The upper bound can be eliminated by chirping the frequency of the drive so that it always remains on resonance with the interaction. These two roles played by the ramp are illustrated in Fig. S9. For a ramp time that is too short (Fig. S9(e)), the initialization is not sufficiently adiabatic to evolve the undriven

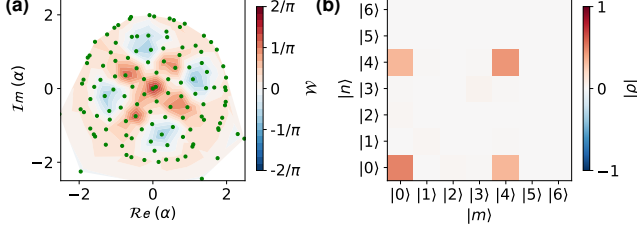


FIG. S10. **Wigner tomography.** (a) Green points are Wigner displacements chosen to minimize the condition number for a chosen Hilbert space cutoff, shown here along with the Wigner data for $\frac{|0\rangle + |4\rangle}{\sqrt{2}}$. (b) The absolute value of the reconstructed density matrix after applying an inversion protocol that enforces positivity and preserves unit trace.

eigenstates to their corresponding Floquet modes. This is seen in the initialized state being a superposition many Floquet modes, including those outside of the interaction subspace. Fig. S9(f) shows the evolution using a ramp that satisfies Eq. S22. Here, the state is initialized into an equal superposition of $|f0\rangle_F$ and $|g1\rangle_F$ which leads to an optimal fidelity interaction.

S8. WIGNER TOMOGRAPHY

We perform single-mode Wigner tomography with slight modifications of the standard protocol consisting of a cavity displacement before a parity measurement ($\hat{\Pi}$). The parity measurement is performed using a transmon $|g\rangle$ - $|e\rangle$ Ramsey sequence after idling for a time $1/(2\chi_i)$, where χ_i is the dispersive shift of the mode. This results in the phase of the transmon superposition state precessing by odd and even multiples of π for the corresponding photon number parity, which is mapped to the transmon state by a final $\pi/2$ pulse.

Ideally, the overall sequence corresponds to the measurement of the displaced Wigner operator, $\hat{W}(\alpha) = \hat{D}_\alpha \hat{\Pi} \hat{D}_{-\alpha}$. The cavity displacements are chosen at an optimal set of points (α_i) and the corresponding Wigner operator measurements are given by $x_i = \langle \langle \hat{W}(\alpha_i) | \rho \rangle \rangle$. We construct a matrix $\mathcal{M}$ with $\mathcal{M}_{ij} = \langle \langle \hat{W}(\alpha_i) | \hat{W}(\alpha_j) \rangle \rangle$ that represents measurements of the Wigner operator at all the displacements, which is inverted while imposing positivity and unit trace constraints, thereby extracting the most likely physical density matrix corresponding to the cavity state [11]. The number of columns of $\mathcal{M}$ is d^2 , where d is the truncated dimension of the Hilbert space of the cavity up to which the tomography is valid, while the rows correspond to the number of Wigner measurements ($> d^2$). The displacements are chosen to minimize the error from the inversion, corresponding to minimizing the ratio of the largest to the smallest eigenvalue of $\mathcal{M}$ (the condition number).

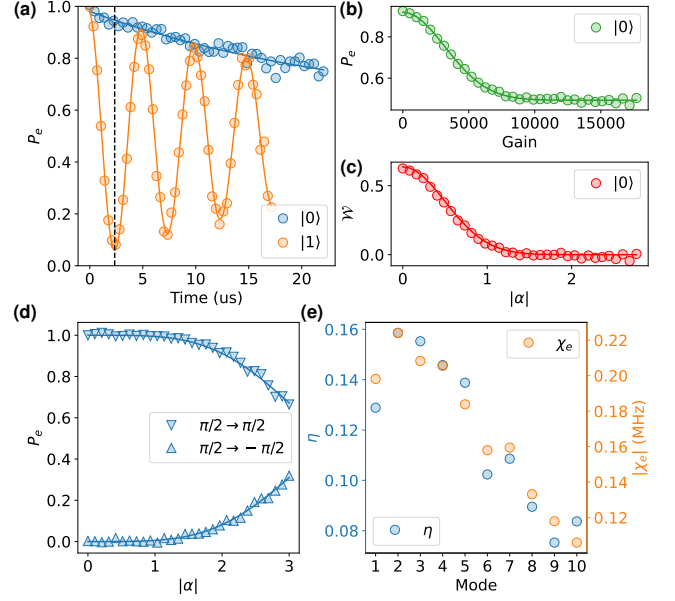


FIG. S11. **Parity measurement error mitigation and contrast correction.** (a) Transmon Ramsey experiment with $\frac{\pi}{2}$ pulses with identical phases following initializing the cavity in the vacuum or $|1\rangle$ state. The dotted black line shows the $1/(2\chi)$ time. (b) Transmon $|e\rangle$ probability (P_e) after the Wigner tomography sequence, as a function of the gain of the gaussian cavity displacement pulse, fit to the functional form $P_e(a) = (1 + e^{-2(sa)^2})(p_e - p_g)/2 + p_g$. (c) Wigner function versus α for $|0\rangle$ of mode 3 after rescaling the measurement using Eqn. S33. (d) Contrast reduction of a parity measurement as a function of α for mode 3 arising from dispersive frequency shift of the transmon and finite pulse bandwidth, fit to the functional form described in the text. (e) The fit value of η is $\propto \chi_e$ as expected.

A. Error mitigation of parity measurement

For weak coupling, the Ramsey time for the parity measurement can constitute a significant fraction of the transmon decoherence time, leading to errors in the parity measurement caused by the transmon's decay and dephasing. This can be seen in Fig. S11(a) where we show a Ramsey experiment after preparing the cavity in $|0\rangle$ or $|1\rangle$. However, since this time is much shorter than the cavity decay time, the resulting error can be calibrated and corrected. This approach represents a form of ancilla error mitigation that enables the recovery of parity measurement expectation values and statistics.

In the following, we rigorously demonstrate the feasibility of this error mitigation strategy by modeling the decoherence during the parity measurement using a Lindblad master equation

$$\frac{d\rho(t)}{dt} = -i[\mathcal{H}, \rho(t)] + \sum_{i=0}^1 \left(L_i \rho(t) L_i^\dagger - \frac{1}{2} \{ L_i^\dagger L_i, \rho(t) \} \right), \quad (\text{S23})$$

where

$$\mathcal{H} = \frac{\chi}{2} b^\dagger b \sigma_z \quad (\text{S24})$$

$$L_0 = \sqrt{\gamma} \sigma_- \quad (\text{S25})$$

$$L_1 = \sqrt{\frac{\gamma_\varphi}{2}} \sigma_z. \quad (\text{S26})$$

Here, γ is the decay rate and γ_φ is the pure dephasing rate of the qubit. Our system evolves under the dispersive Hamiltonian. The general time evolution of the density matrix can be solved to give

$$\begin{aligned} \rho(t) = \sum_{n_1 n_2} & \left(\rho_{n_1 n_2}^{ee}(t) |en_1\rangle \langle en_2| \right. \\ & + \rho_{n_1 n_2}^{eg}(t) |en_1\rangle \langle gn_2| \\ & + \rho_{n_1 n_2}^{ge}(t) |gn_1\rangle \langle en_2| \\ & \left. + \rho_{n_1 n_2}^{gg}(t) |gn_1\rangle \langle gn_2| \right) \end{aligned} \quad (\text{S27})$$

where

$$\rho_{n_1 n_2}^{ee}(t) = e^{-\gamma t} e^{-i\frac{\chi}{2}(n_1 - n_2)t} \rho_{n_1 n_2}^{ee}(0) \quad (\text{S28})$$

$$\rho_{n_1 n_2}^{eg}(t) = e^{-(\frac{\gamma}{2} + \gamma_\varphi)t} e^{-i\frac{\chi}{2}(n_1 + n_2)t} \rho_{n_1 n_2}^{eg}(0) \quad (\text{S29})$$

$$\rho_{n_1 n_2}^{ge}(t) = e^{-(\frac{\gamma}{2} + \gamma_\varphi)t} e^{i\frac{\chi}{2}(n_1 + n_2)t} \rho_{n_1 n_2}^{ge}(0) \quad (\text{S30})$$

$$\begin{aligned} \rho_{n_1 n_2}^{gg}(t) = & \left(\rho_{n_1 n_2}^{gg}(0) + \frac{\gamma \rho_{n_1 n_2}^{ee}(0)}{\gamma + i\chi(n_1 - n_2)} \right) e^{i\frac{\chi}{2}(n_1 - n_2)t} \\ & - \frac{\gamma \rho_{n_1 n_2}^{ee}(0)}{\gamma + i\chi(n_1 - n_2)} e^{-\gamma t} e^{-i\frac{\chi}{2}(n_1 - n_2)t} \end{aligned} \quad (\text{S31})$$

and we have used $t = 0$ as our reference time. Assuming the qubit is in $|g\rangle$ and disentangled with the cavity, we can perform a parity measurement by applying a $\frac{\pi}{2}$ -pulse, waiting $\tau = \frac{1}{2\chi}$, and then applying another $\frac{\pi}{2}$ -pulse. Here, we can see that the probability of measuring the qubit in $|e\rangle$ is

$$P_e = \frac{1}{2} \left(1 - e^{-(\frac{\gamma}{2} + \gamma_\varphi)t} \right) + e^{-(\frac{\gamma}{2} + \gamma_\varphi)t} P_{\text{even}}. \quad (\text{S32})$$

Hence, under decay and dephasing of the qubit we can still recover the parity measurement statistics by offsetting and rescaling the qubit measurement outcomes. Consistent with the expression derived above, we convert between the transmon measurement P and the expected parity measurement $\hat{\Pi}$ using $\hat{\Pi} = \frac{P - p_g}{p_e - p_g}$.

The p_g, p_e values may be calibrated by measuring the decay in contrast of a transmon Ramsey sequence. For our device, due to the spurious dispersively coupled mode excited during transmon readout (see Section S3D), the Ramsey experiment exhibits small micro-oscillations that lead to an additional few percent error if we rely solely on the exponential fit. We therefore estimate p_g and p_e by using the Wigner function measurement for the vacuum state of each mode as a calibration. The displacement

pulses used in the tomography protocol are 4σ Gaussian pulses with $\sigma = 50$ ns for most modes (125 ns for modes 8, 9, and 10). We achieve different cavity displacements (α) by adjusting the pulse gain and phase. We fit the measured transmon probability versus gain ($P(a)$) to extract the scaling ($\beta = \alpha/a$) between the displacement and the pulse gain, as well as p_g and p_e , as shown in SFig. 6(b). We subsequently convert between $P(a)$ and $\mathcal{W}(\alpha)$ using:

$$\mathcal{W}(\alpha) = \mathcal{W}(\beta a) = \frac{2}{\pi} \left(2 \left(\frac{P(a) - p_g}{p_e - p_g} \right) - 1 \right) \quad (\text{S33})$$

In addition to decoherence, the parity measurement also has a systematic error arising from the finite bandwidth of the parity-measurement $\pi/2$ pulse [5]. This leads to a reduction in the contrast in the measurement of the Wigner function for larger cavity displacements, which we calibrate using a sequence comprising two back-to-back $\pi/2$ pulses with either the same or opposite phase, without waiting time between. The result of such a measurement for a given mode (2) is shown in SFig.6(c). The reduction in contrast is consistent with that expected from the off-resonant error in the $\pi/2$ pulse from a frequency shift of $\chi\alpha^2$. The curves are fit to a functional form $C(\alpha) = \frac{1}{1 + \eta^2 \alpha^4}$ and $1 - C(\alpha)$. Here $\eta = \chi/(2\epsilon)$ where ϵ is the strength of the qubit Rabi drive. The η extracted from the fit to the parity measurement is found to be proportional to χ , as shown in SFig. 6(d).

We ignore the additional readout error arising from the cross-Kerr interaction between the storage and readout modes [5], since this is sufficiently small. We compensate for this effect by scaling the measurement of the Wigner operator for a given state and displacement using a linear transformation ($\mathcal{W}(\alpha, \rho) \rightarrow \mathcal{W}(\alpha, \rho)/C(\alpha)$). We note that this calibration was empirically found to only change the fidelity by a few percent.

The reconstruction procedure described above provides the best estimate of the cavity density matrix while mitigating errors in the parity measurement caused by transmon decay and coherent pulse errors. However, several sources of error can still affect the reconstructed ρ , including random measurement noise, uncertainty in the calibration of the parity measurement (p_e, p_g), and errors arising when the transmon is not in the $|g\rangle$ state at the start of the measurement. For the first two sources of error, we estimate the resulting uncertainty in the fidelity and add it to our error bar. To mitigate the latter error, we perform a transmon measurement prior to Wigner tomography and post-select based on the result. We detail this error analysis in the following sections.

B. Wigner tomography with post-selection

The standard protocol for Wigner tomography relies on the transmon being in the $|g\rangle$ state and initially dis-

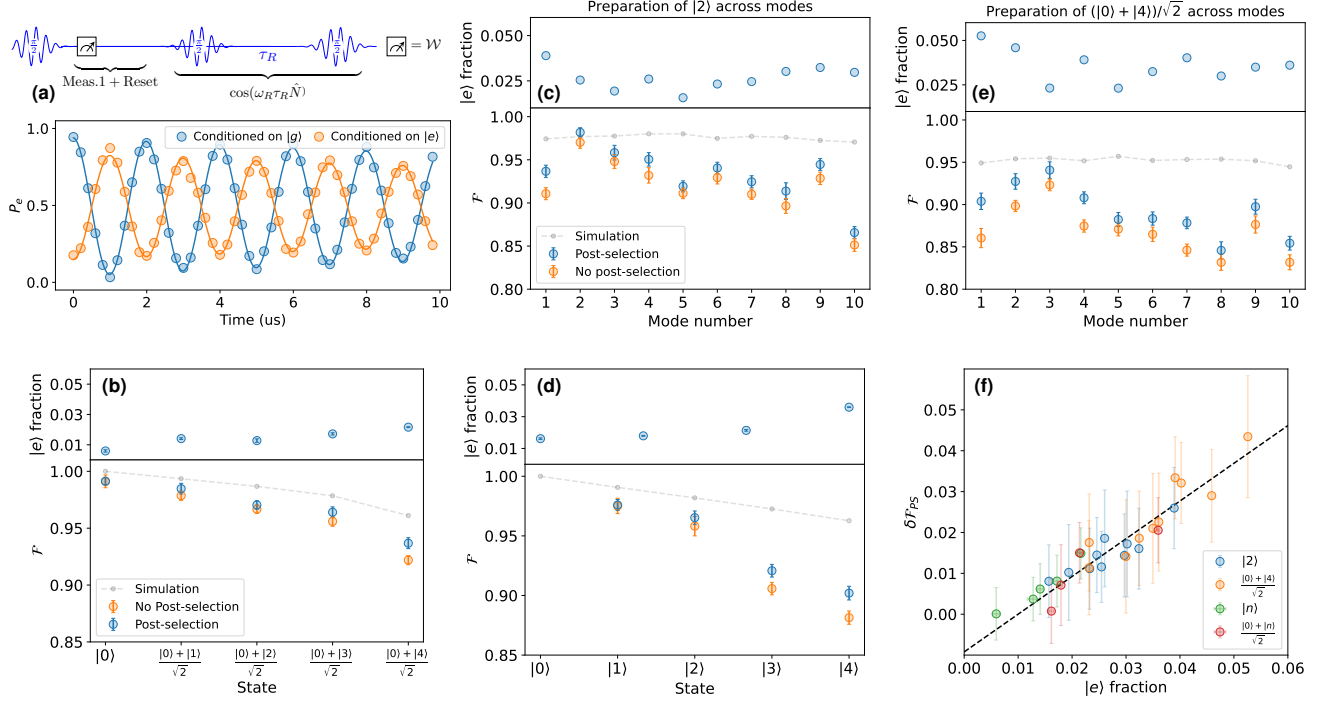


FIG. S12. **Wigner tomography with post-selection.** (a) Validation of post-selection (without cavity displacement), showing the outcome of the second transmon measurement conditioned on the first measurement result while sweeping the idle time τ_R in a subsequent Ramsey experiment. The transmon is initially prepared in the state $\frac{|g\rangle + |e\rangle}{\sqrt{2}}$ (pulse sequence shown at the top). (b) Wigner tomography results with and without post-selection for the preparation of the state $\frac{|0\rangle + |n\rangle}{\sqrt{2}}$ in mode 3. (Top) The fraction of $|e\rangle$ outcomes in the first measurement as a function of n . (Bottom) Comparison of state fidelity from Wigner tomography as a function of n , both without post-selection and with post-selection on the transmon being in $|g\rangle$ at the end of the first measurement. Corresponding results for (top) $|e\rangle$ outcomes in the first measurement and (bottom) state fidelities with and without post-selection (on $|g\rangle$) for (c) the $|n\rangle$ preparation sequence applied to mode 3, (d) the preparation of $|2\rangle$, and (f) the preparation of $\frac{|0\rangle + |4\rangle}{\sqrt{2}}$ across the first ten cavity modes. (f) Improvement in state-fidelity with and without post-selection δF_{PS} as a function of the fraction of shots in $|e\rangle$ in the first post-selection measurement across all the measurements.

entangled from the cavity state. Although the sideband-based state preparation sequences we implement are ideally designed to achieve this, transmon and cavity decoherence and coherent calibration errors can result in the transmon and the cavity modes remaining weakly entangled at the end of state preparation. To mitigate the resulting uncontrolled tomography error, we implement a protocol for Wigner tomography with post-selection. Here, we perform an initial transmon measurement before Wigner tomography and post-select outcomes, where the first measurement confirms that the transmon is in the $|g\rangle$ state. The pulse sequence corresponding to this protocol is shown in Fig. S12(a).

We test the efficacy of post-selection in the context of Wigner tomography through an experiment consisting of a $\pi/2$ pulse, an initial post-selection measurement and readout cavity reset, followed by a Ramsey sequence and a second transmon measurement. The results of this experiment, conditioned on the first measurement outcome being $|g\rangle$ or $|e\rangle$, are shown in Fig. S12(b). We observe the

expected anti-correlated oscillations, and a drop in contrast for an $|e\rangle$ outcome in the first measurement from transmon decay during the measurement and readout reset.

We present a comparison of the measured fidelities with and without post-selection for the preparation of $|n\rangle$ and $\frac{|0\rangle + |n\rangle}{\sqrt{2}}$ superposition states for a given mode (3) in Fig. S12(c) and (d). We also show the corresponding data for state preparation of Binomial code states across the 10 modes of our system in Fig. S12(e) and (f). We see that across all these state preparation experiments, there is an increase in the Wigner tomography fidelity with post-selection linearly correlated with the fraction of shots measured in $|e\rangle$, as shown in Fig. S12(g).

C. Fast reset of the readout cavity

To minimize infidelity from cavity decoherence during post-selection, we developed a fast, integrated readout

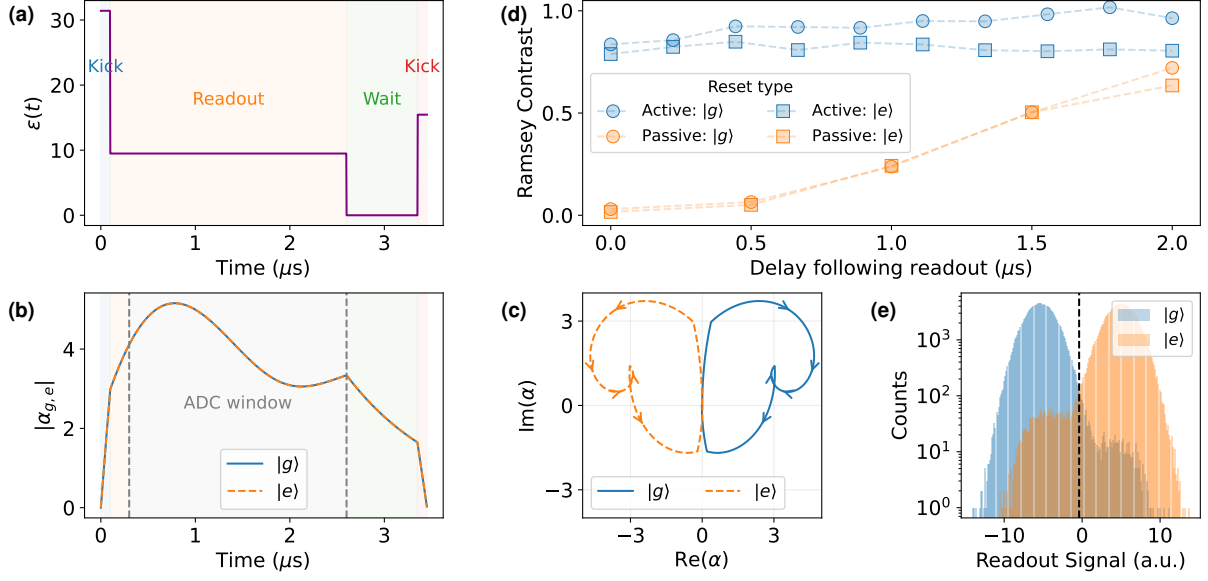


FIG. S13. **Integrated readout and reset pulse.** (a) Pulse sequence for integrated readout and readout reset pulse. (b) Evolution of the readout mode's displacement magnitude for each of the transmon states. (c) Evolution of the readout mode in phase space for each transmon state. (d) Characterization of the readout mode photon population using a transmon Ramsey measurement, illustrating that the reset protocol clears readout photons much faster than passive reset. (e) Histogram of readout signal along the rotated quadrature axis that most distinguishes the single-shot readout histograms, with the threshold line corresponding to a readout fidelity of 97.5%.

and reset pulse similar to the CLEAR pulse [12]. This pulse is particularly necessary due to our relatively low readout $\kappa \approx 1/(600 \text{ ns})$. The schematic of the pulse is shown in Fig. S13(a), with a carrier frequency positioned between the readout resonator frequencies corresponding to the $|g\rangle$ and $|e\rangle$ states. The evolution of the magnitude of the cavity displacement and the cavity state in phase space are illustrated in Fig. S13(b) and (c), respectively.

The pulse consists of an initial ring-up segment that rapidly drives the cavity to the target displacement, followed by a readout segment during which the cavity settles to the nominal values for each state ($\alpha_{g,e} \approx \frac{2\epsilon}{\kappa \pm i\chi}$), corresponding to angles $\theta_{e/g} = \pm \arctan \frac{\chi}{\kappa} = \pm \theta_r$ about the imaginary axis. To reset the cavity states corresponding to both transmon states, the reset pulse includes a wait period during which the cavity states rotate in opposite directions, converging on the opposite side of the cavity phase space. This is followed by a final pulse that pushes both states back to the origin. The total reset time is given by $\tau_r = \frac{2(\pi - \arctan \frac{\chi}{\kappa})}{\chi} \approx 0.7 \mu\text{s}$.

We optimize the amplitude and duration of each component of the readout and reset sequence using the contrast of a transmon Ramsey experiment performed immediately after the readout and reset pulse, along with readout histogram measurements. This optimization minimizes transmon dephasing caused by residual readout photons while simultaneously maximizing readout fidelity. The contrast following the optimized readout pulse is compared with that of a regular uniform read-

out pulse in Fig. S13(d). We observe that the contrast remains near its full value immediately after readout for the active reset pulse, unlike the uniform (passive) readout pulse. Additionally, we include an extra $2 \mu\text{s}$ after the readout and active reset pulse to allow any residual readout photons to fully leak out of the cavity.

With this readout pulse, we are able to achieve a readout fidelity of 97.5% without the use of a parametric amplifier, with an integration time of approximately $1.5 \mu\text{s}$, the corresponding histogram for which is shown Fig. S13(e). This corresponds to a 1.5% (0.5%) assignment error for the transmon measured in $|e\rangle$ ($|g\rangle$). We note that we may further reduce the assignment error to $< 0.5\%$ by post-selecting shots that lie in regions where the assignment error is below that threshold. This necessitates throwing away nearly 50% of the measurements where the first measurement is $|g\rangle$, coming at the expense of increased reconstruction errors from increased measurement noise. This compromise, in conjunction with $> 95\%$ of shots being in $|g\rangle$ resulted in no improvement in Wigner tomography fidelity with further reducing assignment error. We therefore chose to perform post-selected Wigner tomography experiments with the standard threshold voltage, indicated by the dashed black line in Fig S13(e).

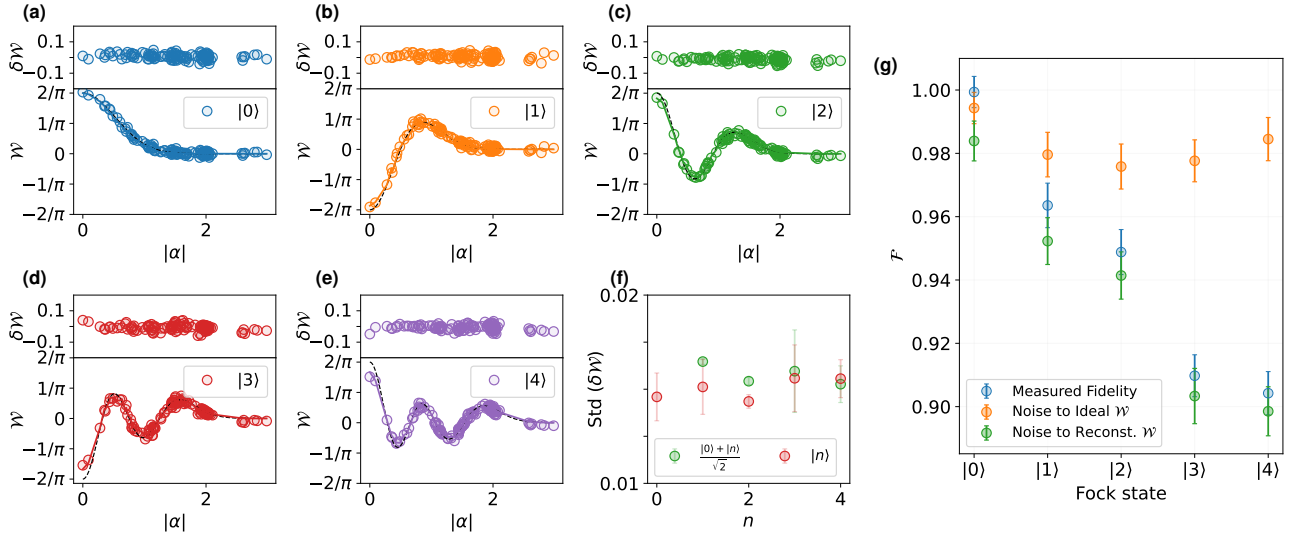


FIG. S14. **Wigner tomography error analysis.** (a-e) Comparison of measured Wigner functions (circle markers) with those of the reconstructed ρ (solid colored lines) and ideal Wigner functions (dashed black lines) as a function of $|\alpha|$ for Fock states ($n = 0$ to 4) in mode 3 of the multimode register. (f) Standard deviation of the difference δW between the measured and reconstructed Wigner functions for data corresponding to the preparation of $|n\rangle$ and $\frac{|0\rangle+|n\rangle}{\sqrt{2}}$ in mode 3. (g) Fidelity and spread obtained from reconstruction following the addition of random noise to ideal and reconstructed Wigner functions.

D. Wigner tomography error analysis

In addition to errors from transmon excitation, Wigner tomography infidelity also arises from measurement noise and fluctuations in the calibration of the parity measurement (p_g, p_e values). We estimate measurement noise by comparing the measured Wigner function with the expected Wigner function for the reconstructed ρ at corresponding phase-space points. Their comparison and their difference for Fock states up to $n = 4$ in mode 3 are shown in Fig. S14(a-e). The standard deviation of the noise, plotted as a function of n for Wigner functions of both $|n\rangle$ and $\frac{|0\rangle+|n\rangle}{\sqrt{2}}$ in Fig. S14(f), shows that the error remains largely consistent across states.

To quantify the fidelity error due to measurement noise, we introduce random noise with the same standard deviation at each data point and perform the reconstruction multiple times (50 iterations). This noise addition is applied to both an ideal Wigner function for $|n\rangle$ and the Wigner function corresponding to the reconstructed ρ from our measurements, with the results shown in Fig. S14(g). Reconstruction following the addition of noise to the ideal Wigner function leads to a drop in the fidelity, as expected. Measurement noise, similarly leads to an underestimate in the fidelity of our measured state by $\sim 2\%$. While the spread in the fidelity is similar for both cases, we take the spread in the fidelity from noise added to the reconstructed Wigner function as a proxy for the error in the fidelity of our reconstructed state in the results presented here.

Finally, we account for fidelity error due to fluctuations in the parity measurement calibration, specifically

in the values of p_e and p_g used to rescale the parity measurement. These fluctuations arise from variations in the Ramsey contrast occurring over the timescale of a Wigner tomography measurement. To quantify this effect, we determine the spread in p_e and p_g through repeated calibrations of the vacuum state and estimate the corresponding fidelity error. For mode 3, this corresponds to $p_g \pm \delta p_g = 0.075 \pm 0.009$ and $p_e \pm \delta p_e = 0.912 \pm 0.006$ over seven datasets. We observe that p_g exhibits nearly a 1% variation, which leads to a comparable error in the reconstructed fidelity. To estimate the fidelity error from this effect, we perform the reconstruction for $(p_g \pm \delta p_g, p_e \pm \delta p_e)$ and evaluate the resulting spread in fidelity. For the data presented in Fig. S12, error bars for individual datasets are obtained by adding the errors from measurement noise and parity calibration in quadrature. The final data is presented as a weighted average of all available datasets using the same calibration routine.

S9. BINOMIAL CODE ENCODING GATE

A. Photon-number-selective sideband calibration

The binomial encoding gate maps the transmon states $\{|g\rangle, |e\rangle\}$ to the binomial code states $\{|2\rangle, \frac{|0\rangle+|4\rangle}{\sqrt{2}}\}$ through a sequence of transmon rotations and fast $|f, n\rangle - |g, n+1\rangle$ sideband pulses, along with two photon-number-selective sideband pulses. The latter leverages the photon-number dependence of the $|fn\rangle - |g, n+1\rangle$ sideband transition frequency, which is dispersively shifted by $n\chi_f$ as shown in Fig. 2(b) of the main

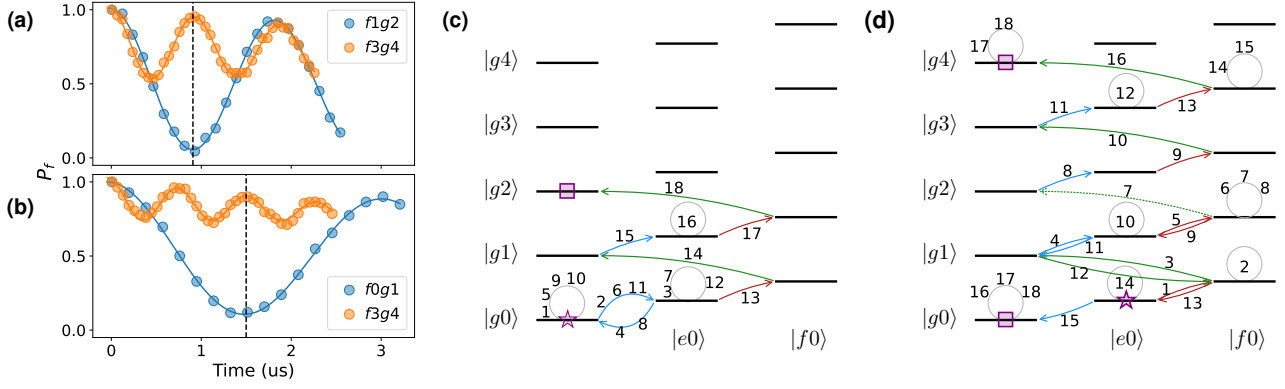


FIG. S15. **Calibration of photon number-selective sideband.** (a) Measured $|f\rangle$ population after applying the sideband on resonance with the $f1-g2$ transition with $g_{sb} = |\chi_f|/\sqrt{2}$ on states prepared in $|f1\rangle$ (blue) and $|f3\rangle$ (orange). (b) Measured $|f\rangle$ population after applying the sideband drive on resonance with the $|f0\rangle$ - $|g1\rangle$ transition with $g_{sb} = \sqrt{3}|\chi_f|/4$ on states prepared in $|f0\rangle$ (blue) and $|f3\rangle$ (orange). (c) Combined Hilbert space of a transmon and one of the cavity modes, illustrating the state transitions induced by the encoding pulse sequence in Eq. S34. The purple star indicates the initial state, and the purple square is the final state ($|g\rangle \rightarrow |2\rangle$). The transition numbers correspond to the pulse numbers in Fig. 5(a) of the main text. (d) The state transitions for the encoding operation with the transmon initialized in $|e\rangle$ ($|e\rangle \rightarrow (|0\rangle + |4\rangle)/\sqrt{2}$).

text. The pulse targets a pair of different photon number states and utilizes the drive detuning to adjust the sideband rates to perform a π -pulse on a given photon number while implementing a $2m\pi$ -rotation on the other. This enables photon-number-selective sideband opera-

tions, significantly faster than achieved by targeting a single photon-number state, which requires $g_{sb} \ll |\chi_f|$ and can be prohibitively slow for weak coupling. The sequence used in this work (shown in Fig. 5 (a) of the main text) is:

$$\begin{aligned}
 U_{\text{enc}} : \pi_{ef} \rightarrow \pi_{ge} \rightarrow \pi_{f0g1} \rightarrow \pi_{ge} \rightarrow \pi_{ef} \rightarrow \pi_{ge} \rightarrow \frac{\pi}{2}_{f1g2} \rightarrow \pi_{ge} \rightarrow \pi_{ef} \rightarrow \pi_{f2g3} \\
 \rightarrow \pi_{ge} \rightarrow \pi_{f0g1} \rightarrow \pi_{ef} \rightarrow \left\{ \begin{array}{l} \pi_{f0g1} \\ 4\pi_{f3g4} \end{array} \right. \rightarrow \pi_{ge} \rightarrow \pi_{f3g4} \rightarrow \pi_{ef} \rightarrow \left\{ \begin{array}{l} \pi_{f1g2} \\ 2\pi_{f3g4} \end{array} \right. .
 \end{aligned} \tag{S34}$$

The two rate-limiting photon-number-selective pulses implement $\pi_{f0g1}-4\pi_{f3g4}$ at a rate $1/\tau = \sqrt{3}|\chi_f|$ and $\pi_{f1g2}-2\pi_{f3g4}$ at a rate $1/\tau = 2\sqrt{2}|\chi_f|$. Although these rates depend on the strength of the dispersive interaction, they scale with χ_f and feature prefactors greater than unity, accelerating selective operations by factors of 3.2 and 5.2 compared to χ_e , respectively. This rate can potentially be further enhanced without introducing additional Purcell loss to the cavity by operating the transmon near the straddling regime [13].

Since sideband interactions are analogous to Rabi oscillations in a two-state subspace, the rates above can be derived considering the Hamiltonian of a driven two-level system in Eq. S18. The Rabi oscillation rate for a detuning Δ from the sideband resonance is given by $\Gamma = \sqrt{g_{sb}^2 + \Delta^2}/4$. Consider sidebands acting on two states with differing photon numbers n_1 and n_2 where $n_2 \geq n_1$. Their resonant sideband rates taken to be g_{sb1} and $g_{sb2} = \sqrt{\frac{n_2+1}{n_1+1}}g_{sb1}$, respectively. By leveraging the

faster rate of a detuned Rabi oscillation, we can calibrate g_{sb1} to perform a π -pulse on the first state (n_1) and a $2m\pi$ -pulse on the second (n_2). Driving on resonance with the first state to perform a high-fidelity π pulse, we derive that

$$g_{sb1} = \frac{(n_2 - n_1)|\chi_f|}{2\sqrt{4m^2 - (n_2 + 1)/(n_1 + 1)}}. \tag{S35}$$

If we were instead to drive on resonance with the second state (which has a faster sideband rate), we require $g_{sb1} = \frac{(n_2 - n_1)|\chi_f|}{2\sqrt{4m^2(n_2 + 1)/(n_1 + 1) - 1}}$, which is strictly slower than driving on the first state.

We calibrate our photon-number-selective sidebands using Eq. S35 to estimate the required sideband rate. We iteratively perform sideband Rabi oscillations as a function of time while finely sweeping the drive amplitude around the expected value, simultaneously performing spectroscopy to determine the resonance frequency at

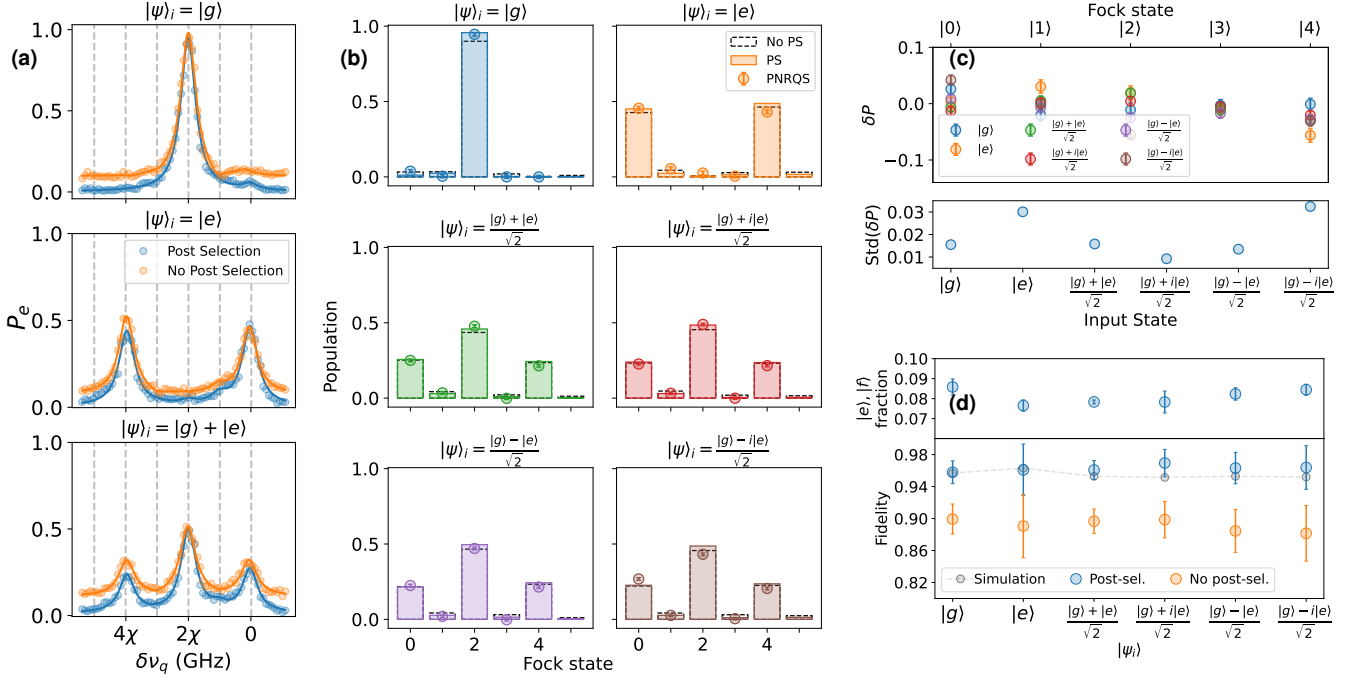


FIG. S16. **Binomial encoding gate characterization.** (a) Photon-number-resolved qubit spectroscopy (PNRQS) with and without post selection after running the encode sequence on $\{|g\rangle, |e\rangle, \frac{|g\rangle+|e\rangle}{\sqrt{2}}\}$, from top to bottom. (b) Comparing the populations extracted from PNRQS and from Wigner tomography (with and without post selection) following acting the encode gate after preparing the transmon in all the cardinal points on the $\{|g\rangle, |e\rangle\}$ Bloch sphere. (c) (Bottom) Improvement in fidelity from Wigner tomography with post selection and (top) the fraction of the shots that are excluded in the post-selection measurement (d) (Top) The differences in the populations extracted from PNRQS and Wigner tomography plotted as a function of the Fock level for the different input states. The standard deviation of the population differences for the different input states are shown in the bottom.

each amplitude. This process is repeated until the durations of the π -pulse and $2m\pi$ -pulse converge to $\sim 1\%$, resulting in a $\sim 0.01\%$ calibration error in the populations.

B. Gate characterization and error budget

The binomial encoding gate is characterized by preparing the transmon in the $|g\rangle, |e\rangle$ states at six cardinal points on the Bloch sphere and measuring the resulting cavity state after the gate operation. The cavity state is analyzed using Photon Number Resolved Qubit Spectroscopy (PNRQS) and Wigner Tomography, both with and without a transmon post-selection measurement prior to the final measurement. The results of PNRQS with and without post-selection are shown in Fig. S17(a). Without post-selection, transmon decay—particularly during the longer photon-number-selective sideband pulses—results in a higher fraction of population remaining in the $|e\rangle$ state. In the PNRQS measurement, this leads to an elevated baseline, which, as expected, is consistent with the excited-state fraction measured in the post-selected Wigner tomography experiment. The PNRQS and Wigner tomography measure-

ments are normalized according to the protocol presented in Section S11.

From the post-selected PNRQS measurements, we extract the cavity population in each Fock state by fitting the spectrum to a sum of Lorentzians while constraining the linewidths to be identical and the center frequencies to shift with photon number according to an independently measured χ . This corresponds to a measurement of the diagonal elements of the cavity density matrix, which we compare to those extracted from Wigner tomography for each input state in Fig. S17(b).

The measured populations from Wigner tomography and PNRQS, post-selected on the transmon being in $|g\rangle$, are found to be consistent, as shown in Fig. S17(c), where we plot the differences in the populations extracted from PNRQS and Wigner tomography for each input state. The standard deviation of the populations extracted from the two methods is found to be within 2%. This further corroborates the fidelities measured via Wigner tomography, shown in Fig. S17(d), both with and without post-selection. As before, the improvement in fidelity with post-selection closely matches the $|e\rangle$ fraction in the post-selection, which is also found to be consistent across Wigner tomography and PNRQS measurements.

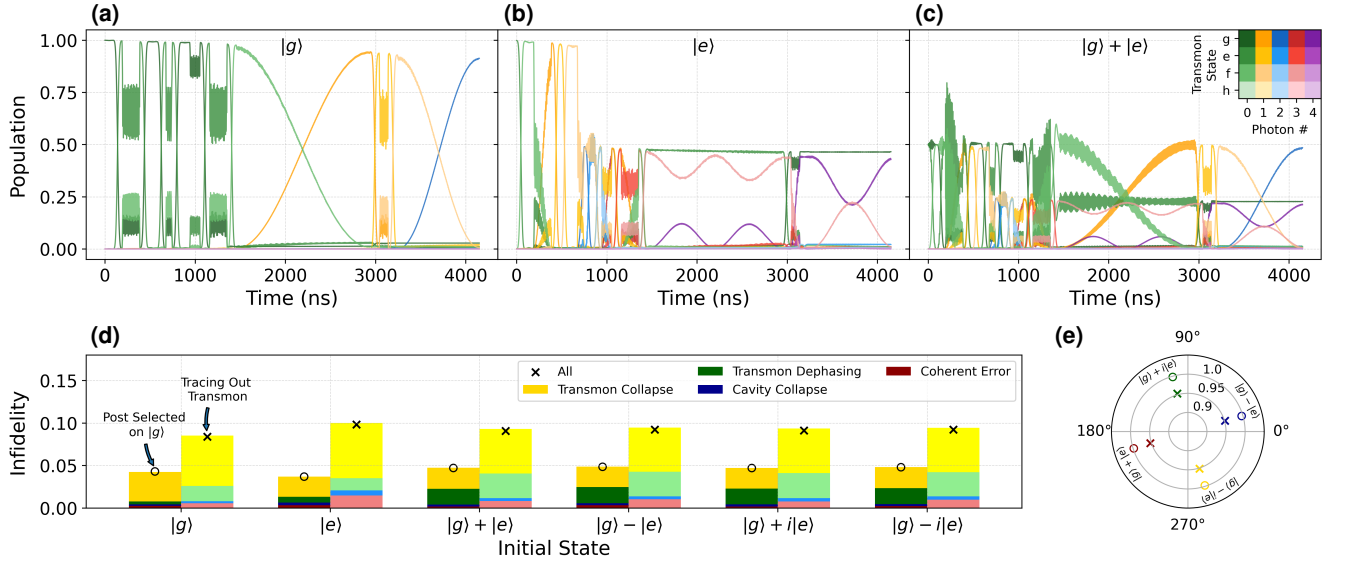


FIG. S17. **Error budget for binomial encoding gate.** (a)-(c) Lindblad master equation simulations for the binomial encoding sequence, starting from the states $|g0\rangle$ (a), $|e0\rangle$ (b), and $\frac{|g0\rangle+|e0\rangle}{\sqrt{2}}$ (c). (d) Error budget for the encoding operation. The primary source of infidelity arises from transmon decay and dephasing, with coherent errors being the next most significant contributor. The contribution from cavity decay is the smallest factor. In the simulations, we ignore cavity decay during the post-selection measurement, which we estimate to contribute an additional 1% to the infidelity. We present both the infidelity of the final cavity state post-selected on the transmon being in $|g\rangle$ and after tracing out the transmon. (e) The relative phases between $|0_L\rangle$ and $|1_L\rangle$ after encoding an initial state of the form $(|g\rangle + e^{i\varphi}|e\rangle)|0\rangle$ for $\varphi = 0, \pm\pi/2$, and π . The encoding preserves the relative phases.

S10. NOON STATE ENCODING GATE

The NOON state encoding gate maps the transmon states $\{|g\rangle, |e\rangle\}$ to the states $\{|N0\rangle, |0N\rangle\}$ in any pair of

cavity modes through a sequence of transmon rotations and alternating fast $|f, n\rangle \rightarrow |g, n+1\rangle$ sideband pulses to each mode and is shown in Fig. 4(c) of the main text and reproduced below. The first and second modes are indexed by m1 and m2, respectively:

$$U_{\text{enc}}(N) : \begin{cases} \pi_{ge} \rightarrow \pi_{ef} \rightarrow \pi_{f0-g1}^{m2} \rightarrow \pi_{ef} \rightarrow \pi_{f0-g1}^{m1} & \text{if } N = 1 \\ \pi_{ge} \rightarrow \pi_{ef} \rightarrow \pi_{f0-g1}^{m2} \rightarrow \pi_{ef} \rightarrow \pi_{ge} \rightarrow \pi_{f0-g1}^{m1} & \\ \rightarrow \sum_{n=1}^{(N-1)*} \left(\pi_{ef} \rightarrow \pi_{ge} \rightarrow \pi_{f,n-g,n+1}^{m2} \rightarrow \pi_{ef} \rightarrow \pi_{ge} \rightarrow \pi_{f,n-g,n+1}^{m1} \right) & \text{if } N > 1. \\ \rightarrow \pi_{ef} \rightarrow \pi_{ge} \rightarrow \pi_{f,N-1-g,N}^{m2} \rightarrow \pi_{ef} \rightarrow \pi_{f,N-1-g,N}^{m1} & \end{cases} \quad (\text{S36})$$

Here, the sum indicates the set of pulses, in parentheses, that must be played sequentially by index n and is not inclusive of the maximum index value $N-1$ (indicated by the asterisk). The resulting cavity state is characterized through three methods: photon number resolved qubit spectroscopy (PNRQS), qubit tomography following an encode and decode operation, and multimode Wigner tomography.

A. Multimode Wigner tomography

Multimode Wigner tomography is implemented using the pulse sequence shown in Figure 4(c) of the main text. In this procedure, both cavities are driven to a set of chosen displacements (α, β) , and we subsequently use a transmon Ramsey sequence to measure a generalized

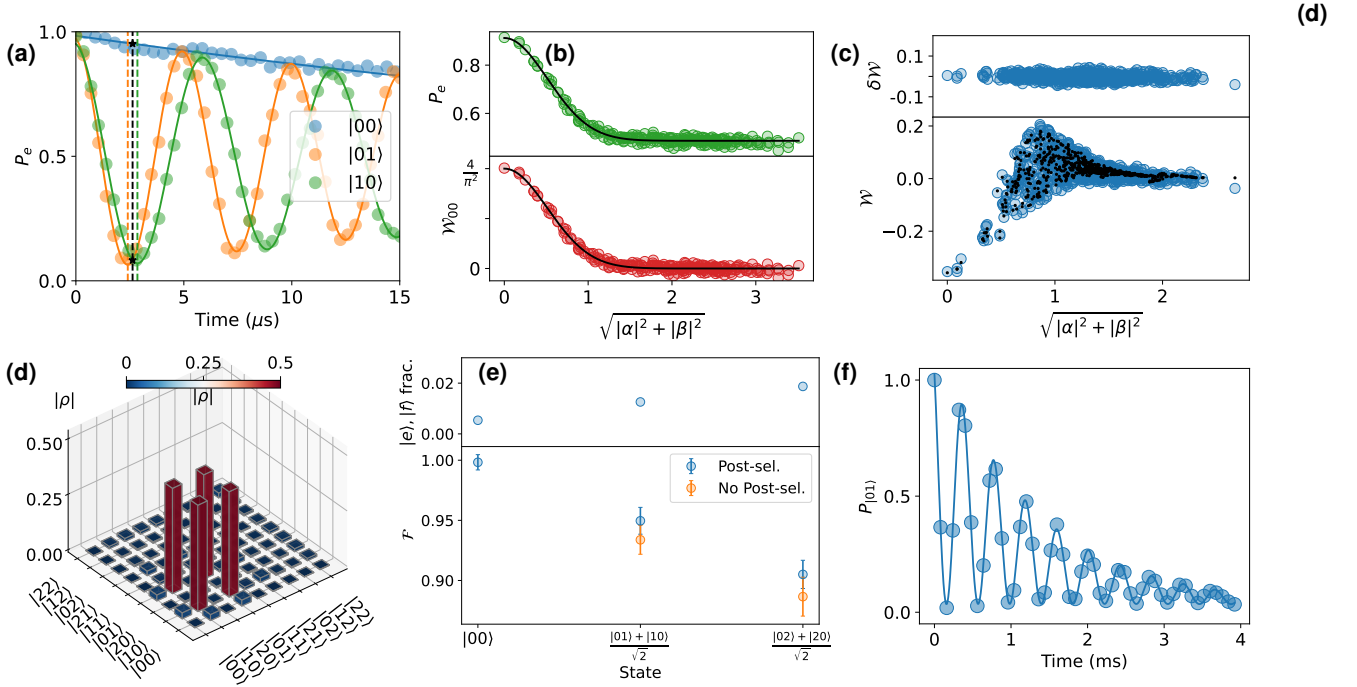


FIG. S18. **NOON state tomography.** (a) Calibration of the wait time for a generalized two-mode parity operation using transmon Ramsey experiments following the addition of a photon in each mode (here, modes three and five). The dashed black line indicates the chosen Ramsey time, which deviates from a parity measurement for both modes. The phases of the Ramsey fringes corresponding to a mode at that time is used to extract θ_1 and θ_2 . (b) (Top) Calibration of the contrast scaling (p_e, p_g) required for error mitigation of the joint parity measurement. (Bottom) Wigner function of the two-mode vacuum state following rescaling, as a function of $\sqrt{\alpha^2 + \beta^2}$. (c) Comparison of the measured data against the Wigner function at the corresponding points of the reconstructed density matrix. Their difference (top) is used to estimate an upper bound on the error in the Wigner measurement. (d) Reconstructed density matrix of a NOON state with $N = 1$. (e) (Top) Fraction of shots not in $|g\rangle$ in the post-selection experiment, found to be $\leq 2\%$ up to $N = 2$. (Bottom) Fidelity extracted from two-mode Wigner tomography for NOON states up to $N = 2$, with and without post-selection. (f) Ramsey experiment for the coherence of $(|10\rangle + |01\rangle)/\sqrt{2}$, resulting in a decay time of 1.27 ± 0.05 ms.

parity-like operator:

$$\begin{aligned} \tilde{\Pi} &= \cos(\theta_1 \hat{n}_1 + \theta_2 \hat{n}_2) \\ &= \cos(\theta_1 \hat{n}_1) \cos(\theta_2 \hat{n}_2) - \sin(\theta_1 \hat{n}_1) \sin(\theta_2 \hat{n}_2), \end{aligned} \quad (\text{S37})$$

where n_i is the photon number operator for mode i . The angles θ_1 and θ_2 are calibrated using Ramsey experiments following the addition of a photon into the respective modes, shown in Fig. S18(a). We calibrate the scale factor between the cavity drive amplitude and the cavity displacement for each mode with the Wigner function of the vacuum state, using the procedure detailed in Section S8. We calibrate the reduction in contrast from transmon decoherence during the parity measurement, due to the weak coupling, using the Wigner function of the two-mode vacuum $|00\rangle$. We fit the transmon population as a function of the distance of the two-mode Wigner function from the origin ($\sqrt{\alpha^2 + \beta^2}$), to extract p_e and p_g , similar to Eqn. S33, as shown in Fig. S18(b). This procedure allows us to extract the Wigner functions for two-mode states at a set of chosen cavity displacements for each mode ($\mathcal{W}(\alpha_i, \beta_j)$), which

we invert to extract the two-mode density matrix using the procedure detailed in [5]. The absolute value of the density matrix $|\rho|$ for the entangled Bell state $\frac{|10\rangle + |01\rangle}{\sqrt{2}}$ is shown in Fig. S18(c).

We perform two-mode Wigner tomography both with and without post-selection for the vacuum and NOON states up to $N = 2$; the results are shown in Fig. S18(e). We achieve fidelities of 0.95 ± 0.011 (0.934 ± 0.012) and 0.905 ± 0.012 (0.887 ± 0.016) with (without) post-selection for NOON states with $N = 1$ and $N = 2$, respectively. The improvement in fidelity with post-selection is consistent with the excited-state fraction shown at the top of Fig. S18(e). We estimate the error in our reconstructed fidelity by extending the methods presented in Section S8D to the two-mode case and assessing how measurement and calibration noise translate to fidelity error. Specifically, we estimate measurement noise by comparing the measured Wigner function with that of the reconstructed ρ at corresponding phase-space points; the results for the $N = 1$ NOON state are shown in Fig. S18(c). The deviation $\langle \delta \mathcal{W} \rangle$ is comparable to that

observed in the single-mode case. We inject Gaussian random noise with the same standard deviation and perform the reconstruction multiple times (10 iterations) to estimate the fidelity error.

B. NOON state coherence using encode and decode

In place of multimode Wigner tomography, we employ a faster method to characterize the state within the NOON subspace—after verifying via photon-number-resolved spectroscopy that less than 2% of the population lies outside this space. The state is characterized by performing qubit tomography following decoding of the NOON state. By idling in the NOON state between the encoding and decoding operations, we can measure its coherence time by sandwiching the encode, idle, and decode operations within a transmon ge Ramsey experiment. The result of this experiment (after resetting to remove measurement errors from populations outside the subspace) is shown in Fig. S18(f), corresponding to a coherence time of 1.27 ± 0.05 ms for the $\frac{|10\rangle + |01\rangle}{\sqrt{2}}$ Bell state.

S11. DATA NORMALIZATION

Probability normalization for transmon measurements was performed using a reference set of single-shot IQ

data, where the transmon was prepared in $|g\rangle$ and $|e\rangle$ after each experiment. A linear rotation transformation was applied to the IQ data to maximize the separation between the rotated median IQ values for $|g\rangle$ and $|e\rangle$ (I_g^{med} and I_e^{med} , respectively) along the I quadrature. Subsequent transmon measurement IQ data were rotated using the same transformation and then normalized along the I quadrature to extract the probabilities using $P_e = (I_{\text{meas}} - I_g^{\text{med}})/(I_e^{\text{med}} - I_g^{\text{med}})$.

Photon-number-resolved qubit spectroscopy (PNRQS) datasets were rescaled to account for SPAM errors arising from decay of the transmon $|e\rangle$ state during the long resolved qubit pulse, which is around 8 μ s long. The scaling was calibrated using PNRQS of the $|0\rangle$ state of the cavity, scaling the measured transmon population to 1.

Datasets for the π_{ef} and $|f0\rangle-|g1\rangle$ sideband pulse trains, the $|f0\rangle-|g1\rangle$ Rabi data in Fig. S15 and Fig. S8, and the NOON state Ramsey experiment in Fig. S18(f) were rescaled to ensure that the initial population at zero pulses or time was 1. This correction mitigates SPAM errors introduced during the reset of transmon states outside the interaction subspace being measured, which was performed before the final transmon measurement.

The reset was implemented using an $|f0_r\rangle-|g1_r\rangle$ sideband to the readout mode for 6 μ s, leading to decay within the interaction subspace under measurement. Corrections for SPAM errors in Wigner tomography due to this effect are detailed in Section S8.

-
- [1] S. Chakram, A. E. Oriani, R. K. Naik, A. V. Dixit, K. He, A. Agrawal, H. Kwon, and D. I. Schuster, *Physical Review Letters* **127**, 107701 (2021).
 - [2] M. J. Reagor, *Superconducting cavities for circuit quantum electrodynamics* (Yale University, 2016).
 - [3] M. Bal, A. A. Murthy, S. Zhu, F. Crisa, X. You, Z. Huang, T. Roy, J. Lee, D. v. Zanten, R. Pilipenko, *et al.*, *npj Quantum Information* **10**, 43 (2024).
 - [4] L. Stefanazzi, K. Treptow, N. Wilcer, C. Stoughton, C. Bradford, S. Uemura, S. Zorzetti, S. Montella, G. Cancelo, S. Sussman, *et al.*, *Review of Scientific Instruments* **93**, <https://doi.org/10.1063/5.0076249> (2022).
 - [5] S. Chakram, K. He, A. V. Dixit, A. E. Oriani, R. K. Naik, N. Leung, H. Kwon, W.-L. Ma, L. Jiang, and D. I. Schuster, *Nature Physics* **18**, 879 (2022).
 - [6] B.-J. Liu, Y.-Y. Wang, T. Sheffer, and C. Wang, *Phys. Rev. Lett.* **133**, 160602 (2024).
 - [7] Z. K. Mineev, Z. Leghtas, S. O. Mundhada, L. Christakis, I. M. Pop, and M. H. Devoret, *npj Quantum Information* **7**, 1 (2021).
 - [8] C. K. Law and J. H. Eberly, *Phys. Rev. Lett.* **76**, 1055 (1996).
 - [9] A. Clerk and D. W. Utami, *Physical Review A—Atomic, Molecular, and Optical Physics* **75**, 042302 (2007).
 - [10] Y. Lu, A. Maiti, J. W. Garmon, S. Ganjam, Y. Zhang, J. Claes, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, *nature communications* **14**, 5767 (2023).
 - [11] P. Reinhold, *Controlling error-correctable bosonic qubits*, Ph.D. thesis, Yale University (2019).
 - [12] D. T. McClure, H. Paik, L. S. Bishop, M. Steffen, J. M. Chow, and J. M. Gambetta, *Physical Review Applied* **5**, 011001 (2016).
 - [13] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, *Physical Review A—Atomic, Molecular, and Optical Physics* **76**, 042319 (2007).