Quantum Random Access Memory Architectures Using 3D Superconducting Cavities

D.K. Weiss[1,2,*] Shruti Puri,[1,2] and S.M. Girvin[1,2]

[1]*Departments of Applied Physics and Physics, Yale University, New Haven, Connecticut 06511, USA*

[2]*Yale Quantum Institute, Yale University, New Haven, Connecticut 06511, USA*

Quantum random access memory (QRAM) is a common architecture resource for algorithms with many proposed applications, including quantum chemistry, windowed quantum arithmetic, unstructured search, machine learning, and quantum cryptography. Here, we propose two bucket-brigade QRAM architectures based on high-coherence superconducting resonators, which differ in their realizations of the conditional-routing operations. In the first, we directly construct cavity-controlled controlled-SWAP (CSWAP) operations, while in the second, we utilize the properties of giant-unidirectional emitters (GUEs). For both architectures, we analyze single- and dual-rail implementations of a bosonic qubit. In the single-rail encoding, we can detect first-order ancilla errors, while the dual-rail encoding additionally allows for the detection of photon losses. For parameter regimes of interest, the postselected infidelity of a QRAM query in a dual-rail architecture is nearly an order of magnitude below that of a corresponding query in a single-rail architecture. These findings suggest that dual-rail encodings are particularly attractive as architectures for QRAM devices in the era before fault tolerance.

I. INTRODUCTION

Many quantum algorithms of interest presume the existence of a QRAM device capable of querying a (classical or quantum) memory in superposition. Perhaps the most well-known application of QRAM is as the oracle in Grover's search algorithm [1,2]. Recently, criticisms have emerged of the utility of QRAM in the context of such "big-data" problems [3], particularly for algorithms such as Grover that require a large number of calls to the QRAM. However, quantum algorithms utilizing QRAM where quantum advantage may still exist [3] include modern versions of Shor's algorithm [4,5], quantum chemistry algorithms [6,7], algorithms for solving the dihedral hidden-subgroup problem [8] and the Harrow-Hassidim-Lloyd (HHL) algorithm [9]. To get a sense of the scale of devices relevant for near-term example demonstrations of quantum advantage, the algorithm presented in Ref. [6] for the quantum simulation of jellium in a classically intractable parameter regime requires a modest-size

QRAM with only eight address qubits and a bus (or "word" length) of 13 qubits.

Existing proposals for QRAM are based on quantum optics [10], Rydberg atoms [11], photonics [12], and circuit quantum acoustodynamics [13](for a more comprehensive review, see Ref. [14] and references therein). Each proposal utilizes the celebrated bucket-brigade architecture [15], which promises a degree of noise resilience as compared to more straightforward algorithmic implementations of QRAM [2]. Nevertheless, actually realizing a QRAM device appears to be extremely difficult, due in part to additional concerns regarding the noise sensitivity of active versus inactive components in a bucket-brigade QRAM [16,17]. Recently, Hann *et al.* [18] have helped to address this issue, showing that the bucket-brigade architecture still enjoys a polylogarithmic scaling of the infidelity of a QRAM query with the size of the memory even if all components are active.

In parallel to this theory work, there has been enormous experimental progress in quantum information processing using three-dimensional (3D) superconducting cavities, including demonstrations of millisecond-scale coherence times [19–21], as well as high-fidelity beam-splitter operations [22,23]. Leveraging these results, here we propose bucket-brigade QRAM implementations based on superconducting cavities. Our proposed architectures utilize recently developed midcircuit error-detection and erasure-detection schemes [24,25], boosting the query fidelity

---

*Corresponding author. daniel.weiss@yale.edu

by utilizing postselection. Among the algorithms utilizing QRAM, some may tolerate such a nondeterministic repeat-until-success procedure, as opposed to those which require interleaved QRAM calls. One example is the HHL algorithm, which can utilize QRAM for state preparation [9].

The key feature of bucket-brigade QRAM that enables its relative noise insensitivity is the conditional routing of quantum information based on the states of address qubits [15,18]. In this work, we propose two QRAM architectures that achieve this gate primitive in distinct ways. In the first architecture, we directly construct a cavity-controlled controlled-SWAP (CSWAP) gate. While an ancilla transmon is required to provide the requisite nonlinearity, we can detect and postselect away first-order transmon errors [24–26]. This proposal benefits from requiring no new hardware components beyond what has already been experimentally implemented in Refs. [22,23,27]. Our second approach utilizes the physics of giant unidirectional emitters (GUEs) [28–30] to realize the conditional routing operation, where again first-order transmon errors are detectable. We term the first the "CSWAP architecture" and the second the "GUE architecture." In both architectures, we explore single- and dual-rail [24,25] implementations. The dual-rail approach doubles the hardware cost; however, it additionally allows for the first-order detection of photon loss in the cavities. This boosts the postselected query fidelity F, with, in particular, $F > 0.8$ for a QRAM with eight address qubits in both architectures (in the remainder of this work, all query infidelities are postselected unless noted otherwise).

Our paper is structured as follows. In Sec. II, we review bucket-brigade QRAM. In Secs. III and IV, we explore the CSWAP and GUE architectures, respectively, detailing how each gate primitive is executed and computing the overall query fidelity. We discuss our results and conclude in Sec. V.

II. BUCKET-BRIGADE QRAM

The purpose of a QRAM device is to realize the unitary operation

$$\sum_i \alpha_i |i\rangle |0\rangle \rightarrow \sum_i \alpha_i |i\rangle |D_i\rangle, \qquad (1)$$

where classical data D_i specified by address $|i\rangle$ are accessed in superposition and stored in the state of the bus. [In this work, we take the memory to be a classical database; however, in general, the memory may be quantum (see, e.g., Refs. [3,16,31]).] The first and second registers in Eq. (1) are the n address qubits and bus qubit, respectively, where we query a database of $N = 2^n$ classical bits.

To appreciate the benefits of the bucket-brigade approach to QRAM, it is useful first to review so-called "fan-out" QRAM [2]. Quantum routers are arranged in a treelike structure [see Fig. 1(a)]. The states of the routers at the ℓth level of the tree are set by the ℓth address qubit. The bus qubit is then routed into the tree and arrives at the memory locations specified by the address qubits. After copying the classical data into the bus and routing it back out, the router-state-setting operation is performed again to uncompute and disentangle the address and bus qubits from the routing tree, thus achieving the operation in Eq. (1).

The issue with this proposal is the high susceptibility to decoherence [18]. Each layer of controlled-NOT (CNOT) gates effectively creates a maximally entangled Greenberger-Horne-Zeilinger (GHZ) state among the routers at each level [see Fig. 1(b)]. If any one router decoheres, this collapses the superposition at that level and reduces the query fidelity on average by a factor of 2 [15]. This is clearly not a scalable approach, as there are exponentially many (in n) quantum routers in the bottom layers of the tree.

The bucket-brigade approach aims to achieve the same operation (1) as fan-out QRAM (and, indeed, also uses a tree structure of quantum routers) but with an algorithm that attempts to minimize as much as possible entanglement between different routers [15]. The address qubits (followed by the bus) are fed into the top of the tree one by one and are conditionally routed based on the states of previous address qubits. The resulting state of the routing tree (after all address qubits have been routed in) is such that routers at the same level of the tree are not entangled with one another. Thus, an error on one router does not cause a catastrophic collapse of the state of the QRAM tree [see Fig. 1(c)]. It instead only disrupts queries along branches passing through the decohered router.

The original bucket-brigade proposal called for three-level routers, with states $|W\rangle$, $|0\rangle$, and $|1\rangle$ [15]. The inactive "wait" state $|W\rangle$ acts trivially (identity) for all conditional routing operations [15,18]. It was originally thought [15–17] that the favorable noise properties of bucket-brigade QRAM were predicated on the wait state being decoherence free. However, it was realized in Ref. [18] that the three-level bucket-brigade QRAM is noise resilient even if the wait state is subject to decoherence. This resiliency is due to the limited entanglement among branches as well as the restricted propagation of errors between branches. Roughly speaking, errors in one branch do not eventually poison (via the conditional routing operations) queries in other branches, due to the trivial action of the conditional routing operations on the $|W\rangle$ state. Hann *et al.* prove in Ref. [18] that the query infidelity for a QRAM with three-level routers is bounded by

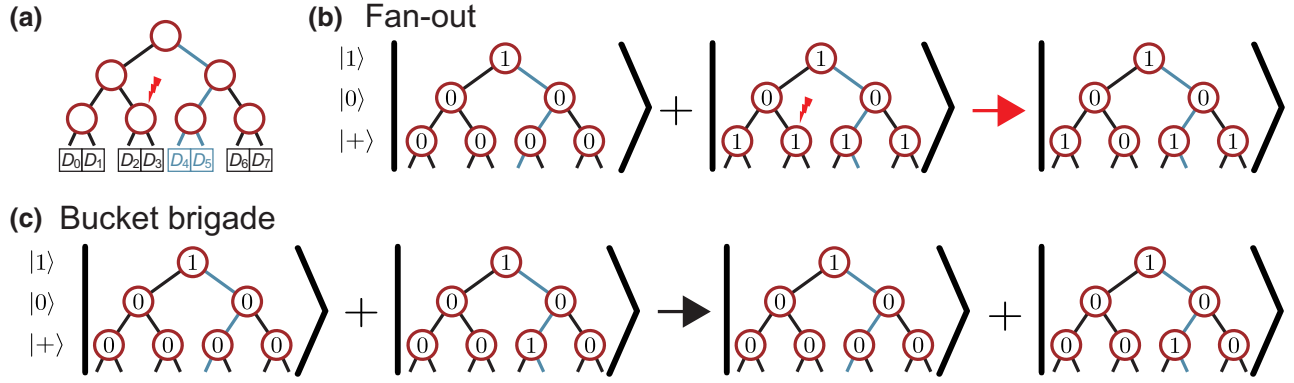$$1 - F \leq A\epsilon N_{ts} \log_2(N), \qquad (2)$$

FIG. 1. A comparison of fan-out and bucket-brigade QRAM with two-level routers. (a) By constructing a tree of quantum routers, one can access data in superposition. We consider the state of the tree in a query to D_4 and D_5 (highlighted) where the indicated router experiences an amplitude-damping error (the general argument also holds for other error types; see Ref. [18]). (b) In a fan-out approach [2], all routers are set at each level according to the state of the address at that level, leading to a GHZ-like state. The resulting state of the QRAM tree is thus highly entangled and susceptible to decoherence. An amplitude-damping event on the indicated router causes the state of the tree to collapse and the query to fail. (c) In the case of bucket-brigade QRAM [15,18], only routers participating in a query to a given branch are activated. This leads to a less entangled state of the QRAM tree as compared to fan-out QRAM. Amplitude damping now cannot occur on the router indicated in (a) and the superposition state is not disrupted.

where $A \approx 4$, ϵ is the error probability per time step, and N_{ts} is the number of time steps in a QRAM query (which we compute below).

Perhaps surprisingly, the noise resilience persists even if two-level routers are utilized (see Fig. 1). Errors can propagate more freely than in the three-level-router case, as conditional routing operations conditioned on router state $|0\rangle$ can cause errors in one branch to interfere with queries to otherwise decoherence-free branches (assuming that the tree is initialized in the vacuum state). However, there is still limited entanglement in the tree and only some errors are allowed to propagate. In this case, the query infidelity is [18]

$$1 - F \leq A \epsilon N_{ts} \log_2(N)[1 + \log_2(N)], \qquad (3)$$

which is still polylogarithmic in the size of the memory. In this work, we consider architectures with two-level routers. We thus utilize Eq. (3) when calculating query infidelities and proceed by calculating the error per time step ϵ.

To realize a bucket-brigade QRAM, three primitive operations are required: (i) setting the state of a router, (ii) conditional routing, and (iii) copying classical data into the state of the bus. We detail how each of these gate primitives is executed for each of our proposals below and analyze the resulting query fidelities.

III. CSWAP ARCHITECTURE

In the CSWAP architecture, quantum information is stored in high-Q superconducting memories that are coupled via beam-splitter elements [22,23,32,33]. In our protocol, the comparatively low-coherence transmons are only excited briefly during gate operations and are disentangled

from the cavities at the conclusion of each gate. Moreover, the transmons are first-order error detected using the techniques described in Refs. [24,25].

A. Gate protocol

Setting the state of a router is performed straightforwardly utilizing SWAP operations [see Fig. 2(b)]. This operation is performed at the top of the tree as well as between output cavities and their neighboring routers [see Fig. 2(e)].

Conditional routing in this architecture is achieved by direct construction of CSWAP operations. It is important to emphasize that in our construction, cavities serve as the controls for conditional routing. This is to be contrasted with previous work that has realized a CSWAP operation using a transmon as the control [33]. Here, transmons are used only as ancillas.

The CSWAP gate is enabled by the joint-parity gate [24, 25,33]

$$\text{JP}_{ab} = |g\rangle\langle g| + \exp(i\pi[\hat{a}^\dagger\hat{a} + \hat{b}^\dagger\hat{b}])|f\rangle\langle f|, \qquad (4)$$

where $\hat{a}$ and $\hat{b}$ are the annihilation operators of the two cavities, referred to as a and b, while $|g\rangle$ and $|f\rangle$ are the two computational states of the transmon. Importantly, the excited state $|e\rangle$ is reserved for first-order error detection of transmon decay [24–26]. We take the transmon to be coupled to cavity a with dispersive Hamiltonian $\frac{\chi}{2}\hat{a}^\dagger\hat{a}\hat{\sigma}_z$, where Pauli matrices are defined in the $|g\rangle, |f\rangle$ manifold and χ denotes the strength of the dispersive interaction. The joint-parity gate is performed by first exciting the ancilla via a Hadamard gate, activating a beam splitter
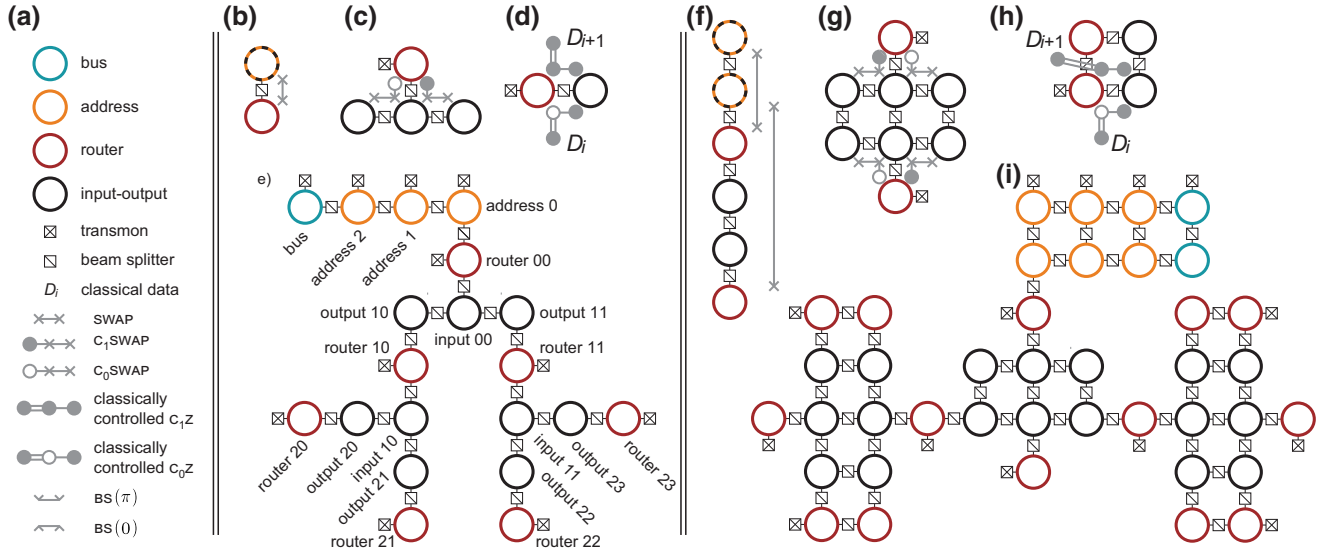
FIG. 2. Single- and dual-rail CSWAP QRAM. (a) We include a legend explaining our coloring scheme, the notation for hardware elements, and gate schematics for this figure as well as later figures. For the single-rail (b)–(e) and dual-rail (f)–(i) implementations, we detail the gates required for (b),(f) setting the state of the router, (c),(g) performing the C_0SWAP and C_1SWAP operations, and (d),(h) copying classical data into the state of the bus. The cavities outlined in dashed orange or black indicate that setting the state of the router occurs both at the top of the tree (involving the address cavity) as well as at intermediate nodes (involving output cavities). We show example hardware layouts for 3-bit (e) single-rail and (i) dual-rail QRAM. The full quantum circuit associated with a QRAM query on the hardware of (e) is shown in Appendix A. The naming scheme for routers and input-output elements is, e.g., router ij, with i corresponding to the layer and j to the position going from left to right. We emphasize that in our proposed bucket-brigade QRAM schemes, the classical data are not stored in any quantum elements before they are copied into the state of the bus. The data instead determine which quantum gates to perform. Thus instead of "memory cells" at the bottom of the tree (which would require additional expensive quantum hardware), we have routers that allow access to data stored at two distinct locations; see (d) and (h).

between cavities a and b for time $2\pi/\chi$, and then applying another Hadamard to the ancilla (for further details, see Refs. [24,25]). Two applications of this gate separated by rotations on the ancilla transmon realizes an entangling operation between the two cavities [24,25]:

$$ZZ_{ab}(\theta) = e^{i\frac{\pi}{4}\hat{\sigma}_y}\mathrm{JP}_{ab}e^{-i\frac{\theta}{2}\hat{\sigma}_x}\mathrm{JP}_{ab}e^{-i\frac{\pi}{4}\hat{\sigma}_y}$$
$$= \cos\frac{\theta}{2}\hat{\mathbb{1}} - i\sin\frac{\theta}{2}e^{i\pi[\hat{a}^\dagger\hat{a}+\hat{b}^\dagger\hat{b}]}\hat{\sigma}_z. \quad (5)$$

If the transmon is detected in $|e\rangle$ or $|f\rangle$ at the conclusion of the gate, we infer that an ancilla decay or dephasing event has occurred, respectively. In the following, we assume that the transmon always begins in its ground state. The effect of this gate on states in the Fock basis is then

$$ZZ_{ab}(\theta)|n_a n_b\rangle = e^{-i\frac{\theta}{2}P_{n_a n_b}}|n_a n_b\rangle, \quad (6)$$

where $P_{ij} = 1, -1$ if the joint parity of the a and b modes is even or odd, respectively. For $\theta = \pi/2$, this gate becomes a CZ_{ab} operation in the computational subspace, up to the single-cavity operations $e^{-i\frac{\pi}{2}\hat{a}^\dagger\hat{a}}e^{-i\frac{\pi}{2}\hat{b}^\dagger\hat{b}}$ and a global phase

[24,25,34]

$$CZ_{ab}|n_a n_b\rangle \equiv e^{i\pi/4}e^{-i\frac{\pi}{2}(\hat{a}^\dagger\hat{a}+\hat{b}^\dagger\hat{b})}ZZ_{ab}(\pi/2)|n_a n_b\rangle$$
$$= e^{-i\frac{\pi}{2}\left[\frac{P_{n_a n_b}}{2}+n_a+n_b-\frac{1}{2}\right]}|n_a n_b\rangle. \quad (7)$$

To utilize this gate toward construction of a CSWAP operation, we take inspiration from the canonical construction of CSWAP [2,35]. This construction sandwiches a Kerr interaction between two cavities, a and b, by 50:50 beam splitters between cavity b and a third cavity c. The Kerr interaction (synthesized here via the CZ_{ab} operation) acts as a phase shifter that either completes or undoes the beam splitter. In this way, we obtain both a C_0SWAP and a C_1SWAP:

$$C_0\mathrm{SWAP}_{abc} = e^{i\pi\hat{b}^\dagger\hat{b}}\mathrm{BS}_{bc}(\pi/2)CZ_{ab}\mathrm{BS}_{bc}(\pi/2),$$
$$C_1\mathrm{SWAP}_{abc} = \mathrm{BS}_{bc}^\dagger(\pi/2)CZ_{ab}\mathrm{BS}_{bc}(\pi/2), \quad (8)$$

where C_0SWAP and C_1SWAP indicate the SWAP is executed if the control is set to $|0\rangle$ or $|1\rangle$, respectively. In this way, we achieve the conditional routing of quantum information

[see Fig. 2(c)]. We have defined

$$\text{BS}_{bc}(\alpha) = \exp\left(-i\frac{\pi}{4}[e^{i\alpha}\hat{c}^\dagger\hat{b} + e^{-i\alpha}\hat{b}^\dagger\hat{c}]\right); \quad (9)$$

thus

$$\text{BS}_{bc}(\pi/2) = \exp\left(\frac{\pi}{4}[\hat{c}^\dagger\hat{b} - \hat{b}^\dagger\hat{c}]\right). \quad (10)$$

The additional single-cavity rotation in C_0SWAP is to correct for unwanted phases. See Appendix B for an explicit verification that the gate functions as intended on the states of interest $|n_a00\rangle$, $|n_a01\rangle$, and $|n_a10\rangle$, $n_a = 0, 1$, where the third index refers to occupation in mode c. We note that these CSWAP gates do not behave as expected if modes b and c are *both* initially occupied. We might expect that the overall operation should be trivial. However, due both to the Hong-Ou-Mandel (HOM) effect [36] and to phase shifts due to the CZ_{ab} gate, population is transferred out of the computational subspace. We stress that in the absence of thermal photons, we never expect these modes to be simultaneously occupied in the course of a QRAM query. Thus all population should ideally remain in the computational subspace [37]. Nevertheless, both modes may become occupied due to thermal excitations; thus we discuss this case in Appendix B.

These operations thus realize logical CSWAP gates in the single-rail case. In the dual-rail case, the logical states are given in terms of the cavity Fock states as $|0_\text{L}\rangle = |10\rangle, |1_\text{L}\rangle = |01\rangle$. Logical CSWAP operations in the dual-rail case are then realized by performing C_0SWAPs and C_1SWAPs on both halves of the dual-rail qubits [see Fig. 2(g)] (taking the cavity with the first index to be the top cavity). The two physical gates making up a single logical C_0SWAP may be performed in parallel; thus the query time is not increased by utilizing a dual-rail architecture. We note here that in the dual-rail case, it is necessary to initialize the QRAM tree in the vacuum state, as opposed to, e.g., initializing each dual rail in $|0_\text{L}\rangle$ (in the single-rail case, we also initialize each cavity in vacuum; however, those are logical states). This initialization choice is due again to the HOM effect discussed above. Initialization of all dual rails in $|0_\text{L}\rangle$ would lead to cases in which both target cavities of a CSWAP are occupied, leading to leakage out of the logical subspace.

The final gate-primitive necessary is the operation for copying classical data into the bus. Following Ref. [18], we route the bus into the tree in the state $|+_\text{L}\rangle$. In the single-rail case, this is the state $(|0\rangle + |1\rangle)/\sqrt{2}$, while in the dual-rail case, this is the state $(|10\rangle + |01\rangle)/\sqrt{2}$. Once the bus reaches the bottom layer of the tree, we perform classically controlled C_0Z and C_1Z operations between the bus and the router as shown in Fig. 2(d), where the subscript refers to conditioning on the state of the router. By definition, we have $\text{C}_1\text{Z} \equiv \text{CZ}$. The C_0Z gate is compiled as

a CZ gate followed by a Z gate on the bus (performed in software), such that overall a Z gate is applied to the bus only if the router is in the state $|0_\text{L}\rangle$. The additional layer of classical control simply means that we execute the C_0Z and C_1Z gates only if the classical data are set to 1. In both the single- and dual-rail cases [see Figs. 2(d) and 2(h)], the gate operations yield

$$[|0_\text{L}\rangle + |1_\text{L}\rangle][\alpha|0_\text{L}\rangle + \beta|1_\text{L}\rangle]$$
$$\rightarrow \alpha[|0_\text{L}\rangle + (-1)^{D_i}|1_\text{L}\rangle]|0_\text{L}\rangle$$
$$+ \beta[|0_\text{L}\rangle + (-1)^{D_{i+1}}|1_\text{L}\rangle]|1_\text{L}\rangle, \quad (11)$$

ordering the states as bus, router, and where $|i_\text{L}\rangle \equiv |i\rangle$ in the single-rail case. Here, D_i and D_{i+1} are the classical data accessed by the router at the bottom of the tree. This scheme satisfies the "no-extra copying" condition [18], where the action of the gate is trivial for any locations where the bus has not been sent.

We emphasize that the error detection of ancilla errors is performed after every CZ and CSWAP gate (see Appendix C). No "which-path" information is revealed by performing measurements on the ancillas, as they are disentangled from the tree after every gate and do not collapse the superposition of queries. This is to be contrasted with error detection of cavity photon-loss events in the dual-rail case, which must wait until the address and bus qubits have been routed out of the tree. We cannot perform midcircuit cavity-photon loss detection because we initialize the tree in the vacuum state, as opposed to dual-rail logical states. Thus, such error detection would reveal which-path information and collapse the superposition state of the tree. For example, detecting that a photon *is present* in a router (indicating occupation of a dual-rail logical state) destroys those terms in the superposition state of the tree where this router was in vacuum. If we hypothetically had access to a CSWAP operation that performed as expected when both target modes were occupied, then we could initialize the tree in logical dual-rail states and perform midcircuit error detection of cavity photon loss. Indeed, detecting cavity photon-loss errors becomes important in the large-n limit, when query times approach cavity lifetimes T_1^c. Thus, constructing such a CSWAP operation is an interesting avenue of further research.

B. Resource estimates

We estimate the hardware cost of our proposed implementations by counting the number of cavities, as the number of required beam-splitter elements and ancilla transmons scale proportionally. We require

$$N_\text{cav}^\text{SR} = \frac{5}{2}N + \log_2(N) - 3, \quad N_\text{cav}^\text{DR} = 2N_\text{cav}^\text{SR}, \quad (12)$$

cavities in the single- and dual-rail case, respectively. N is the size of the memory and is related to the number n

of address qubits by $N = 2^n$. Here and in the following, unless otherwise stated, we assume that $n \geq 2$ and $N \geq 4$.

We now turn to estimating the number of gates. Specifically, we estimate the number of CZ gates, which underlies both the CSWAP and the data-copying operations. We ignore the number of required SWAP (or 50:50 beam-splitter) operations, as these gates are fast compared to the CZ gate execution time: a SWAP can be executed in as little as 50–100 ns [22,23], while a CZ gate takes time $4\pi/\chi$ (ignoring the single-qubit gate times on the transmon) [24,25]. The dispersive coupling strength χ between the cavity and ancilla transmon is typically on the order of one to a few megahertz [38], taken here and in the following to be $\chi/2\pi = 2$ MHz, corresponding to a CZ gate time of $t_{CZ} = 1$ μs. In the single-rail case, the total number of required CZ gates is

$$N_{\text{gates}} = 7N - 4\log_2(N) - 8. \tag{13}$$

In the dual-rail case, the CZ operations necessary for the logical CSWAP gates are executed in parallel. Thus counting the logical operation as a single gate, the total number of logical CZ operations is the same as in the single-rail case (counting the data-copy operations also as logical CZs, though only a single physical CZ is required [see Fig. 2(h)]).

While there are $\mathcal{O}(N)$ gates, the circuit depth scales only as $\mathcal{O}(\log_2[N])$. This is because many of the gates may be executed in parallel: all of the gates at a single horizontal layer of the QRAM may be executed simultaneously. Moreover, we may utilize *address pipelining* [3,39]; once an address qubit has been routed past the first set of output ports, the next address (or bus) qubit may be routed in. Thus the total number of time steps scales only logarithmically in N,

$$N_{\text{ts}} = 4(1 + 2 + 3(n - 3) + 2) + 2, \quad n \geq 3,$$
$$= 12\log_2(N) - 14, \tag{14}$$

and $N_{\text{ts}} = 10$ for $n = 2$. This number is the same for both single rail and dual rail. The factor of 4 in Eq. (14) is two factors of 2, coming from the need to route qubits both in and out, and the need to perform C_0SWAP as well as C_1SWAP. The second address qubit needs to traverse one level of the tree and the third address needs to traverse two levels. Address qubits routed further down the tree may be pipelined and each introduces only a constant factor of additional time steps. The final factor inside of the parentheses accounts for routing the bus, while the final factor in Eq. (14) corresponds to the data-copying steps.

The time to complete a QRAM query is then $t_n = t_{CZ}N_{\text{ts}}$. For the parameters considered in this work, t_n is generally short when compared with T_1^c. For instance, for $n = 12$ and $N = 2^{12}$, we obtain $t_{12} = 130$ μs.

C. Infidelity and error mitigation by postselection

We now turn to estimating the overall infidelity and no-flag probability (the probability of detecting no errors) of a QRAM query. In both the dual-rail and single-rail cases, we utilize the infidelity formula for two-level routers (as opposed to that for three-level routers). This is clear for the single-rail case (as we only utilize Fock states $|0\rangle$ and $|1\rangle$); however, for the dual-rail case, it might appear at first glance that we have indeed implemented a three-level router [and thus can utilize the more favorable infidelity formula (2)]. The physical states $|10\rangle$ and $|01\rangle$ encode logical $|0_L\rangle$ and $|1_L\rangle$, respectively, and the state $|00\rangle$ could then play the role of the logical wait state $|W_L\rangle$ [15,18]. This reasoning is incorrect because we construct the logical CSWAP by utilizing physical C_0SWAP and C_1SWAP operations. On the one hand, errors in target cavities subject to the C_1SWAP do indeed get stuck if the router is in the state $|00\rangle$. On the other hand, errors in target cavities subject to the C_0SWAP can propagate up the tree with the router in the state $|00\rangle$. Thus, we expect the infidelity scaling for two-level routers of Ref. [18] to apply to the dual-rail case as well [40].

We take the probability of error per time step to be $\epsilon = 1 - F_g(\text{CZ})$ [12,18], consistent with neglecting sources of error due to beam-splitter operations. We have defined $F_g(\text{CZ})$, the gate fidelity of the CZ gate. We expect $F_g(\text{CZ})$ to be limited by decoherence and measurement errors (when utilizing postselection), as in the absence of these nonidealities, Eq. (7) realizes a perfect CZ gate. We calculate $F_g(\text{CZ}), \epsilon$ both with and without postselection on first-order errors (for details, see Appendix C) using three sets of parameters (see Table I and Fig. 3). Parameter set 1 (PS1) is based on values recently reported in the literature for a combined transmon and 3D resonator package [42]. For parameter set 2 (PS2), we use state-of-the-art coherence times for transmons [43,44] and 3D resonators [19]. In parameter set 3 (PS3), we utilize the same cavity coherence times as PS2 but make more optimistic assumptions for the transmon coherence times. We additionally include the detrimental effects of measurement errors when calculating $F_g(\text{CZ}), \epsilon$. Such effects occur at second order and thus at the same order as, e.g., two uncaught transmon errors. They are second order because first an error must occur and then it must be misidentified (for more details, see Appendix C).

Without utilizing postselection on first-order transmon errors, the gate infidelity is on the order of $\epsilon \sim 10^{-2}$–10^{-3} for both single and dual rail (see Fig. 3). However, postselected infidelities on the order of 10^{-5} and 10^{-6} are possible for single rail and dual rail, respectively, utilizing estimates from parameter sets 2 and 3.

These low error rates enable high-fidelity QRAM queries [see Fig. 4(a)]. Comparing dual rail and single rail, a dual-rail approach yields nearly an order-of-magnitude
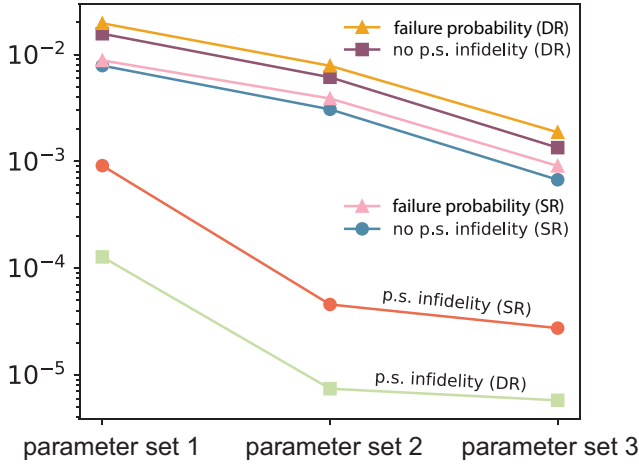
FIG. 3. The infidelity, the postselected (p.s.) infidelity, and the failure probability of the logical CZ operation as a function of the parameter set for both single rail (SR) and dual rail (DR). Utilizing postselection and parameter sets 2 and 3, infidelities on the order of 10^{-5} and 10^{-6} are possible for single and dual rail, respectively.

improvement in postselected infidelity over a single-rail implementation for a given n and parameter set (due to improved infidelities; see Fig. 3). From the perspective of instead setting a desired infidelity of $1 - F < 0.2$ and using PS2 estimates, memories of size $N = 2^4$ and 2^8 can be queried with single- and dual-rail implementations, respectively.

The price to pay for utilizing postselection is the rejection of a large fraction of shots. We thus calculate the "no-flag" probability P_{CZ}, the probability that no error is detected during a single CZ gate (of course, this is not the no-error probability, as two or more transmon errors can go undetected; see Fig. 3 and Appendix C). For an entire QRAM query to succeed, in the worst case all CZ gates

TABLE I. The parameters used in the numerical simulations and the associated jump operators. The decay and dephasing rates of the cavities are $\Gamma_1^c = 1/T_1^c$ and $\Gamma_\phi^c = 1/T_\phi^c$, respectively, implicitly defining the associated coherence times. For the transmon, the decay and dephasing rates are generally state dependent [41]. We make the usual assumption of bosonic enhancement of these rates [24], taking $\Gamma_\phi^{t,gg} = 0$, $\Gamma_\phi^{t,ff} = 4\Gamma_\phi^{t,ee}$, and $\Gamma_1^{t,ef} = 2\Gamma_1^{t,ge}$. The associated coherence times are defined as, e.g., $T_1^{t,ef} = 1/\Gamma_1^{t,ef}$. Finally, we assume a thermal population of $n_{\text{th}} = 0.01$.

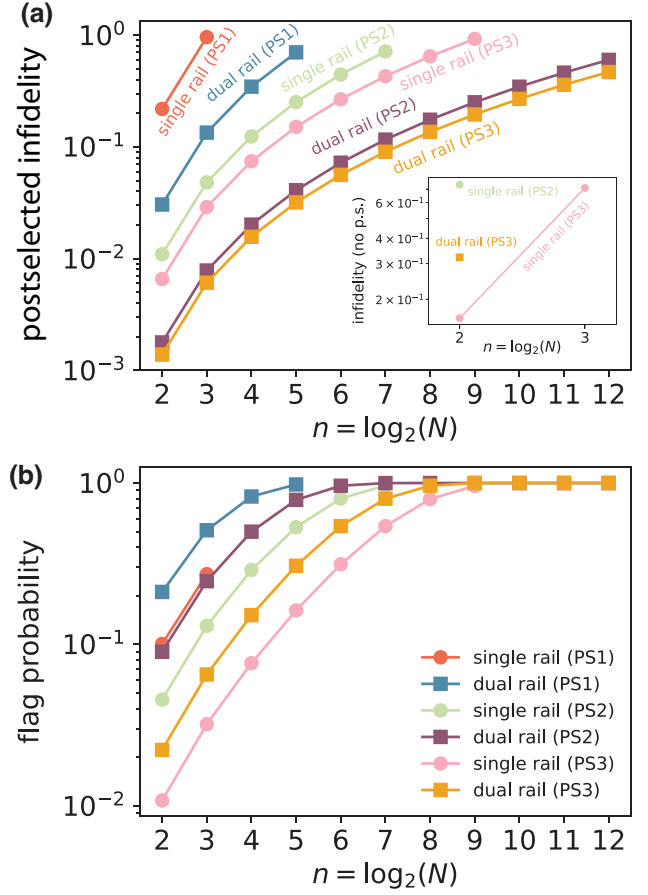|	jump operator	PS1 [42]	PS2	PS3
T_1^c	$\hat{a}$	0.6 ms	25 ms [19]	25 ms
T_ϕ^c	$\hat{a}^\dagger \hat{a}$	5 ms	106 ms [19]	106 ms
$T_1^{t,ge}$	$|g\rangle\langle e|$	0.2 ms	0.5 ms [43]	2 ms
$T_1^{t,ee}$	$|e\rangle\langle e|$	0.4 ms	0.9 ms [43]	4 ms

FIG. 4. The overall infidelity and flag probability of a QRAM query as a function of n. (a) For the three parameter sets shown in Table I, we calculate the postselected infidelity for both the single- and dual-rail architectures (showing only values for which the fidelity is nonzero). In the inset, we plot the fidelity without utilizing postselection. The dual-rail implementation enjoys a decrease in the postselected infidelity with respect to the single-rail case. (b) This comes at the price of increasing the flag probability.

must not flag an error:

$$P_{\text{no flag}} = (P_{\text{CZ}})^{N_{\text{gates}}}, \quad (15)$$

with $P_{\text{flag}} = 1 - P_{\text{no flag}}$. The expected time to obtain a successful query is then

$$T_{\text{success}} = t_{\text{CZ}} N_{\text{ts}} / P_{\text{no flag}}, \quad (16)$$

where $t_{\text{CZ}} N_{\text{ts}}$ is the time per query and $1/P_{\text{no flag}}$ is the expected number of trials until success. The success rate is then $\Gamma_{\text{success}} = 1/T_{\text{success}}$. In the dual-rail case, $P_{\text{no flag}}$ includes the probability of a photon-loss event for the address or bus qubits during the course of a query.

The flag probability is higher for dual rail as compared to single rail for a given parameter set, due to the ability to detect cavity photon-loss events [see Fig. 4(b)]. This

is compensated by the decreased infidelity for dual rail as compared to single rail [see Fig. 4(a)]. For all parameter sets under consideration, the flag probability approaches unity for $n \geq 9$. Again considering PS2 estimates, in the single-rail case for $n = 4$, we obtain $P_{\text{no flag}} = 0.71$ (and $\Gamma_{\text{success}} = 20.9 \text{ kHz}$), indicating that the majority of queries are expected to succeed. On the other hand, in the dual-rail case for $n = 8$, we obtain $P_{\text{no flag}} = 1.0 \times 10^{-6}$ and $\Gamma_{\text{success}} = 0.013 \text{ Hz}$. In this case, the vast majority of queries are expected to fail; however, the fidelity of those that succeed should exceed $F > 0.8$.

It is important to note that our calculated flag probability and success rates are overly pessimistic. In our estimate of P_{flag}, we treat errors at all router locations on equal footing by essentially setting the fidelity to zero in each case. However, a CZ gate failing at the top of the tree is more problematic than one failing at the bottom. This is because queries to all branches through a router that has experienced a gate error are liable to fail [18]. Thus an error on the top router causes all queries to fail, while an error on a router at the bottom of the tree may only affect queries to that branch (with the necessary caveat that with two-level routers, errors at the bottom of the tree may propagate upward and into other branches, as discussed in Sec. II and in Ref. [18]). Therefore, in practice, CZ gate failures near the bottom of the tree may be tolerated with only a mild reduction in fidelity. Investigation of possible trade-offs between query fidelity and flag probability is beyond the scope of this work and left for future research.

We also estimate the fidelity of a QRAM query without utilizing postselection [see the inset of Fig. 4(a)]. High-fidelity queries are only possible for memories of size $2^2 = 4$, even for our most aggressive coherence time estimates PS3. The reported infidelities are worse than what one might expect from a naive estimate based on the flag probabilities: e.g., in the case of single rail and PS2, the flag probability is only 0.05 but the nonpostselected infidelity is $1 - F = 0.26$. The issue is that for small values of N, finite-size effects defeat the favorable asymptotic of $\log_2^3(N)$. Observe that for $n \leq 9$, we have $N_{\text{gates}} < 4N_{\text{ts}}\log_2(N)(1 + \log_2(N))$, despite the linear scaling of N_{gates} with N.

IV. QUANTUM ROUTER VIA DIRECTIONAL PHOTON EMISSION

Recent theoretical [28,29] and experimental [30] works have investigated *giant unidirectional emitters* (GUEs) for use in quantum networks. By coupling two qubits (or cavities) with frequency ω to a waveguide and spatially separating them by $\lambda/4$, the composite system can be made to emit to the left or right by preparing the system in the state $|\psi_L\rangle = (|01\rangle - i|10\rangle)/\sqrt{2}$ or $|\psi_R\rangle = (|01\rangle + i|10\rangle)/\sqrt{2}$, respectively [28–30]. Here, λ is the wavelength

of the emitted photon. We show below that by controlling the direction of emission based on the state of a router (and catching the emitted photon downstream), we implement simultaneous conditional routing operations. This obviates the need that we had in the previous model to do these operations serially.

The simultaneous conditional routing operations are enabled by a pitch-and-catch protocol for quantum state transfer between two GUEs detailed here. In a single-rail architecture composed of one GUE (two cavities) at each node, quantum information is routed down the tree in the superposition state $c_0|00\rangle + c_1|\psi_{R/L}\rangle$. It is necessary to include the state $|00\rangle$ because we require two states to be routed together that act as a qubit (and the vacuum state is trivially routed in our pitch-and-catch protocol). Of course, a photon-loss event from the state $|\psi_{R,L}\rangle$ yields the state $|00\rangle$, amounting to a logical error. This motivates the use of a dual-rail architecture, which utilizes a second pair of GUEs at each send-and-receive node. The quantum information is then encoded in the superposition state $c_0|\psi_{R/L}\rangle|00\rangle + c_1|00\rangle|\psi_{R/L}\rangle$. The error state $|00\rangle|00\rangle$ is outside of the code space and photon-loss events can be detected as in the dual-rail CSWAP architecture.

A. Circuit design and state transfer

We couple the data cavities indirectly to the waveguide via frequency-converting beam-splitter elements coupled to transfer resonators (see Fig. 5). This coupling architecture has two main advantages. First, frequency-converting beam splitters allow us to ensure that the emitted photon packets are indistinguishable in frequency, despite the difference in frequency between the data cavities. Previous work [29] has shown that the directionality properties of a GUE are highly sensitive to frequency differences between the emitted photons. Second, this architecture allows for an effectively *tunable* coupling strength between the data cavities and the waveguide. By turning the coupling strength off, we prevent immediate emission into the waveguide, allowing time for quantum information processing at each node. Once we are prepared to emit a wave packet into the waveguide, we tune the coupling strength to control the shape of the emitted wave packet. By designing the control pulse to emit a time-symmetric wave packet from a sender GUE, a time-reversed pulse on the receiver GUE absorbs the incident wave packet [45–48]. We obtain analytical pulse shapes for state transfer in this architecture by adiabatically eliminating the transfer resonators (see Appendix D). We additionally consider how nonidealities affect the fidelity of state transfer in Appendix E. For PS2 estimates, we predict state-transfer infidelities of 3.3×10^{-4}, 4.0×10^{-6} for single rail and dual rail, respectively (see Table II).
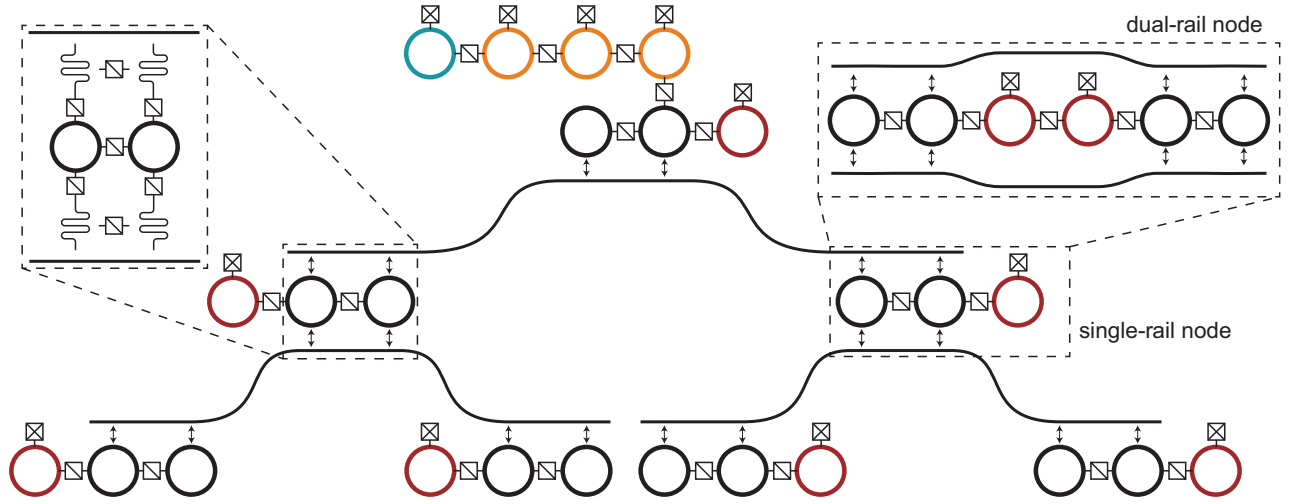
FIG. 5. The 3-bit QRAM implemented using GUEs. We show explicitly the single-rail architecture, with the right inset detailing the structure of an internal node in the dual-rail case. The left inset indicates the coupling mechanism of the data cavities to the waveguide via transfer resonators.

B. GUE-based protocol

We now describe each step of the proposed GUE-based QRAM protocol. As with the CSWAP protocol, to perform a bucket-brigade QRAM query we must implement three primitive operations: (i) setting the state of a router, (ii) conditional routing, and (iii) encoding classical data in the state of the bus. These operations are shown schematically in Fig. 6 for the single- and dual-rail cases. The operations are similar to those used in the CSWAP architecture, with a few differences.

In the case of (i) router-state setting, let us consider a GUE that has just absorbed an incoming wave packet traveling to the right. The state of the GUE-plus-router system is $|\psi\rangle = (\alpha|00\rangle + \beta|\psi_R\rangle)|0\rangle$, ordering the states as a, b, c as labeled in Fig. 6(a). A 50:50 beam splitter $\mathrm{BS}_{ab}(\pi)$ followed by a single-cavity rotation (done in software) places the address state in cavity b that is nearest the router:

$$\exp\left(-i\frac{\pi}{2}\hat{b}^{\dagger}\hat{b}\right)\mathrm{BS}_{ab}(\pi)|\psi\rangle = |0\rangle(\alpha|0\rangle + \beta|1\rangle)|0\rangle. \quad (17)$$

TABLE II. The coherence times (in addition to those used in Table I) used for numerical simulation, infidelities, and failure probabilities for state transfer. Transmon coherence times are not relevant here; thus we only consider PS1 and PS2.

	PS1	PS2
$T_{1,\mathrm{nr}}^{\mathrm{tran}}$	$100\,\mu\mathrm{s}$	$200\,\mu\mathrm{s}$
T_{ϕ}^{tran}	$100\,\mu\mathrm{s}$	$200\,\mu\mathrm{s}$
$1 - F_{\mathrm{st}}$, SR	1.6×10^{-3}	3.3×10^{-4}
$1 - F_{\mathrm{st}}$, DR (p.s.)	8.3×10^{-5}	4.0×10^{-6}
$1 - P_{\mathrm{st}}$, DR	2.8×10^{-3}	6.1×10^{-4}

Finally, a SWAP_{bc} operation places the address state in the router. These gates (omitting the single-cavity rotation) are shown in Fig. 6(a). The generalization to the dual-rail case is straightforward and is shown in Fig. 6(d).

For (ii) conditional routing, we again utilize a phase shift (the logical CZ, performed as in Sec. III for both dual and single rail) to determine the direction in which to send quantum information. Now, the purpose of the CZ operation is to convert, e.g., a left-emitting state into a right-emitting state (and vice versa) conditioned on the state of the router [49]. In the single-rail case, the operation
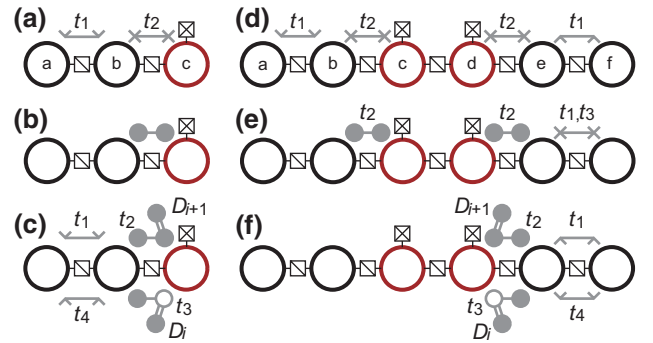

FIG. 6. QRAM operations using GUEs. We have omitted the waveguide and transfer resonators for clarity. The three required primitives for bucket-brigade QRAM are (a),(d) setting the state of the router, (b),(e) controlled routing (the following state transfer into the waveguide is not shown) and (c),(f) copying classical data onto the bus, in the cases of single rail and dual rail, respectively. The gates are labeled $t_1, t_2, \ldots$ to indicate their temporal ordering. The gate schematics and notation utilized here are explained in Fig. 2(a).

shown in Fig. 6(b) yields

$$
\begin{aligned}
&\text{CZ}_{bc}[\alpha|00\rangle + \beta|\psi_{R/L}\rangle][\gamma|0\rangle + \delta|1\rangle] \\
&= \gamma[\alpha|00\rangle + \beta|\psi_{R/L}\rangle]|0\rangle + \delta[\alpha|00\rangle + \beta|\psi_{L/R}\rangle]|1\rangle.
\end{aligned}
\tag{18}
$$

Thus, the qubit state is appropriately entangled with the state of the router and emitted right or left accordingly (after the state-transfer protocol is performed). The generalization to the dual-rail case [see Fig. 6(e)] requires CZ operations on both rails as expected, with additional SWAP operations and a single-cavity rotation to correct for unwanted phases:

$$
\begin{aligned}
&\exp(i\pi\hat{f}^{\dagger}\hat{f})\text{SWAP}_{ef}\,\text{CZ}_{bc}\text{CZ}_{de}\text{SWAP}_{ef} \\
&\quad \times [\alpha|\psi_{R/L}\rangle|00\rangle + \beta|00\rangle|\psi_{R/L}\rangle][\gamma|10\rangle + \delta|01\rangle] \\
&= \gamma(\alpha|\psi_{L/R}\rangle|00\rangle + \beta|00\rangle|\psi_{L/R}\rangle)|10\rangle \\
&\quad + \delta(\alpha|\psi_{R/L}\rangle|00\rangle + \beta|00\rangle|\psi_{R/L}\rangle)|01\rangle,
\end{aligned}
\tag{19}
$$

ordering the states a, b, e, f, c, and d [as labeled in Fig. 6(d)] such that the router states are on the right.

With respect to (iii) data copying, the protocol is as in Sec. III with the modification in (i) that a 50:50 beam splitter and single-cavity rotation are necessary to place the state of the bus in the cavity nearest the router [see Figs. 6(c) and 6(f)].

C. Resource estimates

We estimate the hardware cost of the GUE QRAM architecture by counting the number of high-Q resonators, as before. A simple counting argument yields

$$
\begin{aligned}
N_{\text{cav}}^{\text{SR}} &= 3N + \log_2(N) - 2, \\
N_{\text{cav}}^{\text{DR}} &= 2N_{\text{cav}}^{\text{SR}}.
\end{aligned}
\tag{20}
$$

To estimate the number of gates, we again ignore SWAP or 50:50 beam-splitter operations between high-Q cavities, focusing on the more costly state transfer and CZ gates. We lump together the (logical) CZ [see Figs. 6(b) and 6(e)] and following state transfer as one "operation," yielding the total number of gates,

$$
N_{\text{gates}} = 4N - 2\log_2(N) - 4,
\tag{21}
$$

valid both for single and dual rail (this number includes the N data-copying operations, which do not themselves require a state transfer). The total number of time steps in both cases is

$$
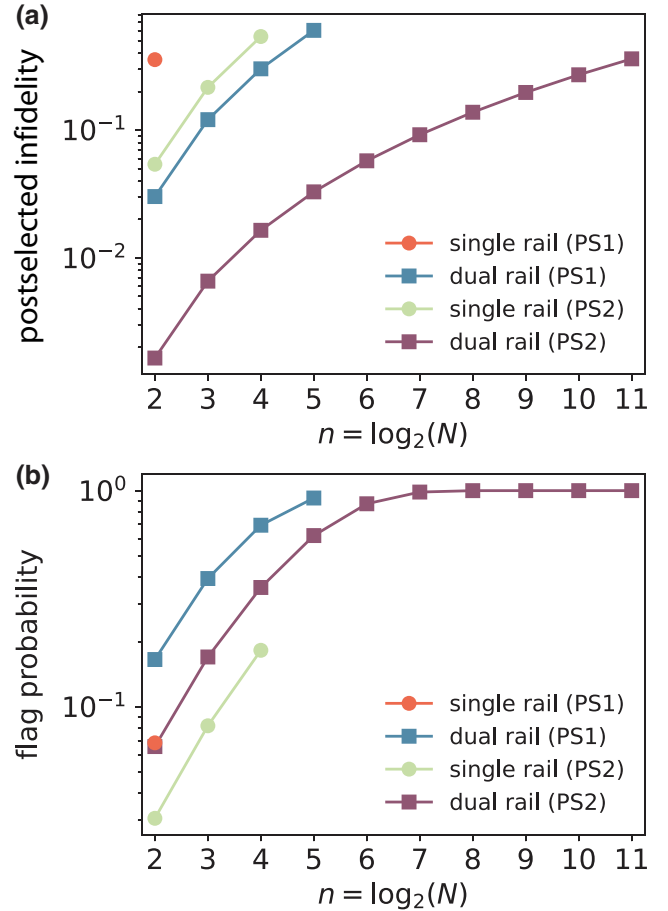N_{\text{ts}} = 6\log_2(N) - 6.
\tag{22}
$$

FIG. 7. The postselected infidelity and failure probability of a QRAM query using GUEs. (a) In the single-rail case, a high-fidelity query is only possible for relatively small n limited by the infidelity of state transfer. In the dual-rail case, we may postselect for photon loss enabling higher-fidelity queries for larger values of n. (b) The flag probability in the single-rail case is due to the ability to postselect away ancilla errors during the CZ operation preceding the state-transfer protocol.

D. Infidelity and error mitigation by postselection

The error probability associated with each time step is $\epsilon = 1 - F_g(\text{CZ})F_{\text{st}}$, grouping the router-controlled operations and following state transfer together, and defining the fidelity of state transfer F_{st}. This expression for the error probability assumes that errors are uncorrelated between the two operations. This may not be exactly true in practice; however, it is likely a good leading-order approximation and it significantly reduces the computational cost of calculating the error probability. We additionally define the no-flag probability

$$
P_{\text{no flag}} = (P_{\text{CZ}}P_{\text{st}})^{N_{\text{gates}}},
\tag{23}
$$

where P_{st} is the success probability for the state-transfer procedure. P_{st} is unity in the single-rail case; however, it

takes nontrivial values in the dual-rail case because photon loss is detectable. Details on the calculation of F_{st} and P_{st} are provided in Appendix D. To calculate the overall infidelity of a QRAM query, we again use Eq. (3) (see Fig. 7).

For PS2 estimates, in the single-rail case only relatively small-scale devices are predicted to yield high fidelities. The issue is that for the state-transfer protocol considered here, we predict state-transfer infidelities on the order of 10^{-4}. This contribution to the overall infidelity overwhelms that due to the CZ operation, which has a predicted postselected infidelity on the order of 10^{-5} for PS2 estimates (see Fig. 3).

The situation is quite different in the dual-rail case, with postselected query fidelities of $F \geq 0.9$ possible for $n \leq 6$ and using PS2 estimates. In the case of $n = 8$, we obtain $F > 0.8$, $P_{no\ flag} = 2 \times 10^{-4}$, and $\Gamma_{success} = 2.2$ Hz. The difference now is that photon loss during the state-transfer procedure is detectable, as the vacuum state is no longer a logical state. The price to be paid is the increased flag probability of dual rail with respect to single rail (for a given parameter set) [see Fig. 7(b)].

V. DISCUSSION AND CONCLUSIONS

We have explored two approaches toward realizing QRAM using superconducting circuits. The first relies on the direct construction of C_0SWAP and C_1SWAP gates between three superconducting 3D cavities. The only gate primitive required (aside from cavity SWAP gates and beam splitters) is the ZZ$(\pi/2)$ gate, which underlies both the conditional routing and the data-copying operations. Such a limited required gate set could significantly alleviate control and calibration overhead in a potential experimental realization. Additionally, this architecture has the advantage of requiring no new hardware beyond what has already been experimentally realized in, e.g., Refs. [22,23,27]. The experimentally feasible architecture and small gate set stand in contrast to previous QRAM proposals, which either still require the experimental implementation of elementary hardware components [10,11,13,50], or utilizes components that have not yet been demonstrated together [12].

The second architecture introduces GUEs, which allow for the simultaneous conditional routing of quantum information down the tree in both directions (whereas in the CSWAP architecture, the routing operations must be performed serially). The coupling of 3D cavities to a waveguide in the manner described in this work has yet to be realized experimentally. However, the tunable coupling of two-dimensional (2D) transmons to a waveguide has been experimentally demonstrated [30]. Many of our results carry over immediately to this case, most notably the state-transfer protocol, which is carried out in the single-excitation manifold. This platform could then implement

the QRAM protocol described in this work, at the cost of using relatively low-coherence transmons as routers.

It is worth emphasizing that in both architectures, the classical data are not stored in any quantum object before they are copied into the state of the bus. Instead, the classical data determine which gates are performed during the QRAM query. This simplification results in quantum hardware savings, and ensures that we need not be concerned with decoherence in "memory cell qubits" [51], which do not exist in our proposal.

On the one hand, the GUE architecture generally enjoys a success-rate advantage over the CSWAP architecture. This is due to the fact that the lossy ancilla is only meaningfully excited during the single CZ gate preceding the state-transfer protocol and not during the state transfer itself. This is to be contrasted with the CSWAP case, where two CZ gates are required at each node to perform the conditional routing. On the other hand, the error per time step is higher in the single-rail-GUE case than for single-rail CSWAP due to the limited fidelity of the state-transfer protocol as compared to the postselected CZ gates (the postselected state-transfer fidelity is improved in the dual-rail case). Thus in attempting to build a QRAM using one of the proposed architectures, the experimenter must carefully consider the trade-offs between hardware complexity, fidelity, success rate, etc.

One could additionally envision merging the two proposed architectures. Laying out the QRAM in a treelike structure, nodes closest to the root are naturally further spatially separated from one another. This motivates perhaps connecting these nodes via the GUE architecture, before reverting to beam-splitter connectivity for the nodes nearer the bottom. Of course, due to the use of 3D cavities with centimeter-scale footprints, scaling up our QRAM architectures will be challenging due to the limited space available in dilution refrigerators. However, we expect small-scale devices to be realizable in the near term [27]. These proof-of-principle experiments will motivate exploring strategies for scaling up if they prove successful.

Recent work on *virtual QRAM* [39] has explored implementing an N-bit QRAM on hardware nominally supporting only an M-bit query, with $M < N$. This technique comes at the cost of increased noise sensitivity, though Xu *et al.* have shown the robustness of virtul QRAM to Z-biased noise [39]. For the 3D cavities considered here, the main noise channel is amplitude damping via photon loss [20,42]. Whether the virtual-QRAM technique can be made robust to these errors is left for future work.

It is worthwhile to place our work in the context of modern quantum algorithms that utilize QRAM. Two examples of recent interest include (i) the quantum simulation of jellium with at least 54 electrons, studied in Ref. [6], and (ii) the factoring of RSA-2048 with a modern version of Shor's algorithm, analyzed in Ref. [4]. In (i), a QRAM of size $N \geq 162$ (at least $n = 8$ address qubits) is required for
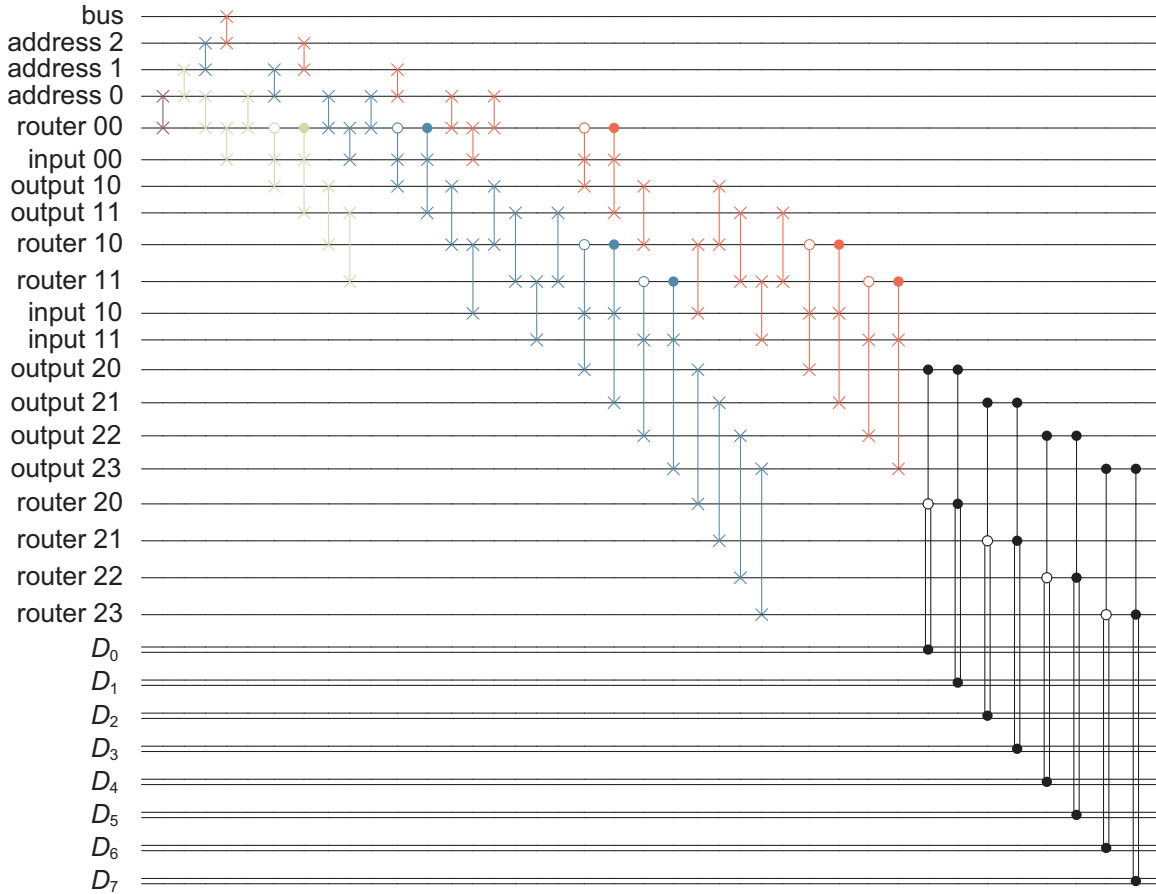
FIG. 8. The quantum circuit for a query on the 3-bit QRAM of Fig. 2(e). The gates required to route in address 0, address 1, address 2, and the bus are colored as purple, green, blue, and orange, respectively. The gates in black copy classical data into the bus. Referring to the routing gates (purple, green, blue, and orange gates) as U_{route} and the data-copy gates (black) as U_{data}, a full QRAM query is given by the unitary $U_{\text{route}}^{\dagger} U_{\text{data}} U_{\text{route}}$. The gates in $U_{\text{route}}^{\dagger}$, which disentangle the bus from the routers, are not shown due to space limitations. This quantum circuit has been prepared using the QCircuit package [53].

loading Hamiltonian coefficients. This device is queried on the order of 10^4 times, with a bus register of about 13 qubits [6]. In (ii), a QRAM of size approximately 2^{10} is utilized for classical precomputation of modular exponentials. This device is queried on the order of 10^5 times and requires a bus register of 2048 qubits (the number of bits of the integer to be factored). These results suggest that in terms of the number of address qubits, QRAMs of the sizes considered in this work may be relevant for quantum algorithms showing quantum advantage. However, query infidelities must decrease at least to the 10^{-4}–10^{-5} level to support $\geq 10^4$–10^5 QRAM queries. Additionally, in our architecture, each qubit in the bus register must be sent into the QRAM one at a time. This lengthens the query time and decreases the fidelity of a query. It is thus worthwhile in future work to consider noise-resilient QRAM architectures that natively support large bus registers.

Further future work could explore larger-scale numerical simulations [52] that go beyond a calculation of the error per time step ϵ and use of the query infidelity formula (3). Such simulations may be possible due to the limited entanglement present in the QRAM device, allowing for the use of, e.g., tensor-network methods. Such simulations would allow for the investigation of effects such as differing decoherence rates and measurement infidelities among different routers.

APPENDIX A: QUANTUM CIRCUIT FOR 3-BIT QRAM

The full quantum circuit associated with a query on the 3-bit QRAM of Fig. 2(e) is shown in Fig. 8. All of the gates are local and are color coded for clarity. We observe the benefits of address pipelining [3,39], where the third address qubit and the bus qubit undergo simultaneous CSWAP operations in different layers of the QRAM device.

APPENDIX B: CSWAP OPERATION

For the computational basis states $|n_a 00\rangle$, $|n_a 01\rangle$, and $|n_a 10\rangle$, $n_a = 0, 1$, the operation $\mathrm{BS}_{bc}^\dagger(\pi/2)\mathrm{CZ}_{ab}\mathrm{BS}_{bc}(\pi/2)$ appropriately functions as a CSWAP operation. To see this, note first that when the target modes are unoccupied, the gate operation is trivial: $\mathrm{BS}_{bc}^\dagger(\pi/2)\mathrm{CZ}_{ab}\mathrm{BS}_{bc}(\pi/2)|n_a 00\rangle = |n_a 00\rangle$. Now if the b mode is occupied, we obtain

$$\mathrm{BS}_{bc}^\dagger(\pi/2)\mathrm{CZ}_{ab}\mathrm{BS}_{bc}(\pi/2)|n_a 10\rangle$$
$$= \frac{1}{2}[1 + (-1)^{n_a}]|n_a 10\rangle + \frac{1}{2}[1 - (-1)^{n_a}]|n_a 01\rangle. \quad \text{(B1)}$$

Thus if $n_a = 0$, no SWAP is performed, while if $n_a = 1$, the photon is swapped to mode c. Similarly, if mode c is initially occupied and mode b is unoccupied, we obtain

$$\mathrm{BS}_{bc}^\dagger(\pi/2)\mathrm{CZ}_{ab}\mathrm{BS}_{bc}(\pi/2)|n_a 01\rangle$$
$$= \frac{1}{2}[1 - (-1)^{n_a}]|n_a 10\rangle + \frac{1}{2}[1 + (-1)^{n_a}]|n_a 01\rangle. \quad \text{(B2)}$$

We have thus obtained an effective CSWAP operation considering only the relevant states.

In the case of the initial state $|n_a 11\rangle$, $n_a = 0, 1$, the 50:50 beam splitter yields

$$\mathrm{BS}_{bc}(\pi/2)|n_a 11\rangle = \frac{1}{\sqrt{2}}(|n_a 02\rangle - |n_a 20\rangle), \quad \text{(B3)}$$

where the photons bunching in one or the other mode are known as the Hong-Ou-Mandel effect. Now, the state $|n_a 20\rangle$ outside of the computational manifold incurs a phase shift as a result of the CZ operation [see Eq. (7)]

and contrast with the case of a pure Kerr interaction, $\exp[i\pi\hat{a}^\dagger\hat{a}\hat{b}^\dagger\hat{b}]]$

$$\mathrm{CZ}_{ab}\frac{1}{\sqrt{2}}(|n_a 02\rangle - |n_a 20\rangle) = \frac{1}{\sqrt{2}}(|n_a 02\rangle + |n_a 20\rangle). \quad \text{(B4)}$$

The resulting state is a dark state of the inverse beam splitter

$$\mathrm{BS}_{bc}^\dagger(\pi/2)\frac{1}{\sqrt{2}}(|n_a 02\rangle + |n_a 20\rangle) = \frac{1}{\sqrt{2}}(|n_a 02\rangle + |n_a 20\rangle), \quad \text{(B5)}$$

and thus population in the state $|n_a 11\rangle$ is transferred out of the computational manifold at the completion of the gate $\mathrm{BS}_{bc}^\dagger(\pi/2)\mathrm{CZ}_{ab}\mathrm{BS}_{bc}(\pi/2)$. In the case of the ideal Kerr interaction, there is no such phase shift on the $|n_a 20\rangle$ state and the inverse beam splitter undoes the Hong-Ou-Mandel effect. We emphasize again that we do not expect both target modes to be occupied in the course of a QRAM query, as the system is initialized in the vacuum state.

APPENDIX C: FIDELITY OF THE CZ OPERATIONS

Based on the ability to perform first-order error detection, we want to calculate the postselected fidelity of the CZ gate. To proceed, we define measurement operators $\mathrm{M}_\checkmark, \mathrm{M}_\times$. A measurement result associated with $\mathrm{M}_\checkmark$ indicates that no error has been detected, while $\mathrm{M}_\times$ indicates that an error has been detected. If before measurement the system was in the state ρ, the postmeasurement state ρ' conditioned on detecting no errors is

$$\rho' = \mathrm{M}_\checkmark \rho \mathrm{M}_\checkmark^\dagger / P_\checkmark(\rho), \quad \text{(C1)}$$

where $P_\checkmark(\rho) = \mathrm{Tr}(\rho \mathrm{M}_\checkmark^\dagger \mathrm{M}_\checkmark)$ is the probability of no errors. As discussed in Refs. [24,25], measurement of the ancilla transmon after the CZ operation provides first-order detection of errors in the ancilla (in the dual-rail case, we measure both transmons associated with the router rails). Measuring the transmon in $|g\rangle$ indicates that no errors have been detected, while measurement in $|e\rangle$ or $|f\rangle$ indicates a bit-flip or a phase-flip error, respectively. (The gate may, of course, still fail if, e.g., two ancilla errors occur during the gate.) In experimental reality, some jumps may occur that do not get flagged; i.e., we may misidentify the $|e\rangle$ and $|f\rangle$ states as $|g\rangle$. We model this by taking as our no-error measurement operator

$$\mathrm{M}_g = \sqrt{\eta_{gg}}|g\rangle\langle g| + \sqrt{\eta_{ge}}|e\rangle\langle e| + \sqrt{\eta_{gf}}|f\rangle\langle f|, \quad \text{(C2)}$$

with measurement coefficients $\eta_{gg} = 1 - 10^{-4}$, $\eta_{ge} = 0.01$, and $\eta_{gf} = \eta_{ge}^2$ [24]. (The opposite problem of misidentifying $|g\rangle$ as $|e\rangle$ or $|f\rangle$ only serves to decrease the success

rate). We can already see that such detrimental effects only enter at second order: first, an error must occur; then, it must be misidentified (rare for cQED experiments with high-fidelity readout such as, e.g., in Ref. [42]).

In the dual-rail case, we can also detect photon-loss events in the cavities by performing a parity check [25]. It is important to note that this parity check can be done only at the end of the QRAM query, once the addresses and bus have been routed in and back out of the tree [54]. Otherwise, we reveal "which-path" information that destroys the superposition state. Thus, in the following, we do not explicitly simulate a parity check at the completion of the CZ operation, which would take additional time of approximately π/χ [24,25] and unrealistically increase our error probability per time step due to additional uncaught transmon errors. Instead, we perform the ancilla measurement(s) as described above and then project onto the dual-rail basis to obtain an estimate of the fidelity boost due to utilizing dual-rail qubits. The overall no-error measurement operator is

$$M_{\checkmark} = M_{DR}M_g, \tag{C3}$$

where

$$M_{DR} = \sum_{i,j=0}^{1} |i_L\rangle|j_L\rangle\langle i_L|\langle j_L|, \tag{C4}$$

written in the logical dual-rail basis. This expression for M_{DR} ignores measurement errors in projecting onto the dual-rail basis; however, such effects are of subleading order compared to ancilla measurement errors and thus can be safely ignored. We have additionally neglected the effects of no-jump back action, which arises if the decay rates are not identical between the two cavities comprising a dual rail [24]. Essentially, superposition states become biased toward occupation of the longer-lived cavity. This effect can be mitigated by periodically performing SWAPs between the two rails of a dual rail, effectively echoing out the no-jump back action.

The normalization by $P_{\checkmark}(\rho)$ in Eq. (C1) ensures that the density matrix ρ' has unit trace. This nonlinear mapping complicates the application of standard fidelity metrics, which assume a linear quantum channel [55,56] that can be decomposed into a set of state-independent Kraus operators [57]. To proceed, we instead view this process as a linear map that yields a subnormalized state. Thus, we can apply formulas that depend on the process being a quantum channel, before correcting for the subnormalization. We utilize Nielsen's formula for the entanglement fidelity associated with a gate U [55]

$$\widetilde{F}_e(U) = \frac{\sum_{jk} \alpha_{jk}\text{Tr}(UU_j^\dagger U^\dagger[M_{\checkmark}\mathcal{E}_U\{\rho_k\}M_{\checkmark}^\dagger])}{d^3}, \tag{C5}$$

where $\mathcal{E}_U(\rho)$ is the quantum channel applied to the density matrix ρ before the error-detecting measurements and $\mathcal{E}_U(\rho) = U\rho U^\dagger$ when the map is unitary. d is the dimension of the Hilbert space (here, $d = 4$), the U_j are an orthonormal operator basis on the d-dimensional space, and the ρ_k are pure-state density matrices consisting of the computational basis states $|0\rangle, \ldots, |d-1\rangle$ as well as their superpositions $(|j\rangle \pm |k\rangle)/\sqrt{2}$ and $(|j\rangle \pm i|k\rangle)/\sqrt{2}$ for $j \neq k$. (In the case $d = 4$, there are four computational basis states and 24 superposition states.) The coefficients α_{jk} are defined by the decomposition $U_j = \sum_k \alpha_{jk}\rho_k$. It is convenient to use the basis $\sqrt{d}|j\rangle\langle k|$, as the decomposition in terms of density matrices is simple $|j\rangle\langle k| = |+\rangle\langle +| + i|-\rangle\langle -| - [|j\rangle\langle j| + |k\rangle\langle k|](1 + i)/2$, where $|+\rangle = (|j\rangle + |k\rangle)/\sqrt{2}$ and $|-\rangle = (|j\rangle + i|k\rangle)/\sqrt{2}$ [58] (note the typographical error in the formula in Ref. [58], fixed here).

The quantity $\widetilde{F}_e(U)$ encodes the product of the average entanglement fidelity and the success probability, as opposed to the average entanglement fidelity alone. We thus divide by the average success probability $P_U = \sum_k \text{Tr}(\mathcal{E}_U(\rho_k))/[2d^2 - d]$ to obtain the average postselected entanglement fidelity $F_e(U) = \widetilde{F}_e(U)/P_U$, summing over the $2d^2 - d$ states utilized in the decomposition of the operator basis. The average postselected gate fidelity $F_g(U)$ is then given by the standard formula [55,59]

$$F_g(U) = \frac{dF_e(U) + 1}{d + 1}. \tag{C6}$$

In the single-rail case, we simulate the CZ operation as described in Sec. III and Ref. [24], utilizing the QuTiP [60, 61] Lindblad-master-equation solver. We observe that the postselected infidelity scales with $(T_1^{t,ge})^{-2}$, $(T_\phi^{t,ee})^{-2}$ due to the ability to detect first-order transmon errors [see Figs. 9(a) and 9(b)]. Of course, the failure probability scales as $(T_1^{t,ge})^{-1}$, $(T_\phi^{t,ee})^{-1}$. Additionally, in the single-rail case there is no ability to detect first-order photon loss errors in the cavities (dephasing events in the cavity cannot be detected in either case). Thus the postselected infidelities and failure probabilities both scale as $(T_1^c)^{-1}$, $(T_\phi^c)^{-1}$.

In the dual-rail case, we make the simplification that the two halves of the dual rails are identical. This allows us to reuse results from the single-rail case and appropriately tensor together single-rail states to obtain dual-rail states. In this way, we simulate the logical CZ operation, which consists of two parallel physical CZ gates as in the logical CSWAP gates schematically represented in Fig. 2(c). Now, with the ability to detect first-order decay events in the cavities, we obtain a postselected infidelity that scales with $(T_1^c)^{-2}$ [see Fig. 9(c)]. Dephasing events in the cavity are still undetectable; thus the postselected infidelity still scales as $(T_\phi^c)^{-1}$ [see Fig. 9(d)].
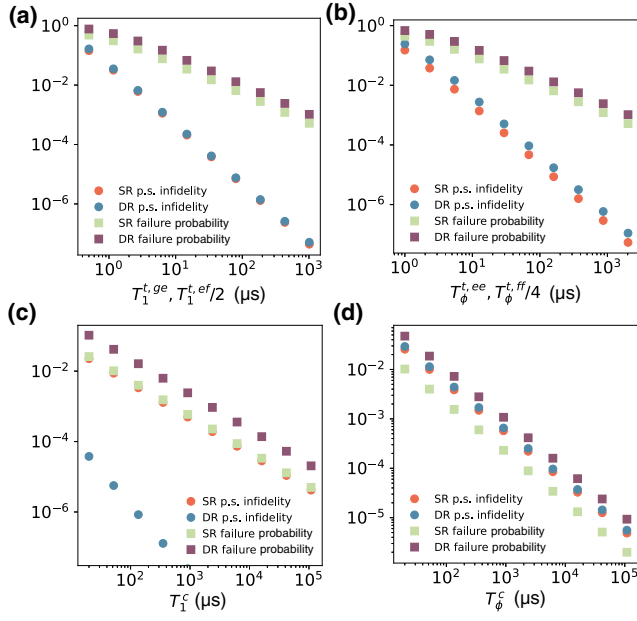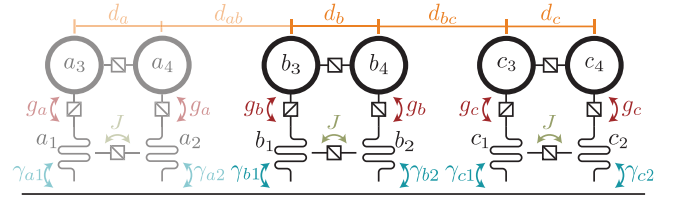
FIG. 10. A schematic of multiple GUEs coupled to a common waveguide. The physics of the state-transfer protocol can be understood by considering only two pairs of GUEs coupled to the waveguide, e.g., GUEs b and c.

left and right data cavities, respectively (see Fig. 10). It is important to note that the waveguide is terminated on each end with a 50-Ω impedance, such that the waveguide supports a continuum of modes as in Eq. (D1).

Including now the tunable beam-splitter interactions and moving into the interaction picture with respect to $\hat{H}_0$, the intra-GUE Hamiltonian is [making the rotating-wave approximation (RWA)]

$$\hat{H}_{\text{sys}}(t) = \sum_{\mu=b,c} [g_\mu(t)(\hat{\mu}_1^\dagger \hat{\mu}_3 + \hat{\mu}_2^\dagger \hat{\mu}_4) + J\hat{\mu}_1^\dagger \hat{\mu}_2 + \text{H. c.}].$$

(D2)

The coefficients $g_\mu(t)$ encode the time-dependent drive envelope of the beam-splitter interaction between the data cavities and the transfer resonators (see Fig. 10). As noted in Refs. [28–30], the static interaction between the transfer resonators is necessary to cancel an effective unwanted exchange interaction mediated by the waveguide. Of course, a beam-splitter interaction between the data cavities [omitted in Eq. (D2), as it is not relevant to the present discussion] is also necessary, e.g., for state preparation. The interaction Hamiltonian between the GUEs and the waveguide is [28]

$$\hat{H}_{\text{I}} = \frac{1}{\sqrt{2\pi}} \int_0^\infty d\omega e^{i(\omega-\omega_0)t} \{\hat{a}_{\text{L}}^\dagger(\omega)[\hat{L}_{\text{L}}^{(b)}(\omega) + \hat{L}_{\text{L}}^{(c)}(\omega)]$$
$$+ \hat{a}_{\text{R}}^\dagger(\omega)[\hat{L}_{\text{R}}^{(b)}(\omega) + \hat{L}_{\text{R}}^{(c)}(\omega)]\} + \text{H.c.},$$

(D3)

where

$$\hat{L}_{\text{R}}^{(b)}(\omega) = \sqrt{\gamma_{b1}}\hat{b}_1 + e^{-i\omega d_b/v}\sqrt{\gamma_{b2}}\hat{b}_2,$$
$$\hat{L}_{\text{L}}^{(b)}(\omega) = \sqrt{\gamma_{b1}}\hat{b}_1 + e^{i\omega d_b/v}\sqrt{\gamma_{b2}}\hat{b}_2,$$
$$\hat{L}_{\text{R}}^{(c)}(\omega) = e^{-i\omega(d_b+d_{bc})/v}(\sqrt{\gamma_{c1}}\hat{c}_1 + e^{-i\omega d_c/v}\sqrt{\gamma_{c2}}\hat{c}_2),$$
$$\hat{L}_{\text{L}}^{(c)}(\omega) = e^{i\omega(d_b+d_{bc})/v}(\sqrt{\gamma_{c1}}\hat{c}_1 + e^{i\omega d_c/v}\sqrt{\gamma_{c2}}\hat{c}_2).$$

(D4)

We define the decay rate $\gamma_{\mu j}$ of transfer resonator μ_j, $j = 1, 2$ into the waveguide (see Fig. 10). The separation between transfer resonators in GUEs b and c are defined

FIG. 9. The postselected gate infidelity and failure probability of the CZ operation. We sweep over transmon (a) relaxation and (b) dephasing times as well as cavity (c) relaxation and (d) dephasing. To observe the scaling of the infidelity and failure probability with each parameter, we set to zero all decay rates that are not being swept. In the single-rail case, postselection allows for the detection of first-order ancilla decay and dephasing errors [(a) and (b)]. Thus the postselected infidelities scale approximately quadratically with the coherence times. In the dual-rail case, we may now additionally postselect on first-order cavity-decay errors (c).

APPENDIX D: GUE STATE TRANSFER

The state-transfer problem involves the sender GUE emitting in both directions simultaneously in superposition. For simplicity, we restrict ourselves to the problem of state transfer between only two GUEs; we show below that in the ideal case of parameter symmetry, the state-transfer problems for both directions decouple. (When performing numerics and including decoherence processes, we analyze the full state-transfer problem of six data cavities and six transfer resonators.) The bare Hamiltonian of the waveguide and the GUEs is

$$\hat{H}_0 = \int_0^\infty d\omega\, \omega[\hat{a}_{\text{R}}^\dagger(\omega)\hat{a}_{\text{R}}(\omega) + \hat{a}_{\text{L}}^\dagger(\omega)\hat{a}_{\text{L}}(\omega)]$$
$$+ \sum_{\substack{\mu=b,c \\ j=1,2,3,4}} \omega_{\mu j}\hat{\mu}_j^\dagger \hat{\mu}_j,$$

(D1)

where $\hat{a}_{\text{L}}(\omega)$ and $\hat{a}_{\text{R}}(\omega)$ are the annihilation operators of left- and right-propagating modes at of the waveguide at frequency ω, respectively, $\hat{\mu}_1$ and $\hat{\mu}_2$ are the annihilation operators of the left and right transfer resonators, respectively, and $\hat{\mu}_3$ and $\hat{\mu}_4$ are the annihilation operators of the

as d_b and d_c, respectively, while d_{bc} denotes the separation between transfer resonators b_2 and c_1. The speed of light in the waveguide is defined as v.

In Eq. (D3), we have made the simplification that all transfer resonators have the same frequency ω_0. This assumption is justified for several reasons. First, the transfer resonators are over-coupled to the waveguide, with line widths of $\gamma_{\mu j}/2\pi \sim 10$–20 MHz assumed in this work. Second, the frequency of the parametric drive that produces the beam-splitter interaction can be tuned to adjust the frequency of the photon that is emitted into the waveguide [29]. Finally, flux-tunable qubits may be utilized as transfer resonators, as experimentally demonstrated in Ref. [30] (the state-transfer protocol discussed here is indifferent to whether the transfer resonators are qubits or are harmonic oscillators, as they are ideally not excited beyond the first excited state).

In the following, we take $d_b = d_c = \lambda/4$ [28,29], where $\lambda = 2\pi v/\omega_0$ is the wavelength of the photon emitted into the waveguide. We also assume symmetric decay rates $\gamma_{\mu j} = \gamma, \forall \mu, j$. We explore the effects of asymmetric decay rates in Appendix E. The authors of Ref. [29] have analyzed the case of deviations from $d = \lambda/4$, finding robustness of the directionality properties to deviations from the ideal value.

As detailed in Refs. [28,29,62], the waveguide can be eliminated in favor of an effective master-equation description in terms of coupled transfer resonators,

$$
\frac{d\rho}{dt} = -i\left[\hat{H}_{\text{eff}}, \rho\right] + \mathcal{D}[\hat{L}_R^{(b)}(\omega_0) + \hat{L}_R^{(c)}(\omega_0)]\rho \\
+ \mathcal{D}[\hat{L}_L^{(b)}(\omega_0) + \hat{L}_L^{(c)}(\omega_0)]\rho, \quad \text{(D5)}
$$

where

$$
\hat{H}_{\text{eff}} = \hat{H}_{\text{sys}}(t) + \sum_{\mu=b,c} \gamma(\hat{\mu}_1^\dagger \hat{\mu}_2 + \hat{\mu}_2^\dagger \hat{\mu}_1) \\
- \frac{i}{2}\Big([\hat{L}_R^{(c)}(\omega_0)]^\dagger \hat{L}_R^{(b)}(\omega_0) \\
+ [\hat{L}_L^{(b)}(\omega_0)]^\dagger \hat{L}_L^{(c)}(\omega_0) - \text{H. c.}\Big) \quad \text{(D6)}
$$

and $\mathcal{D}[\hat{\mathcal{O}}]\rho$ is the standard notation for the dissipator associated with collapse operator $\hat{\mathcal{O}}$ applied to the density matrix ρ. The second term in the first line of Eq. (D6) is the aforementioned effective exchange interaction between transfer resonators within the same GUE that is mediated by the waveguide. Choosing $J = -\gamma$ for the static coupling strength cancels this unwanted interaction. This effective description assumes that the GUEs exchange real photons only at the resonance frequency ω_0, while only virtual photons are exchanged at other frequencies [62].

We now consider the situation where no photons escape either to the left or to the right, known as the *dark-state condition* [28,45,47]. Using the language of quantum

trajectories, a pure state $|\psi(t)\rangle$ then evolves under the non-Hermitian Hamiltonian [63],

$$
\hat{H}_{\text{eff}}^{\text{nh}} = \hat{H}_{\text{eff}} - \frac{i}{2}\hat{L}_R^\dagger \hat{L}_R - \frac{i}{2}\hat{L}_L^\dagger \hat{L}_L, \quad \text{(D7)}
$$

where we have defined the collective decay operators $\hat{L}_{L/R} = \hat{L}_{L/R}^{(b)}(\omega_0) + \hat{L}_{L/R}^{(c)}(\omega_0)$. The non-Hermitian terms in Eq. (D7) interfere with those in $\hat{H}_{\text{eff}}$. After inserting the definitions given in Eq. (D4) and noting that $\omega_0 d_\mu/v = \pi/2, \mu = b, c$, we obtain

$$
\hat{H}_{\text{eff}}^{\text{nh}} = \sum_{\mu=b,c} [g_\mu(t)(\hat{\mu}_1^\dagger \hat{\mu}_3 + \hat{\mu}_2^\dagger \hat{\mu}_4) + \text{H. c.}] \\
- \frac{i}{2} \sum_{\substack{d=\text{R,L} \\ \mu=b,c}} [\hat{L}_d^{(\mu)}(\omega_0)]^\dagger \hat{L}_d^{(\mu)}(\omega_0) \\
- i\Big([\hat{L}_R^{(c)}(\omega_0)]^\dagger \hat{L}_R^{(b)}(\omega_0) + [\hat{L}_L^{(b)}(\omega_0)]^\dagger \hat{L}_L^{(c)}(\omega_0)\Big). \quad \text{(D8)}
$$

The structure of the last line of Eq. (D8) encodes the directionality: a left-propagating state $|\psi_L\rangle$ initialized in GUE c only couples to a left-propagating state in GUE b and a right-propagating state $|\psi_R\rangle$ in GUE b couples only to a right-propagating state in GUE c. That is to say, interference between terms in H_{eff} and those arising from the dissipators leads to the cancellation of terms that would allow for, e.g., the creation of the state $|\psi_L\rangle$ in GUE c along with the annihilation of the state $|\psi_L\rangle$ in GUE b.

We now proceed to derive the necessary control pulses $g_\mu(t), \mu = b, c$ to perform state transfer between GUEs b and c. Without loss of generality, we consider the state-transfer problem beginning with the initial state $|\psi_i\rangle = |\psi_R\rangle|00\rangle|00\rangle|00\rangle$ and ending with the final state $|\psi_f\rangle = |00\rangle|00\rangle|00\rangle|\psi_R\rangle$ after some specified time. We have ordered the kets as $b_3, b_4, b_1, b_2, c_1, c_2, c_3$, and c_4, tracking the progress of population moving from GUE b to GUE c. The state of the system at intermediate times is

$$
|\psi(t)\rangle = \alpha_{Rb}(t)|\psi_{Rb}\rangle + \alpha_{Rbt}(t)|\psi_{Rbt}\rangle \\
+ \alpha_{Rct}(t)|\psi_{Rct}\rangle + \alpha_{Rc}(t)|\psi_{Rc}\rangle, \quad \text{(D9)}
$$

where $|\psi_{R\mu}\rangle$ ($|\psi_{R\mu t}\rangle$) are the states where the data cavities (transfer resonators) of GUE μ are occupied. The time evolution of this state is governed by the time-dependent Schrödinger equation $i\frac{d}{dt}|\psi(t)\rangle = \hat{H}_{\text{eff}}^{\text{nh}}|\psi(t)\rangle$. Inserting Eq. (D9) into the Schrödinger equation yields the following

four coupled differential equations for the coefficients:

$$i\dot{\alpha}_{Rb}(t) = g_b^*(t)\alpha_{Rbt}(t),$$

$$i\dot{\alpha}_{Rbt}(t) = -i\gamma\alpha_{Rbt}(t) + g_b(t)\alpha_{Rb}(t),$$

$$i\dot{\alpha}_{Rct}(t) = -i\gamma\alpha_{Rct}(t) + g_c(t)\alpha_{Rc}(t) + 2\gamma e^{i\phi}\alpha_{Rbt}(t),$$

$$i\dot{\alpha}_{Rc}(t) = g_c^*(t)\alpha_{Rct}(t),$$

$$(D10)$$

defining $\phi = \omega_0 d_{bc}/v$. Analytically solving for the pulses $g_b(t)$ and $g_c(t)$ that satisfy these four differential equations subject to the initial and final conditions does not appear to be straightforward. To proceed, we adiabatically eliminate the transfer resonators. This approximation is motivated by the over-coupling of the transfer resonator to the waveguide, causing any population to immediately be emitted; i.e., the adiabaticity condition is $g_\mu \ll \gamma$. Setting $\dot{\alpha}_{Rbt}(t) \approx 0, \dot{\alpha}_{Rct}(t) \approx 0$, we obtain

$$\dot{\alpha}_{Rb}(t) = -\frac{|g_b(t)|^2}{\gamma}\alpha_{Rb}(t)$$

$$\dot{\alpha}_{Rc}(t) = -\frac{|g_c(t)|^2}{\gamma}\alpha_{Rc}(t) + 2ie^{i\phi}\frac{g_c^*(t)g_b(t)}{\gamma}\alpha_{Rb}(t),$$

$$(D11)$$

describing an effective directional interaction between the two GUEs. Choosing $\phi = \pi/2$ (which amounts to a specific spacing of the GUEs) yields the exact same differential equation for directional state transfer as in Ref. [47, cf. Eq. (47)]. There, Stannigel *et al.* have studied state transfer between two qubits along a one-dimensional waveguide in an optomechanical setting. It is remarkable that we recover the results of Ref. [47], given that we have made no assumption about the directionality of the waveguide itself or preferential coupling to left- or right-propagating modes. Instead, the directionality here is due to appropriate spacing of the transfer resonators along the waveguide.

It is worth emphasizing here that the state-transfer protocol trivially extends to the case of superposition states. We say that it is trivial because the Hamiltonian in Eq. (D8) is number conserving: thus it immediately follows that we can execute the state transfer $(\beta_1|00\rangle + \beta_2|\psi_R\rangle)|00\rangle|00\rangle|00\rangle \rightarrow |00\rangle|00\rangle|00\rangle(\beta_1|00\rangle + \beta_2|\psi_R\rangle)$ if we can perform $|\psi_{Rb}\rangle \rightarrow |\psi_{Rc}\rangle$.

We apply the results of Ref. [47] to obtain pulses that yield a time-symmetric emitted wave packet, allowing for it to be absorbed using a time-reversed pulse. One set of solutions is

$$g_b(t) = \frac{\sqrt{\frac{\gamma}{2}}\exp(-\zeta t^2/2)}{\sqrt{\frac{1}{\xi} - \sqrt{\frac{\pi}{4\zeta}}\mathrm{erf}(\sqrt{\zeta}t)}},$$

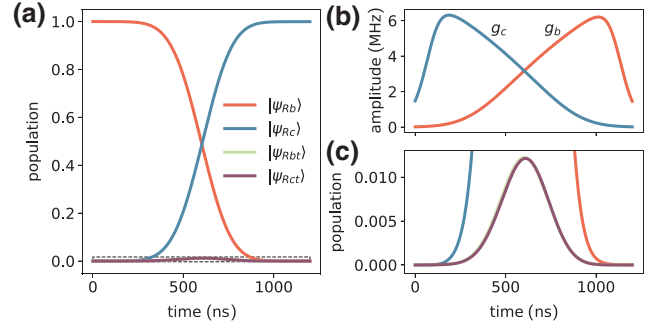$$(D12)$$

$$g_c(t) = g_b(-t),$$

FIG. 11. The state-transfer protocol between two GUEs. (a) Population swaps from GUE b to GUE c due to the control pulses (b) applied between the GUEs and the transfer resonators. (c) The transfer resonators are minimally occupied during the course of the operation [the area plotted in (c) is highlighted by a dashed box in (a)].

where ξ and ζ must be adjusted appropriately to satisfy the boundary conditions

$$\alpha_{Rb}(t_i) = |\alpha_{Rc}(t_f)| = 1, \qquad \alpha_{Rb}(t_f) = \alpha_{Rc}(t_i) = 0,$$

$$(D13)$$

in which t_i and t_f are the initial and final times, respectively, and $t_i = -t_f$. We find empirically that setting $\zeta \lesssim \xi^2\pi/4$ yields good results, leaving ξ as the only variable to tune. Throughout this work, we set $\xi/2\pi = 0.95$ MHz and $\gamma/2\pi = 20$ MHz. Moreover, we find that optimization over scale factors λ_μ, where $g_\mu(t) \rightarrow \lambda_\mu g_\mu(t), \mu = b, c$, improves the state-transfer fidelities (discussed below). We utilize $\lambda_b = 1.018, \lambda_c = 1.017$ in the remainder of this work.

The dark-state condition $\hat{L}_L|\psi(t)\rangle = 0$ is automatically true for the case of rightward state transfer considered here. However, the condition $\hat{L}_R|\psi(t)\rangle = \sqrt{\gamma}[\alpha_{Rbt}(t) - ie^{-i\phi}\alpha_{Rct}(t)] = 0 \forall t$, is only approximately satisfied due to suppressed (by the adiabatic condition) but nonzero occupation of the transfer resonators during the state-transfer protocol (see Fig. 11). Violation of the dark-state condition represents population lost to the waveguide and limits state-transfer fidelities, as we show below.

Toward performing numerical simulations of the full state-transfer protocol, we now include a third GUE (labeled a) to the left of GUE b (see Fig. 10). The master equation is

$$\frac{d\rho}{dt} = -i\left[\hat{H}_{\mathrm{eff}}, \rho\right] + \mathcal{D}[\hat{L}_R^{(a)}(\omega_0) + \hat{L}_R^{(b)}(\omega_0) + \hat{L}_R^{(c)}(\omega_0)]\rho$$

$$+ \mathcal{D}[\hat{L}_L^{(a)}(\omega_0) + \hat{L}_L^{(b)}(\omega_0) + \hat{L}_L^{(c)}(\omega_0)]\rho,$$

$$(D14)$$

where

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{sys}}(t) + \sum_{\mu=a,b,c} \gamma(\hat{\mu}_1^\dagger \hat{\mu}_2 + \hat{\mu}_2^\dagger \hat{\mu}_1) \qquad \text{(D15)}$$

$$- \frac{i}{2}\Big([\hat{L}_R^{(c)}(\omega_0)]^\dagger \hat{L}_R^{(b)}(\omega_0) + [\hat{L}_L^{(b)}(\omega_0)]^\dagger \hat{L}_L^{(c)}(\omega_0)$$

$$+ [\hat{L}_R^{(c)}(\omega_0)]^\dagger \hat{L}_R^{(a)}(\omega_0) + [\hat{L}_L^{(a)}(\omega_0)]^\dagger \hat{L}_L^{(c)}(\omega_0)$$

$$+ [\hat{L}_R^{(b)}(\omega_0)]^\dagger \hat{L}_R^{(a)}(\omega_0) + [\hat{L}_L^{(a)}(\omega_0)]^\dagger \hat{L}_L^{(b)}(\omega_0) - \text{H. c.}\Big),$$

and $\hat{H}_{\text{sys}}(t)$ is as in Eq. (D2) with the sum extended to include GUE a. As before, we obtain the non-Hermitian effective Hamiltonian by assuming the dark-state condition, yielding [28,47]

$$\hat{H}_{\text{eff}}^{\text{nh}} = \sum_{\mu=a,b,c} [g_\mu(t)(\hat{\mu}_1^\dagger \hat{\mu}_3 + \hat{\mu}_2^\dagger \hat{\mu}_4) + \text{H. c.}] \qquad \text{(D16)}$$

$$- \frac{i}{2} \sum_{\substack{d=\text{R,L} \\ \mu=a,b,c}} [\hat{L}_d^{(\mu)}(\omega_0)]^\dagger \hat{L}_d^{(\mu)}(\omega_0)$$

$$- i\Big([\hat{L}_R^{(c)}(\omega_0)]^\dagger \hat{L}_R^{(b)}(\omega_0) + [\hat{L}_L^{(b)}(\omega_0)]^\dagger \hat{L}_L^{(c)}(\omega_0)$$

$$+ [\hat{L}_R^{(c)}(\omega_0)]^\dagger \hat{L}_R^{(a)}(\omega_0) + [\hat{L}_L^{(a)}(\omega_0)]^\dagger \hat{L}_L^{(c)}(\omega_0)$$

$$+ [\hat{L}_R^{(b)}(\omega_0)]^\dagger \hat{L}_R^{(a)}(\omega_0) + [\hat{L}_L^{(a)}(\omega_0)]^\dagger \hat{L}_L^{(b)}(\omega_0)\Big),$$

a relatively straightforward generalization of Eqs. (D7)–(D8). We define the decay operators now utilizing a_1 as the origin; thus we have, e.g., $\hat{L}_R^{(a)}(\omega_0) = \sqrt{\gamma_{a1}}\hat{a}_1 - i\sqrt{\gamma_{a2}}\hat{a}_2$ and $\hat{L}_R^{(c)}(\omega_0) = (-i)^2 e^{-i(\phi_{ab}+\phi_{bc})}(\sqrt{\gamma_{c1}}\hat{c}_1 - i\sqrt{\gamma_{c2}}\hat{c}_2)$, where $\phi_{ab} = \omega_0 d_{ab}/v$, $\phi_{bc} = \omega_0 d_{bc}/v$. We take $g_a(t) = g_c(t)$ to simultaneously catch the emitted wave packets in both receiver GUEs. In the single-rail case, the initial basis states are $\{|\overline{0}\rangle \equiv |00\rangle|00\rangle|00\rangle, |\overline{\psi_R}\rangle \equiv |00\rangle|\psi_R\rangle|00\rangle, |\overline{\psi_L}\rangle \equiv |00\rangle|\psi_L\rangle|00\rangle\}$, (ordering the kets as a_3, a_4, b_3, b_4, c_3, and c_4 and omitting the transfer resonators), while in the dual-rail case, they are $\{|\overline{\psi_R}\rangle|\overline{0}\rangle, |\overline{\psi_L}\rangle|\overline{0}\rangle, |\overline{0}\rangle|\overline{\psi_R}\rangle, |\overline{0}\rangle|\overline{\psi_L}\rangle\}$. The final basis states are similarly defined.

Simulating the full state-transfer protocol is challenging even in the single-rail case due to the 12 subsystems involved (six data cavities and six transfer resonators). We take advantage of the fact that the Hamiltonian is number conserving, where the only non-number-conserving processes are due to decay or heating. Thus we specify a global excitation number cutoff when constructing our basis, as opposed to including, e.g., 3^{12} basis states (allowing for two excitations per mode to include heating effects) [64,65]. We take care to ensure that our results do not depend on the global excitation-number cutoff.

In the dual-rail case, we reconstruct the time evolution of a dual-rail state by appropriately tensoring together the

time evolution of single-rail states [66]. State transfer is now performed pairwise, with the right GUE in the sender dual rail transmitting to the right receiver GUE and likewise for the left GUEs. On the one hand, if these different pairs of GUEs are coupled to separate waveguides (in the experiment of Ref. [67] the authors have coupled two qubits via a 64-m Al cable where the cable left the 2D chip, suggesting an architecture where multiple cables can be braided in 3D), then the state-transfer problems proceeds as in the single-rail case. On the other hand, if all GUEs are connected to the same waveguide, each state-transfer operation includes passing through an "inactive" GUE (see Fig. 5). This pass-through GUE is inactive in the sense that the beam-splitter coupling between the data cavities and transfer resonators is turned off; however, the transfer resonators are still coupled to the waveguide. Perhaps surprisingly, as discussed in Ref. [29] and Appendix E, in the ideal case of parameter symmetry, this GUE serves only to impart a Wigner time delay on the transmitted photon [29,68,69]. Thus the previously derived state-transfer protocol can be applied to this case, provided that the retarded time of the receiver GUEs are redefined to account for the Wigner delay (in addition to the delay due to the finite propagation time of the photon). We discuss the nonideal case of parameter asymmetry in Appendix E.

The average state-transfer fidelity F_{st} is calculated by averaging the individual state-transfer fidelities,

$$F_{\text{st},\mu} = \langle\mu|\text{Tr}_1[\mathcal{E}_{\text{st}}(|\mu\rangle\langle\mu|)]|\mu\rangle, \qquad \text{(D17)}$$

over the initial basis states as well as their X and Y superpositions. The trace is performed over the transfer resonators as well as the initial data cavities. In the single-rail case, $\mathcal{E}_{\text{st}}$ includes only time evolution under the state-transfer protocol, while for dual rail, we also include a projective measurement onto the dual-rail states. Again, we include this measurement only to obtain an estimate for the postselected fidelity of a dual-rail QRAM query and emphasize that such a measurement cannot actually be performed during a query. We include the effects of nonradiative decay of the transfer resonator as well as transfer-resonator dephasing (see Table II, in addition to the data-cavity coherence times in Table I).

In the single-rail case and utilizing PS2 estimates, we obtain a state-transfer infidelity of 3.3×10^{-4} (see Table II). Decay to the vacuum state generally limits the fidelity of the state-transfer protocol, due both to decoherence as well as to violation of the dark-state condition. Thus, upon sweeping the coherence times of the data cavities and transfer resonators, the single-rail infidelity plateaus at the 10^{-4} level (see Fig. 12). Utilizing a dual-rail architecture (see Fig. 5) helps to mitigate this issue, where now decay to the vacuum is detectable. Therefore the infidelity of the dual-rail state-transfer protocol scales
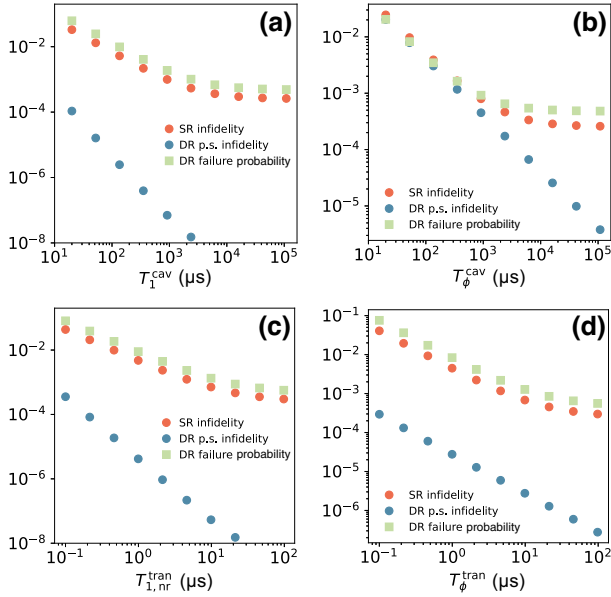
FIG. 12. The state-transfer infidelity and failure probability. We sweep over data cavity (a) T_1^{cav} and (b) T_ϕ^{cav} as well as (c) $T_{1,\text{nr}}^{\text{tran}}$ and (d) T_ϕ^{tran}. The plateau of the single-rail infidelity (and failure probability) is due to the inherent infidelity associated with our state-transfer protocol, at the level of 10^{-4} for our parameters. In the dual-rail case, decay to the vacuum state is detectable; thus the postselected infidelity scales as $(T_1^{\text{cav}})^{-2}$, $(T_{1,\text{nr}}^{\text{tran}})^{-2}$.

quadratically with cavity decay and transfer-resonator non-radiative decay (see Fig. 12). In terms of PS2 estimates, the postselected state-transfer infidelity is 4.0×10^{-6} (see Table II). The failure rate in this case is 6.1×10^{-4}, which is, as expected, roughly twice the single-rail infidelity.

APPENDIX E: NONIDEALITIES IN GUE STATE TRANSFER

Various nonidealities affect the ability to perform high-fidelity state transfer. These include (i) imperfect cancellation of the direct coupling between transfer resonators, (ii) detuning of the transfer resonators from resonance, (iii) deviation from $d = \lambda/4$ of the distance between the transfer resonators, and (iv) disorder in the radiative decay rates of the transfer resonators. Imperfect cancellation (i) can be addressed by using tunable beam-splitter interactions (as discussed in the main text). Nonidealities (ii) and (iii) can be addressed by utilizing, e.g., flux-tunable transmons as the transfer resonators, as in the experiment in Ref. [30]. This allows for the emitters to be tuned into resonance at the specific frequency appropriate for their true separation along the waveguide. Disorder in the decay rates (iv) does not appear to be easily tunable *in situ*; thus we focus on characterizing the directionality properties associated with asymmetric decay rates.

There are three processes that could be affected: emission, absorption, and "pass through." Note that due to

time-reversal symmetry, we need only consider one of emission or absorption. The directionality properties for the other process immediately follows. With respect to the pass-through problem, in the dual-rail architecture both rails may be connected to the same waveguide. Thus it becomes necessary for photons to pass through an inactive GUE. In the case of symmetric decay rates, wave packets pass through undistorted, incurring only a Wigner time delay [29,68]. Disorder in the decay rates of the inactive GUE could result in reflection or distortion of the transmitted signal and we analyze the symmetric and asymmetric cases below.

These problems are most appropriately analyzed in the context of input-output theory [70], considering a wave packet incident on a GUE. The derivation of the associated Langevin equations can be found in, e.g., Refs. [28,29], and we obtain

$$\frac{d\hat{b}_1(t)}{dt} = -\gamma_1 \hat{b}_1(t) - ig_b(t)\hat{b}_3(t)$$
$$- \sqrt{\gamma_1}e^{-i\omega_0 d/v}\hat{a}_R^{\text{in}}(t) - \sqrt{\gamma_1}\hat{a}_L^{\text{in}}(t), \quad \text{(E1)}$$

$$\frac{d\hat{b}_2(t)}{dt} = -\gamma_2 \hat{b}_2(t) - ig_b(t)\hat{b}_4(t)$$
$$- \sqrt{\gamma_2}e^{-i\omega_0 d/v}\hat{a}_L^{\text{in}}(t) - \sqrt{\gamma_2}\hat{a}_R^{\text{in}}(t), \quad \text{(E2)}$$

$$\frac{d\hat{b}_3(t)}{dt} = -ig_b(t)\hat{b}_1(t), \quad \text{(E3)}$$

$$\frac{d\hat{b}_4(t)}{dt} = -ig_b(t)\hat{b}_2(t) \quad \text{(E4)}$$

in terms of the input fields $\hat{a}_{\text{L,R}}^{\text{in}}(t)$ and

$$\frac{d\hat{b}_1(t)}{dt} = \gamma_1 \hat{b}_1(t) - ig_b(t)\hat{b}_3(t) \quad \text{(E5)}$$
$$- \sqrt{\gamma_1}e^{-i\omega_0 d/v}\hat{a}_R^{\text{out}}(t) - \sqrt{\gamma_1}\hat{a}_L^{\text{out}}(t), \quad \text{(E6)}$$

$$\frac{d\hat{b}_2(t)}{dt} = \gamma_2 \hat{b}_2(t) - ig_b(t)\hat{b}_4(t)$$
$$- \sqrt{\gamma_2}e^{-i\omega_0 d/v}\hat{a}_L^{\text{out}}(t) - \sqrt{\gamma_2}\hat{a}_R^{\text{out}}(t)$$

in terms of the output fields $\hat{a}_{\text{L,R}}^{\text{out}}(t)$. Here, d is the distance between the GUEs, $\hat{b}_i$ is the annihilation operator for transfer resonator i, and $\hat{a}_d^{\text{in-out}}$ are the input-output fields in direction d. We first discuss the pass-through protocol, before returning to absorption.

1. Pass through

a. Symmetric decay rates

In this case, we take $\gamma_1 = \gamma_2 = \gamma$. In the pass-through protocol, the beam-splitter interactions are turned off,

$g_b(t) = g_c(t) = 0$. We thus obtain the particularly simple Langevin equations for the right and left collective modes,

$$\frac{d\hat{b}_{R,L}(t)}{dt} = -2\sqrt{\gamma}\hat{a}^{in}_{R,L}(t) - \gamma\hat{b}_{R,L}(t) \qquad (E7)$$

$$= -2\sqrt{\gamma}\hat{a}^{out}_{R,L}(t) + \gamma\hat{b}_{R,L}(t), \qquad (E8)$$

$\hat{b}_R(t) = i\hat{b}_1(t) + \hat{b}_2(t)$ and $\hat{b}_L(t) = \hat{b}_1(t) + i\hat{b}_2(t)$. We read off the input-output relations,

$$\hat{a}^{out}_{R,L}(t) = \hat{a}^{in}_{R,L}(t) + \sqrt{\gamma}\hat{b}_{R,L}(t). \qquad (E9)$$

In this ideal case, the right and left modes completely decouple. We can solve for the output fields in terms of the input fields by working in Fourier space, defining

$$\hat{a}^{in,\,out}_{R,L}(t) = \int_{-\infty}^{\infty} \frac{d\omega}{\sqrt{2\pi}} \hat{a}^{in,\,out}_{R,L}(\omega)e^{-i\omega t}, \qquad (E10)$$

with similar definitions for the fields $\hat{b}_{R,L}(t)$. We emphasize here that because we are working in the rotating frame, zero frequency indicates resonance with the emitter frequencies. Using Eq. (E8), we obtain

$$\hat{b}_{R,L}(\omega) = \frac{2\sqrt{\gamma}}{i\omega - \gamma}\hat{a}^{in}_{R,L}(\omega), \qquad (E11)$$

leading immediately to

$$\hat{a}^{out}_{R,L}(\omega) = \frac{\omega - i\gamma}{\omega + i\gamma}\hat{a}^{in}_{R,L}(\omega). \qquad (E12)$$

Thus we observe that the output fields are related to the input fields by a frequency-dependent prefactor that has unit modulus. Interpreting this prefactor as a phase $e^{-i\phi(\omega)} = (\omega - i\gamma)/(\omega + i\gamma)$, we obtain

$$\phi(\omega) = \arctan\left(\frac{2\gamma\omega}{\omega^2 - \gamma^2}\right). \qquad (E13)$$

Taylor expanding about $\omega = 0$ yields

$$\phi(\omega) = -\frac{2\omega}{\gamma} + \frac{2\omega^3}{3\gamma^3} + \mathcal{O}\left([\omega/\gamma]^5\right). \qquad (E14)$$

For input fields that are nearly resonant with the emitters such that the cubic and higher terms can be neglected, the overall effect is a Wigner time delay [29,68,69]

$$\hat{a}^{out}_{R,L}(t) = \hat{a}^{in}_{R,L}(t - 2/\gamma). \qquad (E15)$$

Thus the inactive GUE can be modeled as merely producing a phase shift, in addition to that produced by the time delay associated with the spatial separation between emitters. Thus all results derived for the GUE architecture (specifically those associated with state transfer between GUEs) are immediately applicable to the dual-rail GUE architecture, where the phase shift between GUEs acquires a contribution from the Wigner delay associated with traversing the pass-through GUE.

b. Asymmetric decay rates

We now return to Eqs. (E1)–(E5) and allow for $\gamma_1 \neq \gamma_2$. The input-output relations are

$$\hat{a}^{out}_R(t) = \hat{a}^{in}_R(t) + i\sqrt{\gamma_1}\hat{b}_1(t) + \sqrt{\gamma_2}\hat{b}_2(t), \qquad (E16)$$

$$\hat{a}^{out}_L(t) = \hat{a}^{in}_L(t) + \sqrt{\gamma_1}\hat{b}_1(t) + i\sqrt{\gamma_2}\hat{b}_2(t), \qquad (E17)$$

where Eqs. (E16)–(E17) are not decoupled as they were in the symmetric case. Again working in Fourier space, we obtain

$$\hat{a}^{out}_{R,L}(\omega) = \pm\frac{\omega(\gamma_1 - \gamma_2)}{(\omega + i\gamma_1)(\omega + i\gamma_2)}\hat{a}^{in}_{L,R}(\omega)$$
$$+ \frac{\gamma_1\gamma_2 + \omega^2}{(\omega + i\gamma_1)(\omega + i\gamma_2)}\hat{a}^{in}_{R,L}(\omega). \qquad (E18)$$

Defining $\gamma = (\gamma_1 + \gamma_2)/2$ and $\delta\gamma = (\gamma_1 - \gamma_2)$ and expanding Eq. (E18) up to second order in $\delta\gamma$ yields

$$\hat{a}^{out}_{R,L}(\omega) = \pm\frac{\omega\delta\gamma}{(\omega + i\gamma)^2}\hat{a}^{in}_{L,R}(\omega)$$
$$+ \left(\frac{\omega - i\gamma}{\omega + i\gamma} - \frac{\omega\delta\gamma^2}{2(\omega + i\gamma)^3}\right)\hat{a}^{in}_{R,L}(\omega) + \mathcal{O}(\delta\gamma^3). \qquad (E19)$$

Thus the leading-order effect of asymmetry in the decay rates is to cause reflection of the input wave form (as opposed to modifying the transmitted wave-form shape or time delay, which is subleading order). To quantify the reduction in fidelity due to this reflection, we compute the probability of reflection,

$$p_{refl} = \frac{\int_{-\infty}^{\infty} d\omega \left|\frac{\omega(\gamma_1 - \gamma_2)}{(\omega + i\gamma_1)(\omega + i\gamma_2)} a^{in}_R(\omega)\right|^2}{\int_{-\infty}^{\infty} d\omega |a^{in}_R(\omega)|^2}, \qquad (E20)$$

assuming an input wave form traveling to the right and utilizing the full formula as in Eq. (E18). We obtain the input wave form from numerical solution of Eq. (D10). If we assume $\gamma/2\pi = 20$ MHz and a decay-rate asymmetry of 10% ($\delta\gamma/2\pi = 2$ MHz), we obtain $p_{refl} = 7.2 \times 10^{-6}$ [see Fig. 13(a)]. Given that we only expect to achieve state-transfer infidelities on the order of 10^{-4}, reflection off of a pass-through GUE due to decay-rate asymmetry is thus not a limiting factor. This robustness to decay-rate asymmetry is promising for a dual-rail architecture where all GUEs at one level of the tree are all connected to the same waveguide.
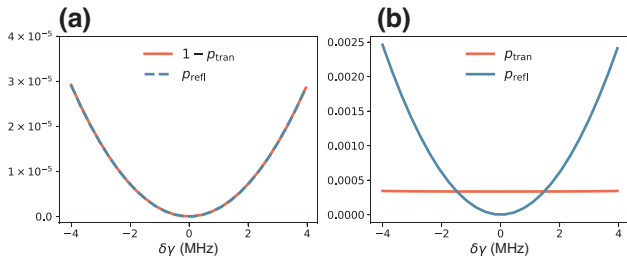
FIG. 13. The reflection and transmission probabilities during pass through and absorption. (a) The pass-through protocol is relatively insensitive to decay-rate asymmetries. The reflection probability and deviation of the transmission probability from unity are both less than 0.005% for $\delta\gamma/\gamma \leq 20\%$. (b) For the absorption process, the transmission probability is nearly constant as a function of $\delta\gamma$. This should ideally be zero and is not due to slight violation of the dark-state condition. The reflection probability varies quadratically with $\delta\gamma$ [see Eqs. (E19)–(E20)] and remains below the transmission probability for $\delta\gamma/\gamma \lesssim 10\%$.

2. Absorption

We now assume time-dependent beam splitter drives appropriate for absorption, as in Eq. (D12). These time-dependent drives complicate the analytics and we proceed by numerically integrating Eqs. (E1)–(E4). The output fields are obtained from the input fields and the internal modes using the input-output relations Eqs. (E16)–(E17). With these solutions in hand, we calculate the reflection and transmission probabilities in the case of a rightward-traveling input wave form,

$$p_{\text{refl}} = \frac{\int_{-\infty}^{\infty} dt |a_{\text{L}}^{\text{out}}(t)|^2}{\int_{-\infty}^{\infty} dt |a_{\text{R}}^{\text{in}}(t)|^2}, \tag{E21}$$

$$p_{\text{tran}} = \frac{\int_{-\infty}^{\infty} dt |a_{\text{R}}^{\text{out}}(t)|^2}{\int_{-\infty}^{\infty} dt |a_{\text{R}}^{\text{in}}(t)|^2}, \tag{E22}$$

respectively. For perfect absorption, both of these probabilities should vanish. In the ideal case of $\gamma_1 = \gamma_2$, the reflection probability indeed vanishes [see Fig. 13(b)]. This is consistent with the decoupling of the input-output relations in Eqs. (E16)–(E17) for $\gamma_1 = \gamma_2$ (these equations are unchanged by the beam-splitter drive). The transmission probability is nonvanishing even for $\gamma_1 = \gamma_2$ [see Fig. 13(b)], due to violation of the dark-state condition. This limits the fidelity of the state-transfer protocol in the symmetric case $\delta\gamma = 0$ and represents population lost to the waveguide. The transmission probability is essentially constant and thus for decay-rate variations of up to approximately 20%, the dark-state violation effect is the leading contributor to infidelity. We have also calculated (not shown) the coherent loss in fidelity due to $\delta\gamma \neq 0$ by simulating the state-transfer protocol using the master equations in Eqs. (D5), (D14). We find that for

modest values of asymmetry $\delta\gamma/\gamma < 0.2$, the violation of the dark-state condition still dominates the contribution to infidelity. It is thus interesting to explore, in future work, protocols for state transfer that respect the dark-state condition, which would improve the fidelities.

[1] L. K. Grover, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96 (Association for Computing Machinery, New York, NY, USA, 1996), p. 212.

[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, Cambridge, England, 2010), Chap. 8.

[3] S. Jaques and A. G. Rattew, QRAM: A Survey and Critique, arXiv:2305.10310.

[4] C. Gidney and M. Ekerå, How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits, Quantum **5**, 433 (2021).

[5] C. Gidney, Windowed quantum arithmetic, arXiv:1905.07682.

[6] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding electronic spectra in quantum circuits with linear T complexity, Phys. Rev. X **8**, 041015 (2018).

[7] D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization, Quantum **3**, 208 (2019).

[8] G. Kuperberg, Another subexponential-time quantum algorithm for the dihedral hidden subgroup problem, arXiv:1112.3333.

[9] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, Phys. Rev. Lett. **103**, 150502 (2009).

[10] V. Giovannetti, S. Lloyd, and L. Maccone, Architectures for a quantum random access memory, Phys. Rev. A **78**, 052310 (2008).

[11] F.-Y. Hong, Y. Xiang, Z.-Y. Zhu, L.-Z. Jiang, and L.-N. Wu, Robust quantum random access memory, Phys. Rev. A **86**, 010306(R) (2012).

[12] K. C. Chen, W. Dai, C. Errando-Herranz, S. Lloyd, and D. Englund, Scalable and high-fidelity quantum random access memory in spin-photon networks, PRX Quantum **2**, 030319 (2021).

[13] C. T. Hann, C.-L. Zou, Y. Zhang, Y. Chu, R. J. Schoelkopf, S. M. Girvin, and L. Jiang, Hardware-efficient quantum random access memory with hybrid quantum acoustic systems, Phys. Rev. Lett. **123**, 250501 (2019).

[14] C. Liu, M. Wang, S. A. Stein, Y. Ding, and A. Li, Quantum memory: A missing piece in quantum computing units, arXiv:2309.14432.

[15] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum random access memory, Phys. Rev. Lett. **100**, 160501 (2008).

[16] S. Arunachalam, V. Gheorghiu, T. Jochym-O'Connor, M. Mosca, and P. V. Srinivasan, On the robustness of bucket brigade quantum ram, New J. Phys. **17**, 123010 (2015).

[17] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, Quantum machine learning: A classical perspective, Proc. R. Soc. A: Math. Phys. Eng. Sci. **474**, 20170551 (2018).

[18] C. T. Hann, G. Lee, S. M. Girvin, and L. Jiang, Resilience of quantum random access memory to generic noise, PRX Quantum **2**, 020311 (2021).

[19] O. Milul, B. Guttel, U. Goldblatt, S. Hazanov, L. M. Joshi, D. Chausovsky, N. Kahn, E. Çiftyürek, F. Lafont, and S. Rosenblum, Superconducting cavity qubit with tens of milliseconds single-photon coherence time, PRX Quantum **4**, 030336 (2023).

[20] S. Rosenblum, P. Reinhold, M. Mirrahimi, L. Jiang, L. Frunzio, and R. J. Schoelkopf, Fault-tolerant detection of a quantum error, Science **361**, 266 (2018).

[21] S. Chakram, A. E. Oriani, R. K. Naik, A. V. Dixit, K. He, A. Agrawal, H. Kwon, and D. I. Schuster, Seamless high-Q microwave cavities for multimode circuit quantum electrodynamics, Phys. Rev. Lett. **127**, 107701 (2021).

[22] 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, S. O. Mundhada, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, High-on-off-ratio beam-splitter interaction for gates on bosonically encoded qubits, PRX Quantum **4**, 020355 (2023).

[23] Y. Lu, A. Maiti, J. W. O. Garmon, S. Ganjam, Y. Zhang, J. Claes, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, High-fidelity parametric beamsplitting with a parity-protected converter, Nat. Commun. **14**, 5767 (2023).

[24] 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, Error-detectable bosonic entangling gates with a noisy ancilla, PRX Quantum **4**, 020354 (2023).

[25] J. D. Teoh, P. Winkel, H. K. Babla, B. J. Chapman, J. Claes, S. J. de Graaf, J. W. O. Garmon, W. D. Kalfus, Y. Lu, A. Maiti, K. Sahay, N. Thakur, T. Tsunoda, S. H. Xue, L. Frunzio, S. M. Girvin, S. Puri, and R. J. Schoelkopf, Dual-rail encoding with superconducting cavities, PNAS **120**, e2221736120 (2023).

[26] A. Kubica, A. Haim, Y. Vaknin, F. Brandão, and A. Retzker, Erasure qubits: Overcoming the t_1 limit in superconducting circuits, arXiv:2208.05461.

[27] K. S. Chou *et al.*, Demonstrating a superconducting dual-rail cavity qubit with erasure-detected logical measurements, arXiv:2307.03169.

[28] P. O. Guimond, B. Vermersch, M. L. Juan, A. Sharafiev, G. Kirchmair, and P. Zoller, A unidirectional on-chip photonic interface for superconducting circuits, npj Quantum Inf. **6**, 32 (2020).

[29] N. Gheeraert, S. Kono, and Y. Nakamura, Programmable directional emitter and receiver of itinerant microwave photons in a waveguide, Phys. Rev. A **102**, 053720 (2020).

[30] B. Kannan, A. Almanakly, Y. Sung, A. Di Paolo, D. A. Rower, J. Braumüller, A. Melville, B. M. Niedzielski, A. Karamlou, K. Serniak, A. Vepsäläinen, M. E. Schwartz, J. L. Yoder, R. Winik, J. I.-J. Wang, T. P. Orlando, S. Gustavsson, J. A. Grover, and W. D. Oliver, On-demand directional microwave photon emission using waveguide quantum electrodynamics, Nat. Phys. **19**, 394 (2023).

[31] A. Ambainis, Quantum walk algorithm for element distinctness, arXiv:quant-ph/0311001.

[32] Y. Y. Gao, B. J. Lester, Y. Zhang, C. Wang, S. Rosenblum, L. Frunzio, L. Jiang, S. M. Girvin, and R. J. Schoelkopf, Programmable interference between two microwave quantum memories, Phys. Rev. X **8**, 021073 (2018).

[33] Y. Y. Gao, B. J. Lester, K. S. Chou, L. Frunzio, M. H. Devoret, L. Jiang, S. M. Girvin, and R. J. Schoelkopf, Entanglement of bosonic modes through an engineered exchange interaction, Nature **566**, 509 (2019).

[34] N. Schuch and J. Siewert, Natural two-qubit gate for quantum computation using the XY interaction, Phys. Rev. A **67**, 032301 (2003).

[35] I. L. Chuang and Y. Yamamoto, Simple quantum computer, Phys. Rev. A **52**, 3489 (1995).

[36] C. K. Hong, Z. Y. Ou, and L. Mandel, Measurement of subpicosecond time intervals between two photons by interference, Phys. Rev. Lett. **59**, 2044 (1987).

[37] We are assuming that the QRAM device is initialized in the vacuum state, as opposed to some proposed QRAM architectures that can be initialized in an arbitrary state [18].

[38] 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, Charge-insensitive qubit design derived from the Cooper pair box, Phys. Rev. A **76**, 042319 (2007).

[39] S. Xu, C. T. Hann, B. Foxman, S. M. Girvin, and Y. Ding, Systems architecture for quantum random access memory, arXiv:2306.03242.

[40] The arguments of Ref. [18] with regard to two-level routers do not exactly apply, as the reasoning there has been based on the ability of errors to only propagate up through routers on the left. They can be made to apply by considering separately each half of the dual rail and noting that each of these halves is occupied approximately half the time.

[41] A. Morvan, V. V. Ramasesh, M. S. Blok, J. M. Kreikebaum, K. O'Brien, L. Chen, B. K. Mitchell, R. K. Naik, D. I. Santiago, and I. Siddiqi, Qutrit randomized benchmarking, Phys. Rev. Lett. **126**, 210504 (2021).

[42] V. V. Sivak, A. Eickbusch, B. Royer, S. Singh, I. Tsioutsios, S. Ganjam, A. Miano, B. L. Brock, A. Z. Ding, L. Frunzio, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Real-time quantum error correction beyond break-even, Nature **616**, 50 (2023).

[43] C. Wang *et al.*, Towards practical quantum computers: Transmon qubit with a lifetime approaching 0.5 milliseconds, npj Quantum Inf. **8**, 3 (2022).

[44] A. P. M. Place *et al.*, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, Nat. Commun. **12**, 1779 (2021).

[45] J. I. Cirac, P. Zoller, H. J. Kimble, and H. Mabuchi, Quantum state transfer and entanglement distribution among distant nodes in a quantum network, Phys. Rev. Lett. **78**, 3221 (1997).

[46] A. N. Korotkov, Flying microwave qubits with nearly perfect transfer efficiency, Phys. Rev. B **84**, 014510 (2011).

[47] K. Stannigel, P. Rabl, A. S. Sørensen, M. D. Lukin, and P. Zoller, Optomechanical transducers for quantum-information processing, Phys. Rev. A **84**, 042341 (2011).

[48] P. Kurpiers, P. Magnard, T. Walter, B. Royer, M. Pechal, J. Heinsoo, Y. Salathé, A. Akin, S. Storz, J. C. Besse, S. Gasparinetti, A. Blais, and A. Wallraff, Deterministic quantum state transfer and remote entanglement using microwave photons, Nature **558**, 264 (2018).

[49] The router thus serves as a "switch" that communicates to incoming data either "keep going" or "change direction." Compiling that information into the actual location encoded by the address qubits can be done easily by tracking phases in software.

[50] U. von Lüpke, I. C. Rodrigues, Y. Yang, M. Fadel, and Y. Chu, Engineering phonon-phonon interactions in multimode circuit quantum acousto-dynamics, arXiv:2303.00730.

[51] D. Kim, S. Heng, S. Heng, and Y. Han, Maximizing the yield of bucket brigade quantum random access memory using redundancy repair, arXiv:2312.17483.

[52] L. Bugalho, E. Z. Cruzeiro, K. C. Chen, W. Dai, D. Englund, and Y. Omar, Resource-efficient simulation of noisy quantum circuits and application to network-enabled qram optimization, npj Quantum Inf. **9**, 105 (2023).

[53] B. Eastin and S. T. Flammia, Q-circuit tutorial, arXiv:quant-ph/0406003.

[54] Note that the QRAM query times $t_n = t_{CZ} N_{ts}$ are shorter than T_1^c for the parameters considered in this paper. Thus a parity check performed only at the end of the circuit is sufficient and we do not expect to be limited by uncaught cavity decay or heating events [71].

[55] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Phys. Lett. A **303**, 249 (2002).

[56] C. Dankert, Efficient simulation of random quantum states and operators, arXiv:quant-ph/0512217.

[57] L. H. Pedersen, N. M. Møller, and K. Mølmer, Fidelity of quantum operations, Phys. Lett. A **367**, 47 (2007).

[58] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, Phys. Rev. A **77**, 032322 (2008).

[59] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A **60**, 1888 (1999).

[60] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP: An open-source PYTHON framework for the dynamics of open quantum systems, Comput. Phys. Commun. **183**, 1760 (2012).

[61] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP 2: A PYTHON framework for the dynamics of open quantum systems, Comput. Phys. Commun. **184**, 1234 (2013).

[62] K. Lalumiere, B. C. Sanders, A. F. van Loo, A. Fedorov, A. Wallraff, and A. Blais, Input-output theory for waveguide QED with an ensemble of inhomogeneous atoms, Phys. Rev. A **88**, 043806 (2013).

[63] From the perspective of a dark state, this Hamiltonian actually is Hermitian.

[64] D. K. Weiss, W. DeGottardi, J. Koch, and D. G. Ferguson, Variational tight-binding method for simulating large superconducting circuits, Phys. Rev. Res. **3**, 033244 (2021).

[65] J. M. Zhang and R. X. Dong, Exact diagonalization: The Bose-Hubbard model as an example, Eur. J. Phys. **31**, 591 (2010).

[66] This assumes truly independent time evolution, e.g., if the different rails are connected to different waveguides such that the state transfer can be done in parallel. In the case in which they are connected to the same waveguide, the state transfer would need to be accomplished serially.

[67] J. Qiu, Y. Liu, J. Niu, L. Hu, Y. Wu, L. Zhang, W. Huang, Y. Chen, J. Li, S. Liu, Y. Zhong, L. Duan, and D. Yu, Deterministic quantum teleportation between distant superconducting chips, arXiv:2302.08756.

[68] E. P. Wigner, Lower limit for the energy derivative of the scattering phase shift, Phys. Rev. **98**, 145 (1955).

[69] E. H. Hauge and J. A. Støvneng, Tunneling times: A critical review, Rev. Mod. Phys. **61**, 917 (1989).

[70] C. W. Gardiner and M. J. Collett, Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation, Phys. Rev. A **31**, 3761 (1985).

[71] H. Levine, A. Haim, J. S. C. Hung, N. Alidoust, M. Kalaee, L. DeLorenzo, E. A. Wollack, P. A. Arriola, A. Khalajhedayati, Y. Vaknin, A. Kubica, A. A. Clerk, D. Hover, F. Brandão, A. Retzker, and O. Painter, Demonstrating a long-coherence dual-rail erasure qubit using tunable transmons, arXiv:2307.08737.