Spectator-Aware Frequency Allocation in Tunable-Coupler Quantum Architectures

Evan McKinney,[1, *] Israa G. Yusuf,[1, 2] Girgis Falstin,[1] Gaurav Agarwal,[2] Michael Hatridge,[2] and Alex K. Jones[3, †]

[1]*University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[2]*Yale University, New Haven, Connecticut, USA*
[3]*Syracuse University, Syracuse, New York, USA*

This paper addresses frequency crowding in SNAIL-based superconducting quantum modules. First, we present design constraints by describing a physical model for realizable gates within a module, and building a fidelity model using error budgeting derived from device characteristics. Second, we tackle the allocation problem by analyzing the impact of frequency crowding on gate fidelity as the radix of the module increases. We explore whether the heuristic gate fidelity can be optimized with a discrete set of qubit frequencies while adhering to defined separation thresholds. By leveraging a combination of analytical and numerical techniques, we demonstrate scalable frequency allocation strategies that minimize spectator-induced errors. Our results further indicate that removing edges leads to improved gate fidelities while maintaining sufficient connectivity, suggesting that edge density is not a limiting factor for NISQ-scale benchmarks. The findings have implications for designing robust, high-fidelity quantum systems with practical constraints on hardware and connectivity.

I. INTRODUCTION

Quantum circuits are executed by mapping gates to pulse controls on physically coupled qubit pairs. Accurate execution requires precise control of qubit-qubit couplings, activating interactions only when needed [2–4]. However, as coupling density increases, frequency crowding makes it harder to isolate specific interactions, degrading gate fidelity. Still, high connectivity is needed to reduce `SWAP` overhead and support topological error correction [5]. Modular superconducting architectures address this by enabling selective, localized coupling within modules.

A central challenge in Noisy Intermediate-Scale Quantum (NISQ) architecture design is balancing qubit density with gate fidelity (Fig. 1). Sparse layouts reduce crosstalk and yield higher base fidelities, but require more routing overhead. Denser layouts reduce `SWAP`s but suffer from increased parasitic interactions. We analyze this trade-off by incorporating physical constraints—qubit frequency ranges, coupler bandwidths, and hybridization strengths—into an error budgeting framework to propose architectures meeting fidelity targets under realistic chip constraints [6–9].

Another consideration is coupler-to-qubit ratio. Unlike fixed qubit-qubit links, tunable couplers can mediate multi-qubit interactions. In particular, the Superconducting Nonlinear Asymmetric Inductive eLement (SNAIL) enables all-to-all parametric coupling within a module, at the cost of weak static cross-Kerr terms [10]. This flexibility expands the architectural design space, provided spectator effects remain controllable.

We study how increasing the number of qubits per SNAIL affects the ability to selectively activate `iSWAP`
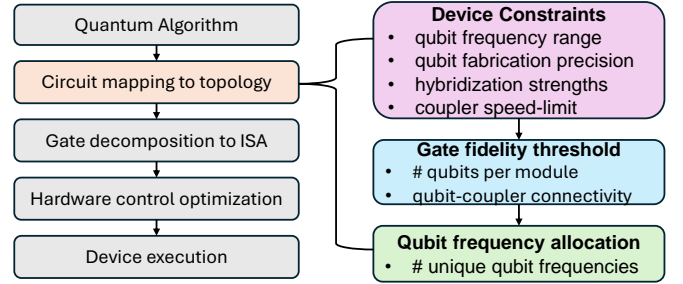

Figure 1. High-level workflow for the co-design of quantum computer architectures [1].

gates. While prior work has optimized frequency allocation and fabrication-aware layout [11–17], we focus on gate fidelity constraints induced by coupling all module qubits to a single SNAIL using its third-order mixing. Our contributions include:

- We show that gate fidelity is **100× more sensitive to detuning for SNAIL-qubit conversion than for qubit-qubit conversion**, establishing fundamental limits on modular architecture design.

- We solve a structured frequency allocation problem to determine the largest viable module size under fidelity constraints. We find **4 qubits per SNAIL** is the practical upper bound for preserving > 0.99 fidelity.

II. BACKGROUND

This section provides a brief overview of two-qubit gates such as `iSWAP`, introduces the SNAIL-based modular architecture used in our study, and defines the fidelity metrics used to evaluate architectural trade-offs.

Quantum computation is performed using unitary operations. Single-qubit gates are typically fast and low-
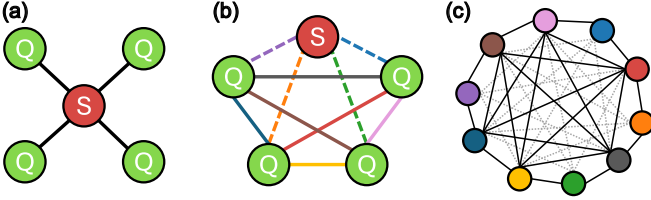
Figure 2. (a) SNAIL-qubit coupling graph (b) 2Q conversion connectivity graph, with unique colored edges for activated exchanges. (c) Compatibility graph, with nodes as interactions and edges as interferences.
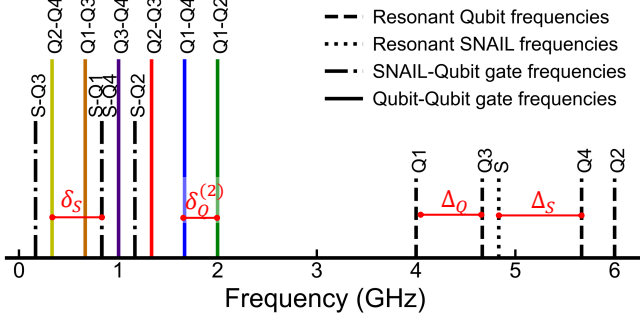

Figure 3. Spectral positioning of a SNAIL and 4 qubit bare modes and their interacting resonant frequencies.

error, while two-qubit gates introduce entanglement but are more error-prone. Though `CNOT` is common in algorithms, the `iSWAP` gate is more naturally implemented in tunable coupler architectures, enabling direct exchange of quantum states:

$$\sqrt[n]{\texttt{iSWAP}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\pi/2n) & i\sin(\pi/2n) & 0 \\ 0 & i\sin(\pi/2n) & \cos(\pi/2n) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

Gate dynamics are governed by the system Hamiltonian. Assuming time-independent evolution, the unitary is:

$$U(t) = e^{-i\int_0^t H(t')dt'/\hbar} = e^{-i\hat{H}t/\hbar}. \quad (2)$$

Understanding the ideal gate unitary helps identify and mitigate deviations due to parasitic interactions or decoherence.

We implement two-qubit gates using a SNAIL as a tunable coupler [18]. The SNAIL allows selective activation of specific interaction terms via targeted parametric pumping. Each SNAIL "Corral" module [10, 19] contains four fixed-frequency transmons, each with a single Josephson junction (JJ) that provides dominant fourth-order nonlinearity [20]. The SNAIL has an asymmetric loop of JJs, which when flux-biased to a critical value creates a device with strong third-order nonlinearity [18]. Each qubit is capacitvely coupled to a readout resonator and controlled via dedicated drive lines.

Parametric drives at resonant conversion frequencies induce hybridization-mediated exchange between qubit pairs. Gate performance thus depends on carefully managing interaction frequencies to avoid unintended exchanges and spectator errors (Fig. 2).

Quantum circuits incur both coherent (unitary) and incoherent (dissipative) errors. Coherent errors stem from unwanted Hamiltonian terms—e.g., crosstalk and spectator interactions. Incoherent errors include energy relaxation (T_1) and dephasing (T_ϕ), which irreversibly degrade state fidelity. We define fidelity heuristically to capture both error types. The average gate infidelity quantifies deviation from the target unitary [21]:

$$\epsilon_{\text{avg}} = 1 - \frac{d + \text{Tr}[UV^\dagger]}{d+1}, \quad (3)$$

where U is the ideal gate, V is the noisy implementation, and d is the Hilbert space dimension. Assuming additive contributions from spectator effects, total coherent infidelity is:

$$\epsilon_{\text{coh}} \approx \sum_i \epsilon_{\text{avg}}^{(i)}, \quad (4)$$

where $\epsilon_{\text{avg}}^{(i)}$ denoting infidelity from the ith spectator. We model incoherent decay as an exponential suppression: $\epsilon_{\text{inc}} \approx 1 - e^{-t/\tau}$. The combined fidelity is:

$$\epsilon_{\text{gate}} \approx 1 - (1 - \epsilon_{\text{inc}}) \times (1 - \epsilon_{\text{coh}}). \quad (5)$$

This decomposition isolates unitary and non-unitary effects. Spectator-induced infidelities add, while decoherence applies a multiplicative penalty. This model underlies our cost function for frequency allocation (Sec.IV) [22–24].

III. INFIDELITY IN A MODULE

In this section, we characterize how different types of infidelity contribute to errors within a module. We first quantify infidelity per detuning for various interactions. The dominant error sources are spectator terms, which accumulate as coherent infidelity. Additionally, subharmonic constraints of the SNAIL coupler impose limits on gate speed, contributing to incoherent infidelity. These effects together establish the frequency separation constraints used later in chip-level allocation.

A. Hamiltonian Expansion

To derive the two-qubit basis gates, we begin with a static system Hamiltonian and apply standard transformations to move into the interaction frame, allowing each term to acquire a pump-frequency-dependent coefficient.

The SNAIL-Corral architecture implements an `iSWAP` instruction set [25] using photon-conversion interactions.
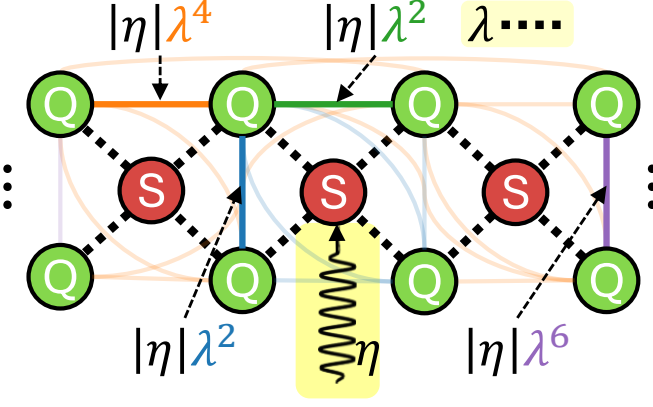
Figure 4. In this diagram, the central SNAIL is driven with the target interaction denoted by a green edge. Spectator terms diminish based on how many orders of hybridization they are removed from the driven SNAIL. (Blue) Both qubits coupled to the driven SNAIL (in the same module); (Orange) One qubit directly coupled to the driven SNAIL; (Purple) Neither qubit directly coupled to the driven SNAIL.

Table I. Order of magnitude analysis for driven, intra-module, and inter-module spectator terms sorted by normalized prefactor. Here, q_a and q_b are generic qubits in the driven module; s denotes the SNAIL; and q_c, q_d, s_n denote neighboring module elements.

Driven Term			
Term	Coefficient	$\omega_p =$	Normalized Prefactor
$(q_a^\dagger q_b + q_a q_b^\dagger)$	$6|\eta|\lambda^2 g_3$	$|\omega_{q_b} - \omega_{q_a}|$	1.0
Intra-Module Spectator Terms			
$(s^\dagger + s)$	$3|\eta|^2 g_3$	$\omega_s/2$	100.0
$(s^\dagger q_a + s q_a^\dagger)$	$6|\eta|\lambda g_3$	$|\omega_s - \omega_{q_a}|$	10.0
$(q_a^\dagger + q_a)$	$3|\eta|^2\lambda g_3$	$\omega_{q_a}/2$	10.0
$(s^\dagger q_a + s q_a^\dagger)$	$\alpha|\eta|^2\lambda^3$	$|\omega_s - \omega_{q_a}|/2$	0.067
$(q_a^\dagger + q_a)$	$\alpha|\eta|^3\lambda^3/3$	$\omega_{q_a}/3$	0.044
$(s^\dagger + s)$	$N_q\alpha|\eta|^3\lambda^4/3$	$\omega_s/3$	0.018
Inter-Module Spectator Terms			
$(s_n^\dagger + s_n)$	$3|\eta|^2\lambda^2 g_3$	$\omega_{s_n}/2$	1.0
$(s^\dagger q_c + s q_c^\dagger)$	$6|\eta|\lambda^3 g_3$	$|\omega_s - \omega_{q_c}|$	0.1
$(q_c^\dagger + q_c)$	$3|\eta|^2\lambda^3 g_3$	$\omega_{q_c}/2$	0.1
$(q_a^\dagger q_c + q_a q_c^\dagger)$	$6|\eta|\lambda^4 g_3$	$|\omega_{q_c} - \omega_{q_a}|$	0.01
$(s_n^\dagger q_a + s_n q_a^\dagger)$	$6|\eta|\lambda^5 g_3$	$|\omega_{s_n} - \omega_{q_a}|$	0.001
$(q_c^\dagger q_d + q_c q_d^\dagger)$	$6|\eta|\lambda^6 g_3$	$|\omega_{q_d} - \omega_{q_c}|$	0.0001

The full Hamiltonian is decomposed as $H = H_{0L} + H_{NL} + H_c$, with H_{0L} capturing the bare mode frequencies:

$$H_{0L} = \omega_s s^\dagger s + \sum_i \omega_i q_i^\dagger q_i, \qquad (6)$$

where ω_s and ω_i denote the SNAIL and qubit frequencies. The nonlinear terms are:

$$H_{NL} = g_3(s^\dagger + s)^3 + \sum_i \frac{\alpha_i}{12}(q_i^\dagger + q_i)^4, \qquad (7)$$

with g_3 the third-order SNAIL nonlinearity and α_i the transmon anharmonicity. Coupling between SNAIL and qubits is defined by:

$$H_c = \sum_q g_{sq}(s^\dagger q + s q^\dagger), \qquad (8)$$

where g_{sq} is the coupling strength.

The interaction frame is useful for describing an effective Hamiltonian where each interaction term is modulated by a rate (the strength of the pump) and a complex phase factor (how far off-resonance the interaction is from the pump's frequency). Transforming to the interaction frame via the Bogoliubov diagonalization and displacement transform [10, 26, 27], the Hamiltonian becomes:

$$\tilde{H}_I = g_3\left(se^{-i\tilde{\omega}_s t} + \eta e^{-i\omega_p t} + \sum_i \lambda_{si} q_i e^{-i\tilde{\omega}_{q_i} t} + \text{h.c.}\right)^3$$
$$+ \sum_i \frac{\alpha_i}{12}\left(q_i e^{-i\tilde{\omega}_{q_i} t} - \lambda_{si}(se^{-i\tilde{\omega}_s t} + \eta e^{-i\omega_p t}) + \text{h.c.}\right)^4. \qquad (9)$$

Here, $\tilde{\omega}$ are dressed mode frequencies, and $\lambda_{si} = g_{sq}/\Delta$ quantifies hybridization between the SNAIL and each

qubit. The pump strength $\eta = \sqrt{n_s}$ reflects the coherent pump amplitude in the SNAIL mode. Terms are resonant when $\tilde{\omega}_x \approx \omega_p$, and vanish when far-detuned. The Rotating-Wave Approximation (RWA) retains only those terms near resonance, this *assumes that the unwanted terms are actually far enough off-resonance* and thus could be considered sufficiently small [10, 26–28].

We use OpenFermion and SymPy to symbolically expand the cubic and quartic Hamiltonian polynomials (Eq. 9), in addition to normal-ordering and expression simplification. We retain dominant terms: two-qubit `iSWAP`s, SNAIL-qubit conversions, and subharmonic single-mode drives (Table I). Not included, *e.g.*, are gain-process terms, conjugate operators created by frequency additions rather than differences, because they are far from the pump.

To isolate a desired two-qubit gate, we set the pump on-resonance, $\omega_p = \tilde{\omega} q_2 - \tilde{\omega} q_1$, selecting the term:

$$H_{\text{target}} = 6|\eta|g_3\lambda^2(q_1 q_2^\dagger + q_1^\dagger q_2). \qquad (10)$$

The required pulse duration t_f follows from Eq. 2, yielding:

$$\frac{\pi}{2n} = 6t_f|\eta|g_3\lambda^2 \quad (\sqrt[n]{\text{iSWAP}}). \qquad (11)$$

Spectator interactions result from residual hybridization between qubits and the SNAIL. We distinguish intra-modular spectators—where the qubit is directly coupled to the driven SNAIL—from inter-modular ones, which involve indirect pathways through neighboring SNAILs (Fig.4). The strength and detuning of these interactions determine their contribution to infidelity, as detailed in TableI.
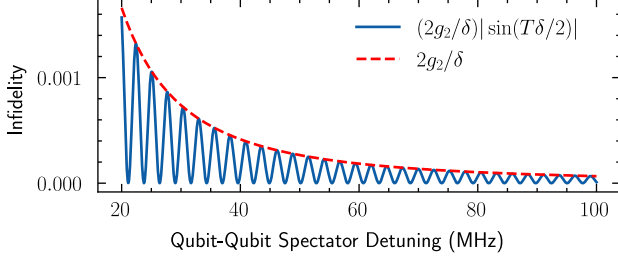
Figure 5. Infidelity versus detuning for spectator amplitude with Rabi oscillations compared with spectator amplitude using Rabi magnitude bound.

B. Spectator Infidelities

To establish frequency separation constraints, we evaluate how detuning impacts gate fidelity. Coherent errors arise from unintended spectator interactions, while incoherent errors are linked to SNAIL coupler speed limits.

1. Coherent loss from spectators

To quantify spectator-induced infidelity, we isolate each interaction term and simulate its impact using Eq. 3. A representative Hamiltonian with one target and one detuned spectator gate is:

$$H(t) = 6|\eta|g_3\lambda^2\left(q_1^\dagger q_2 + e^{-it\delta_Q}q_3^\dagger q_4 + \text{h.c.}\right). \quad (12)$$

Here, δ_Q denotes the detuning between the spectator term and the target gate.

In certain cases, the target gate may coincide with the anti-node of the spectator oscillation (Fig. 5), partially mitigating its effect. However, in realistic devices where interaction rates vary and gate durations are not as precisely synchronized nor fabricated. Instead, we apply a conservative amplitude bound:

$$\left|\int_0^T e^{-it\delta}dt\right| = 2|\sin(T\delta/2)|/\delta \le 2/\delta. \quad (13)$$

This avoids dependence on precise Rabi synchronization and simplifies the simulation by removing time dependence from the interaction term. The resulting unitary becomes:

$$U(t) = e^{-i\left(g_1 t(q_1^\dagger q_2) + 2g_2(q_3^\dagger q_4)/\delta + \text{h.c.}\right)}, \quad (14)$$

with g_1 and g_2 determined by the target and spectator coefficients in TableI. While the example above uses a spectator qubit-qubit term $q_3^\dagger q_4 + q_3 q_4^\dagger$, we apply the same procedure to other off-resonant interactions in the same way. The corresponding infidelity is then computed across a range of δ_Q values and fit to:

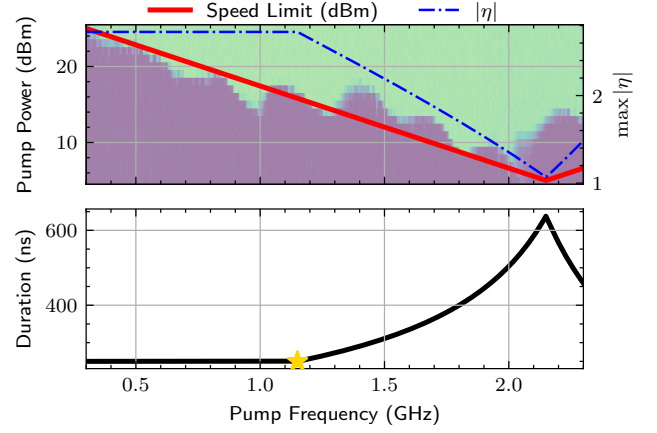$$\epsilon_{\text{coh}}(\delta) := \frac{2x_0}{(x_1 + \delta)^2} \quad (15)$$

Figure 6. (a) Maximum pump power vs. pump frequency. Overlaid is the maximum achievable $|\eta|$, fitted under the assumption that a viable gate duration exists at the star marker. (b) Gate duration vs. pump frequency, using the star marker to establish a scaling factor between pump power and $|\eta|$.

as shown in Fig. 7.

2. Incoherent loss from speed limits

We next consider incoherent loss from pump power constraints. Figure 6 shows experimental SNAIL breakdown data [26, 29], where exceeding a pump threshold causes loss of parametric behavior. Although the detailed physics are complex, we fit this data to approximate the maximal achievable η as a function of pump frequency.

Driving the SNAIL resonantly enables gate operations by populating it with photons. However, excessive pump power causes breakdown, particularly in subharmonic modes that are $\sim 100\times$ stronger than typical interaction terms (Table I). While these do not directly introduce unitary error, they destabilize gate execution and affect layout density and parallelization [30].

From Eq. 11, gate duration t_f is determined by η, which itself depends on ω_p and ω_s:

$$|\eta| = \frac{\epsilon\omega_s}{\omega_p^2 - \omega_s^2}, \quad (16)$$

Maintaining fixed gate time at low ω_p requires increasing drive amplitude ε, trading off gate speed against stability.

To model this tradeoff, we assume a linear relationship between dBm and ε, justified by attenuation being roughly linear across the drive chain. This allows us to define a breakdown threshold where increasing ε further would destabilize the coupler. Above this threshold, additional speedup is infeasible. The critical photon number [31] limits how strongly the SNAIL can be driven. Beyond this, nonlinear effects and chaotic behavior disrupt gate performance [27]. Other second-order effects—such as pump-induced frequency shifts and higher-order Kerr
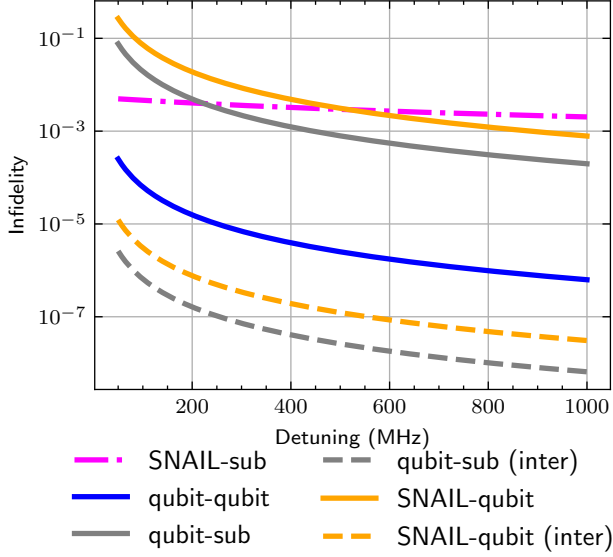
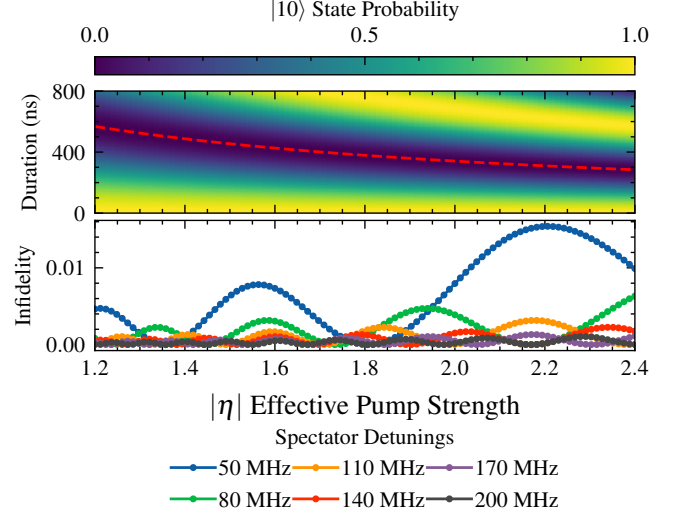Figure 7. Infidelity scaling versus detuning for different spectator types.

Figure 8. (a) Population exchange between $|01\rangle$ and $|10\rangle$ as a function of time and pump strength. The dashed red line marks a full `iSWAP`. (b) Coherent infidelity from a qubit-qubit spectator vs. effective pump strength at various δ_Q.
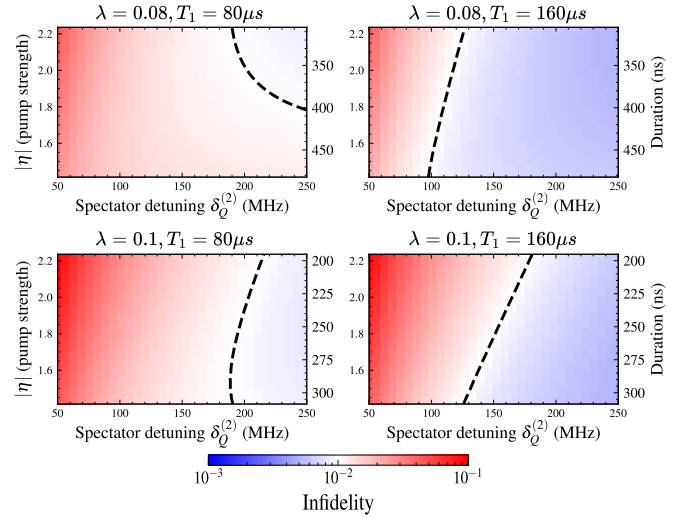
terms—further limit fidelity but are not modeled explicitly here [32].

Assuming a viable gate exists at the star marker in Fig.6 (250 μs at 1 GHz detuned from $\omega_S/2$), we fit the maximum usable η curve and compute t_f. Using Eq.5, we convert gate duration into decoherence-induced loss and fit the result with the heuristic:

$$\epsilon_{\mathrm{inc}}(\delta) := \frac{x_0}{(x_1 + \delta)} \tag{17}$$

which is plotted in Fig. 7.

C. Design Tradeoffs

Optimizing gate fidelity requires balancing gate speed against spectator interactions. Stronger pumps shorten gate time but amplify unwanted interactions (Fig. 8). From Eq. 14, achieving $g_1 t_f \gg 2g_2/\delta_Q$ ensures the target gate dominates. Increasing δ_Q suppresses spectators but lengthens t_f, increasing T_1-induced loss.

To evaluate both error types, we simulate noisy evolution using QuTiP's Lindblad solver with amplitude damping for T_1. The resulting operator $U(t_f)$ defines the fidelity threshold region in Fig. 9, plotted over λ and T_1. High-fidelity gates ($F \geq 0.99$) appear after around 150–200 MHz detuning. Weaker couplings shift the threshold left, into T_1-dominated loss; stronger pumps shift it upward, requiring more detuning to suppress spectators. This reveals an optimal operating region where both error sources are balanced—supporting the bounds in Fig. 6, which limit η accordingly.

We assume uniform SNAIL-qubit coupling ($\lambda_i \equiv \lambda_j$), though fabrication variability causes nonuniform hy-

Figure 9. Comparison of `iSWAP` fidelity threshold boundaries for different values of λ and T_1.

bridization. This affects both gate duration and spectator interference (Fig. 5). We also neglect higher-order nonlinearities in the transmon potential, which can cause AC Stark shifts and access higher excitation levels. Finally, we ignore direct qubit-qubit and SNAIL-SNAIL couplings ($\lambda_{qq} = \lambda_{ss} = 0$), though these may be non-negligible when modules are densely packed. In particular, couplers sharing qubits can introduce residual cross-coupling due to their physical proximity, complicating interaction isolation.
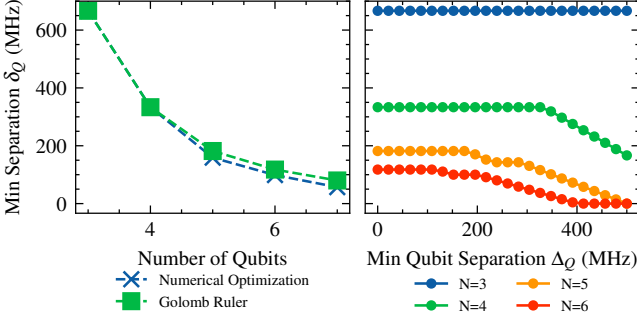
Figure 10. (Left) Minimum Difference Separation vs. Number of Qubits. Numerical optimization scales with the analytical Golomb Ruler but trails the optimal solution due to slight instability in convergence. (Right) Separation results with additional constraint of minimum allowed spacing between bare qubit frequencies.

IV. QUBIT FREQUENCY ALLOCATION

We now use the infidelity models from Section III B to construct frequency allocations that maximize gate fidelity. We formalize the frequency allocation task as a variant of the classical Frequency Assignment Problem (FAP), describe our numerical optimization strategy, and evaluate trade-offs that emerge in multi-module layouts.

FAP is a well-known NP-complete problem in which discrete frequency channels are assigned to transmitters while avoiding interference [33–35]. It is typically cast as a graph coloring problem: nodes are transmitters, colors are frequencies, and edges encode minimum required spacing:

$$|f(i) - f(j)| \geq q_{ij} \quad \forall(i,j), i \neq j \qquad (18)$$

While the classical objective is often to minimize total bandwidth, alternate formulations maximize the minimum pairwise spacing::

$$\max \min_{i,j \in X, i \neq j} |f(i) - f(j)| \qquad (19)$$

Quantum frequency allocation differs fundamentally from classical FAP in that physical constraints act on differences between frequencies, not their absolute values. Two-qubit gates are mediated by frequency detunings, and nonlinearities like subharmonic and hybridization effects impose coupled constraints across all interactions. These dependencies create a rugged optimization landscape that cannot be decomposed into independent frequency assignments.

To quantify achievable interaction separation, we establish a theoretical baseline using a Golomb ruler as an optimal bound. A Golomb ruler is a sequence of marks (interpreted here as frequencies) where all pairwise differences are unique, minimizing spectral crowding [36]. One construction is given by:

$$f_k = f_{\min} + c(2pk + (k^2 \bmod p)), \quad k = 0, 1, \ldots, p-1 \qquad (20)$$

where p is the number of qubits and c a scaling factor. Though suspected NP-hard [37], optimal sequences are known for small p, providing a benchmark for interaction separation.

However, the Golomb sequence requires marks to become increasingly close together, focusing only on relative differences while ignoring absolute placement constraints. In quantum systems, overly-close qubit frequencies create new challenges; for example, if qubits are too close in frequency, single-qubit gates become difficult to address, and higher-order state transitions introduce additional interference. To account for this, we introduce a constraint Δ_Q on the minimum allowed spacing between bare qubit frequencies.

Algorithm 1 Minimize Spectator Infidelity

Require: Module graph, frequency bounds $\omega_Q, \omega_S, \epsilon_{\mathrm{coh}}, \epsilon_{\mathrm{inc}}$, worst-gate exclusion k
Ensure: Optimized frequency assignment minimizing total infidelity
1: Initialize random frequencies $\{\omega_{q_i}\}, \omega_S$
2: **repeat**
3:     Collect all spectator frequencies $\mathcal{S} = \{\omega_{\mathrm{spec}}\}$ for all SNAIL-qubit, qubit-qubit pairs and qubit subharmonics
4:     **for** each gate (q_i, q_j) in module **do**
5:         $\epsilon_{\mathrm{coh}}(q_i, q_j) \leftarrow 0$
6:         **for** each spectator $\omega_{\mathrm{spec}} \in \mathcal{S}$ **do**
7:             $\epsilon_{\mathrm{coh}}(q_i, q_j) \mathrel{+}= \epsilon_{\mathrm{coh}}^{\mathrm{spec}}(|\omega_{q_i} - \omega_{\mathrm{spec}}|)$
8:         **end for**
9:         $\epsilon_{\mathrm{inc}}(q_i) \leftarrow \epsilon_{\mathrm{inc}}(|\omega_{q_i} - \omega_S/2|)$
10:        $\epsilon_{\mathrm{gate}}(q_i, q_j) \leftarrow 1 - (1 - \epsilon_{\mathrm{coh}}(q_i, q_j))(1 - \epsilon_{\mathrm{inc}}(q_i))$
11:     **end for**
12:     **for** each qubit **do**
13:         Apply penalty if $\min |\omega_{q_i} - \omega_{q_j}| < \Delta_Q$
14:     **end for**
15:     Sort $\{\epsilon_{\mathrm{gate}}\}$ in descending order and drop worst k gates
16:     $\mathcal{L} \leftarrow \sum_{\mathrm{gates\ remaining}} \epsilon_{\mathrm{gate}}$
17:     Update $\{\omega_{q_i}\}, \omega_S$ using Nelder-Mead to minimize $\mathcal{L}$
18: **until** convergence or max iterations reached
19: **return** Optimized $\{\omega_{q_i}\}, \omega_S$

Figure 10 compares our results to the Golomb bound. For small modules, the numerical solutions generally match the ideal separation. When Δ_Q is enforced, the Golomb bound itself begins to degrade. With four qubits, the optimizer maintains > 300 MHz spacing; with five, the worst pair drops below 200 MHz. This highlights the challenge of balancing single-qubit isolation against detuning for two-qubit gates.

Unlike the linear FAP, our cost function is nonlinear and involves absolute values, detuning-dependent infidelity models, and coupled penalty terms. We use a numerical optimizer (Algorithm 1) to assign qubit and coupler frequencies that minimize the total two-qubit gate infidelity, incorporating both coherent (spectator) and incoherent (lifetime) loss terms. Related methods have also been explored using graph neural networks [38] and perturbative corrections [39].

The scaling of spectator effects is derived from order-

of-magnitude estimates in Table I, which categorize the impact of different interaction types. Figure 7 translates these estimates into infidelity scaling, allowing the optimizer to focus on mitigating the most significant error sources. The largest contributors are coupler-qubit and qubit subharmonic, followed by qubit-qubit interactions. The full optimization process minimizes total infidelity for each gate:

$$\epsilon_{\text{gate}}(q_i, q_j) = 1 - (1 - \epsilon_{\text{coh}}(q_i, q_j))(1 - \epsilon_{\text{inc}}(q_i)). \quad (21)$$

under the following constraints:

$$3.3 \text{ GHz} \leq \omega_Q \leq 5.7 \text{ GHz} \quad (22)$$
$$4.2 \text{ GHz} \leq \omega_S \leq 4.7 \text{ GHz} \quad (23)$$
$$|\omega_{q_i} - \omega_{q_j}| \geq \Delta_Q = 200 \text{ MHz}. \quad (24)$$

The optimization proceeds via Nelder-Mead, with heavy penalties imposed for violations of Δ_Q. Figure 11 shows the optimized frequency placements, and Fig. 12 reports gate fidelities across module sizes. The geometric mean for gate fidelity for each module size:

- $N = 2$: $1 - \epsilon_{\text{gate}} \approx 0.996$

- $N = 3$: $1 - \epsilon_{\text{gate}} \approx 0.994$

- $N = 4$: $1 - \epsilon_{\text{gate}} \approx 0.991$

- $N = 5$: $1 - \epsilon_{\text{gate}} \approx 0.940$

Fidelity degrades as expected with module size due to increasingly tight spectral constraints.

V. DISCUSSION

A. Reducing Connectivity to Improve Fidelity

Full connectivity is not always necessary within a module. A module with N qubits supports $\binom{N}{2}$ interactions, but selectively omitting one or two gates can significantly improve overall fidelity. As shown in Fig. 12, a 4-qubit module with one gate removed (5 active) improves average fidelity from 0.991 to 0.993, and with two gates removed (4 active) reaches 0.994. These gains arise because omitting spectrally crowded interactions reduces pressure, enabling cleaner frequency separation for remaining gates. The optimizer reallocates detunings to prioritize fidelity across the active subset.

In NISQ devices, where fidelity is a limiting factor, our results suggest that selectively reducing connectivity may be preferable to enforcing full all-to-all coupling. However, realizing sparse physical layouts introduces practical challenges. Dedicated couplers for each interaction increase hardware complexity, chip area, and cryogenic routing demands. Instead, shared couplers—such as SNAILs—offer a more scalable alternative, reducing overhead while enabling compact designs and flexible control.
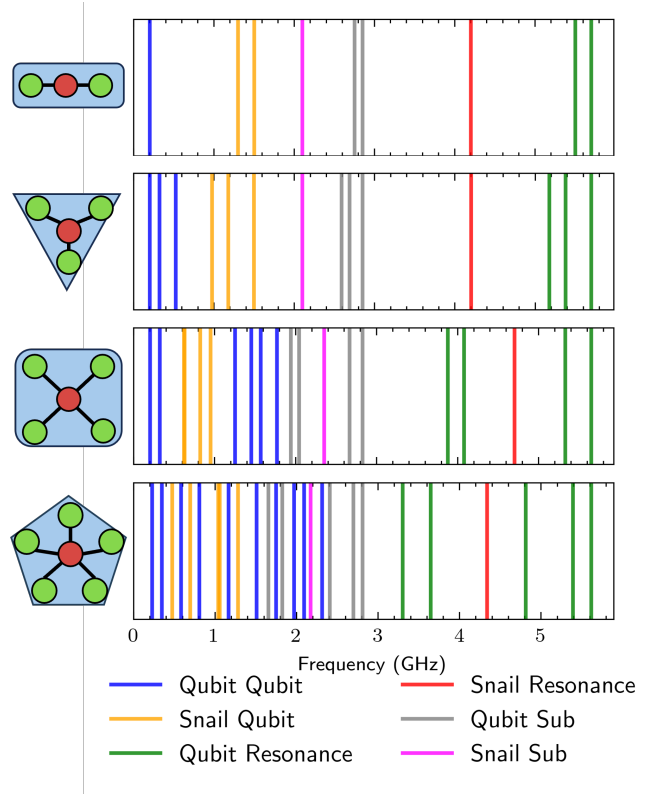
Figure 11. Optimized Frequency Stack: Frequencies of qubit and SNAIL resonances alongside interaction terms, grouped by module size $N = 2, 3, 4, 5$ (from top to bottom).
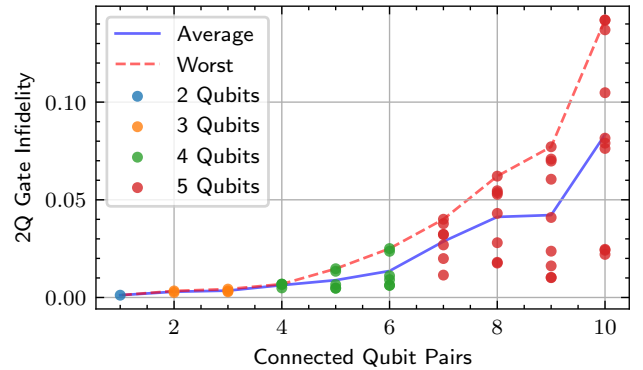
Figure 12. Average two-qubit gate infidelities across module sizes with and without selective edge removal.

B. Optimization Results and Design Considerations

Frequency allocation for two-qubit gates in SNAIL-based architectures presents a nonconvex and highly variable optimization landscape. Individual runs of the optimizer often converge to distinct local minima, and high-fidelity solutions typically require hundreds of random restarts. Despite this variability, the procedure is prac-

tically sufficient: the goal is not to guarantee a global minimum but to find any configuration that satisfies design and fabrication requirements. While better solutions may exist, the marginal benefit of exhaustive search is outweighed by its computational cost.

Since gate durations are typically short relative to qubit lifetimes, incoherent errors contribute minimally. Figure 7 confirms that infidelity is dominated by coherent spectator effects when gate-mode detunings fall below roughly 200 MHz. Thus, fidelity improvements will likely come from suppressing coherent errors, through more precise frequency allocation, improved control, or reduced coupling strengths (η, λ, g_3), albeit at the cost of slower gates.

Table II summarizes the gate fidelities for one such optimized frequency stack in a 4-qubit module. Achieving average gate fidelities above 0.99 required many hundreds of attempts, with typical solutions falling just below that threshold. In the reported case, average fidelities were $F_{\mathrm{coh}} = 0.9947$ and $F_{\mathrm{gate}} = 0.9918$, reaffirming that coherent errors remain the dominant limitation. Although the SNAIL-subharmonic penalty in our cost function is heuristic, it helps steer gate frequencies away from known unstable regions associated with SNAIL breakdown. While omitting this term does not drastically degrade results, including it ensures more robust solutions by avoiding poorly understood operational regimes.

Table II. Two-qubit gate parameters after frequency optimization. F_{coh} reports coherent gate fidelity (excluding lifetime loss), and F_{gate} includes incoherent decay.

Gate	Freq (GHz)	F_{coh}	F_{gate}
(Q0, Q1)	0.2000	0.9980	0.9967
(Q1, Q2)	0.3200	0.9939	0.9926
(Q0, Q2)	0.5200	0.9864	0.9849
(Q0, Q3)	1.8797	0.9930	0.9894
(Q1, Q3)	2.0797	0.9958	0.9912
(Q2, Q3)	2.3997	0.9909	0.9870

Because the optimization landscape is rugged, solutions are not unique. For example, some runs place two qubits above and two below the SNAIL frequency, while others put three above and one below. These variations allow some layout flexibility and can be selected to accommodate fabrication tolerances or simplify chip design. In more complex configurations with overlapping modules [40], additional frequency-sharing constraints must also be satisfied for qubits participating in multiple couplers—posing new optimization challenges that remain an active direction for future refinement. As overlapping complexity increases, it may become preferable to abandon continuous cost minimization in favor of discrete constraint satisfaction approaches.

Looking forward, we plan to explore noise-aware edge deletion and topological rewiring, including qubit-graph isomorphisms that preserve code distance while improv-

ing gate quality [41]. These ideas align with recent architectural proposals emphasizing modular, fault-tolerant layouts [42, 43].

Finally, direct hybridization between couplers remains a key limitation. In future designs, destructive interference via symmetric detuning may help suppress unwanted SNAIL–SNAIL interactions [32, 44]. Increasing hybridization strength λ or nonlinear coupling g_3 may enable faster gates, but must be balanced against increased crosstalk and instability. Mitigation strategies such as pulse shaping, compensatory drive cancellation [45–48], and error-aware scheduling could further reduce residual spectator terms. While our current model captures key hardware constraints, refined characterization of nonlinear dynamics and breakdown thresholds will be essential as next-generation devices emerge.

VI. CONCLUSION

We have presented a method for optimizing frequency allocation in superconducting qubit modules that balances the competing effects of spectator interactions and T_1-limited incoherent loss. By modeling both coherent and incoherent infidelity contributions and incorporating empirical constraints, we produce frequency layouts that achieve high-fidelity gate sets even in spectrally dense modules.

To further improve performance, we demonstrated that selectively removing low-quality gates can raise overall fidelity with minimal impact on connectivity. These techniques offer a path forward for both near-term quantum devices and future fault-tolerant systems. Future work will incorporate additional nonidealities—such as AC Stark shifts, higher-order couplings, and advanced pulse synthesis—to support scalable, high-performance quantum architectures under realistic hardware constraints.

[1] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, Layered architecture for quantum computing, Phys.Rev.X 10.1103/PhysRevX.2.031007 (2012).

[2] M. E. Beverland, P. Murali, M. Troyer, K. M. Svore, T. Hoefler, V. Kliuchnikov, G. H. Low, M. Soeken, A. Sundaram, and A. Vaschillo, Assessing requirements to scale to practical quantum advantage (2022).

[3] T. Tomesh and M. Martonosi, Quantum codesign, IEEE Micro **41**, 33 (2021).

[4] P. Murali, D. C. McKay, M. Martonosi, and A. Javadi-Abhari, Software mitigation of crosstalk on noisy intermediate-scale quantum computers, in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems* (2020) pp. 1001–1016.

[5] P. Murali, N. M. Linke, M. Martonosi, A. J. Abhari, N. H. Nguyen, and C. H. Alderete, Full-stack, real-system quantum computer studies: Architectural comparisons and design insights, in *Proceedings of the 46th International Symposium on Computer Architecture* (2019) pp. 527–540.

[6] M. Brink, J. M. Chow, J. Hertzberg, E. Magesan, and S. Rosenblatt, Device challenges for near term superconducting quantum processors: frequency collisions, in *IEDM* (2018).

[7] V. Tripathi, M. Khezri, and A. N. Korotkov, Operation and intrinsic error budget of a two-qubit cross-resonance gate, Physical Review A **100**, 012301 (2019).

[8] X. Ni, Z. Wang, R. Chao, and J. Chen, Superconducting processor design optimization for quantum error correction performance, arXiv preprint arXiv:2312.04186 (2023).

[9] E. A. Sete, V. Tripathi, J. A. Valery, D. Lidar, and J. Y. Mutus, Error budget of a parametric resonance entangling gate with a tunable coupler, Physical Review Applied **22**, 014059 (2024).

[10] C. Zhou, P. Lu, M. Praquin, T.-C. Chien, R. Kaufman, X. Cao, M. Xia, R. S. Mong, W. Pfaff, D. Pekker, *et al.*, Realizing all-to-all couplings among detachable quantum modules using a microwave quantum state router, npj quantum information **9**, 54 (2023).

[11] G. Li, Y. Ding, and Y. Xie, Towards efficient superconducting quantum processor architecture design, in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems* (2020) pp. 1031–1045.

[12] Y. Ding, P. Gokhale, S. F. Lin, R. Rines, T. Propson, and F. T. Chong, Systematic crosstalk mitigation for superconducting qubits via frequency-aware compilation, in *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)* (IEEE, 2020) pp. 201–214.

[13] K. N. Smith, G. S. Ravi, J. M. Baker, and F. T. Chong, Scaling superconducting quantum computers with chiplet architectures, in *2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO)* (IEEE, 2022) pp. 1092–1109.

[14] A. Morvan, L. Chen, J. M. Larson, D. I. Santiago, and I. Siddiqi, Optimizing frequency allocation for fixed-frequency superconducting quantum proces-

[15] sors, Phys.Rev.Res. 10.1103/PhysRevResearch.4.023079 (2022).

[15] A. Osman, J. Fernández-Pendás, C. Warren, S. Kosen, M. Scigliuzzo, A. Frisk Kockum, G. Tancredi, A. Fadavi Roudsari, and J. Bylander, Mitigation of frequency collisions in superconducting quantum processors, Phys.Rev.Res. 10.1103/PhysRevResearch.5.043001 (2023).

[16] J. Zhang, H. Wang, Q. Ding, J. Gu, R. Assouly, W. D. Oliver, S. Han, K. R. Brown, H. Li, Y. Chen, *et al.*, Qplacer: Frequency-aware component placement for superconducting quantum computers (2024).

[17] Z. Zhang, P. Gokhale, and J. M. Larson, Efficient frequency allocation for superconducting quantum processors using improved optimization techniques (2025).

[18] N. Frattini, U. Vool, A. Narla, K. Swila, and M. Devoret, 3-wave mixing josephson dipole element, Applied physics letters **110** (2017).

[19] E. McKinney, M. Xia, C. Zhou, P. Lu, M. Hatridge, and A. K. Jones, Co-designed architectures for modular superconducting quantum computers, in *2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA)* (IEEE, 2023) pp. 759–772.

[20] J. Koch, M. Y. Terri, J. Gambetta, A. Houck, D. Schuster, J. Majer, A. Blais, M. Devoret, S. Girvin, *et al.*, Charge-insensitive qubit design derived from the cooper pair box, Physical Review Letters A (2007).

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

[22] P. Hopf, N. Quetschlich, L. Schulz, and R. Wille, Improving figures of merit for quantum circuit compilation, arXiv preprint arXiv:2501.13155 (2025).

[23] P. Gokhale, T. Tomesh, M. Suchara, and F. Chong, Faster and more reliable quantum swaps via native gates, in *Proceedings of the 2024 International Conference on Parallel Architectures and Compilation Techniques* (2024) pp. 351–362.

[24] L. Schmid, D. F. Locher, M. Rispler, S. Blatt, J. Zeiher, M. Müller, and R. Wille, Computational capabilities and compiler development for neutral atom quantum processors—connecting tool developers and hardware experts, Quantum Science and Technology **9**, 033001 (2024).

[25] J. Chen, D. Ding, W. Gong, C. Huang, and Q. Ye, One gate scheme to rule them all: Introducing a complex yet reduced instruction set for quantum computing, in *ASPLOS* (2024).

[26] C. Zhou, *Superconducting Quantum Routers, Modules, Gates, and Measurements Based on Charge-pumped Parametric Interactions*, Ph.D. thesis, University of Pittsburgh (2023).

[27] M. Xia, C. Zhou, C. Liu, P. Patel, X. Cao, P. Lu, B. Mesits, M. Mucci, D. Gorski, D. Pekker, *et al.*, Fast superconducting qubit control with sub-harmonic drives, arXiv preprint arXiv:2306.10162 (2023).

[28] K. D. Barajas and W. C. Campbell, Quantum averaging for high-fidelity quantum logic gates, arXiv preprint arXiv:2503.08886 (2025).

[29] E. McKinney, C. Zhou, M. Xia, M. Hatridge, and A. K. Jones, Parallel driving for fast quantum computing under speed limits, in *Proceedings of the 50th Annual Interna-*

tional Symposium on Computer Architecture (2023) pp. 1–13.

[30] M. F. Dumas, B. Groleau-Paré, A. McDonald, M. H. Muñoz Arias, C. Lledó, B. D'Anjou, and A. Blais, Measurement-Induced Transmon Ionization, Phys. Rev. X **14**, 041023 (2024), arXiv:2402.06615 [quant-ph].

[31] N. E. Frattini, *Three-wave mixing in superconducting circuits: stabilizing cats with SNAILs*, Ph.D. thesis, Yale University (2021).

[32] P. J. Larkin, *Infrared and Raman Spectroscopy (Second Edition)* (Elsevier, 2018).

[33] T. Park and C. Y. Lee, Application of the graph coloring algorithm to the frequency assignment problem, Journal of the Operations Research Society of Japan 10.15807/jorsj.39.258 (1996).

[34] R. J. Waters, *Graph colouring and frequency assignment*, Ph.D. thesis, University of Nottingham (2005).

[35] D. Orden, J. M. Gimenez-Guzman, I. Marsa-Maestre, and E. De la Hoz, Spectrum Graph Coloring and Applications to Wi-Fi Channel Assignment, Symmetry 10.3390/sym10030065 (2018).

[36] N. Memarsadeghi, Nasa computational case study: Golomb rulers and their applications, Computing in Science & Engineering **18**, 58 (2016).

[37] C. Meyer and P. A. Papakonstantinou, On the complexity of constructing golomb rulers, Discrete applied mathematics **157**, 738 (2009).

[38] H. Ai and Y.-x. Liu, Graph neural networks-based parameter design towards large-scale superconducting quantum circuits for crosstalk mitigation, arXiv preprint arXiv:2411.16354 (2024).

[39] A. Mammola, Q. Schaeverbeke, and M. M. Desjardins, Optimal connectivity from idle qubit residual coupling cross-talks in a cavity mediated entangling gate, arXiv preprint arXiv:2503.07455 (2025).

[40] I. Yusuf, E. McKinney, G. Agarwal, G. Falstin, M. Xia, A. Jones, and M. Hatridge, A densely connected snail-based 6 qubit quantum module, in *APS March Meeting Abstracts* (2025).

[41] P. D. Nation and M. Treinish, Suppressing quantum circuit errors due to system variability, PRX Quantum **4**, 010327 (2023).

[42] S. Singh, F. Gu, S. de Bone, E. Villaseñor, D. Elkouss, and J. Borregaard, Modular architectures and entanglement schemes for error-corrected distributed quantum computation, arXiv preprint arXiv:2408.02837 (2024).

[43] E. Sutcliffe, B. Jonnadula, C. L. Gall, A. E. Moylett, and C. M. Westoby, Distributed quantum error correction based on hyperbolic floquet codes, arXiv preprint arXiv:2501.14029 (2025).

[44] J. Majer, J. Chow, J. Gambetta, J. Koch, B. Johnson, J. Schreier, L. Frunzio, D. Schuster, A. A. Houck, A. Wallraff, *et al.*, Coupling superconducting qubits via a cavity bus, Nature **449**, 443 (2007).

[45] D. Zajac, J. Stehlik, D. Underwood, T. Phung, J. Blair, S. Carnevale, D. Klaus, G. Keefe, A. Carniol, M. Kumph, *et al.*, Spectator errors in tunable coupling architectures, arXiv preprint arXiv:2108.11221 (2021).

[46] A. Seif, H. Liao, V. Tripathi, K. Krsulich, M. Malekakhlagh, M. Amico, P. Jurcevic, and A. Javadi-Abhari, Suppressing correlated noise in quantum computers via context-aware compiling, in *2024 ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA)* (IEEE, 2024) pp. 310–324.

[47] B.-h. Lu, Y.-c. Wu, P. Wang, and G.-p. Guo, A compilation scheme for suppressing crosstalk and decoherence in superconducting quantum chips with tunable coupling and tunable qubits, arXiv preprint arXiv:2311.18160 (2023).

[48] S. Vallés-Sanclemente, T. Vroomans, T. van Abswoude, F. Brulleman, T. Stavenga, S. van der Meer, Y. Xin, A. Lawrence, V. Singh, M. Rol, *et al.*, Optimizing the frequency positioning of tunable couplers in a circuit qed processor to mitigate spectator effects on quantum operations, arXiv preprint arXiv:2503.13225 (2025).

[49] E. McKinney, G. Falstin, I. G. Yusuf, G. Agarwal, M. Hatridge, and A. K. Jones, Towards error budgeting for superconducting modular quantum architecture designs, arXiv preprint arXiv:2409.18262v3 (2024).