YOUTIAO: Hybrid Multiplexing with Dynamic Qubit Grouping for Low-cost and Scalable Quantum Wiring

Wuwei Tian
Zhejiang University
Hangzhou, China
sonder@zju.edu.cn

Liqiang Lu*
ZJU-Ningbo Global Innovation Center
Zhejiang University
Ningbo, China
liqianglu@zju.edu.cn

Siwei Tan
ZJU-Ningbo Global Innovation Center
Zhejiang University
Ningbo, China
siweitan@zju.edu.cn

Shiyu Li
Zhejiang University
Hangzhou, China
shiyuli@zju.edu.cn

Hengyi Li
Zhejiang University
Hangzhou, China
lihengyi@zju.edu.cn

Tianyao Chu
Zhejiang University
Hangzhou, China
tianyao_chu@zju.edu.cn

Xuhong Zhang
ZJU-Ningbo Global Innovation Center
Zhejiang University
Ningbo, China
zhangxuhong@zju.edu.cn

Mingshuai Chen
Zhejiang University
Hangzhou, China
m.chen@zju.edu.cn

Jianwei Yin*
ZJU-Ningbo Global Innovation Center
Zhejiang University
Ningbo, China
zjuyjw@zju.edu.cn

Abstract

With continuous advances in physical technology, the number of qubits has increased from just a few to several thousand. To further extend the scale, the density of control lines has become one of the major limitations that decides the cost and control fidelity. Specifically, each superconducting qubit requires dedicated control lines to manipulate its state; however, these lines reach the spatial upper limit of the cryostat when scaling up the quantum chip. Inspired by collinear signal transmission, a promising solution is to adopt multiplexing methods—such as frequency-division multiplexing (FDM) and time-division multiplexing (TDM)—to share control lines among superconducting qubits. However, existing methods lack a systematic architectural approach to support multiplexing-aware wiring, which leads to low parallelism and high crosstalk during deployment. In this work, we propose a multiplexing-aware design for the peripheral control lines of superconducting quantum processors, combining cryostat-level wiring optimization with on-chip routing. Our key novelty lies in a hybrid multiplexing architecture that adopts FDM for XY control and readout lines, and TDM for Z control lines. This enables high utilization of natural non-parallel operations, thereby cutting the additional circuit depth for TDM control. Finally, we develop an interaction model that co-optimizes qubit layout and multiplexed channel allocation. Our experiments use the data collected from self-developed Xmon quantum chips. The results show that YOUTIAO achieves a 67.7% reduction in cryostat-level coaxial wiring complexity and overall superconducting quantum system costs, while reducing on-chip routing area by 23%. With these improvements, we still keep the 1q-gate fidelity at 99.98%, and only introduce 5% extra circuit latency compared to the partial-multiplexing system.

CCS Concepts

• **Computer systems organization** → **Quantum computing**; • **Hardware** → **Quantum technologies**.

Keywords

Quantum wiring system, Signal multiplexing

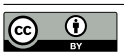
ACM Reference Format:
Wuwei Tian, Liqiang Lu, Siwei Tan, Shiyu Li, Hengyi Li, Tianyao Chu, Xuhong Zhang, Mingshuai Chen, and Jianwei Yin. 2025. YOUTIAO: Hybrid Multiplexing with Dynamic Qubit Grouping for Low-cost and Scalable Quantum Wiring. In *58th IEEE/ACM International Symposium on Microarchitecture (MICRO '25), October 18–22, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3725843.3756061

1 Introduction

Quantum computing exhibits a high potential to outperform classical computing in dealing with tough computational challenges, such as many-body physics [16, 43], and cryptography [33]. Scalability is one of the main bottlenecks faced by current quantum systems [10, 31]. Algorithms such as quantum error correction [36], quantum molecular simulation [14], quantum integer factorization [32], and database search [15] all require large-scale quantum systems to realize their practical advantages. Due to the complex physical interconnections and the low-fidelity state teleportation across cryogenic platforms (i.e., below 80%) [12, 30], the challenge of improving superconducting system integration and orchestration within a single cryostat becomes increasingly prominent [31].

One integration challenge results from wiring and control electronics costs, as well as superconducting quantum chip packaging

*Corresponding Author(s)



MICRO '25, Seoul, Republic of Korea

ACM ISBN 979-8-4007-1573-0/25/10
https://doi.org/10.1145/3725843.3756061

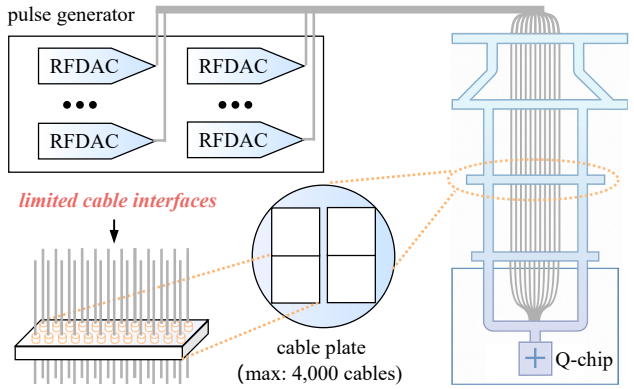
[31]. Among various sources of limitation, the thermal noise introduced from the superconducting qubit control wiring is a primary factor that limits the wiring density allowed for a single quantum device. Taking Google's superconducting quantum chip control scheme as an example, due to physical constraints related to wiring layout and vacuum sealing [37], even the advanced Bluefors KIDE cryostat only permits a maximum of 4,000 coaxial cables [6], thus limiting the future integration density of quantum chips. Moreover, wiring accounts for the largest portion of the investment in superconducting quantum hardware, taking around 80% of the total cost [31]. For instance, the hardware investment for a 150-qubit quantum system is approximately $5M, with $4M spent on wiring [31]. As a result, reducing the number of wiring controls becomes a critical task to extend the scale of quantum devices further.

The control wiring of a superconducting qubit includes XY- Z- and readout lines. The reduction can be achieved by CMOS-based cryogenic control [41] and quantum signal multiplexing [2, 13]. Considering the large-scale quantum control and cooling budget, multiplexing optimization for superconducting quantum chips is a feasible solution, as it allows for the collinear transmission of control pulses [23, 38]. Quantum signal multiplexing is generally categorized into two types: (i) frequency division multiplexing (FDM) [7, 17, 25], which superimposes the control signals with different frequencies in a single wiring [13]. (ii) time division multiplexing (TDM) [2, 9, 18, 19], which adopts time-separated control through a demultiplexer (DEMUX) [2] or multi-level switch [2, 19, 38].

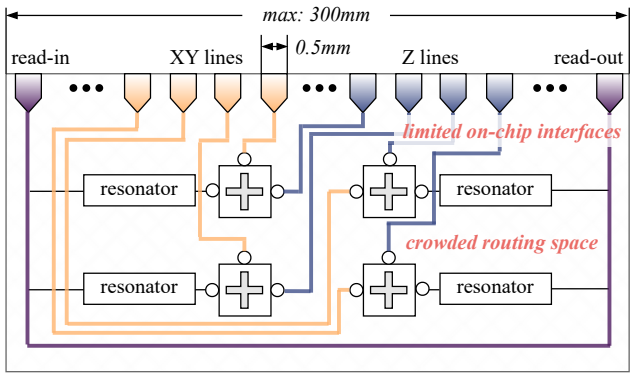
Current multiplexing inevitably faces the dilemma that the superconducting chip wiring scheme is decoupled from the signal multiplexing solution. On the one hand, the qubits that rely on the same DEMUX cannot be operated by gate simultaneously. For instance, in the case of an 8-qubit Deutsch-Jozsa circuit, using a 1:4 DEMUX increases the circuit latency by 2.1×, making fidelity drop from 87.6% to 77.3%. On the other hand, when considering frequency division multiplexing, the unoptimized grouping strategy, for the XY line of the processor, produces high crosstalk due to pulse leakage on the same coaxial cable or from the qubits of adjacent lines. For example, when the transmon qubit groups share the same frequency settings, the combination of multiple qubits results in heavier crosstalk. The fidelity of parallel X-gate operations drops to 98.9%.

In this paper, we tackle the challenges by introducing YOUTIAO. Instead of local clustering, YOUTIAO conceptualizes the superconducting chip layout as multipartite subgraphs, which represent noise-aware grouping optimization zones for signal multiplexing, including FDM on XY/readout lines and TDM on Z lines. Then we use a crosstalk characterization model to guide qubit grouping for FDM control and two-level noise-aware frequency allocation. For TDM control, we exploit topological and noisy non-parallel operations to group qubits and couplers with a greedy search-based graph coloring approach, maintaining gate parallelism while achieving time-separated signal multiplexing. The contributions of this paper are summarized as follows:

- We propose YOUTIAO, a multiplexing control system that hybridizes FDM and TDM control with an efficient qubit grouping method, which attempts to address the scalability challenge by reducing superconducting quantum wiring burden.



(a) Cryostat-level wiring architecture.



(b) Chip-level control line routing.

Figure 1: Wiring breakdown of superconducting quantum hardware.

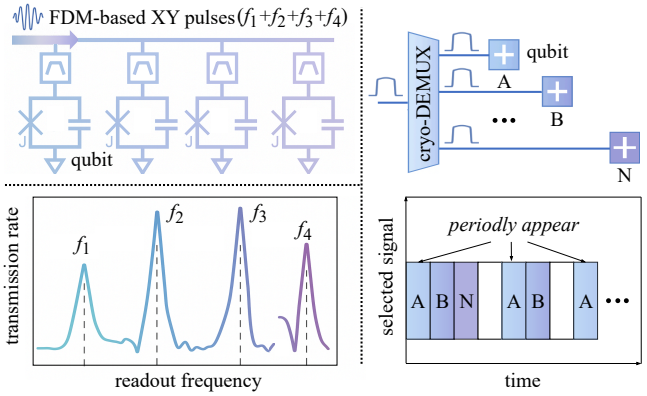
- We propose the first grouping method that exploits natural non-parallel operations to reduce additional circuit depth for TDM control.
- We propose a generative superconducting chip partition scheme to reduce the search space for grouping on large-scale quantum chips.

The state-of-the-art FDM approach is presented by George et al. [13] for multiplexing superconducting readout control lines. For TDM-based control, the leading work is by Acharya et al. [2], featuring a cryogenic demultiplexer. Experiments suggest that YOUTIAO achieves a 67.7% reduction in cryostat-level coaxial wiring complexity and overall superconducting quantum system costs, while reducing on-chip routing area by 23%. These improvements are achieved while maintaining a single-qubit gate fidelity of 99.98% and incurring only a 1.05× increase in circuit latency compared to partial-multiplexing systems such as Google Sycamore [4, 36], which adopts readout multiplexing only.

2 Background

2.1 Superconducting Quantum Wiring

In superconducting quantum architectures, qubits are stored inside cryostats. The wiring system serves as a bridge to transmit analog signals to on-chip qubits [1, 4, 22, 36]. This system comprises



(a) FDM-based qubit control. (b) TDM-based qubit control.

Figure 2: Multiplexing-based qubit control scheme. (The frequency separation data is derived from [13].)

two types of wiring configurations: (1) Cryostat-level wiring using coaxial cables to carry analog signals from room-temperature environments to quantum chips operating at ultra-low temperatures (~20 mK). (2) Chip-level routing employing coplanar waveguides fabricated through sapphire-substrate etching techniques to direct signals from on-chip interfaces to individual qubits. However, the integration and scalability of these systems are constrained by spatial limitations.

Cryostat-level wiring. Cryostat-level wiring is designed to provide sufficiently strong coupling rates to the integrated quantum processor, while minimizing decoherence due to coupling of the quantum processor via these lines to its environment [24] under spatial limitations. Therefore, only high-density coaxial cables with low-temperature tolerance and excellent electromagnetic interference shielding are capable of operating in such an environment, but they come at an extremely high cost due to the use of high-purity materials and complex manufacturing processes [24].

Figure 1 (a) shows cryostat-level wiring architecture. One end of the cable is connected to the radio-frequency DAC (RFDAC) output interface on the pulse generator, while the other end passes through the various temperature stages of the dilution refrigerator to reach the chip layer at the bottom. Due to space limitations and strict sealing requirements, there is an upper limit to the integration density of cables. Even the advanced Bluefors KIDE cryostat allows a maximum of only 4,000 coaxial lines to operate with approximately 1,300 qubits [6]. The wiring density represents one of the key bottlenecks restricting the chip's scalability.

Chip-level routing. Chip-level routing aims to design the shortest control signal paths from the on-chip interfaces to the qubits, while minimizing signal loss, such as crosstalk [3, 5, 26, 42]. Figure 1 (b) illustrates a chip-level control line routing scheme. For a transmon qubit, three control lines are required: (1) the radio-frequency (RF) phase control line (XY line); (2) the low-frequency control line (Z line combined with the direct-current bias line); and (3) the RF readout line, which is connected to the resonators coupled with qubits, and can be shared among multiple qubits.

However, as the quantum chip scales up, it faces the challenge of limited on-chip interfaces. For the plane-integrated quantum chip, the width of each interface is approximately 0.5 mm [42]. With a maximum wafer size of 300 mm [31], only up to 2,400 interfaces can be accommodated around the perimeter of a rectangular chip. Additionally, compared to traditional chip designs, quantum chips have to integrate many large components, such as a transmon qubit with a diameter of approximately 0.65 mm and a resonator length of about 4.3 mm [42]. These large devices limit the available routing space on the chip. With a sapphire substrate and 20 μm width wiring, at a 30 μm pitch, around 3,300 lines can be routed across the wafer in a single wiring layer [31, 42]. Such a routing space bottleneck could potentially make high-density quantum chips unmanageable.

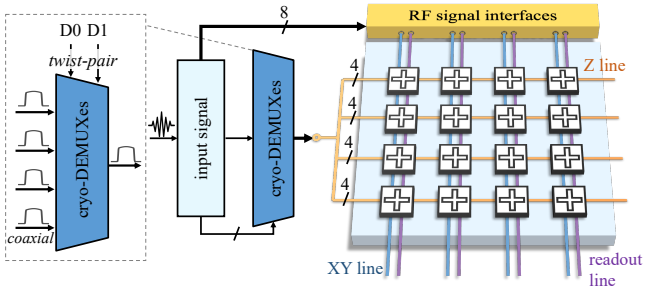


Figure 3: Multiplexing control scheme.

2.2 Quantum Signal Multiplexing

FDM-based qubit control. Frequency-division multiplexing (FDM) enables the simultaneous transmission of signals with different frequencies through shared coaxial cables. Figure 2 (a) depicts an FDM XY control architecture where four qubits share a common readout line. Typically, the XY line shares a common FDM line with dedicated bandpass filters for signal isolation, while readout multiplexing can operate without filters (Figure 2 (b)). As the frequencies of readout pulses of different qubits vary, frequency-encoded pulses are coherently combined in the control line for parallel qubit measurement. Detection efficiency mismatch (DEM) imperfections of hardware induce resonance broadening [13], necessitating meticulous frequency spacing design to suppress inter-channel crosstalk below -30 dB. Current implementations support FDM exclusively on frequency-tunable XY and readout lines, achieving single-shot readout fidelity of 99.0% [36].

TDM-based qubit control. Time-division multiplexing (TDM) shares the control lines at the temporary scale via cryogenic microwave switches (cryo-DEMUX). Figure 2 (b) showcases a TDM scheme using cryo-DEMUX at base temperature (~20 mK). As reported by Acharya et al [2], High-frequency signals can be transmitted collinearly and switched among qubits (from qubit A to qubit N) within 2.6 ns. The time-axis analysis at the bottom of Figure 2 (b) further demonstrates the control flow across different windows. All of the XY line, Z line, and readout line are compatible with TDM for mutually exclusive operations, enabling single-qubit gate applications with a fidelity of 99.9% [2].

3 Multiplexing Control System

3.1 System Fomulation

Since there are two types of pulse signals (RF signals such as XY and readout pulses and non-RF signals such as Z pulses) involved in the

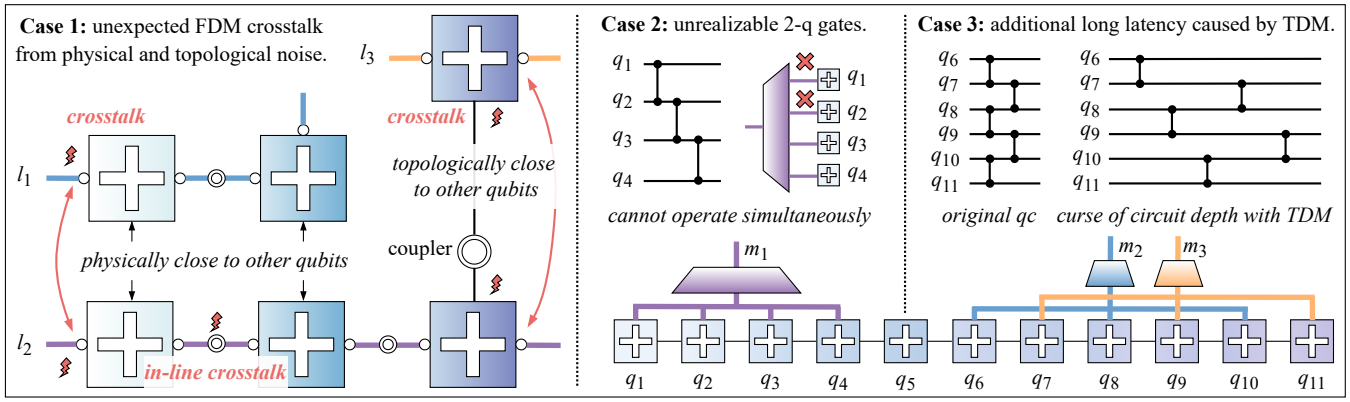


Figure 4: Motivational examples of YOUTIAO.

control of qubits, we incorporate the two types of quantum signal multiplexing techniques in a multiplexing control scheme (Figure 3) based on the frequency characteristics of control pulses. Given that the fine-tuned frequencies of different qubits are naturally configured to be different from each other within the same control line, the RF signal of the target frequency can be automatically extracted by qubit resonance from the mixed signals. Hence, FDM-based qubit control is adopted for the RF signal transmitted to the XY or readout lines by directly mixing the signal of qubits connected to the same line, and then the FDM signal is delivered by the RF signal interfaces to the corresponding line. In the figure, *"4"* represents a bundle of 4 TDM lines, while *"8"* denotes a bundle of 8 FDM lines. Notably, the demand for transmitting signals on the Z line to tune the qubit frequency is relatively sparse in temporal for each qubit. Meanwhile, the frequency of signals transmitted on Z lines is relatively low, which can hardly be distinguished in the frequency field. Hence, TDM-based qubit control is adopted for the Z line by loading in turn, and then the cryo-DEMUXes select the signal for each qubit in the corresponding time window. The cryo-DEMUX is controlled by the digital signal provided by room-temperature DACs via cheaper twisted-pair cables at cryostat-level, which can use much smaller interfaces than coaxial cables. Time-domain signal filtering is implemented through digital gating of the D0 and D1 signals.

3.2 Challenges and Opportunities

Despite the significant hardware advancements introduced in Section 2.2, directly employing unoptimized multiplexing techniques can lead to routing difficulties and imperfect qubit performance. We summarize three cases to illustrate the current challenges faced by multiplexing technologies, as shown in Figure 4:

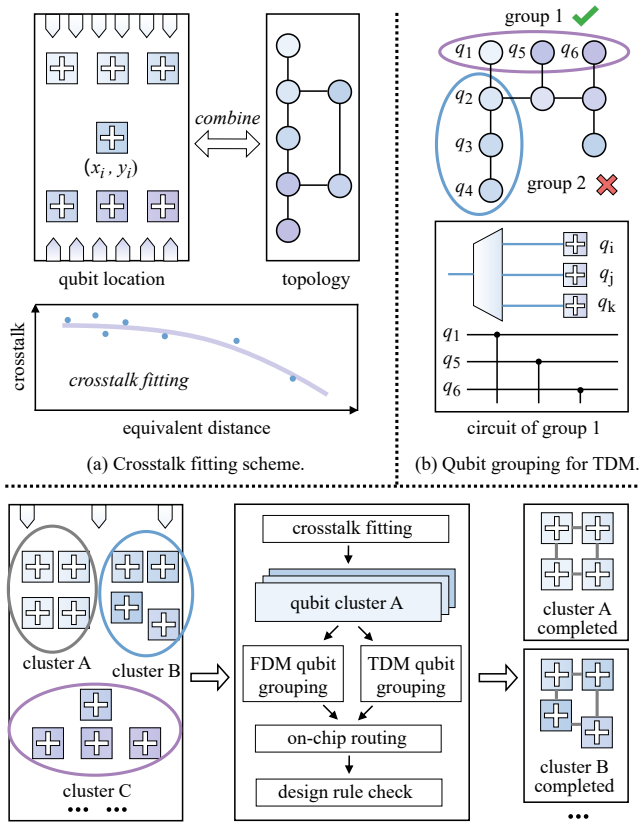
(1) **Inaccurate FDM control.** In Figure 4, Case 1, when qubits have similar frequencies (indicated by the same color in Case 1) and are positioned closely or have direct topological connections via couplers, unexpected crosstalk occurs. It also happens via pulse leakage when qubits share the same line. Previous multiplexing studies focus on reducing in-line crosstalk by frequency overlap on the same shared FDM line, failing to optimize the other two types of crosstalk [13, 38]. For example, even if qubits on FDM line l_2 are modulated at distinct frequencies, noise from adjacent FDM lines l_1 and l_3 can still induce crosstalk. Such interference sometimes causes gate or readout errors by more than 1%, or even failure [44].

(2) **Unrealizable two-qubit gates for TDM control.** In Case 2, when qubits with direct topological connections are linked to the same cryo-DEMUX m_1, it becomes impossible to operate both qubits simultaneously, leading to unrealizable two-qubit gates. For instance, when qubits q_1, q_2, q_3, and q_4 are under the TDM control scheme, all two-qubit gates in this group will fail because the DEMUX can only control one qubit at a time, while two-qubit gates generally require pulses to be applied to both qubits simultaneously. This situation significantly restricts the diversity of the qubit's ability and makes the chip no longer appropriate for circuit execution.

(3) **Additional long latency caused by TDM.** In Case 3, when two cryo-DEMUXs are employed to cross-control q_6 to q_{11}, although two-qubit gates can be applied (assuming couplers are effectively controlled), the parallelism of the quantum circuit is also compromised. For instance, in the original quantum circuit, five two-qubit gates could be executed in just two layers in around 120 ns. However, due to the picking time constraints of DEMUX, these gates must go sequentially, resulting in *curse of increased circuit depth with TDM*. Given the qubit lifetime (T_1) limitation, this additional latency increases exposure to various sources of error for qubits, thereby leading to a substantially higher error rate.

Apparently, the anharmonicity between unoptimized multiplexing techniques and the practical applications is becoming the biggest challenge. To build the bridge, a straightforward approach is to (1) first characterize the crosstalk characteristics between qubits; (2) optimize the frequency space both within and among FDM lines through two-level qubit frequency allocation to improve fidelity; (3) develop appropriate grouping strategies to enhance the parallelism of TDM-based two-qubit gate execution; and (4) design a global multiplexing routing scheme for large-scale quantum chip systems. With this idea, YOUTIAO observes several opportunities to address these challenges.

Observation 1: *Crosstalk can be approximately characterized with chip layout and topology.* There is hardly a precise noise model that captures all aspects of crosstalk [11]. However, the placement



(a) Crosstalk fitting scheme.
(b) Qubit grouping for TDM.
(c) Scalable chip routing with divide-and-conquer method.

Figure 5: Opportunities for YOUTIAO.

of qubits on a chip and their connectivity are crucial factors that determine the magnitude of crosstalk [11, 40]. As shown in Figure 5 (a), we have found that combining the qubit on-chip position (x_i, y_i) with the topological connectivity among qubits can be transformed into an *equivalent distance* for crosstalk fitting. The trained model then provides guidance for subsequent grouping strategies, making the system more tolerant to scaling.

Observation 2: *It's a better choice to group non-parallel nearby qubits for effective TDM control.* As shown in Figure 5 (b), under the rules of TDM, qubits with execution dependencies (group 2) cannot be connected to the same DEMUX. Qubits that are topologically close within a group will face an increase in circuit depth, thus necessitating the grouping of qubits with lower connectivity, such as qubits in group 1. However, considering ZZ crosstalk, qubits that are topologically distant exhibit less crosstalk and are more likely to require parallel execution. When devising a grouping strategy, it is crucial to balance qubit connectivity while minimizing circuit depth and maintaining parallelism.

Observation 3: *Divide-and-conquer scheme will optimize over scalable routing.* A large-scale quantum chip can be partitioned into multiple multiplexing clusters using a divide-and-conquer approach to avoid an excessively large parameter space and facilitate scalability. Figure 5 (c) illustrates how the chip is partitioned into numerous clusters, with FDM and TDM strategies implemented within each cluster, followed by routing optimization and design

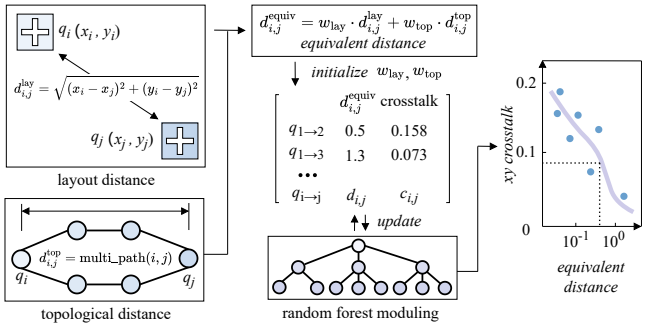


Figure 6: Crosstalk fitting with qubit layout and topology.

rule check (DRC). By combining these clusters, the superconducting chip effectively ensembles signal multiplexing routing in parallel.

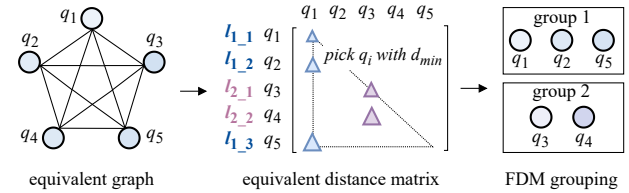
4 Method

4.1 Crosstalk Characterization

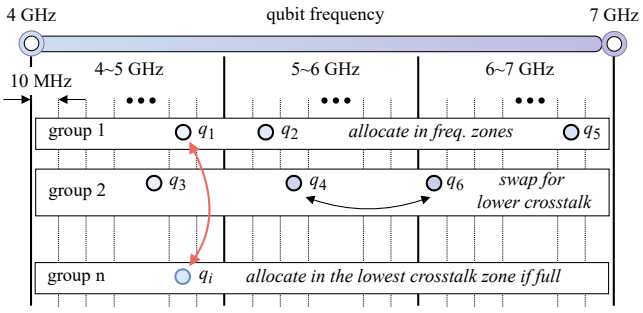
The insight of designing a crosstalk characterization model is to understand the crosstalk characteristics and incorporate them as a quantified constraint in the multiplexing-aware qubit grouping and routing procedures, thereby achieving better chip performance. Taking the superconducting Xmon quantum chip as an example, imperfections in the chip substrate and frequency collisions between adjacent qubits can lead to crosstalk [11, 21, 39], causing unexpected resonances during the application of quantum pulses while performing gate operations. In general, qubits that are physically close to each other or have topological connections with similar frequencies are more likely to experience crosstalk. With suppressed minimal two-level system (TLS) defects, an approximate fitting relationship for the chip's crosstalk can be derived.

For crosstalk fitting, we utilize a self-developed quantum processor (details are introduced in Section 5.1) to obtain qubit layout, topological structure, and crosstalk data. Since establishing a crosstalk fitting relationship in isolation is insufficient and burdens subsequent optimization, YOUTIAO adopts a joint matrix representation method that combines qubit physical distance and connectivity, using an equivalent distance to fit the crosstalk data.

As shown in the Figure 6, according to the layout of the quantum chip, the physical distance $d_{i,j}^{\text{phy}}$ between qubits q_i and q_j can be expressed as $d_{i,j}^{\text{phy}} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, where q_i and q_j are located in (x_i, y_i) and (x_j, y_j). Simultaneously, considering chip topology, the topological distance $d_{i,j}^{\text{top}}$ between q_i and q_j can be defined as the multiple shortest path distance derived using Dijkstra's algorithm. A topological distance of 1 indicates that q_i and q_j are adjacent, while a topological distance of 2 implies that q_i and q_j are separated by a qubit. When there are n paths for the minimum length l, the value of d is set to $d_{i,j}^{\text{top}} = nl$. This is because multi-path metrics are more robust, especially for chips arranged in a square topology. Although the crosstalk effect decays exponentially with increasing topological distance, different paths with the same topological distance can still lead to varying levels of crosstalk. Therefore, we can calculate the equivalent distance



(a) Qubit grouping scheme for FDM control.



(b) Two-level frequency allocation for FDM control.

Figure 7: Noise-aware FDM qubit grouping.

$d_{i,j}^{\text{equiv}}$ as follows:

$$d_{i,j}^{\text{equiv}} = w_{\text{phy}} \cdot d_{i,j}^{\text{phy}} + w_{\text{top}} \cdot d_{i,j}^{\text{top}}$$

where w_{phy} and w_{top} represent the weight between physical distance and topological distance, which can be determined by the fitting model. We generate XY and ZZ crosstalk data from our real-world superconducting quantum chip. For XY crosstalk, we collect the probability of energy level transitions of uncontrolled qubits when operating gates on the target qubits. For ZZ crosstalk, we calibrate the frequency shift of the uncontrolled qubits. We use random forest algorithm to fit the relationship between $d_{i,j}^{\text{equiv}}$ and the crosstalk, and employ mean squared error $E(a,b) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ to measure the discrepancy between predicted crosstalk y_i and actual one $\hat{y}_i$. Using 5-fold cross-validation, we train to obtain the optimal parameter set $(w_{phy}^{best}, w_{top}^{best})$ along with the crosstalk characterization model that yields the minimum error. The fitted crosstalk model provides valuable information for chip routing with similar substrates and processes.

4.2 FDM Grouping Strategy

Considering incremental quantum chip routing, the additional constraint introduced by the FDM-based control scheme during routing is to determine which qubits should share the same line. The previous FDM schemes are prone to crosstalk because they only focus on in-line frequency allocation, while neglecting nearby qubits with similar frequencies. Therefore, the FDM grouping strategy should include two steps: noise-aware grouping and two-level coarse-grained frequency allocation, to ensure accurate control.

Noise-aware qubit grouping. In FDM, to prevent pulse leakage, the frequency spacing between qubits on the same line is typically adjusted to be large to meet the cryogenic bandpass filters' requirement. As a result, the insightful suggestion is that qubits with adjacent physical or topological distances should be allocated to the same FDM line, because they should be naturally separated in frequency during the chip design. We utilize the equivalent distance matrix and apply a greedy search algorithm to determine the qubit groups for FDM lines, as shown in Figure 7 (a). The equivalent distance between qubits can be modeled as an *equivalent graph*, where vertices represent the qubits q_1, q_2, q_3, q_4, and q_5, and edges represent their distance $d_{i,j}^{\text{equiv}}$. The equivalent distance matrix is regarded as the adjacency representation of the equivalent graph.

We put forward the 3-step grouping flow for a single FDM line with a capacity of 3 qubits. (1) Grouping of FDM line l_1 starts randomly with q_1. (2) The qubit q_2, which has minimum equivalent distance $d_{1,2}^{\text{equiv}}$ to q_1, is added to *group 1*. (3) Next, the qubit q_5, which is the closest to q_1 with distance of $d_{1,5}^{\text{equiv}}$, is compared to q_3, the qubit closest to q_2 with distance of $d_{2,3}^{\text{equiv}}$. The shortest of the two distances $d_{1,5}^{\text{equiv}}$ and $d_{2,3}^{\text{equiv}}$ is selected, and q_5 is added to *group 1*. This process is then repeated to form *group 2* of l_2, which finally contains q_3 and q_4.

Two-level coarse-grained frequency allocation. Even for frequency-tunable transmon qubits, the frequency range adjusted via the Z-line is quite limited (typically within 50 MHz). The base frequency of the qubits and their coupling strength are determined during fabrication. To ensure the practical usability of the chip, we still need to check that qubits within the same group, as well as between groups, do not experience significant crosstalk due to closely spaced frequencies before routing.

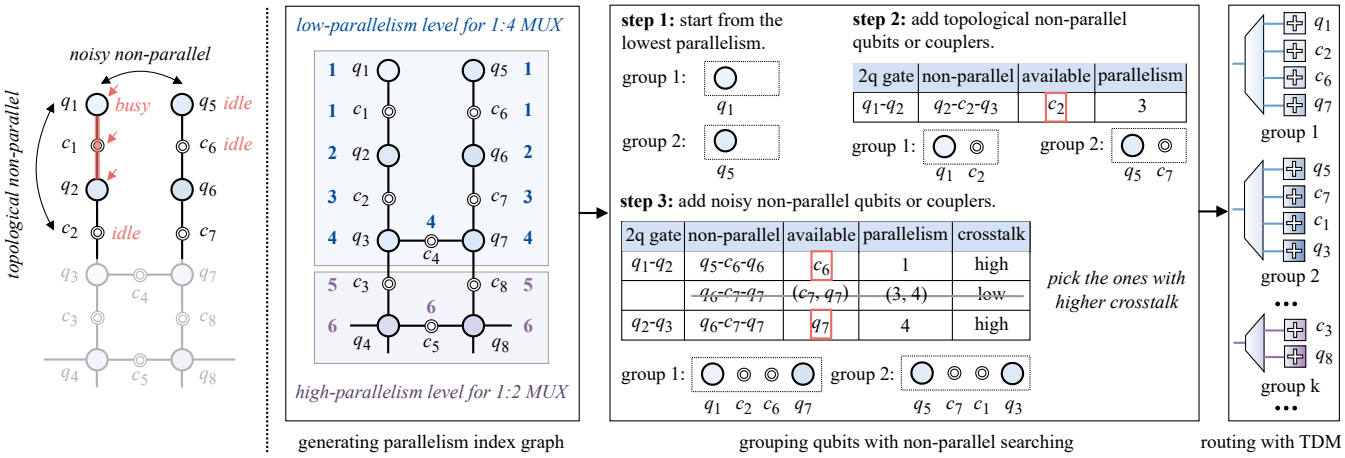
Figure 7 (b) illustrates a noise-aware two-level frequency allocation assumption based on heuristic search. According to the FDM line capacity, we divide the effective frequency range for qubits (4~7 GHz) into three zones: 4~5 GHz, 5~6 GHz, and 6~7 GHz. Each zone is further subdivided into frequency cells with a 10 MHz spacing. Qubits within the same group are allocated to different zones, ensuring large frequency spacing for in-line frequency allocation. For example, q_1, q_2, and q_3 are assigned to different zones. When allocating qubits in *group 2*, three constraints must be satisfied:

(1) Similarly, the three qubits should be allocated to different zones.
(2) Within each zone, qubits should be assigned to unoccupied frequency cells. For instance, q_4 is assigned to a cell in the 5~6 GHz frequency zone, different from the one assigned to q_2.
(3) The lowest crosstalk can be achieved by swapping qubits within the group according to the crosstalk model, such as swapping q_4 and q_6 for lower crosstalk.

For large-scale quantum chips, frequency crowding may occur. In such cases, when selecting a cell for qubit q_i in *group n*, the choice should prioritize cells with the lowest crosstalk, based on their equivalent distance $d_{i,j}^{\text{equiv}}$. For example, within the 4~5 GHz zone, q_i and q_1 have minimal crosstalk and can be located in the same cell for frequency reuse.

4.3 TDM Grouping Strategy

The biggest challenge for TDM control is maintaining the legality and parallelism of qubit control. Determining which qubits are connected to the same cryo-DEMUX is the primary issue in TDM routing. A key insight is that circuits inherently exhibit non-parallelism



(a) Natural non-parallel operations.

(b) TDM routing strategy.

Figure 8: Noise-aware TDM qubit grouping.

during execution, and this natural time separation is tailored for TDM control, orchestrating with the operational characteristics of the cryo-DEMUX. With this idea, better grouping can be achieved by identifying the non-parallel relationships between qubits and couplers.

Figure 8 (a) illustrates the two types of natural non-parallel operations on the topology graph. (1) *Topological non-parallelism*: For instance, during a two-qubit gate operation $q_1 - c_1 - q_2$, the qubits q_1, q_2, and coupler c_1 receive square pulses to tune to the same frequency. While they are busy, q_2 is occupied, preventing the execution of the $q_2 - c_2 - q_3$ gate. Thus, during this time, c_2 is topologically isolated and remains idle. (2) *Noisy non-parallelism*: When $q_1 - c_1 - q_2$ is performing a two-qubit gate, due to adjacency in topology or physical distance, executing the $q_5 - c_6 - q_6$ gate simultaneously would result in significant crosstalk, which reduces the fidelity of the gate execution. Therefore, q_5 and c_6 must remain idle. These two natural non-parallelism characteristics provide an opportunity to mitigate the curse of circuit depth with TDM.

Additionally, qubits located at different topological positions have varying levels of connectivity. For qubits with high connectivity, TDM significantly limits their degrees of gate freedom. Considering a qubit with a connectivity of 4 connected to a 1:4 DEMUX with average gating times, the number of two-qubit gates it can perform will decrease to 1/16 of its original capacity. Therefore, YOUTIAO proposes two suggestions to maintain qubit parallelism: (1) select multi-level DEMUX configurations to handle grouping scenarios involving qubits or couplers with different parallelism; (2) group qubits or couplers that exhibit natural non-parallelism.

Figure 8 (b) illustrates the TDM grouping strategy. First, since connectivity only characterizes the degree of freedom of a qubit and does not account for the parallelism of couplers, we define a *parallelism index* to represent the non-coexistence characteristics of qubits or couplers with respect to neighboring two-qubit gates, shown as numbers in blue or purple in the figure. This index measures the number of adjacent two-qubit gates that do not coexist topologically when a qubit q_i or coupler c_i is occupied. The

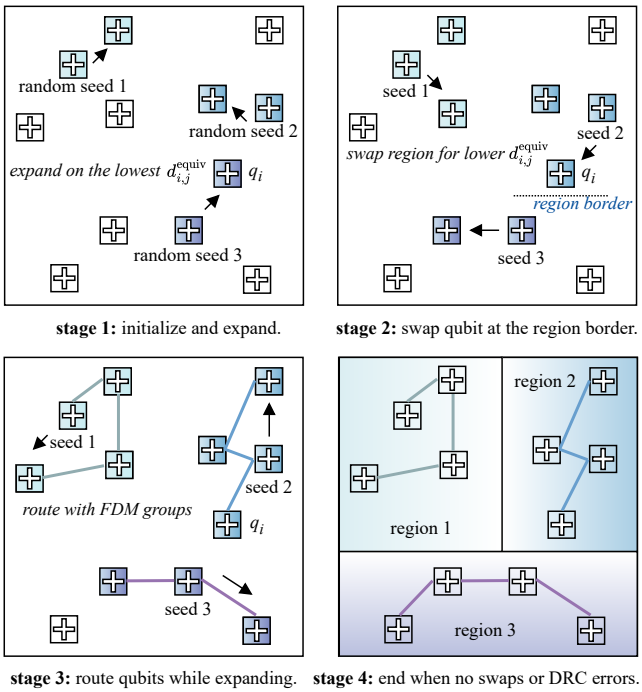


Figure 9: Generative quantum chip partition.

parallelism index is defined as:

$$parallelism\ index\ (q_i/c_i) = \frac{\sum_c non - coexisted\ 2q\ gates}{connectivity\ of\ q_i/c_i}$$

where c represents the connectivity, and the connectivity of the coupler is always adjusted to 1.

For example, when c_1 is occupied, the only two-qubit gate choice is $q_1 - c_1 - q_2$, and $q_2 - c_2 - q_3$ is the only topologically non-parallel two-qubit gate. Therefore, the parallelism index of c_1 is parallelism index$(c_1) = \frac{1}{1} = 1$. For q_3, when $q_2 - c_2 - q_3$ is active, there are 3 topologically non-parallel two-qubit gates—$q_1 - c_1 -$

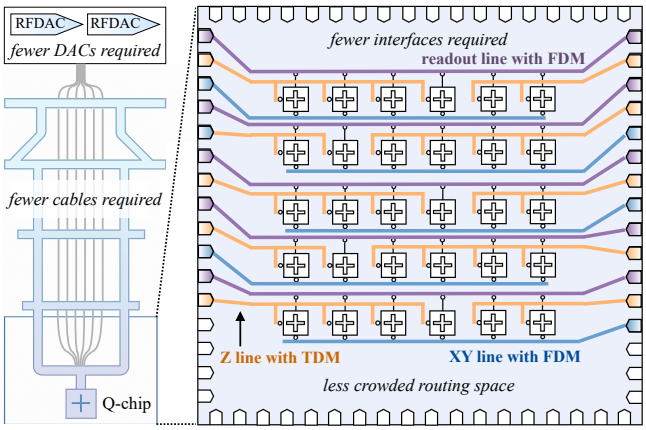


Figure 10: Multiplexing wiring system.

q_2, $q_3 - c_4 - q_7$, and $q_3 - c_3 - q_4$—that cannot operate. When q_3 is occupied for $q_3 - c_4 - q_7$ and $q_3 - c_3 - q_4$, the non-parallel gates are 4 and 5, respectively. Therefore, the parallelism index of q_3 is: parallelism index$(q_3) = \frac{(3+4+5)}{3} = 4$. A threshold θ is defined for two-level parallelism separation. In this example, the parallelism level is separated by $\theta = 4$. For qubits and couplers with low parallelism, we use a 1:4 cryo-DEMUX, while for those with higher parallelism, we use a 1:2 cryo-DEMUX to reduce the limitations on gate freedom.

After generating the parallelism index graph, for the set of qubits and couplers $[q_1, q_2, \ldots, q_n] \cup [c_1, c_2, \ldots, c_m]$, we have designed a 3-step, greedily search-based graph coloring method for TDM grouping. The design is based on the following principles: (1) Since different devices are used during routing, the grouping considers only qubits and couplers within each level. (2) Group legal qubits and couplers that are non-parallel to each other. (3) For the remaining qubits and couplers that can potentially execute in parallel, partition them with similar group parallelism indices for balancing. The grouping scheme is represented in the following steps:

- **Step 1:** Grouping starts with qubits or couplers that have the lowest parallelism. For example, qubits q_1 and q_5 with parallelism index = 1 are respectively added to *group 1* and *group 2*.
- **Step 2:** Add topologically non-parallel qubits or couplers. For instance, in *group 1*, based on the topological relationships, when q_1 is busy, the only two-qubit gate in the relevant gate list is $q_1 - c_1 - q_2$, and the only topologically non-parallel gate in the list is $q_2 - c_2 - q_3$. Since q_2 is indirectly coupled to q_1 and q_3 has other coupling relations, there is no non-coexistence with q_1. Therefore, only c_2 is available and is added to *group 1*. Similarly, c_7 is added to *group 2*.
- **Step 3:** Add noisy non-parallel qubits or couplers. In *group 1*, based on the crosstalk model, we pick the ones with higher crosstalk. $q_5 - c_6 - q_6$ and $q_6 - c_7 - q_7$ both have high crosstalk, so we add the available c_6 and q_7 to *group 1*. Similarly, the best match for *group 2* is found.

After graph coloring, TDM routing is executed for all the groups. For an n-qubit large-scale quantum chip, if the entire chip is treated as the search space and divided into k groups, the worst-case computational complexity would reach $O(n^k)$. Therefore, a chip partitioning method is required to reduce the search space for TDM grouping, allowing to perform routing within a smaller range.

4.4 Generative Chip Partition

Traditional clustering methods based on chip layout overlook the impact of crosstalk, often leading to local optima, and are not suitable for multiplexing routing scenarios. However, YOUTIAO employs a 4-stage generative chip partition scheme to set multiple routing regions for better management of multiplexing constraints. This method dynamically expands regions by randomly selecting seed points and considering topological and crosstalk constraints, while optimizing qubit grouping using TDM and FDM technologies to enhance parallelism and reduce conflicts. Through region partitioning and qubit swapping, generative chip partition ensures efficient routing and optimized execution of the chip, followed by these stages, as shown in Figure 9:

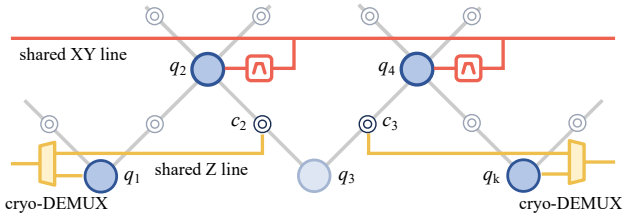
(1) stage 1: initialize and expand. To accommodate different quantum chip layouts, we randomly select multiple qubits as seeds, such as the three random seeds shown in Figure 9, and use the qubit with the lowest equivalent distance $d_{i,j}^{\text{equiv}}$ as the starting qubit for expansion. By expanding from these seeds, we gradually form the initial routing regions.
(2) stage 2: swap qubit at the region border. To avoid getting stuck in local optima during partitioning and to adhere to the FDM line routing logic, the topology of the regions is optimized by swapping qubits. As shown in the figure, qubit q_i is initially grouped with seed 3, but when seed 2 expands to q_i, due to $d_{seed_2,\, q_i}^{\text{equiv}} < d_{seed_3,\, q_i}^{\text{equiv}}$, q_i is reassigned to the region of q_2.
(3) stage 3: route qubits while expanding. Because the grouping in FDM is based on a greedy algorithm, the routing and region expansion processes are pipelined, thus improving routing efficiency.
(4) stage 4: end when no swaps or DRC errors. In the final stage, the routing process continues to expand and adjust until all qubits are successfully allocated to their respective regions. At this point, regions 1, 2, and 3 have been successfully assigned, and all qubits are controlled by valid control lines (including FDM and TDM), with no conflicts in the routing between regions. The chip's routing optimization is complete.

Based on the multiplex grouping and wire routing methods presented above, we complete the design of a multiplexing wiring system with fewer physical resources in Figure 10. Owing to the application of quantum signal multiplexing, fewer signal channels are used to transmit the control signal. Then in the perspective of physical resources, fewer DACs and cables are required in the cryostat and fewer interfaces are occupied on the quantum chip (Q-chip). Meanwhile, the routing space becomes less crowded under the same qubit placement density.

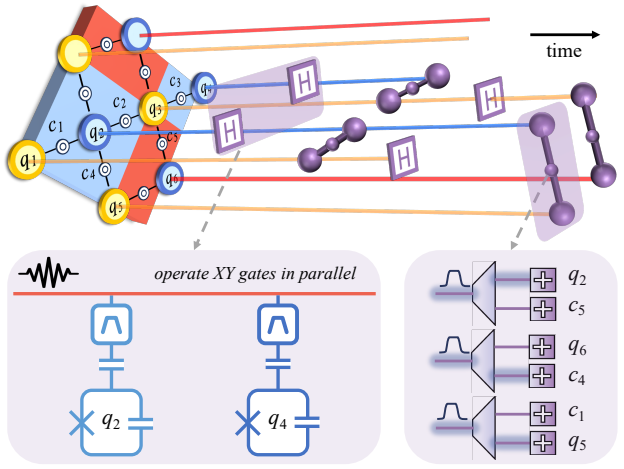
5 Evaluation

5.1 Experimental Setup

Hardware. We use two self-developed quantum chips with 36 Xmon-qubits arranged in a 6×6 grid topology and an 8×8 chip with



(a) Multiplexing in fault-tolerant chip routing.



(b) Shared control line during error correction cycles.

Figure 11: Fault-tolerant quantum chip designed with YOUTIAO. (The surface code scheme is derived from [36].)

Table 1: Wiring results of fault-tolerant quantum chip.

	distance	#XY line	#Z line	wiring cost	2q gate depth
Google [36]	3	17	41	\$413K	600
	5	49	129	\$1.25M	675
	7	97	265	\$2.53M	600
	9	161	449	\$4.26M	675
	11	241	681	\$6.43M	600
YOUTIAO	3	4	16	\$164K	625
	5	10	57	\$524K	800
	7	20	122	\$1.09M	750
	9	33	211	\$1.87M	800
	11	49	324	\$2.84M	750

the same qubit type, topology, and fabrication process. The average qubit relaxation time (T_1) is 90 μs. The basis gates of the device are RX, RY, RZ, and CZ gates. The single-qubit gates and two-qubit gates are calibrated to reach the fidelities up to 99.99% and 99.73%, respectively. The XY crosstalk and ZZ crosstalk data for each qubit are collected for analysis. Information on each sapphire-based chip topology and on-chip qubit placement is also collected.

Software configurations. All programs are implemented with `Python` (3.12.1) and the `Numpy` package (1.23.1). We use `Qiskit` to simulate the noisy execution results. We also use `Qutip` package to perform pulse-level quantum gate simulation. All software experiments are performed on an AMD Ryzen 9 9950X 4.30GHz 16-core CPU with 32GB of memory.

Baseline. We compare YOUTIAO against Google Sycamore [36], the state-of-the-art quantum wiring architecture. For the fidelity test, we compare FDM technique with the state-of-the-art FDM approach presented by George et al. [13] and TDM technique with legal qubit clustering operated with Acharya et al. [2].

Benchmark. The evaluation is performed on 5 algorithms, including Variational Quantum Classifier (VQC) [34], Linear Ising Model (ISING) [29], Deutsche-Jozsa Algorithm (DJ) [28], Quantum Fourier Transformation (QFT) [20] and Quantum K-nearest Neighbors Algorithm (QKNN) [27].

5.2 Case Study: Fault-Tolerant Chip Design

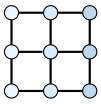
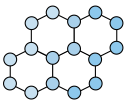
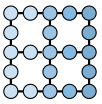
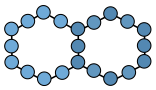
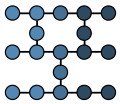
Towards the fault-tolerant quantum computing era, the design of quantum chips has to consider the deployment of quantum error correction code. Here we evaluate the feasibility of designing the fault-tolerant quantum chip with YOUTIAO. As shown in Figure 11 (a), the parity-check qubits ($q_2, q_4, \cdots$), where parallel X gates are performed, are controlled by FDM on a shared XY line. Meanwhile, the data qubits ($q_1, \cdots, q_k$) and the couplers ($c_2, c_3, \cdots$), where Z gates are performed, are controlled by TDM with the signals transmitted alternatively on the Z lines. To be more specific, the gate-performing process of the surface code is presented in Figure 11 (b). It is shown that the Hadamard gates can be operated in parallel for q_2 and q_4 in the error correction cycle. And the CZ gate for $q_2 - c_4 - q_5$ can be efficiently performed with TDM multiplexing without additional circuit depth.

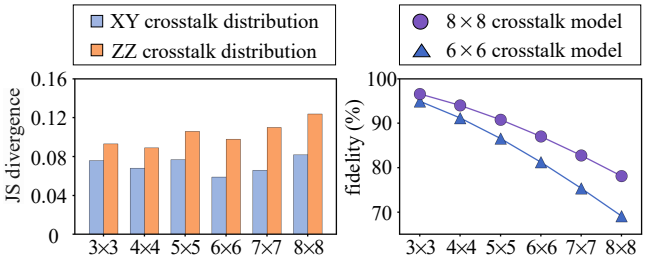
Compared with the quantum chip design method employed by Google [36], Table 1 presents the wiring results for fault-tolerant quantum chips that support surface code. The metrics are evaluated for the realization of a single logic qubit. YOUTIAO achieves a 2.35× reduction in wiring burden, driving the cost down from \$6.43M to \$2.84M for the chip supporting surface code with distance=11. YOUTIAO results in only a 1.18× increase in the two-qubit gate depth for TDM control within a 25-cycle error correction circuit, effectively balancing resource reuse and circuit execution efficiency. YOUTIAO boosts the suppression of wiring cost in view of a certain operation mode of the surface code.

5.3 Quantum Wiring

Table 2 illustrates the cryostat-level and chip-level wiring analysis of the 5 topologies, including square, heavy square, hexagon, heavy hexagon, and low-density qubit arrangement. At cryostat-level, overall, YOUTIAO achieves 3.1× in both cryostat-level coaxial wiring reduction and wiring cost reduction compared to Google's state-of-the-art wiring architecture [36], driving the economic burden from \$470K down to \$151K for a 21-qubit quantum wiring system. Additionally, YOUTIAO obtains 4.2× XY line reduction and 3.7× Z line reduction on average with FDM line capacity of 5 qubits and TDM line using 1:2 and 1:4 cryo-DEMUXes. This improvement can be attributed to the effective integration of FDM on the XY lines and TDM on the Z lines. With technological advances, new materials have reduced the cost of coaxial cables [8, 24], though they remain expensive. YOUTIAO offers a cost-effective and orthogonal

Table 2: Evaluation of quantum wiring system.

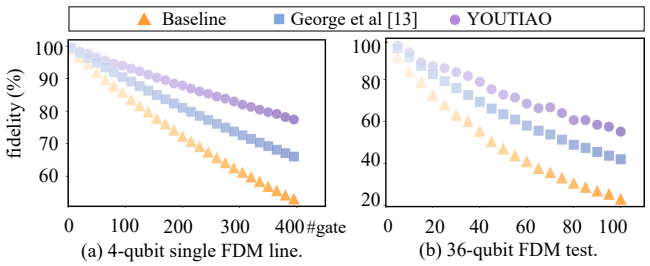
topology		Square		Hexagon		Heavy Square		Heavy Hexagon		Low-density	
											
		Google [36]	YOUTIAO	Google [36]	YOUTIAO	Google [36]	YOUTIAO	Google [36]	YOUTIAO	Google [36]	YOUTIAO
cryostat level	#qubit	9		16		21		21		18	
	#XY line	9	2 (4.5×)	16	4 (4×)	21	5 (4.2×)	21	5 (4.2×)	18	4 (4.5×)
	#Z line	37	7 (3×)	35	9 (3.9×)	45	12 (3.8×)	43	11 (3.9×)	36	9 (4×)
	DEMUX control	-	11	-	18	-	24	-	22	-	18
	#DAC	33	23 (1.4×)	55	35 (1.6×)	72	47 (1.5×)	70	44 (1.6×)	59	36 (1.6×)
	wiring cost	$216K	$79K (2.8×)	$359K	$111K (3.3×)	$470K	$151K (3.2×)	$457K	$143K (3.2×)	$385K	$118K (3.3×)
chip level	#interface	32	22 (1.4×)	53	33 (1.6×)	69	44 (1.6×)	67	41 (1.6×)	57	34 (1.7×)
	routing area (mm^2)	4.62	3.68 (1.2×)	7.98	6.12 (1.3×)	10.15	7.97 (1.3×)	9.20	7.07 (1.3×)	8.30	6.47 (1.3×)



(a) JS divergence of crosstalk characterization for two chips.

(b) FDM fidelity test at different chip scales.

Figure 12: Crosstalk model generality on similar chips.



(a) 4-qubit single FDM line.

(b) 36-qubit FDM test.

Figure 13: Evaluation of FDM grouping with random gates.

wiring solution that continues to deliver practical benefits even as technology evolves. By bridging the quantum co-processor (FPGA) with FDM-based and TDM-based quantum devices, this co-linear transmission approach significantly reduces the high-density coaxial cable cost in dilution refrigerators. Consequently, the same cable density can control more qubits, addressing the wiring bottleneck and enhancing the qubit density on-chip. As shown in the table, different topologies achieve varying degrees of line reduction. Compared to the square topology, the low-density topology achieves the greatest reduction in wiring costs. This is because qubits in the low-density topology generally exhibit lower parallelism indices, making them more suitable for multiplexing with higher-density 1:4 DEMUXes. Considering twist-pair control lines (a lot cheaper than coaxial cables [8, 24]) for DEMUX digital control signal transmission, the control line overhead is also evaluated, as well as DACs and on-chip interfaces. In contrast, qubits within the square topology require higher parallel freedom, resulting in a relatively lower ratio of line reduction.

At chip-level, overall, YOUTIAO achieves 1.3× in routing cost with 20 μm width line, at a 30 μm pitch under the same fabrication condition as the target 36-qubit chip. YOUTIAO also reduces on-chip signal interfaces by 1.6× on average. This experiment is implemented using path-based simulations, where routing paths are represented by a grid with a resolution of 10 μm. After defining the physical coordinates (denoted as (x, y)) of the devices on the chip, the shortest routing paths are determined by applying an A* algorithm, subject to standard EDA constraints—specifically, prohibiting routing intersections and maintaining adequate spacing between adjacent lines. The achieved reductions in both interface complexity and routing area are primarily attributed to the layout approach involving FDM-based XY lines, which reduces the requirement for extensive XY-plane wiring on-chip, and TDM-based Z lines, which substantially minimize the interface demands.

5.4 Evaluation of FDM Grouping.

Figure 12 illustrates the generality of the crosstalk model on a similar chip, characterized by the same qubit type, topology, and fabrication process. In Figure 12 (a), we train crosstalk models separately on 6×6 and 8×8 chips, and generate the corresponding predicted noise distributions. The Jensen–Shannon (JS) divergence between the two distributions reaches a minimum of 0.06, indicating a high degree of similarity, suggesting that the crosstalk model demonstrates generality across similar hardware.

Figure 12 (b) shows the result of applying the crosstalk model trained on the 6×6 chip to an 8×8 chip for FDM-based qubit grouping, followed by a fidelity test. On each 4-qubit FDM line, we execute 10 layers of random XY gates per qubit and measure the fidelity at various scales. Overall, FDM achieves 99.94% single-qubit gate fidelity across all tested scales. In comparison, using a noise model trained directly on the 8×8 chip achieves 99.96% fidelity. These results demonstrate that the crosstalk model retains generality with acceptable fidelity loss. However, the fidelity degradation becomes

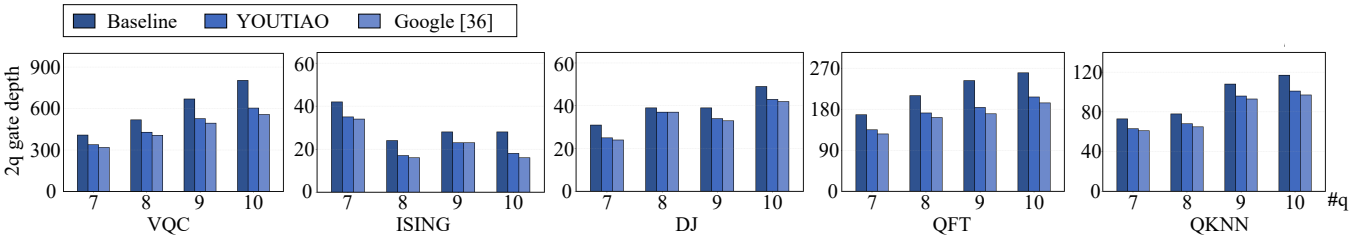


Figure 14: Two-qubit gate depth optimized with TDM grouping across 5 benchmarks.

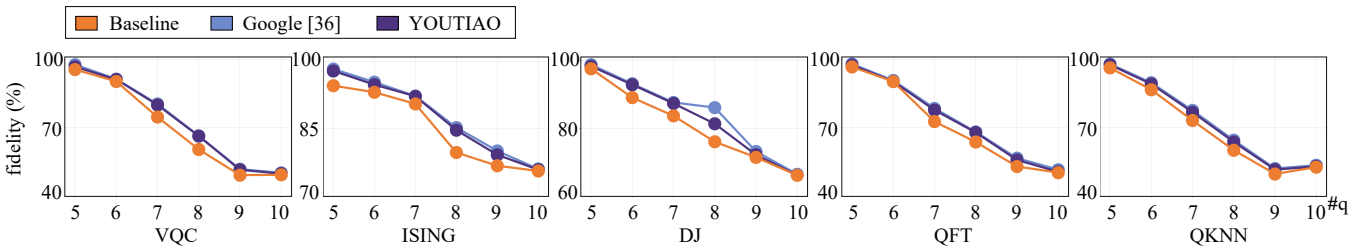


Figure 15: Circuit fidelity optimized with TDM-based routing across 5 benchmarks.

more pronounced as the scale increases, highlighting the need for more accurate and robust noise models in the future.

Figure 13 (a) depicts the fidelity of random single-qubit gates operated on a 4-qubit FDM line on the target 36-qubit chip. Overall, YOUTIAO achieves 1.37× FDM fidelity improvement with the state-of-the-art FDM practice [13] considering in-line crosstalk. YOUTIAO also achieves a 2.25× improvement in fidelity compared to the baseline, which employs an unoptimized FDM approach with chip-local clustering. Specifically, single-qubit gate fidelity reaches 99.98% in YOUTIAO, compared to an average fidelity of 99.96% for George et al. [13], as verified through Qutip-based pulse simulations. These simulations incorporate realistic parameters from the actual quantum processor, including fine-tuned qubit parameters such as frequencies, pulse attributes (e.g., pulse amplitude), and measured XY crosstalk data. The observed fidelity enhancement results from the proposed FDM grouping strategy, effectively mitigating unexpected FDM crosstalk at the chip routing stage.

Additionally, Figure 13 (b) presents the fidelity of random single-qubit gates executed across a 36-qubit quantum processor with 9 FDM lines. After 100 gate layers, the fidelity of the baseline method declines significantly to 22.9%, whereas YOUTIAO maintains a fidelity of 55.1%, thereby highlighting its robust capability to resist crosstalk-induced performance degradation.

5.5 Evaluation of TDM Grouping.

Figure 14 presents two-qubit gate depth analysis for 5 benchmark circuits implemented with Google's architecture [36], YOUTIAO, and state-of-the-art TDM practice [2], considering local clustering. Overall, YOUTIAO achieves a 1.23× reduction in two-qubit gate layer depth compared to the baseline, while incurring only a modest 1.05× increase relative to Google's architecture. The effectiveness of YOUTIAO stems from its well-balanced trade-off between TDM-based wiring reduction and circuit depth minimization, enabled by leveraging the intrinsic non-parallel characteristics during circuit execution through the optimized TDM grouping strategy. Specifically, by strategically grouping non-parallel qubits and couplers,

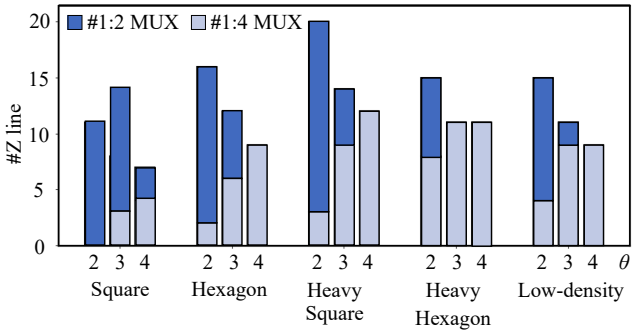
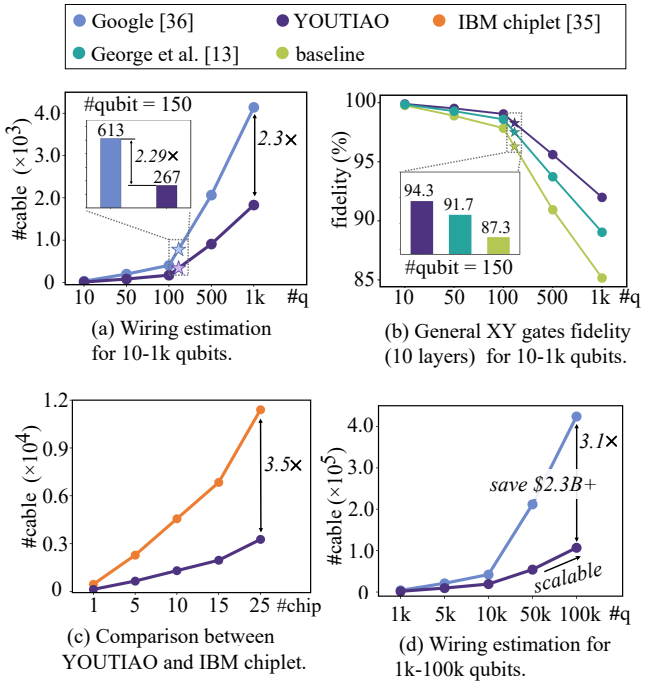


Figure 16: Cryo-DEMUX proportion for various topologies.

YOUTIAO successfully maintains parallelism of two-qubit gate operations while simultaneously reducing the wiring overhead. For instance, in highly parallelizable circuits such as the VQC, YOUTIAO demonstrates an even more pronounced advantage, achieving a 1.36× improvement in two-qubit gate layer depth compared to the Acharya et al.

Figure 15 presents the circuit fidelity results for TDM-based routing across 5 benchmarks. Overall, YOUTIAO achieves a 1.23× improvement in fidelity compared to Acharya et al. [2], while incurring a slight fidelity degradation of 1.06× relative to Google's architecture [36]. This improvement primarily arises from the proposed noise-aware TDM grouping strategy, which significantly reduces the concurrent execution of high-crosstalk two-qubit gates. Moreover, the strategy of grouping non-parallel qubits and couplers effectively mitigates the detrimental effects associated with excessive circuit depth, thus minimizing qubit exposure to potential noise sources. However, the unavoidable increase in circuit depth still introduces a modest fidelity loss due to decoherence. Fortunately, this reduction remains relatively small, demonstrating that YOUTIAO effectively balances multiplexing efficiency with fidelity preservation.



(a) Wiring estimation for 10-1k qubits.

(b) General XY gates fidelity (10 layers) for 10-1k qubits.

(c) Comparison between YOUTIAO and IBM chiplet.

(d) Wiring estimation for 1k-100k qubits.

Figure 17: Wiring estimation for the large-scale quantum system.

5.6 System Scalability

Different qubit topologies. Figure 16 illustrates the usage proportions of different-level cryo-DEMUXes across various topologies for different parallelism thresholds θ. It is evident that the qubit connectivity within the topology plays a significant role in this behavior. For instance, the square topology, which exhibits the highest parallelism, consistently utilizes the largest proportion of 1:2 cryo-DEMUXes. By adjusting the parallelism threshold, a trade-off between the Z-line multiplexing efficiency and gate parallelism can be made, ultimately optimizing the benefits of the TDM grouping strategy.

Large-scale quantum system. Figure 17 (a) illustrates the wiring estimation for large-scale quantum systems with a square topology [35]. When the system scales up to 10-1k qubits, YOUTIAO will reduce the number of coaxial cables in Google's wiring architecture [36] for over 2.3×. For a 150-qubit system, the coaxial cables budget will decrease from 613 to 267, with fidelity maintaining 94.3% for XY gates on all qubits, as shown in Figure 17 (b). Figure 17 (c) conducts the comparison between YOUTIAO and IBM's chiplet scale-out strategy [35]. Overall, YOUTIAO achieves 3.4× cable reduction compared to IBM chiplet. When the IBM chiplet framework scales to the interconnected 25 copies of 133-qubit chips, YOUTIAO can achieve a 3.5× reduction in cable usage, significantly lowering the wiring cost within the cryogenic system.

As shown in Figure 17 (d), when the system scales up to 1k-100k qubits, YOUTIAO offers 3.1 × cable reduction, driving from 4.4×10^5 to 32% of the original count. This reduction not only increases wiring density but also saves over $2.3 billion. This achievement is attributed to the optimized multiplexing collinear transmission scheme proposed by YOUTIAO. As the number of qubits increases, the cost savings achieved through multiplexing become more significant.

6 Related Work

Signal multiplexing techniques, including FDM and TDM, are widely recognized as key strategies for addressing the wiring bottleneck and power consumption challenges in scalable quantum computing. Previous research has primarily focused on applying individual multiplexing schemes to specific tasks, achieving significant progress. For instance, FDM has been widely and successfully adopted for parallel readout of superconducting qubits, substantially reducing the number of readout lines [13, 17]. With the parametric amplifiers fabrication technique and in-line frequency allocation fine-tuning, the FDM-based readout line is capable of controlling up to 8 qubits at a time [13]. Meanwhile, TDM techniques based on cryogenic CMOS switches have demonstrated strong potential as efficient channel-switching solutions for qubit control [2, 19]. De Korte et al. [9] also used balanced SQUID pairs and digital flux-lock loops in a 32-channel time-division multiplexer, achieving improved noise performance and efficient channel multiplexing for superconducting sensor arrays.

Additionally, several works have proposed shared-control-line architectures—such as row-column addressing schemes—to enable large-scale parallel qubit addressing with sublinear resource scaling [7, 9, 18, 25]. However, these pioneering efforts often treat FDM or TDM as isolated, component-level solutions or propose a single mode of multiplexing, lacking a unified system-level architecture capable of co-optimized design based on signal physical characteristics. This disjointed design approach leads to mismatches between signal types and multiplexing strategies, ultimately limiting overall performance. In contrast, YOUTIAO introduces a hybrid multiplexing control system architecture with qubit grouping. Rather than merely combining existing techniques, YOUTIAO constructs a signal-driven multiplexing system through detailed analysis of signal properties, and enhances overall system performance via noise-aware frequency allocation and non-parallel searching.

7 Conclusion

Scalability is one of the main bottlenecks faced by current quantum systems. However, the density and high economic costs of qubit control wiring are the primary factors limiting current system integration. Multiplexing is a promising approach to address the current wiring bottleneck, as it allows for the collinear transmission of control pulses. However, it also faces crosstalk issues and the curse of circuit depth. In this paper, we propose YOUTIAO to introduce a multiplexing wiring system that integrates noise-aware FDM and TDM qubit grouping for multiplexing to reduce the wiring cost. YOUTIAO also conceptualizes the chip topology as multi-partite subgraphs to further reduce the on-chip routing area.

Acknowledgments

This work was funded by the National Key Research and Development Program of China (No. 2023YFF0905200) and Zhejiang Pioneer (Jianbing) Project (No. 2023C01036). This work was also supported by the National Natural Science Foundation of China (No.62472374).

References

[1] Rajeev Acharya, Dmitry A. Abanin, Laleh Aghababaie-Beni, Igor Aleiner, Trond I. Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Nikita Astrakhantsev, Juan Atalaya, Ryan Babbush, Dave Bacon, Brian Ballard, Joseph C. Bardin, Johannes Bausch, Andreas Bengtsson, Alexander Bilmes, Sam Blackwell, Sergio Boixo, Gina Bortoli, Alexandre Bourassa, Jenna Bovaird, Leon Brill, Michael Broughton, David A. Browne, Brett Buchea, Bob B. Buckley, David A. Buell, Tim Burger, Brian Burkett, Nicholas Bushnell, Anthony Cabrera, Juan Campero, Hung-Shen Chang, Yu Chen, Zijun Chen, Ben Chiaro, Desmond Chik, Charina Chou, Jahan Claes, Agnetta Y. Cleland, Josh Cogan, Roberto Collins, Paul Conner, William Courtney, Alexander L. Crook, Ben Curtin, Sayan Das, Alex Davies, Laura De Lorenzo, Dripto M. Debroy, Sean Demura, Michel Devoret, Agustin Di Paolo, Paul Donohoe, Ilya Drozdov, Andrew Dunsworth, Clint Earle, Thomas Edlich, Alec Eickbusch, Aviv Moshe Elbag, Mahmoud Elzouka, Catherine Erickson, Lara Faoro, Edward Farhi, Vinicius S. Ferreira, Leslie Flores Burgos, Ebrahim Forati, Austin G. Fowler, Brooks Foxen, Suhas Ganjam, Gonzalo Garcia, Robert Gasca, Élie Genois, William Giang, Craig Gidney, Dar Gilboa, Raja Gosula, Alejandro Grajales Dau, Dietrich Graumann, Alex Greene, Jonathan A. Gross, Steve Habegger, John Hall, Michael C. Hamilton, Monica Hansen, Matthew P. Harrigan, Sean D. Harrington, Francisco J. H. Heras, Stephen Heslin, Paula Heu, Oscar Higgott, Gordon Hill, Jeremy Hilton, George Holland, Sabrina Hong, Hsin-Yuan Huang, Ashley Huff, William J. Huggins, Lev B. Ioffe, Sergei V. Isakov, Justin Iveland, Evan Jeffrey, Zhang Jiang, Cody Jones, Stephen Jordan, Chaitali Joshi, Pavol Juhas, Dvir Kafri, Hui Kang, Amir H. Karamlou, Kostyantyn Kechedzhi, Julian Kelly, Trupti Khaire, Tanuj Khattar, Mostafa Khezri, Seon Kim, Paul V. Klimov, Andrey R. Klots, Bryce Kobrin, Pushmeet Kohli, Alexander N. Korotkov, Fedor Kostritsa, Robin Kothari, Borislav Kozlovskii, John Mark Kreikebaum, Vladislav D. Kurilovich, Nathan Lacroix, David Landhuis, Tiano Lange-Dei, Brandon W. Langley, Pavel Laptev, Kim-Ming Lau, Loïck Le Guevel, Justin Ledford, Joonho Lee, Kenny Lee, Yuri D. Lensky, Shannon Leon, Brian J. Lester, Wing Yan Li, Yin Li, Alexander T. Lill, Wayne Liu, William P. Livingston, Aditya Locharla, Erik Lucero, Daniel Lundahl, Aaron Lunt, Sid Madhuk, Fionn D. Malone, Ashley Maloney, Salvatore Mandrà, James Manyika, Leigh S. Martin, Orion Martin, Steven Martin, Cameron Maxfield, Jarrod R. McClean, Matt McEwen, Seneca Meeks, Anthony Megrant, Xiao Mi, Kevin C. Miao, Amanda Mieszala, Reza Molavi, Sebastian Molina, Shirin Montazeri, Alexis Morvan, Ramis Movassagh, Wojciech Mruczkiewicz, Ofer Naaman, Matthew Neeley, Charles Neill, Ani Nersisyan, Hartmut Neven, Michael Newman, Jiun How Ng, Anthony Nguyen, Murray Nguyen, Chia-Hung Ni, Murphy Yuezhen Niu, Thomas E. O'Brien, William D. Oliver, Alex Opremcak, Kristoffer Ottosson, Andre Petukhov, Alex Pizzuto, John Platt, Rebecca Potter, Orion Pritchard, Leonid P. Pryadko, Chris Quintana, Ganesh Ramachandran, Matthew J. Reagor, John Redding, David M. Rhodes, Gabrielle Roberts, Eliott Rosenberg, Emma Rosenfeld, Pedram Roushan, Nicholas C. Rubin, Negar Saei, Daniel Sank, Kannan Sankaragomathi, Kevin J. Satzinger, Henry F. Schurkus, Christopher Schuster, Andrew W. Senior, Michael J. Shearn, Aaron Shorter, Noah Shutty, Vladimir Shvarts, Shraddha Singh, Volodymyr Sivak, Jindra Skruzny, Spencer Small, Vadim Smelyanskiy, W. Clarke Smith, Rolando D. Somma, Sofia Springer, George Sterling, Doug Strain, Jordan Suchard, Aaron Szasz, Alex Sztein, Douglas Thor, Alfredo Torres, M. Mert Torunbalci, Abeer Vaishnav, Justin Vargas, Sergey Vdovichev, Guifre Vidal, Benjamin Villalonga, Catherine Vollgraff Heidweiller, Steven Waltman, Shannon X. Wang, Brayden Ware, Kate Weber, Travis Weidel, Theodore White, Kristi Wong, Bryan W. K. Woo, Cheng Xing, Z. Jamie Yao, Ping Yeh, Bicheng Ying, Juhwan Yoo, Noureldin Yosri, Grayson Young, Adam Zalcman, Yaxing Zhang, Ningfeng Zhu, Nicholas Zobrist, and Google Quantum AI and Collaborators. 2025. Quantum error correction below the surface code threshold. *Nature* 638, 7973 (2025), 920–926.

[2] R. Acharya, S. Brebels, A. Grill, J. Verjauw, Ts. Ivanov, D. Perez Lozano, D. Wan, J. Van Damme, A. M. Vadiraj, M. Mongillo, B. Govoreanu, J. Craninckx, I. P. Radu, K. De Greve, G. Gielen, F. Catthoor, and A. Potočnik. 2023. Multiplexed superconducting qubit control at millikelvin temperatures with a low-power cryo-CMOS multiplexer. *Nature Electronics* 6, 11 (2023), 900–909.

[3] Carmen G Almudever, Robert Wille, Fabio Sebastiano, Nadia Haider, and Eduard Alarcon. 2024. From designing quantum processors to large-scale quantum computing systems. In *2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 1–10.

[4] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando G. S. L. Brandao, David A. Buell, Brian Burkett, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Andrew Dunsworth, Edward Farhi, Brooks Foxen, Austin Fowler, Craig Gidney, Marissa Giustina, Rob Graff, Keith Guerin, Steve Habegger, Matthew P. Harrigan, Michael J. Hartmann, Alan Ho, Markus Hoffmann, Trent Huang, Travis S. Humble, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Paul V. Klimov, Sergey Knysh, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Mike Lindmark, Erik Lucero, Dmitry Lyakh, Salvatore Mandrà, Jarrod R. McClean, Matthew McEwen, Anthony Megrant, Xiao Mi, Kristel Michielsen, Masoud Mohseni, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Murphy Yuezhen Niu, Eric Ostby, Andre Petukhov, John C. Platt, Chris Quintana, Eleanor G. Rieffel, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Kevin J. Sung, Matthew D. Trevithick, Amit Vainsencher, Benjamin Villalonga, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Hartmut Neven, and John M. Martinis. 2019. Quantum supremacy using a programmable superconducting processor. *Nature* 574, 7779 (2019), 505–510.

[5] Bikash K Behera, Tasnum Reza, Angad Gupta, and Prasanta K Panigrahi. 2019. Designing quantum router in IBM quantum computer. *Quantum Information Processing* 18 (2019), 1–13.

[6] Bluefors. 2025. KIDE Cryogenic Platform. https://bluefors.com/products/kide-cryogenic-platform/ Accessed: 2025-03-17.

[7] Yu Chen, D. Sank, P. O'Malley, T. White, R. Barends, B. Chiaro, J. Kelly, E. Lucero, M. Mariantoni, A. Megrant, C. Neill, A. Vainsencher, J. Wenner, Y. Yin, A. N. Cleland, and John M. Martinis. 2012. Multiplexed dispersive readout of superconducting phase qubits. *Applied Physics Letters* 101, 18 (2012).

[8] Keycom Corporation. 2024. Keycom Corporation Official Website. https://keycom.co.jp/index-e.htm Accessed: 2025-06-20.

[9] Piet AJ De Korte, Joern Beyer, Steve Deiker, Gene C Hilton, Kent D Irwin, Mike MacIntosh, Sae Woo Nam, Carl D Reintsema, Leila R Vale, and Martin E Huber. 2003. Time-division superconducting quantum interference device multiplexer for transition-edge sensors. *Review of Scientific Instruments* 74, 8 (2003), 3807–3815.

[10] Nathalie P De Leon, Kohei M Itoh, Dohun Kim, Karan K Mehta, Tracy E Northup, Hanhee Paik, BS Palmer, Nitin Samarth, Sorawis Sangtawesin, and David W Steuerman. 2021. Materials challenges and opportunities for quantum computing hardware. *Science* 372, 6539 (2021), eabb2823.

[11] Yongshan Ding, Pranav Gokhale, Sophia Fuhui Lin, Richard Rines, Thomas Propson, and Frederic T Chong. 2020. Systematic crosstalk mitigation for superconducting qubits via frequency-aware compilation. In *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 201–214.

[12] Kirill G. Fedorov, Michael Renger, Stefan Pogorzalek, Roberto Di Candia, Qiming Chen, Yuki Nojiri, Kunihiro Inomata, Yasunobu Nakamura, Matti Partanen, Achim Marx, Rudolf Gross, and Frank Deppe. 2021. Experimental quantum teleportation of propagating microwaves. *Science advances* 7, 52 (2021), eabk0891.

[13] RE George, J Senior, O-P Saira, JP Pekola, SE de Graaf, T Lindström, and Yu A Pashkin. 2017. Multiplexing superconducting qubit circuit for single microwave photon generation. *Journal of Low Temperature Physics* 189 (2017), 60–75.

[14] Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. 2019. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature communications* 10, 1 (2019), 3007.

[15] Lov K Grover. 1996. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 212–219.

[16] Qiujiang Guo, Chen Cheng, Zheng-Hang Sun, Zixuan Song, Hekang Li, Zhen Wang, Wenhui Ren, Hang Dong, Dongning Zheng, Yu-Ran Zhang, Rubem Mondaini, Heng Fan, and H. Wang. 2021. Observation of energy-resolved many-body localization. *Nature Physics* 17, 2 (2021), 234–239.

[17] Rutian Huang, Xiao Geng, Xinyu Wu, Genting Dai, Liangliang Yang, Jianshe Liu, and Wei Chen. 2022. Cryogenic multiplexing control chip for a superconducting quantum processor. *Physical Review Applied* 18, 6 (2022), 064046.

[18] KD Irwin, MD Niemack, J Beyer, HM Cho, WB Doriese, GC Hilton, CD Reintsema, DR Schmidt, JN Ullom, and LR Vale. 2010. Code-division multiplexing of superconducting transition-edge sensor arrays. *Superconductor Science and Technology* 23, 3 (2010), 034004.

[19] Lei Jiang, Yu Xu, Shaowei Li, Zhiguang Yan, Ming Gong, Tao Rong, Chenyin Sun, Tianzuo Sun, Tao Jiang, Hui Deng, Chen Zha, Jin Lin, Fusheng Chen, Qingling Zhu, Yangsen Ye, Hao Rong, Kai Yan, Sirui Cao, Yuan Li, Shaojun Guo, Haoran Qian, Yisen Hu, Yulin Wu, Yuhuai Li, Gang Wu, Xueshen Wang, Shijian Wang, Wenhui Cao, Yeru Wang, Jinjin Li, Cheng-Zhi Peng, Xiaobo Zhu, and Jian-Wei Pan. 2024. In situ qubit frequency tuning circuit for scalable superconducting quantum computing: Scheme and experiment. *arXiv preprint arXiv:2407.21415* (2024).

[20] Yuwei Jin, Xiangyu Gao, Minghao Guo, Henry Chen, Fei Hua, Chi Zhang, and Eddy Z Zhang. 2023. Quantum fourier transformation circuits compilation. *arXiv preprint arXiv:2312.16114* (2023).

[21] Andreas Ketterer and Thomas Wellens. 2023. Characterizing crosstalk of superconducting transmon processors. *Physical Review Applied* 20, 3 (2023), 034065.

[22] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout van den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, and Abhinav Kandala. 2023. Evidence for the utility of quantum computing before fault tolerance. *Nature* 618, 7965 (2023), 500–505.

[23] Philip Krantz, Morten Kjaergaard, Fei Yan, Terry P Orlando, Simon Gustavsson, and William D Oliver. 2019. A quantum engineer's guide to superconducting qubits. *Applied physics reviews* 6, 2 (2019).

[24] Sebastian Krinner, Simon Storz, Philipp Kurpiers, Paul Magnard, Johannes Heinsoo, Raphael Keller, Janis Luetolf, Christopher Eichler, and Andreas Wallraff. 2019. Engineering cryogenic setups for 100-qubit scale superconducting circuit systems. *EPJ Quantum Technology* 6, 1 (2019), 2.

[25] TM Lanting, Hsiao-Mei Cho, John Clarke, Matt Dobbs, Adrian T Lee, PL Richards, AD Smith, and HG Spieler. 2003. A frequency-domain SQUID multiplexer for arrays of transition-edge superconducting sensors. *IEEE transactions on applied superconductivity* 13, 2 (2003), 626–629.

[26] Fei-Yu Li and Li-Jing Jin. 2023. Quantum chip design optimization and automation in superconducting coupler architecture. *Quantum Science and Technology* 8, 4 (2023), 045015.

[27] Jiaye Li, Jian Zhang, Jilian Zhang, and Shichao Zhang. 2024. Quantum KNN Classification With K Value Selection and Neighbor Selection. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 43, 5 (2024), 1332–1345. doi:10.1109/TCAD.2023.3345251

[28] Zhiqiang Li, Juan Dai, Suhan Pan, Wei Zhang, and Jiajia Hu. 2020. Synthesis of Deutsch-Jozsa circuits and verification by IBM Q. *International Journal of Theoretical Physics* 59 (2020), 1668–1678.

[29] Filip B. Maciejewski, Stuart Hadfield, Benjamin Hall, Mark Hodson, Maxime Dupont, Bram Evert, James Sud, M. Sohaib Alam, Zhihui Wang, Stephen Jeffrey, Bhuvanesh Sundar, P. Aaron Lott, Shon Grabbe, Eleanor G. Rieffel, Matthew J. Reagor, and Davide Venturelli. 2024. Design and execution of quantum circuits using tens of superconducting qubits and thousands of gates for dense Ising optimization problems. *Phys. Rev. Appl.* 22 (Oct 2024), 044074. Issue 4. doi:10.1103/PhysRevApplied.22.044074

[30] P. Magnard, S. Storz, P. Kurpiers, J. Schär, F. Marxer, J. Lütolf, T. Walter, J.-C. Besse, M. Gabureac, K. Reuer, A. Akin, B. Royer, A. Blais, and A. Wallraff. 2020. Microwave quantum link between superconducting circuits housed in spatially separated cryogenic systems. *Physical review letters* 125, 26 (2020), 260502.

[31] Masoud Mohseni, Artur Scherer, K. Grace Johnson, Oded Wertheim, Matthew Otten, Navid Anjum Aadit, Yuri Alexeev, Kirk M. Bresniker, Kerem Y. Camsari, Barbara Chapman, Soumitra Chatterjee, Gebremedhin A. Dagnew, Aniello Esposito, Farah Fahim, Marco Fiorentino, Archit Gajjar, Abdullah Khalid, Xiangzhou Kong, Bohdan Kulchytskyy, Elica Kyoseva, Ruoyu Li, P. Aaron Lott, Igor L. Markov, Robert F. McDermott, Giacomo Pedretti, Pooja Rao, Eleanor Rieffel, Allyson Silva, John Sorebo, Panagiotis Spentzouris, Ziv Steiner, Boyan Torosov, Davide Venturelli, Robert J. Visser, Zak Webb, Xin Zhan, Yonatan Cohen, Pooya Ronagh, Alan Ho, Raymond G. Beausoleil, and John M. Martinis. 2024. How to build a quantum supercomputer: Scaling challenges and opportunities. *arXiv preprint arXiv:2411.10406* (2024).

[32] Thomas Monz, Daniel Nigg, Esteban A Martinez, Matthias F Brandl, Philipp Schindler, Richard Rines, Shannon X Wang, Isaac L Chuang, and Rainer Blatt. 2016. Realization of a scalable Shor algorithm. *Science* 351, 6277 (2016), 1068–1070.

[33] Christopher Portmann and Renato Renner. 2022. Security in quantum cryptography. *Reviews of Modern Physics* 94, 2 (2022), 025008.

[34] Jun Qi, Chao-Han Yang, and Pin-Yu Chen. 2023. QTN-VQC: an end-to-end learning framework for quantum neural networks. *Physica Scripta* 99, 1 (dec 2023), 015111. doi:10.1088/1402-4896/ad14d6

[35] IBM Quantum. 2025. How IBM will build the world's first large-scale, fault-tolerant quantum computer. https://www.ibm.com/quantum/blog/large-scale-ftqc#fn-2 Accessed: 2025-06-16.

[36] Google quantum AI. 2023. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* 614, 7949 (2023), 676–681.

[37] RJ Schoelkopf and SM Girvin. 2008. Wiring up quantum systems. *Nature* 451, 7179 (2008), 664–669.

[38] Pan Shi, Jiahao Yuan, Fei Yan, and Haifeng Yu. 2023. Multiplexed control scheme for scalable quantum information processing with superconducting qubits. *arXiv preprint arXiv:2312.06911* (2023).

[39] Peter A Spring, Shuxiang Cao, Takahiro Tsunoda, Giulio Campanaro, Simone Fasciati, James Wills, Mustafa Bakr, Vivek Chidambaram, Boris Shteynas, Lewis Carpenter, Paul Gow, James Gates, Brian Vlastakis, and Peter J Leek. 2022. High coherence and low cross-talk in a tileable 3D integrated superconducting circuit architecture. *Science Advances* 8, 16 (2022), eabl6698.

[40] J Wenner, M Neeley, Radoslaw C Bialczak, M Lenander, Erik Lucero, A D O'Connell, D Sank, H Wang, M Weides, A N Cleland, and John M Martinis. 2011. Wirebond crosstalk and cavity modes in large chip mounts for superconducting qubits. *Superconductor Science and Technology* 24, 6 (2011), 065001.

[41] Xiao Xue, Bishnu Patra, Jeroen P. G. van Dijk, Nodar Samkharadze, Sushil Subramanian, Andrea Corna, Brian Paquelet Wuetz, Charles Jeon, Farhana Sheikh, Esdras Juarez-Hernandez, Brando Perez Esparza, Huzaifa Rampurawala, Brent Carlton, Surej Ravikumar, Carlos Nieva, Sungwon Kim, Hyung-Jin Lee, Amir Sammak, Giordano Scappucci, Menno Veldhorst, Fabio Sebastiano, Masoud Babaie, Stefano Pellerano, Edoardo Charbon, and Lieven M. K. Vandersypen. 2021. CMOS-based cryogenic control of silicon quantum circuits. *Nature* 593, 7858 (2021), 205–210.

[42] Tian Yang, Chen Liang, Weilong Wang, Bo Zhao, Lixin Wang, Qibing Xiong, Xuefei Feng, and Zheng Shan. 2025. A scalable routing method for superconducting quantum processor. *EPJ Quantum Technology* 12, 1 (2025), 20.

[43] Yunyan Yao, Liang Xiang, Zexian Guo, Zehang Bao, Yong-Feng Yang, Zixuan Song, Haohai Shi, Xuhao Zhu, Feitong Jin, Jiachen Chen, Shibo Xu, Zitian Zhu, Fanhao Shen, Ning Wang, Chuanyu Zhang, Yaozu Wu, Yiren Zou, Pengfei Zhang, Hekang Li, Zhen Wang, Chao Song, Chen Cheng, Rubem Mondaini, H. Wang, J. Q. You, Shi-Yao Zhu, Lei Ying, and Qiujiang Guo. 2023. Observation of many-body Fock space dynamics in two dimensions. *Nature Physics* (2023), 1–7.

[44] Peng Zhao, Kehuan Linghu, Zhiyuan Li, Peng Xu, Ruixia Wang, Guangming Xue, Yirong Jin, and Haifeng Yu. 2022. Quantum crosstalk analysis for simultaneous gate operations on superconducting qubits. *PRX quantum* 3, 2 (2022), 020301.