Evaluating Optical-Loss and Connectivity Effects on Q-Fly Performance in Distributed Fault-Tolerant Quantum Computing

Daisuke Sakuma*¶, and Shota Nagayama¶‡
*Graduate School of Media and Governance, Keio University Shonan Fujisawa Campus, Kanagawa, Japan
¶Keio University Graduate School of Media Design, Hiyoshi Campus, Kanagawa, Japan
sakuma@qitf.org shota@qitf.org

***Abstract*—Distributed fault-tolerant quantum computing (FTQC) is a promising approach to scaling quantum computers beyond a single device. In such systems, quantum processing units (QPUs) are connected by a quantum interconnect, whose topology determines the cost of remote logical operations. However, system performance depends not only on connectivity but also on optical loss along communication paths and within switches, which affects entanglement-generation latency and total program execution time.**

In this work, we evaluate the system-level impact of optical loss in distributed FTQC by incorporating path and switch losses into a remote-operation latency model. Using QFT and the VBE adder as benchmarks, we compare Q-Fly with a 2D grid and analyze the interplay among topology, optical loss, network size, and workload communication demand. Q-Fly reduces hop counts for long-range operations, but its advantage depends on whether this reduction offsets switch insertion loss. Assuming Spanke–Benes group switches composed of elementary 2×2 switches, Q-Fly outperforms the 2D grid for QFT with fewer than 100 nodes when each elementary switch has 97% transmittance. For larger systems, accumulated switch loss increases remote-gate latency and reduces Q-Fly' s advantage, while queueing delay remains negligible. In contrast, for the VBE adder, Q-Fly approaches but does not outperform the 2D grid even with lossless switches because the workload is local and sequential. These results show that interconnect performance strongly depends on workload communication patterns. For QFT, they also identify target loss regimes for optical-switch development and provide initial guidelines for workload-aware quantum interconnects.

***Index Terms*—Distributed Quantum Computing, Fault-Tolerant Quantum Computing, Interconnect, Q-Fly, 2D grid, Latency, Performance Evaluation**

I. Introduction

Practical quantum computing requires many qubits; for example, factoring a 2048-bit RSA integer may require fewer than one million noisy qubits [1]. Distributed quantum computing is a promising approach for scaling beyond a single device by connecting multiple quantum processing units (QPUs) through a quantum interconnect. Candidate topologies include 2D grids, Clos networks, BCube, and Q-Fly [2]–[4].

For optical quantum interconnects, communication performance depends on both logical connectivity and optical loss, since photon loss increases remote Bell-pair generation latency. Q-Fly was proposed as a low-diameter optical interconnect for modular quantum computers [2]. However, despite its low-diameter design, Q-Fly performance still depends on the chosen parameters and optical-switch characteristics.

A Q-Fly topology is characterized by four parameters [2]: p, g, b, and L_{switch}, representing end nodes per group, groups, Bell-state analyzers (BSAs) per group, and group-switch insertion loss, respectively. The parameters p and g determine total end nodes $N = pg$ and group-switch radix, affecting photon loss. BSAs b determine communication resources and can affect resource contention. If a group switch is implemented using elementary 2×2 optical switches, such as a Spanke–Benes switch [5], its total insertion loss depends on elementary-switch loss and switching stages. Thus, end-to-end performance depends on topology, grouping parameters, BSA resources, and optical device characteristics.

In this work, we introduce an optical-loss-aware execution-time model for quantum interconnects and compare Q-Fly with a 2D-grid baseline to identify the parameter regime in which Q-Fly provides a performance advantage.

II. Remote Gate Execution Model

We assume that each end node employs the surface code for quantum error correction and lattice surgery for logical two-qubit operations [6]. A remote logical CX gate consists of local physical gate operations and remote Bell-pair generation. Its execution time is modeled as

$$\tau_{\mathrm{remote-CX}} = \tau_{\mathrm{local}} + \tau_{\mathrm{Bell}}, \tag{1}$$

where τ_{local} is the local gate execution time and τ_{Bell} is the latency required to generate the logical Bell pairs.

Following remote lattice surgery [7], we model $\tau_{\mathrm{Bell}} = \tau_{\mathrm{gen}} N_{\mathrm{req}} d(2d-1)$, where τ_{gen} is the latency for generating one physical Bell pair, N_{req} is the physical Bell pairs required to obtain one logical Bell pair, and d is the code distance. Bell-pair generation latency depends on the trial rate r_{trial}, Bell-state measurement success probability P_{Bell}, and the end-to-end optical transmittance [2], [4]. We model

$$\tau_{\mathrm{gen}} \propto \left(P_{\mathrm{Bell}} \eta_{\mathrm{prop}} \eta_{\mathrm{switch}}^{N_{\mathrm{hop}}} \right)^{-1}, \tag{2}$$

where η_{prop} is the propagation transmittance, η_{switch} is the transmittance of a single group switch, and N_{hop} is the number

This work was supported by JST [Moonshot R&D] [JPMJMS226C] and [JPMJMS256K].

of traversed group switches. Since the group-switch radix k determines the number of switching stages, different choices of p and g produce different optical losses even for the same network size N. Throughout this paper, we adopt $p = \lfloor g/2 \rfloor$, which minimizes the required radix for SPHD. The same rule is applied to DPHD and DPFD for consistency.

Equation (1) models only remote logical CX execution time. Additional time waiting arises when lattice-surgery paths or communication resources are unavailable. This waiting depends on logical-qubit placement, available communication resources, and network connectivity. Here, we evaluate the combined impact of optical-switch transmittance and topology-dependent resource contention on total execution time.

III. Evaluation

We compare three Q-Fly variants, single-path half-duplex (SPHD), dual-path half-duplex (DPHD), and dual-path full-duplex (DPFD), with a 2D-grid topology. For each network size, we set $p = \lfloor g/2 \rfloor$ and $N = pg$, following the parameter scaling rule described in Section II. Each QPU contains one logical data qubit and three ancillae for lattice-surgery paths.

The evaluation framework schedules each gate of an input quantum circuit in program order while allocating computation and communication resources. Allocated resources remain occupied until completion; for inter-node gates, execution time follows Eq. (1). The final-gate completion time is reported as the total program execution time. Since τ_{Bell} in Eq. (2) depends on the network architecture, total execution time also depends on topology and parameter configuration.

We use QFT and the VBE adder [8] as benchmark because they represent contrasting communication patterns. QFT requires frequent long-range interactions, whereas the VBE adder mainly uses short-range sequential interactions. Q-Fly reduces hop count between arbitrary node pairs: two to three hops for SPHD and DPHD, and one to two hops for DPFD. Therefore, QFT is well suited for evaluating the benefit of low-diameter connectivity. In contrast, the VBE adder has limited opportunity to exploit this benefit, because its dependency structure is largely sequential and its communication is mostly local. Thus, even with low switch loss, Q-Fly is not expected to substantially outperform the 2D grid for the VBE adder.

Figure 1(a) compares the topologies with elementary-switch transmittance fixed at 97%, showing that DPFD performs best among the Q-Fly variants and outperforms the 2D grid for fewer than 100 end nodes, while Fig. 1(b) shows that higher transmittance extends this operating region: DPFD outperforms the 2D grid up to 595 end nodes at 99% transmittance and at least 903 end nodes with lossless switches. This degradation at larger scales is dominated by the increase in τ_{Bell} from accumulated optical loss, while queueing delay from resource contention remains negligible, indicating that DPFD scalability is limited mainly by switch loss.

In contrast, Fig. 1(c) and (d) show that Q-Fly does not outperform the 2D grid for the VBE adder, whose short-range sequential interactions benefit little from reduced network diameter. Finite switch loss therefore increases remote-gate latency; nevertheless, with lossless switches, Q-Fly reaches nearly the same performance as the 2D grid. These results show that Q-Fly's advantage is workload dependent.

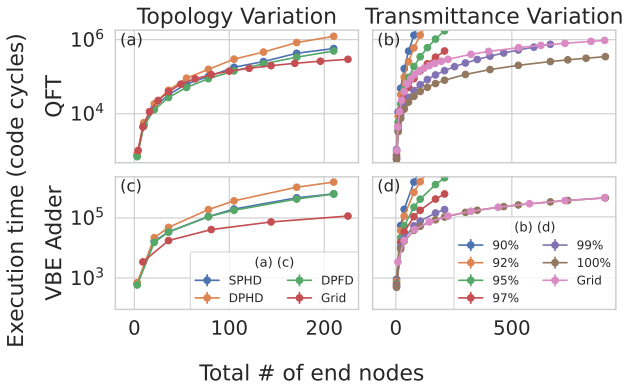

Fig. 1. Execution time scalability. (a)(b) are for QFT, and (c)(d) are for VBE adder. (a)(c) Comparison among three Q-Fly variants and the 2D grid when each elementary 2×2 switch in the Q-Fly group switches has 97% transmittance. (b)(d) Transmittance dependence of DPFD Q-Fly compared with the 2D grid. Percentage values in legend label indicate transmittance of the elementary 2×2 switch.

IV. Conclusion and Future Work

In this work, we evaluated how optical loss and workload communication structure affect Q-Fly execution time against a 2D-grid baseline. For QFT, Q-Fly outperforms the 2D grid when each elementary 2×2 switch exceeds 97% transmittance, giving a concrete target for low-loss switch development. For the mostly local and sequential VBE adder, Q-Fly does not outperform the grid, but approaches grid-level performance when switch loss is sufficiently small. These results motivate workload-aware interconnect design and algorithm-specific compilation, mapping, routing, and scheduling. Future work will evaluate broader workloads and QPU-size effects.

References

[1] C. Gidney, "How to factor 2048 bit RSA integers with less than a million noisy qubits," 2025, arXiv preprint arXiv:2505.15917.

[2] D. Sakuma *et al.*, "Q-Fly: An Optical Interconnect for Modular Quantum Computers," 2025, arXiv preprint arXiv:2412.09299.

[3] H. Shapourian, E. Kaur, T. Sewell, J. Zhao, M. Kilzer, R. Kompella, and R. Nejabati, "Quantum Data Center Infrastructures: A Scalable Architectural Design Perspective," 2025, arXiv preprint arXiv:2501.05598.

[4] S. Pouryousef, E. Kaur, H. Shapourian, D. Towsley, R. Kompella, and R. Nejabati, "Benchmarking Quantum Data Center Architectures: A Performance and Scalability Perspective," 2026, arXiv preprint arXiv:2601.01353.

[5] R. A. Spanke and V. E. Benes, "N-stage planar optical permutation network," *Appl. Opt.*, vol. 26, no. 7, pp. 1226–1229, Apr 1987, doi:10.1364/AO.26.001226.

[6] D. Litinski, "A Game of Surface Codes: Large-Scale Quantum Computing with Lattice Surgery," *Quantum*, vol. 3, p. 128, Mar. 2019, doi:10.22331/q-2019-03-05-128.

[7] A. G. Fowler, D. S. Wang, C. D. Hill, T. D. Ladd, R. Van Meter, and L. C. L. Hollenberg, "Surface code quantum communication," *Phys. Rev. Lett.*, vol. 104, p. 180503, May 2010, doi:10.1103/PhysRevLett.104.180503.

[8] V. Vedral, A. Barenco, and A. Ekert, "Quantum networks for elementary arithmetic operations," *Phys. Rev. A*, vol. 54, pp. 147–153, Jul 1996, doi:10.1103/PhysRevA.54.147.