cisco
Quantum
Labs

Network-aware Distributed Quantum Computing Compiler

Cisco's Software Technology for Quantum Computing Scale out

Contents

Quantum data center 3

Quantum compiler 4

Performance and design analysis tools and use cases 6

©2025 Outshift by Cisco and/or its affiliates. All rights reserved.

Summary

Cisco has developed the world's first network-aware Quantum Compiler prototype, purpose-built for quantum data centers. This compiler delivers several industry-first capabilities, including network-aware circuit partitioning and support for distributed quantum error correction. It operates seamlessly across simulated, emulated, and real hardware infrastructures. For algorithm developers, it enables circuit optimization for distributed execution by incorporating qubit allocation, network latency, and error rates. For data center designers, it provides powerful tools to model and refine architectures based on target quantum processing unit (QPU) technologies, network topologies, and application demands. By bridging software, networking, and hardware, the Quantum Compiler establishes a true co-design platform that accelerates the realization of scalable, fault-tolerant, and efficient quantum data centers.

Quantum data center

A practical quantum computer capable of solving real-world problems must operate with full error correction. This, in turn, requires multiple millions and potentially tens of millions of physical qubits. Today's quantum processing units (QPU) are limited to only tens or hundreds of qubits, and even the most optimistic vendor roadmaps predict only a few thousand-qubit processors in the next three to five years.

The most practical way to build a useful quantum computer with all perspectives in mind—engineering, scientific, capital expenditures (CAPEX) and operational expenses (OPEX)—is to scale through distributed quantum computing in a quantum data center (QDC).

In this architecture, multiple quantum processors are networked together, enabling a distributed system that can grow to meet the demands of practical quantum computing applications.

QDCs not only provide the scalability required for large-scale quantum computation but also offer economic and operational benefits by centralizing quantum resources in a controlled environment.

Cisco is actively developing the architectures and technologies necessary for quantum data centers. Our architectural approach and analysis have been extensively described in Quantum data center Infrastructures: A Scalable Architectural Design Perspective. From a technology standpoint, we recently released Cisco's Quantum Network Entanglement Chip operating at telecom wavelengths with an ultra-high pair generation rate (>200M entangled pairs/second) and exceptional fidelity (~99%). We have also invested in a startup building a dual-frequency entanglement source capable of generating entangled pairs at both telecom and near-infrared (NIR) wavelengths, paving the way for linking heterogeneous quantum computers into a unified Quantum Network. In parallel, we are aggressively developing a novel quantum switch that will significantly accelerate Cisco's capability to realize its proposed QDC architecture.

©2025 Outshift by Cisco and/or its affiliates. All rights reserved.

As hardware advances rapidly, we recognize that building a quantum data center also requires hardware-software co-design. A key component of this effort is developing a distributed quantum computing compiler that explicitly accounts for Quantum Network connectivity and inter-processor communication within the data center. These capabilitiesmake Cisco's Quantum Compiler the world's first network-aware distributed quantum computing compiler that is optimized for data center operation.

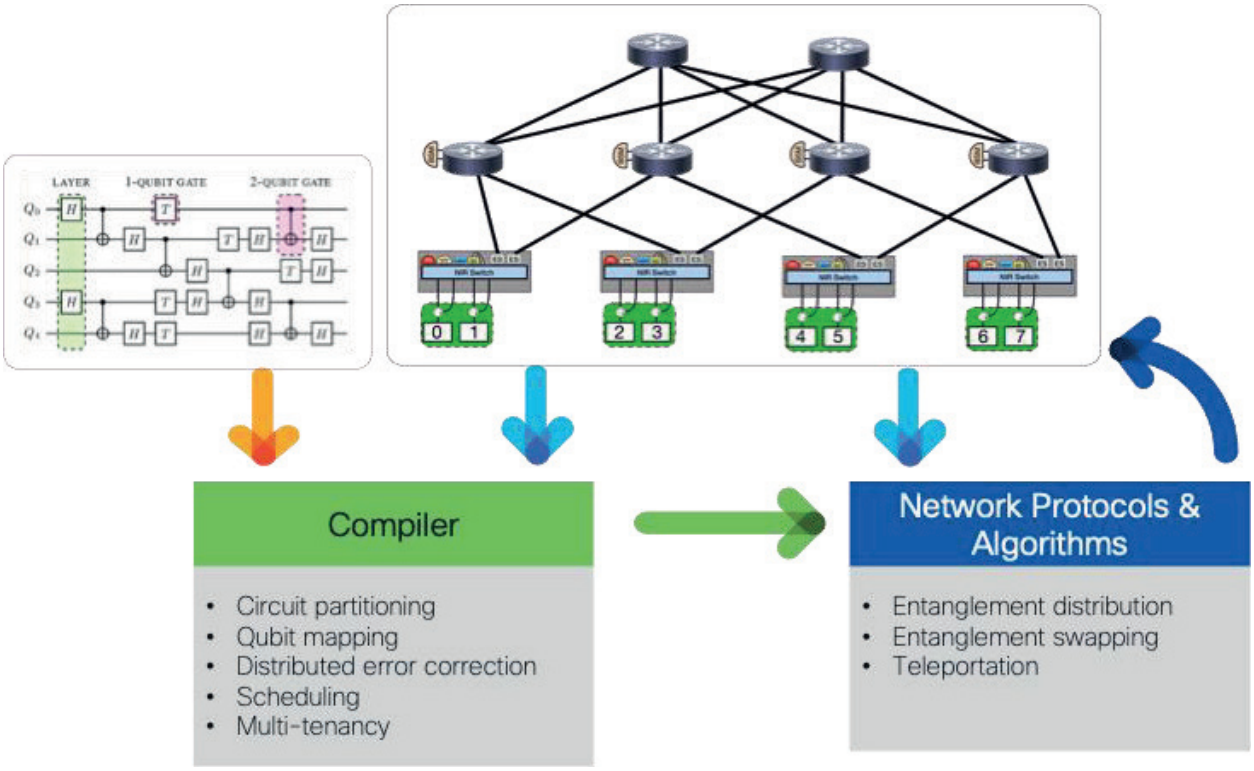

Fig 1. The industry's first Quantum Compiler prototype from Cisco.

Quantum Compiler

Cisco has developed the world's first network-aware Quantum Compiler prototype purpose-built for quantum data centers. The Quantum Compiler introduces several unique, industry-first capabilities including being the first compiler platform that supports distributed quantum error correction, allowing you to split quantum algorithms across multiple processors to achieve scalable computation.

The compiler's main functionalities also include:

> **Network-aware circuit partitioning:** The compiler intelligently partitions quantum circuits across multiple quantum processors, taking into account arbitrary interconnect topologies within a distributed quantum computing environment. Our circuit partitioning algorithm addresses the core challenge of distributing quantum workloads across networked QPUs, minimizing nonlocal operations. The method frames the circuit as an interaction graph and performs windowed

©2025 Outshift by Cisco and/or its affiliates. All rights reserved.

partitioning, optimizing qubit placement per window using a modified Kernighan-Lin (KL) or multilevel partitioner. Crucially, it integrates gate packing and selective state teleportation within a window, applying inter-window qubit remapping only when the expected entanglement generation and distribution cost improvement justifies the additional movement overhead.

In practice, this approach reduces entanglement consumption by up to an order of magnitude compared to static partitioning, while remaining computationally tractable for circuits with hundreds of qubits. (Optimized Quantum Circuit Partitioning Across Multiple Quantum Processors)

Qubit mapping: The qubit mapping process converts the circuit-level qubit assignments from the partitioner into a concrete placement on physical qubits across multiple QPUs. Its objective is to minimize reliance on high-noise connections by mapping qubits onto hardware locations that preferentially use the most reliable interconnects. The mapping operates on interaction graphs annotated with gate frequencies and network characteristics—including link fidelity, entanglement generation rates, and bandwidth—and produces a placement that reduces the cost of remote gates. Based on the Integer Linear Programming (ILP) formulation, the mapping process delivers a globally optimized, network-aware assignment that helps preserve circuit fidelity. (Ref 1: Optimized Quantum Circuit Partitioning Across Multiple Quantum Processors, Ref 2: Quantum data center Infrastructures: A Scalable Architectural Design Perspective)

Distributed quantum error correction: Error correction is essential for scaling quantum computation, as it protects circuits from noise and enables fault-tolerant execution. Once a circuit is defined, we encode it using an error-correction code—such as bivariate bicycle or surface code—and run the resulting error corrected circuit through our compilation and scheduling pipeline. This end-to-end flow allows us to execute the error correction in distributed environment and quantify the impact of error correction on network resource utilization, including entanglement demand, communication cost, and execution latency. We are extending this approach to incorporate code-specific scheduling strategies and decoder performance models, enabling a more detailed and realistic assessment of distributed quantum workloads.

Scheduling: In distributed quantum computing, scheduling is the orchestration layer that determines when to generate entanglement and execute nonlocal gates so that a circuit completes efficiently under hardware constraints. The scheduler accounts for network topology, available communication qubits, probabilistic entanglement generation, and Bell-state measurement (BSM) availability. The scheduling mechanism supports static scheduling, which processes circuits step by step, and dynamic scheduling, which triggers entanglement generation as soon as dependencies are met. Static scheduling offers easier performance analysis, while dynamic scheduling improves resource utilization and responsiveness under variable network conditions. (Ref 1: Network-Aware Scheduling for Remote Gate Execution in quantum data centers, Ref. 2: Quantum data center Infrastructures: A Scalable Architectural Design Perspective)

©2025 Outshift by Cisco and/or its affiliates. All rights reserved.

Multi-tenancy: Quantum data centers offer a practical path to scaling quantum computation, but their performance depends on both hardware limits and orchestration efficiency. When multiple jobs compete for limited QPUs and network bandwidth, scheduling decisions directly impact throughput and fidelity. We address the challenge of scheduling quantum workloads in a hierarchical, resource-constrained architecture by modeling job costs in terms of local and non-local gate execution, with the latter incurring higher latency and entanglement overhead.

Our approach introduces a constraint-aware resource allocation algorithm and evaluates job reordering heuristics based on qubit count, circuit depth, and hybrid strategies. Simulation results demonstrate that strategic job ordering and QPU placement can significantly reduce total execution cost and improve system utilization. (Multi-Tenant Job Scheduling in quantum data centers. IEEE International Conference on quantum computing and Engineering (QCE25)).

As discussed above, the Quantum Compiler is inherently network-aware and, therefore, relies on a unified software stack consisting of common network protocols and intelligence that are also leveraged across other applications. These include entanglement distribution, entanglement swapping, teleportation, and quantum measurement.

Performance and design analysis tools and use cases

The Quantum Compiler incorporates a comprehensive suite of analysis tools specifically designed to serve both quantum algorithm developers and quantum data center architects. These tools provide a unified framework in which the Quantum Compiler and the Quantum Network Controller can operate seamlessly across simulated, emulated, or real hardware infrastructures or any hybrid combination thereof.

For quantum algorithm developers, the toolkit enables detailed performance modeling and optimization of circuits in a distributed quantum computing environment. Developers can assess trade-offs between qubit allocation, network latency, error rates, and resource scheduling. This allows them to adapt their algorithms for execution across heterogeneous and interconnected quantum processors.

For quantum data center designers, the tools provide powerful capabilities for designing and validating data center architectures. They can model and evaluate infrastructure based on their target QPU technologies, network topologies, and intended applications or user workloads. This allows hyperscale operators to explore different design choices—such as interconnect strategies, error correction schemes, and resource allocation policies—and optimize their architectures to achieve maximum scalability, reliability, and efficiency.

Together, these capabilities make the Quantum Compiler not only a programming and optimization environment but also a co-design platform that tightly couples software, networking, and hardware layers for the development of next-generation distributed quantum data centers.

©2025 Outshift by Cisco and/or its affiliates. All rights reserved.

outshift.com

@ outshift.com