NVIDIA DGX Quantum

The first unified system for quantum-classical computing

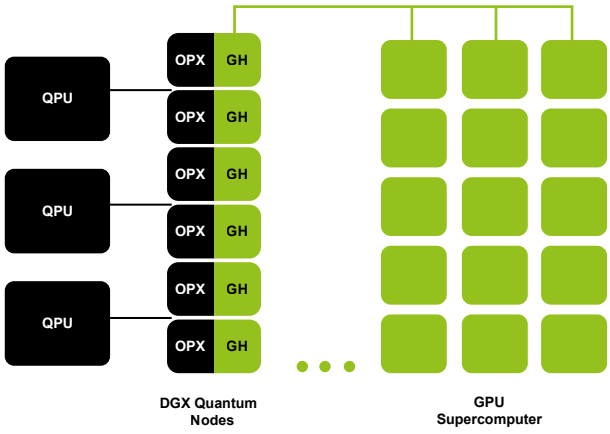
Quantum computing is the newest frontier of scientific computing, with researchers and enterprises around the globe exploring how to tap into its capabilities to build the algorithms and applications that can supercharge science. From drug development to financial modeling, the promise of quantum computing is vast.

It has long been understood that all useful quantum computing will require tightly integrated high-performance classical computing, from hybrid algorithms and applications to quantum error correction, calibration, and control. As quantum processors scale and improve, these workloads become increasingly computationally intensive, and begin to require low-latency coupling between quantum and classical. In order to run quantum error correction and unlock fault-tolerant quantum computing, the classical decoding must be done much faster than the coherence time of the qubit. Superficial integration will no longer be sufficient - integrated hardware is called for.

To meet this need and unlock the promise of quantum computing, NVIDIA and Quantum Machines are proud to introduce DGX Quantum, the first unified system for integrated quantum-classical computing.

DGX Quantum

DGX Quantum integrates Grace Hopper with OPX through low-latency PCIe, enabling data movement from the quantum control system to the GPU in under 500ns, a 10-100x improvement in latency over previous state-of-the-art integrations. A single DGX Quantum fits in a 4U 19" rack form factor, and is scalable in two directions: more OPX+ can be added for a larger QPU, and more Grace Hopper can be added for larger classical compute requirements. DGX Quantum is the system to scale from small experiments to a full quantum-accelerated supercomputer.



Hardware

Grace Hopper

The classical chip driving DGX Quantum is the NVIDIA Grace Hopper superchip. Grace Hopper — which integrates the high-performance NVIDIA Hopper architecture GPU with the company's new Grace CPU — is supercharged for giant-scale AI and HPC applications. It delivers up to 10x higher performance for applications running terabytes of data, providing unprecedented acceleration for error correction decoding, calibration, control, and other quantum-classical workloads.

OPX

OPX is an all-in-one device for ultra low latency quantum control and readout. OPX hosts a unique Pulse Processing Unit (PPU) that optimizes the integration of quantum operations (pulses and measurements) with ultra-fast classical processing and control flow. It combines parametric pulse generation, real-time, Turing-complete math calculations, on-the-fly pulse manipulations, and control flow. OPX is programmed using QM's intuitive quantum-classical pulse-level language - QUA, which opens the way for running complex quantum algorithms right out of the box, including quantum error correction, error-mitigation, ultra-fast multi-qubit calibration, and more. The OPX platform includes integration with gate level via industry standards such as OpenQASM3 and now Cuda Quantum.

Get Started with DGX Quantum

For more information about the DGX Quantum partner program, contact:

info@quantum-machines.co

Software

Optimized quantum-classical hardware requires optimized quantum-classical software. DGX Quantum comes with NVIDIA's CUDA Quantum platform, an open-source programming model and compiler toolchain for unified application-level quantum-classical programming, as well as Quantum Machines' QUA framework, a universal pulse-level programming language enabling any quantum protocol.

CUDA Quantum

With a unified and open programming model, NVIDIA CUDA Quantum is a first-of-its-kind platform for integrating and programming quantum processing units (QPUs), GPUs, and CPUs in a single system. CUDA Quantum features a flexible model that is familiar to domain scientists and interoperates with the best tools for both classical and quantum computing, as well as a high-performance compiler built for hybrid workloads.

QUA

QUA is a first-of-its-kind pulse-level programming language that integrates classical processing into the lowest levels of quantum programming in an unprecedented way. QUA unifies universal quantum operations in their 'raw' format – all the pulse-level stuff used to control and measure qubits – with universal classical operations used in classical processing and all the benefits of Python, Matlab, Java, etc. Highly scalable, QUA enables coding 1,000 qubits as easily as coding one, removing any limitations in implementing user protocols, from the simplest to the most complex. With DGX Quantum, QUA will allow calling classical acceleration routines right from the pulse-level quantum program and have them executed on NVIDIA's GPU with controller-GPU latencies as low as 100s of nanoseconds. QUA will be integrated with CUDA Quantum to allow hybrid algorithms and workflows to be productively programmed and run with industry leading performance.