# Co-Packaged Silicon Photonics Switches for Gigawatt AI Factories

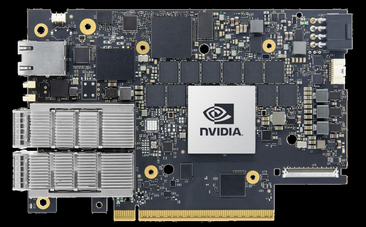
Gilad Shainer, Senior Vice President of Networking | Hot Chips 2025

# The Data Center Is the Computer, the Network Defines the Data Center

Compute Scale-Up

NVLink Switch

Secured User Access
Storage/Management




BlueField-3 DPU

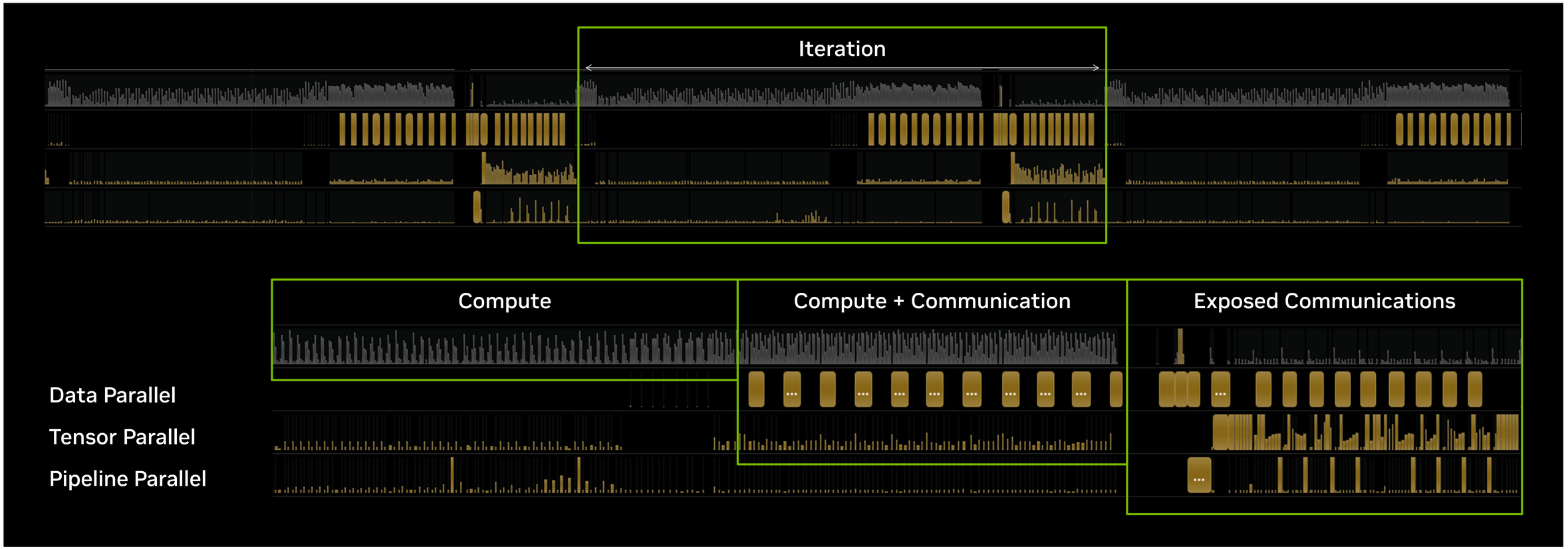
Compute Scale-Out

**Quantum-X**
InfiniBand Switch

**SuperNIC**
InfiniBand/Ethernet

**Spectrum-X**
Ethernet AI Switch

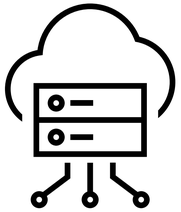
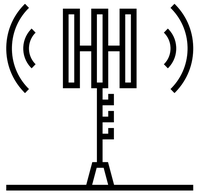
# AI Requires Zero Jitter Communications




Representative profile from a large-scale training run

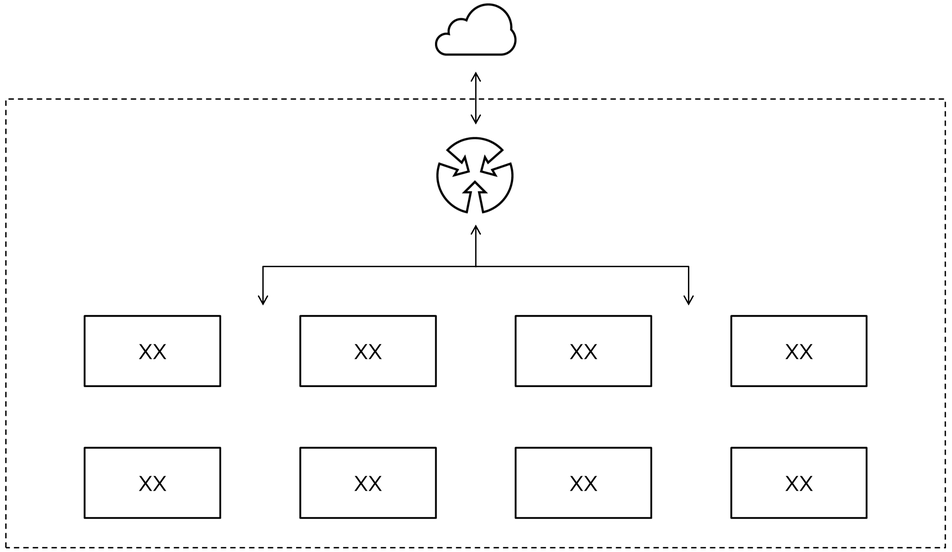
Communications are bursty in nature; an average bandwidth utilization is not a good network criterion

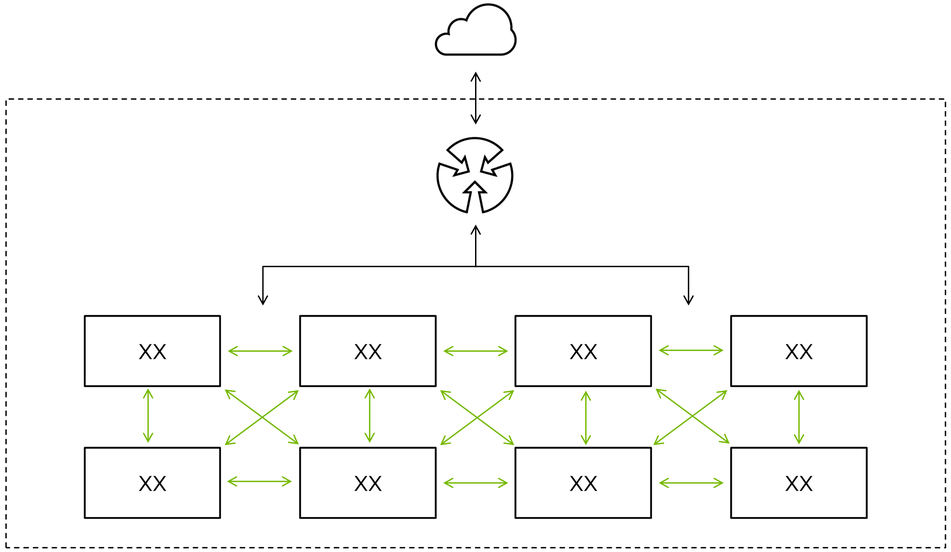
# The Different Ethernet Architectures

| Enterprise | Hyperscale Spine | AI Factories | Service Provider |
| --- | --- | --- | --- |
|  |  |  |  |
| Enterprise, feature-rich DC | Hyperscale DC, cloud spine | High-performance, distributed computing | Service provider core, carrier, DCI |

# Spectrum-X Ethernet Brings High Performance to Ethernet

The network defines the data center

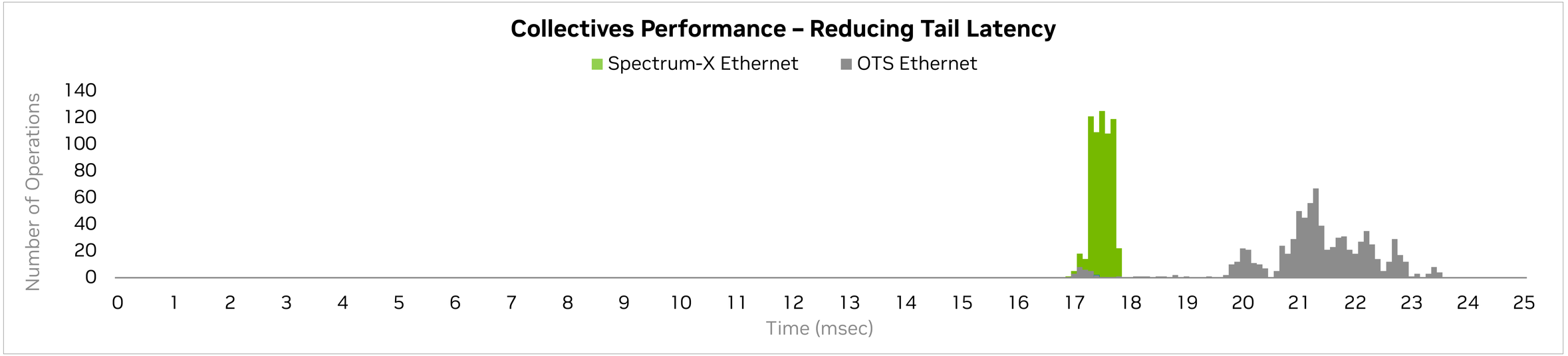



| OTS Ethernet - Hyperscale Clouds | | Spectrum-X Ethernet – AI Factories |
|---|---|---|
| Loosely Coupled Applications | | Distributed Tightly-Coupled Processing |
| TCP (Low Bandwidth Flows and Utilization) | | RoCE (High Bandwidth Flows and Utilization) |
| High Jitter Tolerance | | Low Jitter Tolerance (Long Tail Kills Performance) |
| Heterogeneous Traffic Average Multi-Pathing | | Bursty Network Capacity Predictable Performance |

# SpectrumX Ethernet Low Jitter Communications for AI

## Switch-to-SuperNIC, end-to-end network processing, bringing high performance to Ethernet




**Spectrum-X Ethernet SuperNIC**

Schedule Data Transmission
to Avoid Congestion

**Spectrum-X Ethernet Switch**

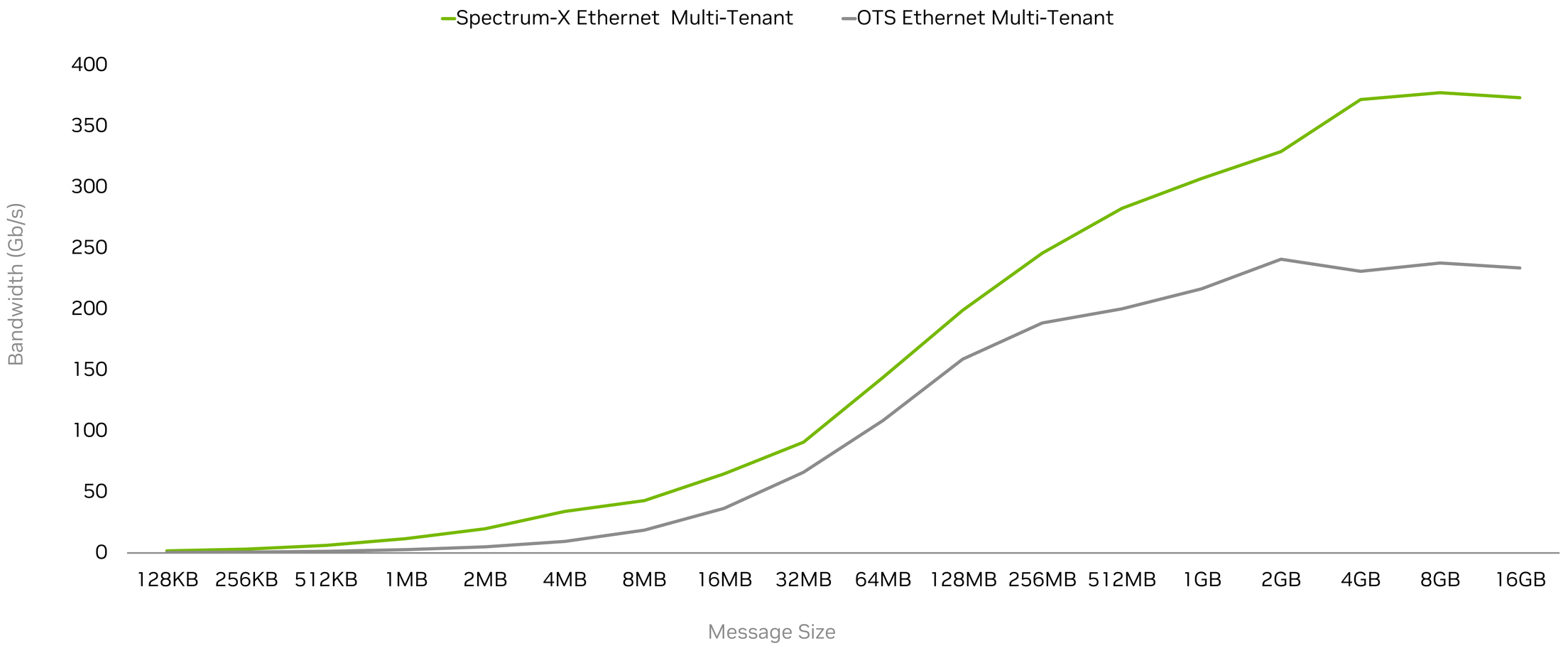
Ultra-High-Speed Traffic Monitoring Distribute Data
Across All Switch Ports Ignoring Data Ordering

**Spectrum-X Ethernet SuperNIC**

Reordering - Receive Data
and Place it Back in Order

# 35% Higher NCCL Multi-Tenant Performance

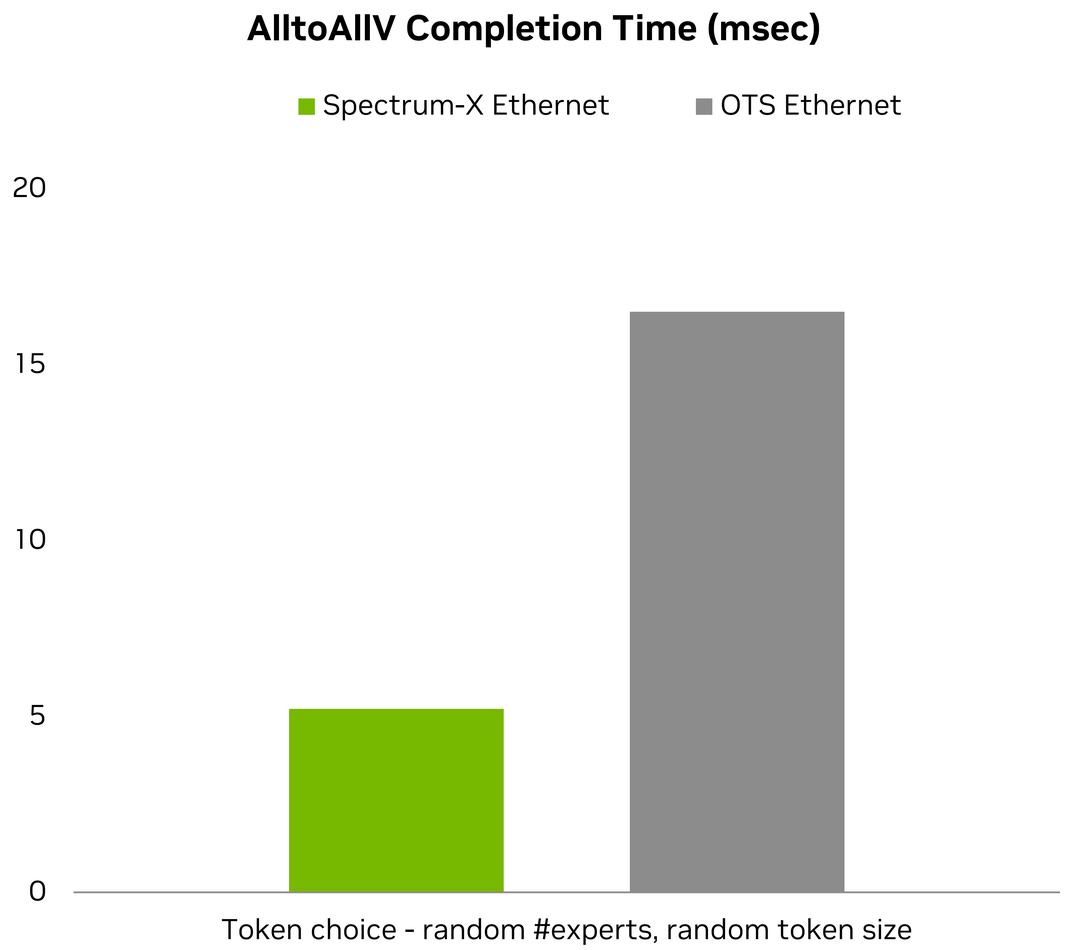
Spectrum-X Ethernet isolates multi-tenant communication noise, 35% higher performance vs. OTS Ethernet



# 3X Higher Expert Dispatch Performance

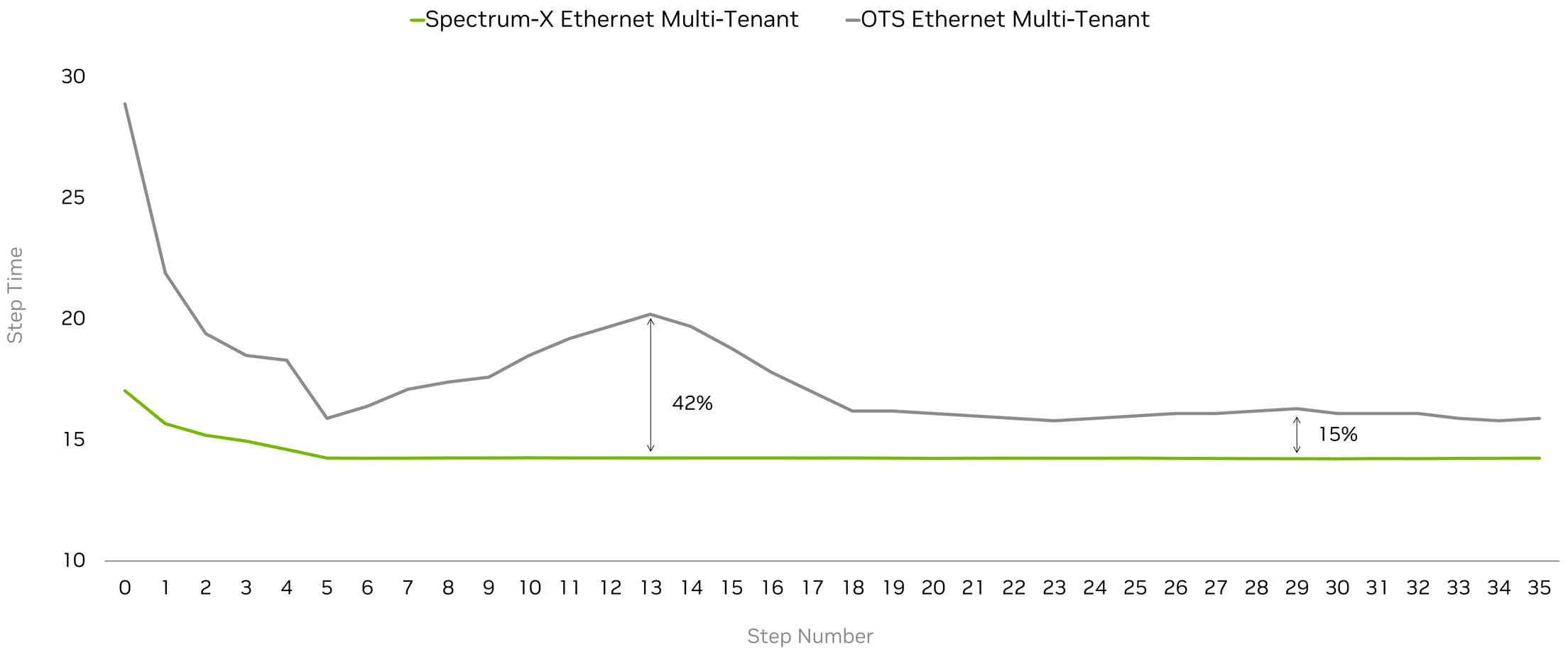
Spectrum-X Ethernet load balancing and congestion control show 3X higher performance

- Modern MoE applications (training and inference) use asymmetric and stochastic Expert token dispatch
- This exposed communication stage can be modeled with variable size all to all V collective (A2AV)
- This pattern is largely exposed and can create extreme congestion that can cause packet drops and victim flows



# 42% Higher LLAMA3 70B Training in Multi-Tenant Data Center

Spectrum-X Ethernet isolates multi-tenant application noise, higher performance with deterministic results

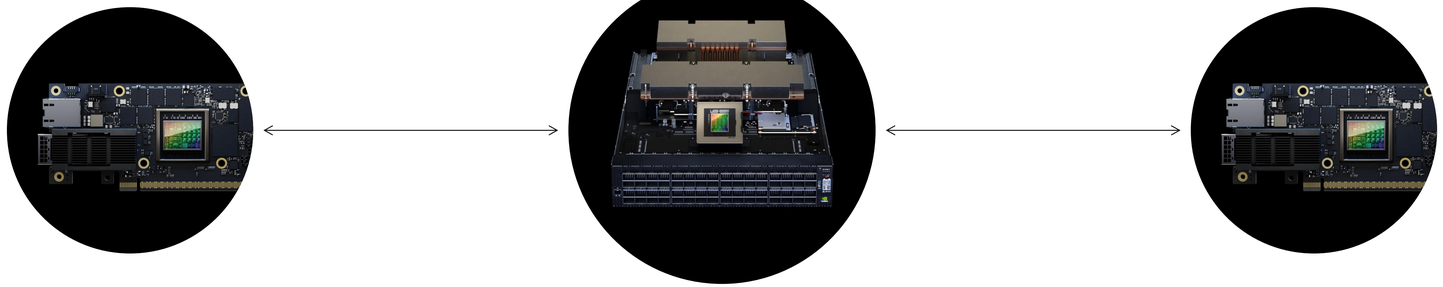


# What Makes Spectrum-X Ethernet Special

Switch-to-SuperNIC, end-to-end network processing, bringing high performance to Ethernet

| Load Balancing | Tail Latency | Noise Isolation | Resilient Performance | High-Frequency Telemetry | Datapath Programmability |
|---|---|---|---|---|---|
| **1.6X** Higher effective bandwidth | **1.3X** Higher collectives bandwidth | **2.2X** Higher all-reduce bandwidth | **1.3X** Higher all-to-all bandwidth | **1,000X** Faster telemetry collection | **16 Cores** **256 Threads** In-network configurable computing |
| Optimal usage of all network connections | Reducing collectives tail latency | Eliminating network hot spots | Rebalances data flows avoiding failed links | Real-time adaption to network conditions | Support for customized end-user AI frameworks |

SONiC | Cumulus | NetQ | NCCL | NIXL | Air | SAI/SPSDK | DOCA | Magnum IO



**Spectrum-X Ethernet SuperNIC** **Spectrum-X Ethernet Switch** **Spectrum-X Ethernet SuperNIC**

# Scale-Out and AI Density Depend on Optical Connectivity

The optical network power consumption represents 10% of compute resources

Traditional Cloud Data Center

**100K**
Servers

**2.3 MW**
Transceiver Power

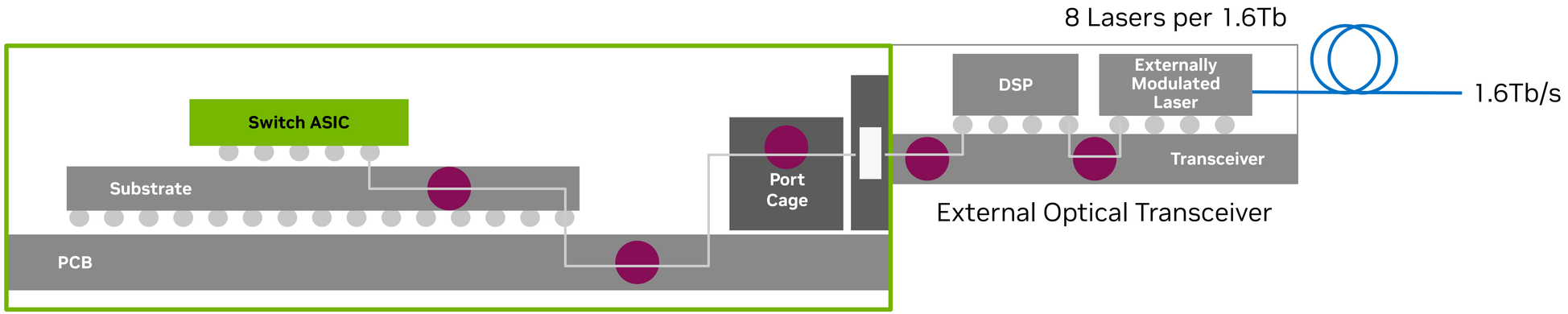
AI Factory

**100K**
Servers

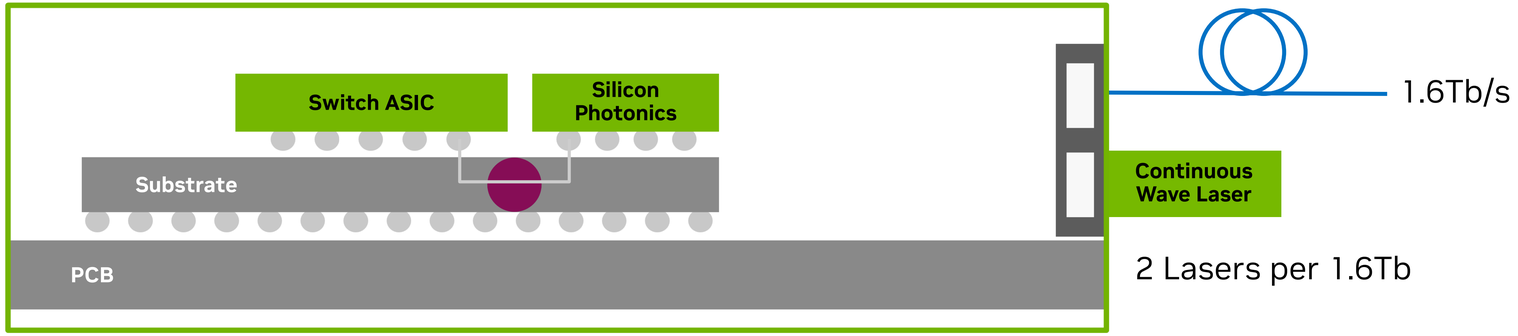
**40 MW**
Transceiver Power

# Introducing Spectrum-X Ethernet Photonics

World's first 200G/SerDes co-packaged optics




Traditional Pluggable Switch

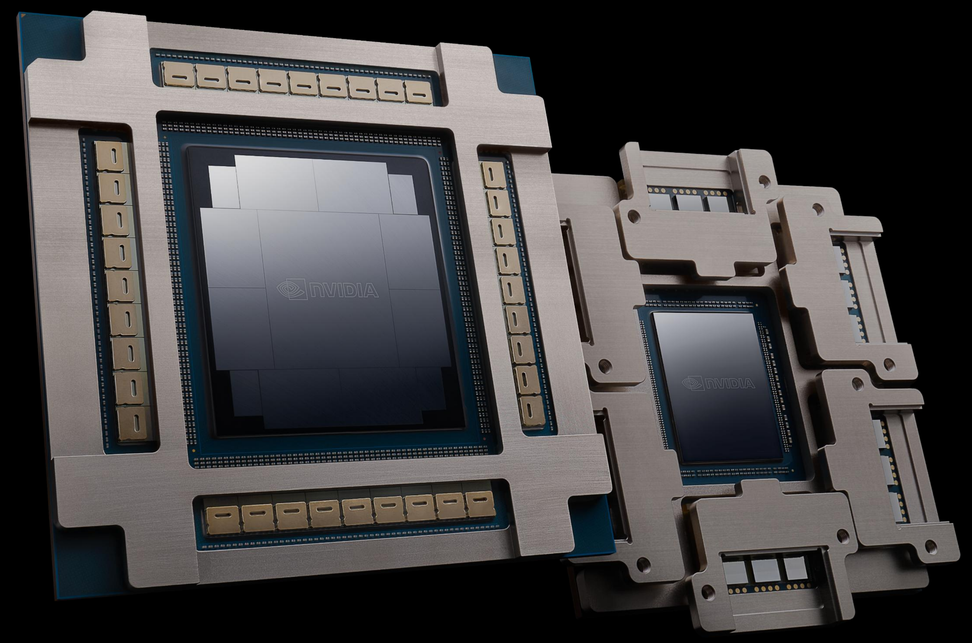



Spectrum-X Ethernet Photonics: Co-packaged Silicon Photonics

# NVIDIA Photonics

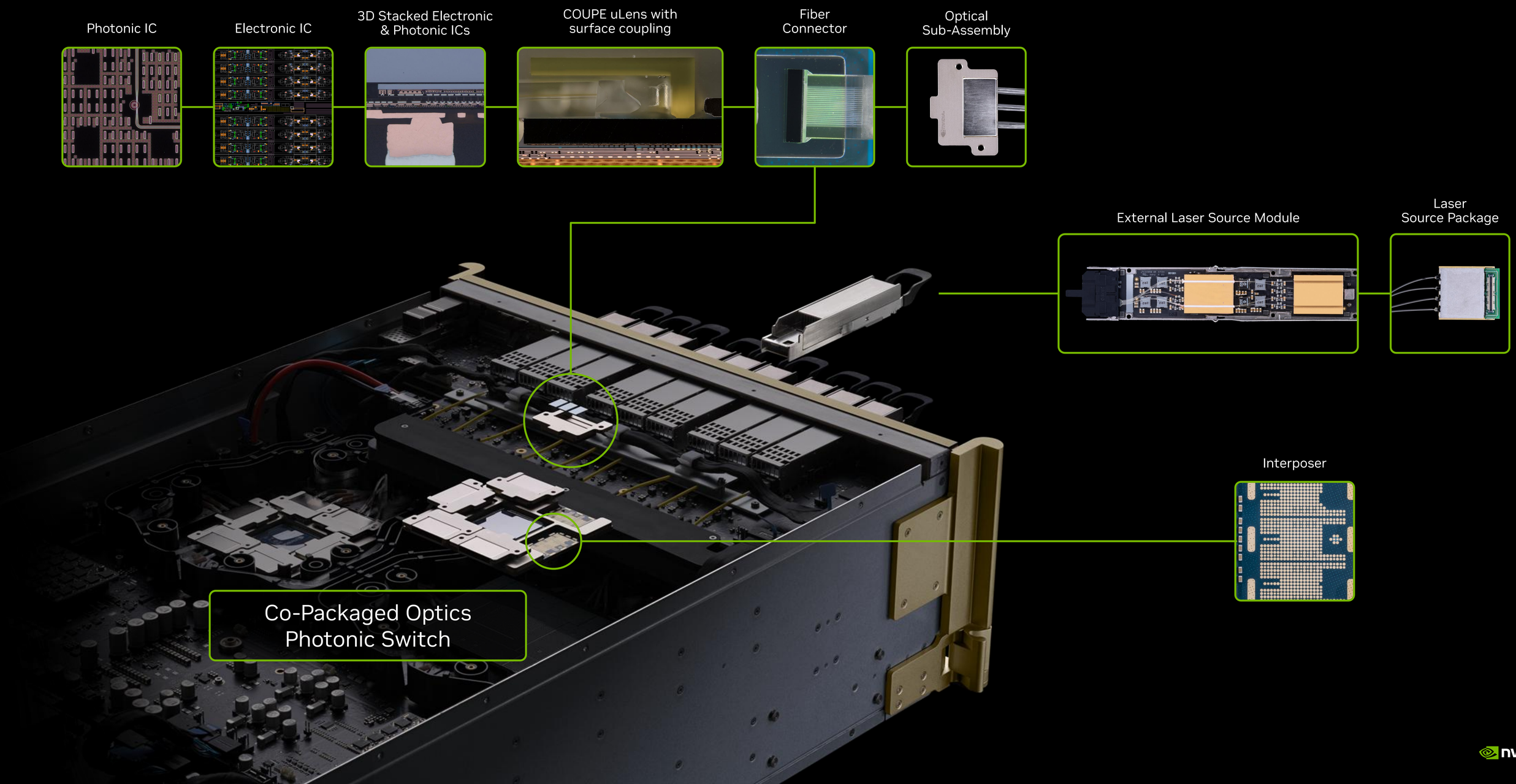
## CPO co-invention with ecosystem partners

- 1st 1.6T Silicon Photonics CPO Chip - new Micro Ring Modulators (MRM)
- 1st 3D-Stacked Silicon Photonics Engine with TSMC process
- High-power, high-efficiency lasers
- Detachable fiber connectors
- 100’s of patents, licensed to partners

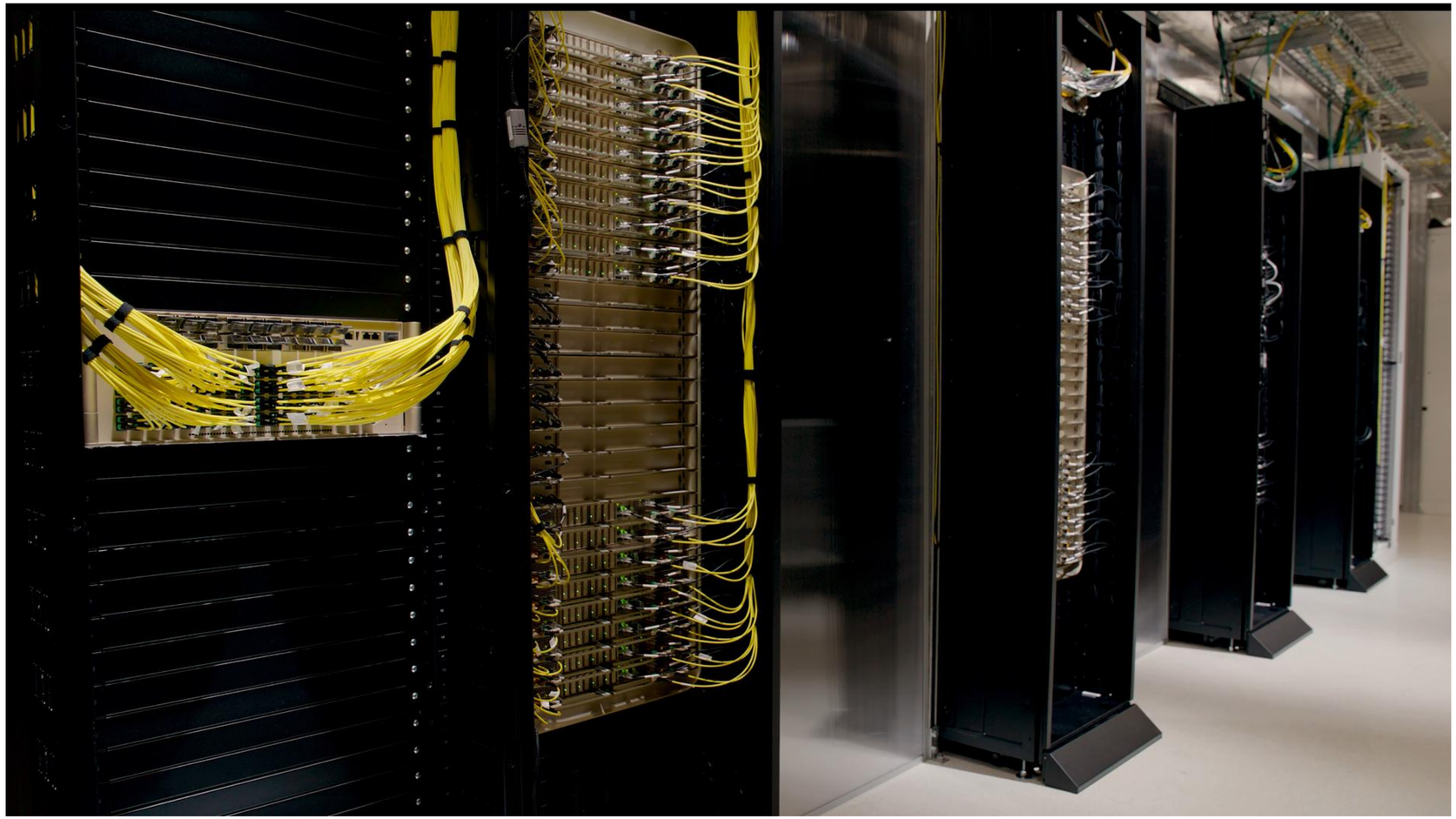


**Spectrum-X Ethernet Integrated Silicon Photonics**

**Quantum-X InfiniBand Integrated Silicon Photonics**


Photonic IC
Electronic IC
3D Stacked Electronic & Photonic ICs
COUPE uLens with surface coupling
Fiber Connector
Optical Sub-Assembly
External Laser Source Module
Laser Source Package
Interposer
Co-Packaged Optics Photonic Switch

# NVIDIA Photonics at NVIDIA Data Center



**3.5X**
Power efficiency

**10X**
Higher resiliency

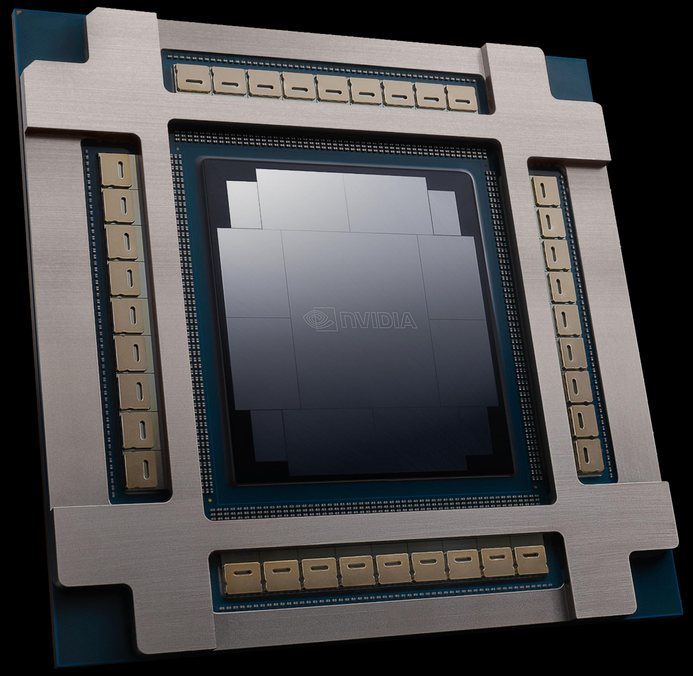
**1.3X**
Faster time to operation

# NVIDIA Spectrum-X Ethernet Photonics

World’s most advanced 200G co-package optics

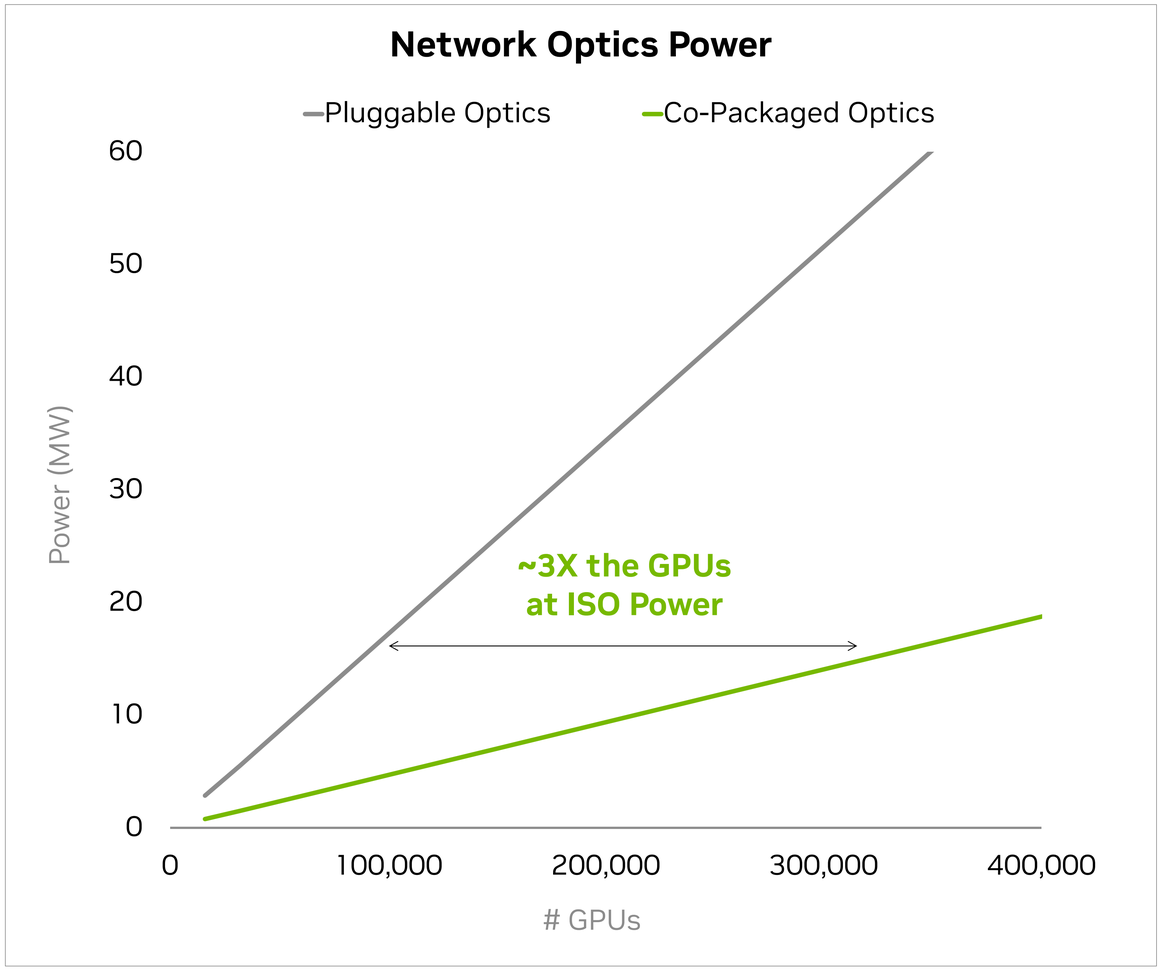
## Spectrum-6 102T Integrated Silicon Photonics

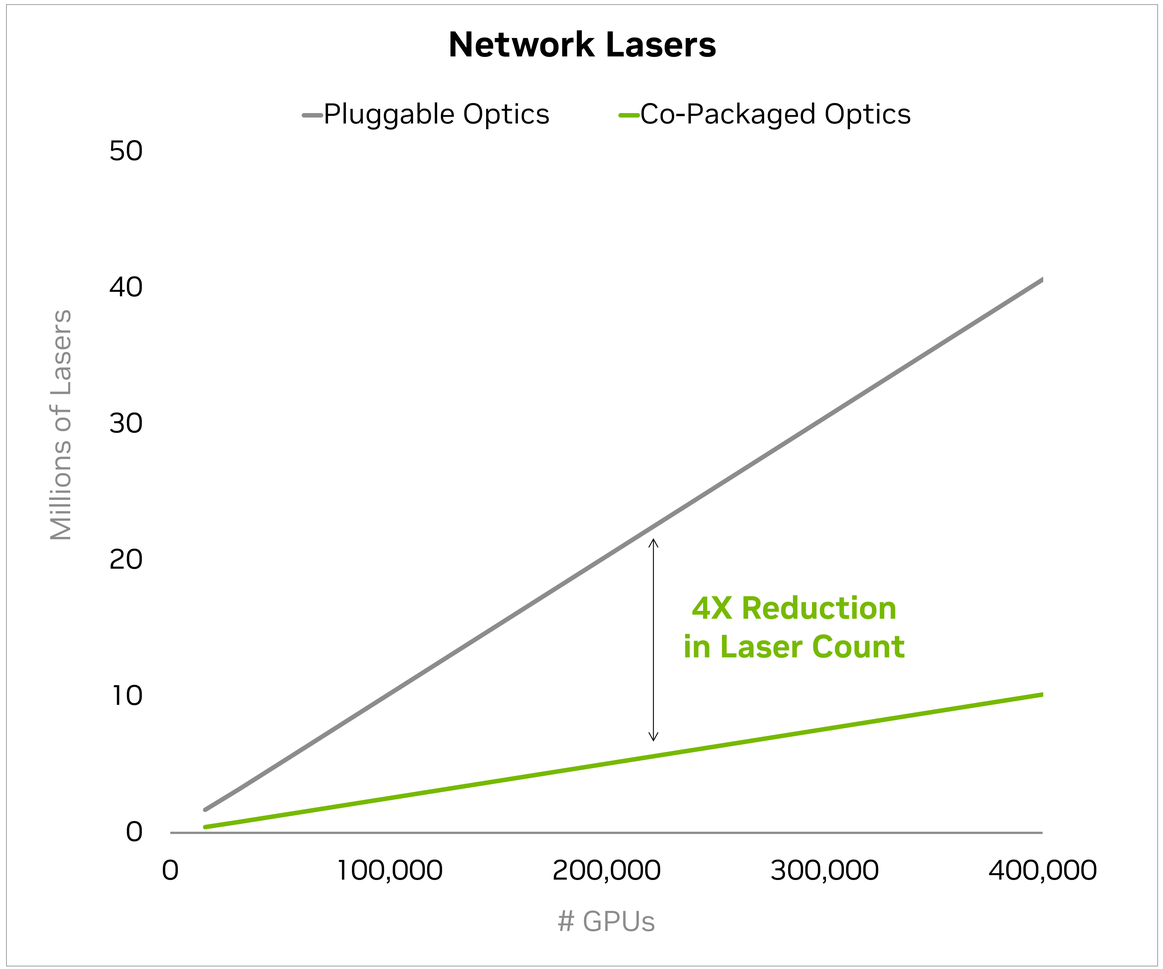
- **2X** Throughput
- **63X** Signal integrity
- **4X** Less Lasers
- **1.6X** Bandwidth density
- **13X** Better laser reliability
- **64** Transceivers replaced




**Spectrum-X Ethernet**
**Integrated Silicon Photonics**

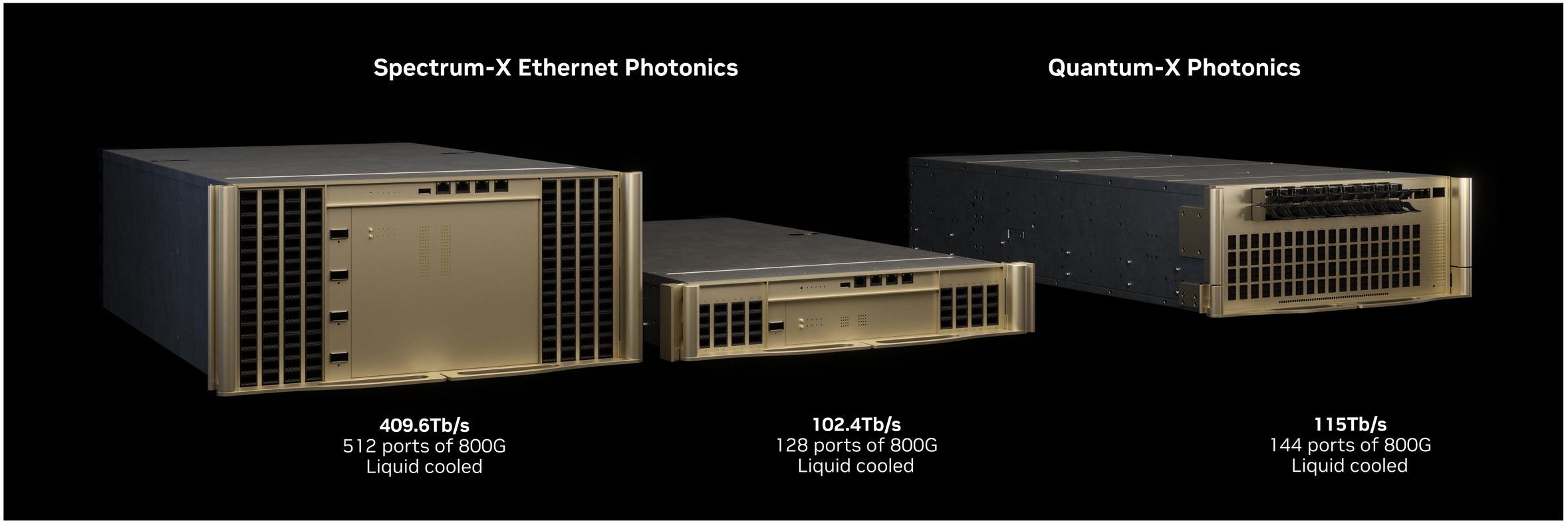
# NVIDIA Photonics Solves Power & Reliability Challenges of AI Scale-out



# NVIDIA Photonics Switch Systems

Co-packaged optics networking switches to scale AI factories to millions of GPUs



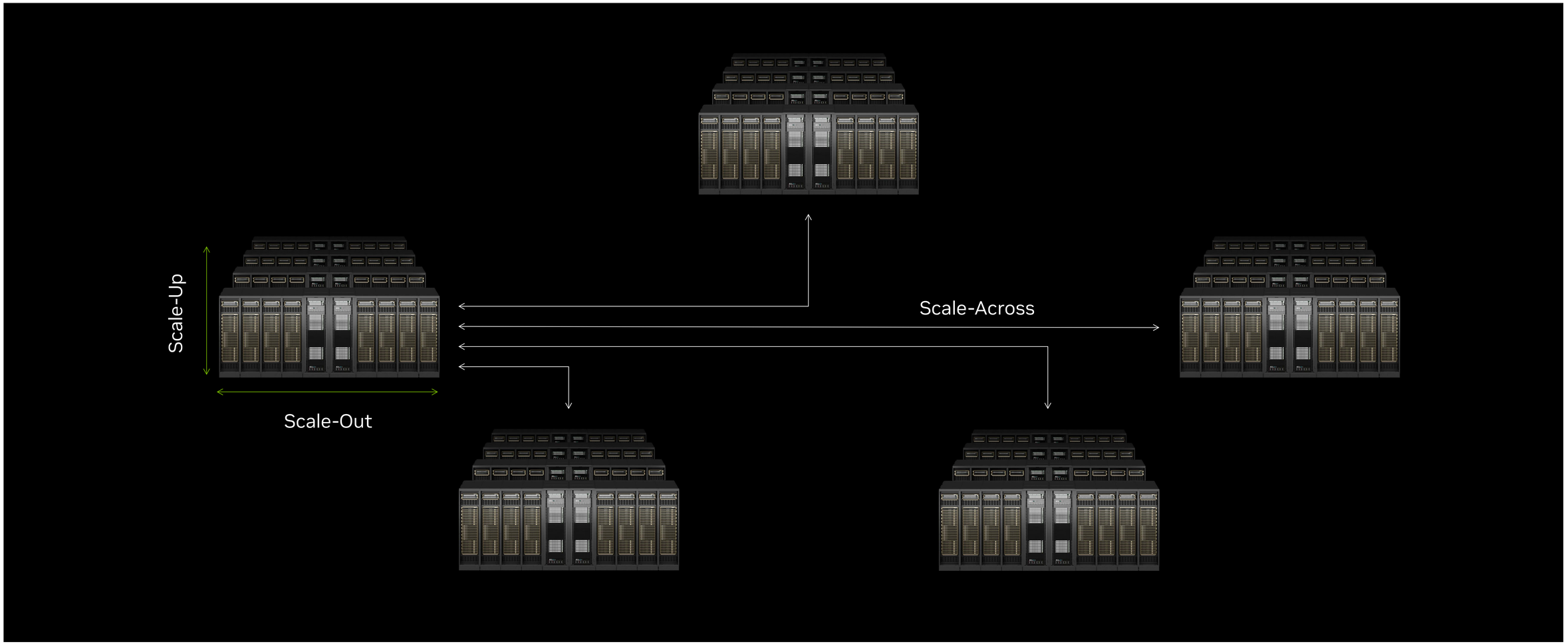

**3.5X**
Power efficiency

**10X**
Higher resiliency

**1.3X**
Faster time to operation

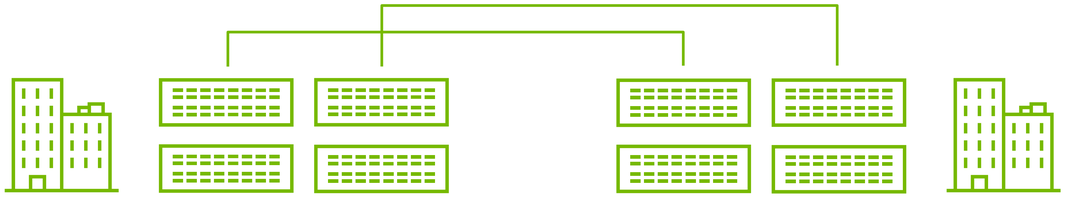
# Scaling AI Beyond The Data Center Requires New Infrastructure

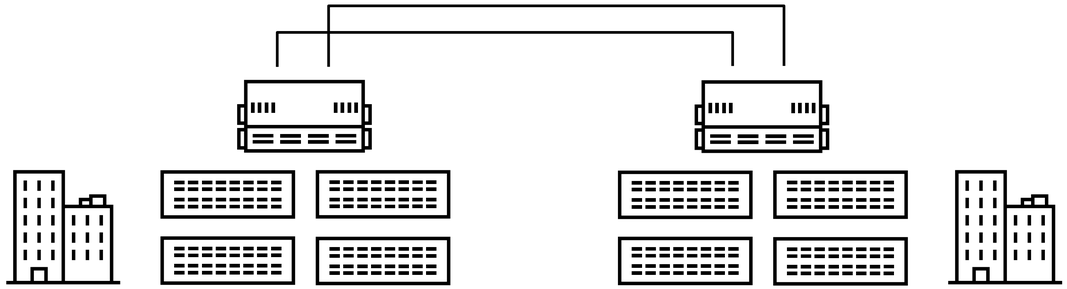
Distributed AI between remote locations, overcoming power and physical limitations



# Announcing Spectrum-XGS Ethernet Clustering Data Centers Into Giga-Scale AI

Unifying multiple data centers into the world's largest supercomputers

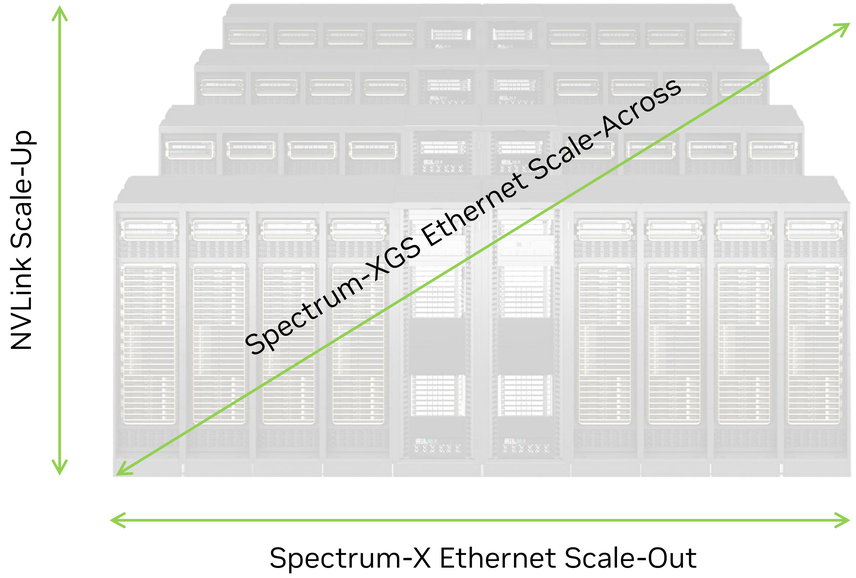


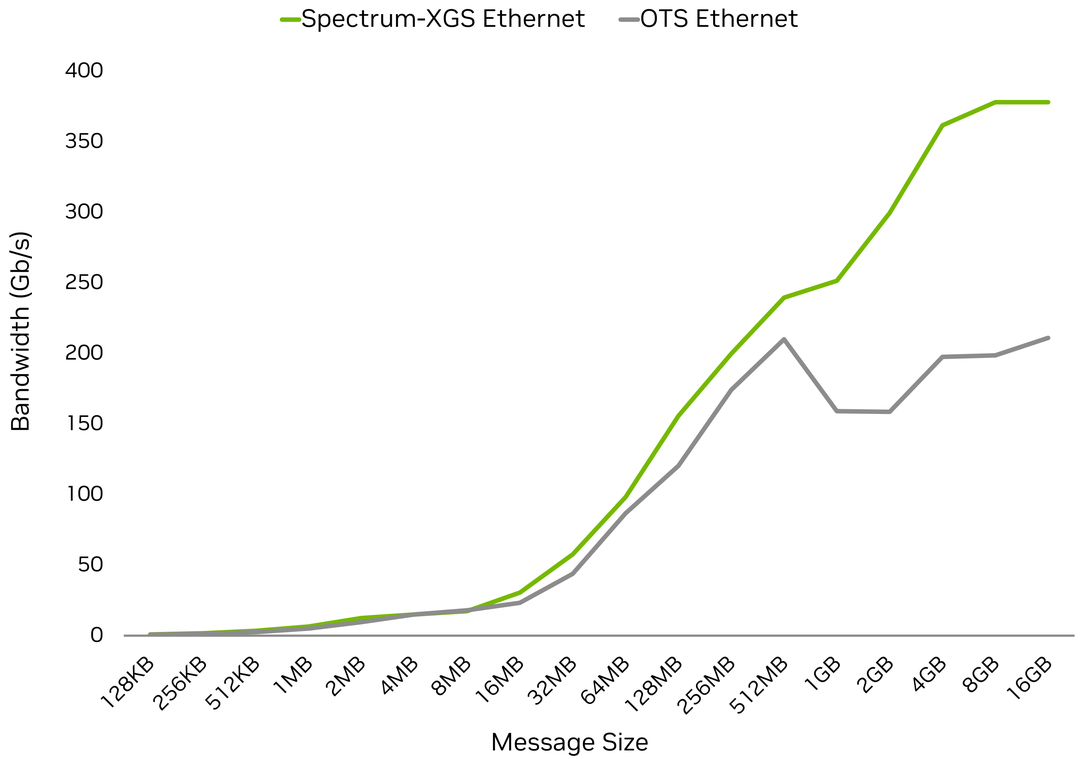


| Spectrum-XGS Ethernet Scale-Across AI | OTS Ethernet |
| --- | --- |
| Fully integrated scale-out and scale-across infrastructures | Separated discrete networks and management |
| End-to-end global telemetry-based congestion control architecture | Deep buffer switch architecture |
| Unlimited scale due to end-to-end architecture | Limited scale due to pre-determined buffer size |
| Auto-adjust load balancing based on scale-across distance | 100s Kilometers of performance penalty regardless of distance |
| Minimize cross-data-center latency | Extreme jitter due to deep buffers |

# 1.9X Higher NCCL Multi-Site Performance With Spectrum-XGS Ethernet

## Auto-adjust load balancing and topology-aware congestion control maximize NCCL throughout



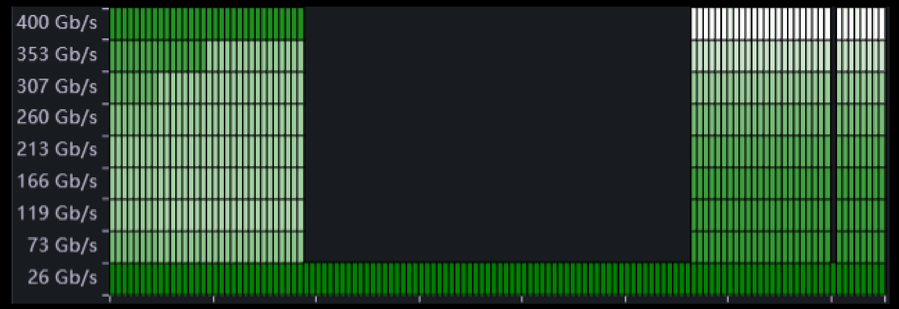




10km distance NCCL All-Reduce between two data centers

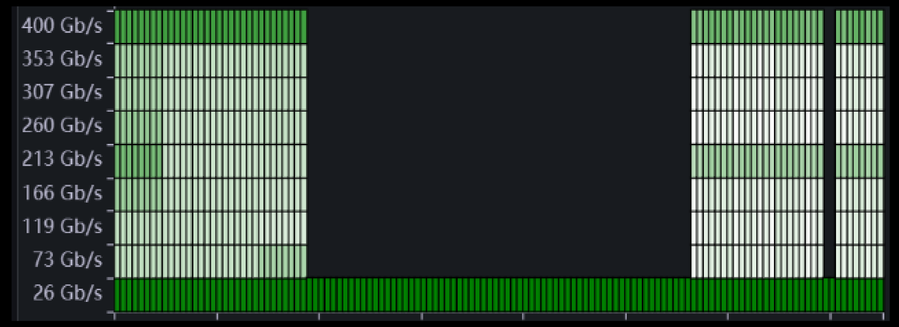
# 1.9X Higher NCCL Multi-Site Performance With Spectrum-XGS Ethernet

Auto-adjust load balancing and topology-aware congestion control maximize NCCL throughout

Spectrum-XGS Ethernet



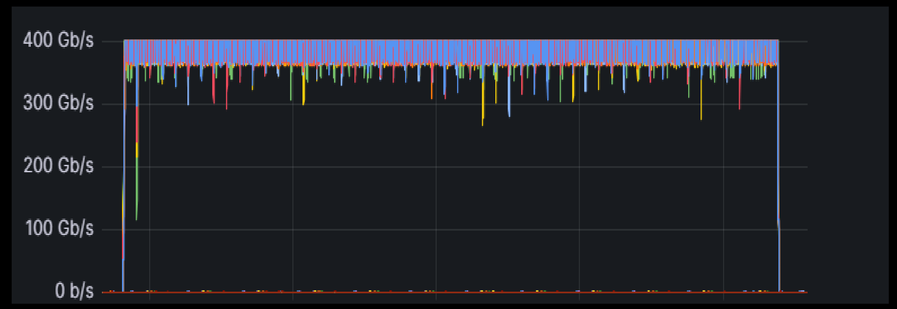

OTS Ethernet




Switch downlink ports

Switch uplink ports

Spectrum-XGS Ethernet

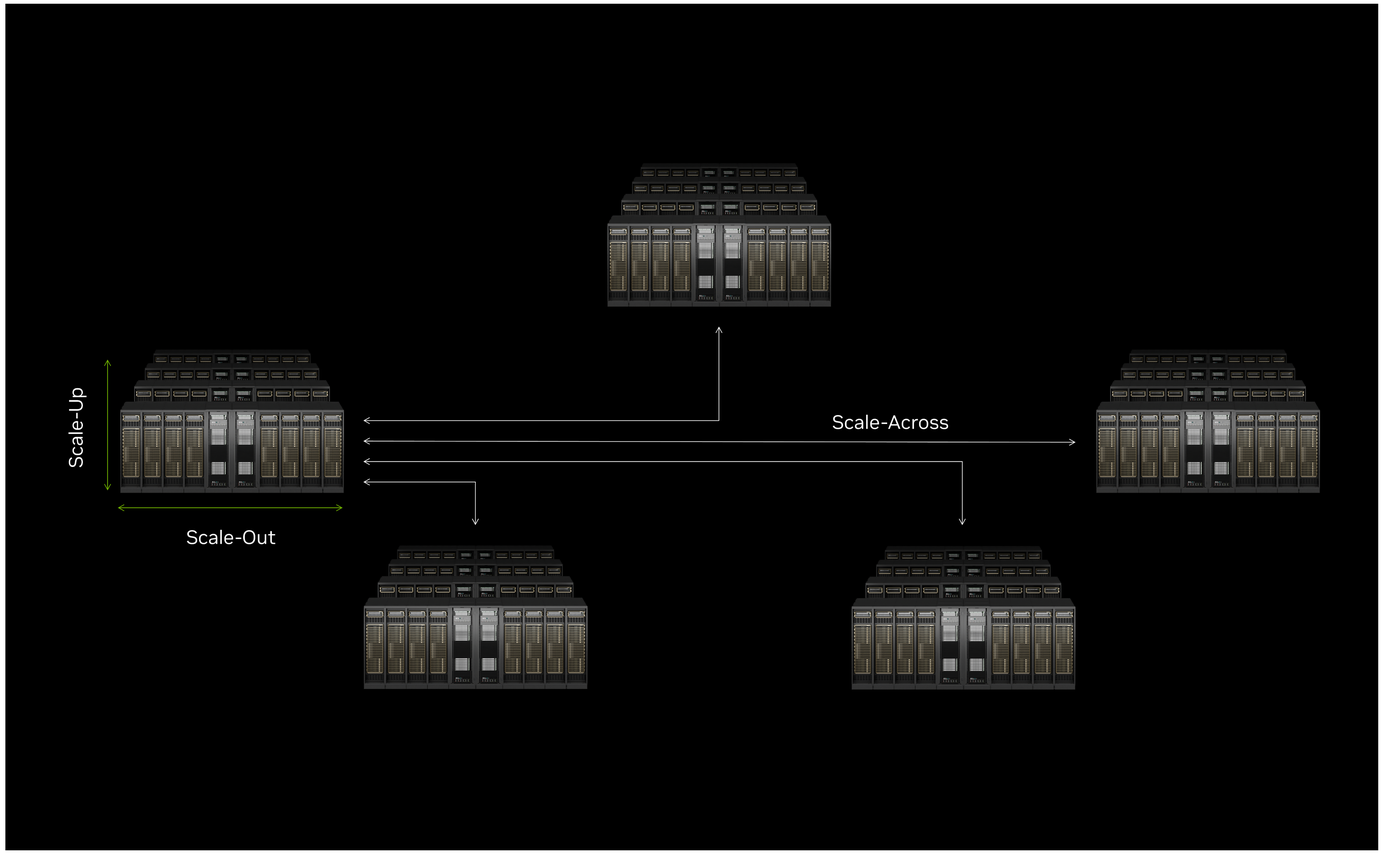



OTS Ethernet


Host bandwidth

# Spectrum-XGS Ethernet Clustering Data Centers Into Giga-Scale AI

1.9X higher multi-site NCCL performance compared to OTS Ethernet




- Scaling AI beyond the data center requires new infrastructure
- Spectrum-XGS Ethernet unifying multiple data centers into the world's largest supercomputers
- Integrating scale-out and scale-across
- Auto-adjust load balancing based on scale-across distance
- 1.9x higher NCCL multi-site performance