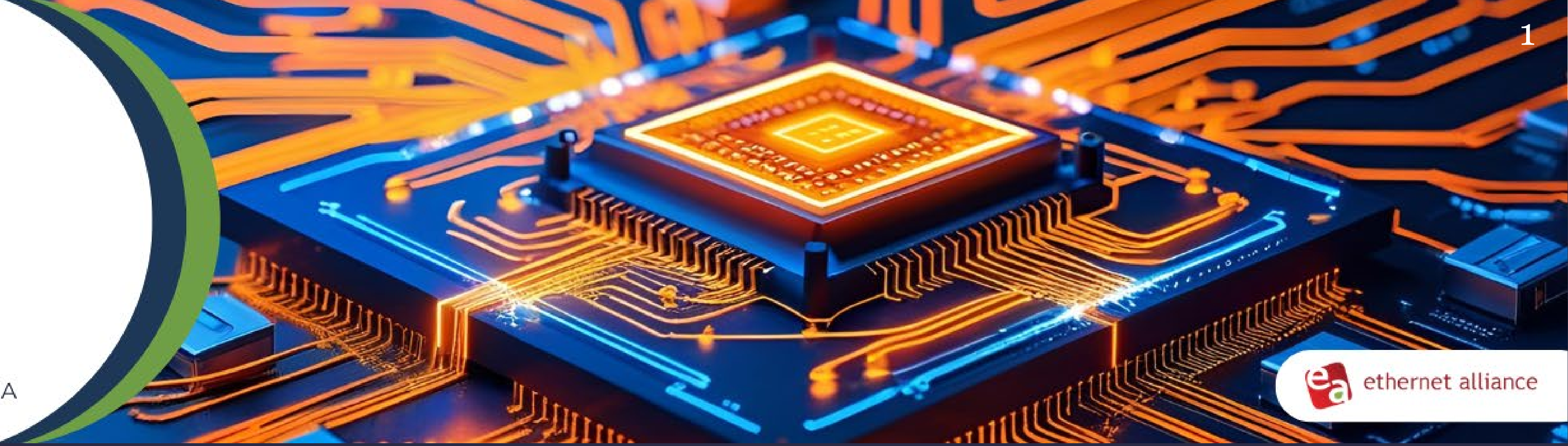
TEF 2025
Ethernet for AI

December 2-3, 2025
Hyatt Centric Mountain View, CA, USA



400 Gb/s Optics for AI Networks

December 2-3, 2025

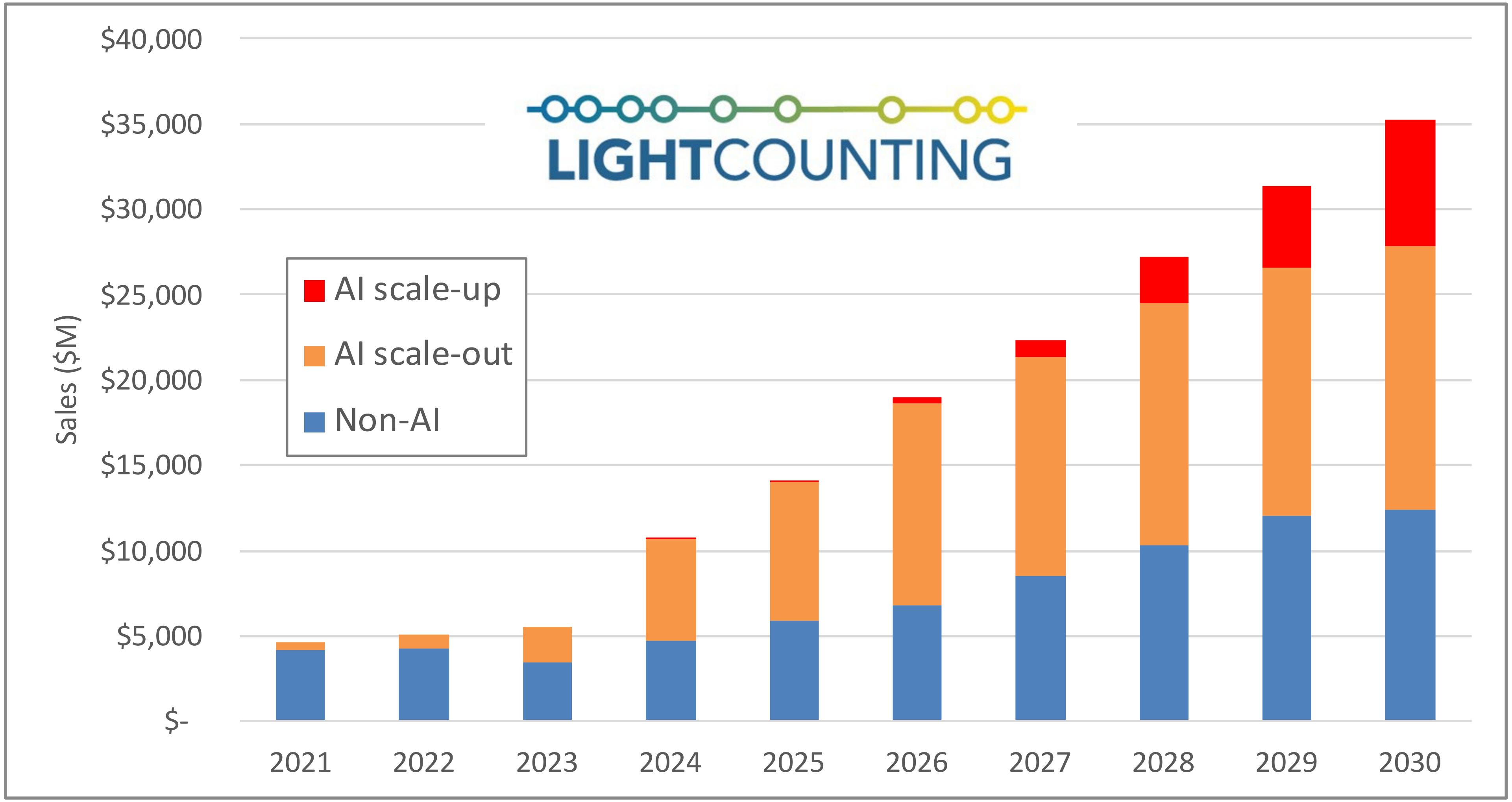
This presentation has been developed within the Ethernet Alliance, and is intended to educate and promote the exchange of information. Opinions expressed during this presentation are the views of the presenters, and should not be considered the views or positions of the Ethernet Alliance

Integrated optics for AI Clusters

Vladimir Kozlov, CEO of LightCounting

Impact of AI on the Ethernet Transceiver Market

Two key factors: Investments into AI skyrocketed and Nvidia started to use transceivers instead of AOCs in 2023. Growth in investments continues to exceed our forecast.

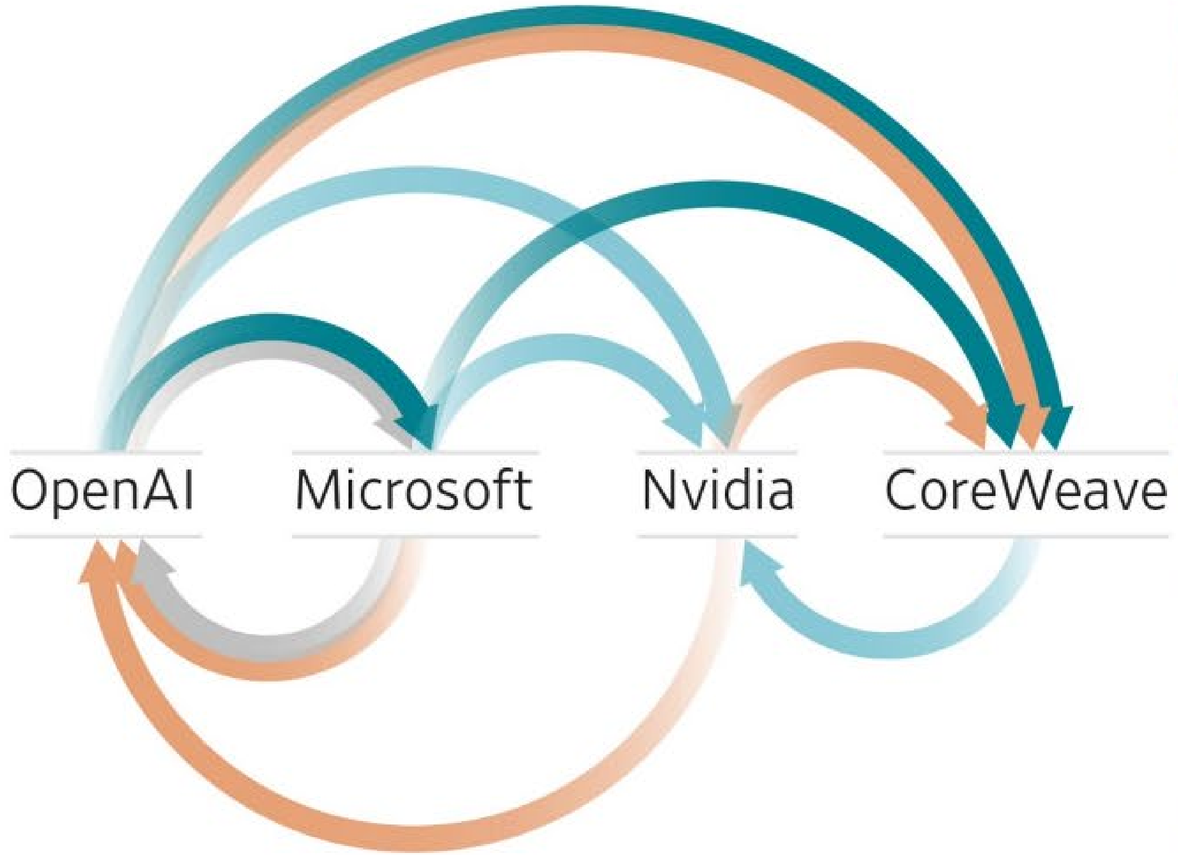


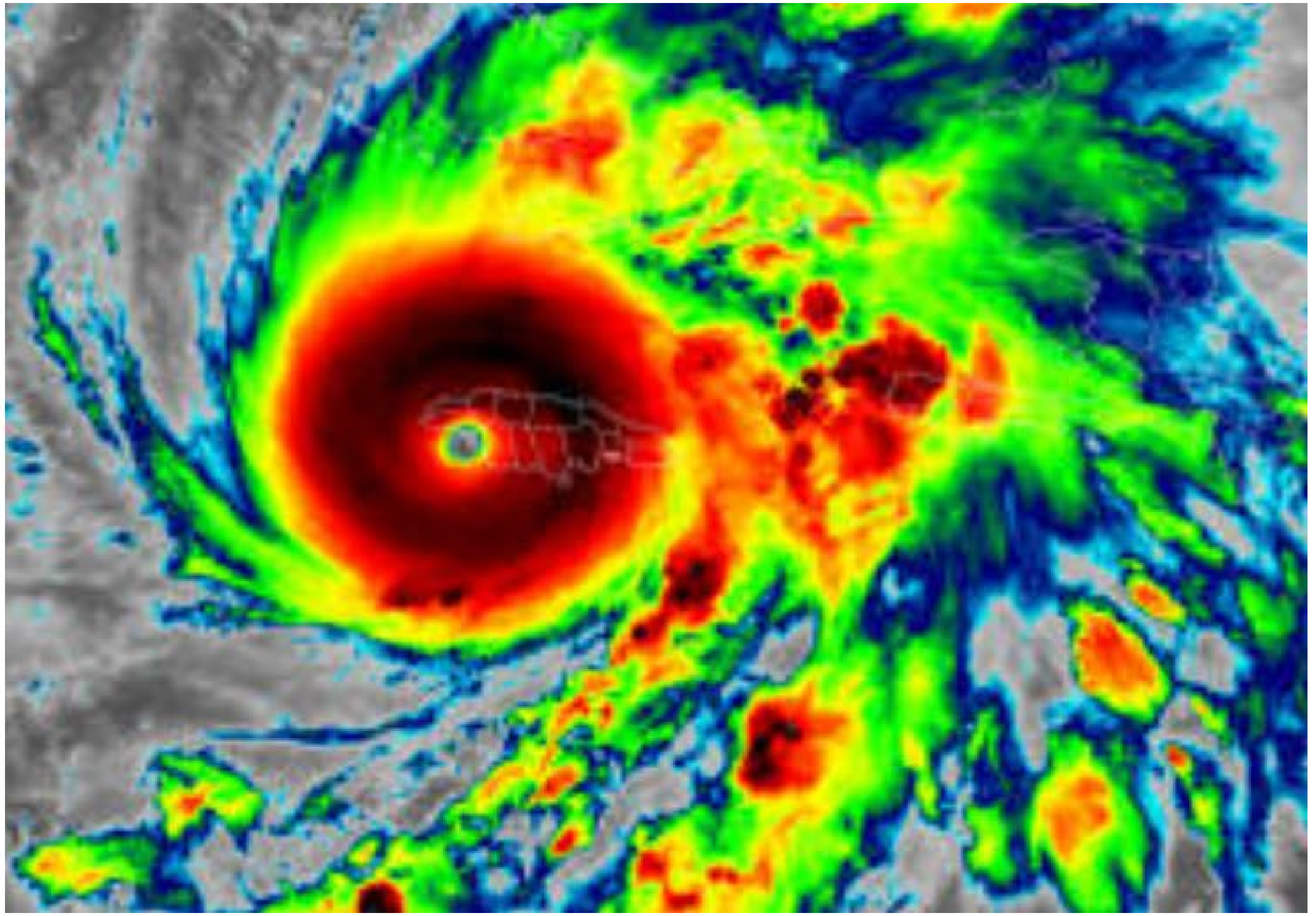
Source: Cloud Data Center Optics – July 2025

TEF 2025
Ethernet for AI

The AI Hurricane

How will this end?



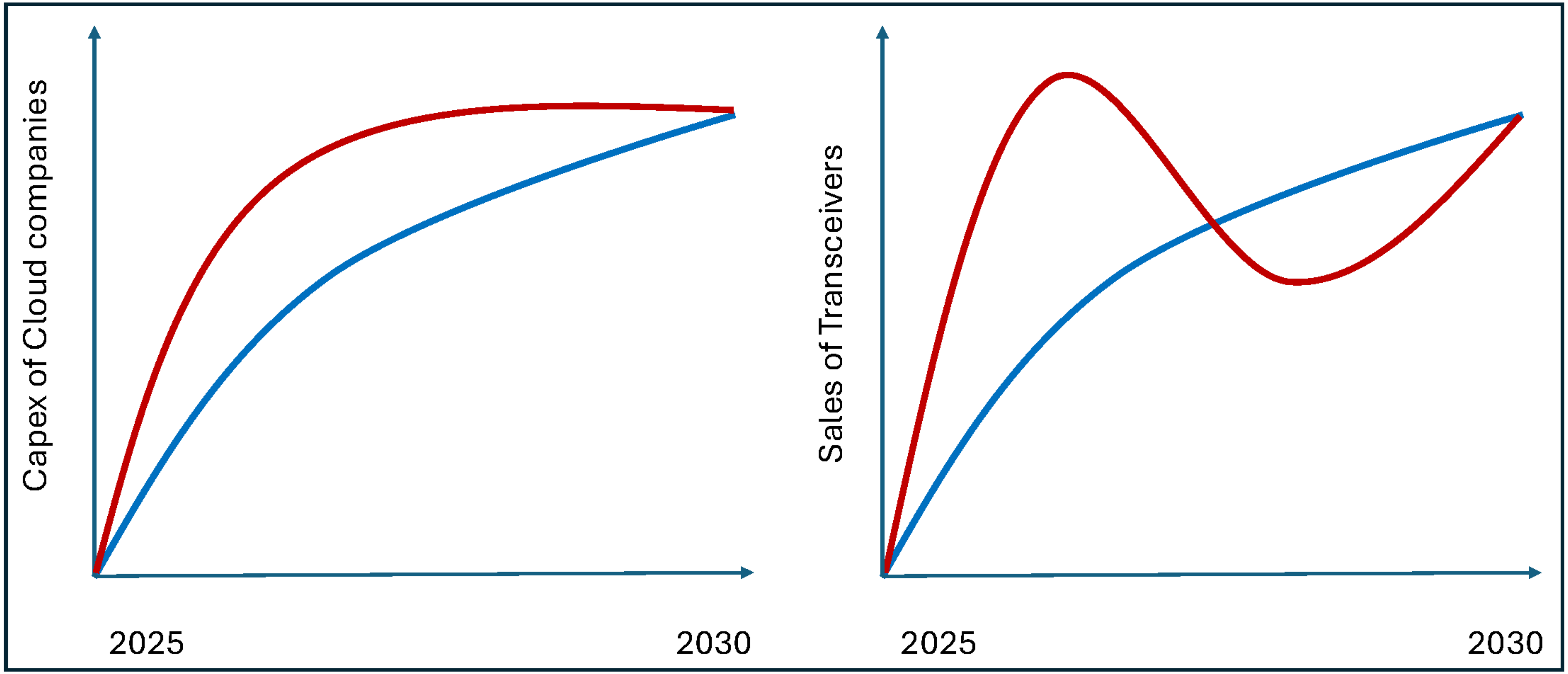


TEF 2025 Ethernet for AI

Source: Market Forecast – October 2025

Two scenarios for the forecast

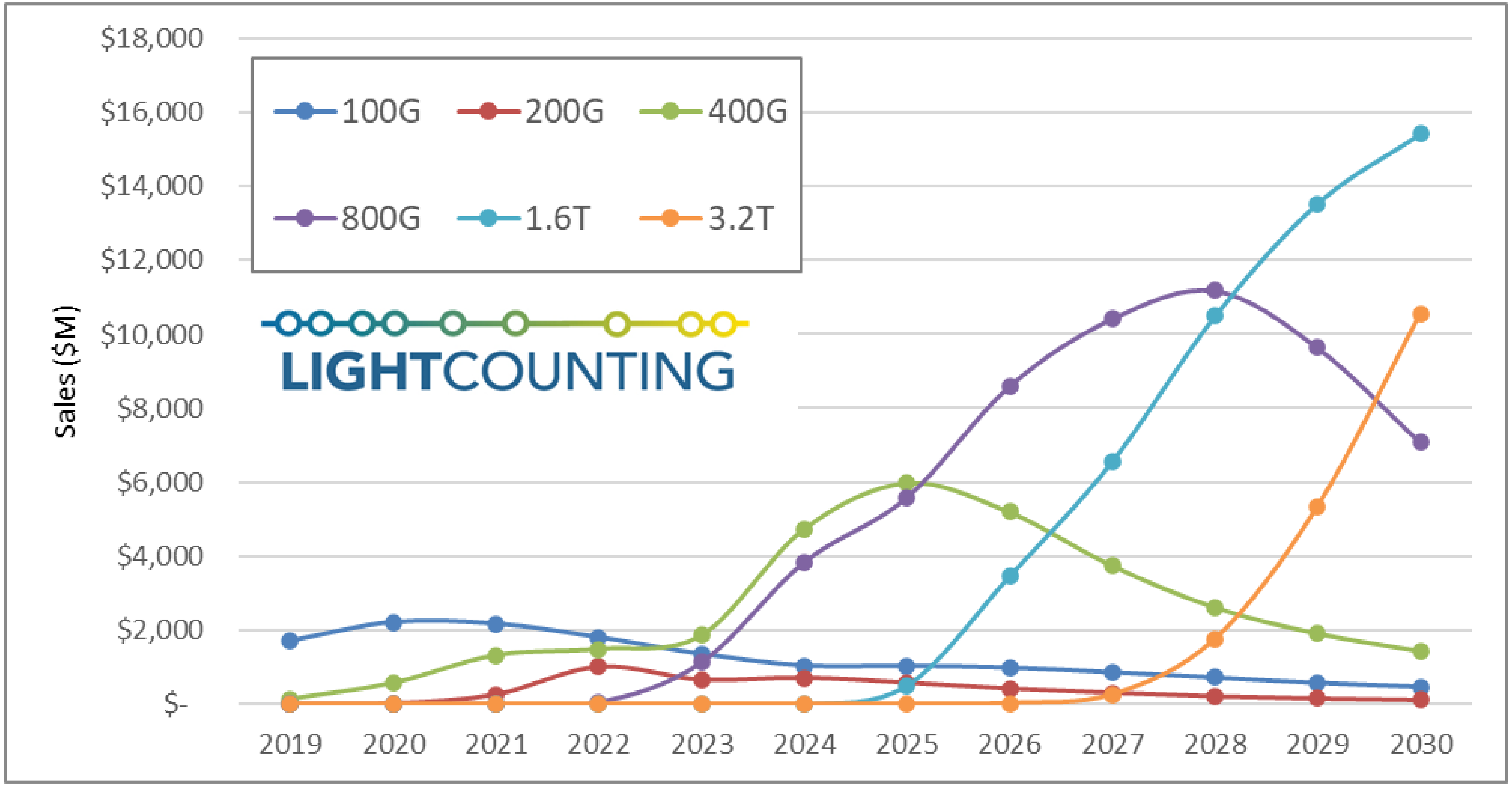
The industry supply chain chain is non-linear.



Source: Ethernet Optics – September 2025

Ethernet Transceivers by data rate (total market)

This data includes LPO and CPO for scale-out and scale-up applications.



TEF 2025 Ethernet for AI

Source: Cloud Data Center Optics – July 2025

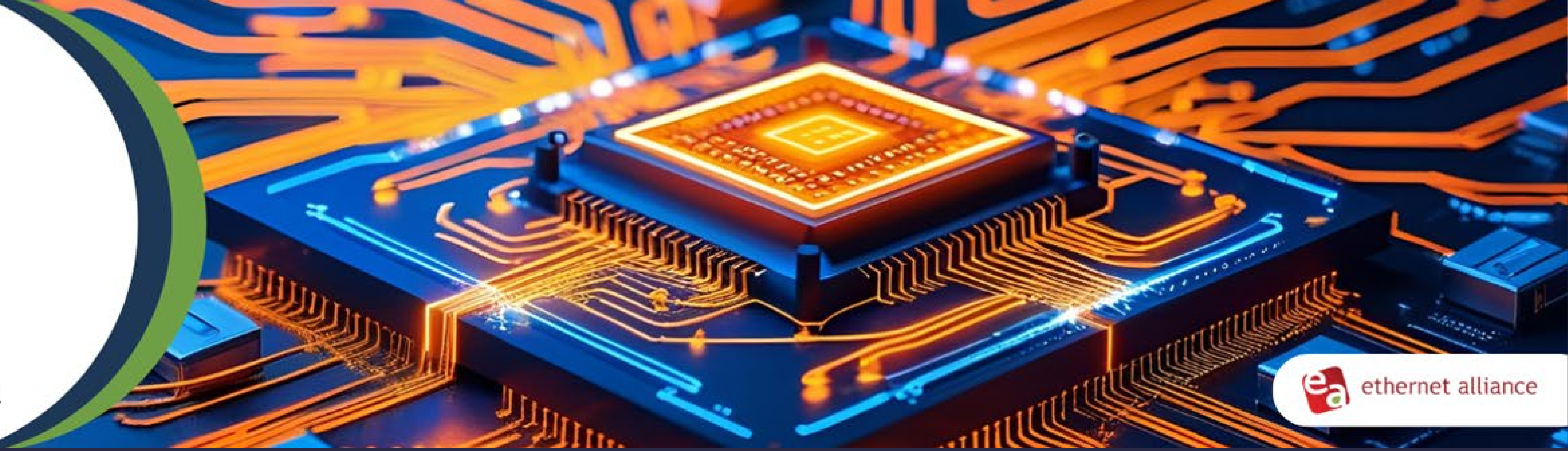
Panelists

- Gilad Shainer, NVIDIA, “Co-Packaged Silicon Photonics Switches for Gigawatt AI Factories”
- Jose M. Castro, Panduit, “Enabling Massive Scale-Out AI Networks with Ethernet and Optical Lane Breakouts”
- Naim Ben-Hamida, Ciena, “448G technology for next gen networking”

QUESTIONS?

TEF 2025
Ethernet for AI

December 2-3, 2025
Hyatt Centric Mountain View, CA, USA



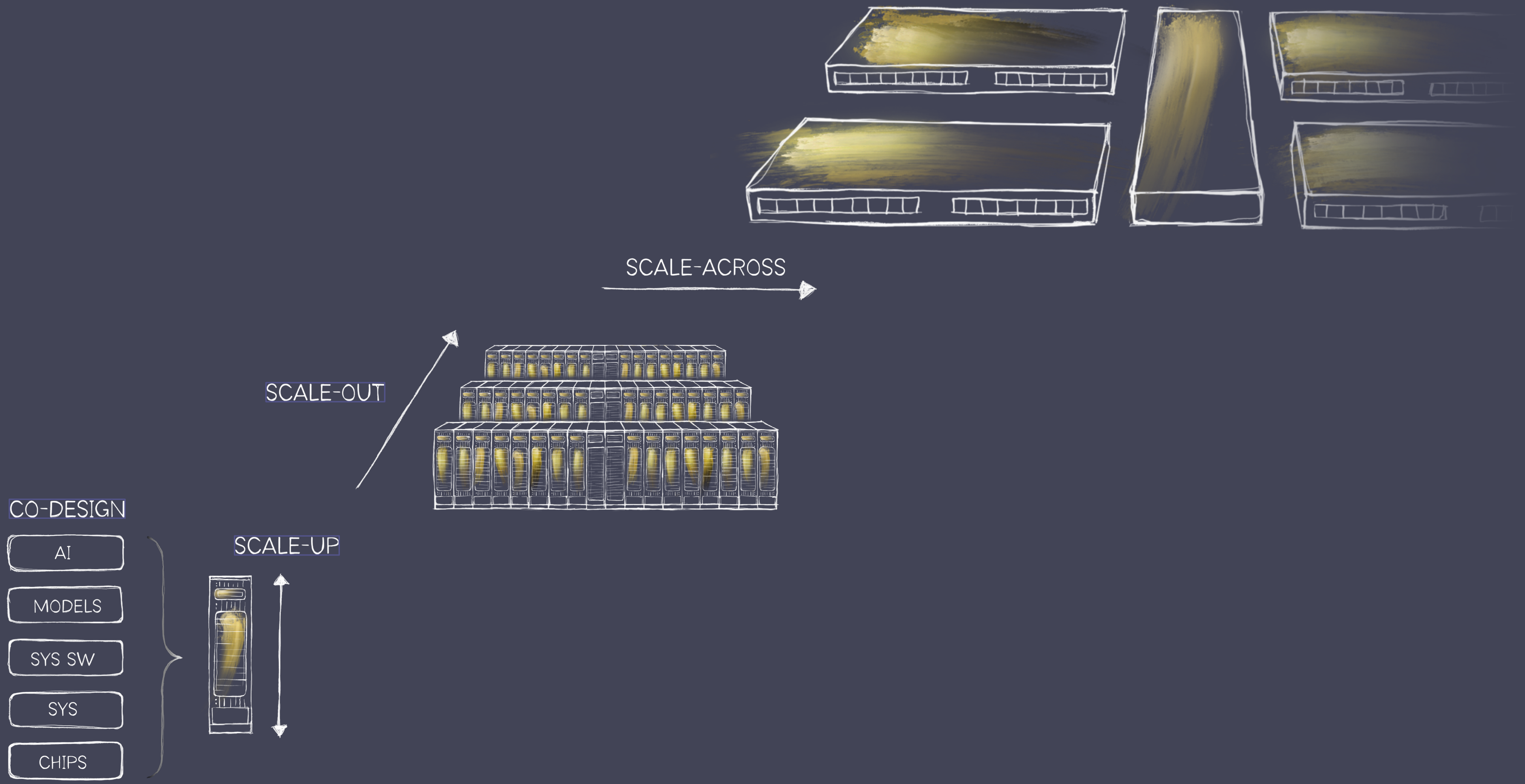
ethernet alliance

Co-Packaged Silicon Photonics Switches for Gigawatt AI Factories

Gilad Shainer, Senior Vice President of Networking | Dec 2025

NVIDIA

The Giga-Scale AI Factory



TEF 2025 Ethernet for AI

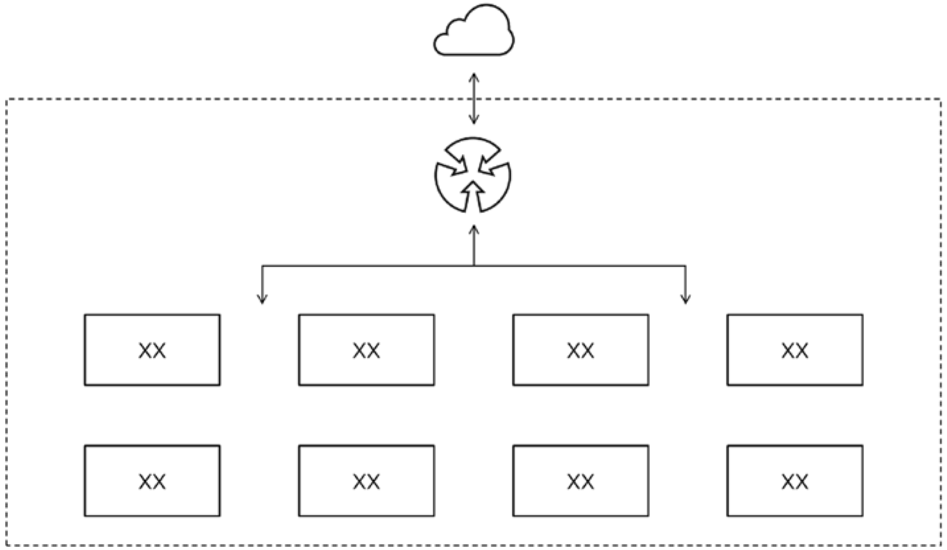
NVIDIA

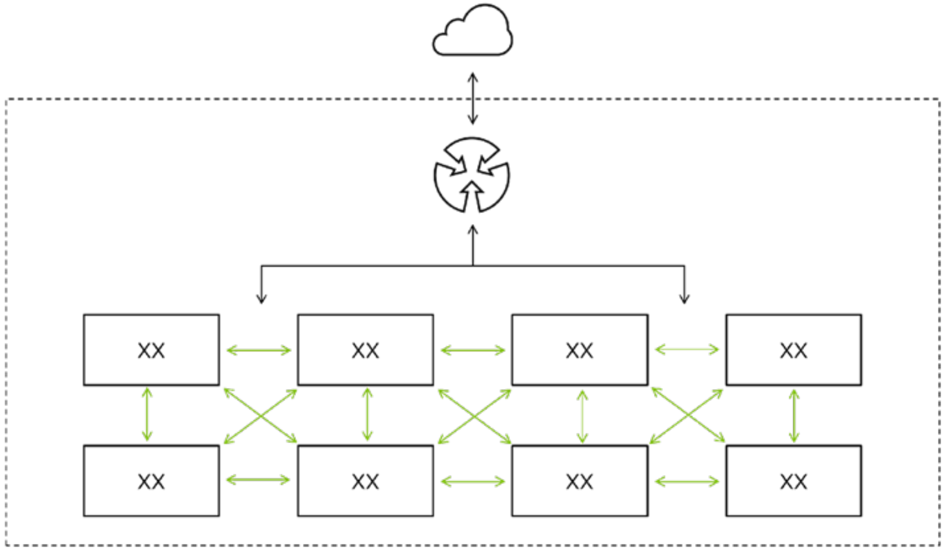
The Different Ethernet Architectures

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

AI Ethernet

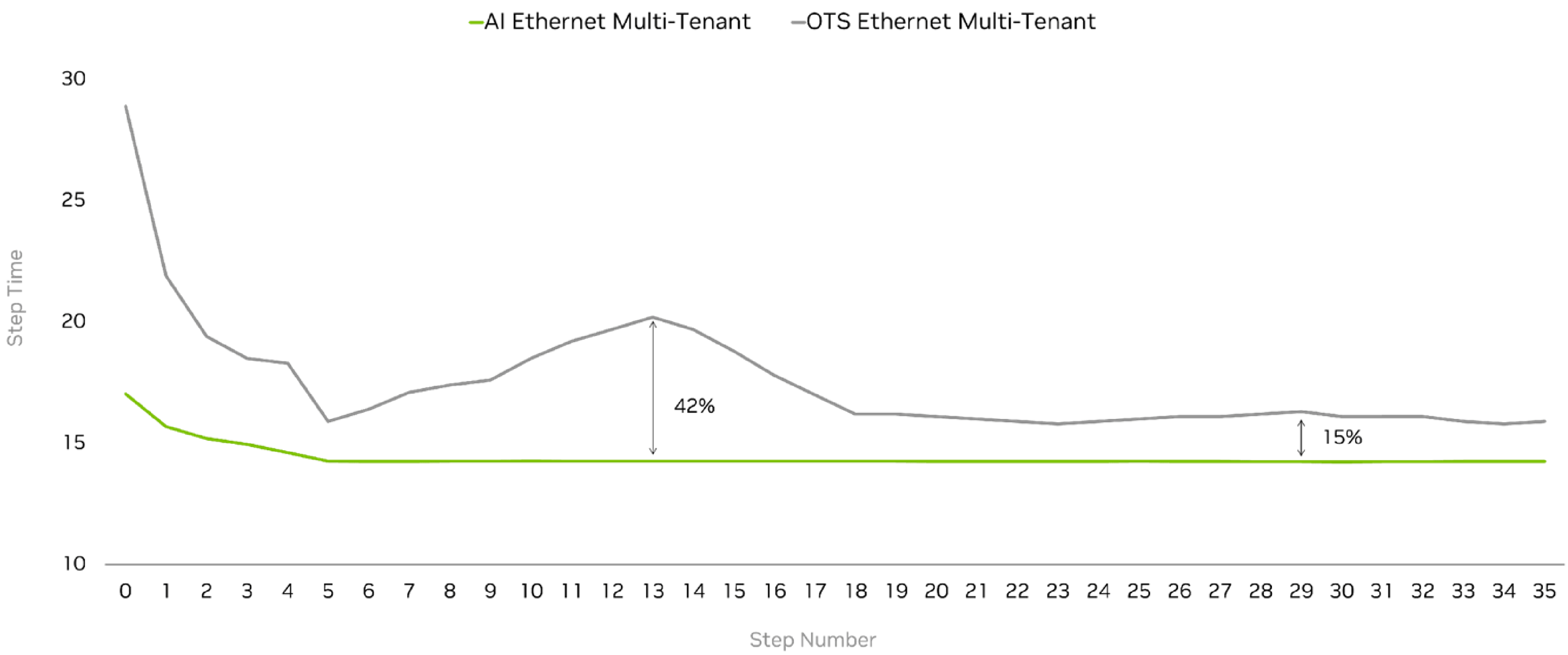
The network defines the data center





OTS Ethernet - Hyperscale Clouds	AI Ethernet
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

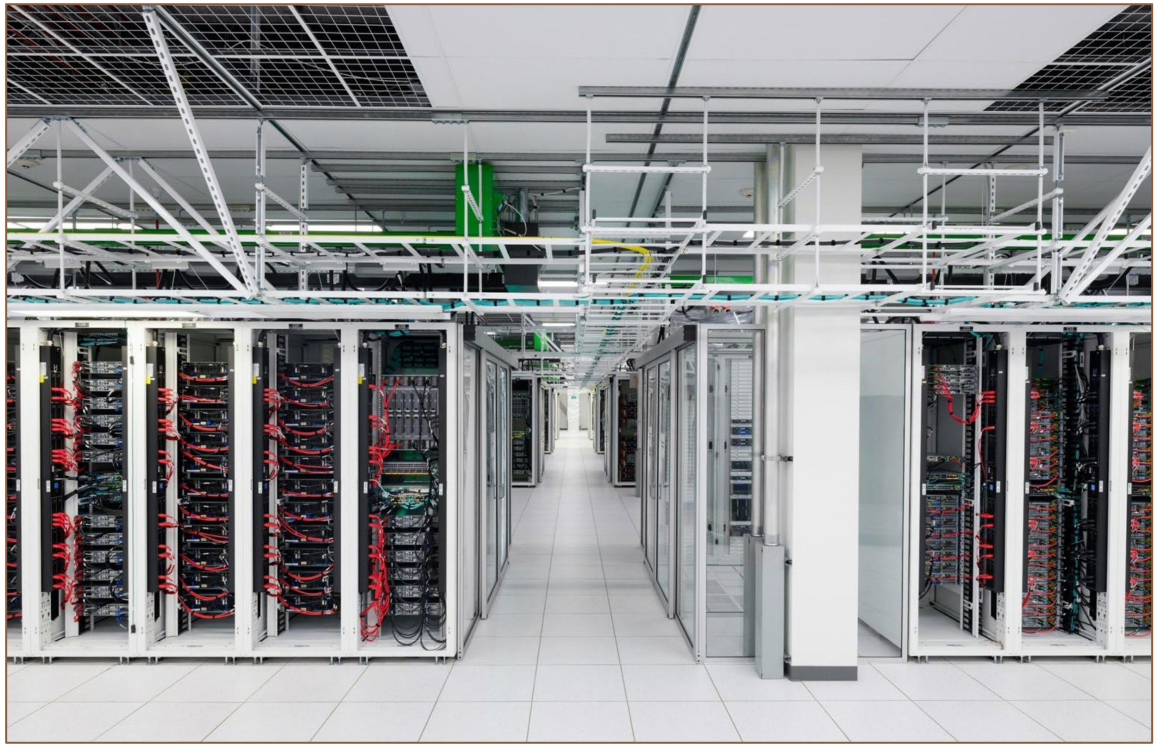
1.4X Higher LLAMA3 70B Training Performance in Multi-Tenant Data Center



TEF 2025
Ethernet for AI

Scale-Out and AI Density Depend on Optical Connectivity

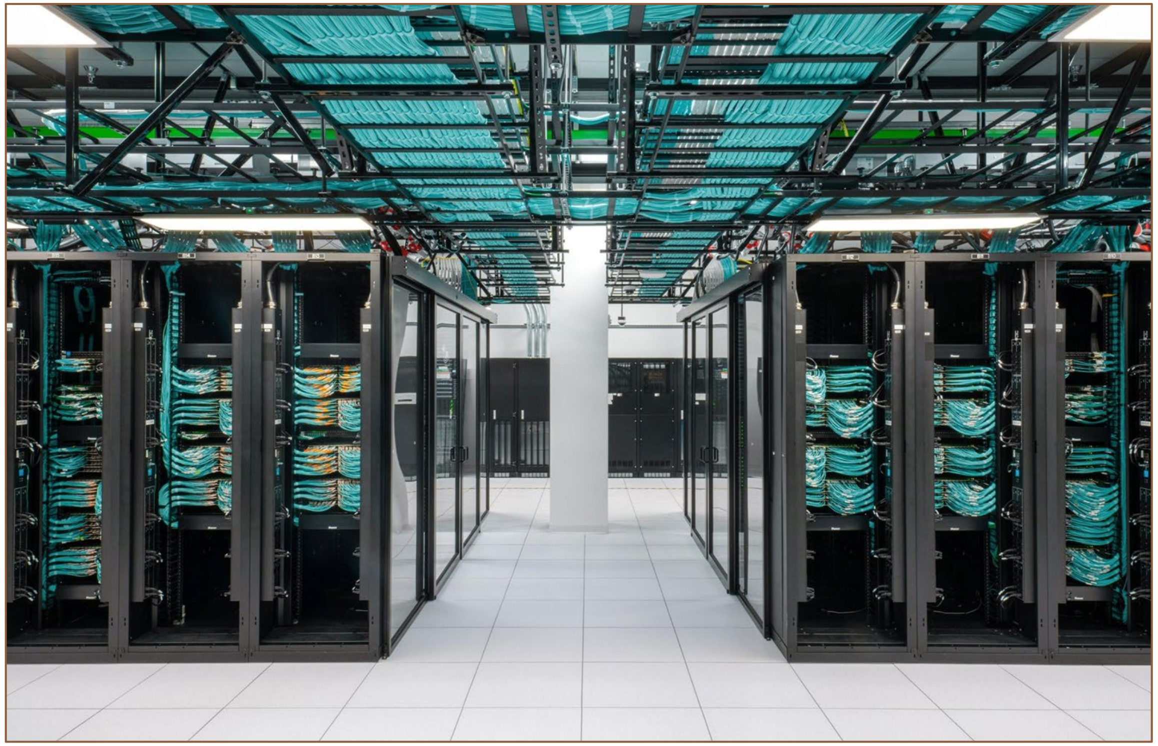
The optical network power consumption represents 10% of compute resources



Traditional Cloud Data Center

100K
Servers

2.3 MW
Transceiver Power



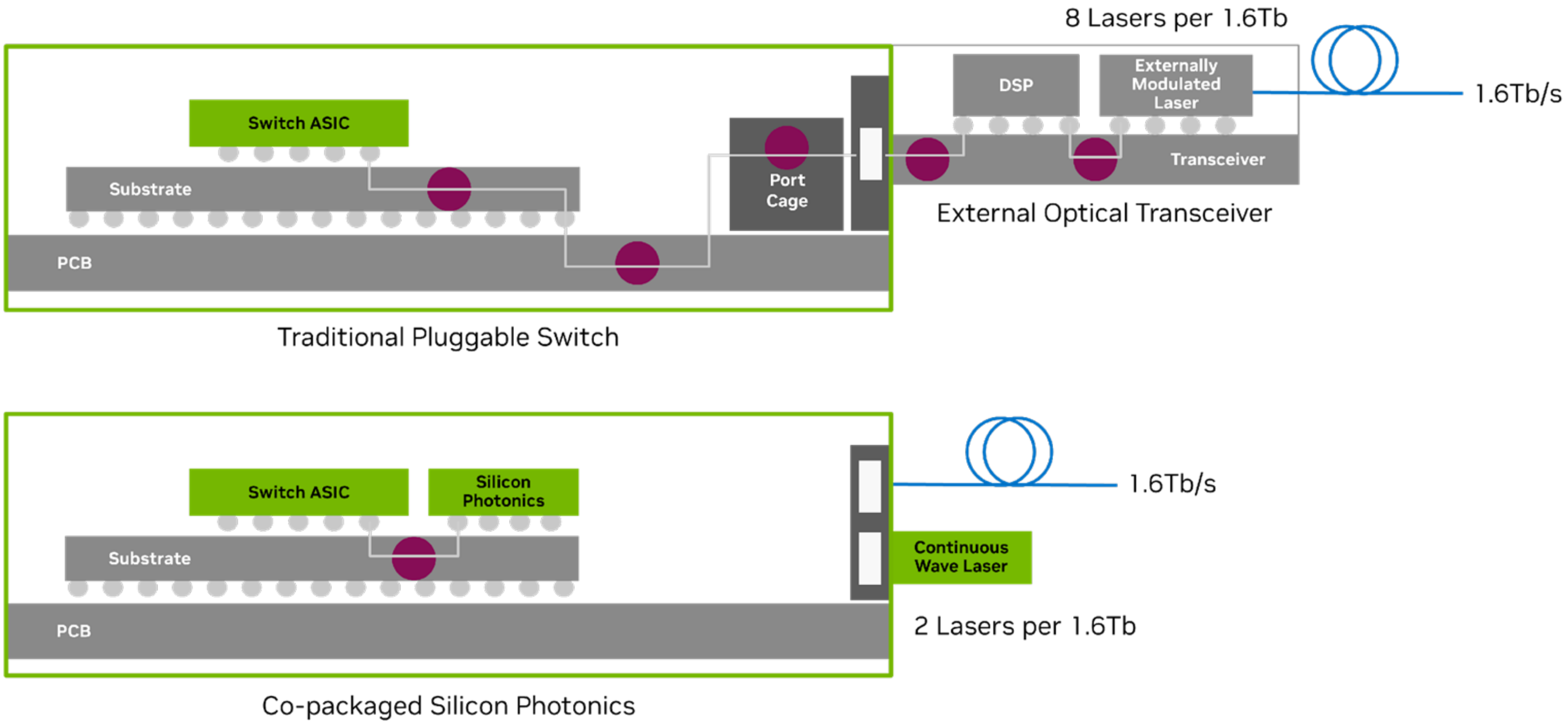
AI Factory

100K
Servers

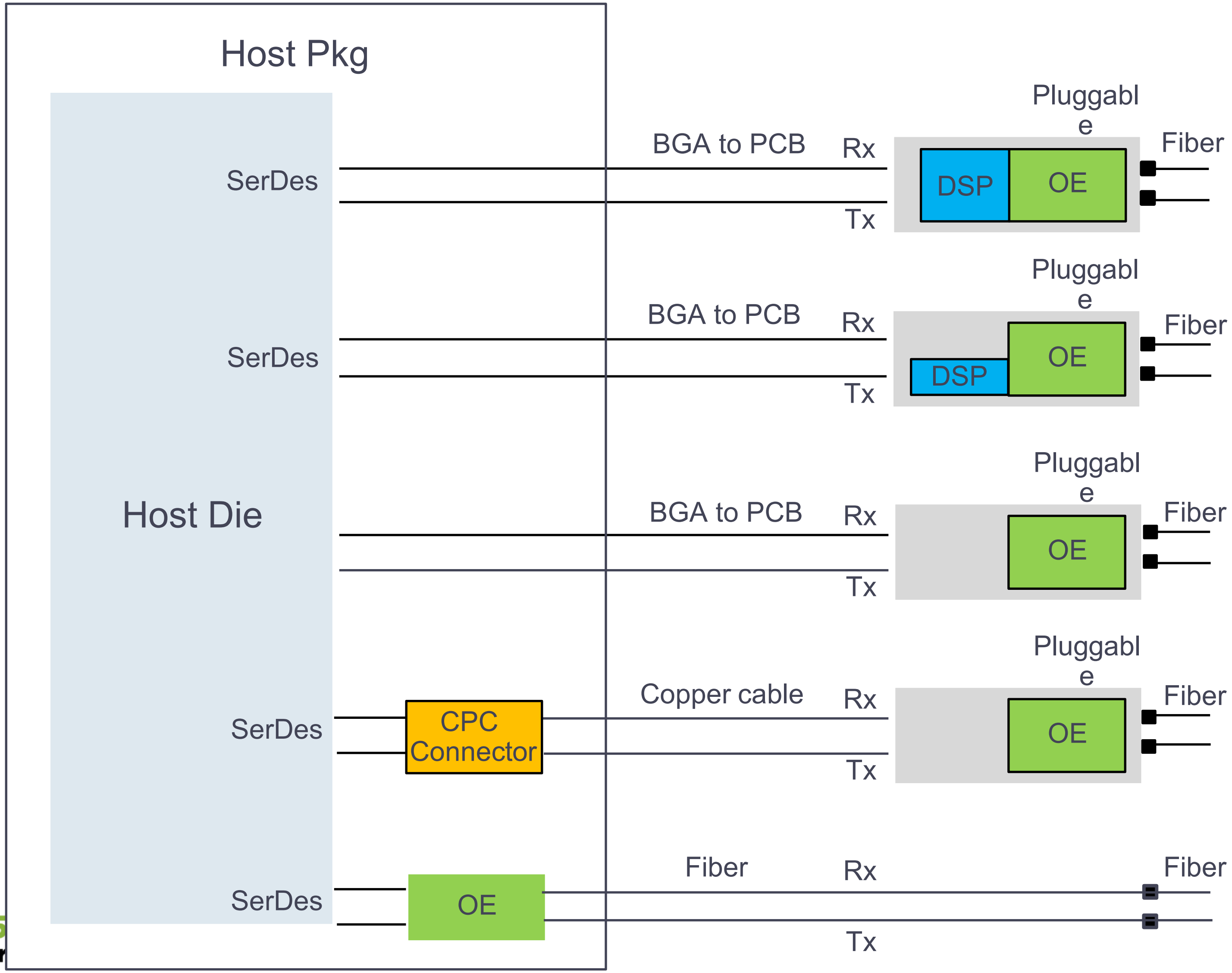
40 MW
Transceiver Power

TEF 2025 Ethernet for AI

Introducing Co-Packaged Silicon Photonics



Optics Options



	pJ/b	Power @1.6T
FRO (Full- Retimed Optics)	14	25W
TRO (Tx-Retimed Optics)	10	18W
LPO (Linear Pluggable Optics)	6.4	11W
LPO w/ CPC (Co-Packaged Copper)	6.4	11W
CPO (Co-Packaged Optics)	4	7W

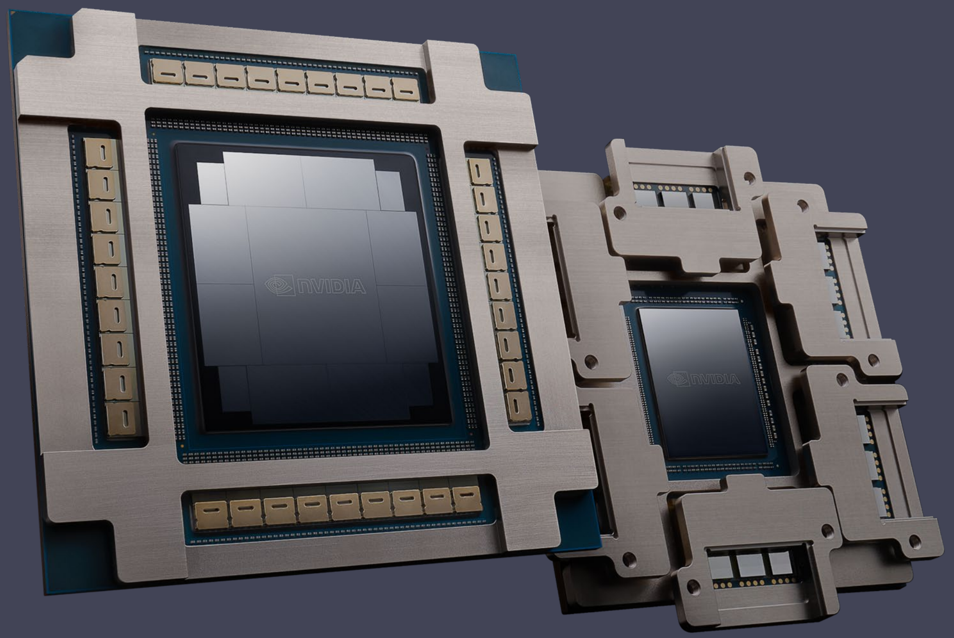
NVIDIA Photonics

CPO co-invention with ecosystem partners

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

波若威科技 BROWAVE CORPORATION | COHERENT | CORNING | fabrinet | FOXCONN

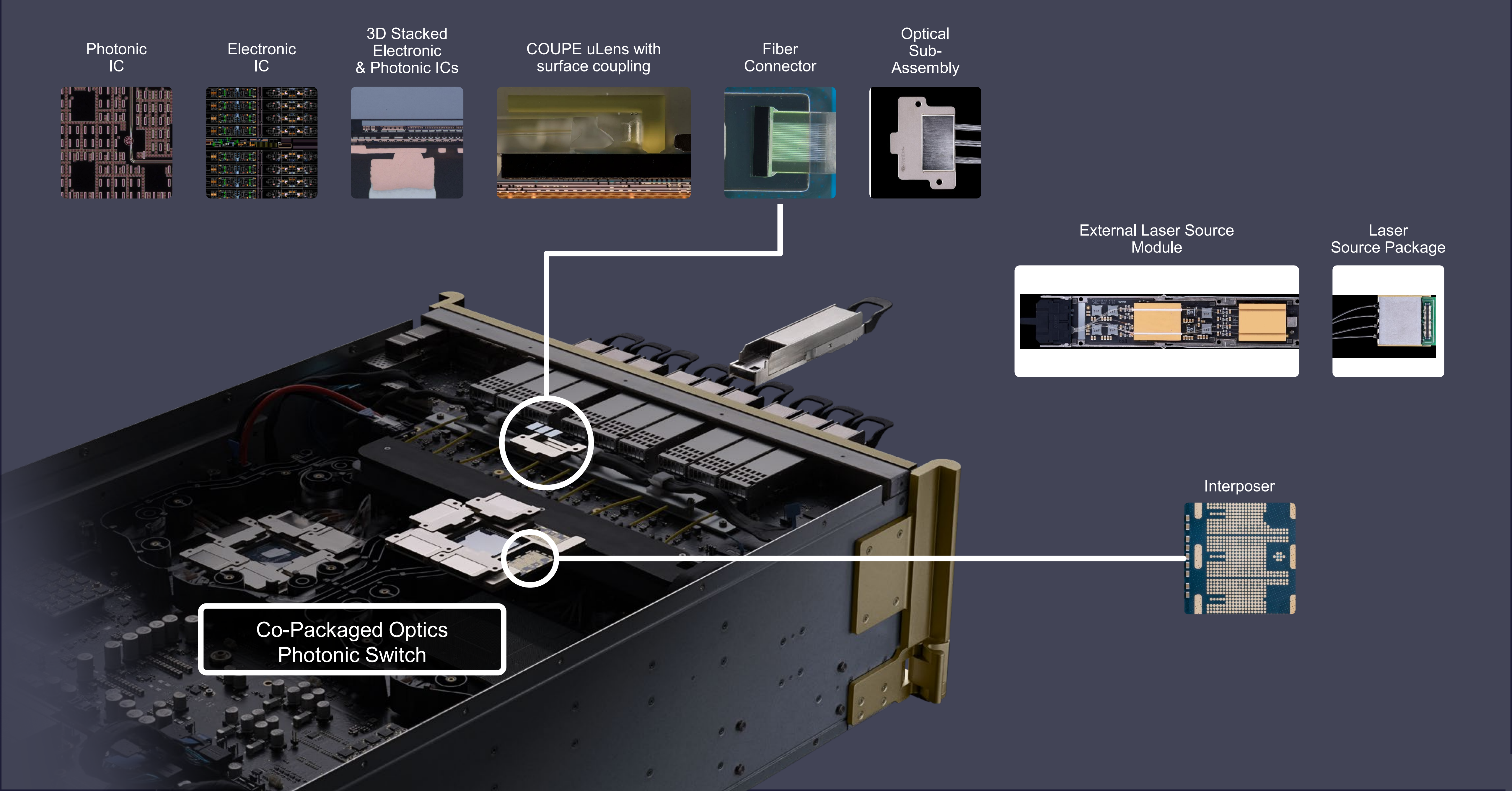
LUMENTUM | SENKO Advanced Components | SPIL | SUMITOMO ELECTRIC | TFC | tsmc



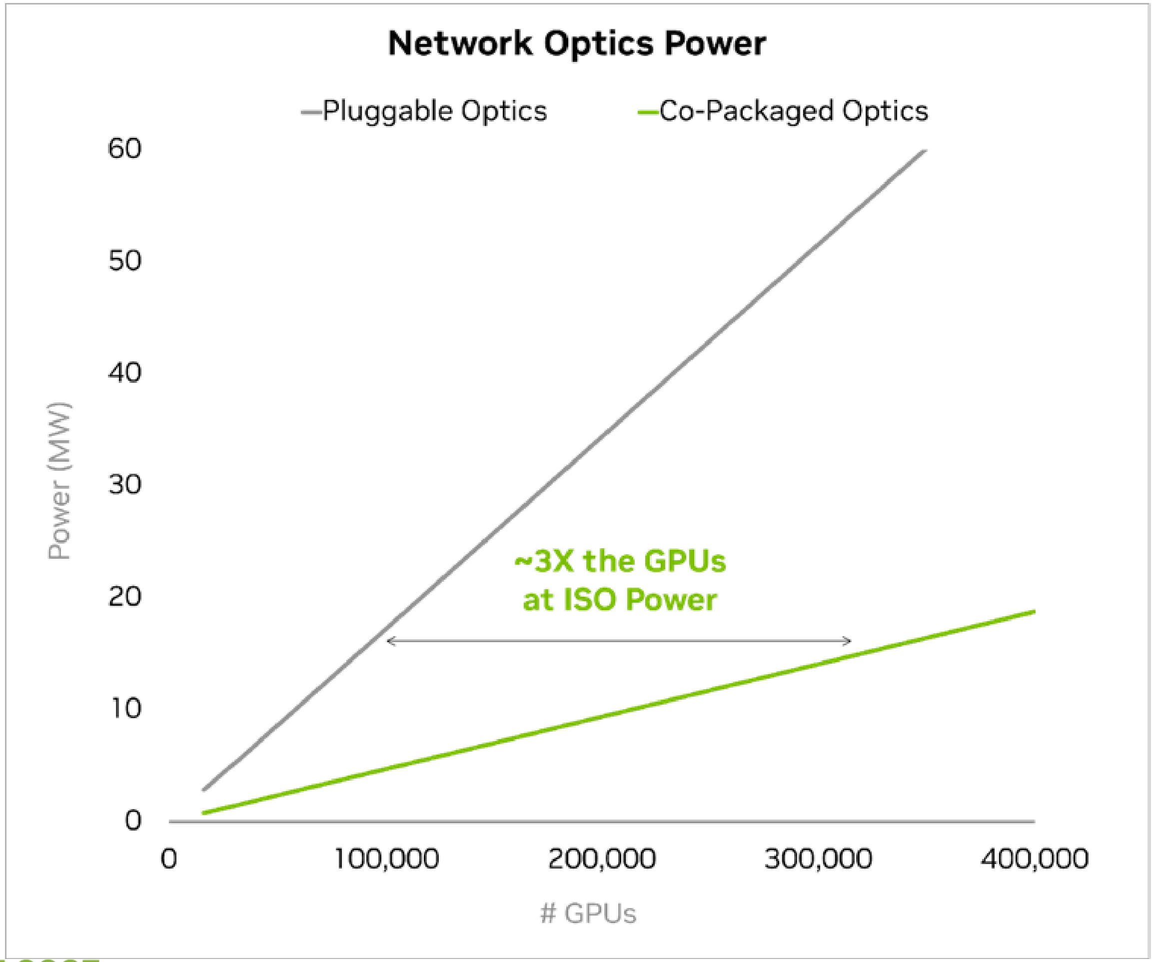
Ethernet
Integrated Silicon Photonics

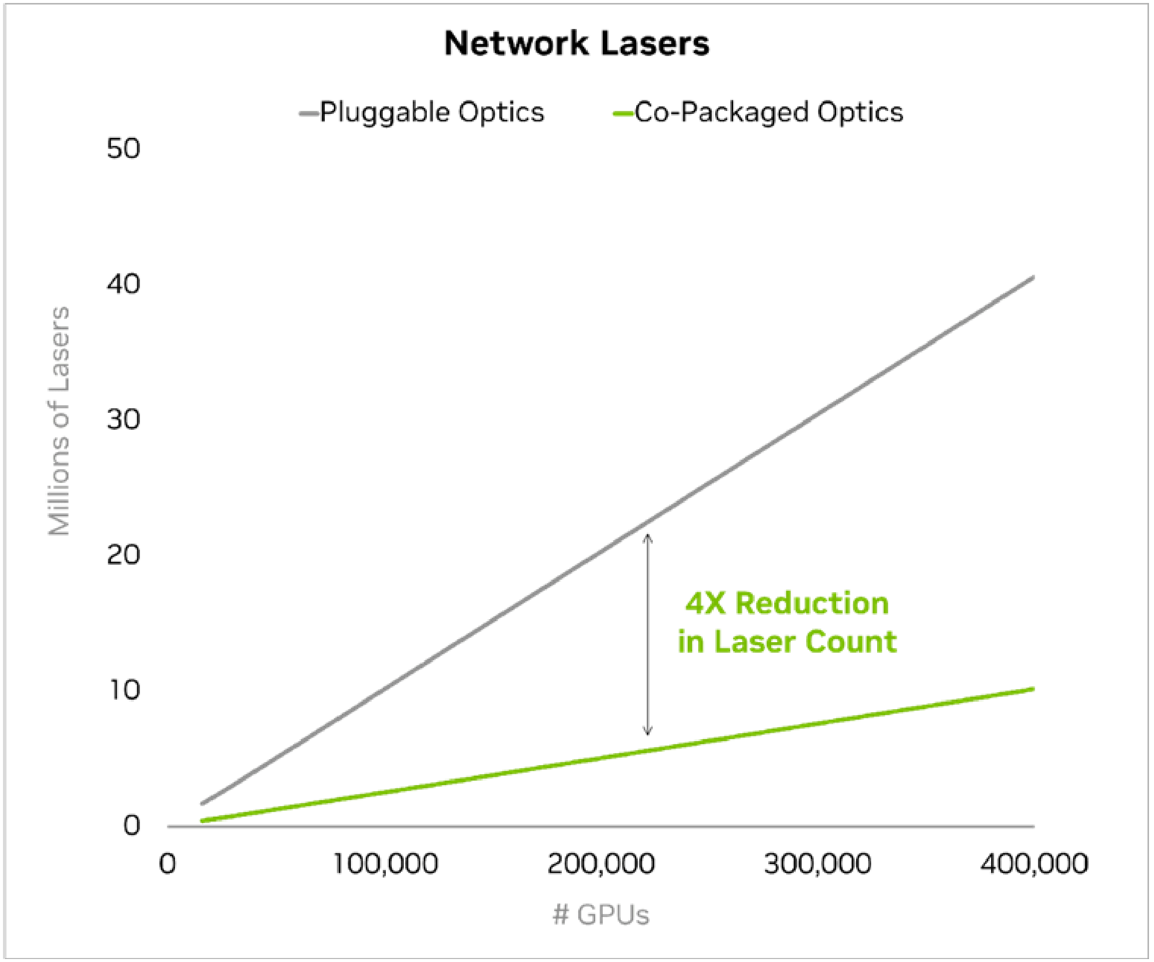
InfiniBand
Integrated Silicon Photonics

TEF 2025 Ethernet for AI

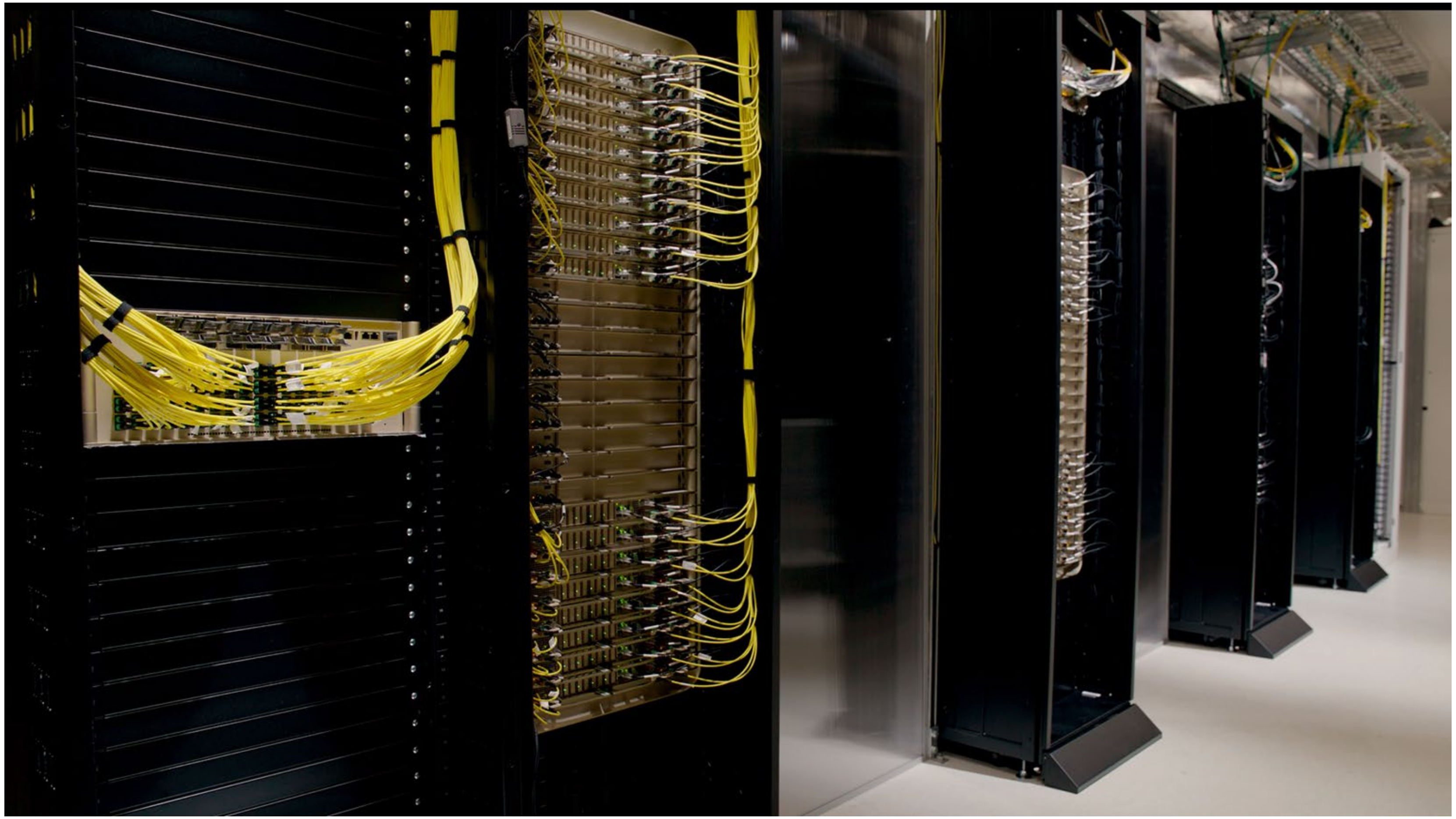

Photonic IC
Electronic IC
3D Stacked Electronic & Photonic ICs
COUPE uLens with surface coupling
Fiber Connector
Optical Sub-Assembly
External Laser Source Module
Laser Source Package
Interposer
Co-Packaged Optics Photonic Switch

CPO Solves Power & Reliability Challenges of AI Scale-out





CPO at NVIDIA Data Center

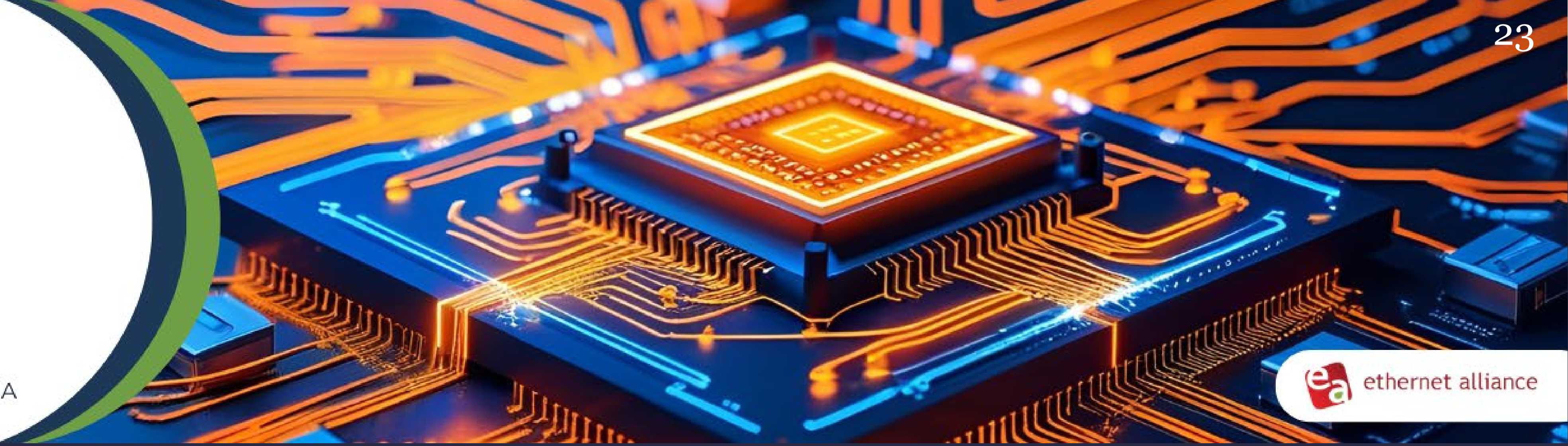


3.5X	10X	5X
Power efficiency	Higher resiliency	Higher Uptime

QUESTIONS?

TEF 2025
Ethernet for AI

December 2-3, 2025
Hyatt Centric Mountain View, CA, USA



Enabling Massive Scale-Out AI Networks with Ethernet and Optical Lane Breakouts

Jose M. Castro
Fiber R&D Manager, Panduit

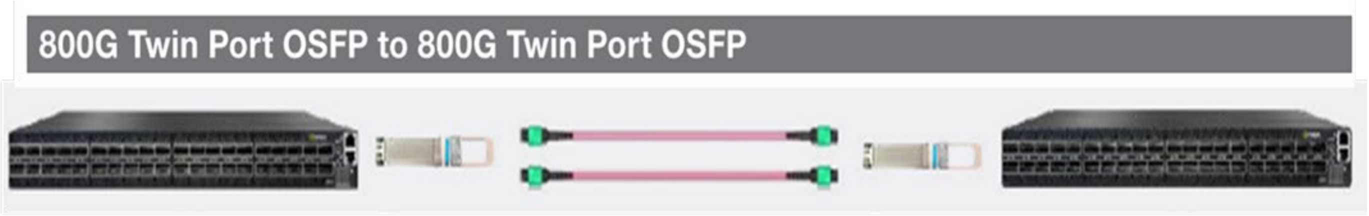
December 2-3, 2025

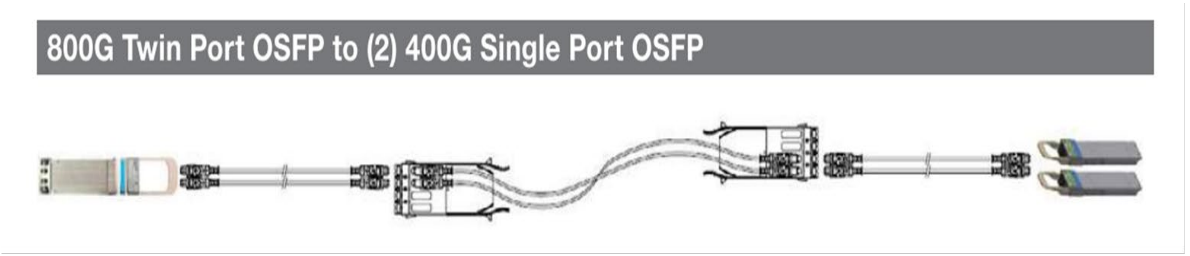
Background

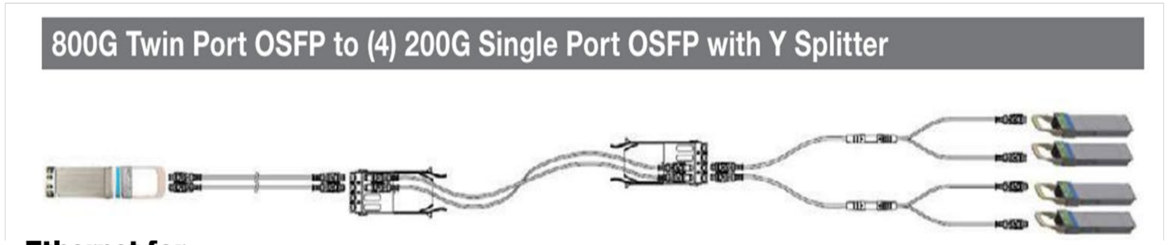
- AI networks with several hundred thousand GPUs are now being deployed, targeting zetascale computing capacity.
- Spraying messages across multiple paths (path/flow diversity) reduces the impact of network congestion or component failures.
  - Example: 1.6 Tb/s → 8×200 G lanes meshing switches from different tiers.
- Flatter network architectures also offer cost advantages while reducing power consumption and communication latency.
- A method known as meshing or shuffling using lane breakouts simplifies deployment.

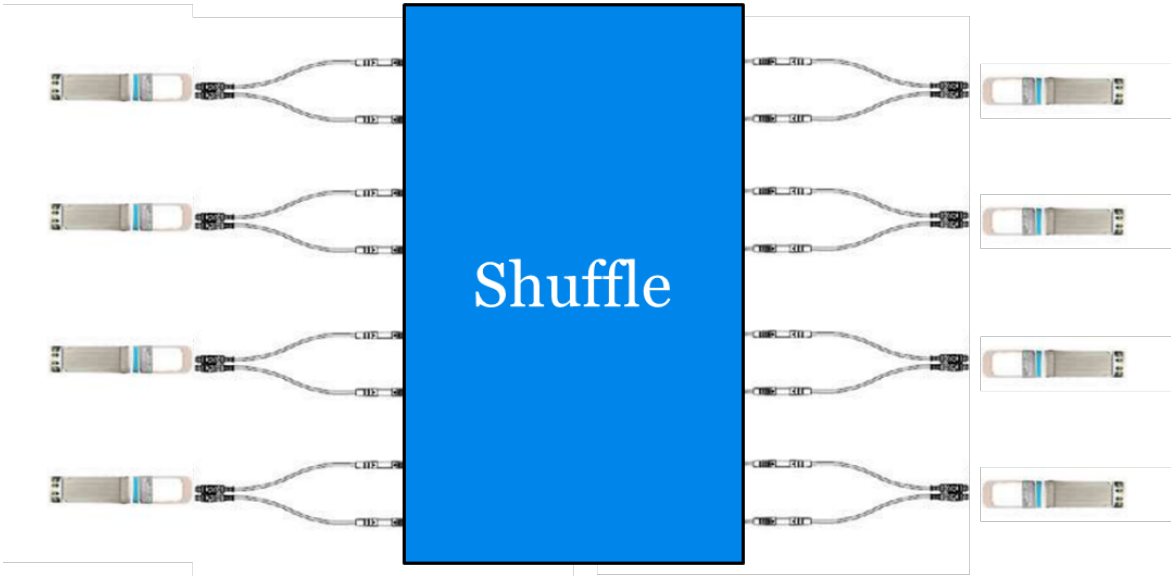
DR Transceivers Optical Lane Breakouts

Without Optical Lane Breakouts









Ethernet for AI

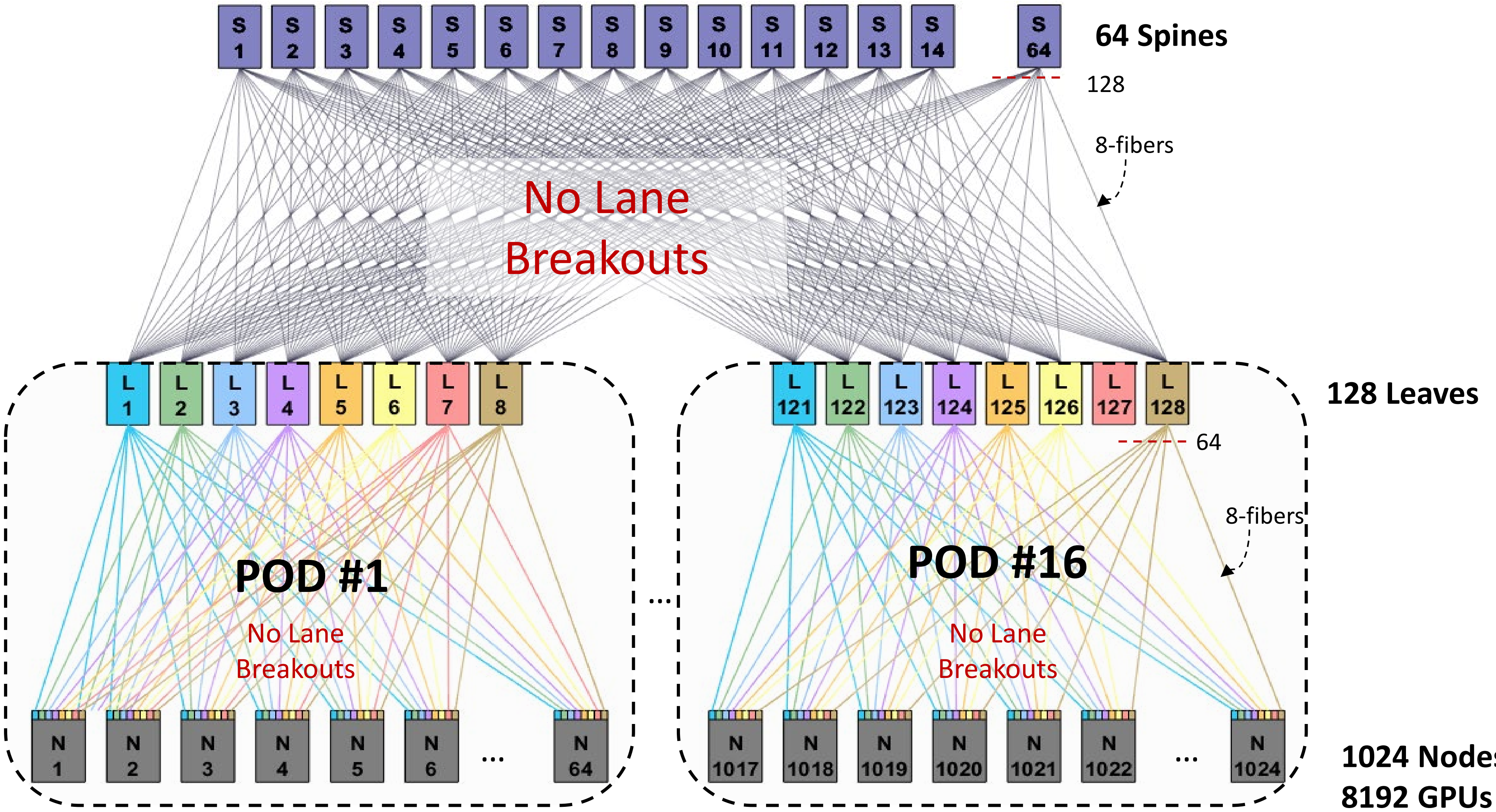
Impact of Lane Breakout on Scaling AI Networks

- Future AI networks utilizing lane breakouts can reduce cost, latency, and power consumption assuming a similar number of GPUs.
  - A 2-tier network requires ~40% less switches, ~50% less transceivers than a 3-tier network
- However, the complexity of deployment could increase.
  - Shuffle modules/harnesses used to reduce deployment complexity.

Switch Radix	Servers/ POD	GPUs/POD	SW Quad Ports (800G)	POD Breakouts	Leaf-Spine Breakouts	Max # Leaves	Max. # Spines	Max # PODs	Max # GPUs	Notes
512	64	512	128	1	1	128	64	16	8192	No Lane Breakouts
512	64	512	128	1	4	512	256	64	32768	Breakouts between Switches
512	256	2048	128	4	4	512	256	64	131072	Full breakouts
1024	64	512	128	1	1	128	64	16	8192	No Lane Breakouts
1024	64	512	128	1	8	1024	512	128	65536	Breakouts between Switches
1024	512	4096	128	8	8	1024	512	128	524288	Full breakouts

AI Cluster with 8192 GPUs

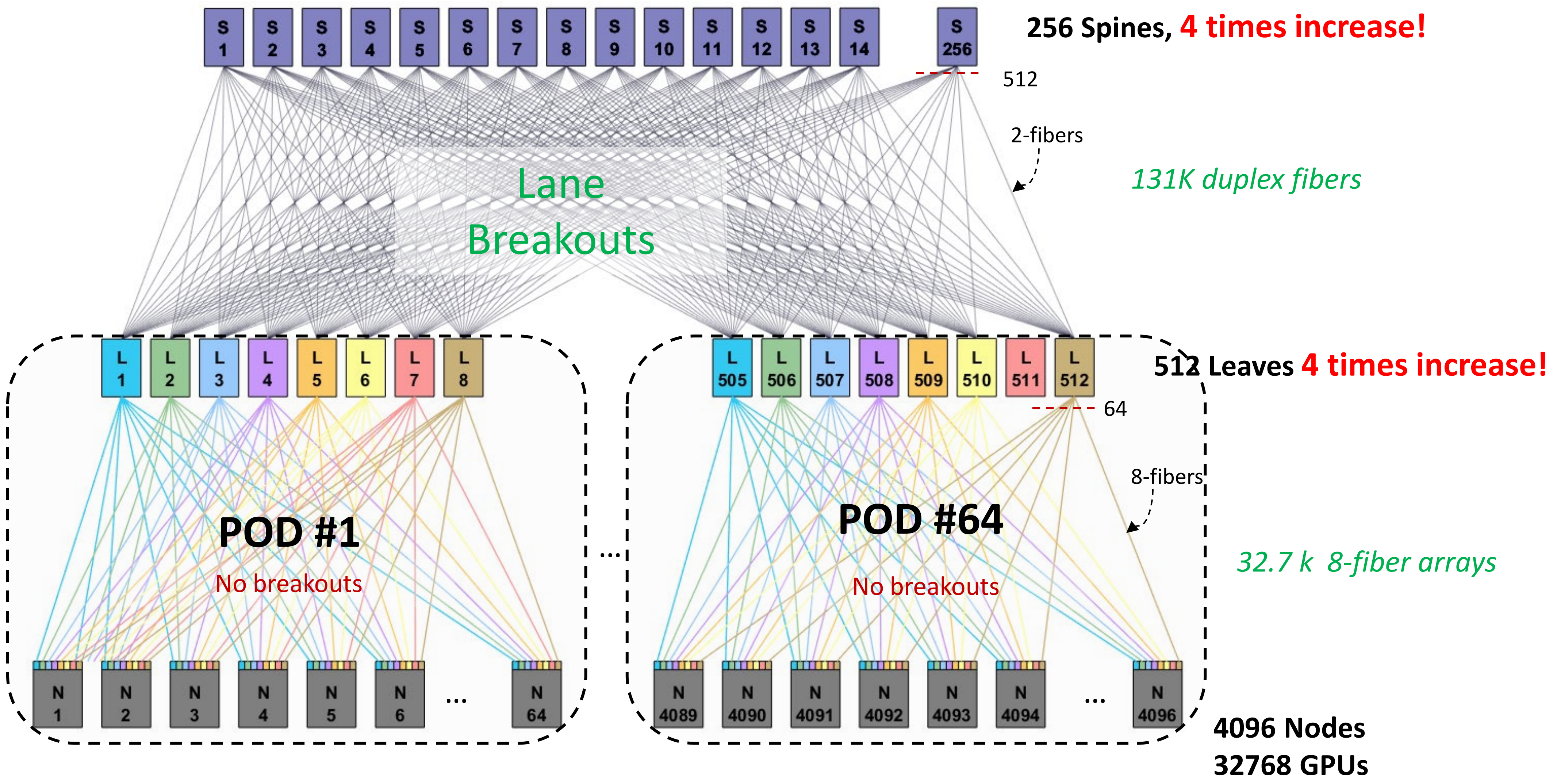
N: Node with 8 GPUs (4 OSFP ports)
L: Leaf , S: Spine
All switches have a radix of 512



TEF 2025 Ethernet for AI

AI Cluster with 32768 GPUs

N: Node with 8 GPUs
L: Leaf , S: Spine
All switches have a radix of 512

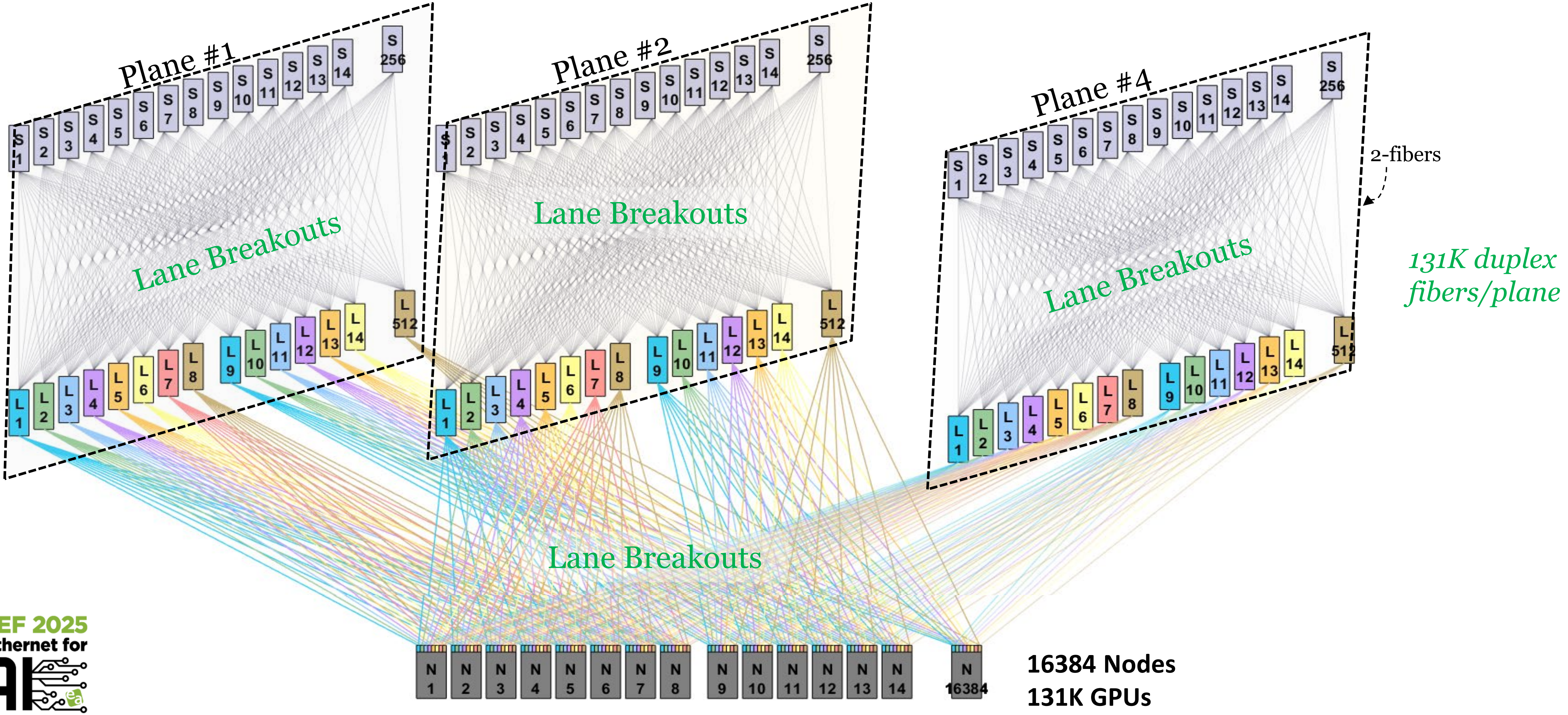


TEF 2025
Ethernet for AI

AI Cluster with 131 k GPUs

Adding node-to-leaf lane breakouts increases the number of GPUs by 4×

N: Node with 8 GPUs
L: Leaf , S: Spine
All switches have a radix of 512

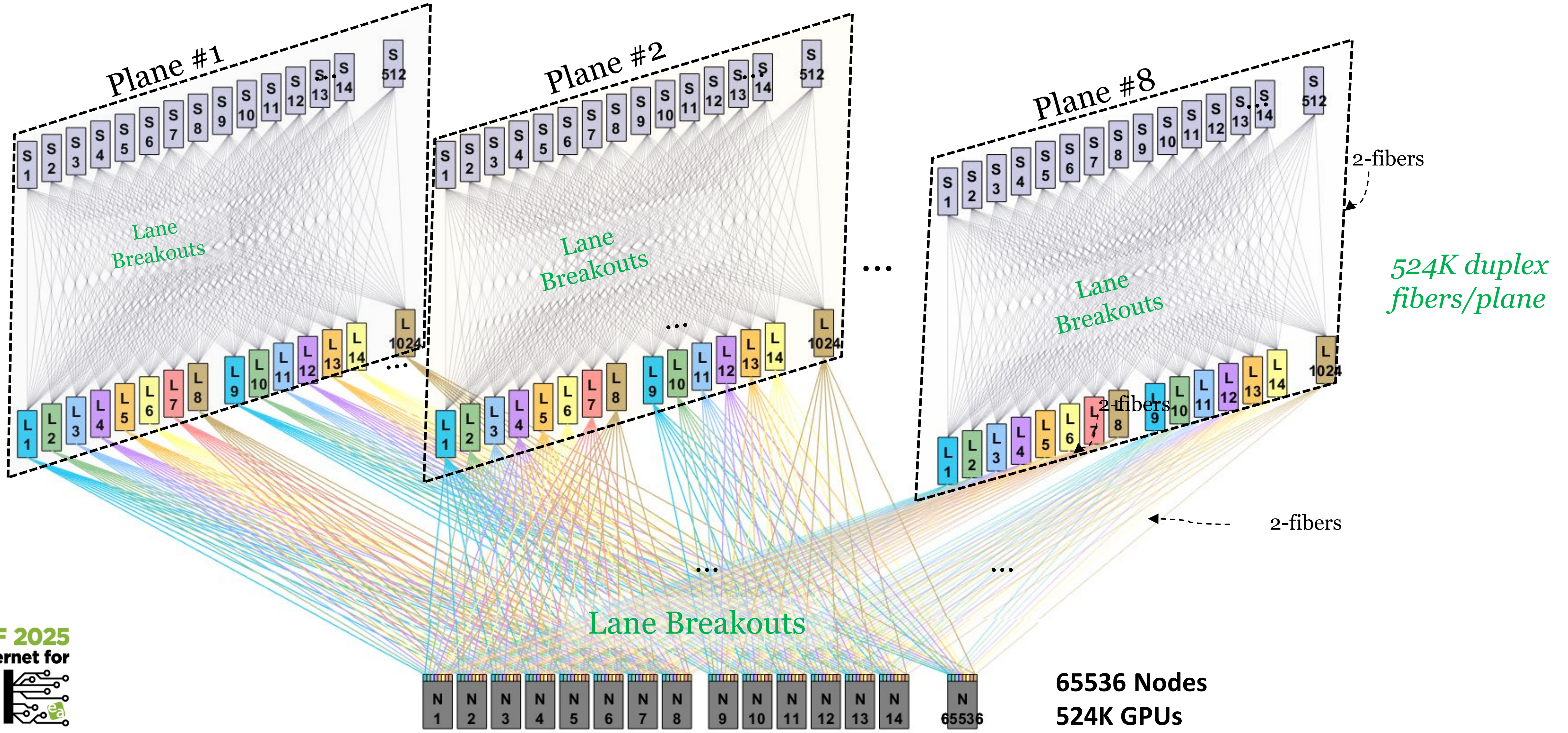


TEF 2025 Ethernet for AI

AI Cluster with 524 k GPUs

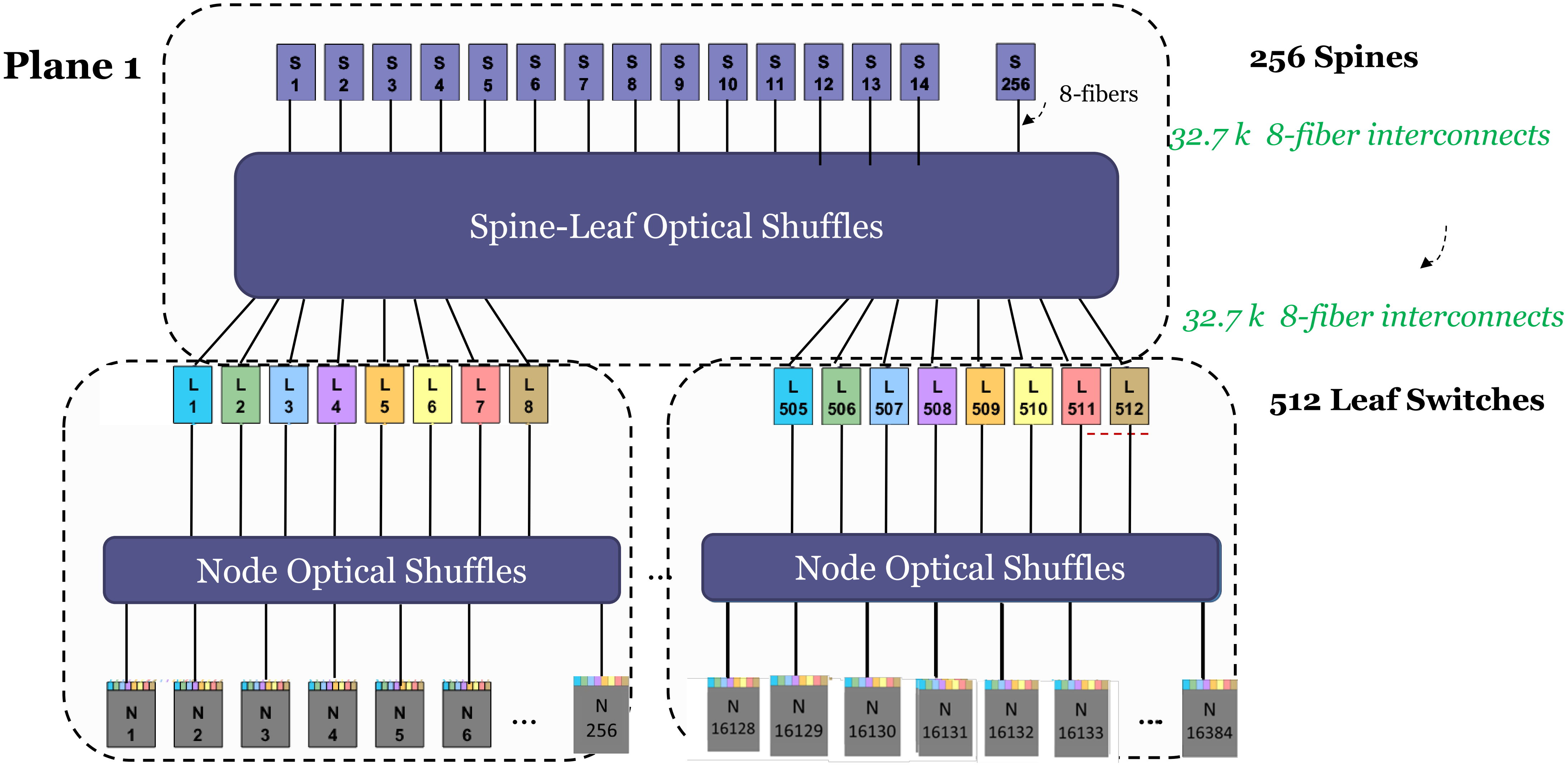
Adding node-to-leaf lane breakouts increases the number of GPUs by 8×

N: Node with 8 GPUs
L: Leaf , S: Spine
All switches have a radix of 1024

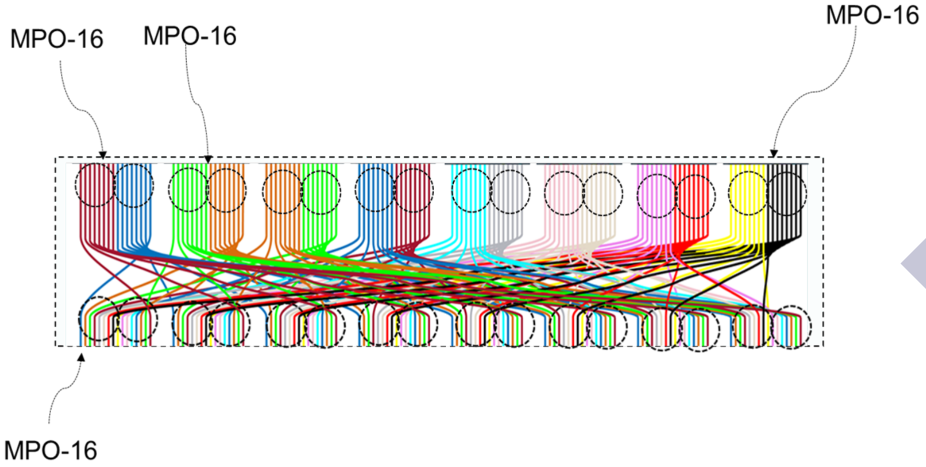


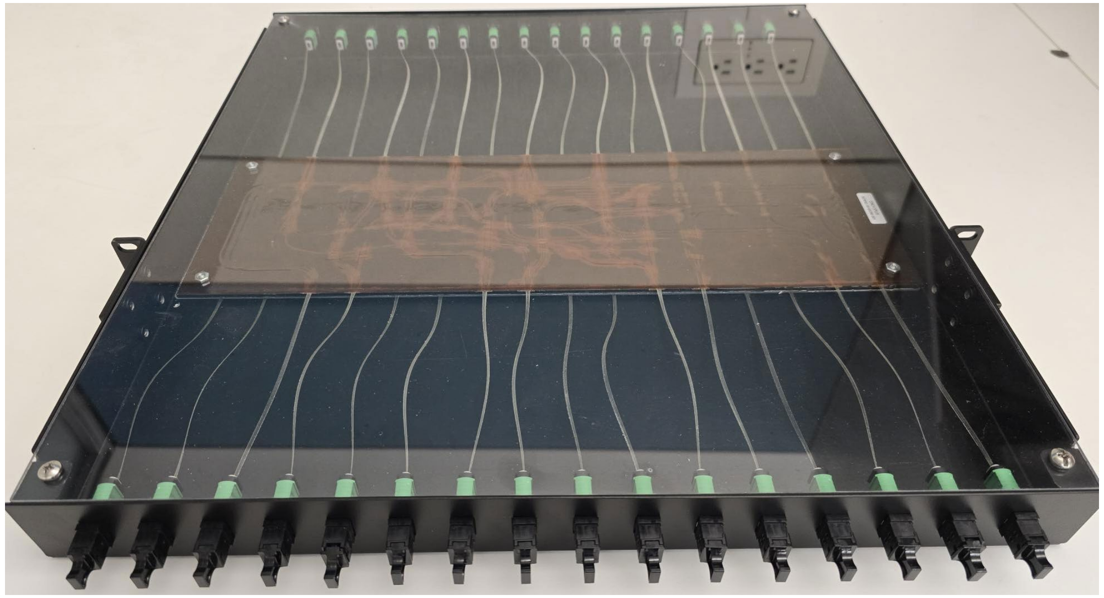
65536 Nodes
524K GPUs

Shuffle Modules Reduce Deployment Complexity



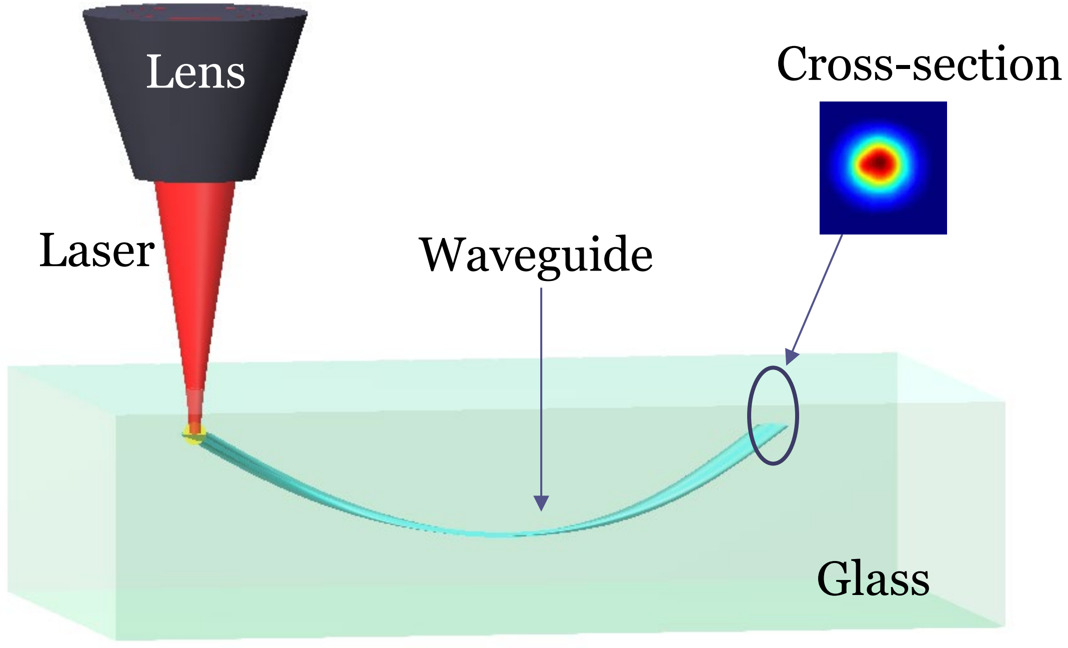
Shuffle Solutions/Technology: Optical Flex Circuits



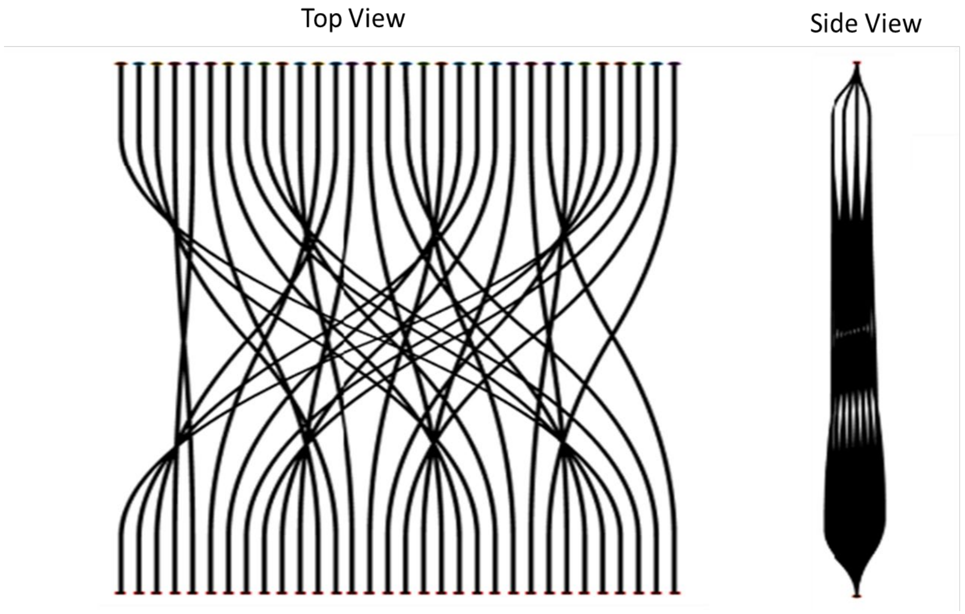


Shuffle Technology: 3D Fabric Embedded in Glass

- Directly written 3D waveguides on glass minimize crossovers, crosstalk, and fabrication time by eliminating masking and exposure.
  - However, state-of-the-art technology still produces higher loss than flex-based shuffle modules.

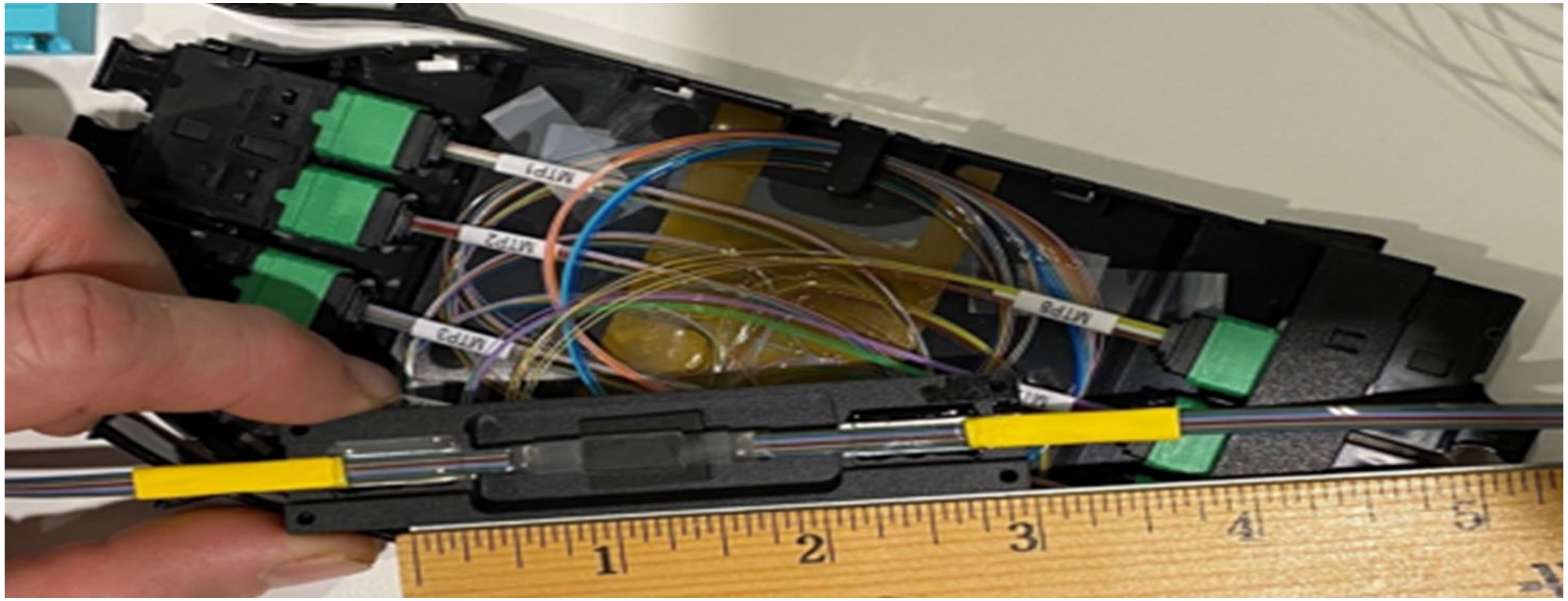


Fabrication Technology



3D Shuffle Design

Shuffle Technology: 3D Fabric embedded in Glass



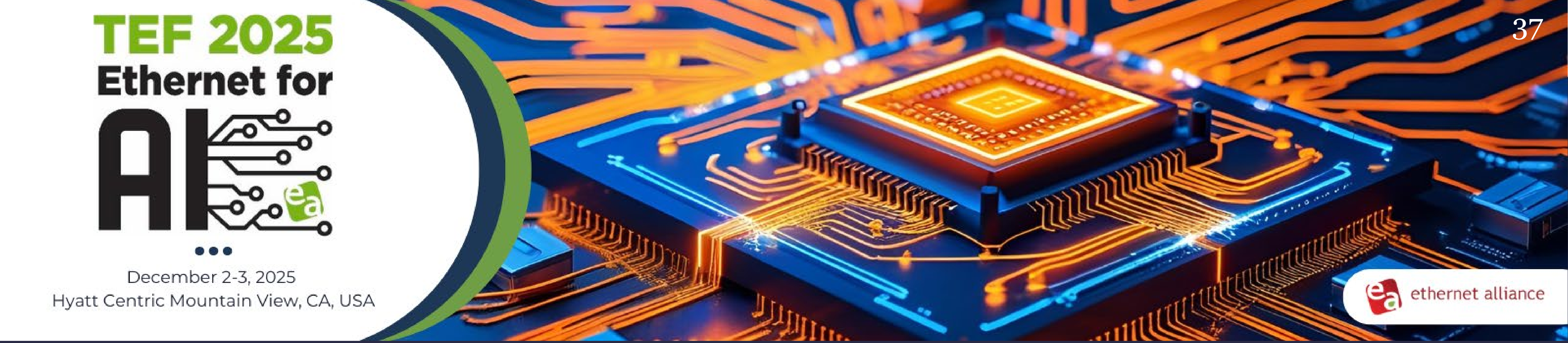
-Photonic Chip with Embedded Data Center Fabric," J. Castro et. Al., **OFC 2025.**

TEF 2025
Ethernet for
AI

Summary and Discussion

- Optical lane breakouts enable massive scale of AI systems.
  - 2-layer fabric supporting 100s of thousands of GPUs.
  - 3-layer fabric supporting several millions of GPUs.
- Reduce latency and provide path diversity
  - Distribute the GPU messages across many switches (packet spraying).
- Reduce networking cost and power consumption
  - More GPUs for a given allocated power.
  - 40% fewer switches.
- Shuffles modules or harnesses simplify the network deployment
  - Enable direct well-organized connections from node to switch ports.

QUESTIONS?



Ethernet for AI:

448G technology for next gen networking

Naim Ben-Hamida
Sr Director Analog Design
Ciena

December 2-3, 2025

ciena

Motivation

- Urgent Need for 448G per lane for scale up and scale out
- Time frame is earlier than expected: 2027-2028
- Reasons
  - ➢ Shoreline density and scalability
  - ➢ Power and cost reduction
- PAM4 vs PAM6
  - ➢ Same cardinality for both electrical and optical: No gear-shifting
  - ➢ Gear-shifting excludes CPO, NPO, LPO and LRO
  - ➢ PAM4 is needed for the optical side regardless, why not use for both sides?

Outline

AI scaling: Need for speed

- Acceleration of the 448G definition
- Acceleration agents: Drivers and applications
- PAM6 vs PAM4: retimed vs non retimed

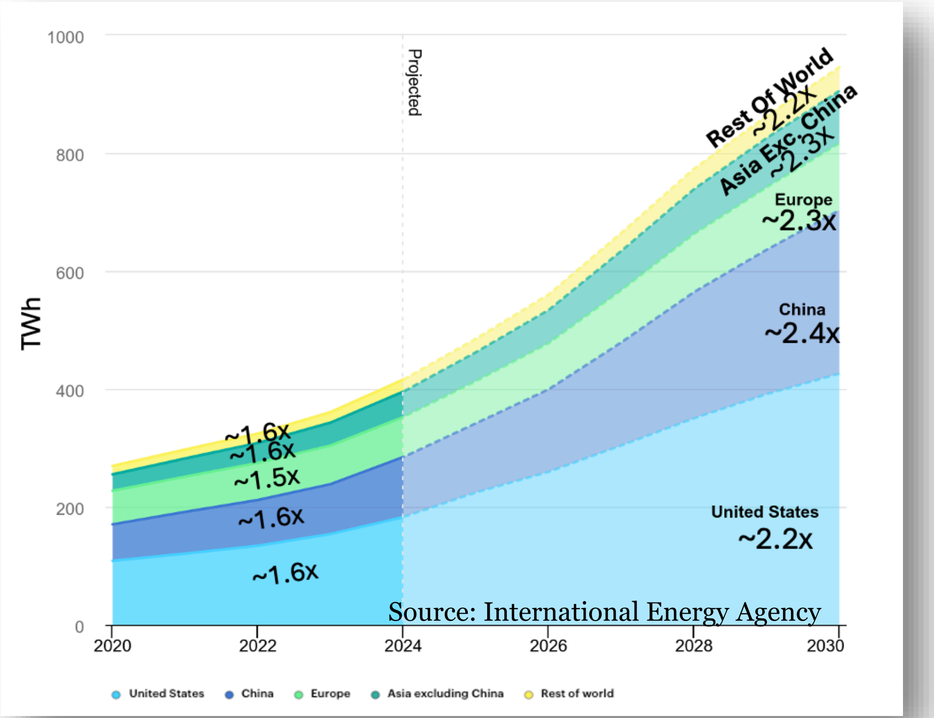
448G Ecosystem for AI scaling

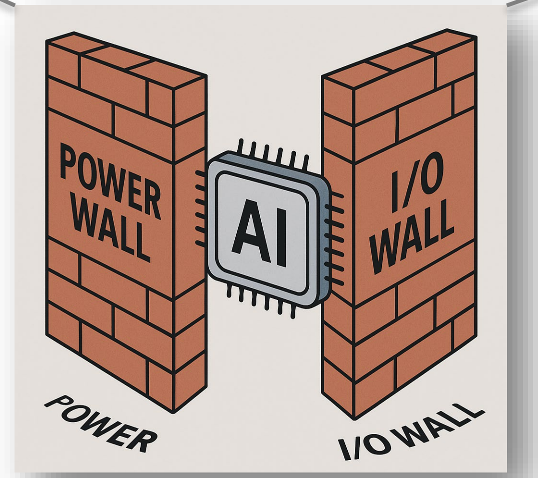
- Silicon Challenges and Readiness
- Optical Engine Challenges and Readiness
- Connector Challenges and Readiness
- Standards bodies

Summary & Key Takeaways

Scaling Strategies for AI Clusters

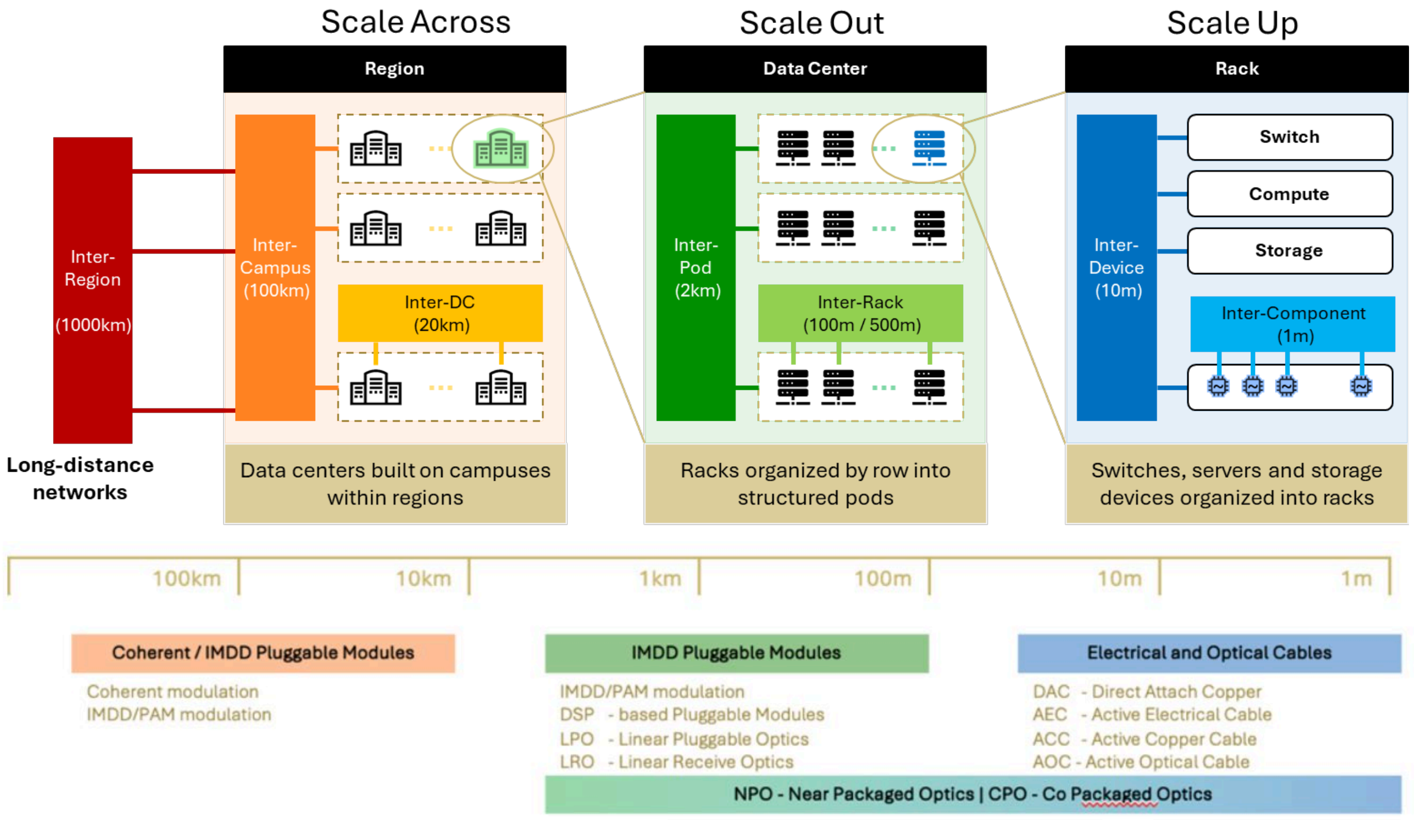
- AI/ML cluster driving the exponential growth in data center traffic and bandwidth
- Scale-Up: 224G ->448G per lane SerDes, higher symbol rates, better packaging and connectors, higher shoreline density
- Scale-Out: 224G -> 448G per lane optical; Add more parallel lanes (8, 16, 32, 64 lanes per module)
- Scale-Across: Build large Coherent fabrics with predictable latency and fault-tolerance





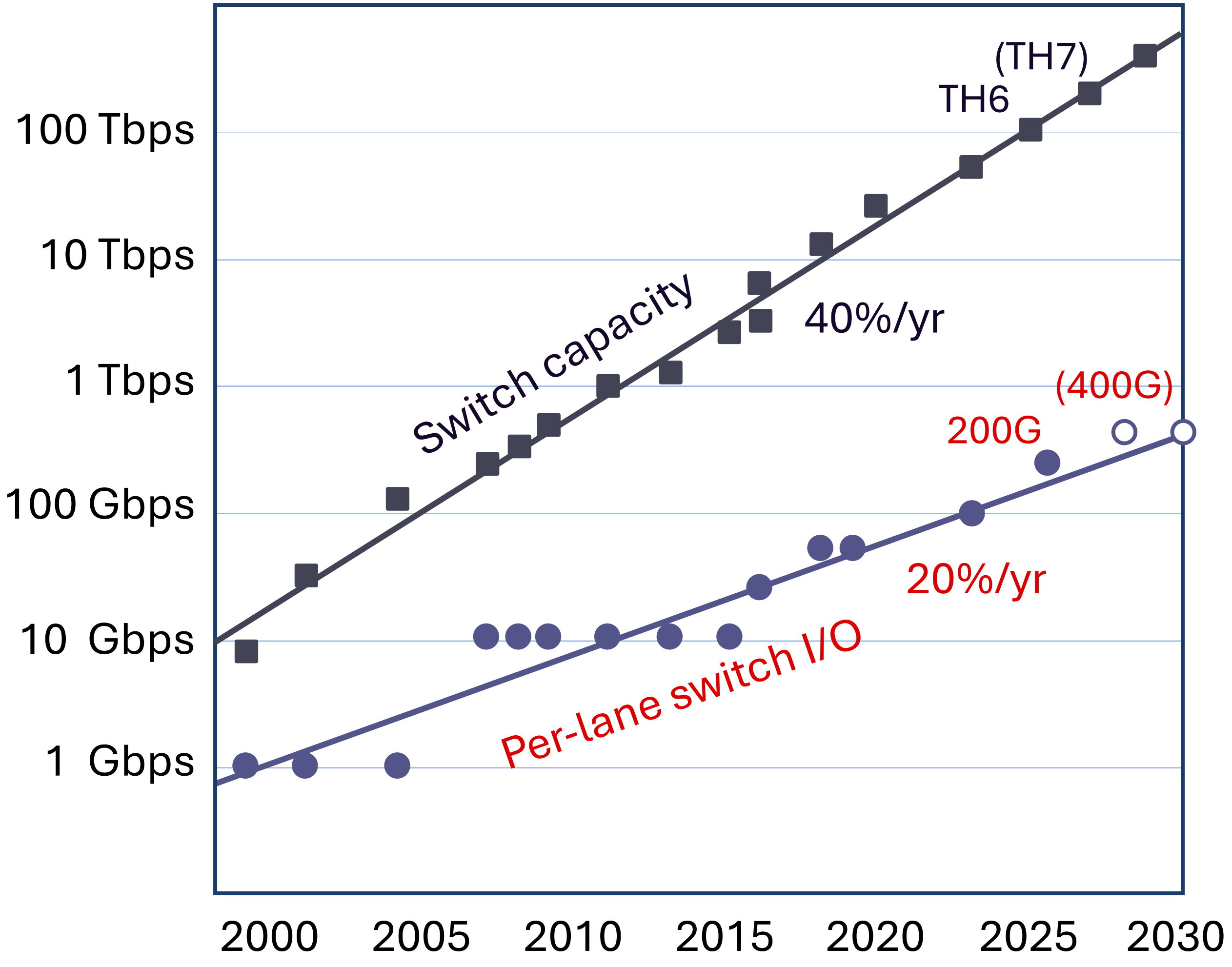
Copper or Light? Choosing the Right Highway for AI

- Copper: cheap, low power, short reach
    - Extremely challenging at 200G, active technology required for most applications
- Optical: long reach, high bandwidth, higher cost/power
    - Many different architectural approaches with CPO, NPO, LPO,LRO etc.
- AI workloads are pushing copper beyond limits. Gradual transition to optical for ultra-high speeds and reach
- **We have already entered the terabit and petabit era**



Ongoing evolution of technology in and around the data center, leveraging adjacency

High speed SerDes: currency of the Data Center



Mixed-media is core requirement
"Copper when you can, optics when you must"

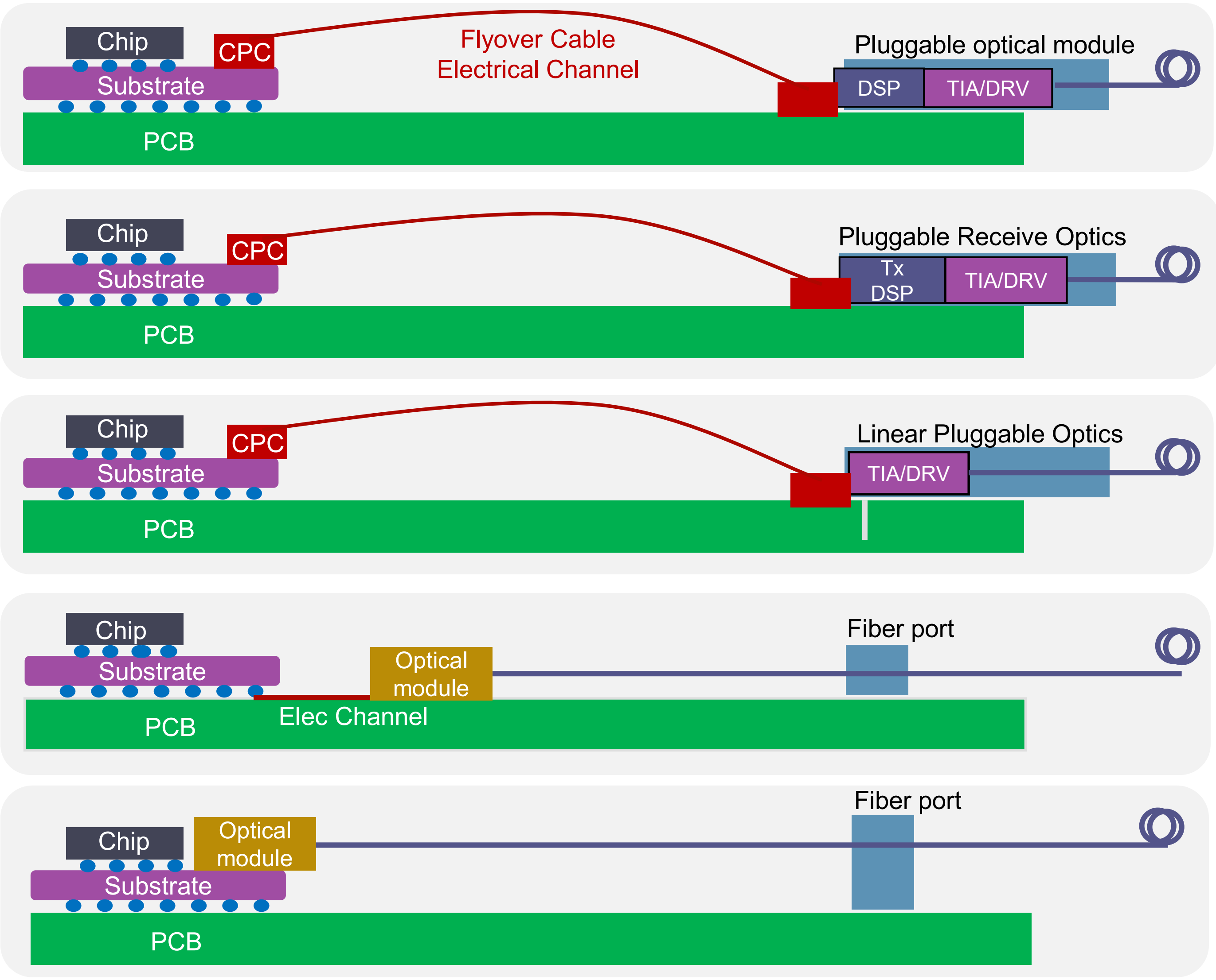
Requires high-speed SerDes I/O

Paradigm extends to 400G

➔ Low power, high density optics must interface with high speed SerDes I/O

SerDes show no signs of slowing down

Optical scale out: Retimed vs non retimed



PAM6-PAM4

Fully Retimed (DSP)
Power: 18 pJ/bit

PAM4-PAM4
PAM6-PAM6

Half Retimed (LRO)
Power: 12 pJ/bit)

PAM4-PAM4
PAM6-PAM6

Linear Pluggable (LPO)
Power: 7 pJ/bit

PAM4-PAM4
PAM6-PAM6

Near Packaged Optics (NPO)
Power: 5 pJ/bit

PAM4-PAM4
PAM6-PAM6

Co-Packaged Optics (CPO)
Power: 3 pJ/bit

30-60% power saving but shorter reach

70-85% power saving but shorter reach

Highlights - 448G Ecosystem Progress

Aug 2024
First 1.6T coherent link
(224 Gbaud)
in live network
WaveLogic 6 Extreme

Oct 2024 – TEF 1
First 448G networking industry conference

April 2025 - OFC
Multiple 448G vendor demos:
- DSPs
- EMLs

April 2024 –
OIF 448G Workshop
Second 448G networking industry conference

Sold Out!

Oct 2025 – OCP
- 448G CPC
- Live demo over 500m link

Dec 2025 – TEF

Much, much more going on!

2024

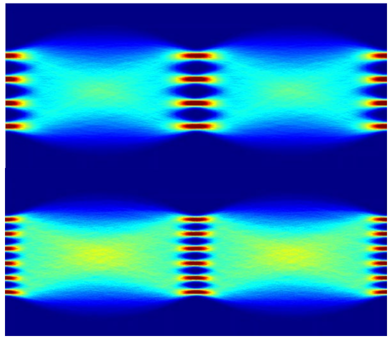
2025

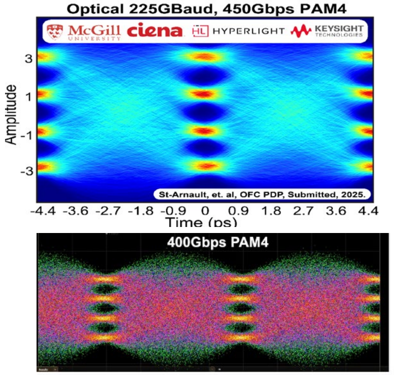
Oct 2024 - OCP
First 448G PAM4 Demonstration

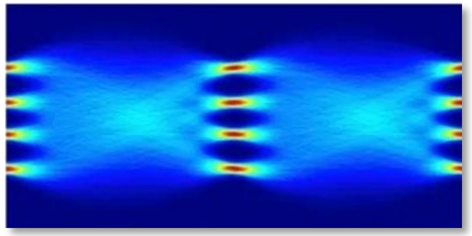
April 2025 – OFC
Post deadline paper
3.2T – 8 x 448G over 2km TFLN

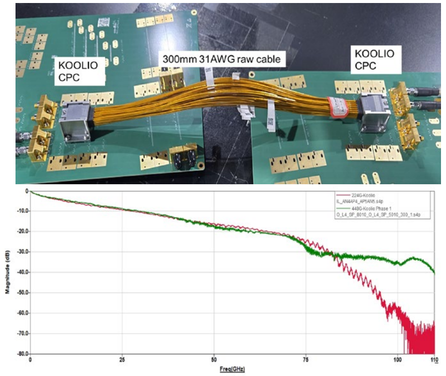
Sep 2025 – ECOC
Low Vpi TFLN
448G with 750mV drive

Nov 2025 – OIF
100GHz connector







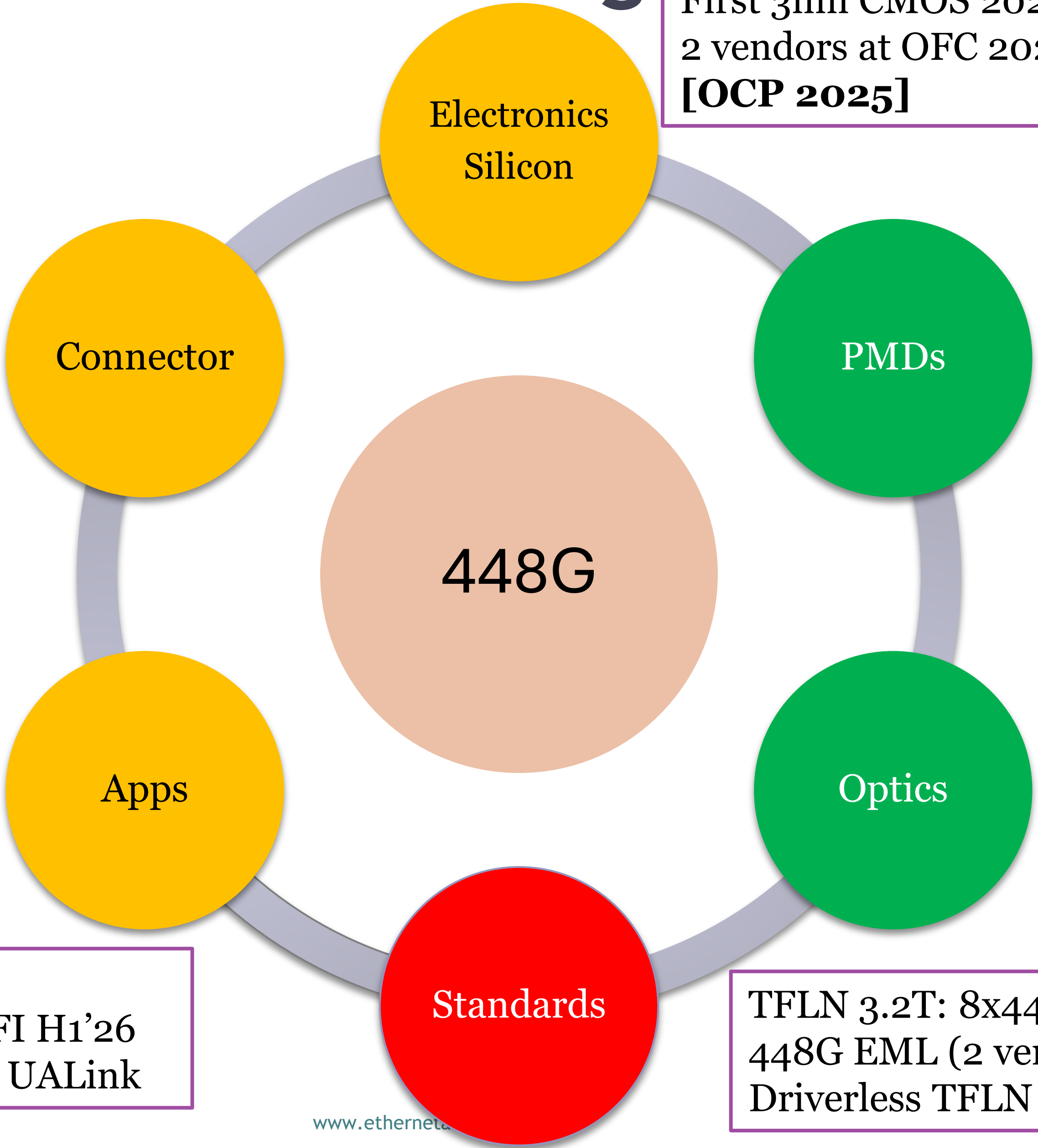


TEF 2025
Ethernet for AI

448G ecosystem for AI scaling

First 3nm CMOS 2024 OCP
2 vendors at OFC 2025
[OCP 2025]

100GHZ Flex interconnect
PAM6 85GHz connectors
2D Connectors



PDs
Driver protos available
TIA protos available

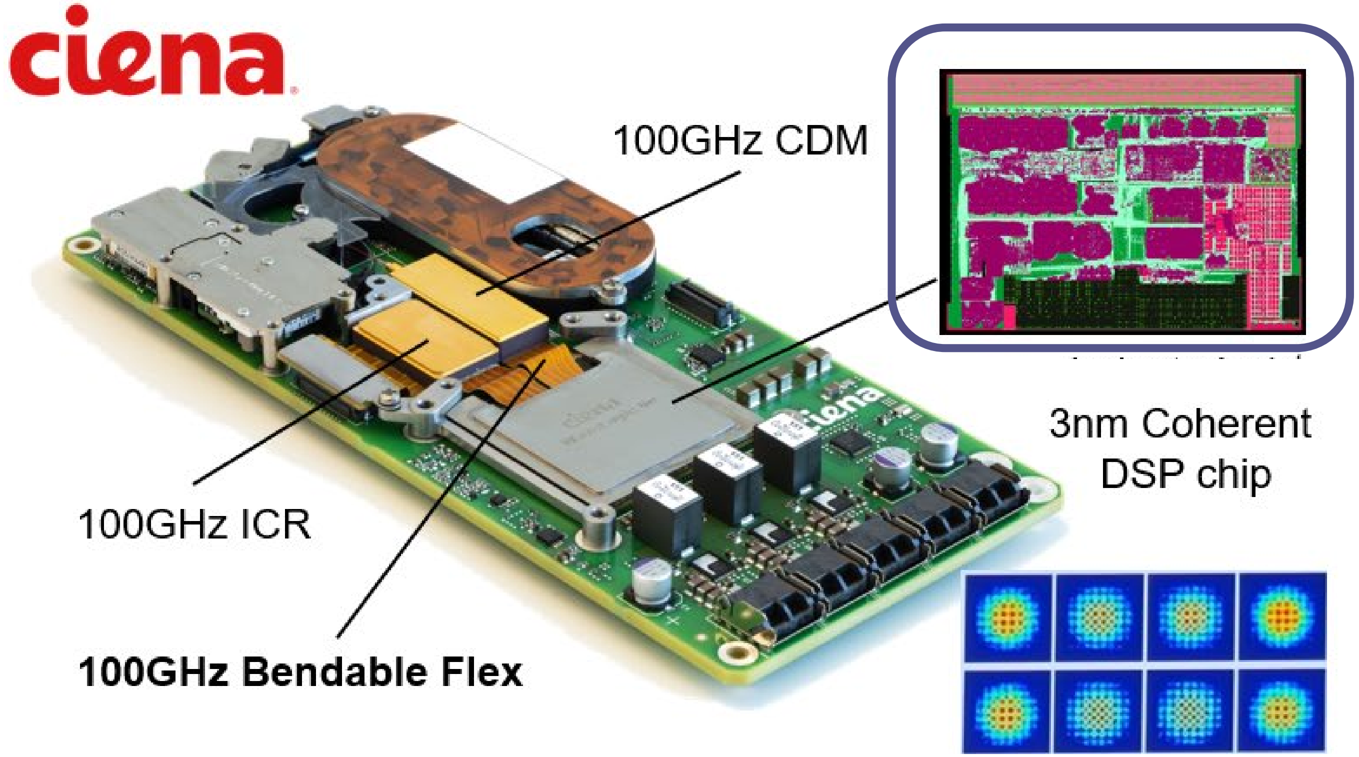
GPU Connectivity: Scale up
Rack connectivity: Scale out
High Density Switches (1Pbps)

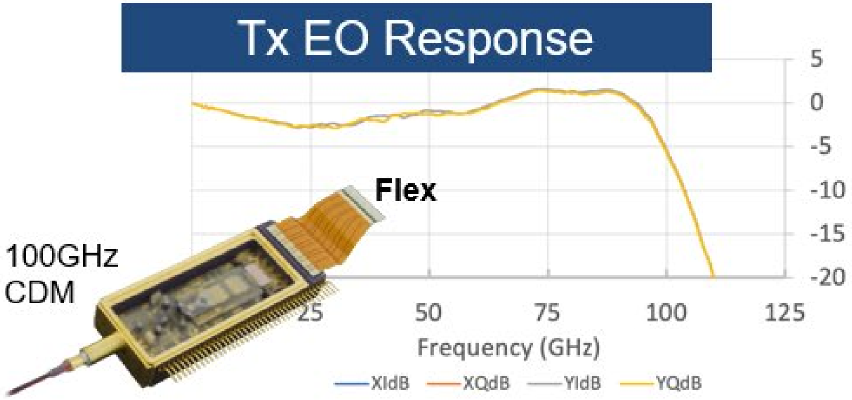
Pre-demo
Demo
Protos

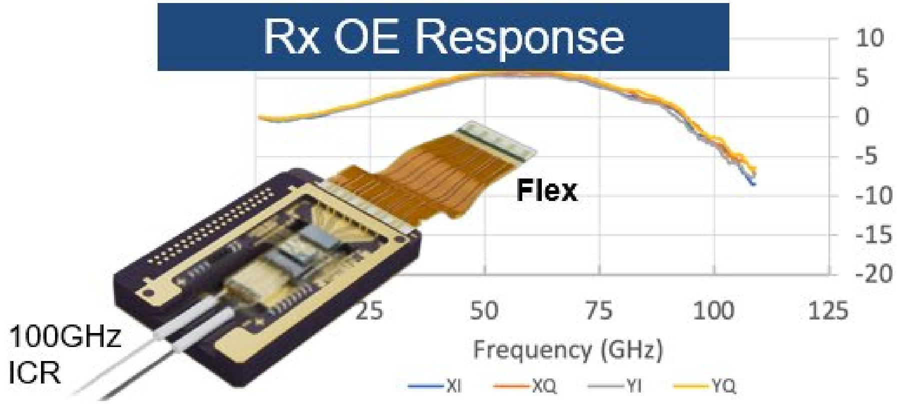
OIF: CEI-448G
IEEE: Building consensus, CFI H1'26
Ultra Ethernet Consortium & UALink

TFLN 3.2T: 8x448G (OFC 25/**OCP 2025**)
448G EML (2 vendors at OFC 25)
Driverless TFLN (ECOC 25/**OCP 2025**)

1.6T DSP chip and Flex interconnect

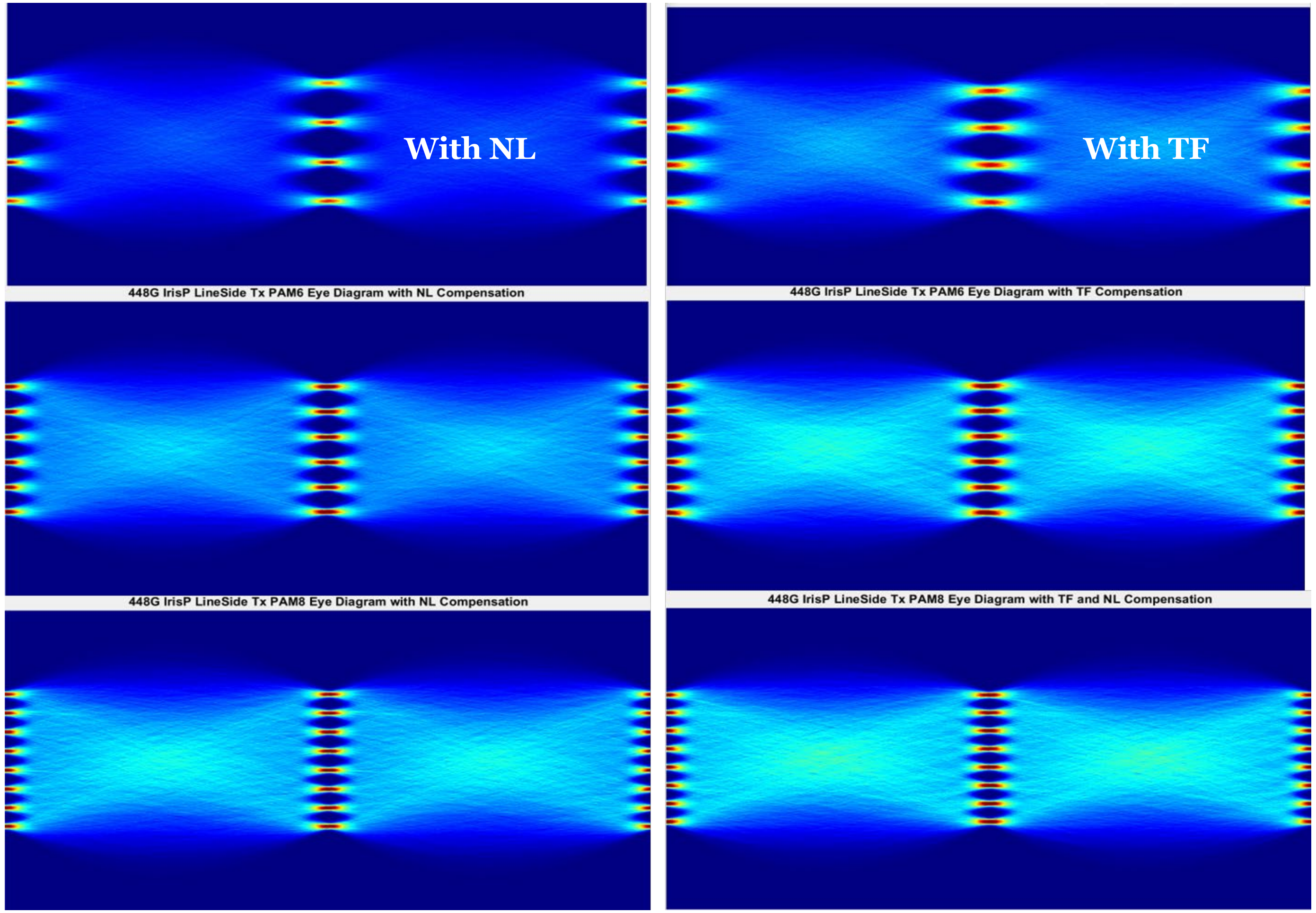






1.6 Tb/s coherent modem: 200GBaud DSP and E/O
Use of NPO interfaces to minimize interconnect losses and crosstalk for 100GHz+ BW

448Gb/s in PAM4, PAM6 and PAM8



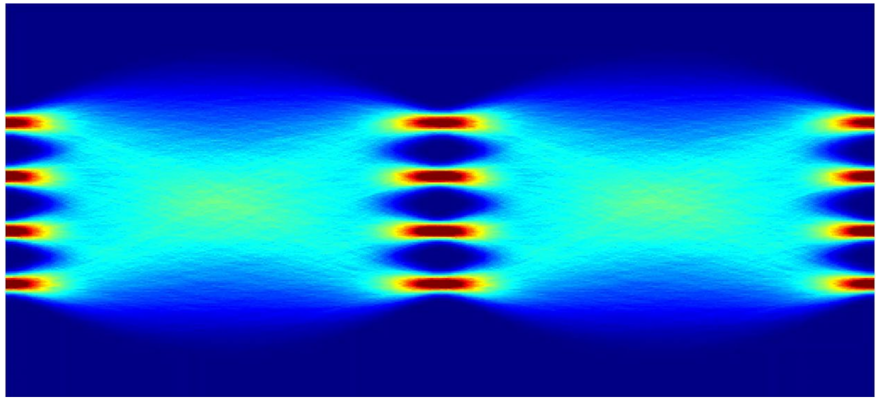
PAM4
Baud rate: 225G
SNDR: 25.6dB

PAM6
Baud rate: 175G
SNDR: 27.9dB

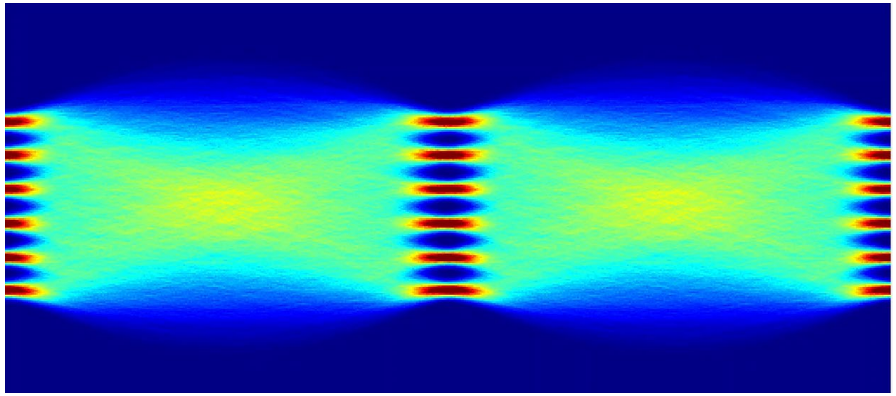
PAM8
Baud rate: 150G
SNDR: 28.5dB

Improvement of SNR does not justify the increase in cardinality

Loopback Results for PAM4 and PAM6



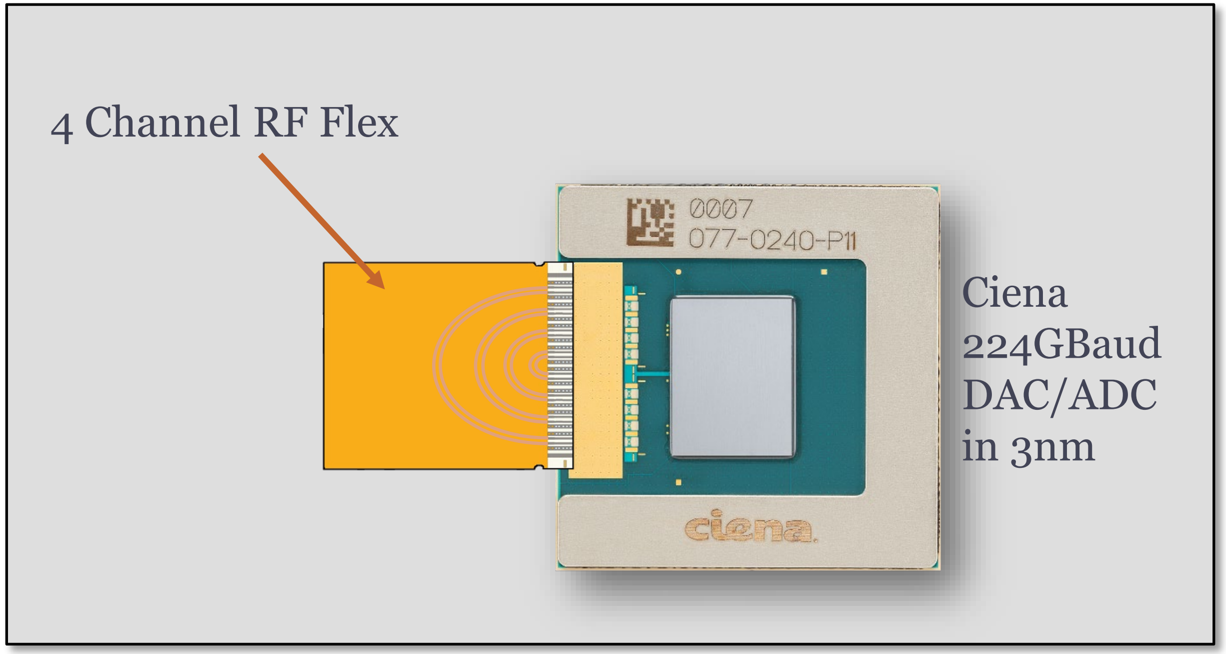
PAM4: 448Gb/s Received eye
224GBaud/112GHz



PAM6: 448Gb/s Received eye
Baud rate: 173.4G/86GHz

448Gb/s PAM4, PAM6, PAM8 generated

Flexibility to generate PAM4, PAM6, PAM8 provides ability to analyze for trade-offs between performance, power, cost.



Improvement of SNR does not justify the increase in cardinality

448G Serdes Design Challenges

Feasibility:

- Demonstrated live 3nm silicon delivering 448G and beyond in PAM4/6/8 at OCP24

Tx Challenges

- Delivering 112GHz Electrical Bandwidth
- Sampling at 224GS/s => 4.4ps Sampling Period
- Extreme Sensitivity to Jitter: <70fs RMS Jitter
- Deliver High Tx SNDR

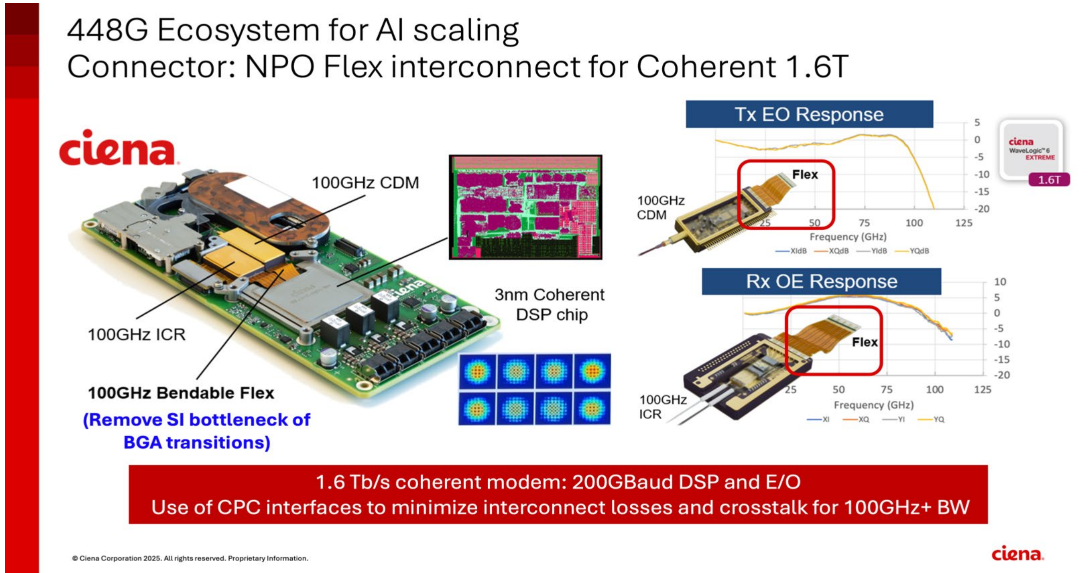
Rx Challenges

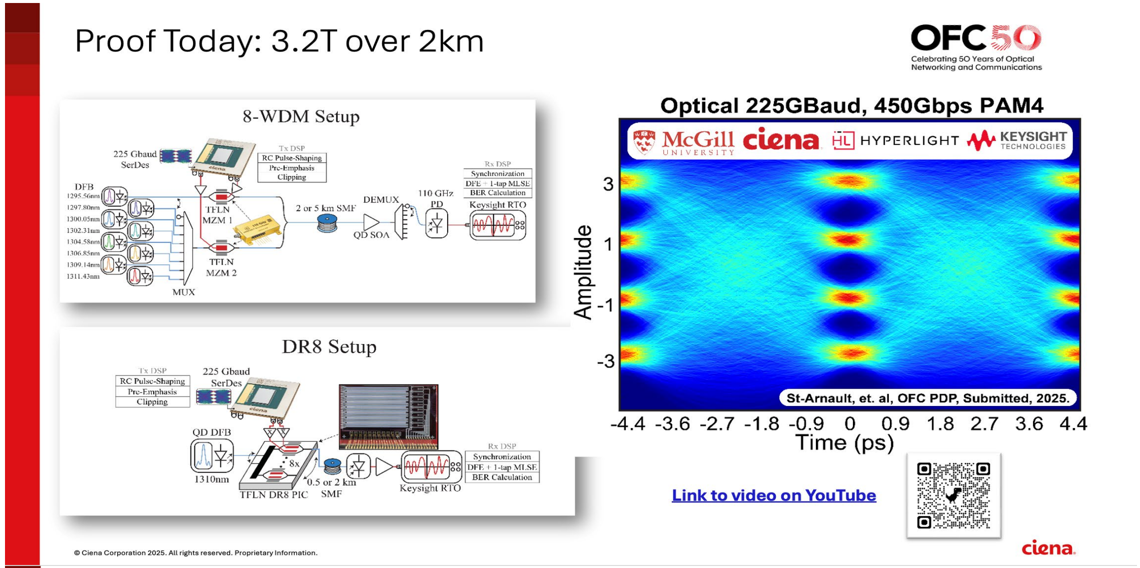
- Analog Front End (CTLE) Delivering Peaking at 112GHz
- Reduce RMS Jitter to < 70fs
- Timing Correction down to 10fs
- Achieving High Bandwidth Clock and Data Recovery (CDR) Loop
- Delivering High Rx SNDR

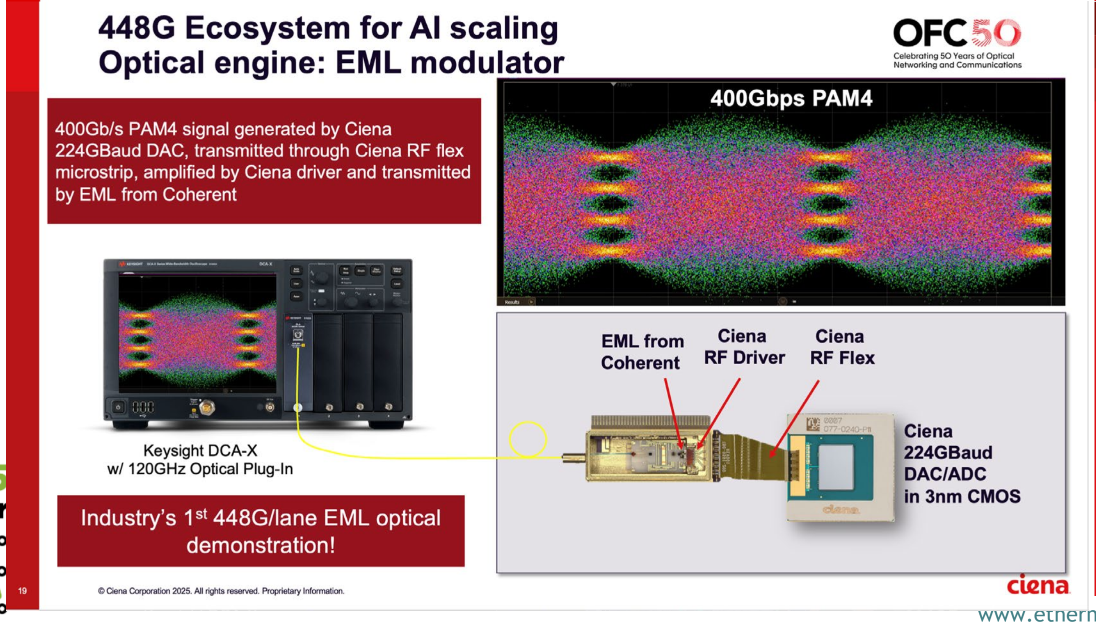
Cardinality: PAM4, PAM6, PAM8

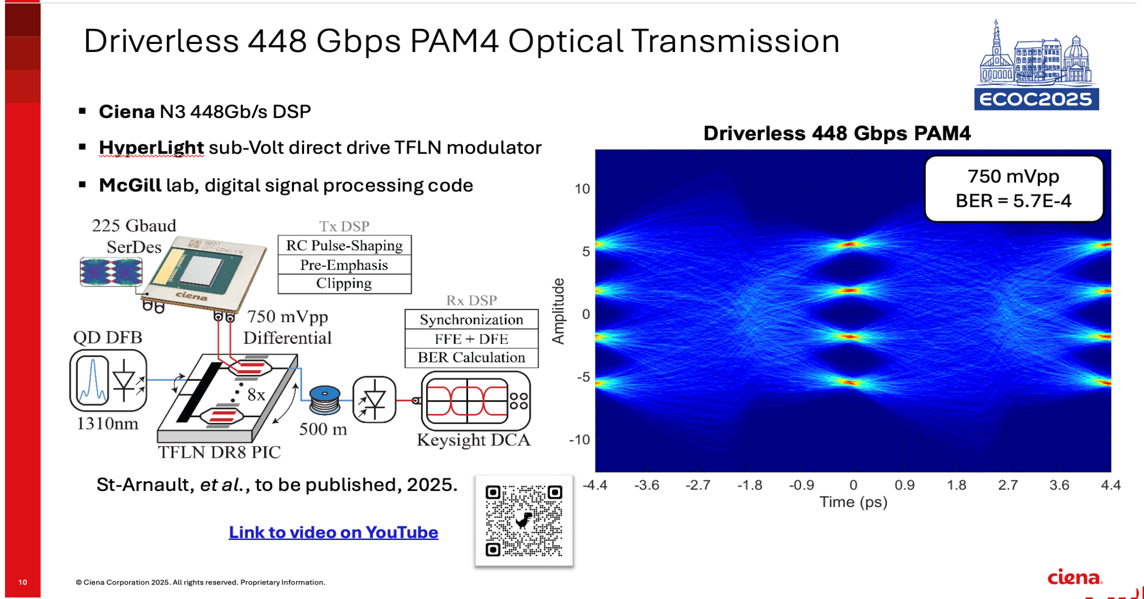
- PAM4: Lower required SNDR, higher bandwidth
- PAM6: High required SNDR, lower bandwidth, stronger FEC

448G with different optical engines







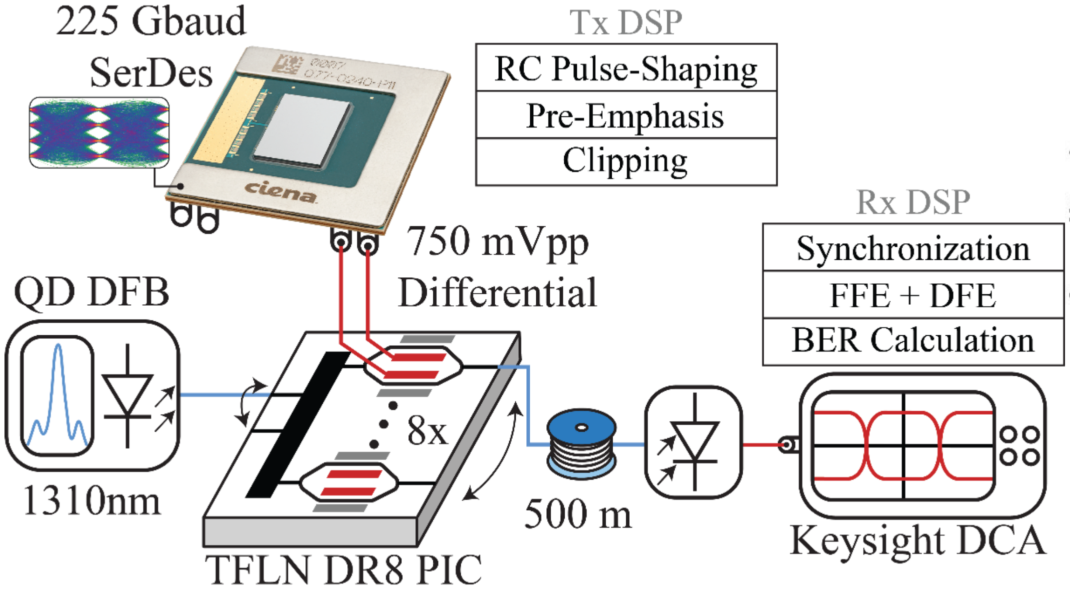


TEF 2025
Ethernet for AI

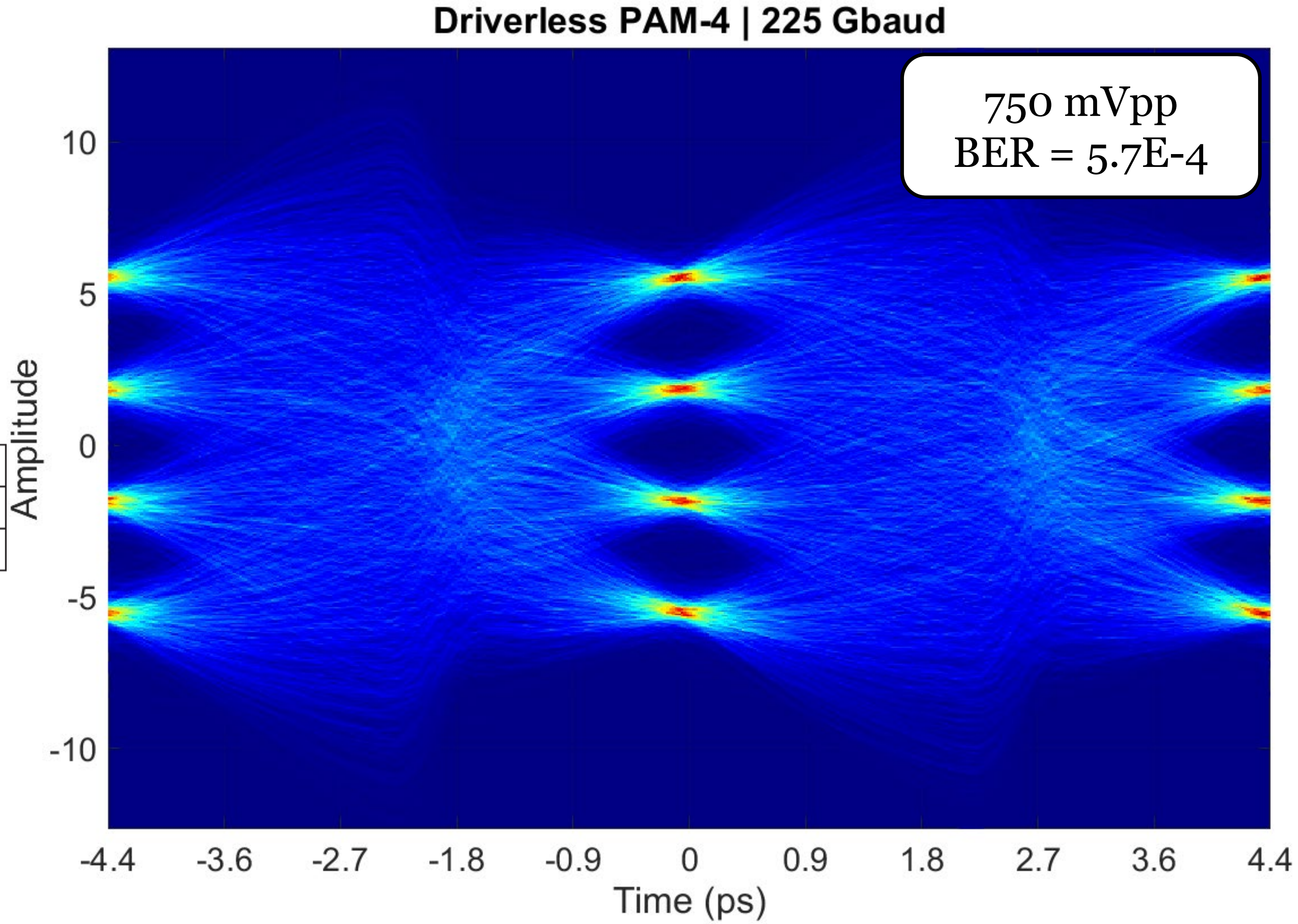
ciena

Driverless 448 Gbps PAM4

- Press-release: ECOC 2025
  - McGill, Ciena, HyperLight



St-Arnault, *et al.*, to be published, 2025.

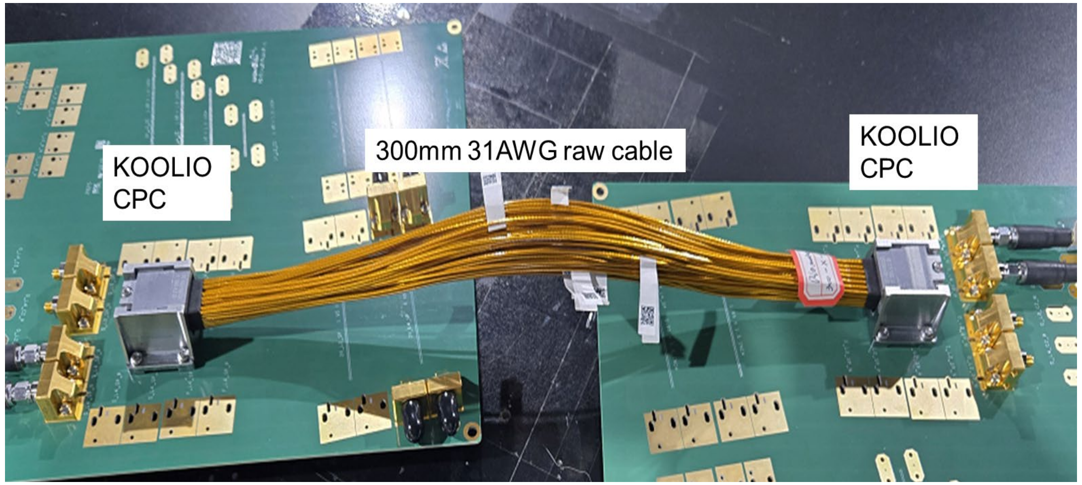


Modulator technology for 448G

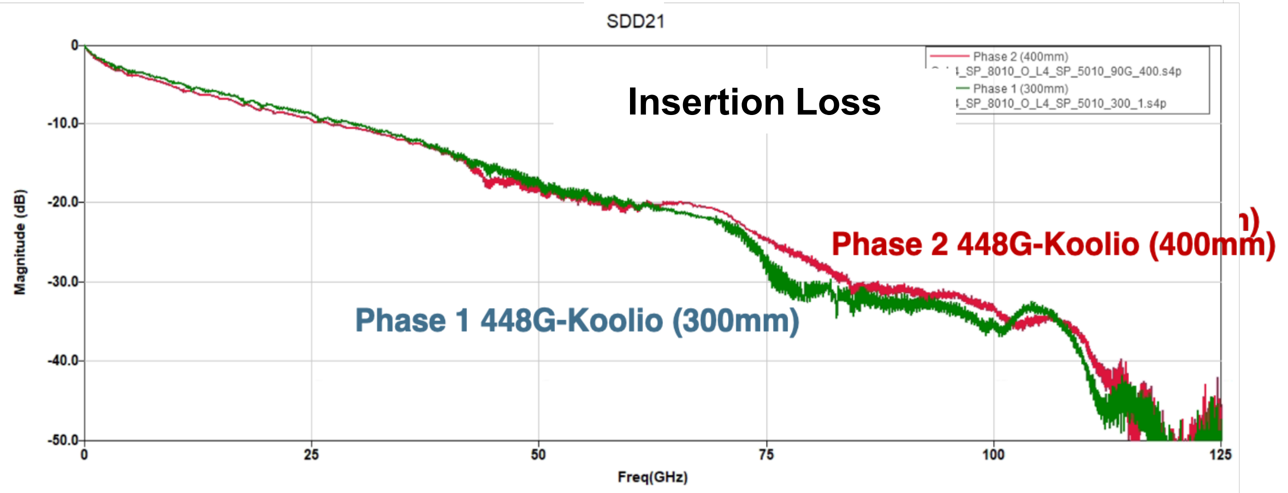
TFLN and EML are well positioned to capture the low density pluggable (8x448G) application
There is no leading contender for high density application (64x448G)

448G-KOOLIO Phase 1 Measurement

Measurement Setup

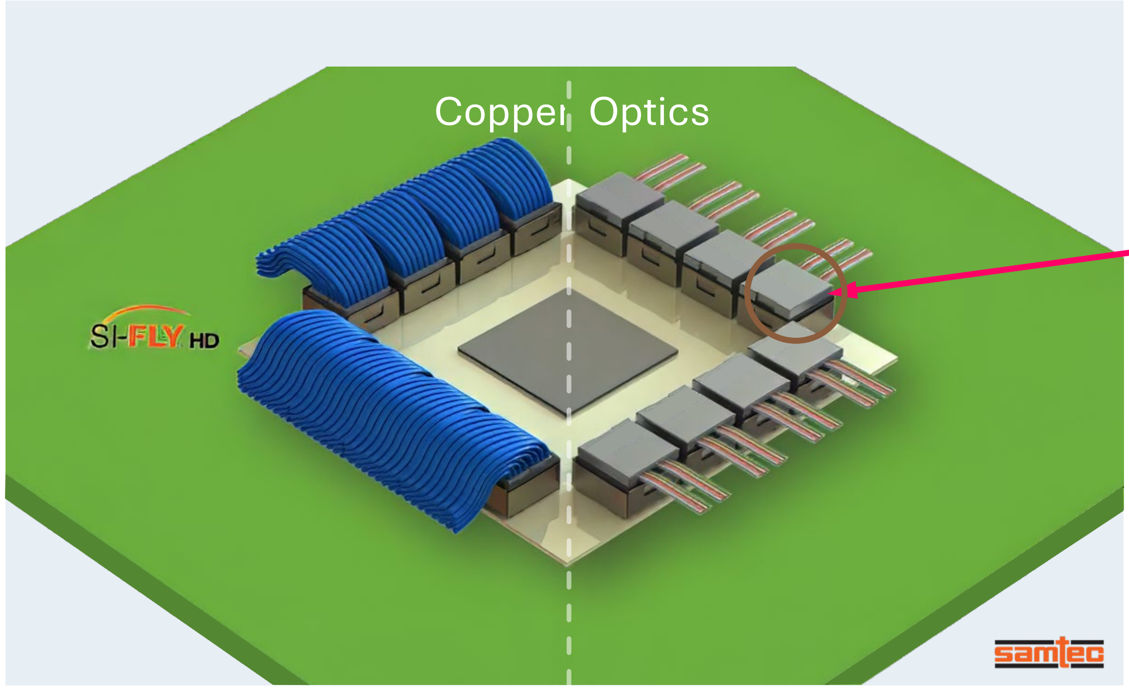


New MCB w/ 1mm conn — KOOLIO CPC — 300mm 31AWG — KOOLIO CPC — New MCB w/ 1mm conn



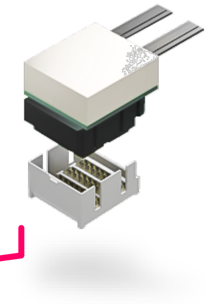
- Today's OSFP Test fixture can marginally support PAM6 at 175Gbaud
- PAM4 is impossible with yesterday's connector but hopeful with connector improvements
- CPC connector can deliver >100GHz BW enabling PAM4 448G: 30dB loss at 100GHz

2D connector: co-packed copper & optics



Source: Samtec

Ciena CPO



- 6.4T removable connector
- Compatible with copper
- Retimer-free linear interface
- 5 Watts per Tbps (5 pJ/b)
- Matched to SerDes I/O
- Multi-vendor

High-density low-power (retimer-free) systems need optics in CPC sockets

Summary

Advanced CMOS nodes can deliver >100GHz BW for 448G PAM4

- Demonstrated live 3nm silicon delivering 448G and beyond in PAM4/6/8 at OCP24

Optical engines are 448G capable

- Delivering 112GHz Electrical Bandwidth

Connectors are improving but not there for PAM4

- We cannot address tomorrow's application with today's technology

Standards bodies

Applications:

- Scale Up and Scale Out
- High density Switches (Petabit switch)

Aggressive time frame: PAM4 may be the only viable solution for 2028

- PAM6 will need a new FEC

QUESTIONS?