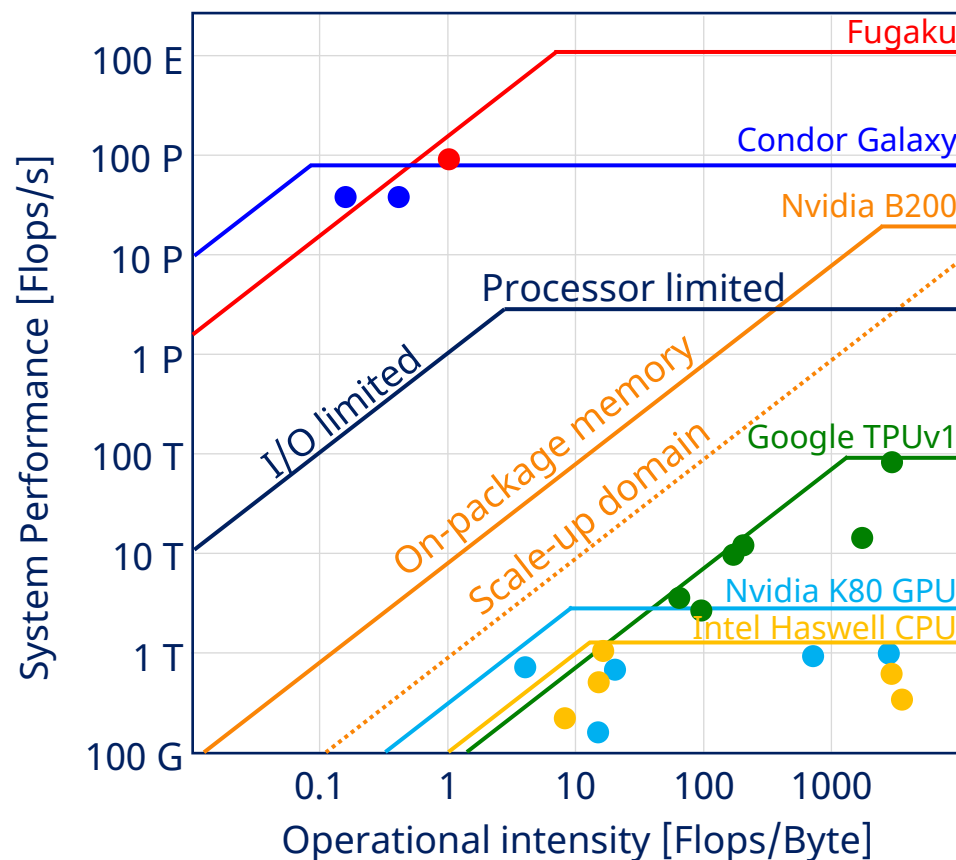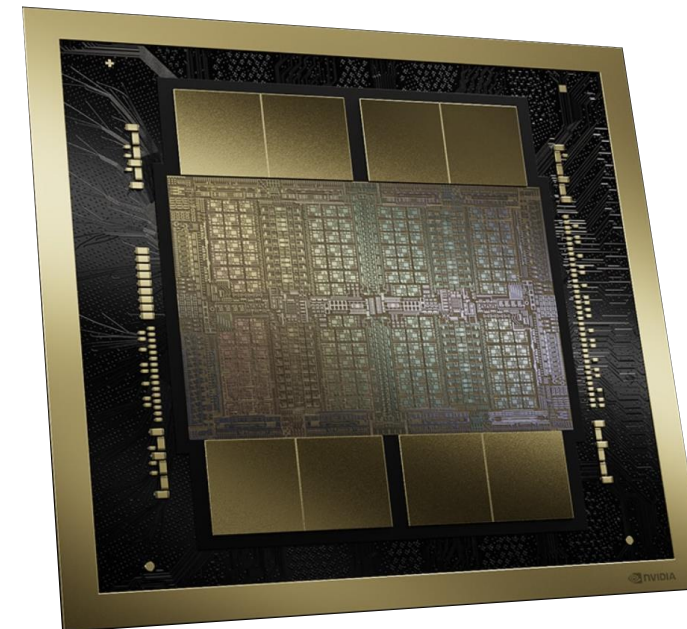# CPX – Operationalizing CPO for AI Clusters

Hot Interconnects, August 2025

AI Engines Need Very High I/O Bandwidths

Chart: System Performance [Flops/s] vs Operational intensity [Flops/Byte]

Y-axis labels (System Performance [Flops/s]): 100 E, 100 P, 10 P, 1 P, 100 T, 10 T, 1 T, 100 G

X-axis labels (Operational intensity [Flops/Byte]): 0.1, 1, 10, 100, 1000

Labels on chart:
- Fugaku
- Condor Galaxy
- Nvidia B200
- Processor limited
- I/O limited
- On-package memory
- Scale-up domain
- Google TPUv1
- Nvidia K80 GPU
- Intel Haswell CPU

Very hungry
Simple calculations ← How data-hungry is your algorithm → Not so hungry
Complex calculations

Sources:
[1] S. Williams, A. Waterman, and D. Patterson "Roofline: An Insightful Visual Performance Model for Floating-Point Programs and Multicore Architectures," Communications of the ACM, 52(4), 65-76 (2009).
[2] N. P. Jouppi et al., "In-Datacenter Performance Analysis of a Tensor Processing Unit TM," Proc. 44th annual int. symp. on computer architecture, 1-12 (2017).
[3] H. Ltaief et al., "Scaling the "Memory Wall" for Multi-Dimensional Seismic Processing with Algebraic Compression on Cerebras CS-2 Systems," ACM/IEEE Int. Conf. High Performance Computing, Networking, Storage, and Analysis (SC'23) (2023).
[4] Nvidia NVL72; online: https://www.nvidia.com/en-us/data-center/gb200-nvl72/

xPUs could use orders of magnitude more I/O BW

High Speed SerDes: Currency of the Datacenter
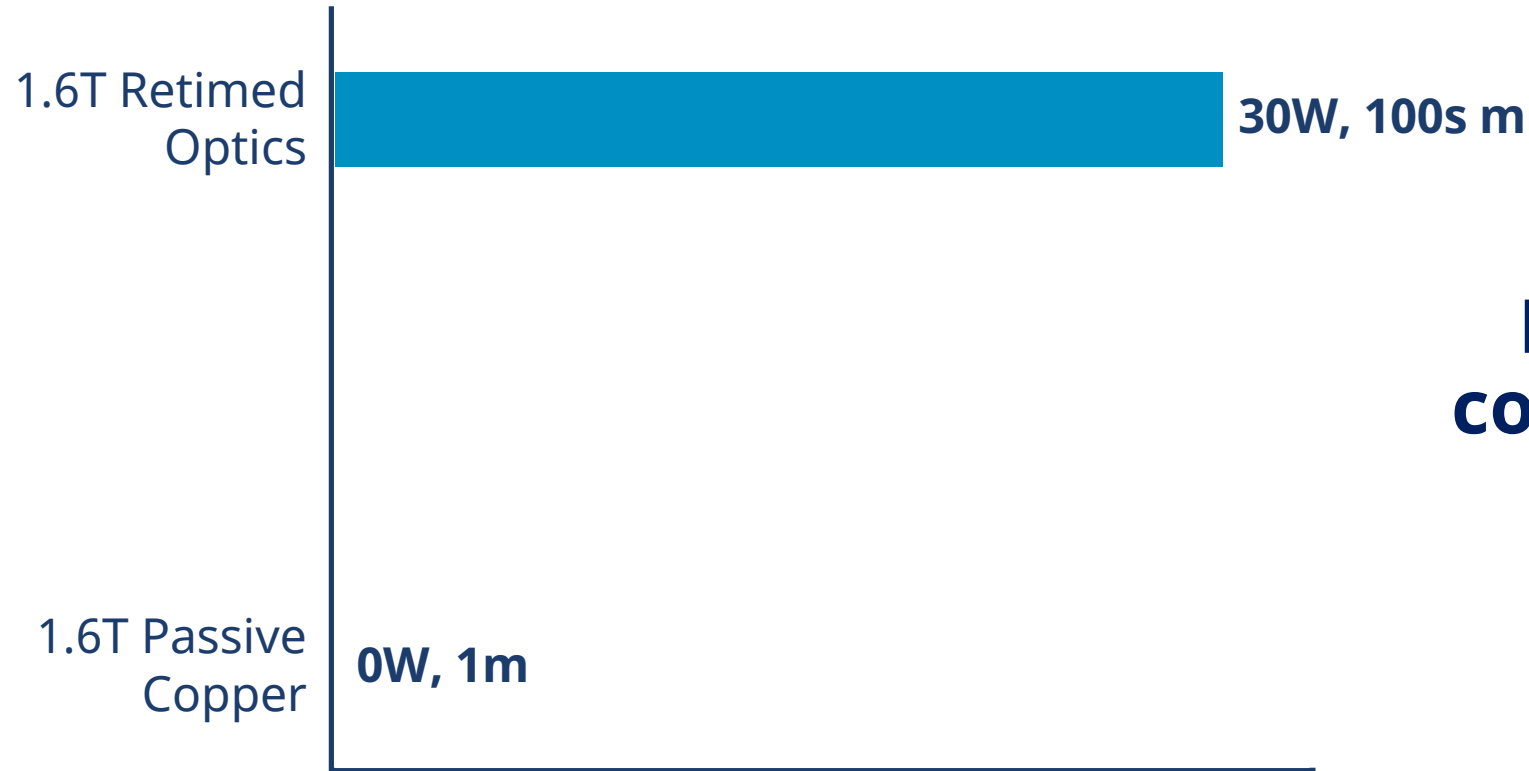## No signs of slowing down next 5-10 years: 200G → 400G
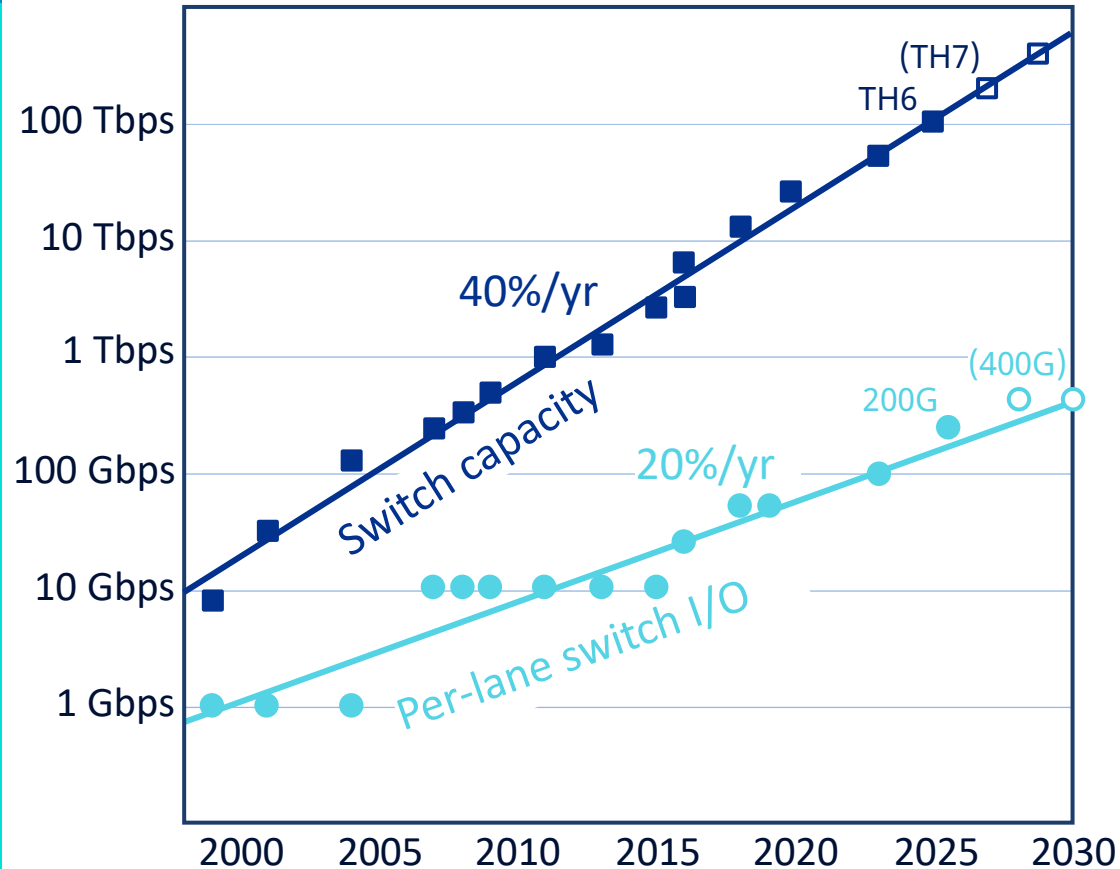
Mixed-media is core requirement
"Copper when you can, optics when you must"

Requires high-speed SerDes I/O

Paradigm extends to 400G

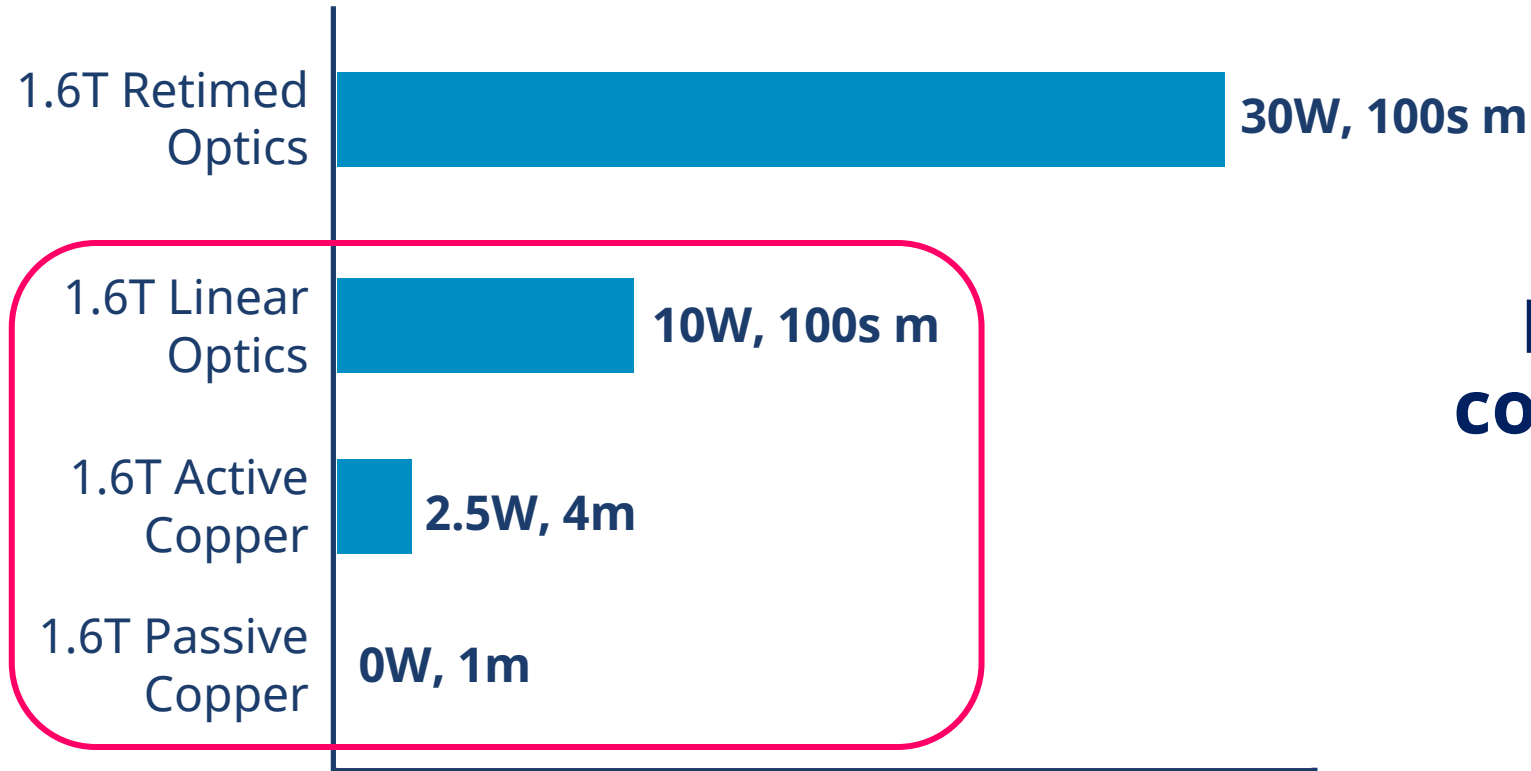➔ Low power, high density optics must interface with high speed SerDes I/O

Mixed Media Support Saves Power in AI Clusters

1.6T Retimed Optics — 30W, 100s m

Retimer DSP power consumption costs $Bs in large cluster operations

1.6T Passive Copper — 0W, 1m

Copper when you can, optics when you must

Mixed Media Support Saves Power in AI Clusters

1.6T Retimed Optics — 30W, 100s m

1.6T Linear Optics — 10W, 100s m

1.6T Active Copper — 2.5W, 4m

1.6T Passive Copper — 0W, 1m

Retimer DSP power consumption costs $Bs in large cluster operations

Copper when you can, optics when you must

Avoid retimer DSPs at all costs

Example Mixed-Media AI Clusters

	Passive Copper & Retimed Optics	Passive Copper & Linear Optics	Passive Copper & Retimed Optics	Active Copper & Linear Optics
# xPUs	150k	150k	600k	600k
BW per GPU (any-to-any connectivity)	25.6 Tbps	25.6 Tbps	25.6 Tbps	25.6 Tbps
Switches & Links (as % of cluster power)	240 MW 48%	150MW 37%	1.7 GW 61%	600MW 37%

~ $320M savings over 5 years   **~ $3.8B savings over 5 years**

xPUs   •••   xPUs
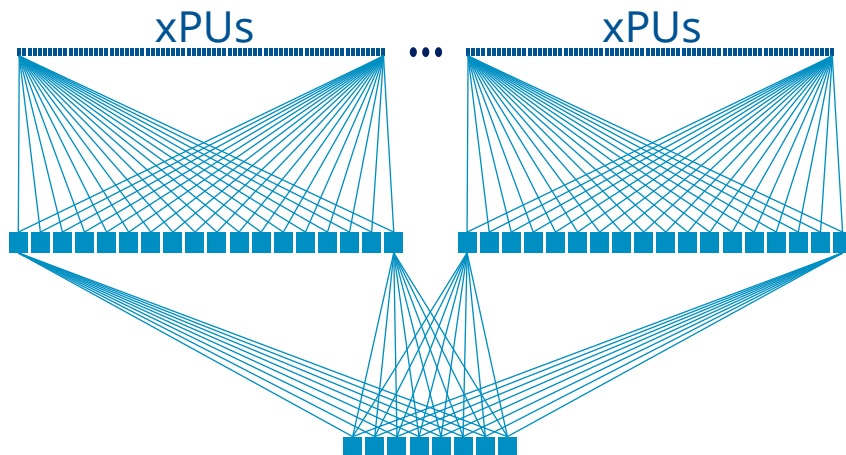
Active linear copper enables larger clusters without % power impact

Densification Needed to Avoid Retimer DSPs

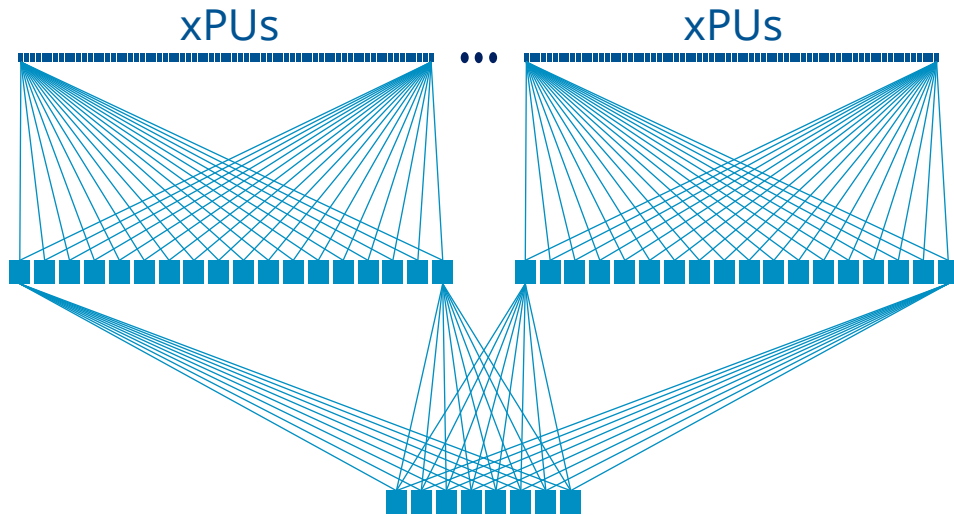|	I/O density (TX+RX) [Tbps/mm]
OSFP 1.6T	0.14
200G SerDes	1.5

5x – 10x I/O densification
- ➜ Shorter on-board traces
- ➜ Denser systems, shorter cables
- ➜ No retimer DSPs needed
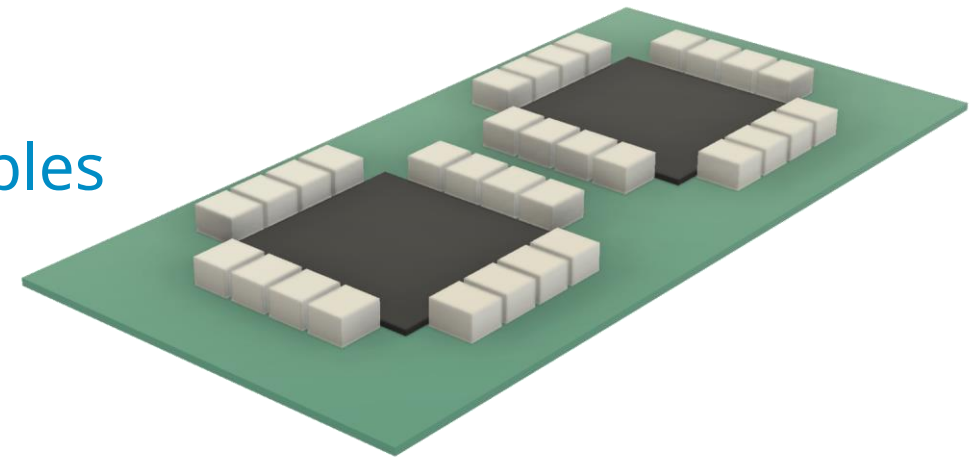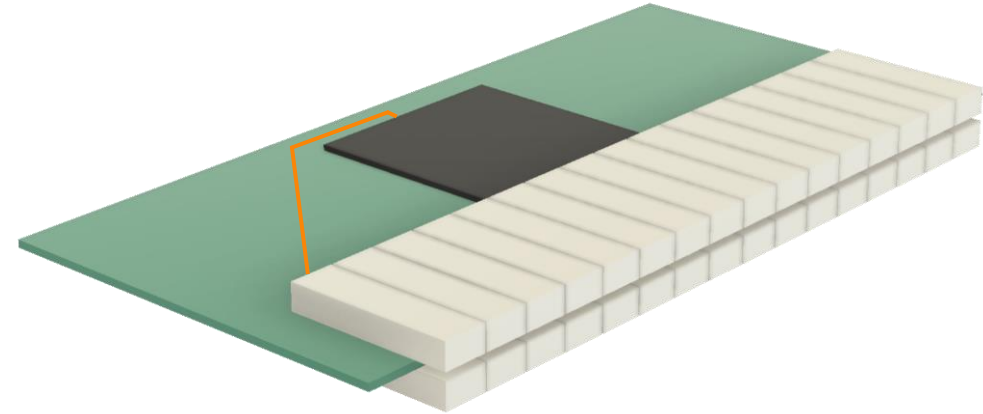
Power efficient I/O for AI needs to approach SerDes densities

AI Cluster Scaling Needs Full Fan-Out Networking
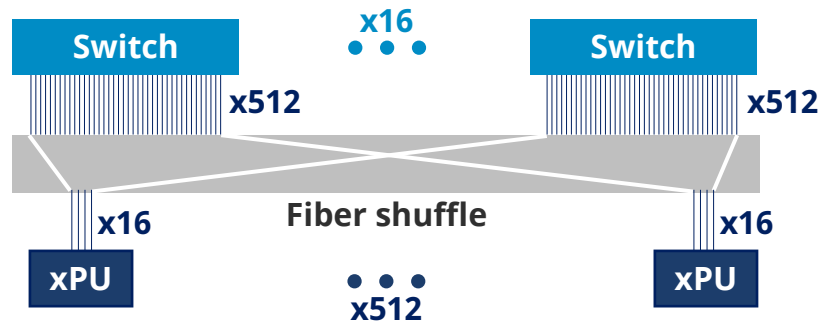
xPUs ... xPUs

- Clustered Arch → Mixed media I/O

- Large # of xPUs → Large switch radix

- High xPU I/O BW → Large # of switches
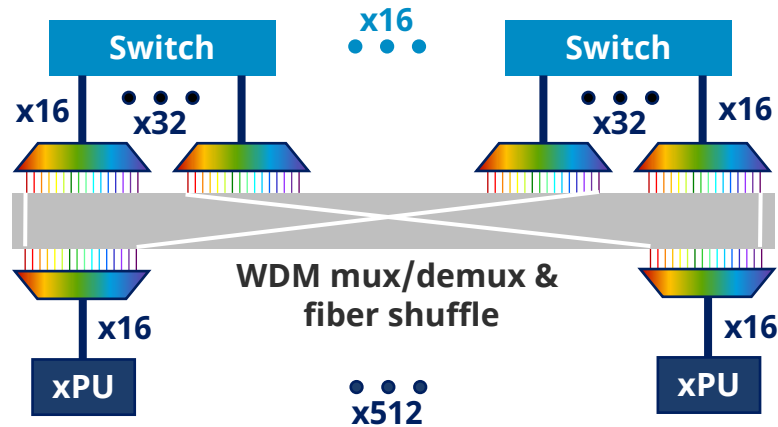
Full fan-out at SerDes rates

No big fat point-to-point pipes
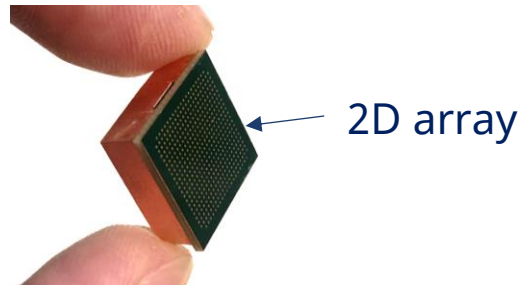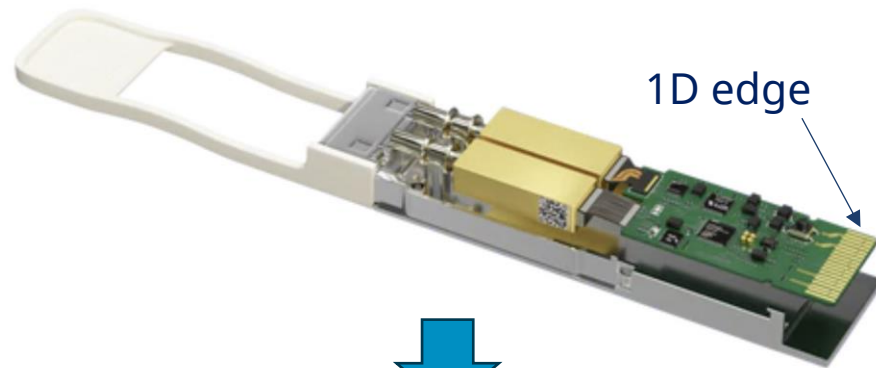
Single-Wavelength Optical I/O Beats WDM

Single-wavelength fiber shuffle
➜ Simple, low-loss

Multi-wavelength (WDM) fiber shuffle
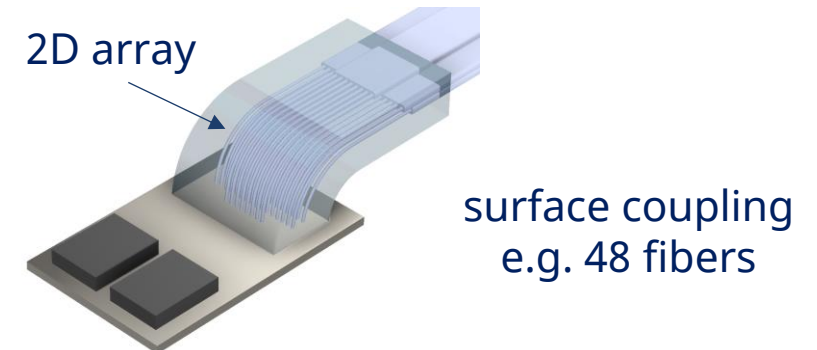➜ Complex, lossy, expensive

2D Arrays Enable Scalable High-Density I/O
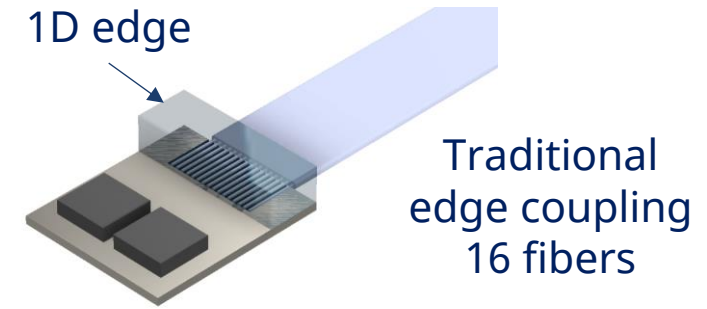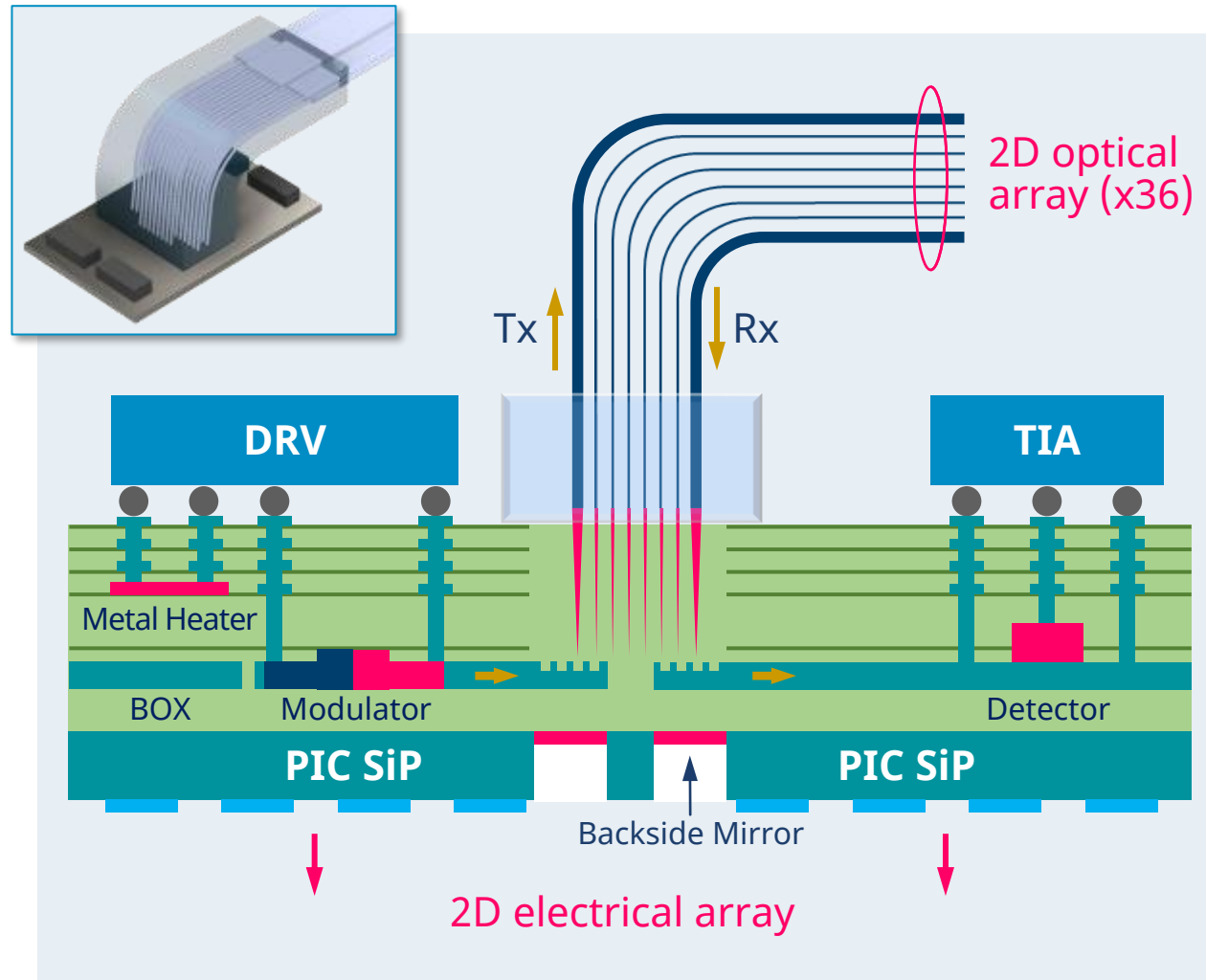
Electrical

1D edge

2D array

5-10x better density

Optical

1D edge

Traditional edge coupling 16 fibers

2D array

surface coupling e.g. 48 fibers

5-10x better density

A 2D Optical Engine – Details

2D optical array (x36)

Tx ↑ ↓ Rx

DRV

TIA

Metal Heater

BOX

Modulator

Detector

PIC SiP

PIC SiP

Backside Mirror

2D electrical array

Driver & Trans-Impedance Amplifier
224G per lane using SiGe technology
Advanced equalization

Photonic IC (Silicon Photonics)
Polarization insensitive surface couplers
High-speed modulators (> 50 GHz)
High-speed detectors (> 50 GHz)

Advanced Packaging
3D stacked electronics and photonics ICs
2D Fiber coupling with backside mirror

It Is Clear What AI Wants From Optics

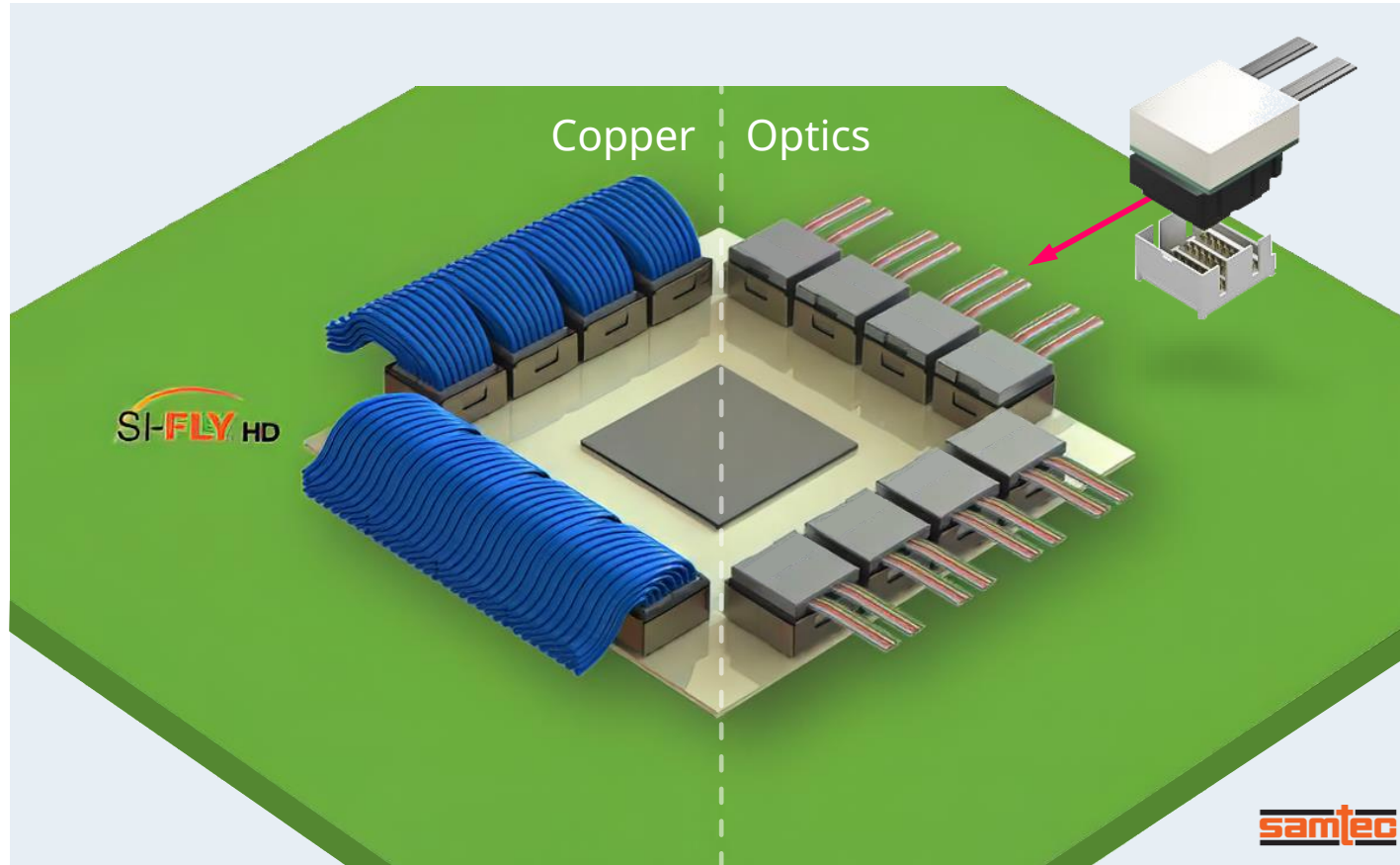Low power – Support for retimer-free linear optics

High density – Matched to SerDes I/O

Full fan-out – Single wavelength per fiber (DR optics)

Operations – Serviceability and multi-vendor ecosystem

How?

Common Connector for Co-Packed Copper & Optics

Copper | Optics

SI-FLY HD

samtec

- 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

Key Takeaways

Mixed-Media support
→ Optics must support high speed SerDes for next 5-10 years

Full fan-out for scale-up/scale-out
→ Scale-up & scale-out converging on DR optics

Linear Optics paradigm for low-power I/O
→ Co-packaged I/O eliminates retimer DSPs

Multi-vendor ecosystem
→ CPX paradigm replicates front-panel pluggable paradigm

Thank You