

2024 Hot Chips

# 4 Tb/s Optical Compute Interconnect Chiplet for XPU-to-XPU Connectivity

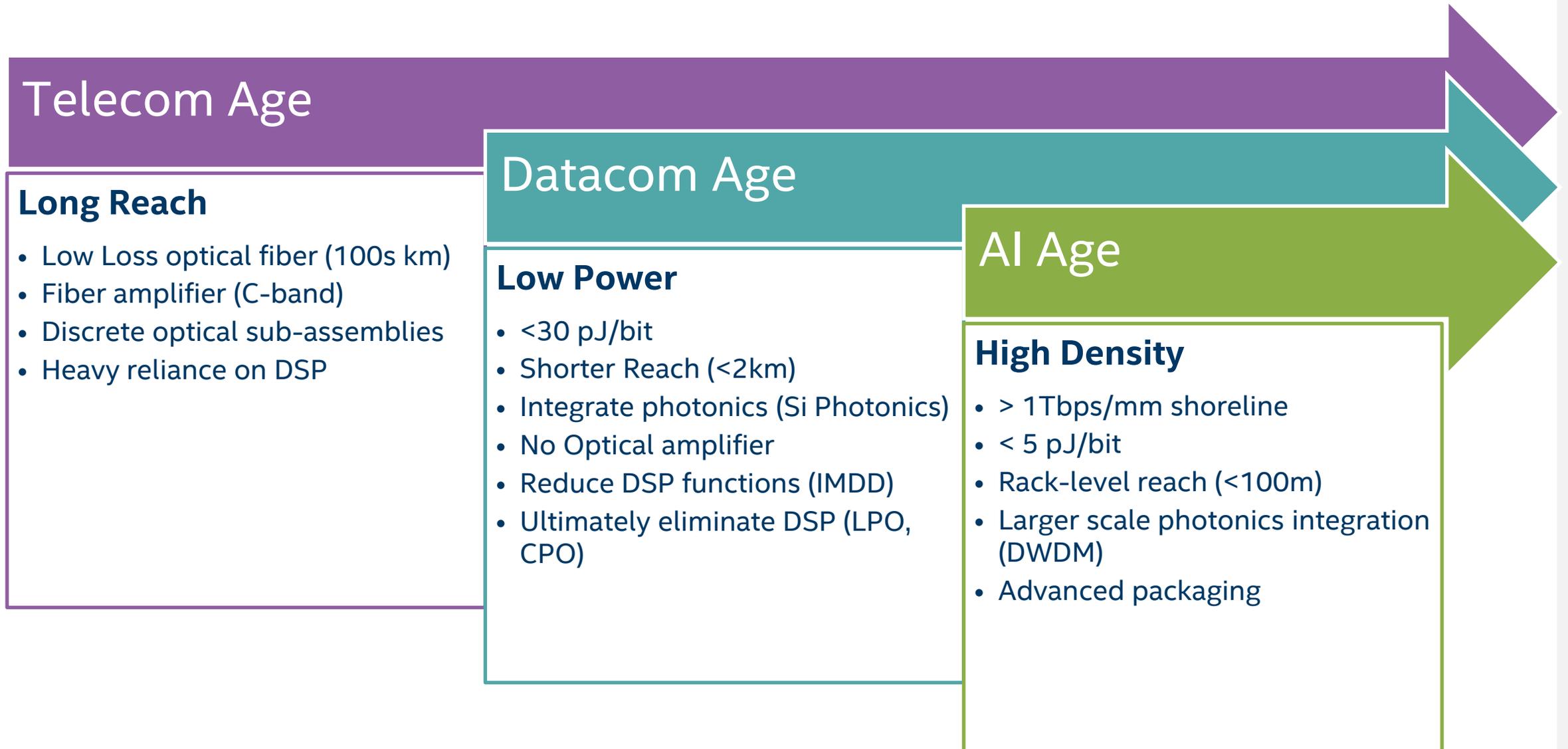
Saeed Fathololoumi, Intel Corporation

August 2024

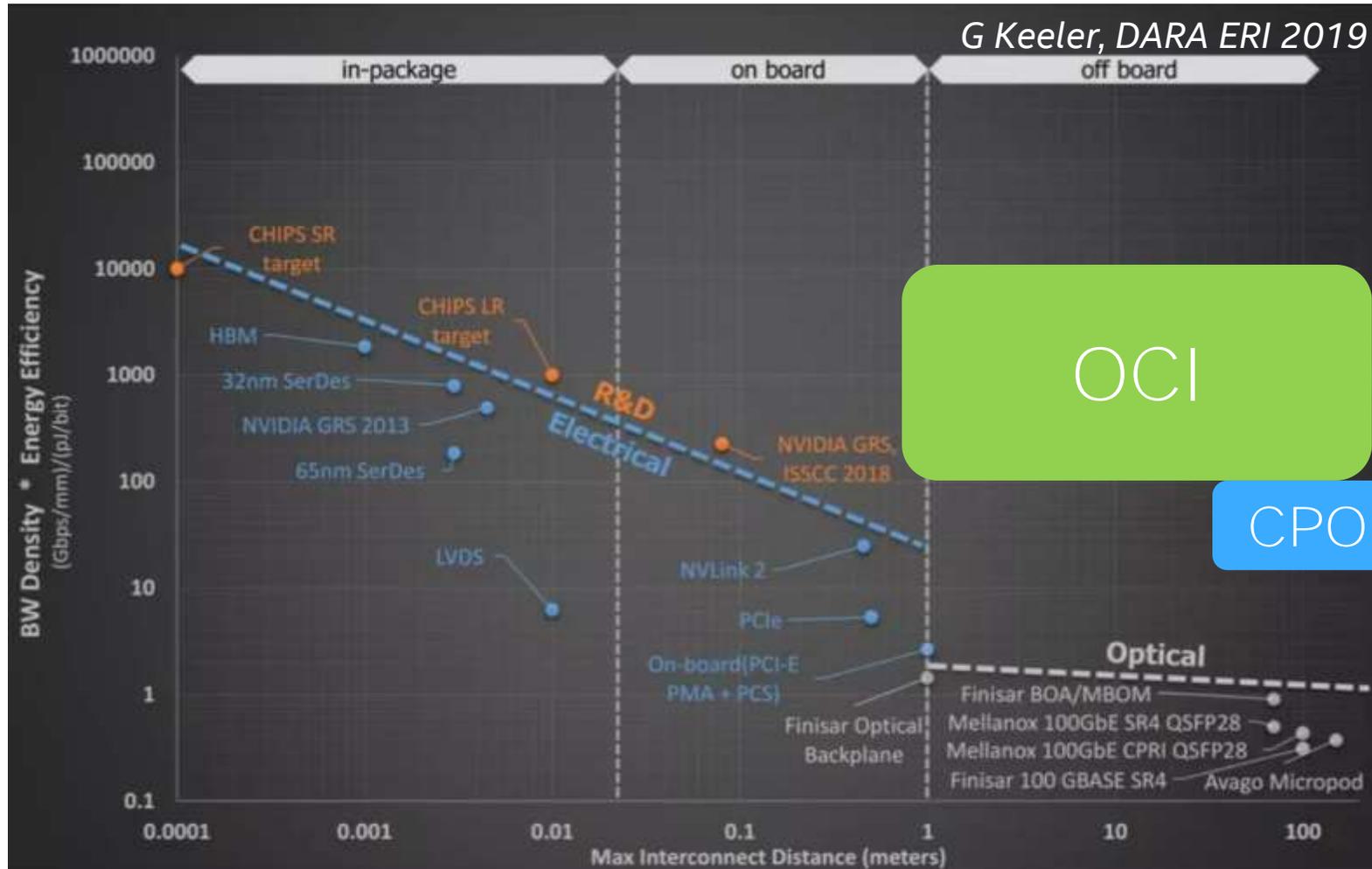
The Intel logo, consisting of the word "intel" in a lowercase, sans-serif font, with a registered trademark symbol (®) to its upper right. The logo is white and positioned in the bottom left corner of the slide.

intel<sup>®</sup>

# Evolution of Optical Communications



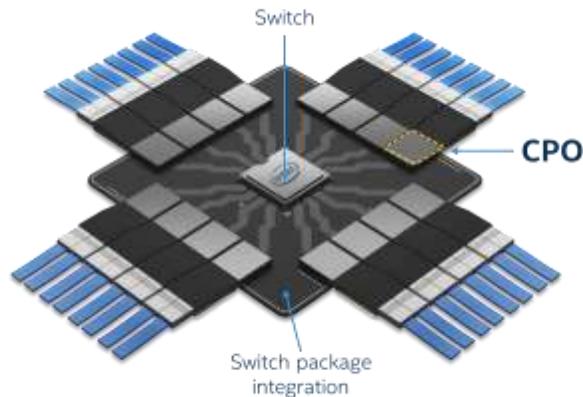
# Interconnect Technologies and Co-packaged Optics



- Co-Packaged Optics (CPO): Networking applications
- Optical Compute Interconnect (OCI): compute fabrics

# Co-Packaged Optics: Use Cases

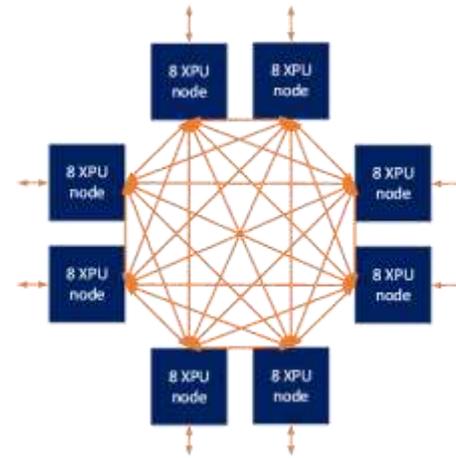
## Ethernet CPO



### Ethernet networking

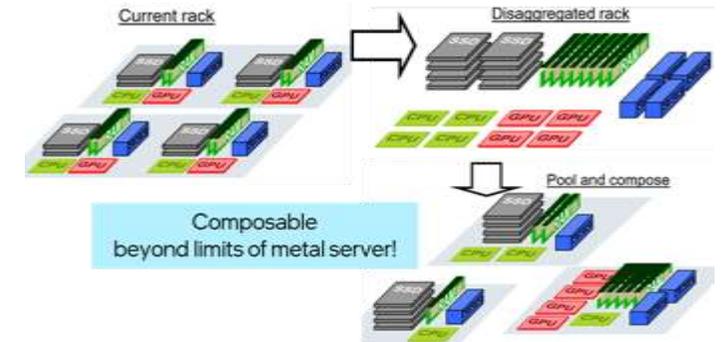
- Reduce power and cost vs. pluggable
- Standards compliance and legacy interop
- High bandwidth density
- Not sensitive to latency

## Optical Compute Interconnect (OCI)



### Compute Fabric (AI/ML)

- Copper replacement; book-end - no existing optical standard
- More bandwidth density
- Longer reach for larger clusters
- Low power
- Same or better latency



### Resource disaggregation

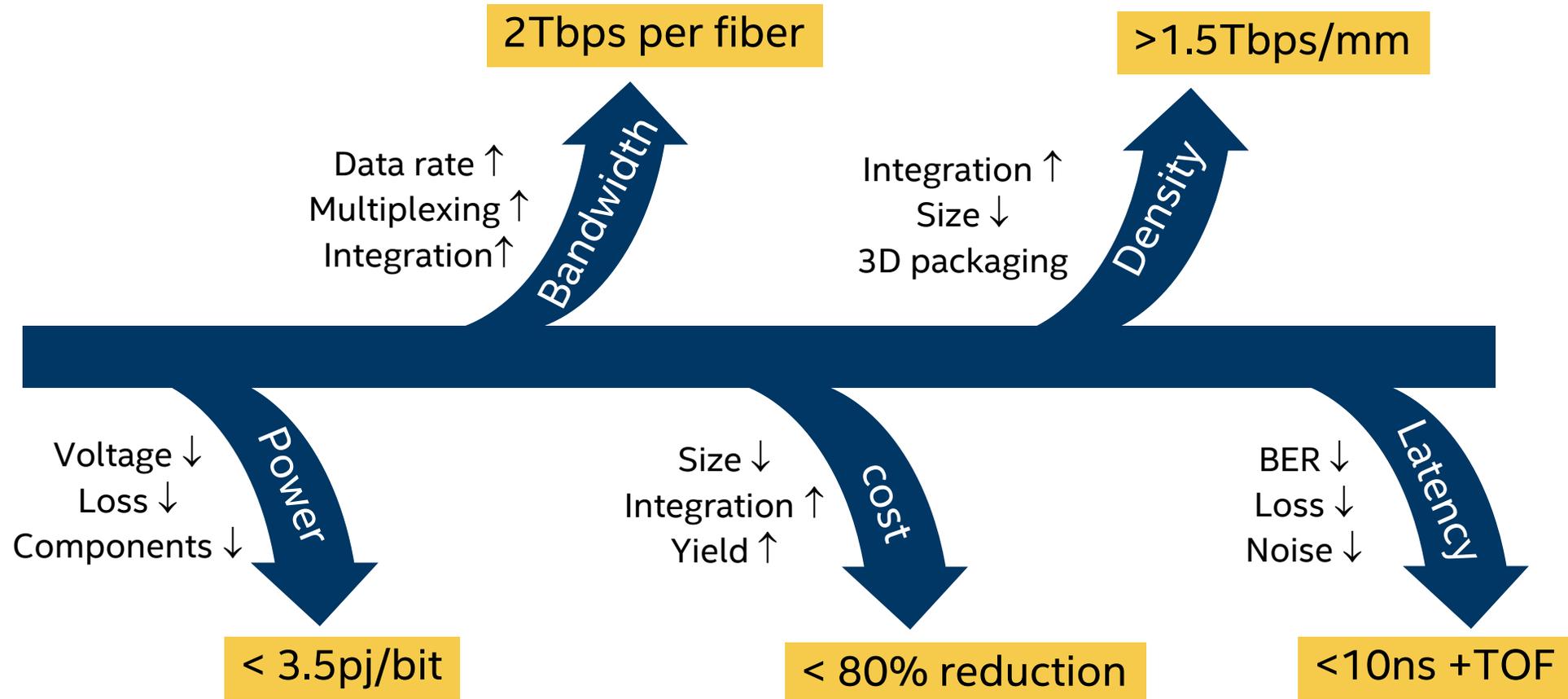
- Unconstrain resources from package/socket limitations
- Create pools of resources
- Latency sensitive
- High bandwidth density
- Low power

# Co-packaged Optics for Ethernet vs. Compute

	Ethernet Networking	Compute Connectivity
Incumbent technology	Pluggable optics	Cu SERDES i/f (i.e. PCIe/CXL)
Typical use cases	Switch-to-switch NIC-to-switch	Compute fabric scale-up/scale-out Resource disaggregation/pooling
Typical reach	Up to kms	Up to 100m
Typical Latency	>100ns	<10ns
FEC	KP	None or light-FEC
Optical interface standard Wavelength grid Modulation speed and format	MSA 20nm O-band FR 100&200Gb/s PAM4	None yet MSA emerging: 100s GHz 16-64Gb/s NRZ is popular
Host interface standard	IEEE/OIF	UCle/PCI-SIG

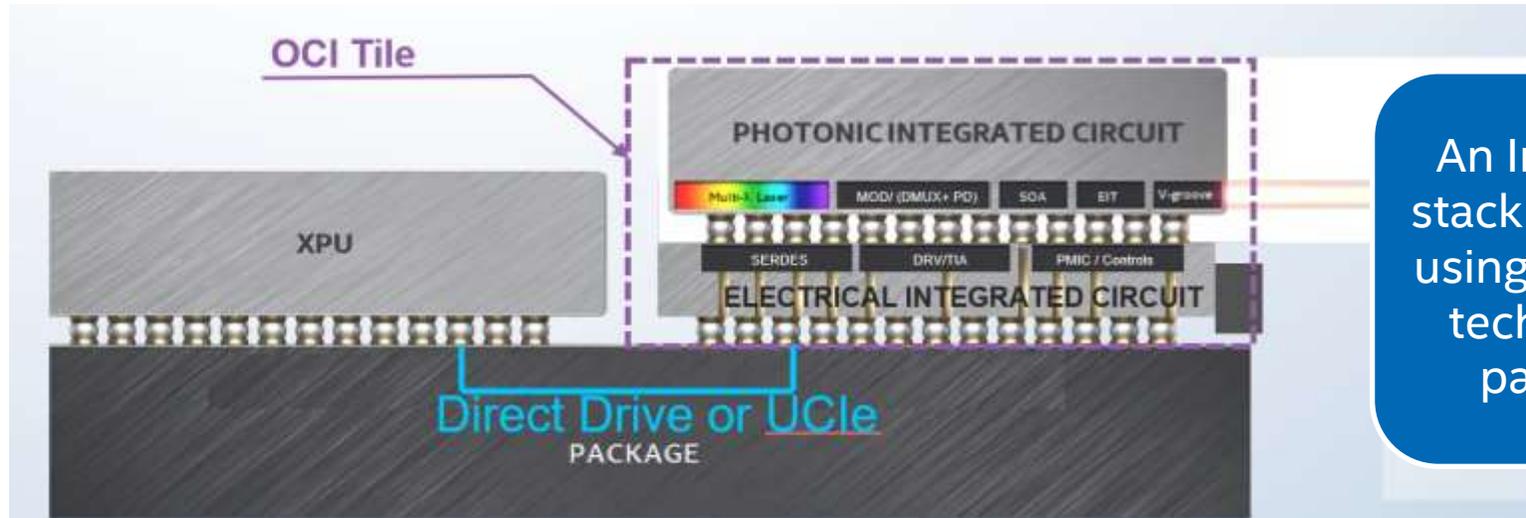
Different requirements drive different architectures and solutions

# OCI Scaling Vectors and KPIs Points



In addition to power and cost, solution needs to address total BW, BW density and latency

# Integrated Optics: Our Approach

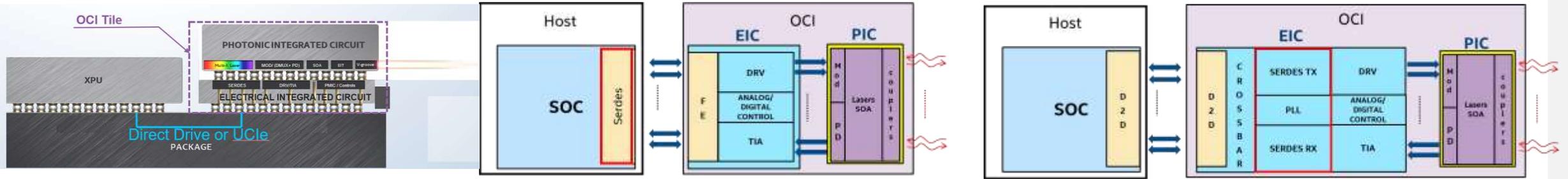


An Intel OCI chiplet is a die-stack that provides optical I/O using **Intel Silicon Photonics** technology and can be co-packaged with the XPU.

- Greater photonic functionality integrated on the Photonic Integrated circuits (PICs) → **best done in silicon photonics**
- Integration of PICs with best of breed ICs using Advanced Packaging to create Optical Engines → **heterogenous integration trumps monolithic integration**
- Tighter Integration of Optical engines with the host (xPU, Switch) → **enabling new systems and applications**

We refer to this approach as Co-Packaged Optics (CPO) when applied to networking applications and Optical Compute Interconnect (OCI) when applied to compute fabrics

# Co-Packaged Optics: Serial or die-to-die (D2D)



	<b>High-Speed Serial (Un-retimed direct drive)</b> Ethernet, IB, PCIe, ....	<b>Wide-slow D2D interface</b> UCIe
Serdes	On host: VSR/MR or more	In chiplet: Purpose-built, optimized for optical channel
Shoreline bandwidth density	~500-800 Gbps/mm (per direction) Limited by Serdes	>1.2 Tbps/mm (per direction) Unconstrained by host interface
Power efficiency (end-to-end, including host I/O)	<10pJ/b Serdes at 3-5pJ/b + I/O chiplet	<5pJ/b Power efficient D2D interface + optimized Serdes
Optical signaling	Rate and format set by host	Independent of host – can be optimized for use-case
Ecosystem	Broad and healthy; standards exist; multi sourcing and interop; Chiplet to enable existing ports on current Silicon	Nascent Need standards and interop First UCIe ports in volume a few years out

*D2D attach provides best performance, but ecosystem is not there yet*

# 4Tbps OCI and System Overview

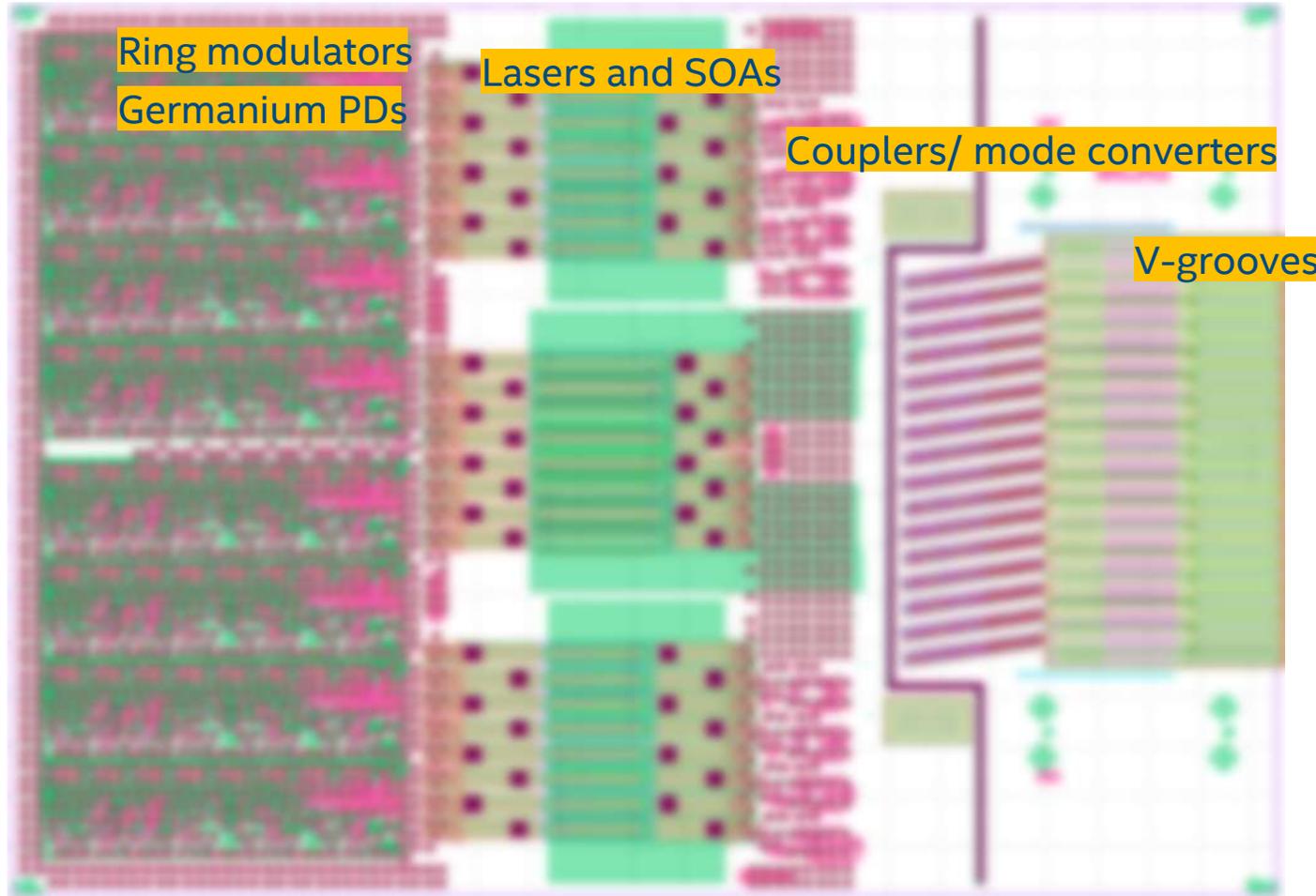


Enable optical connectivity within existing compute eco-system

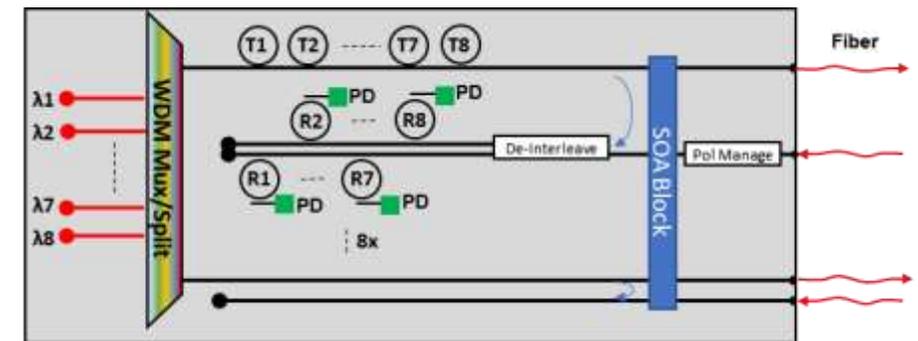
- Attach to standard I/O ports on compute platform
- Un-retimed, direct drive from host PCIe5 (and UPI) SERDES
- 2 Tbps BW per direction ( $8f \times 8\lambda \times 32 \text{ Gb/s}$ )
- Enabling for next-gen un-retimed PCIe6\* (64 Gb/s PAM4) attach
- Enabling for future co-optimized parallel i/f design

\* PCI-SIG Optical WG baseline proposal for ECN to PCIe Base Specification Rev6.2 (aka Gen6) to enable PCIe compliant optical links will require new features in PHY logical block, nominally implemented in a PCIe Retimer (i.e., ECN will focus on updates to section 4.3, PCIe Retimer specification.)

# 4 Tb/s (8 Tb/s bi-directional) Silicon Photonic IC



- Supports parallel and serial host i/f
- Optimized for power and size
- High volume SiPh platform
  - 8f x 8λ x 64 Gbaud (future proof)
  - Laser and SOA sharing
  - High speed ring modulators and Ge-PDs
  - Micro-ring drop filter DMUX
  - Polarization diverse Rx
  - V-groove passive fiber coupling or detachable optical connector

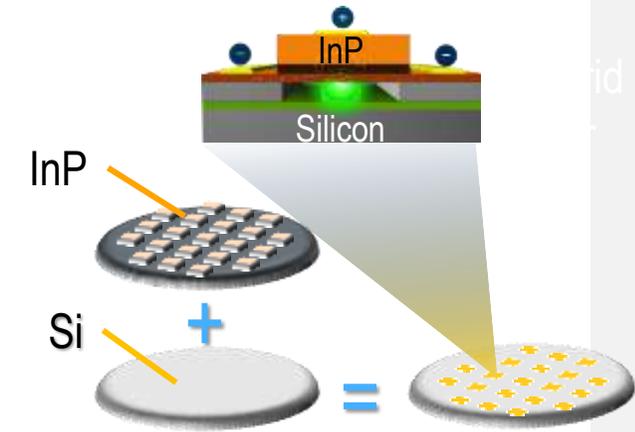
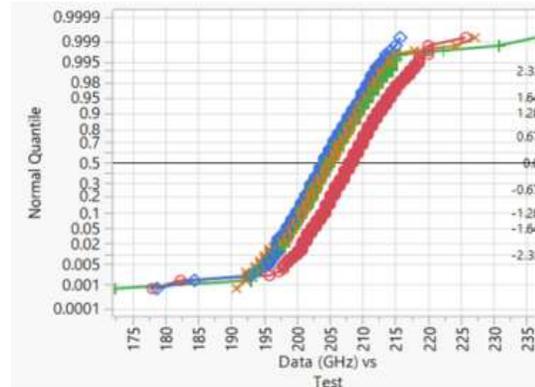
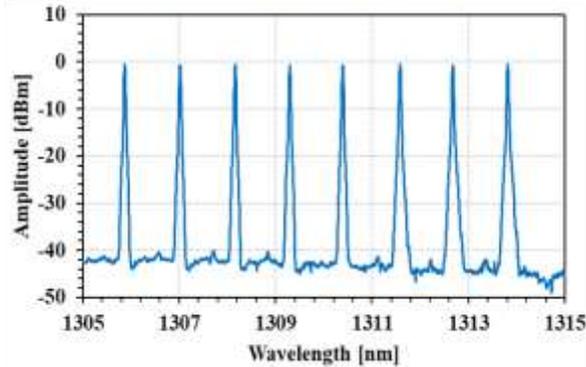


S. Fatholouloumi et al., "Highly Integrated 4 Tbps Silicon Photonic IC for Compute Fabric Connectivity," 2022 IEEE Symposium on High-Performance Interconnects (HOTI), CA, USA, 2022

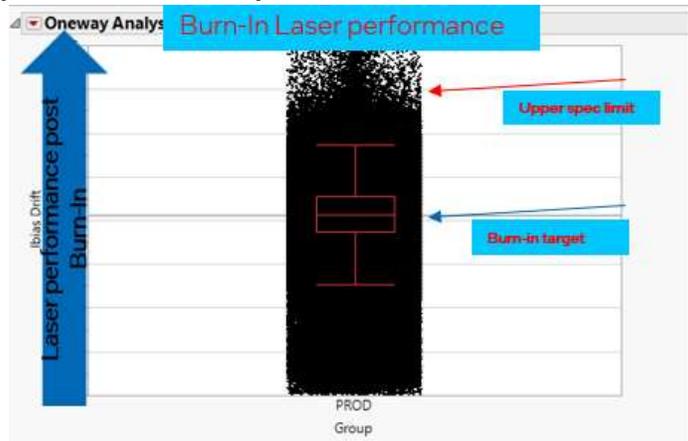
# Value of Integrated III-Vs

Wafer-level fabrication, burn-in, SORT, and reliability testing

- Performance (CMOS control):  $< \pm 15\text{GHz}$  variation for WDM grid (1400 8- $\lambda$  arrays)

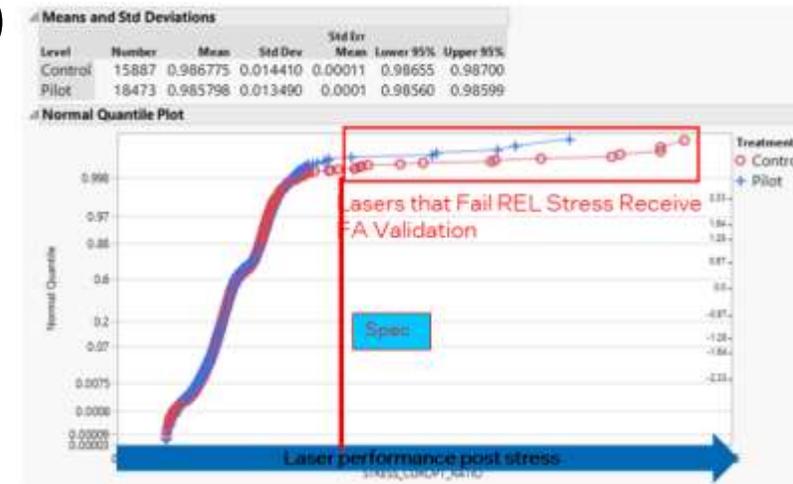


- Cost (wafer scale):  $> 11\text{M}$  lasers BI on wafer



- Reliability (rigor & scale)

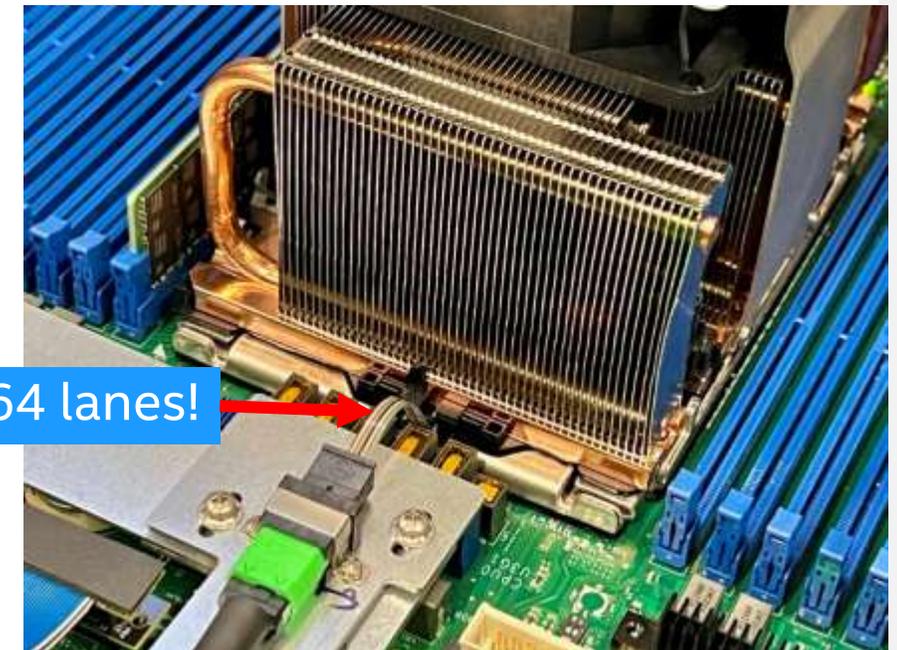
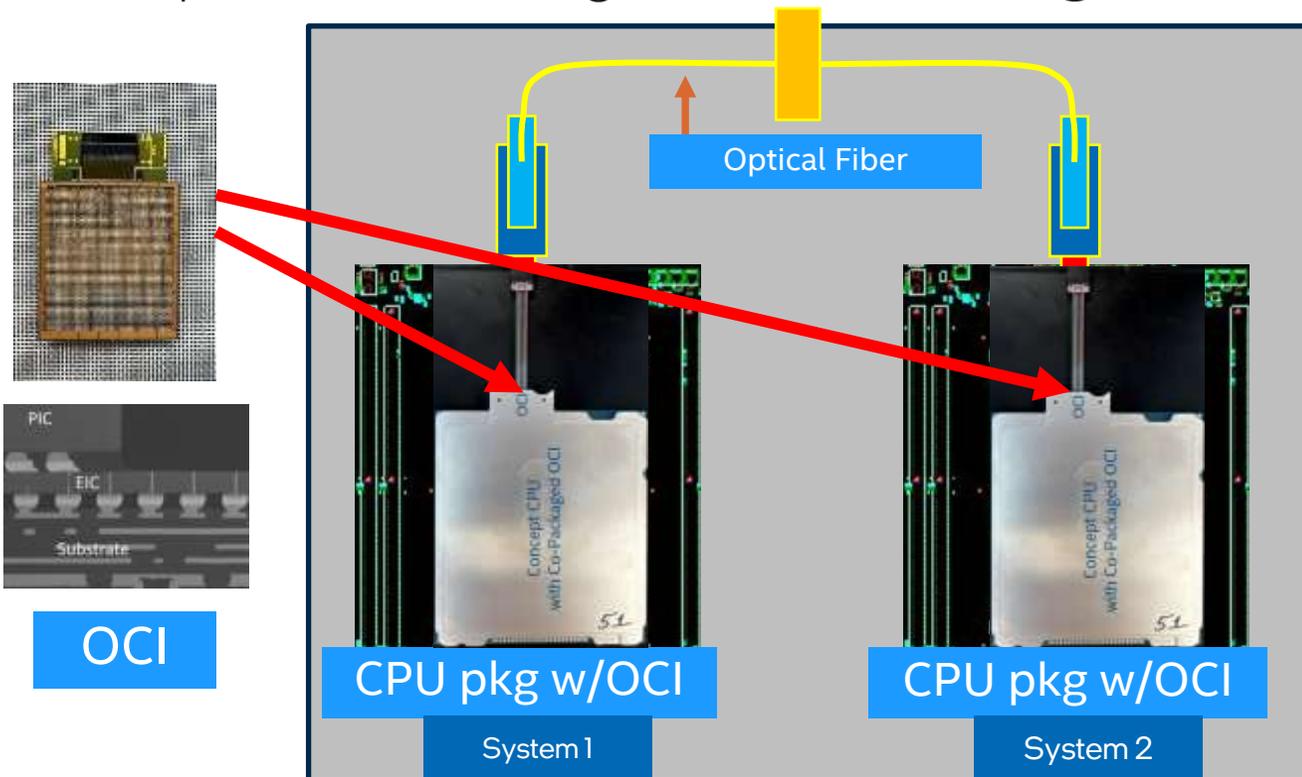
- 0.1FIT from field data
- Enabled by rigorous process certification, wafer-level BI, SORT and reliability testing, and ability to collect large data sets
- DOE of  $> 30\text{K}$  lasers  $\rightarrow$



# Concept CPU with Co-Packaged OCI

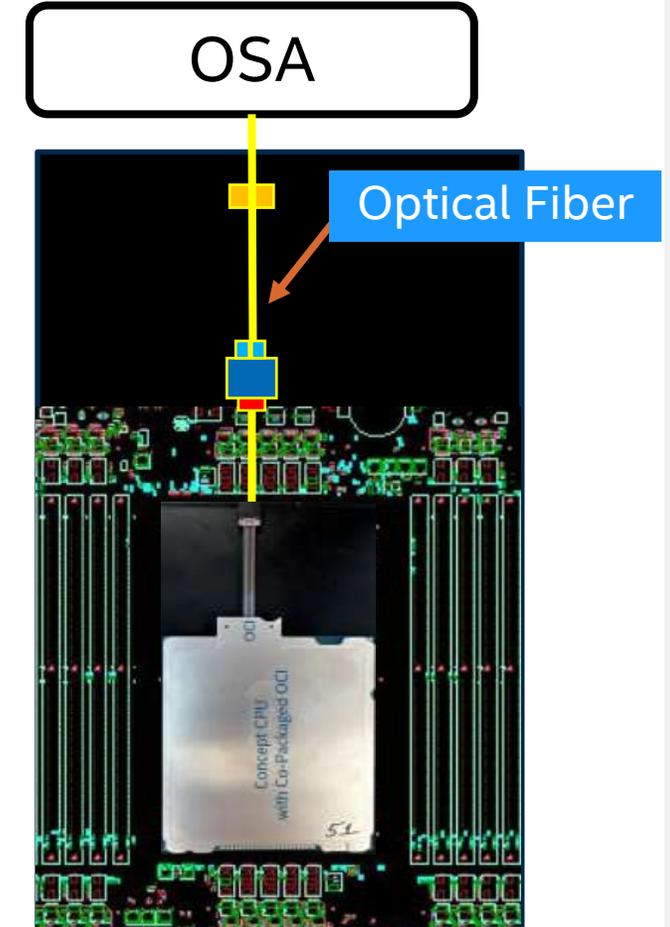
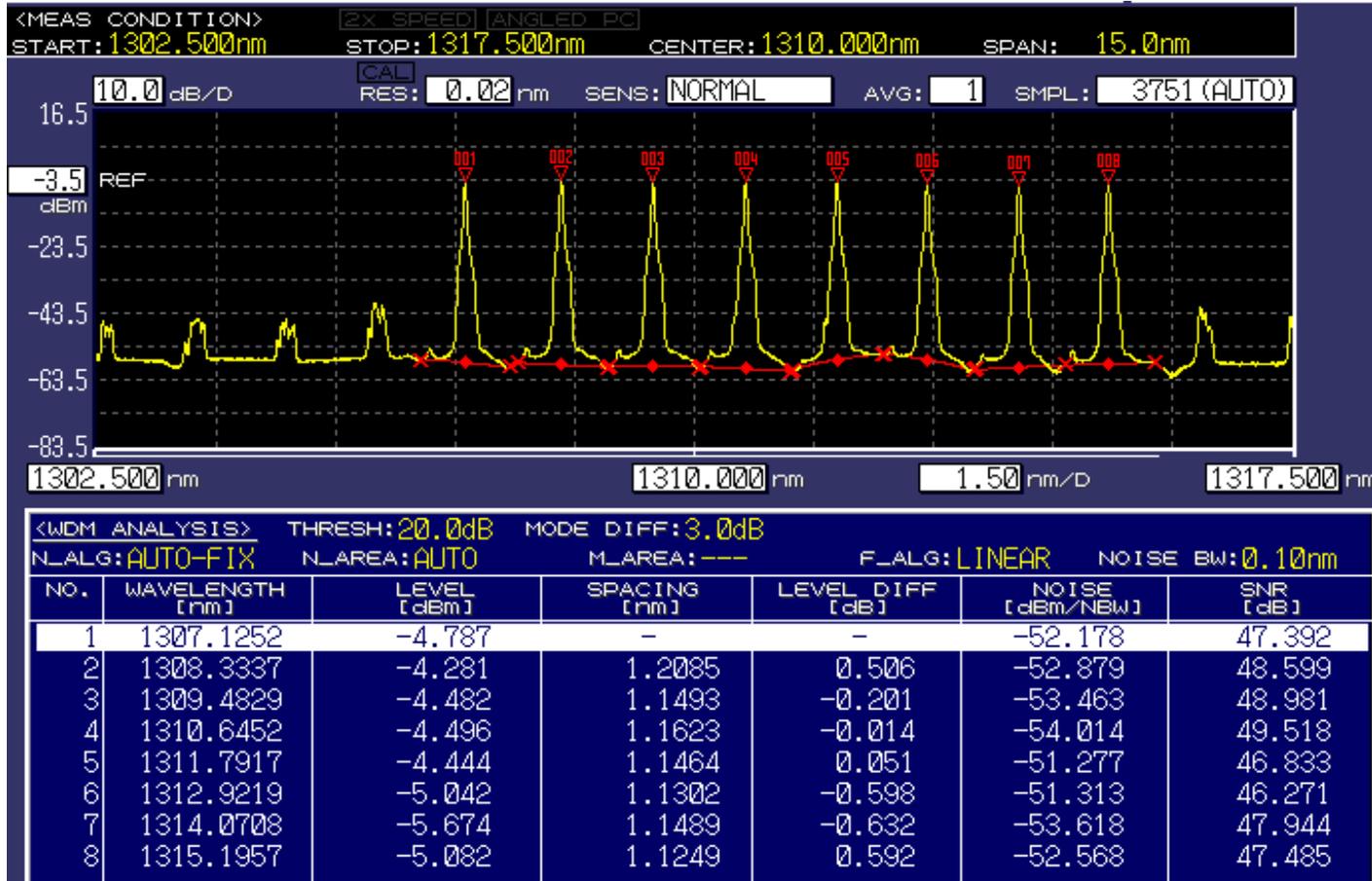
## Package and Platform Integration

- ✓ Gen5/CXL2 links
- ✓ OCI (fully integrated optical engine with integrated lasers), 8 SMF fiber pairs (no PMF required), 8 wavelengths/fiber, 64 lanes @ 32Gb/s = 4Tbps (bidirectional)



# Measured Integrated Laser Spectrum

## 8-wavelength integrated laser

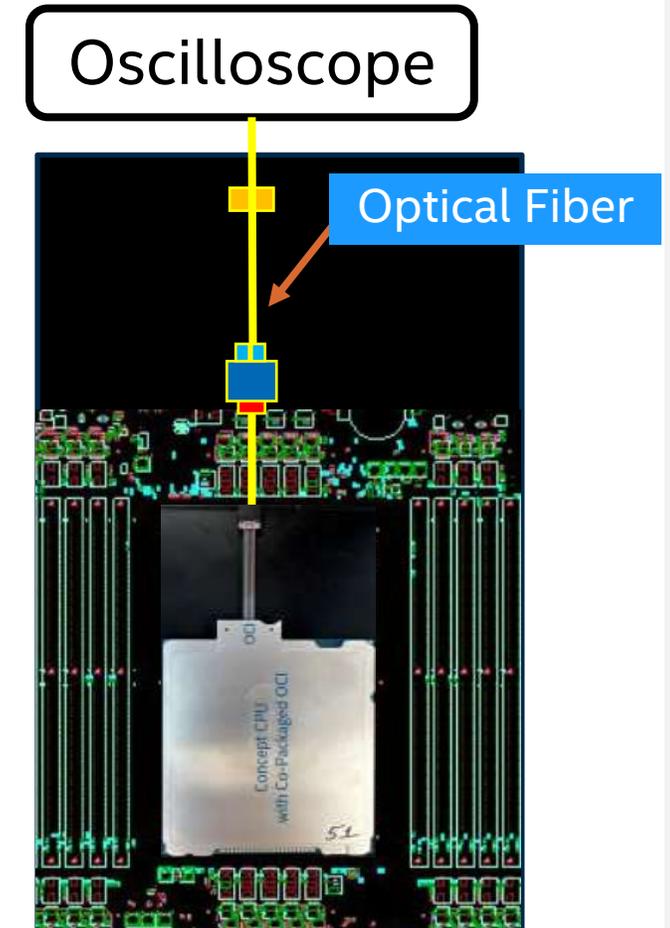
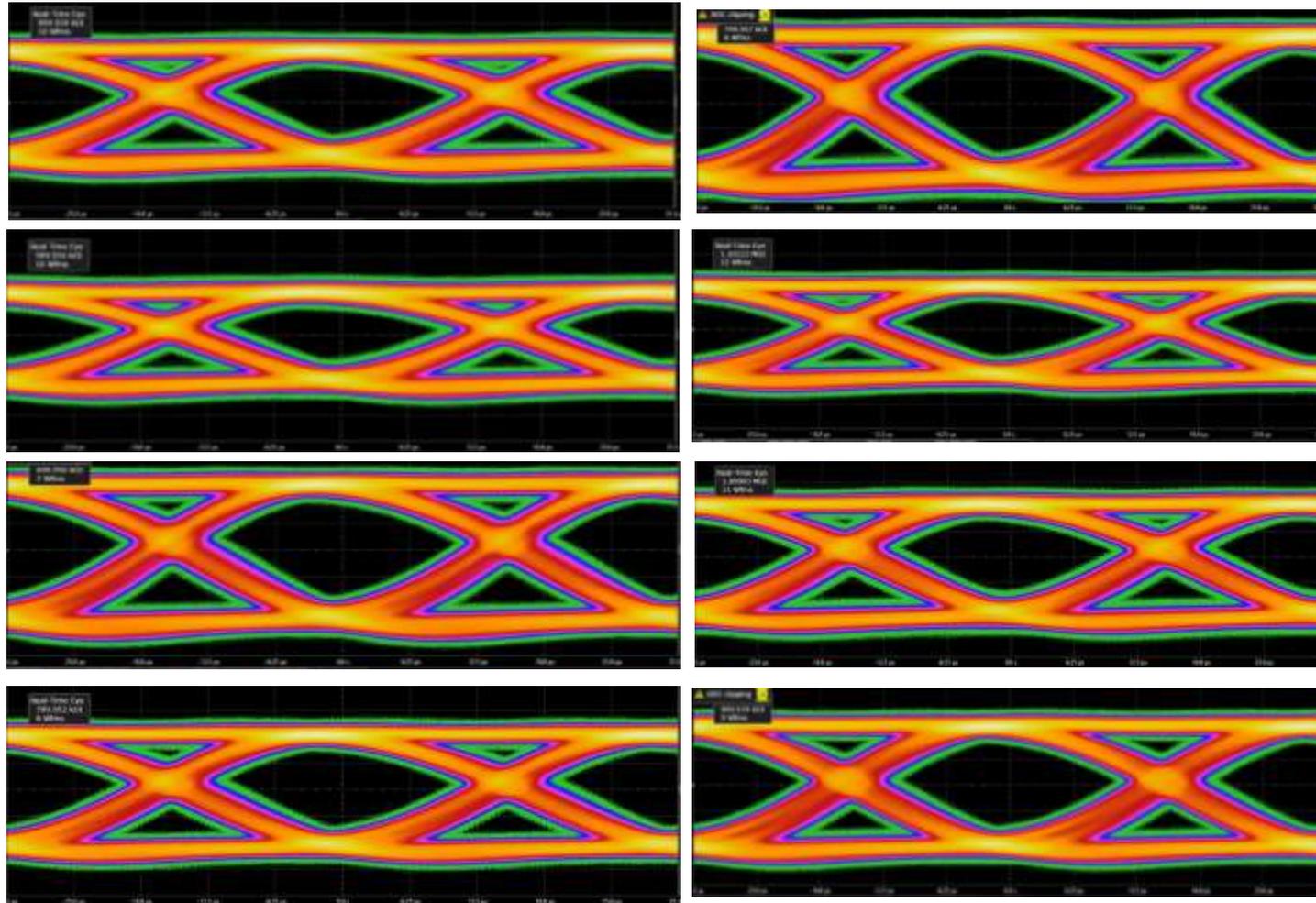


Output power uniformity:  $\pm 0.7$  dBm

Wavelength spacing uniformity:  $\pm 15$  GHz

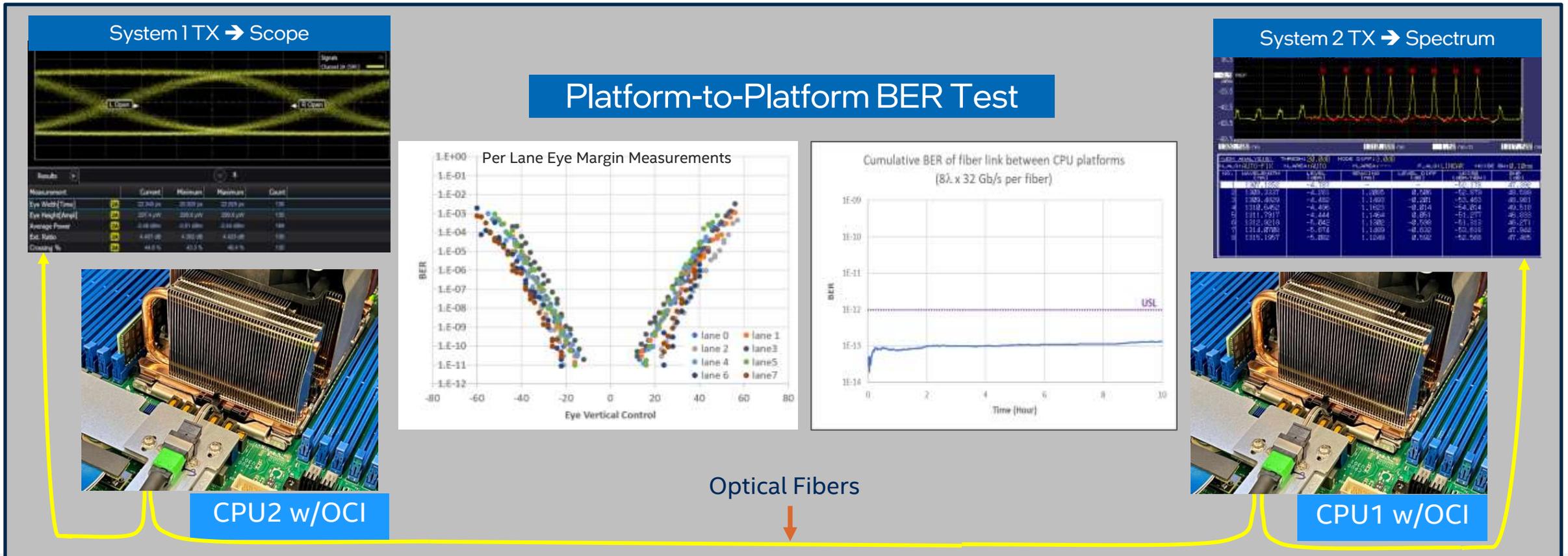
# Measured Transmitter Output Eye

Data rate = 32Gb/s



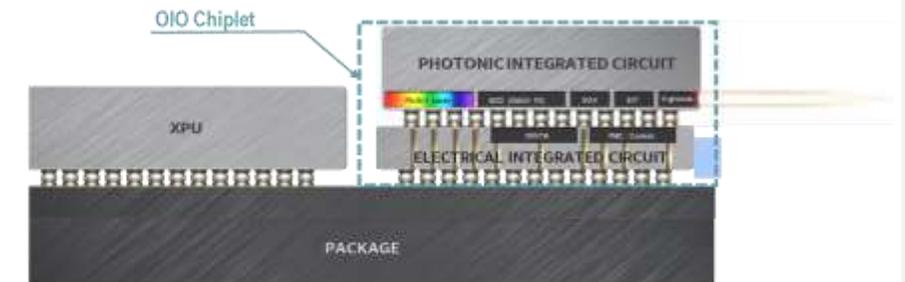
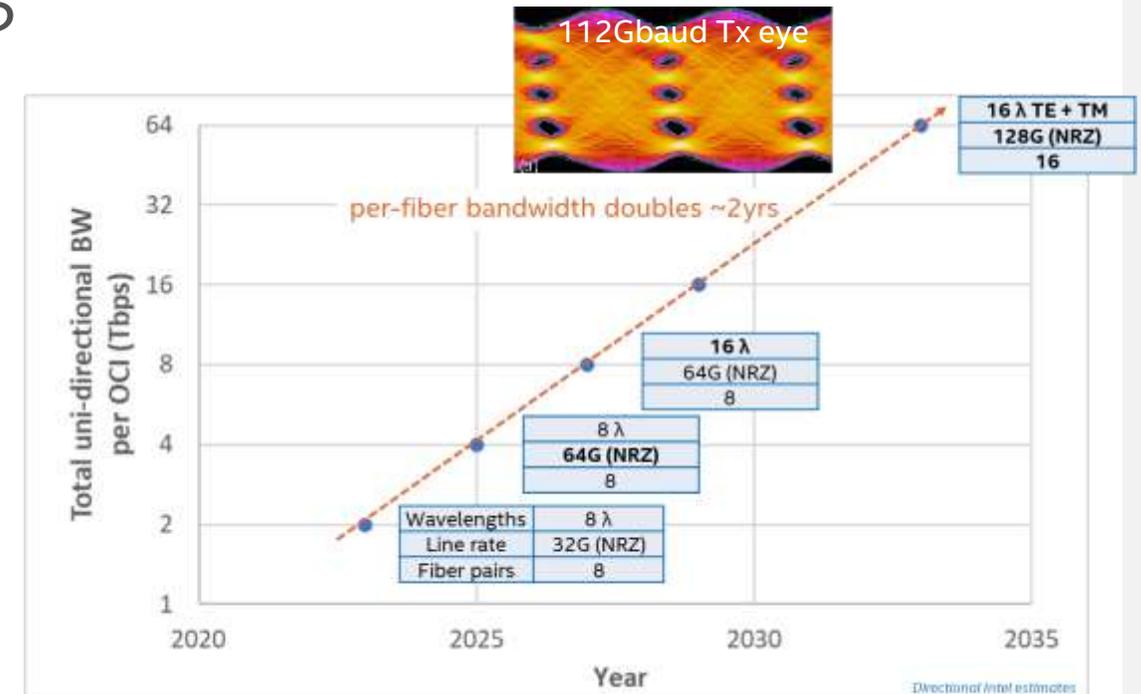
# CPU to CPU Optical Communication

- First demonstration of CPU-to-CPU communication over co-packaged OCI and fiber
- CPU1 injects 32Gb/s/lane PRBS31 data, CPU2 receives and detects errors



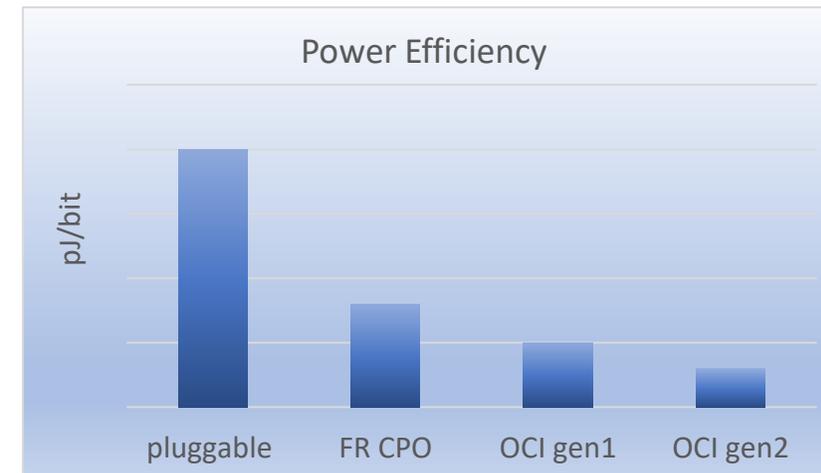
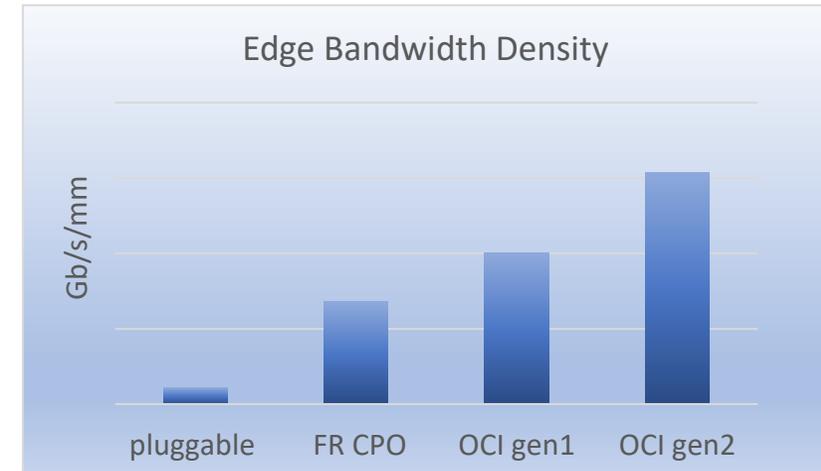
# Future Bandwidth Scaling

- Bandwidth scaling along multiple vectors:
  - Wavelengths: 16- $\lambda$  in development
  - Modulation rate
    - 32G $\rightarrow$ 64G $\rightarrow$ 128G compute
    - $\rightarrow$ 112G $\rightarrow$ 224G networking
    - Demonstrated 128GBaud MRR-modulator
  - # fibers (more bandwidth, more radix)
    - Enabled by compact connector
- Architecture also supports 224G/lane Ethernet CPO
  - High-speed linear interface
  - Ethernet PMD like DR8 or FR4 at 200G/lane



# Summary

- Demonstrated first fully functional OCI link between two CPUs
- SiPh drives power, bandwidth density and cost scaling
- SiPh is key enabler of co-packaged optics



# Thank You!

[www.intel.com/siliconphotonics](http://www.intel.com/siliconphotonics)

## Notices & Disclaimers

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit [www.intel.com/benchmarks](http://www.intel.com/benchmarks).

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure. Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.