

Building Reliable at Scale AI Clusters with Co-packaged Optics

OCP Educational Webinar Series

June 10, 2025



Community-driven hyperscale innovation for all

Building Reliable at Scale AI Clusters with CPO

Cliff Grossner, Ph.D.

Chief Innovation Officer
OCP Foundation



Bob Wheeler

Analyst at Large
LightCounting



Manish Mehta

VP Marketing & Operations
Broadcom



Andrew Qu

Chief Technology Officer
Micas Networks



Agenda Overview

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

Field Observed Advantages with CPO for Scale Up

Sponsor Products

Conclusion

Live Q&A

Agenda

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

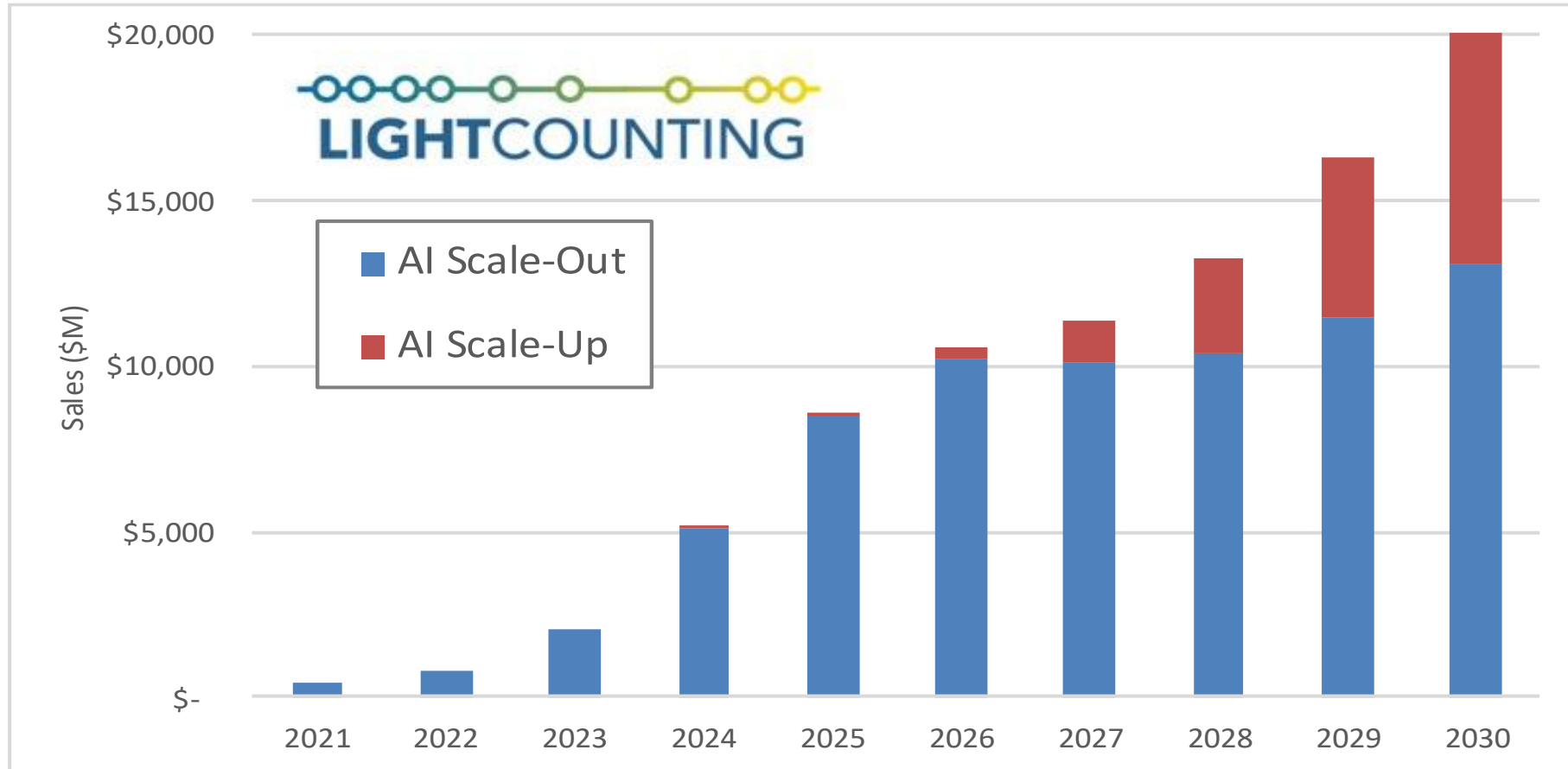
Field Observed Advantages with CPO for Scale Up

Sponsor Products

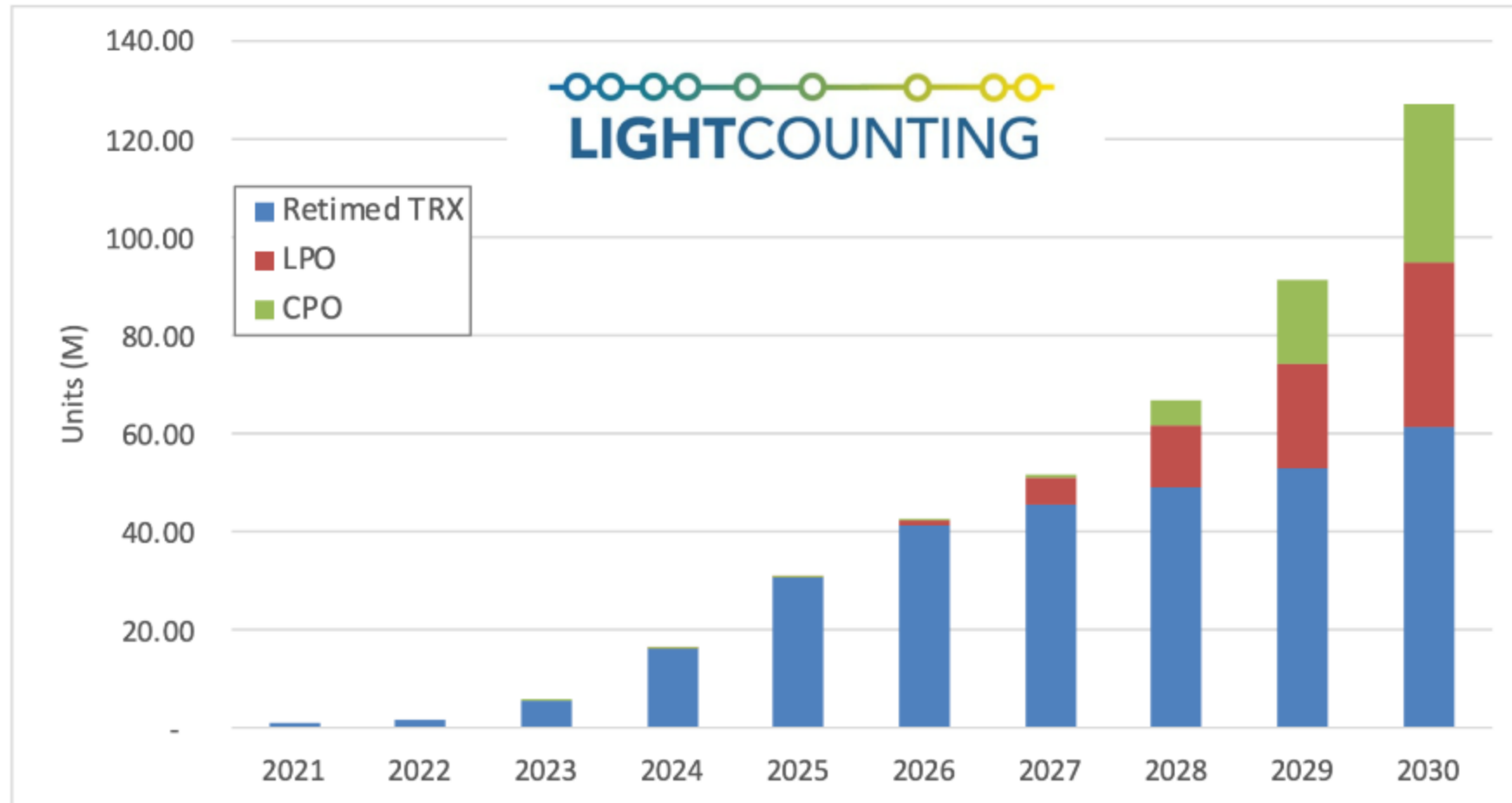
Conclusion

Live Q&A

Scale-Up Grows as Scale-Out Slows



800G LPO Stop-gap Delays CPO Adoption



1.6T DSPs Require Sub-5nm Process

Vendor	Broadcom	Marvell	MaxLinear
Product	Sian3 (BCM83628)	Ara (MV-CD833)	Rushmore
Announced	March 2025	December 2024	March 2025
Host lanes	8 x 224Gbps	8 x 224Gbps	8 x 224Gbps
Optical lanes	8 x 224Gbps	8 x 224Gbps	8 x 224Gbps
Integration	Drivers	Drivers	Drivers
Process	TSMC 3nm CMOS	TSMC 3nm CMOS	Samsung 4nm CMOS
Power claims	23W	24W	25W
Availability	Sampling	Sampling	Samples May 2025

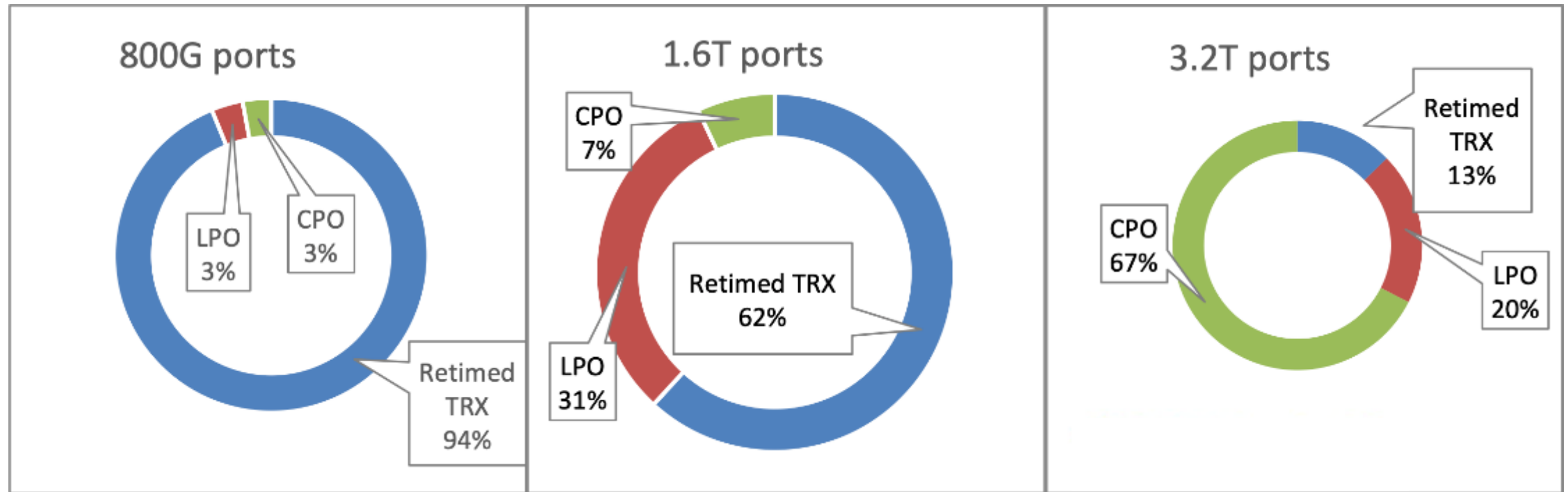
The Road Ahead

- 200G/lane LPO is unproven
 - Likely requires LRO and/or co-packaged copper (CPC)
- 400G/lane pluggable optics face high power dissipation
 - First 3.2T modules will gearbox 16x200G to 8x400G lanes
- ELSFP reduces CPO reliability concerns
 - Lasers are the least reliable component in optics
- Liquid cooling eases thermals and improves efficiency
 - Switch-chip densities approaching the limit of air cooling



CPO is Inevitable

- Meaningful penetration at 200G/lane
- Dominant technology at 400G/lane
- Growing investment in CPO for scale-up interconnects beyond 3.2T



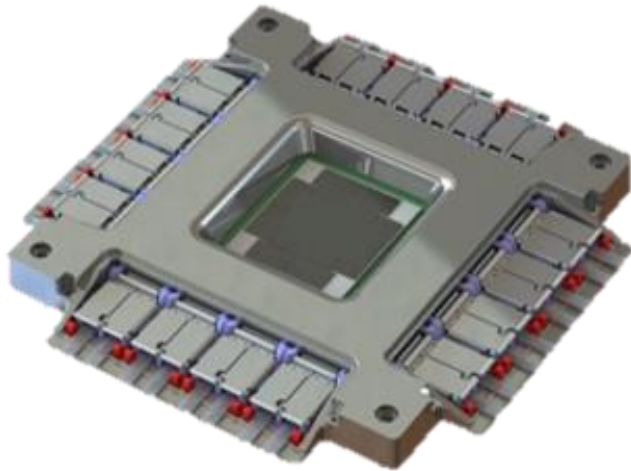
2029 Port Shipments by Type

Current State of CPO Technology

Broadcom's 3rd Generation CPO

102.4Tbps TH6-Davisson

- Tomahawk 6 Ethernet switch ASIC
- Davisson 200G/lane optical engine
- 16x6.4T OEs for full CPO
- Available: TBA

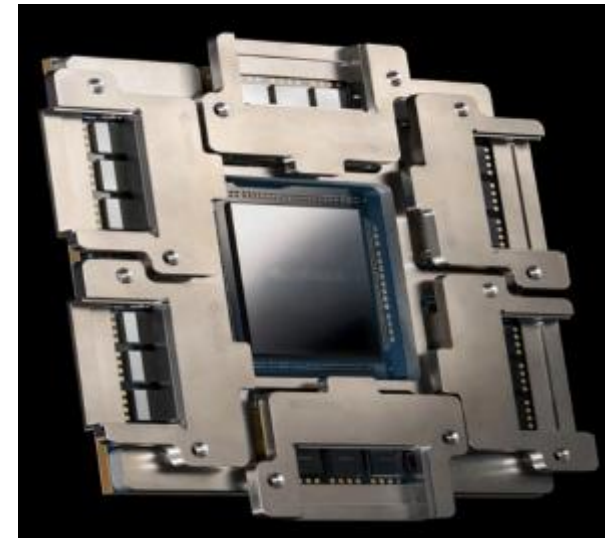


Source: Broadcom

Nvidia's 1st Generation CPO

115.2Tbps Quantum-X Photonics

- 4x Quantum-3 InfiniBand switch ASICs
- 200G/lane optical engine
- 72x1.6T OEs for full CPO
- Available: 2H25



Source: Nvidia



Agenda

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

Field Observed Advantages with CPO for Scale Up

Sponsor Products

Conclusion

Live Q&A

Current CPO Focus: Reliability and Serviceability

- General CPO Progress
 - Volumes increasing
 - Supply chains maturing
 - Encouraging early results for reliability
- Next steps
 - More extensive system-level Reliability and Serviceability validation
 - Need an industry level methodology for CPO reliability with broader collaboration for acceptance and standardization
- Reliability
 - Example methodology on next page
- Serviceability
 - Encouraging results with Pluggable Laser Sources



CPO Reliability – Baseline Methodology

Test	Component	Comments	Status
1	Laser	100 x 10 ⁹ device hours in the field. Failure rate <0.1 FIT.	Complete
2	Pluggable Laser Source (PLS)	All qualification tests are complete and have passed.	Complete
3	Photonic IC (PIC)	All qualification tests are complete and have passed. Extended reliability tests continue.	Complete
4	Optical Connector	All qualification tests are complete and have passed.	Complete
	Unbiased Damp Heat	• 85C, 85%RH, 1000hr, 48 systems	Complete
	Temperature Cycling	• -40C to 85C, 500cy, 48 systems	Complete
	Mechanical Shock	• 12G, 11ms half-sine, 10 shocks/axis, 3 axes, 48 systems	Complete
	Mechanical Vibration	• Sine; 0.1G, 5-100-5Hz, 0.1 oct/min, 90min/axis, 3 axes, 48 systems	Complete
	Dust	• A2 dust, 2kg/m ³ , 6 second bursts, 20cy, 16 systems	Complete
5	Electrical Interconnect	Completion is complete	Complete
	HTOL Rack Systems	• 11 systems running to 100,000 stress hours by 12/31	On-going
	HTOL Mezz Card Systems	• 19 systems running to 100,000 stress hours by 12/31	On-going

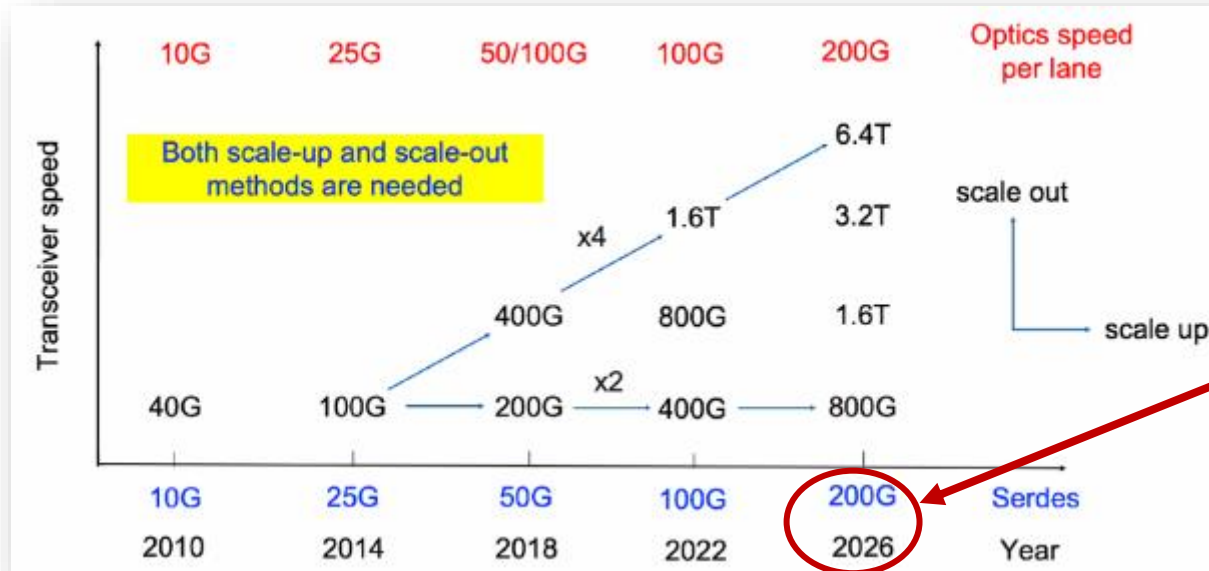
Challenges in Building High Performance Switches

High-speed SerDes challenges: This year, 102.4T ASIC 112G/224G SerDes power higher, distance shorter, future higher count of 224G ASIC or higher rate SerDes, like 448G, will post more challenges in conventional PCB routing approach, given much higher cost of materials

Module Power and Cost: With higher throughput next-gen ASICs, the high count of 800G/1.6Tbps optic module systems cost will be high as well as total power contributed by module, estimated 102T system may consume 5000W+ mostly consumed by ports

Port Density Limitation: Port density constrained by space and external module cooling.

High speed port stability: Pluggable transceiver unstable becomes obvious in large scale network



By 2026 200Gbps/lane
will post significant challenges in
making conventional switches

Agenda

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

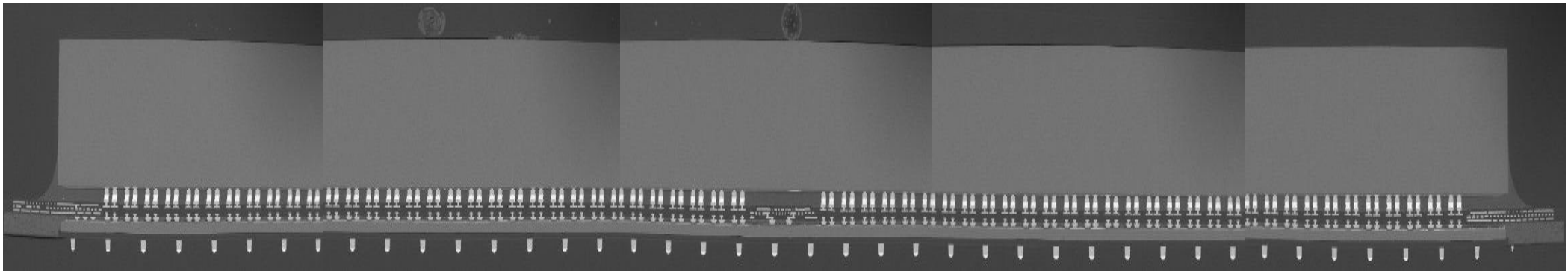
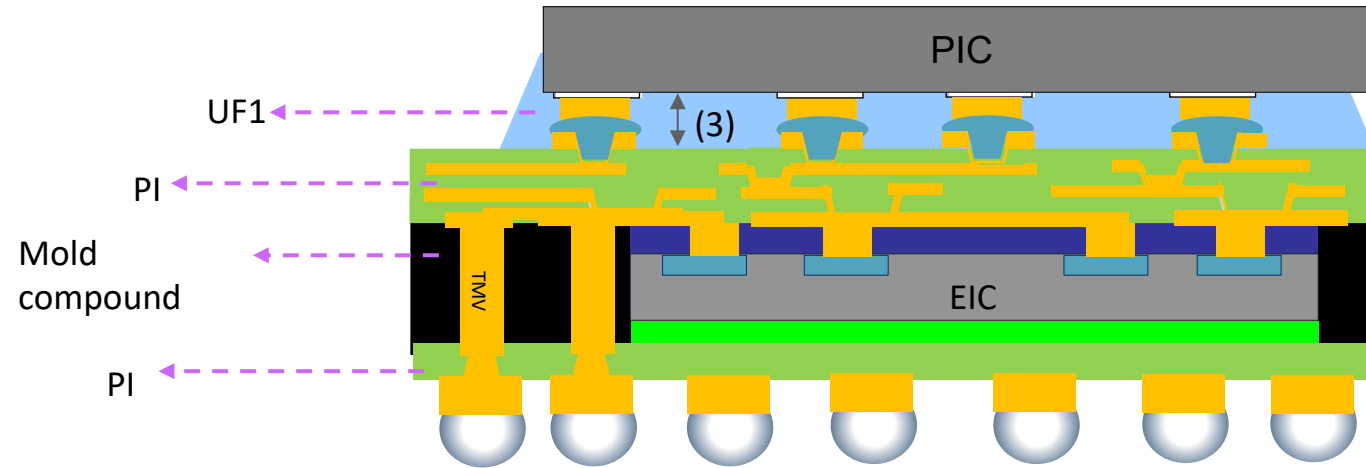
Field Observed Advantages with CPO for Scale Up

Sponsor Products

Conclusion

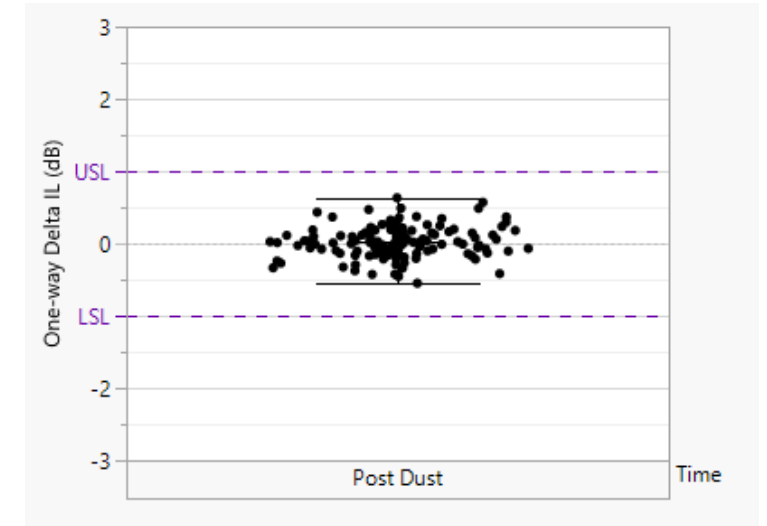
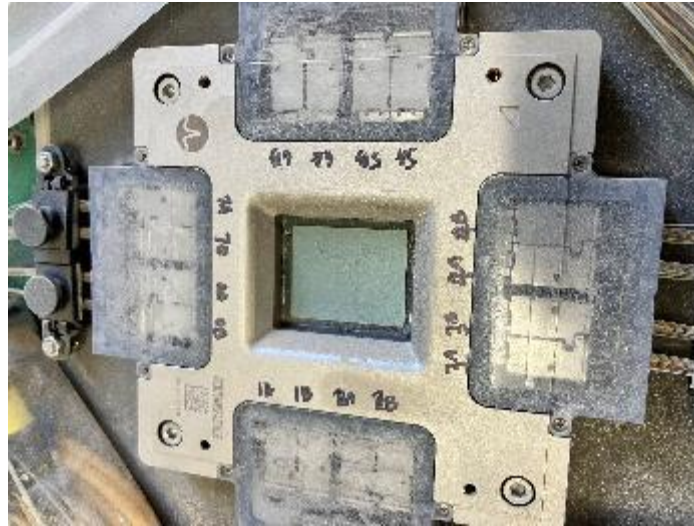
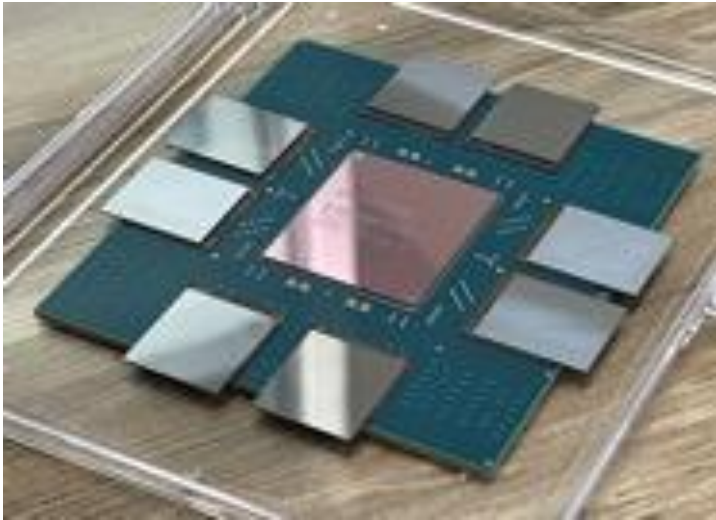
Live Q&A

Construction of a 3D Optical Engine



RDL Based High Density Interconnect one option (used on TH5-Bailly)

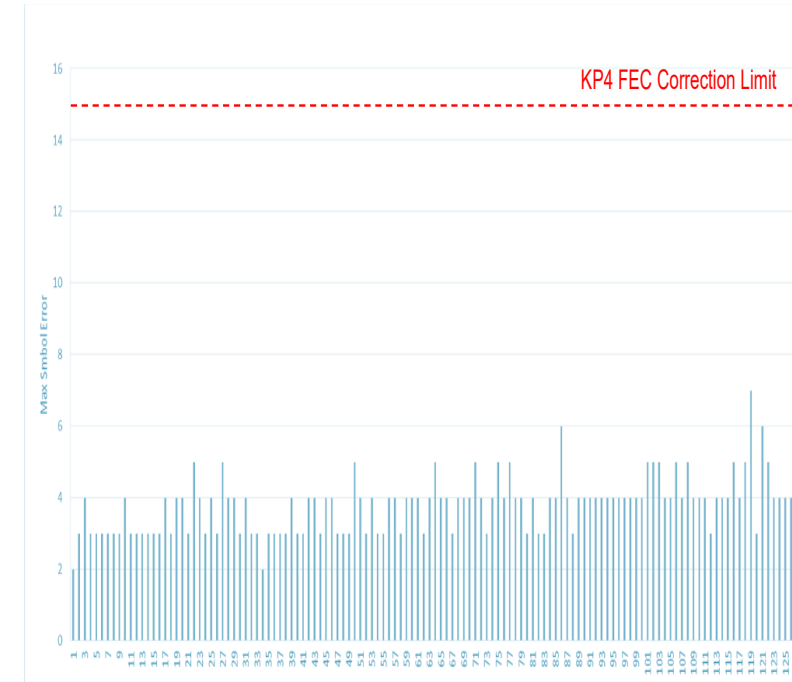
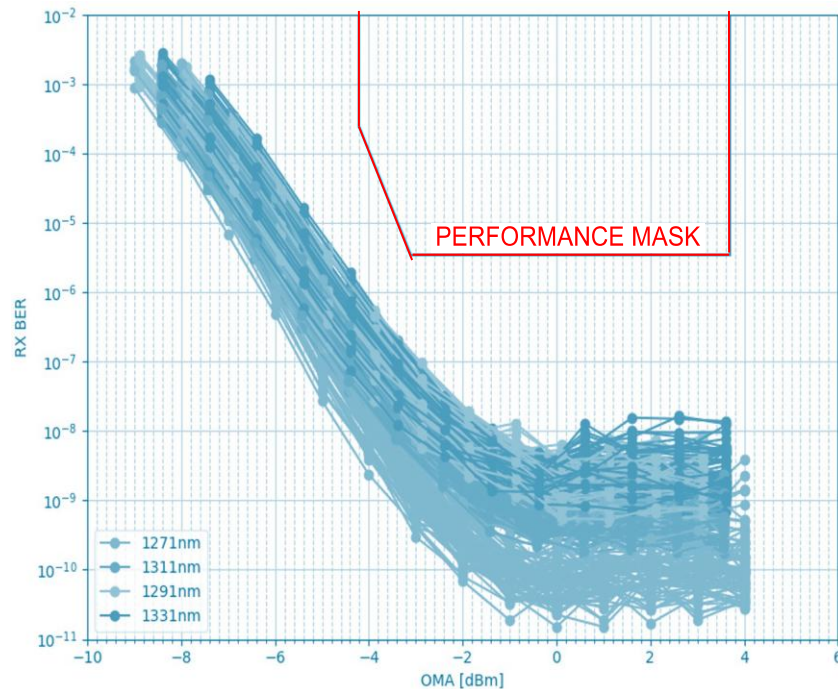
From Engines-on-Substrate to Performant Devices



- Some areas that required iteration to improve reliability:
 - Dust protection (ports to future generations)
 - Detachable fiber insertion mechanical structure (ports to future generations)
 - LGA socket design (future generations likely to migrate to BGA solder attach)

CPO - Link Performance Validation

- Enabling traffic testing during CPO production test
- Test for both Pre-FEC BER and Post-FEC T-count statistics
- Known TP1 and TP4 links at manufacturing

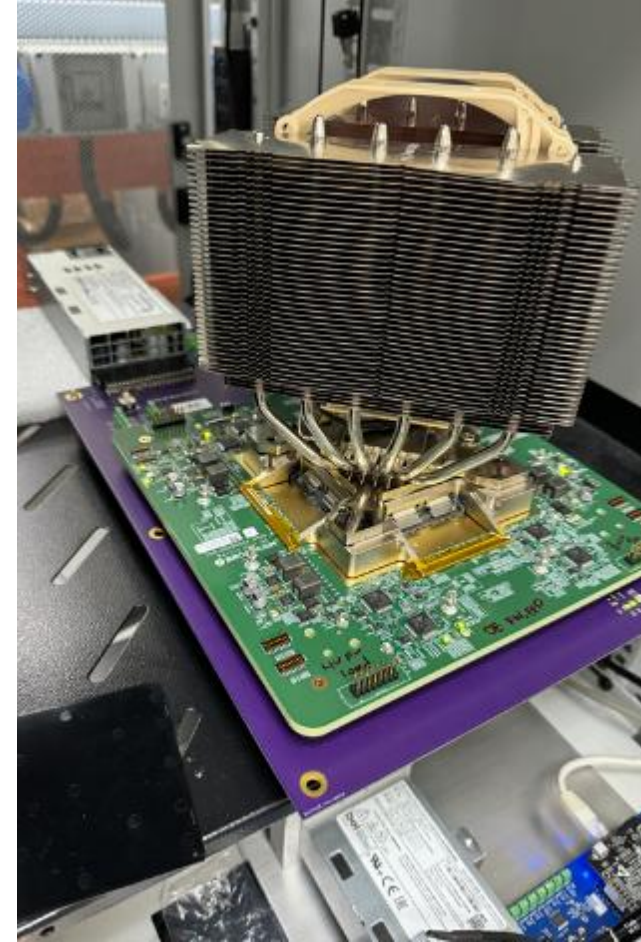


CPO Reliability – Rack Level Infrastructure



System Racks

- Full system traffic testing
- Limited temperature acceleration
- Stresses all parts of system

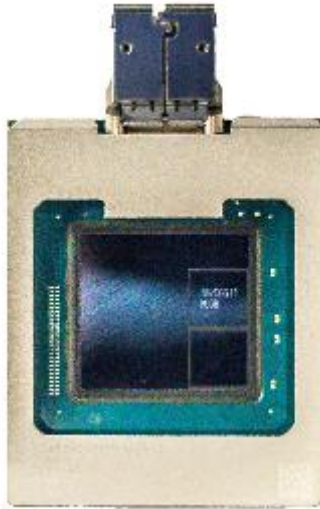


Mezz Card

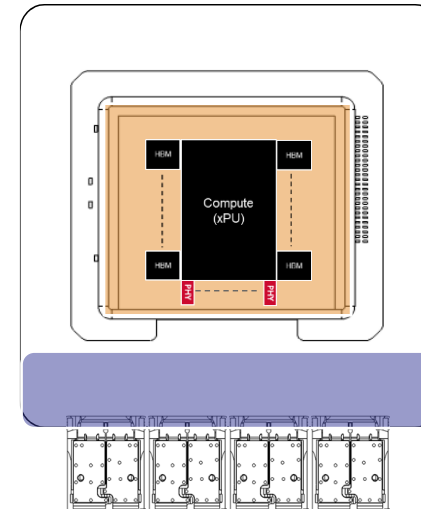
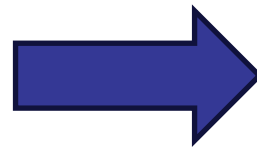
- CPO testing only
- Higher temperature operation
- Increased acceleration factor

Scale-Up Diversity

- Industry rapidly evolving to three protocols for scale-up: Ethernet, UALink, Nvlink
- CPO based Optical interconnects provide an alternative to mega-rack architectures and enable much larger scale-up architectures
- Will be generally agnostic to protocol

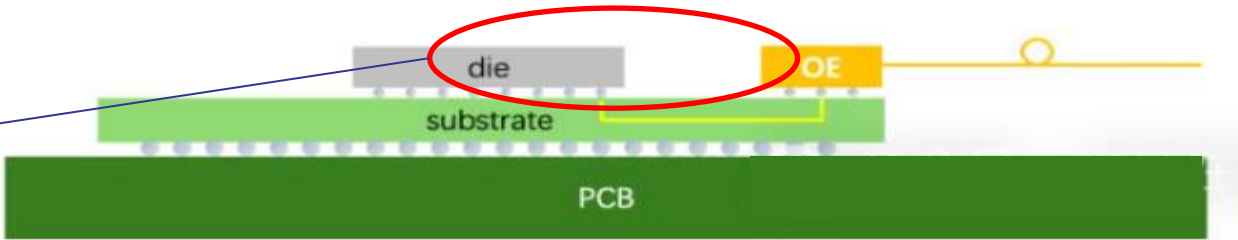
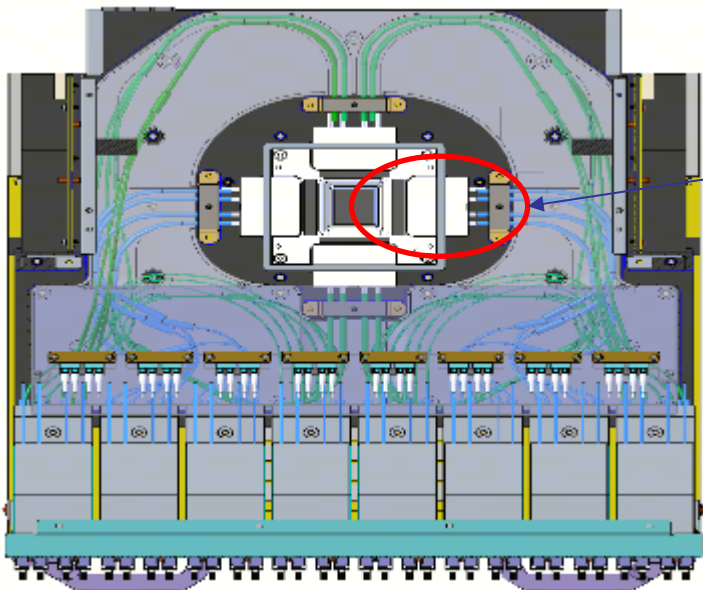


Prototype xPU + CPO



w/ ability to scale to high package BW density

CPO Enables Future Higher Bandwidth 224/448G per Lane Technology



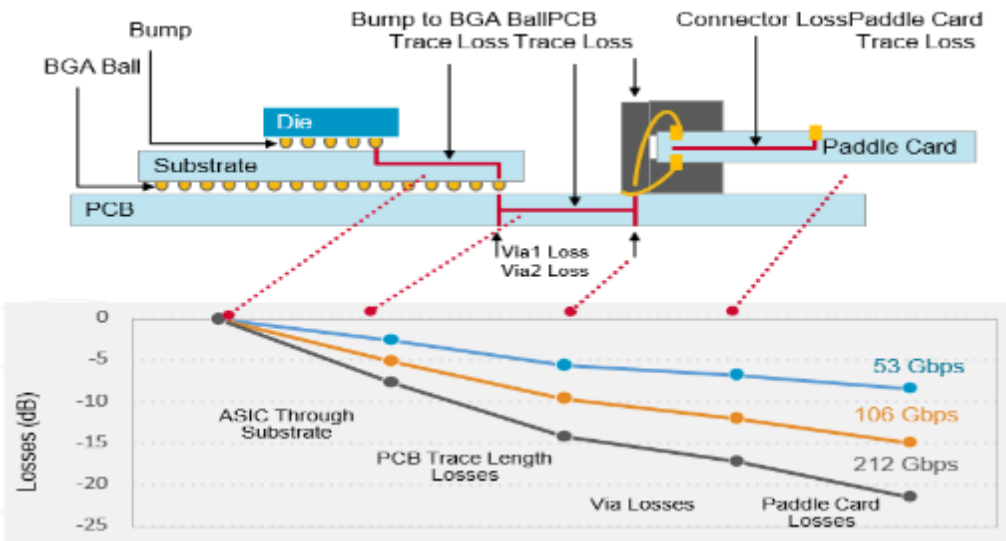
Greatly reduced distance meet high speed SerDes requirements without new component like CPC



TH5 CPO system PCB

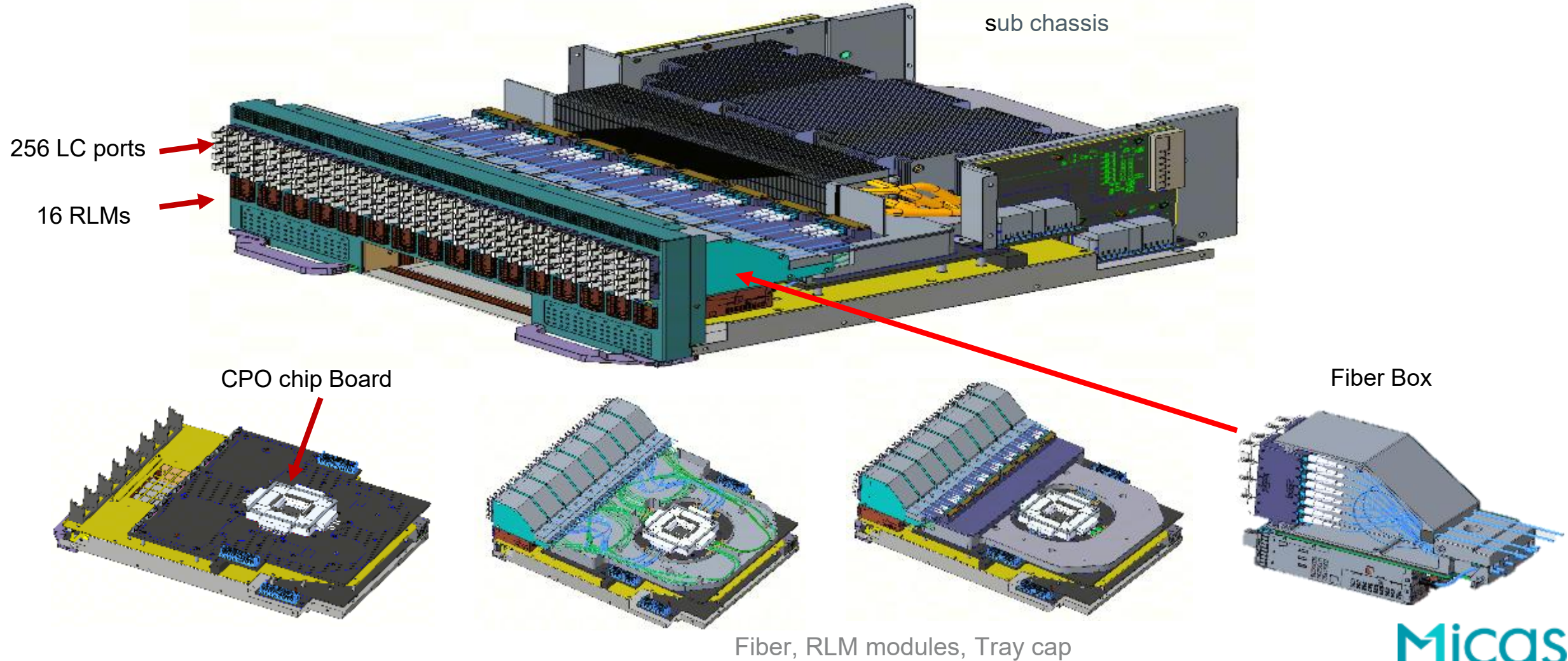


TH5 system PCB



Signal loss chart for regular system

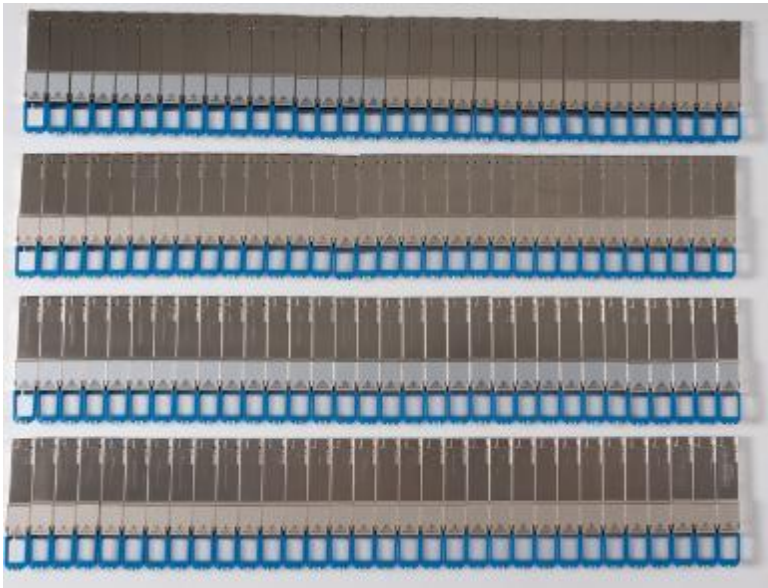
CPO Requires New Chassis Layouts



CPO Drives Lower System Cost

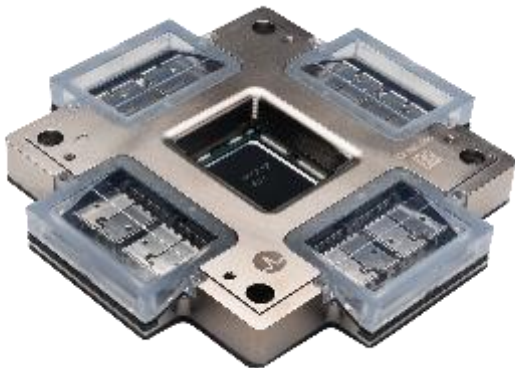


PCB board cost less, less high-quality connectors.



128x 400G Pluggable Transceivers

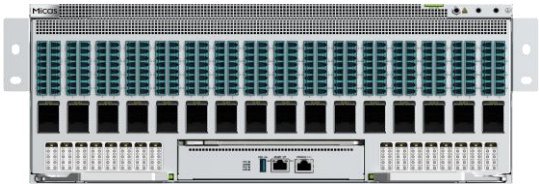
VS



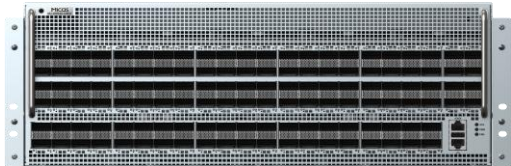
Number of transceivers Is large,
Manufacture process risk enlarged

128x ports 400G (Bailly CPO)

128x400G CPO

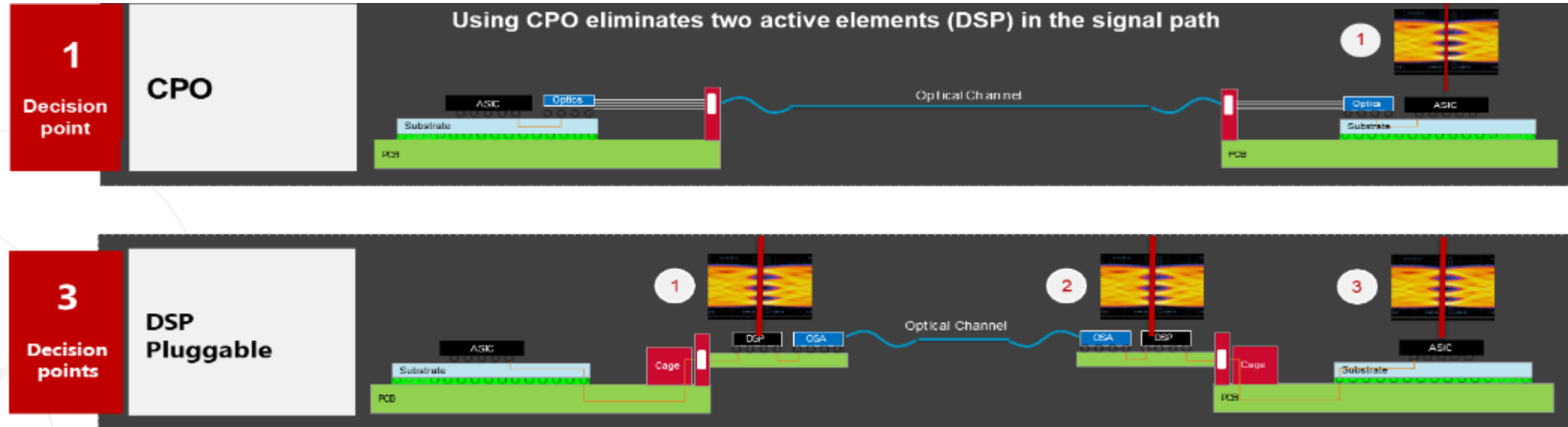


128x400G Pluggable



Improved Port Density, Latency and Resiliency

- Enable 1RU high port density with LC ports for high performance and high redix count
- Latency is lower (very short trace, a smaller number of DSPs)
- Port should be more stable and last longer
(Laser source far away from heat source with low FIT field use smaller number of connectors)



Audience Poll Question

When do you believe CPO will be ready for data-center-scale deployments?

1. It's ready now.
2. Within 2 years.
3. In 2 - 5 years.
4. Next decade (or never).

Audience Poll Results

- Some discussion facilitated by host and analyst of the results with optional comments from sponsors

Agenda

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

Field Observed Advantages with CPO for Scale Up

Sponsor Products

Conclusion

Live Q&A

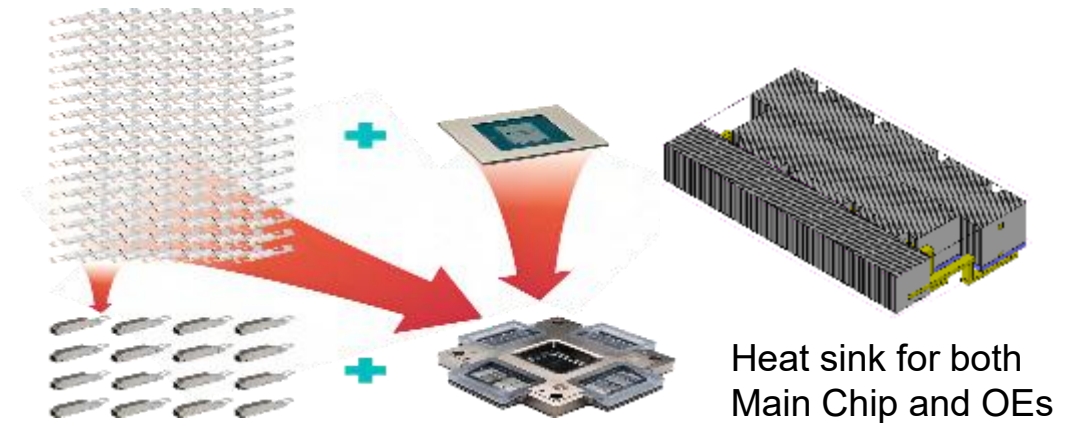
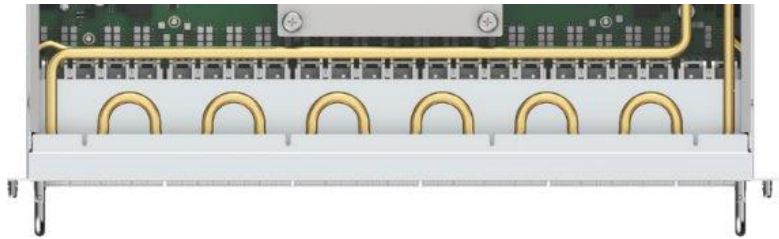
Significant Power Saving Measured from Real Systems

OE save 70% optical interconnect power,

Easy thermal requirements, FAN power saving

Cooling OEs is more efficient than individual cages

Advanced cooling associated cost, complexity in failures



Micas 64 Port TH5



2xFR4 **DSP** Based
Transceiver (**20W per 800G**)

2780 W

Micas 64 Port TH5



DR8 **LPO** Transceiver
(**8W per 800G**)

2020 W
Estimated

BAILLY

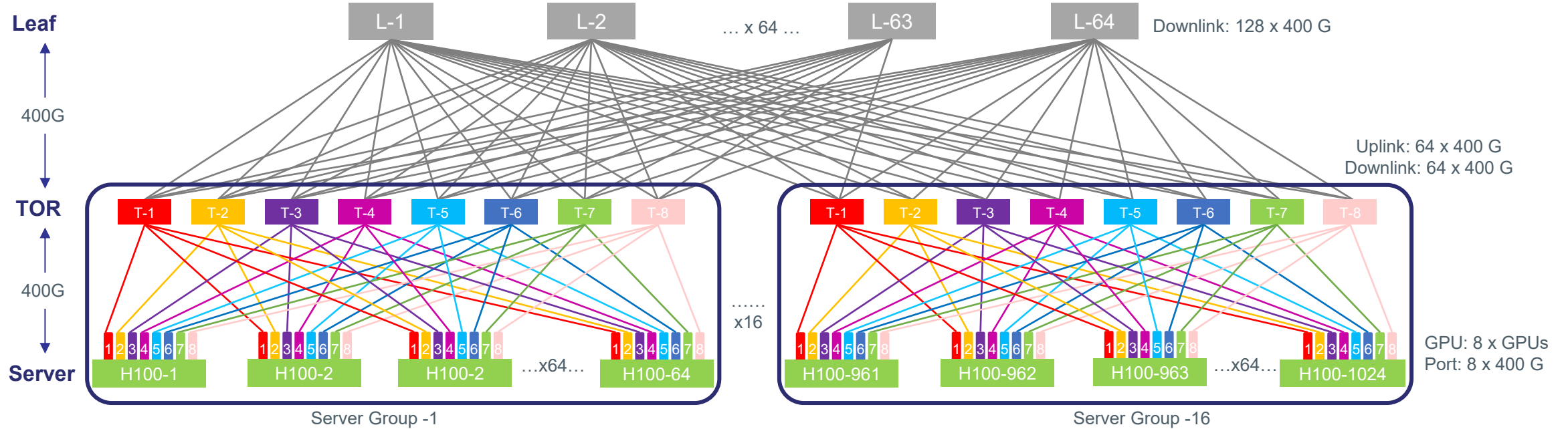


CPO (**5.5W per 800G**)

1650 W

Sample Deployment of AI/ML Network

Small to Medium Scale ($\leq 8,192$ GPUs, 1024 Server)



- This is 192 128x400G system network
- Power consumption saving ($1.4 \times 192 \times 24 \times 365 = 2,296,627$ kWh)
- Lower latency for AI/ML payload
- Much lower TCO when all ports fully populated with optics
- But more stable network

Agenda

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

Field Observed Advantages with CPO for Scale Up

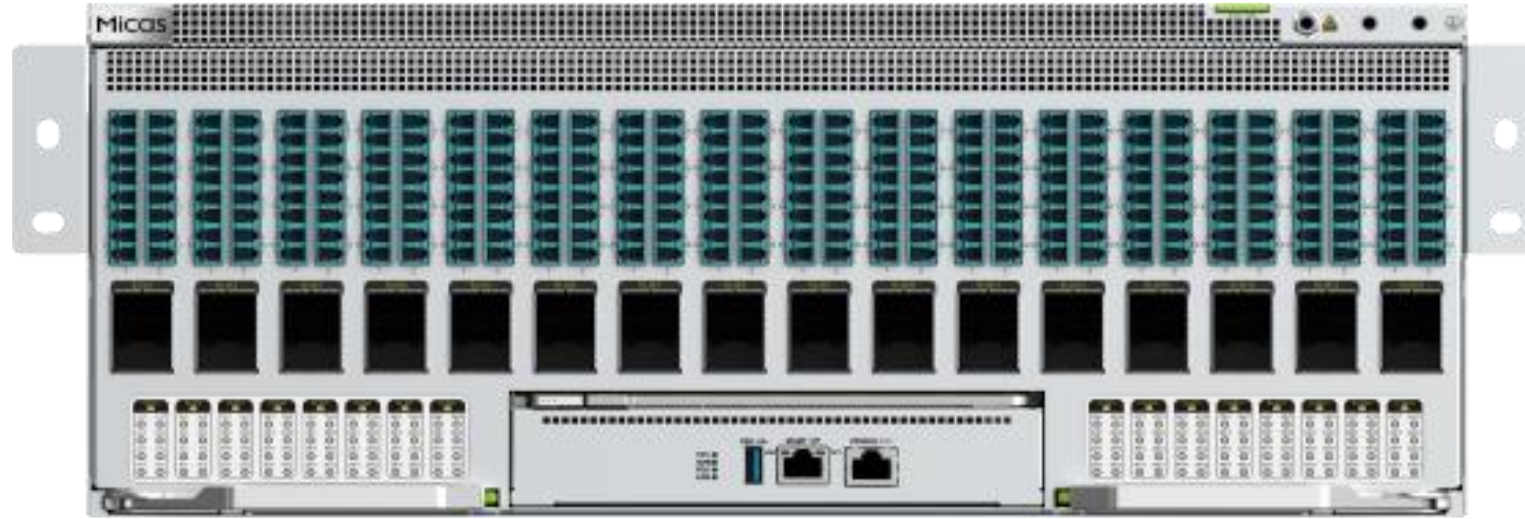
Sponsor Products

Conclusion

Live Q&A

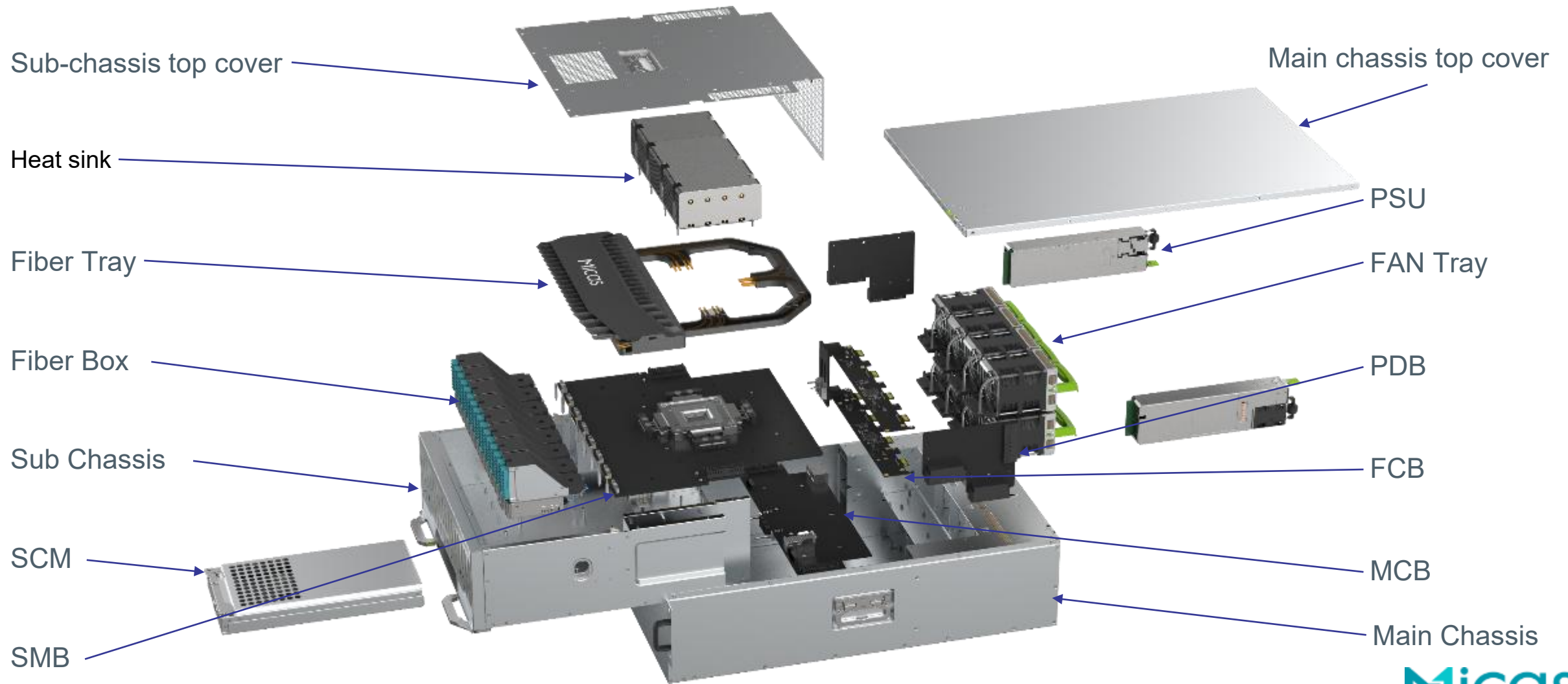
Micas Networks: 128x400G CPO System

(Broadcom Tomhawk5 + Bailly OE engine)



Micas
NETWORKS

CPO Chassis View



Agenda

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

Field Observed Advantages with CPO for Scale Up

Sponsor Products

Conclusion

Live Q&A

Take Aways

- Front-panel pluggable optics are running out of runway
- Power per bit no longer falling due to rising DSP complexity
- Tomahawk5 integrated SerDes handle direct-drive for LPO and CPO
- TH5-Bailly proves many aspects of CPO manufacturability and reliability
 - Field reliability data still lacking
- 200G/lane CPO coming over the next 12 months
 - Nvidia's QuantumX Photonics brings CPO to 800G InfiniBand
 - TH6-Davisson will enable CPO for 1.6T Ethernet
- Scale-up interconnects will ultimately drive the highest volumes



Continue Learning About CPO

Tracking Progress in Co-Packaged Optics

Virtual conference: July 15th 5pm-8pm PT

- Session 1: Established suppliers: Broadcom, Coherent, Marvell
- Session 2: CPO start-ups: Avicena, Celestial Ai, LightMatter, Nubis Communications
- Session 3: Supporting technologies: Alfalume, Lumentum, Poet, Scintil Photonics, Senko
- Closing discussion with Brad Booth: Was COBO ahead of its time?

Register Here: https://us02web.zoom.us/webinar/register/WN_IWHSZAS-Sp6-RMMH8EkCFg



Agenda

CPO is it Finally Here?

Reliability and Serviceability Challenges

System Wide CPO Validation and Deployment

Field Observed Advantages with CPO for Scale Up

Sponsor Products

Conclusion

Live Q&A

Live Q&A

Cliff Grossner, Ph.D.

Chief Innovation Officer
OCP Foundation



Bob Wheeler

Analyst at Large
LightCounting



Manish Mehta

VP Marketing & Operations
Broadcom



Andrew Qu

Chief Technology Officer
Micas Networks



Thank you for attending

Building Reliable at Scale AI Clusters with Co-packaged Optics

OCP Educational Webinar Series

June 10, 2025



Community-driven hyperscale innovation for all