Energy Efficiency in AI Applications – Making Sense of Multiple Requirements

Jeff Hutchins / Ranovus
OIF PLL WG EEI Vice Chair

Wednesday 01 October 2025

Copyright © 2025 OIF

Accelerating Market Adoption of Optical Networking Technologies

160+ Member Companies ▶ 25+ Years of Service ▶ Member Driven Global Organization

COHERENT OPTICAL

Multi-Vendor Interoperability in Client Form Factors

1600ZR+
- <1000km Multi-Span Coherent DWDM

1600LR, 800LR
- <10km Coherent Point-to-Point

1600ZR, 800ZR, 400ZR
- >80km Coherent DWDM

ELECTRO-OPTICAL

Energy Efficient Interfaces (EEI) –Low Latency/ Optimized Energy Interfaces for AI/ML
- Compute Optics Interface (COI)
- Retimed Tx, Linear Rx (RTLR)
- External Laser Sources (ELSFP)
- Co-Packaged Modules (3.2T)

Common Electrical I/O (CEI)
- High-Speed Building Blocks
- 448G, 224G, 112G, 56G, 28G
- LR, MR, VSR, XSR+, XSR, MCM, Linear
- Protocol Agnostic Link Training

OIF

- Identifies Industry Needs and Gaps
- Publishes Implementation Agreements (specifications) (100+), Requirements and White Papers
- Performs Interoperability Demonstrations (65+)
- Advances industry consensus via workshops, webinars, etc.

MANAGEMENT

Common Management Interface Specification (CMIS)
- Single Solution Ranging From Copper to Coherent
- Simplified Bring up Between Host and Module
- Supports Standard and Custom Interfaces

Transport SDN APIs
- Automation, Programmability

Enhanced Network Operations
- Artificial Intelligence
- Digital Twin
- DC Storage and Optical Multi-Layer Coordination

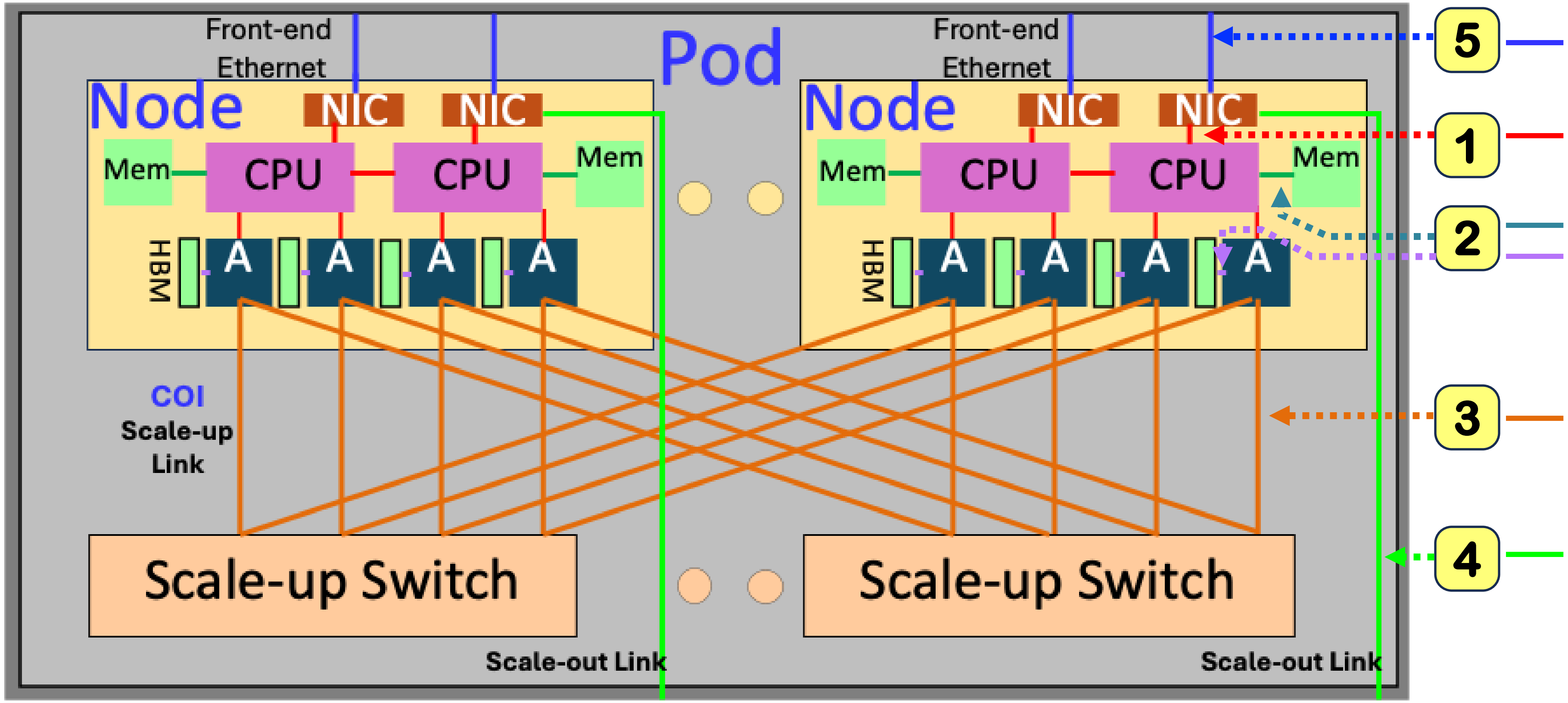
PROTOCOL

Flex Ethernet (FlexE)
- 800 Gb/s Ethernet PHY support

For more information, visit www.oiforum.com

Copyright © 2025 OIF

Links in an AI Compute Pod



Ethernet

Compute I/O (PCIe/CXL/D2D/UCIe)

Memory I/F (CXL/PCIe)

HBM I/F

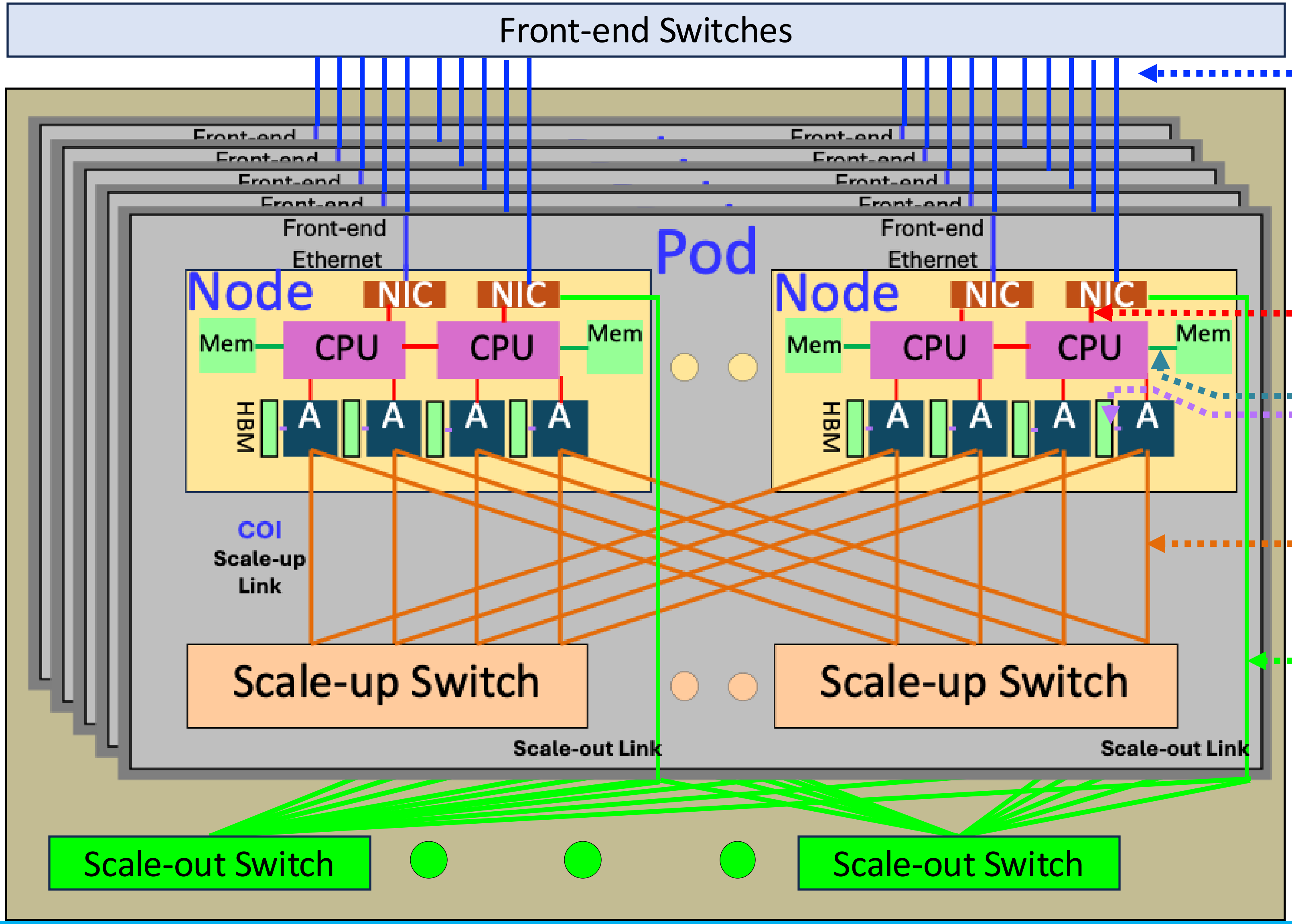
Scale-up
(UALink / UEC / NVLink / SUE, etc.)

Scale-out
(UEC / InfiniBand / Ethernet / etc.)

There are broadly 5 types of links used in an AI Compute Pod

Copyright © 2025 OIF

AI Compute aggregates Pods with Scale-out links



5 Ethernet

1 Compute I/O (PCIe/CXL/D2D/UCIe)

2 Memory I/F (CXL/PCIe)
HBM I/F

3 Scale-up (UALink / UEC / NVLink / SUE / etc.)

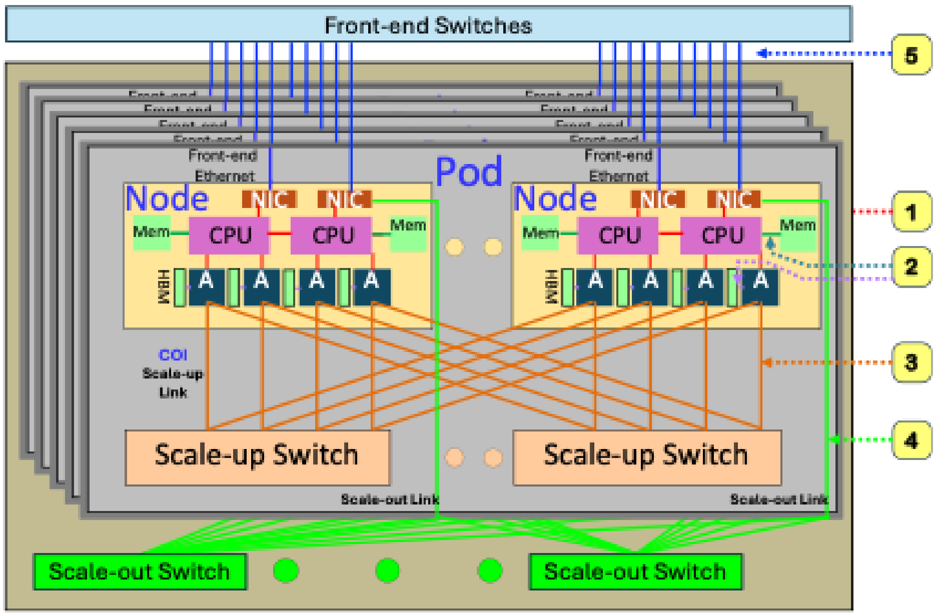
4 Scale-out (UEC / InfiniBand / Ethernet / etc.)

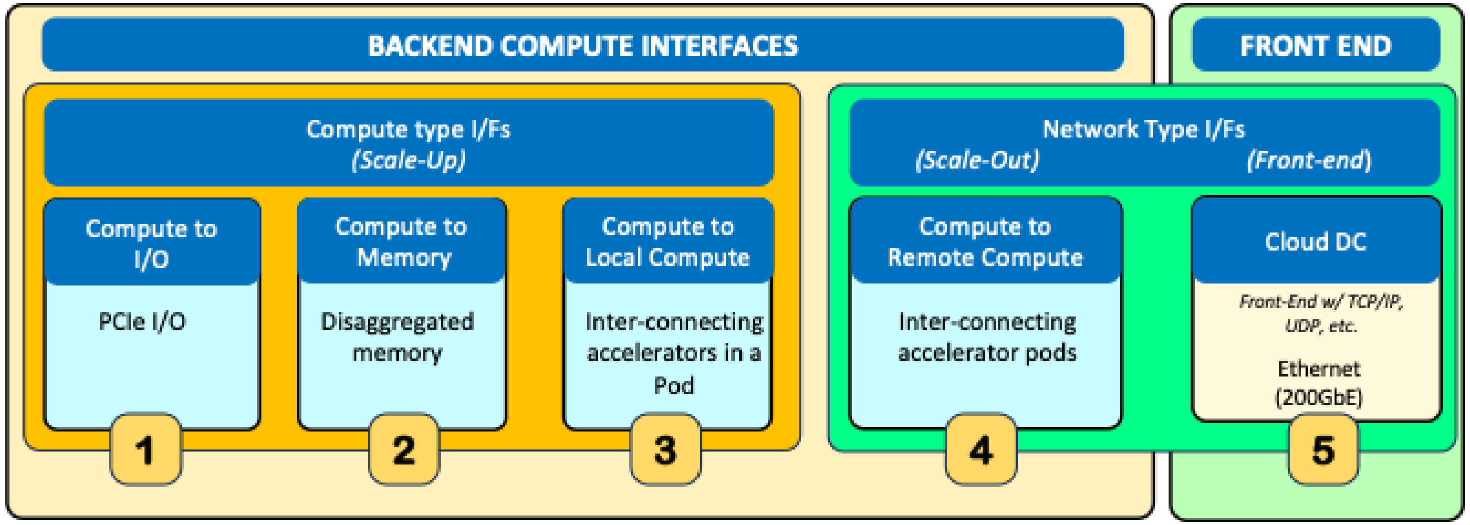
There are broadly 5 types of links used in AI Compute

Copyright © 2025 OIF

RANOVUS®

AI Compute Requirements





Parameter	3: COI: Scale-up Compute - Local	4: Scale-out Compute - Network	5: Front-end Compute Servers
Scale	Local	10's of racks	Data center
Reach [m]	~ 20	~ 100	~ 1000
ASIC BW Edge Escape Density [Tbs/mm] *(Tx + Rx)/(edge width)*	> 2.0	> 2.0	Std Ethernet
Transceiver Latency (Tx + Rx) [ns]	latency-A: < 5 latency-B: < 100	latency-A: < 60 latency-B: <100	< 100
Energy Utilization [pJ/b] *(Tx + Rx with lasers)*	EU-A: < 4 EU-B: < 12	EU-A: < 10 EU-B: < 12	< 12
Reliability (link errors)	very high	high	same as std Ethernet
Reliability (hw failures)	inside: very high front panel: high	inside: very high front panel: high	same as std Ethernet

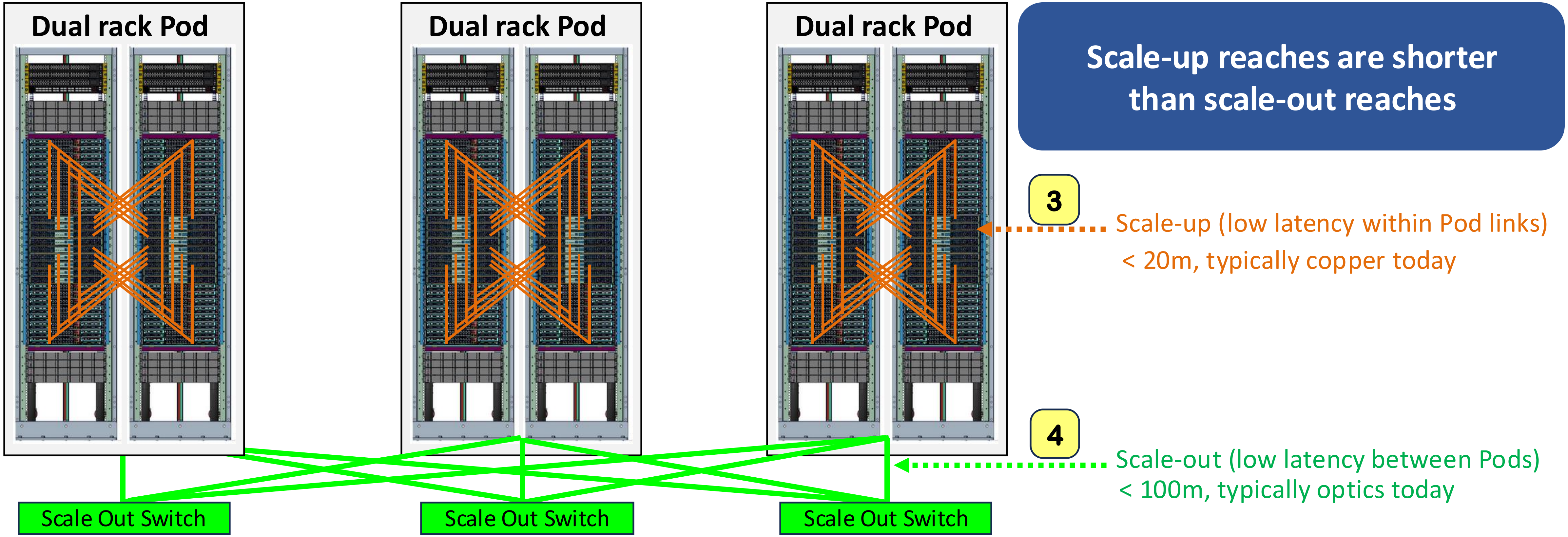
Key requirements from hyperscalers are summarized in the table

We will focus on the requirements for link types 3 & 4

Copyright © 2025 OIF

What reaches are needed?

Parameter	3: COI: Scale-up Compute - Local	4: Scale-out Compute - Network	5: Front-end Compute Servers
Scale	Local	10's of racks	Data center
Reach [m]	~ 20	~ 100	~ 1000



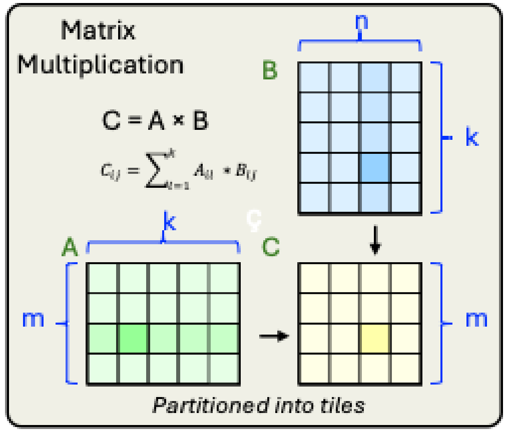
Scale-up reaches are shorter than scale-out reaches

Copyright © 2025 OIF

Latency

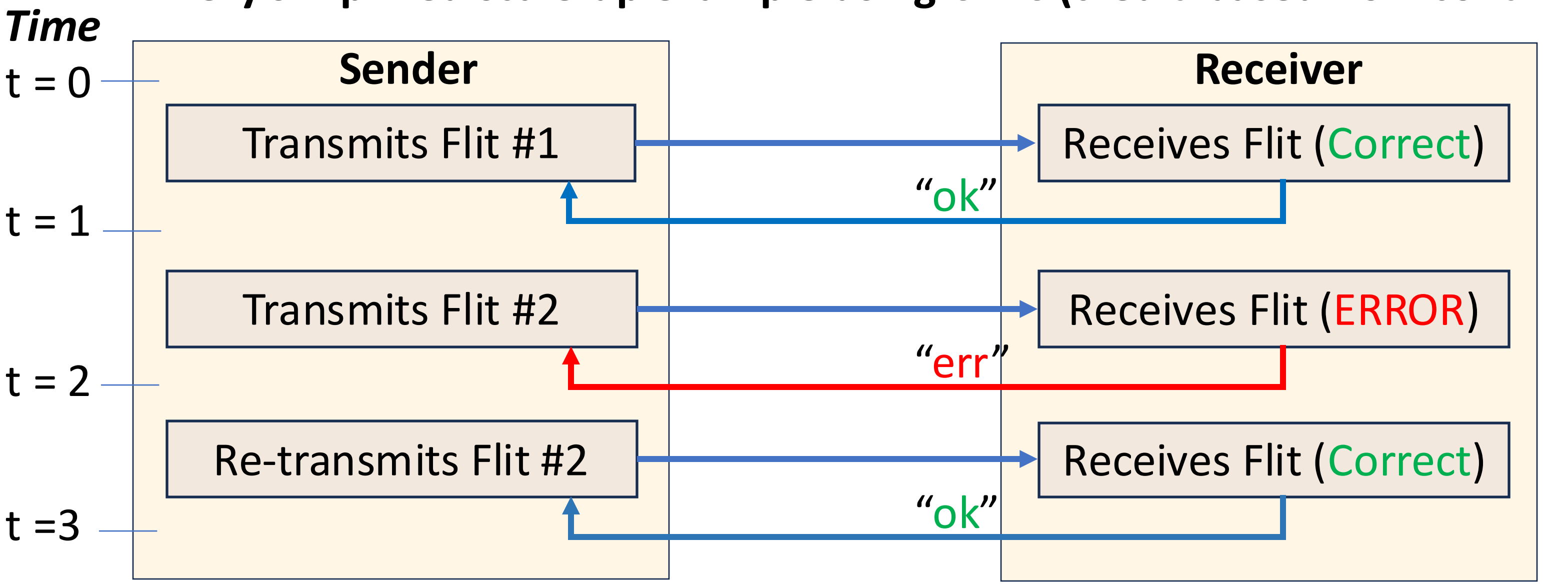
RANOVUS

Parameter	3: COI: Scale-up Compute - Local	4: Scale-out Compute - Network	5: Front-end Compute Servers
Transceiver Latency (Tx + Rx) [ns]	latency-A: < 5 latency-B: < 100	latency-A: < 60 latency-B: <100	< 100



Job partitioned into tiles assigned to accelerators (exchanging large blocks of data)

Very simplified scale-up example using CBFC (credit-based flow control)



Hypothetical example:

2 Flits received in the time for 3 transmissions

Effectively ~ 66% of maximum transmission rate due to received Flits with uncorrectable errors

Minimizing tail latency coupled with low BER improves effective link bandwidth

Copyright © 2025 OIF

Reliability

RANOVUS

Parameter		3: COI: Scale-up Compute - Local	4: Scale-out Compute - Network	5: Front-end Compute Servers
Reliability	(link errors)	very high	high	same as std Ethernet
Reliability	(hw failures)	inside: very high front panel: high	inside: very high front panel: high	same as std Ethernet

Link Reliability

- AI architectures requires higher reliability
  As failures (BER) can reduce the effective link bandwidth reducing compute performance

Hardware Reliability

- AI architectures requires higher reliability
  As work may be paused while awaiting repair
- Transceivers located behind the front panel are more difficult to replace and therefore need even higher reliability

Replacing the longer copper links with optics enables larger Pods
But requires improved reliability over the typical optical transceiver

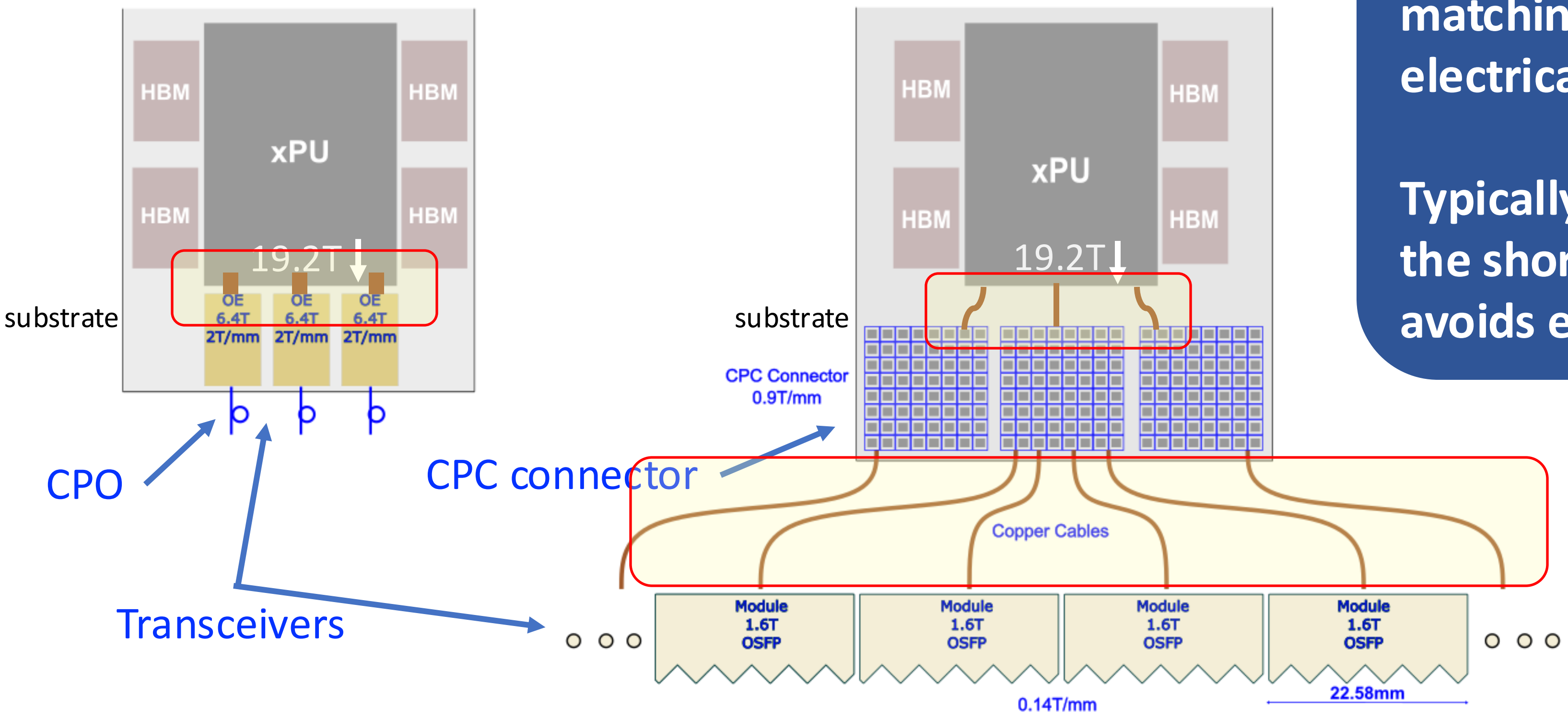
Copyright © 2025 OIF

ASIC Escape Bandwidth Edge Density

Parameter	3: COI: Scale-up Compute - Local	4: Scale-out Compute - Network	5: Front-end Compute Servers
ASIC BW Edge Escape Density [Tbs/mm] *(Tx + Rx)/(edge width)*	> 2.0	> 2.0	Std Ethernet

~ 2 T/mm with CPO (co-packaged optics)

~ 0.9 T/mm with CPC (co-packaged copper)



The "edge density" requirement comes from matching the transceiver density to the electrical interface density at the ASIC

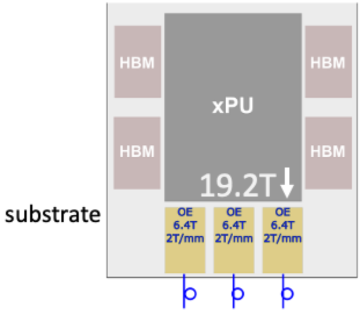
Typically, this results in lower power due to the shortened electrical channels which avoids electrical fanout

Copyright © 2025 OIF

Energy Utilization

Parameter	3: COI: Scale-up Compute - Local	4: Scale-out Compute - Network	5: Front-end Compute Servers
Energy Utilization [pJ/b] *(Tx + Rx with lasers)*	EU-A: < 4 EU-B: < 12	EU-A: < 10 EU-B: < 12	< 12

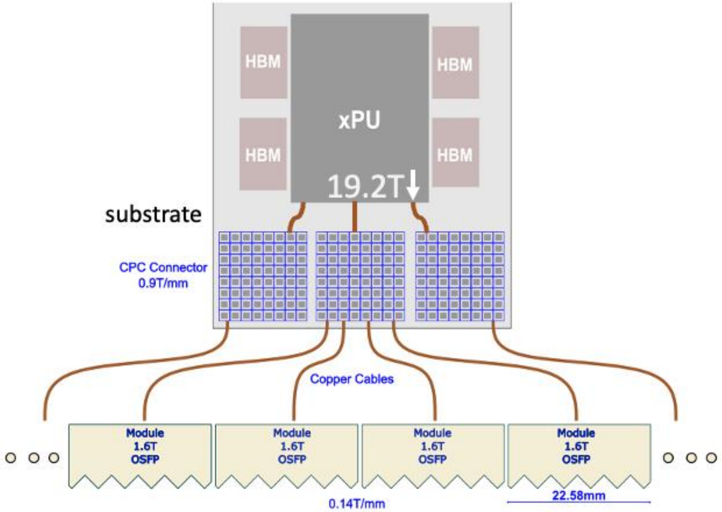
A



~ 2 T/mm with CPO
(co-packaged optics)

Very short electrical channel
Use non-retimed (linear) interface (~4-6 pJ/b)

B



~ 0.9 T/mm with CPC
(co-packaged copper)

Longer electrical channel
Non-retimed (linear) with a good SerDes and channel
Or Tx retimed (RTLR) or fully retimed (< 12 pJ/b)

Copyright © 2025 OIF

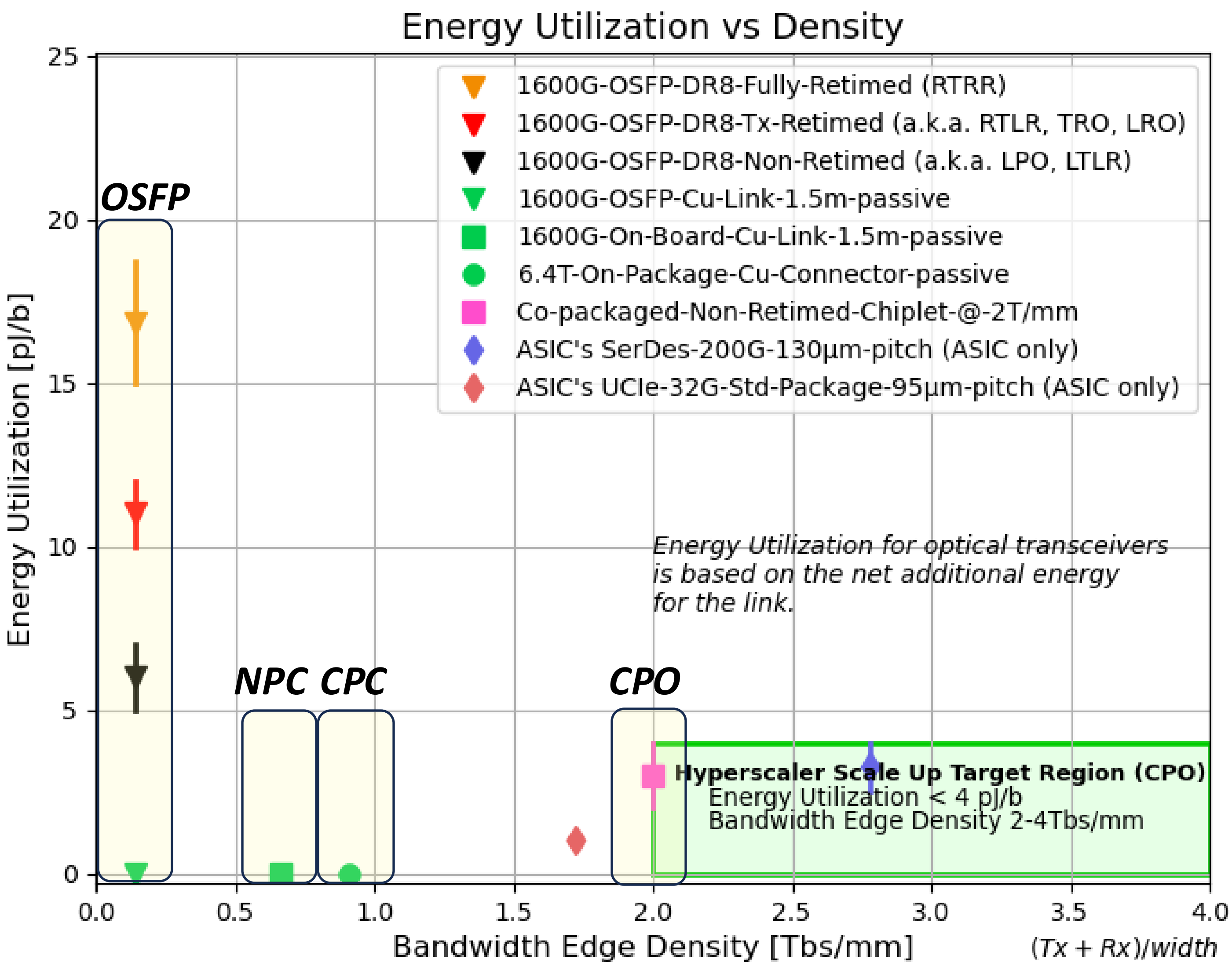
Energy Utilization and Edge Density

RANOVUS®

Visualizing the Tradeoffs

Mapping technologies against the requirements:

- Energy Utilization
- ASIC Escape Bandwidth Edge Density



Copyright © 2025 OIF

Summary

The next generation of AI Compute will need larger Pods

- Optical links can enable these larger Pods

The OIF is working with its members from across the eco-system to find the best solutions among the various trade-offs while balancing scalability and manufacturability:

- CPO vs CPC
- Minimizing latency
- Lower energy utilization
- Improved hardware and link reliability

Come see the OIF's Energy Efficient Interfaces (EEI) demos at C3425

Copyright © 2025 OIF

Thank-you

Copyright © 2025 OIF