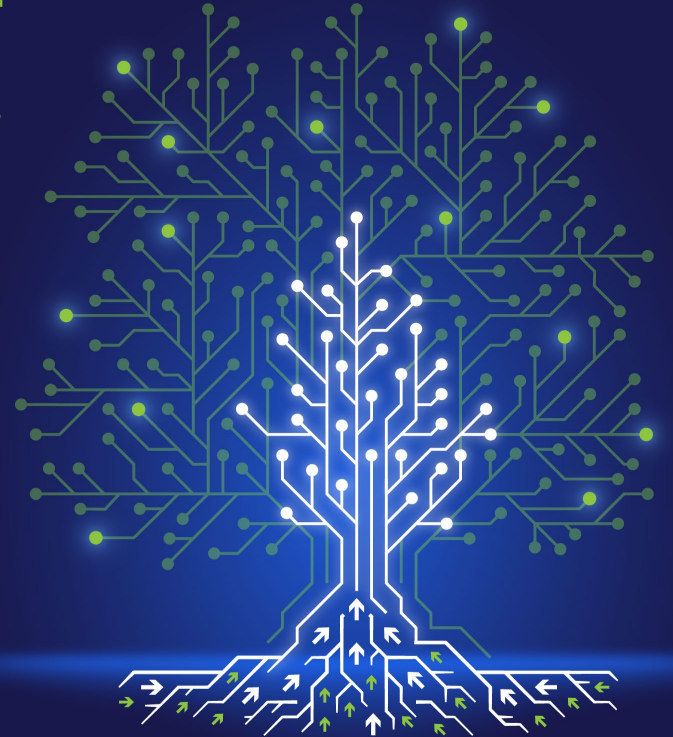


Coherent Optics for Artificial Intelligence

Generative artificial intelligence (AI) models are transforming several industries, and data centers are no exception. AI models are computationally heavy, and their increasing complexity will require faster and more efficient interconnections than ever between GPUs, nodes, server racks, and data center campuses. These interconnects will have a major impact on the ability of data center architectures to scale and handle the demands of AI models sustainably.

In this presentation, EFFECT Photonics will discuss how coherent optics and photonics technologies can improve the cost and power per bit of these data center interconnects and what role they will play at different interconnect scales (intra-node, intra-rack, intra-campus). This talk will also address the advantages of combining optical technologies with digital signal processing to improve these coherent optical interconnects.



Scaling Innovation Through Collaboration



OCP
REGIONAL
SUMMIT

24-25 APRIL 2024
LISBON, PORTUGAL



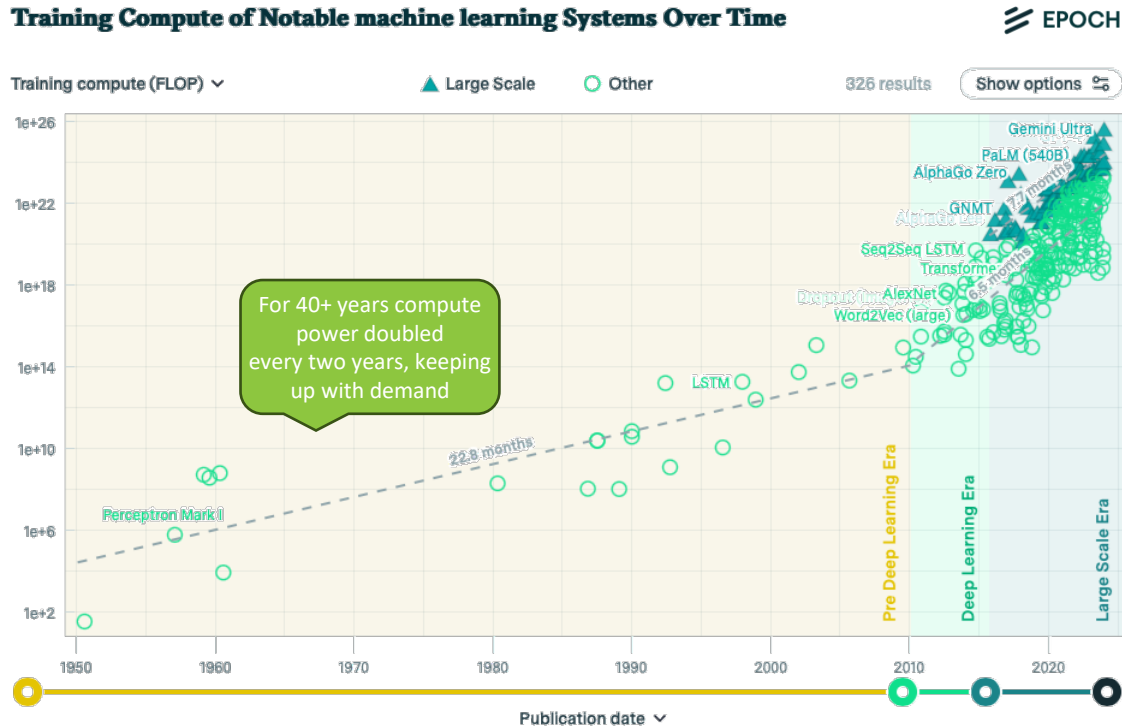
OPTICS

Coherent Optics for Artificial Intelligence

Joost Verberk, Vice President of Product Management, EFFECT Photonics



Moore's law is rapidly running out of steam



The Deep Learning era broke that paradigm with demand for training compute far outstripping growth in compute power

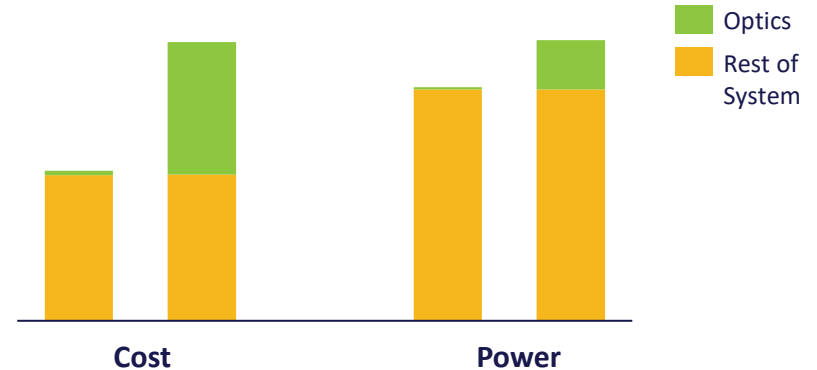
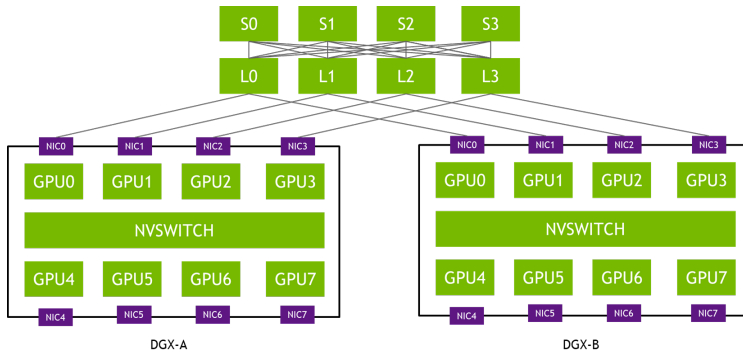
Source: [EPOCH](#), accessed: 21st of March 2024

But that's OK, we can parallelize, right?!

Yes, we can! But we create an optical bottleneck.

Electrical (memory) bandwidth is already outpacing optical bandwidth growth, and the need for connectivity driven by parallelization increases this bandwidth gap.

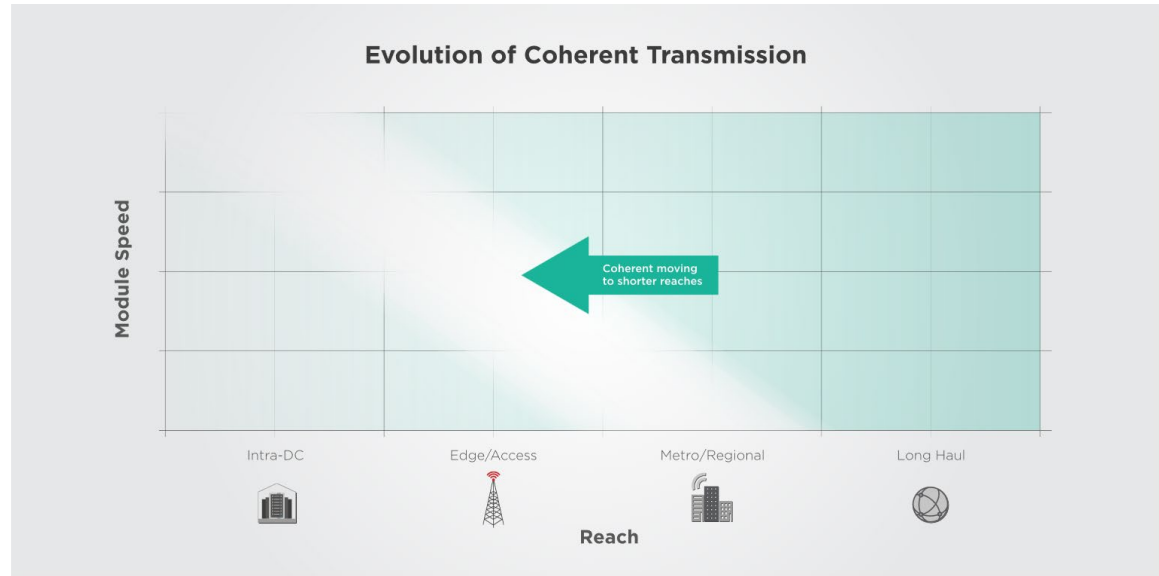
Scaling with present-day optical transceivers creates a cost and power problem.



Source: [NVIDIA](#)

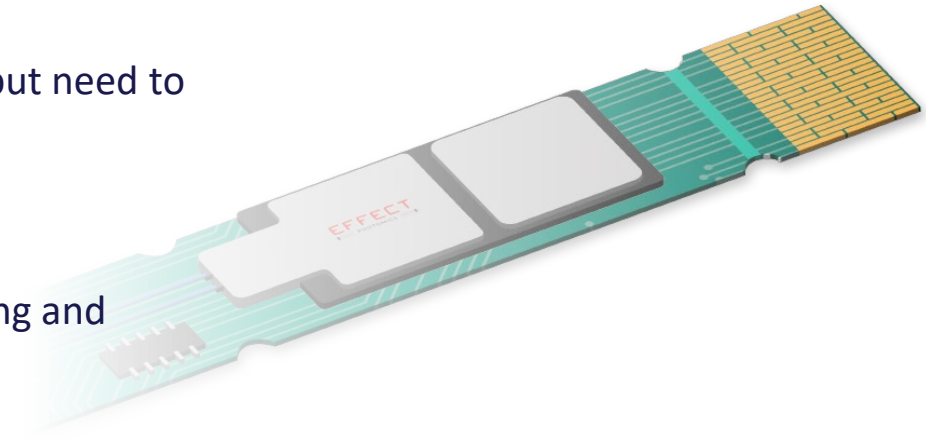
Proven technology to increase data rate

- Coherent optical communication has been used in core networks for decades.
- It is a proven technology that allows for 4x data rate at the same baud rate compared with direct detection.
- Clear trend toward the use of coherent transmissions for shorter distances
- Coherent miniaturization enables pluggable use cases
 - 100G and 400G ZR



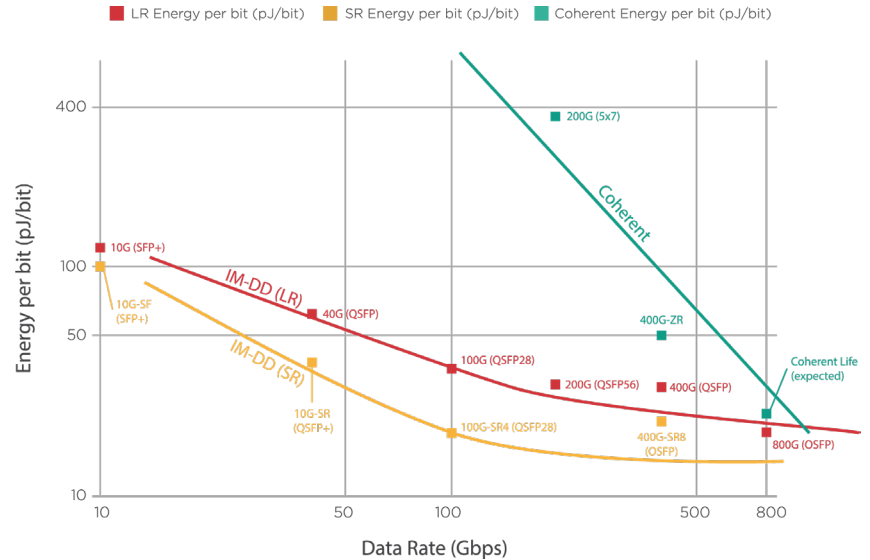
Co-packaged coherent DSPs drive down \$/bit

- \$/bit
 - Parallel optics don't drive down \$/bit and create a reliability concern
 - Coherent optics allow for higher data rates but need to be designed for cost
- Co-packaging drives down manufacturing costs.
 - Using existing microelectronics manufacturing and packaging processes
 - Shared real-estate
 - Lower losses allow for simplified designs while maintaining design margins.



Low-power coherent DSPs drive down pJ/bit

- Optics don't scale with CMOS speed
 - No Moore's law in optics
- Low-power coherent DSPs can squeeze out more performance from the same optics.
 - Exploiting the rapid advances in CMOS to drive down the energy per bit (pJ/bit)
 - Designed for data center applications (e.g. , O-band and latency)



Source: [Google](#)

Use existing technology to scale AI networks

We need to find a sustainable way to scale AI clusters (in terms of cost and power)

- Don't reinvent the wheel! Coherent optical technology has been around for decades in the network core and datacenter interconnect.
- Let's join forces and find ways for coherent optics to help address the challenges of AI-interconnect.

🏠 > Projects > Future Technologies Initiative > Short Reach Optical Interconnect



OCP
FUTURE
TECHNOLOGIES
INITIATIVE

Project Lead
Ron Swartzentruber

[Short Reach Optical Interconnect Group](#)

Where to find additional information (URL links)

<https://effectphotonics.com/insights/coherent-optics-for-ai/>

<https://effectphotonics.com/insights/tunable-lasers-and-dsps-in-the-age-of-ai/>

<https://effectphotonics.com/insights/transceivers-in-the-age-of-ai/>

Slides will be available for download at www.effectphotonics.com

Thank you!

Scaling Innovation Through Collaboration



OCP
REGIONAL
SUMMIT

24-25 APRIL 2024
LISBON, PORTUGAL

