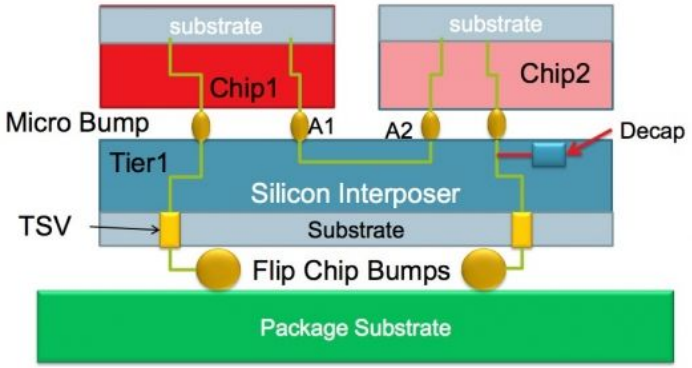
Co-Packaged Optics

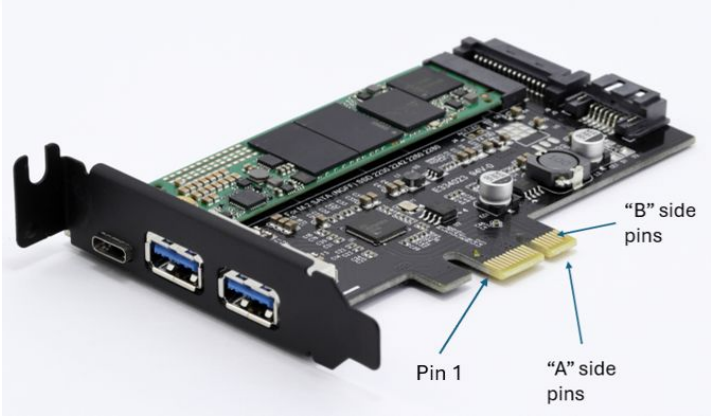
Eric Ding

Corning exhibits, OFC2025

The Journey of a Bit

1. CPU/GPU/ASIC
2. Substrate
3. PCB trace, (PCIe)
4. NIC
5. Copper / Fiber / Air
6. NIC / Mirror / Antenna
7. …
8. CPU/GPU/ASIC/Switch…
9. …





https://www.nwengineeringllc.com/article/how-interposers-are-used-in-ic-substrates.php

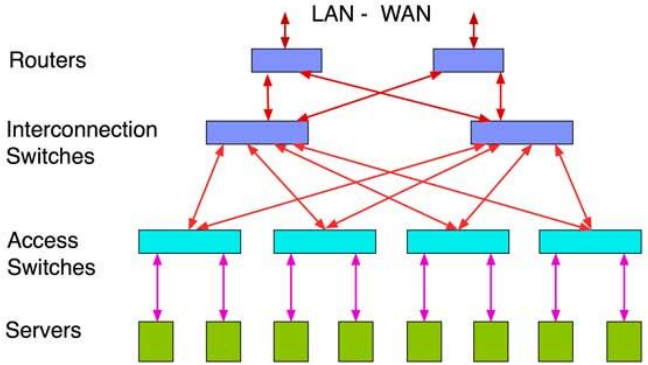
Outline

1. Review of optical networks in datacenters
2. Digital to optical conversion
3. A **gentle** introduction of CPO
    a. Components
    b. Packaging
    c. SOTA implementation
4. Challenges

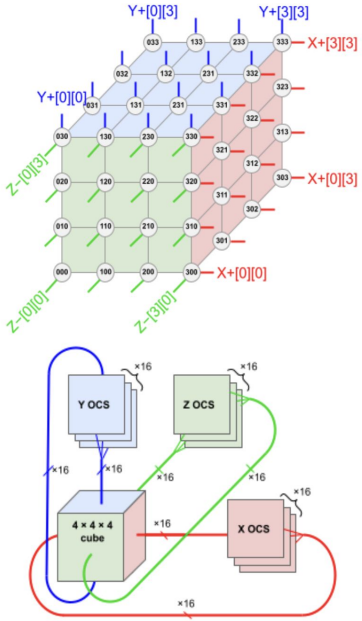
I learned a lot from her posts on optics.

Sharada Yeluri
Sr. Director of Engineering, Silicon and Systems Technology, @ Juniper Networks

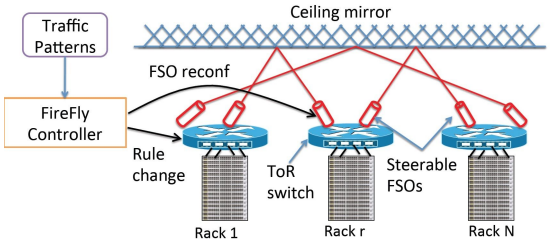
Optics in Datacenter



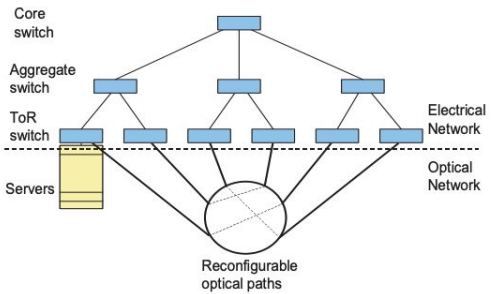
Clos



TPU Torus

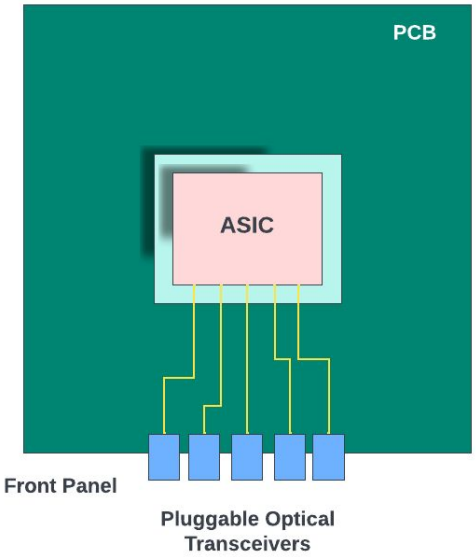


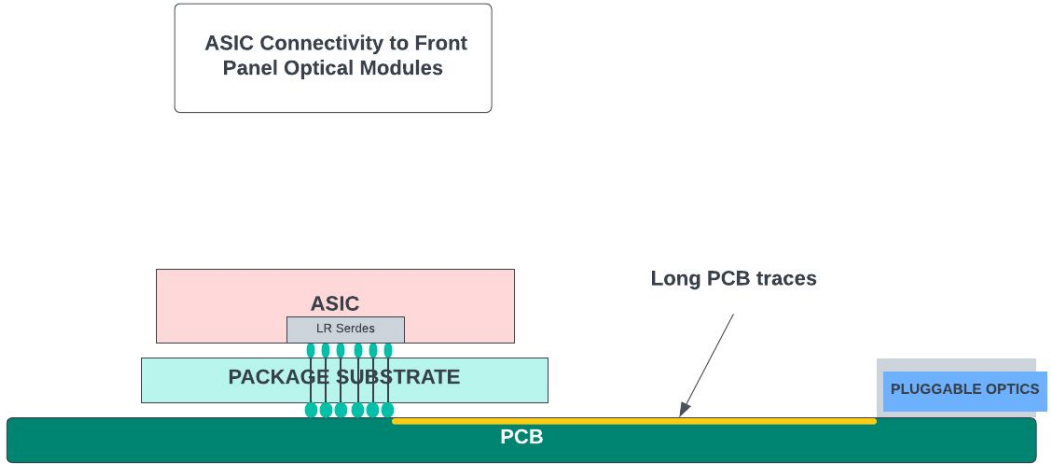
Firefly



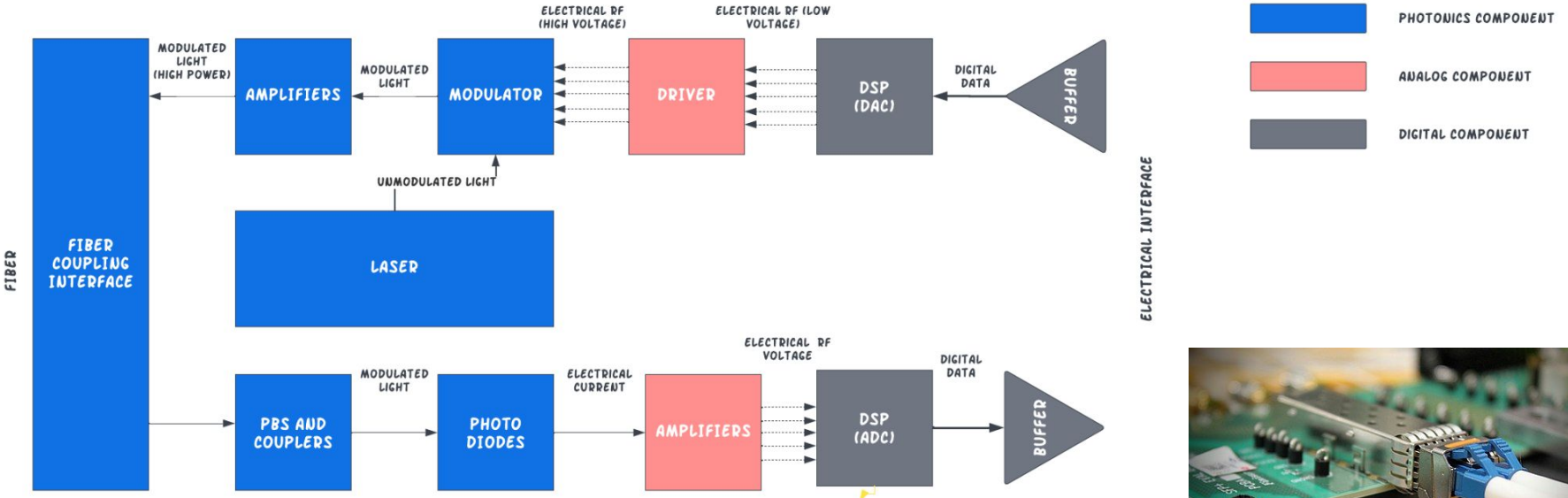
C-Through

How Do We Translate Data to Wavelengths

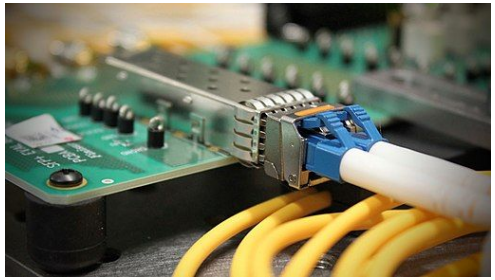


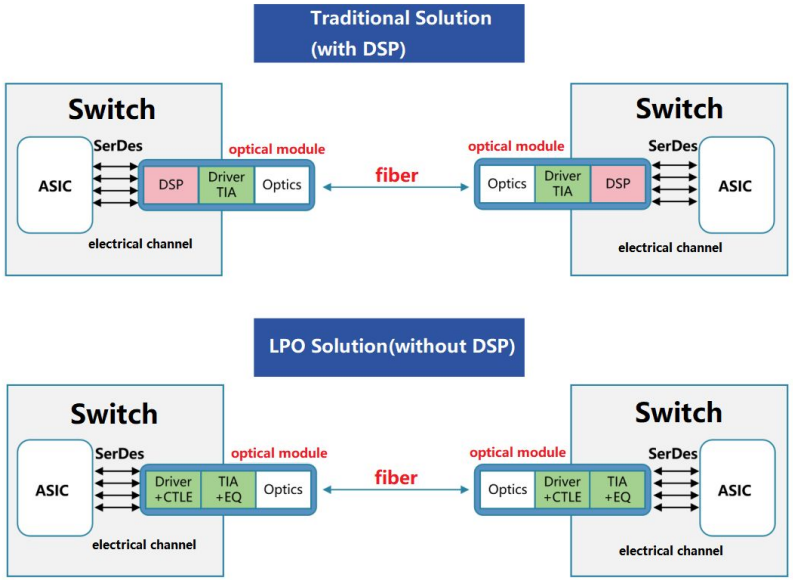


How Do We Translate Data to Wavelengths



How Do We Translate Data to Wavelengths

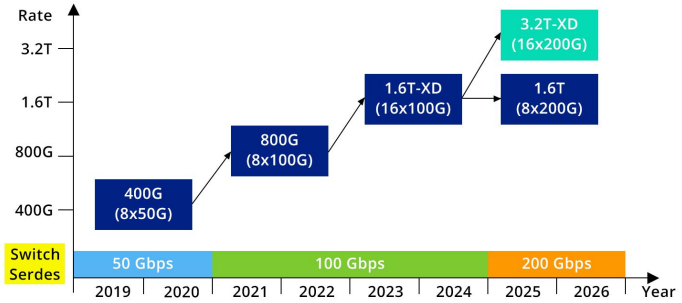


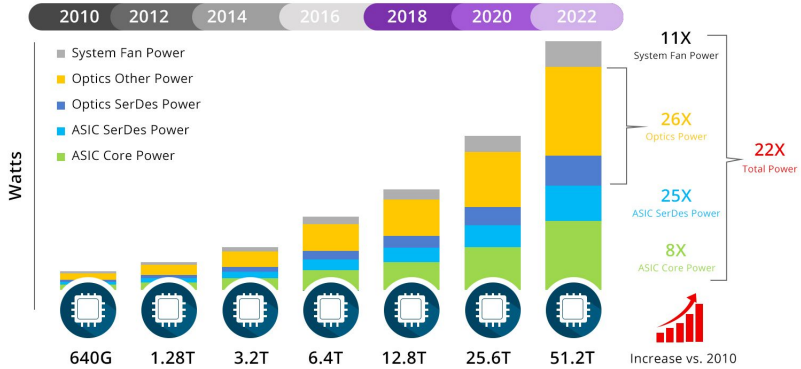


Component	Traditional Pluggable Optics	LPO (Linear Pluggable Optics)
DSP	Integrated in module (high power)	Host-based or eliminated
Driver	DSP-driven with retiming	Linear-drive with CTLE integration
TIA	Basic amplification	Adaptive EQ + CTLE + digital diagnostics
Power	10–14W/module	2–4W/module
Latency	Higher (DSP processing)	Lower (direct drive)

https://www.fs.com/blog/what-is-the-lpo-transceiver-8.html

From Pluggable to CPO





Higher bandwidth and bandwidth density

Front panel size constraint for pluggables

Energy consumption

Cisco's and Broadcom's early CPO implementations show 30–50% power savings

Latency

FEC adds latency of ~100ns for >50Gbps serdes

Higher Signal Integrity

Eliminate much of the electrical path and associated losses

Scalability and Future-Proofing

https://www.fs.com/blog/what-is-the-lpo-transceiver-8.html

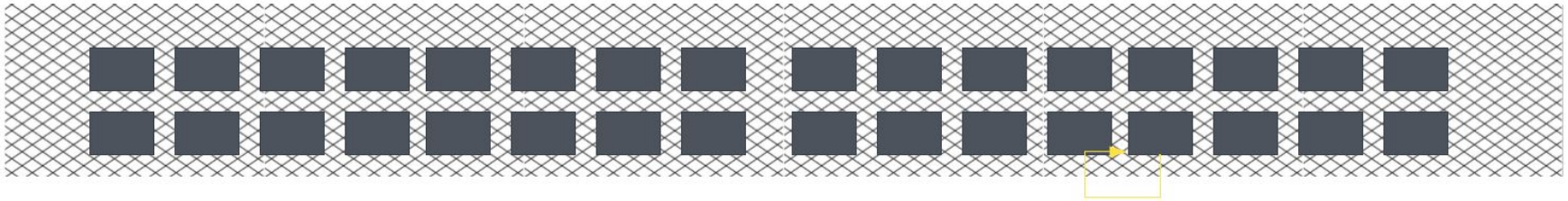
From Pluggable to CPO

Higher bandwidth and bandwidth density

Front panel size constraint for pluggables

In a typical single rack unit (1RU) switch box, one can fit between 32-36 pluggable optics

Airflow into the system is obstructed



From Pluggable to CPO

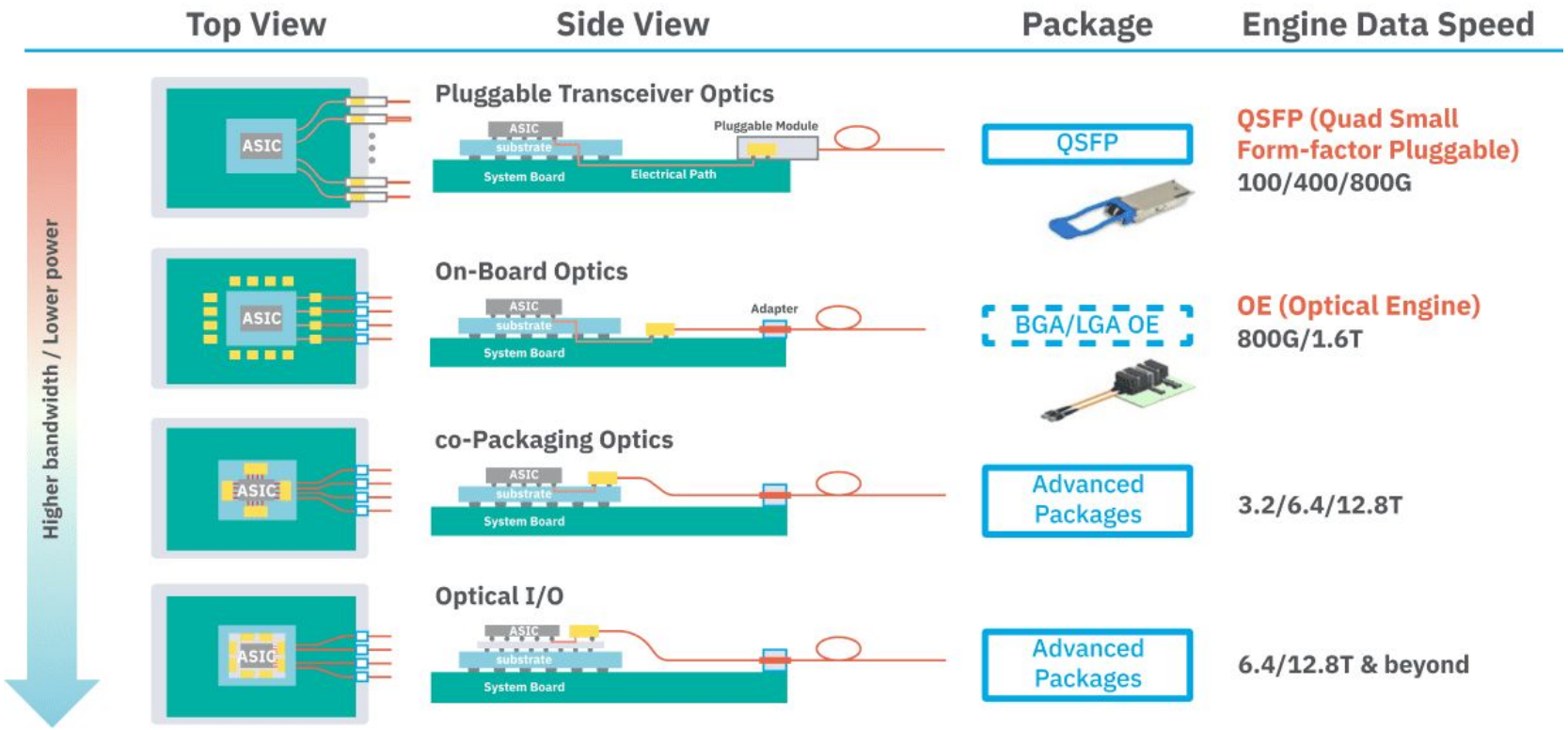


Illustration of various integration techniques for optics. ASE

Key Metrics

Bandwidth: Tbps, Gbps

Bandwidth density: Tbps/mm

Core-die: 500-1000 Gbps/mm x 625 mm^2 = 100 Tbps

Energy per bit: pJ/bit

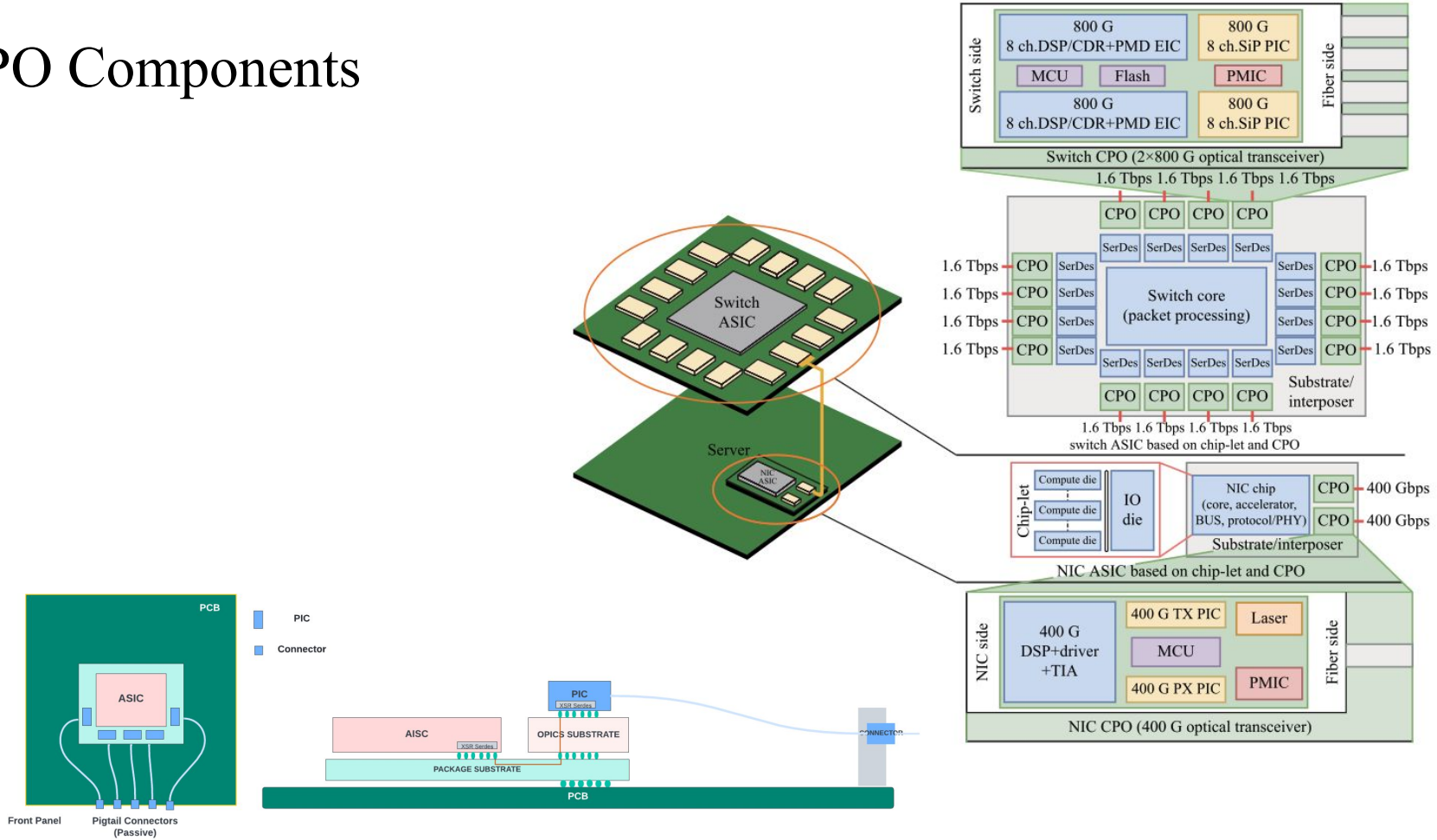
Failure in time (1 billion hours): FIT

Defects: DPPM (defective parts per million)

Temperature: degree C

Signal power: dBm

CPO Components



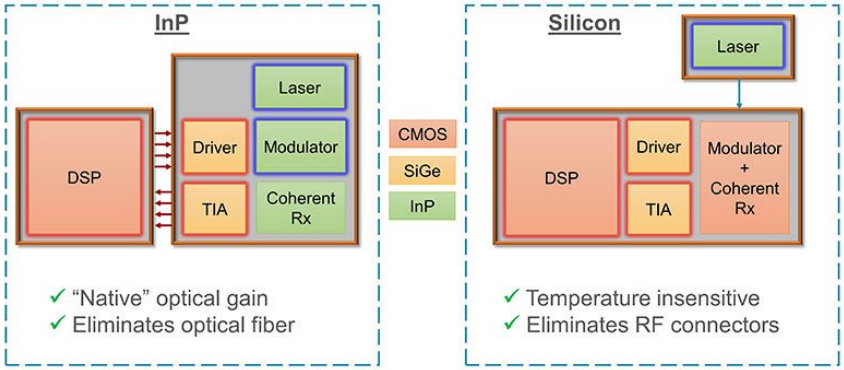
Photonic Integrated Circuits

Integrate many of the optical and electrical components in transceivers

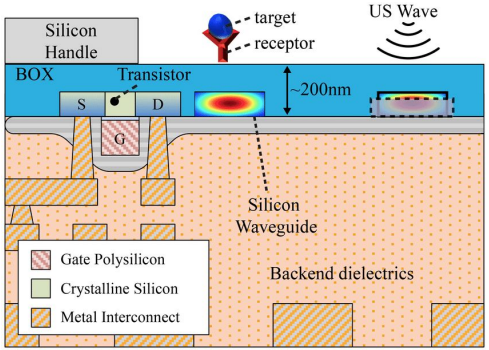
- Reduce power due to fewer coupling effect

Two mainstream platform

- **InP** (Indium Phosphide), more mature
- **Si** (Silicon/CMOS-based), heavy investment, can undergo wafer-level testing, lowering defect rates to 30 dppm

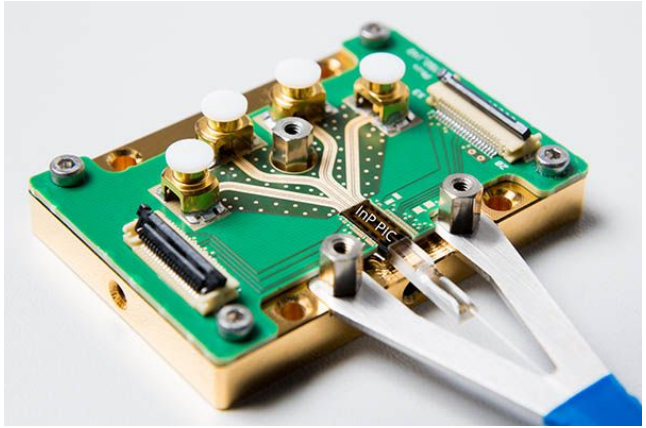


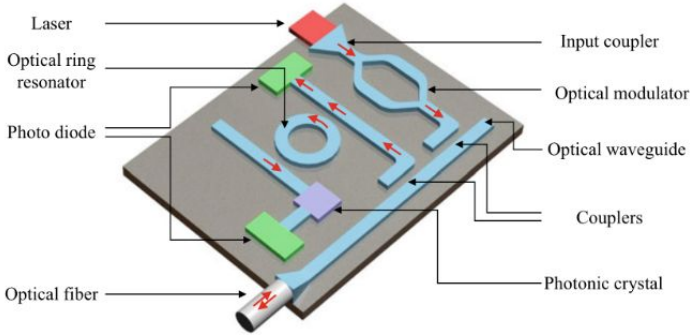
https://acacia-inc.com/acacia-resources/100gbaud-silicon-photonics-solutions-drive-optical-network-evolution/

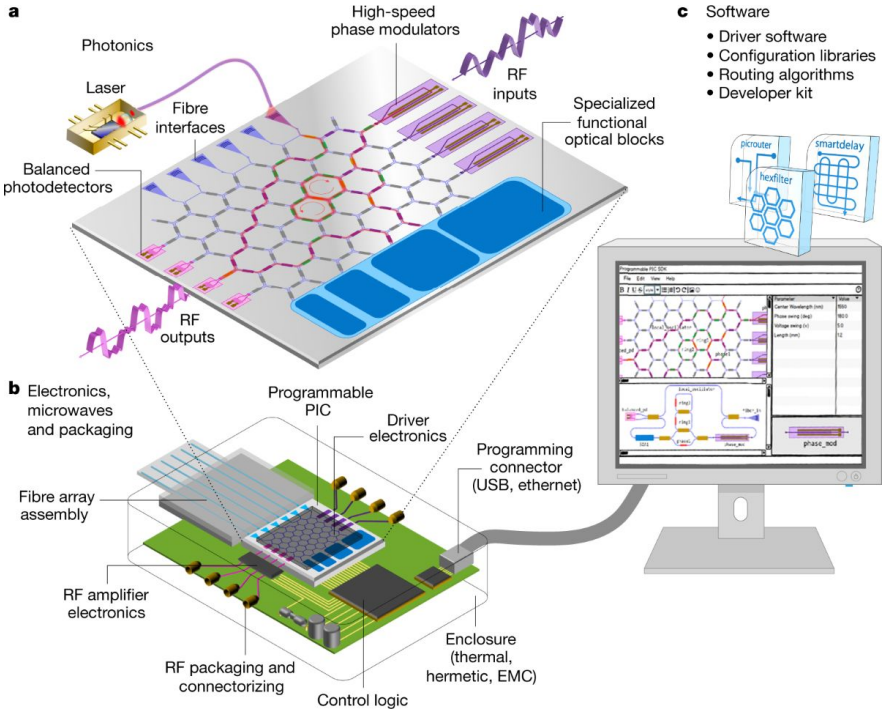


https://www.researchgate.net/figure/nm-Monolithic-Platform-Cross-Section-The-thin-Si-waveguide-and-BOX-layer-allows-a-low_fig4_355749927

Photonic Integrated Circuits



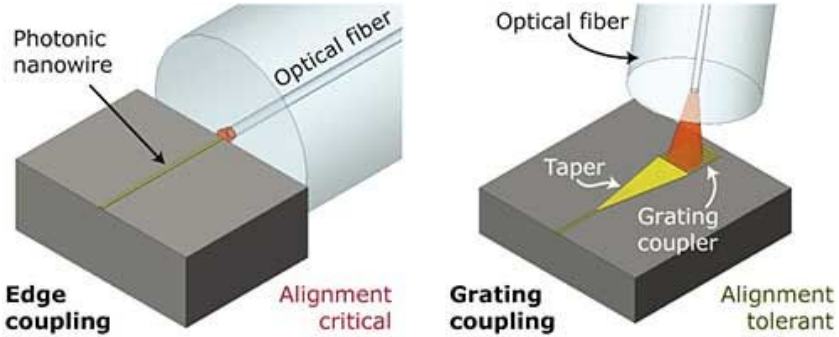




https://www.ovaga.com/blog/transistor/photonic-integrated-circuit-definition-disadvantage-fabrication-application
https://www.nature.com/articles/s41586-020-2764-0

Fiber Coupling

- Grating coupling
  - Use on-chip diffractive grating coupler
  - Fiber could be placed anywhere above the photonic die
  - High loss
- Edge coupling
  - V-groove fiber arrays for precisely spacing fibers (50~250 um)
  - Low insertion loss
  - Constraint by edge length



Both Broadcom and NVIDIA choose edge coupling

https://www.kth.se/is/mst/research/photonics/projects/apodized-waveguide-to-fiber-surface-grating-couplers-1.315473

Modulators

Components inside the optical engine that convert electrical signals into light

Micro-ring

- NVIDIA's approach
- Smaller footprint
- Less power consumption, 1–2 pJ/bit
- Support WDM. Each ring targets one wavelength

Mach-Zehnder modulator

- Applying electric fields to the arms changes optical path lengths resulting in phase modulation
- Has better tolerance to temperature variations
- Higher power consumption, 5–10 pJ/bit
- Larger foot-print
- Good for 100 Gbps, but will hit density and power limits when scale to ~200G lanes

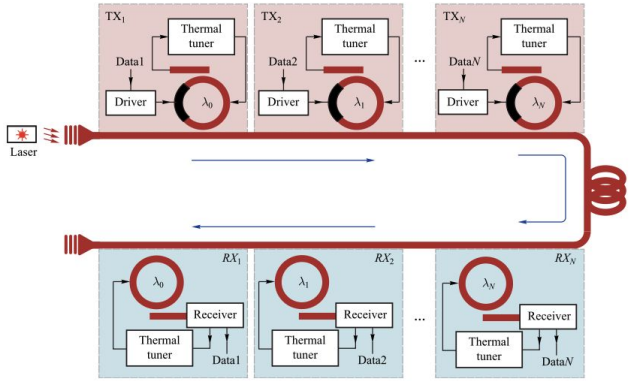
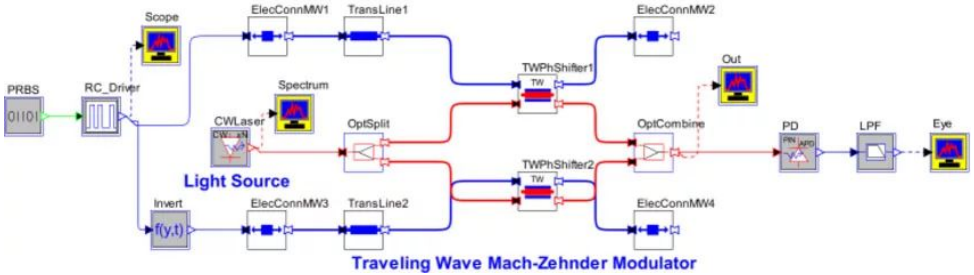


Fig. 14 The diagram of MRR-based transceiver consisting of drivers, MRMs, receivers, MRR DEMUX, and thermal tuners

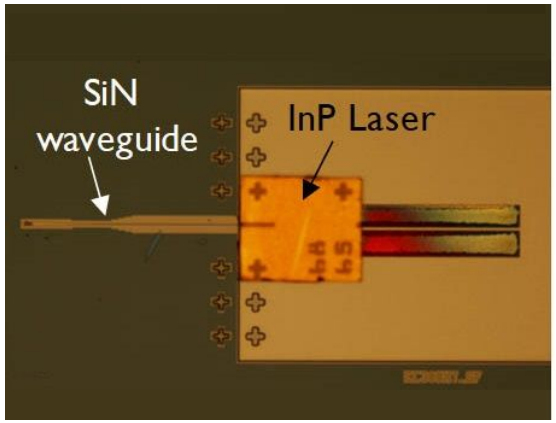


https://www.synopsys.com/glossary/what-is-a-mach-zehnder-modulator.html

Laser Source

On-chip laser

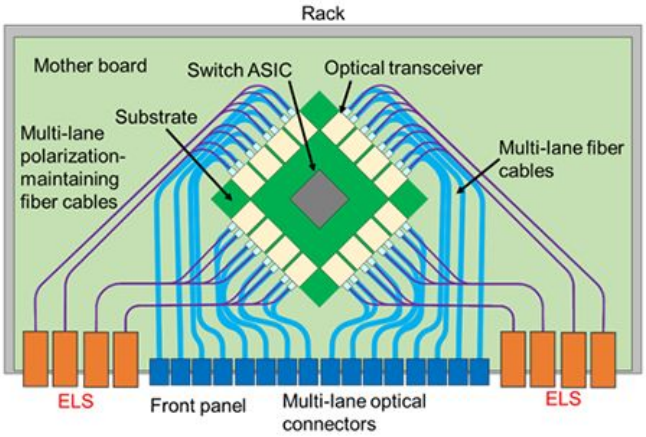
- Strict thermal/cooling requirements
- Redundancy, adding to the cost/area



Microscope image of an InP DFB laser assembled on a Si Photonics chip.

External laser source (ELS):

- Plugged into the front panel LC ports
- Fiber patch cords deliver the light from laser module into CPO engines
- Can be easily replaced
- Higher power consumption

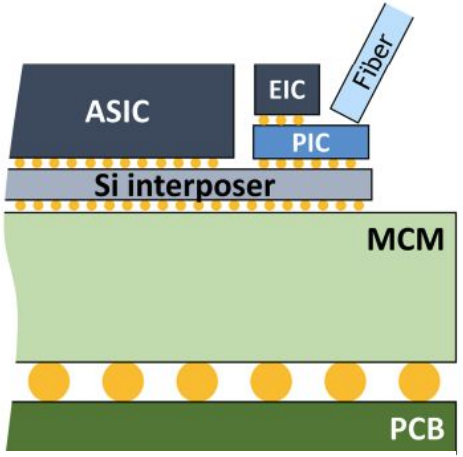


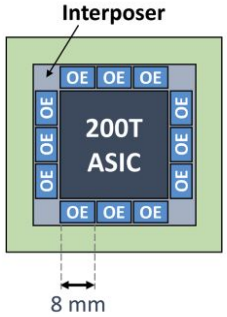
https://www.furukawa.co.jp/en/release/2022/comm_20220307.html

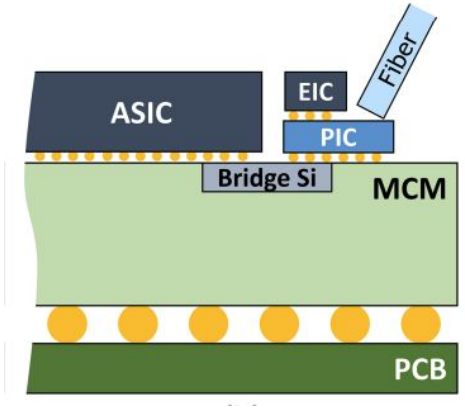
CPO Integration

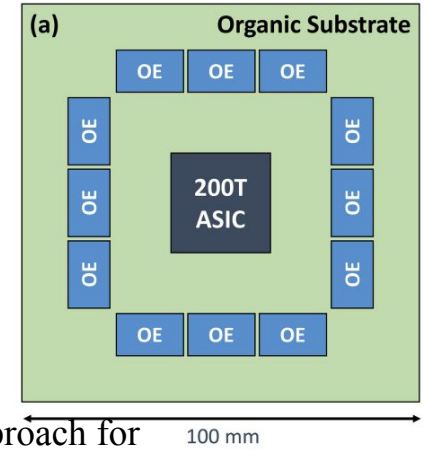
Two approaches:

- Silicon interposer
  - Co-located on a silicon interposer
  - A smaller package
  - Thermal issue (151.76 ℃)
  - Interposer size constraint
  - High bandwidth density requirement
- Organic substrate approach
  - Allows relaxed packaging
  - Thermal isolation



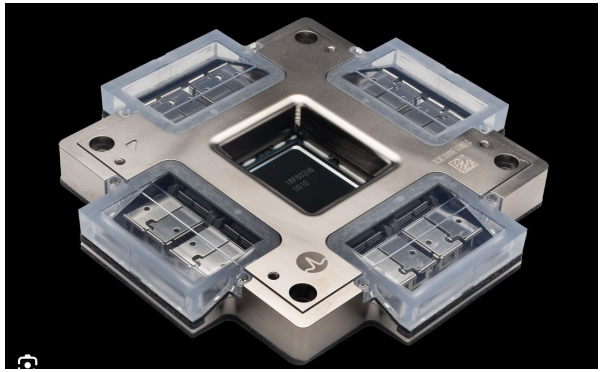






Beyond CPO: A Motivation and Approach for Bringing Optics Onto the Silicon Interposer, JOLT 2023

CPO Implementation



Broadcom's Bailly CPO ASIC, 2024

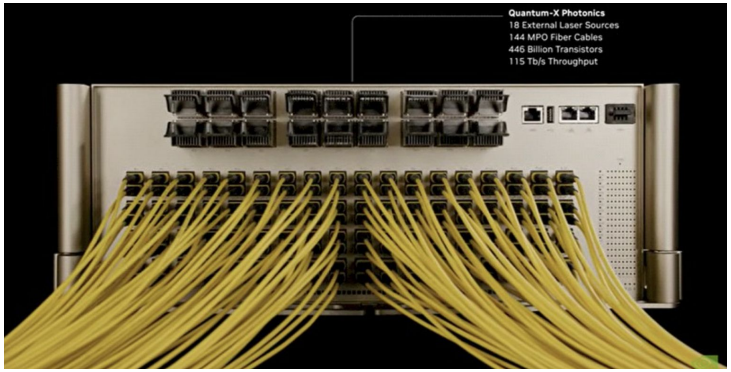
Tomohawk-5 ASIC

8 x 6.4 Tbps optical engine

Organic substrate

51.2 Tbps total optical bandwidth

30% power saving



NVIDIA Quantum-X Photonics Switch, 2025

Quantum X800 ASIC

4 x 28.8 Tbps switch ASIC core

6 x 4.8 Tbps detachable optical sub-assemblies, replacable for each core

115.2 Tbps total optical bandwidth

From 2.5D to 3D Packaging

Base layer: lasers, waveguides, and optical switching/routing

Top: compute and memory chiplets

A continuous 2D surface for optical I/O!

Thermal management could be tricky

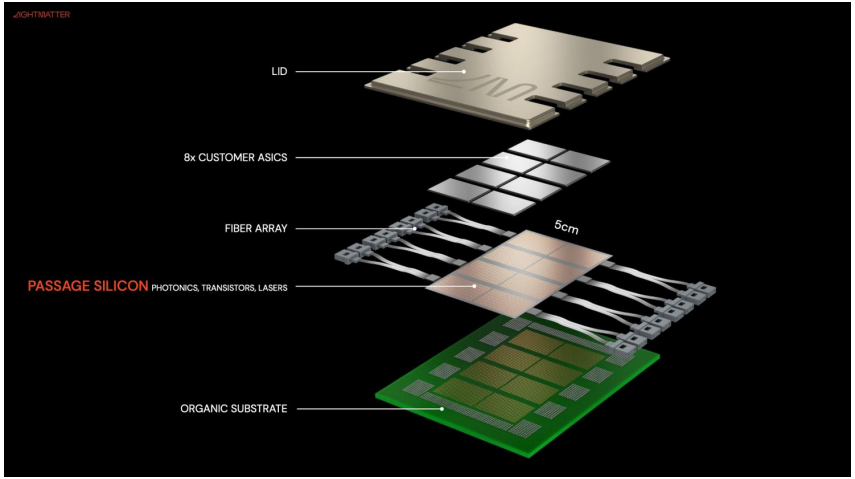
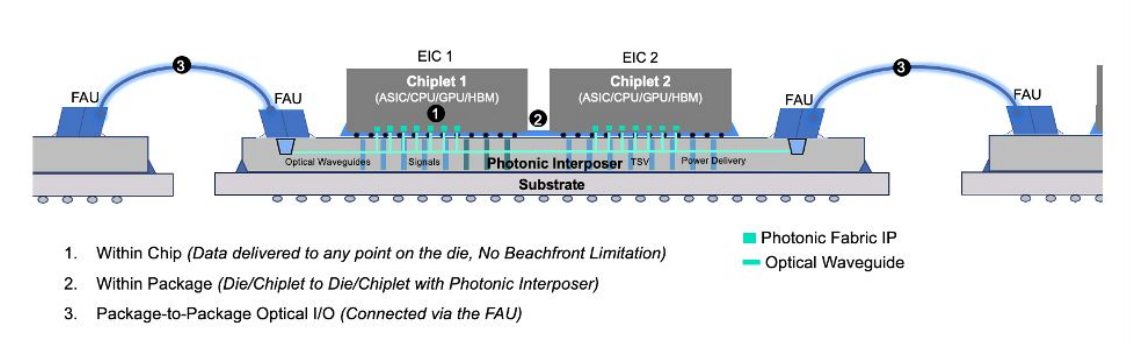


Illustration of Photonic Interposer. LightMatter



Photonic Interposer. Celestial.ai

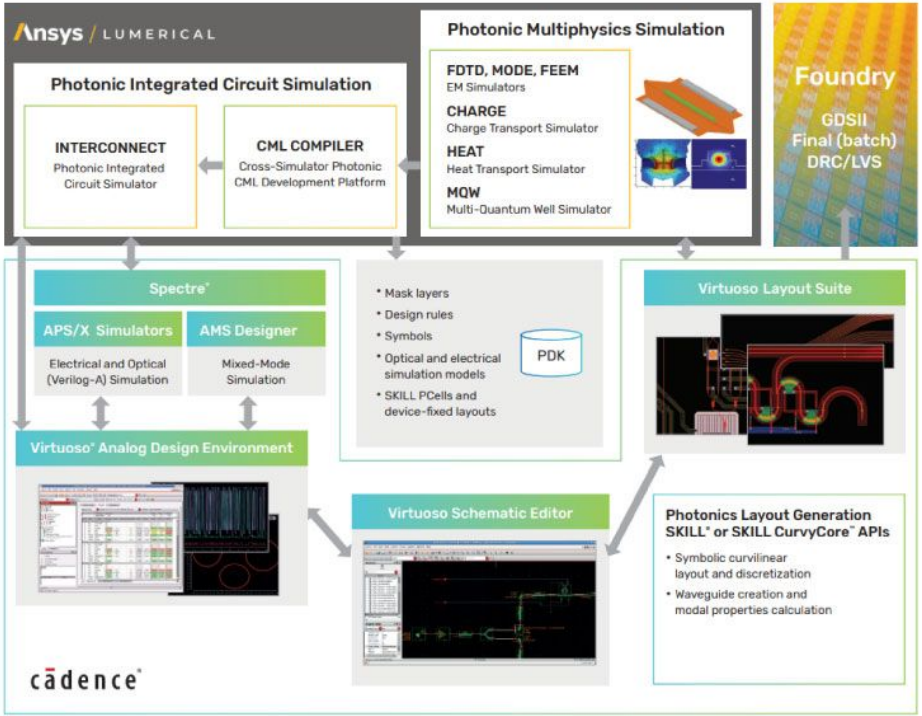
Integrated Design Flows for Photonic Circuits

verilog

```
`include "constants.vams"
`include "disciplines.vams"

module optical_attenuator(in, out);
  electrical in, out;
  parameter real attenuation_dB = 3.0;
  real attenuation_lin;

  analog begin
    attenuation_lin = 10**(-attenuation_dB/20);
    V(out) <+ V(in) * attenuation_lin;
  end
endmodule
```



Principles to Achieve Faster Datarate

WDM: get more channels per fiber

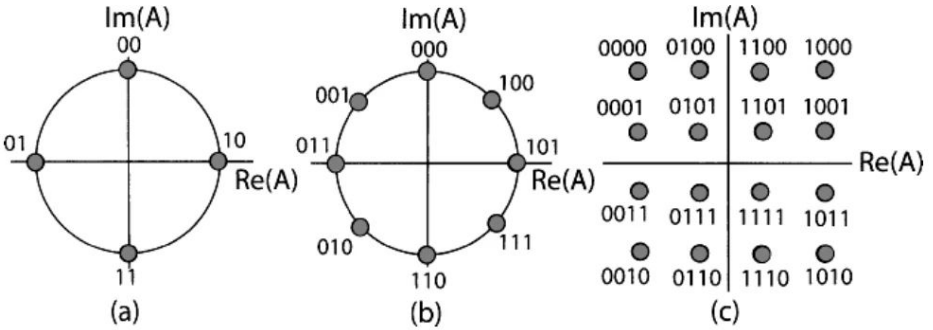
Multi-core fiber: packs several independent cores within one fiber cladding

Dense fiber: more fibers per mm

Grating coupling: more fibers per mm^2

Better cooling

Better modulation: QPSK, 8-PSK, 16-QAM



Constellation diagrams for (a) QPSK, (b) 8-PSK, and (c) 16-QAM modulation formats showing how multibit combinations are assigned to different symbols.

CPO Challenges

Interoperability

Cannot choose different cables (speed/range) for different ports like pluggables

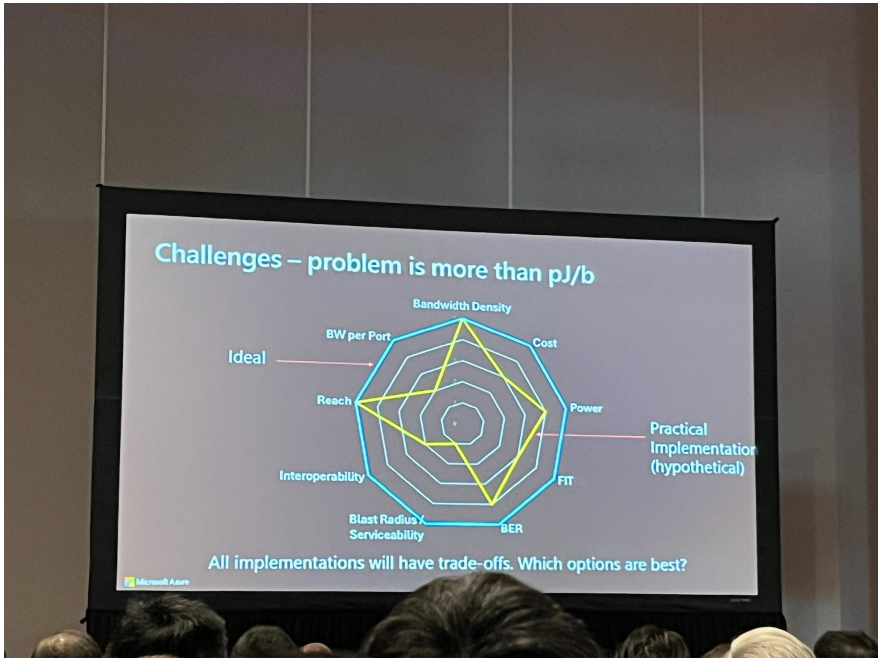
Thermal issue with electronic chip

Fault tolerance

What if one optical engine fails?

Complexity

Founder of Arista, Andy Bechtolsheim, supports LPO



Microsoft, OFC2025

Conclusion

Power saving: 30% reduction for 51.2Tbps switch

Higher bandwidth

Interoperability issues

Wide applications:

- Scale-out datacenter networking
- High speed interconnect in HPC systems
- LiDAR chip for autonomous driving