Interconnect trends for AI datacenters

Sara Zebian
OpenAI
April 15-16, 2025

Agenda

- Key Takeaways
- ChatGPT in a nutshell
- Scaling Infrastructure
- Interconnects Trends and Challenges
- Conclusions

Key Takeaways

Rapid Scaling of AI infrastructures will likely require:

- **Faster transition to NextGen signaling**
- **Reliable AI systems and Interconnects**
- **Resilient Supply Chains**
- **Both Copper and Optics Interconnects**

ChatGPT in a nutshell

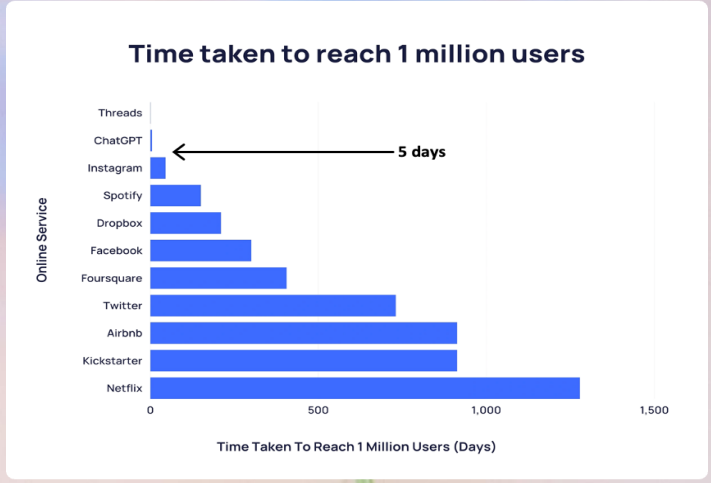
OIF 448Gbps Signaling for AI Workshop
April 15-16, 2025

ChatGPT in a daily life

>400M users weekly

Educational Enhancement
Content Creation
Programming assistance
Language learning

Sora video creation
Imagegen



ChatGPT in a nutshell

- **Collect a dataset including:**
  - Text
  - Code
  - Images
  - Audio
  - Problems in Math and Logic
- **Pre-train** a model to predict the next word
- **Post-train** it to:
  - Follow instructions and safety policies
  - Be conversational
  - Use tools

➢ **Extremely large scale of data being processed**

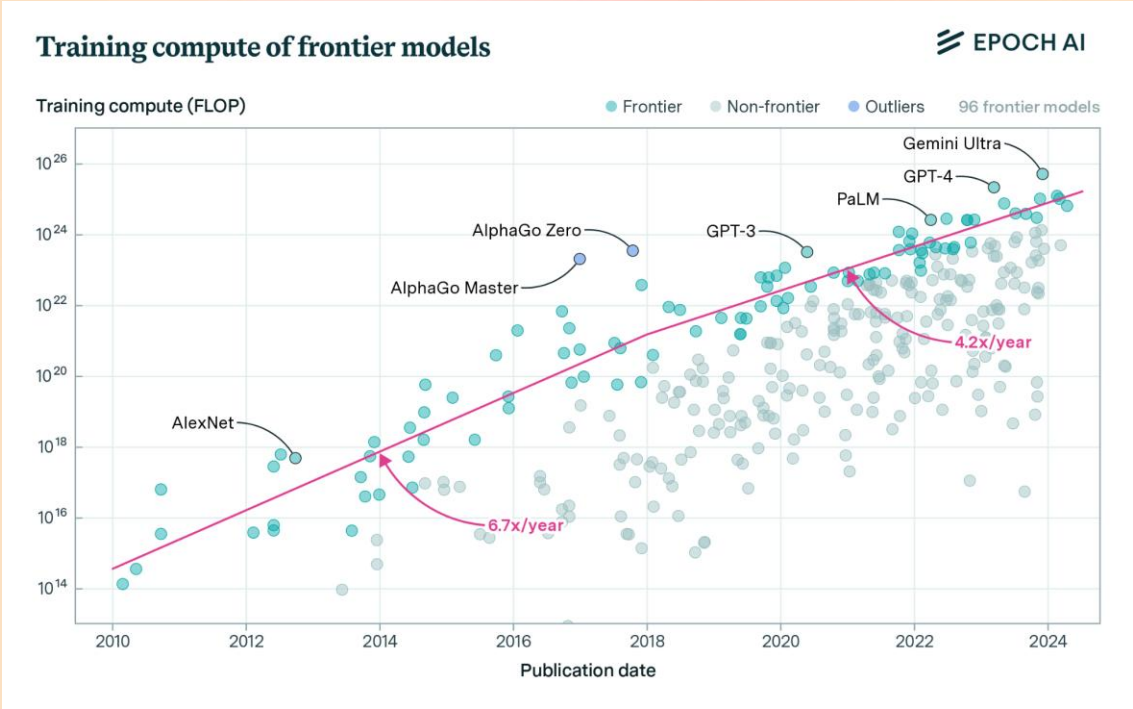
Scaling Infrastructure

OIF 448Gbps Signaling for AI Workshop
April 15-16, 2025

Demand for Compute Continues to Grow

Compute to Train Frontier models

- Grew by 6.7x/year until 2018
- Grew over 4x/year since
- Moore's Law helped
- Primarily enabled by:
  - Precision innovations
  - Architecture optimization
  - System scale
  - Run times

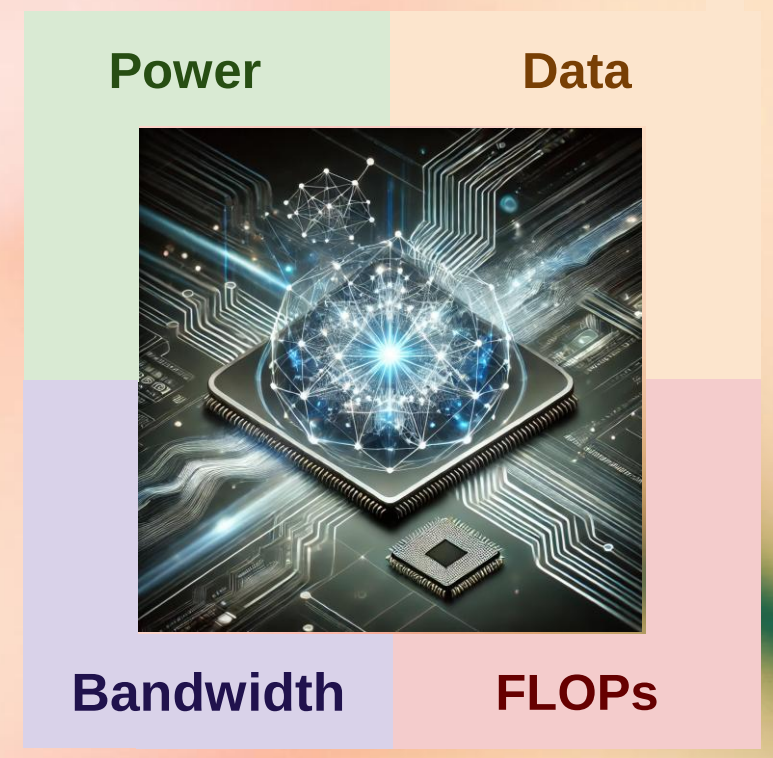


Training Compute of Frontier AI Models Grows by 4-5x per Year, Sevilla and Roldán.
https://epochai.org/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year (2024)

➢ **Required compute resources grow very fast**

Continued Scaling of AI Systems

1. **Data**
   - There are ~500T tokens of data (epoch.ai)
2. **Power**
   - Available renewable green energy
3. **Hardware FLOPs**
   - Limits to GPU/Accelerator production
4. **Bandwidth**
   - Memory - moving data to compute
   - **Interconnect** - parallelization



➢ **Bandwidth & interconnect growth need to keep up with compute**

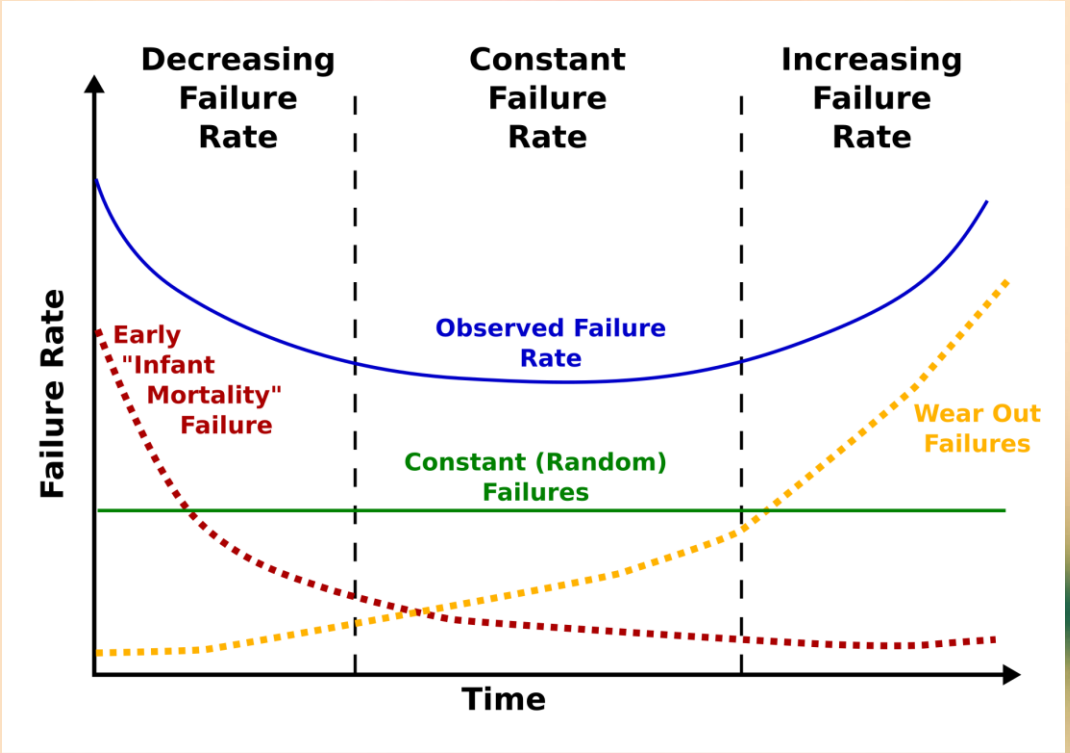
Infrastructure development

➢ **Infrastructure will need to be developed at very large scale with a resilient supply chain**

Optimizing AI Computation

Availability of large-systems

- Small drop in MTBF → big drop in uptime
- Time to recovery is important

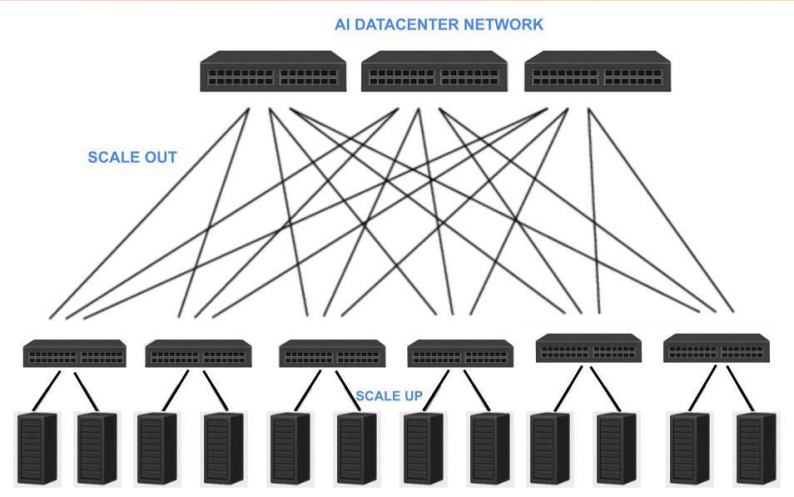


➢ **Reliability of the interconnect is a key aspect of a stable Network**

Example of a Datacenter with AI Systems

AI “factories” different than Cloud datacenters:

- Ability to run single synchronized workloads
- Much higher power and cooling densities



2 main networks:

1. **Scale-Up**
   a. Highest Bandwidth, Low Latency, Low Power
2. **Scale-Out**
   a. Resilient, High Bandwidth, High Scale

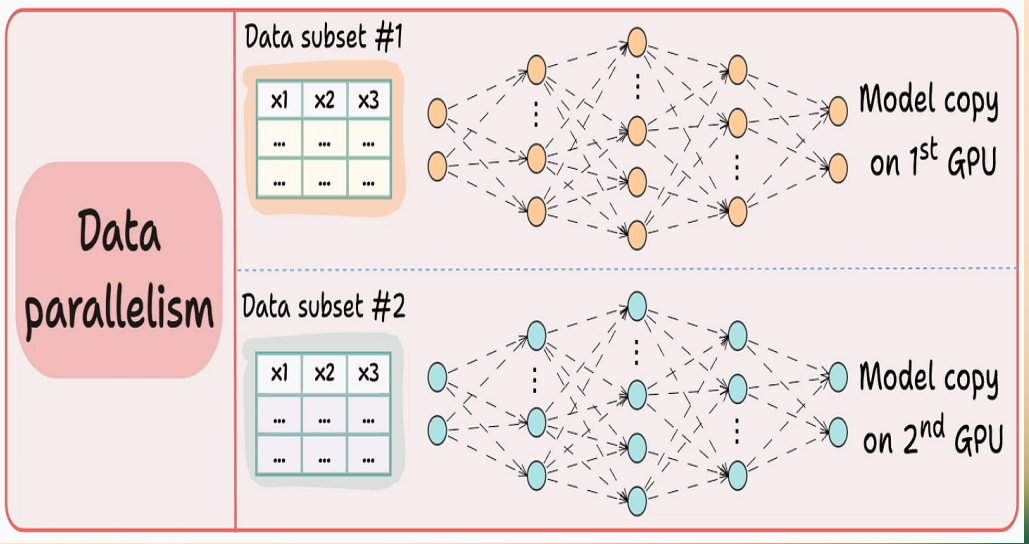
➢ **Each network type has different Interconnects requirements**

Scale up: Rack Scaling

Today's scale-up domain sizes likely to hit a wall

Accelerator chips to talk to each other A LOT with parallelism
As many accelerator chips as possible to feel like one big one

Denser architecture



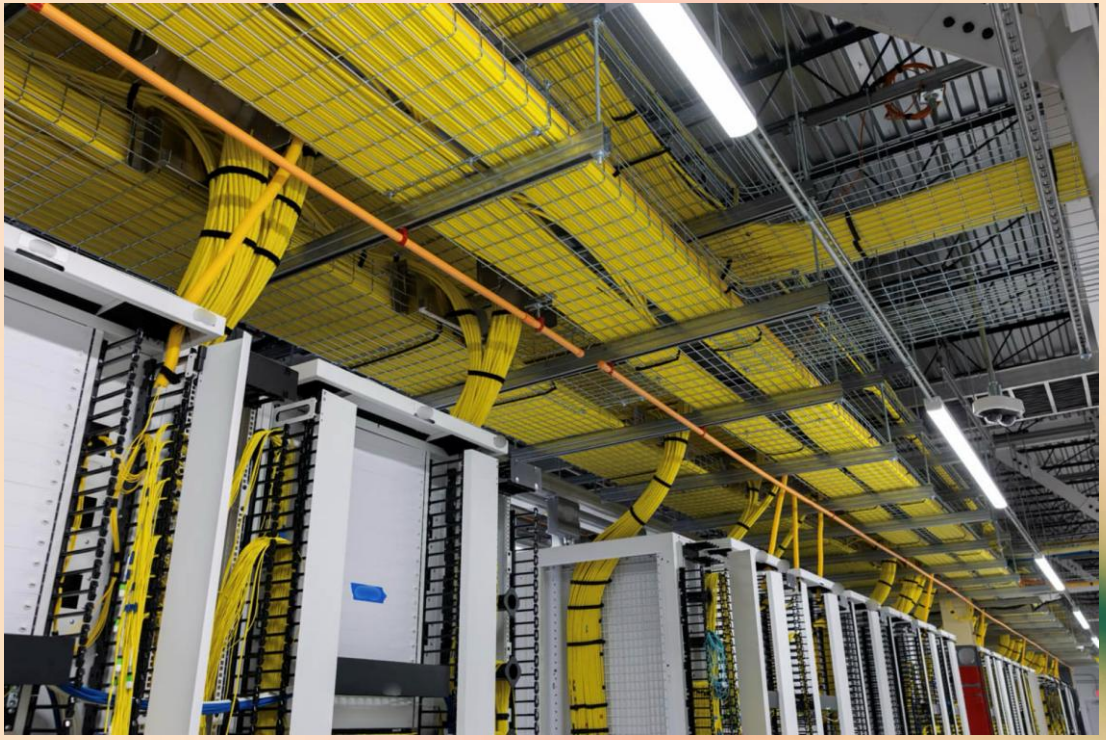
➢ **AI Scale-up clusters require significantly large number of interconnects**

Scale Out: Beyond the Rack

Non-blocking fabric clos topologies with mainly Ethernet interconnects

Pluggable optical modules have been the norm (longer reach)

Slower network relative to Scale up



https://www.servethehome.com/

➢ **Scale-out network most likely to continue to use Optics**

Interconnects Trends & Challenges

OIF 448Gbps Signaling for AI Workshop
April 15-16, 2025

Datacenter Interconnects

Copper

- Passive: DAC, Backplane, …
- Active: ACC, AEC, retimers, …

Optics

- DR4, FR4, LPO, …

Flexibility, faster repairability and more resilient supply chain with **Pluggables**

https://developer.nvidia.com/blog/nvidia-gb200-nvl72-delivers-trillion-parameter-llm-training-and-real-time-inference/

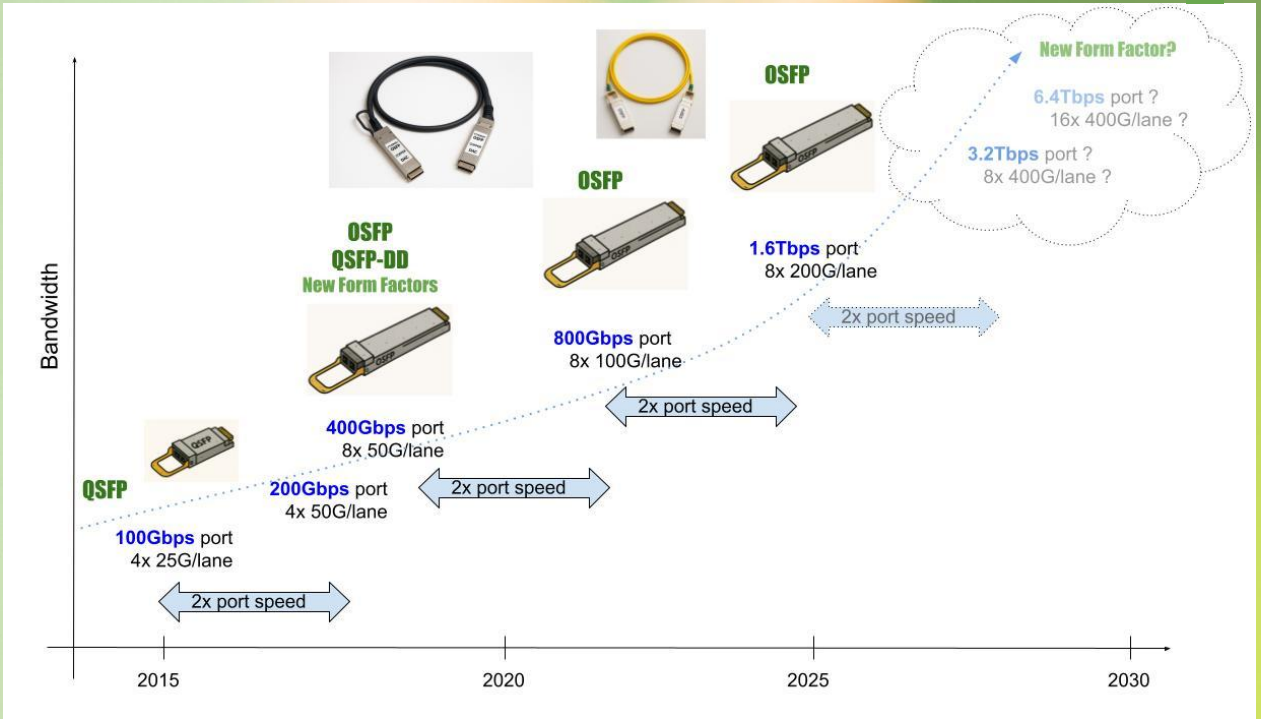
➢ **Interconnect type selection: Reach vs Power vs Reliability**

Evolution of Ethernet Port Speeds in Datacenter

Pluggables have been a constant in today's datacenters

Evolution of port speed continues, doubling every ~3 years

Likely faster trajectory now due to AI



➢ **Need to accelerate transition to nextGen port speed**

How to get to 3.2Tbps/port?

Today

1.6Tbps port
In 8 lanes of 200Gbps

Next Generation

3.2Tbps in 16x 200Gbps

- Lower risk technology
- Beachfront package challenge
- System complexity due to density challenges

3.2Tbps in 8x 400Gbps

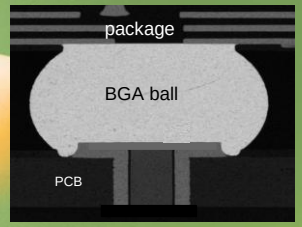
- Fewer number of interconnects
- New technology with higher bandwidth

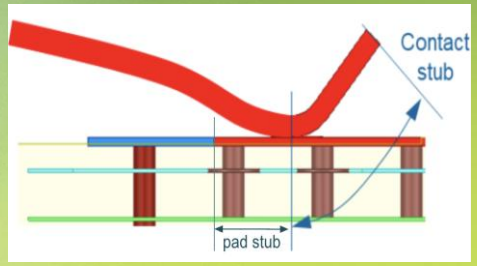
➢ **Doubling the lane speed to 400Gbps/lane will likely prevail**

Interconnect Challenges with 400Gbps/lane

Signal Integrity

- BGA balls and vias
- Crosstalk
- Mating stub effect
- Manufacturing variations





From TE/Shaji/Tracy (EA TEF Oct 2024)

Power

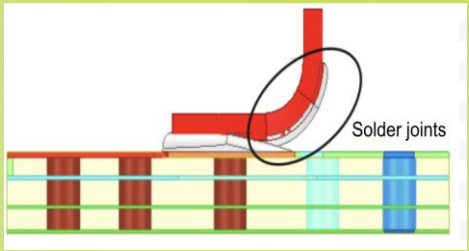
- Plateaued pJ/bit efficiency

Latency

- Stronger FEC may be needed

Reliability

- More active components on the channels



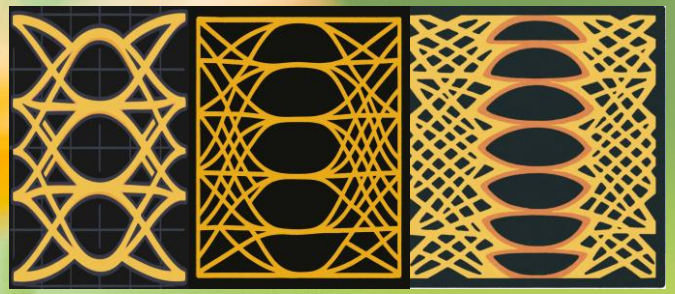
➢ **Mitigations with Innovative ASIC and Channel Technologies**

Potential Mitigations to get to 400Gbps/lane

1. ASIC

While trying to minimize power and latency …

- **Differentiated SerDes modulations**
  - Higher-order PAMx for lower bandwidth (Copper)
  - Lower-order PAMx for better SNR (Optics)
- **Extended reach**
  - More capable DSP-based SerDes (45dB?)
- **Advanced process node and packaging**
  - 2nm with better performance and lower power
  - Co-packaged technologies
- **Link Level Retry**



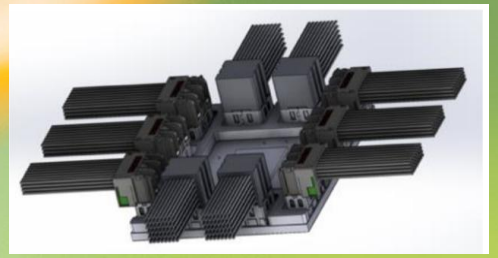
Bandwidth: 106GHz 85GHz **71GHz**

➢ **ASIC improvements involving modulations, reach and new packaging technologies can help**

Potential Mitigations to get to 400Gbps/lane

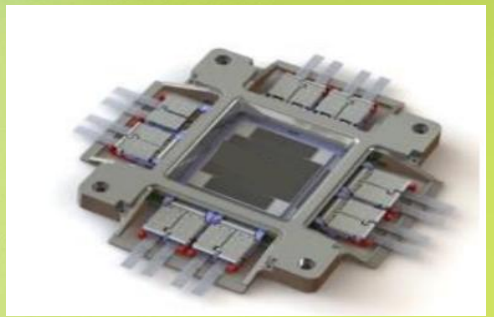
2. Channels

- Advanced connectors and cables designs
- Advanced manufacturing technologies
- Multi-sourcing options for resilient supply chain
- Innovative system architectures
  - Shorter interconnect reach within rack



From Broadcom

Co-Packaged Copper



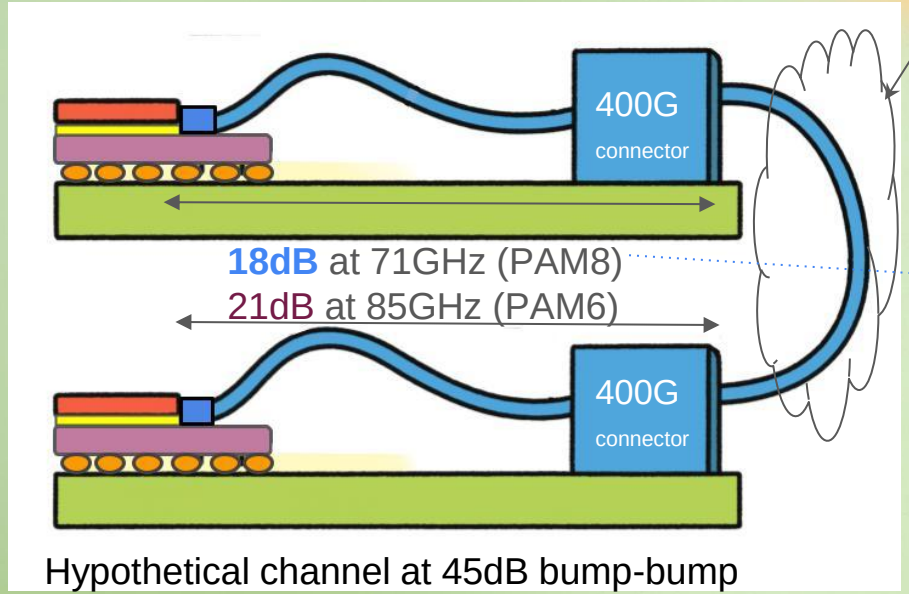
From Broadcom

Co-Packaged Optics

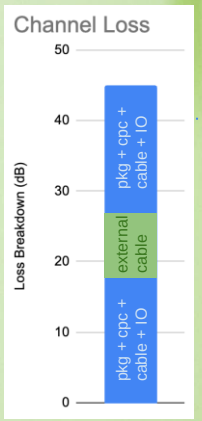
➢ **Channels: Advanced connector, cable and innovative systems can help**

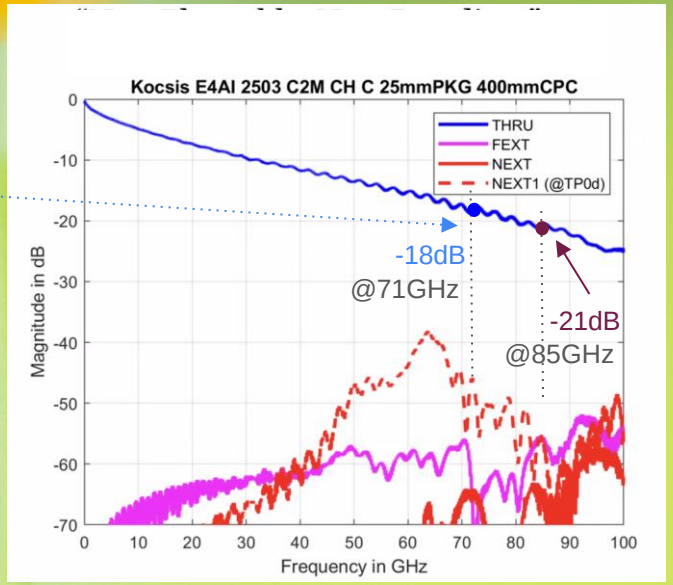
Hypothetical Example of a 400Gbps Channel for Scale-up

Optics pluggables or
Copper pluggables or backplane…
Short intra-rack channels
Likely with PAM8 and 45dB max reach?



Hypothetical channel at 45dB bump-bump





From Amphenol/Kocsis (IEEE March 2025)

➢ **Continue to support Copper for short Scale-up links**

Conclusions

As we look ahead, the following will likely be required:

- Faster deployment and faster scaling
- Reliable interconnects
- Resilient supply chain
- Enable both CPC and CPO for nextGen AI networks

THANK YOU