SNIA | Data, Storage & Networking

Accelerating AI Infrastructure: The Role of 400G and PCIe 8.0 in Next-Gen Interconnects

Live Webinar
January 29, 2026
10:00 am PT / 1:00 pm ET

Today's Presenters

Erik Smith
Distinguished Engineer
Dell Technologies
Moderator

Anthony Constantine
Distinguished Member of Technical Staff
Micron Technology

Tom Palkert
Consultant
Samtec

The SNIA Community

200
industry leading organizations

2,000
active contributing members

50,000
IT end users & storage pros worldwide

SNIA | DSN A SNIA Community

SNIA | Data, Storage & Networking

What We Do

Drive the awareness and adoption of a broad set of technologies, including:

- Storage Protocols (Block, File, Object)
- Traditional and software-defined storage
- Disaggregated, virtualized and hyperconverged
- AI, including storage and networking considerations
- Edge implementation opportunities and factors
- Storage and networking security
- Acceleration and offloads
- Programming frameworks
- Sustainability

How We Do It

By delivering:

- Expert webinars and podcasts
- White papers
- Articles in trade journals
- Blogs
- Social Media
- Presentations at industry events

Logistics

- The slides are available under the attachments tab at the bottom of your console.
- Questions are welcome!
- Please rate the session and provide feedback!
- Want more sessions like this or other topics, let us know!
  - JOIN US! We meet on Thursday mornings at 11:00 AM eastern.
  - Email dsn-chair@snia.com if you have questions.

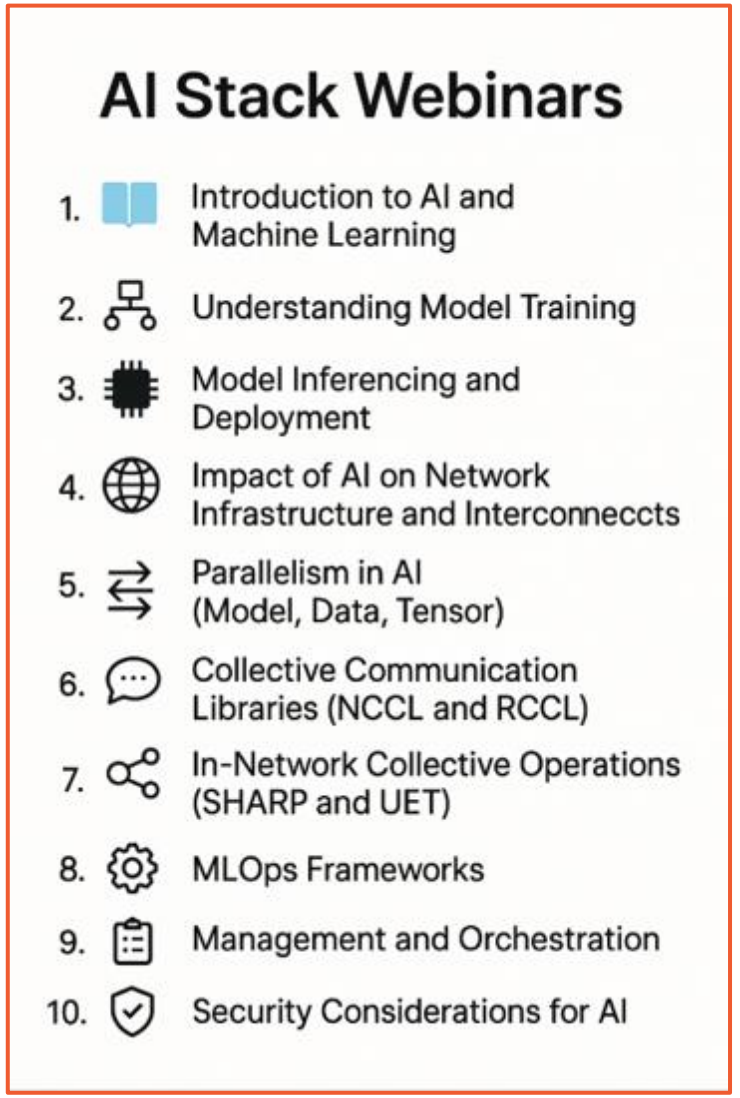
SNIA Legal Notice

- The material contained in this presentation is copyrighted by SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

The “AI Stack” Webinar Series

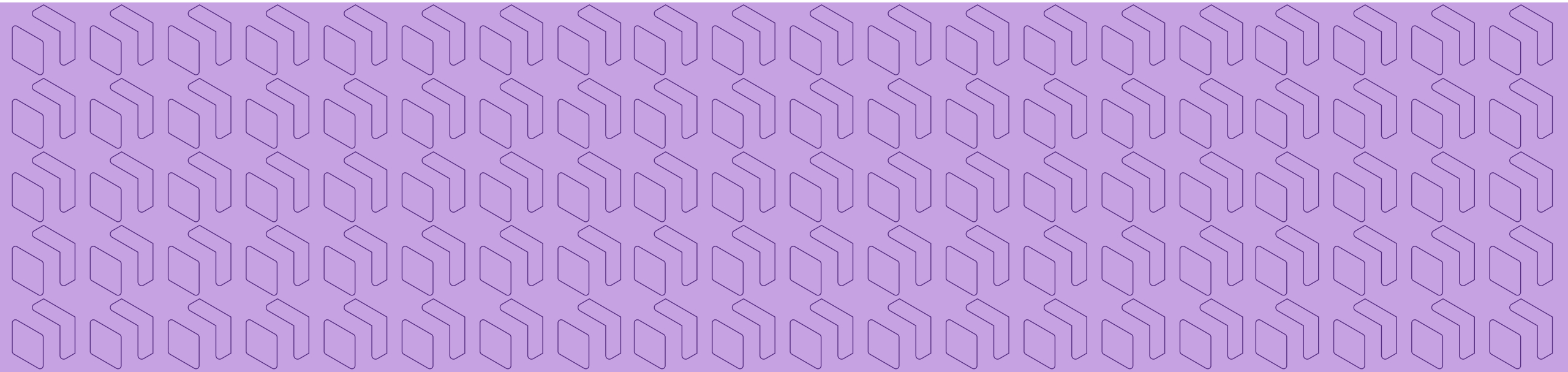
- Building a Strong Foundation for All Experience Levels:
  - Starting from the basics
  - Building steps-by-step
  - Connecting theory to practice
  - Demonstrations
  - Preparing for real-world challenges



Agenda

- Introduction into Physical Infrastructure
- Copper connections and the challenges to 400G
- PCIe Cabling to SSDs and the road to PCIe 8.0
- The inflection point of Copper and Optical
- AI needs of SSDs

Introduction into Physical Infrastructure

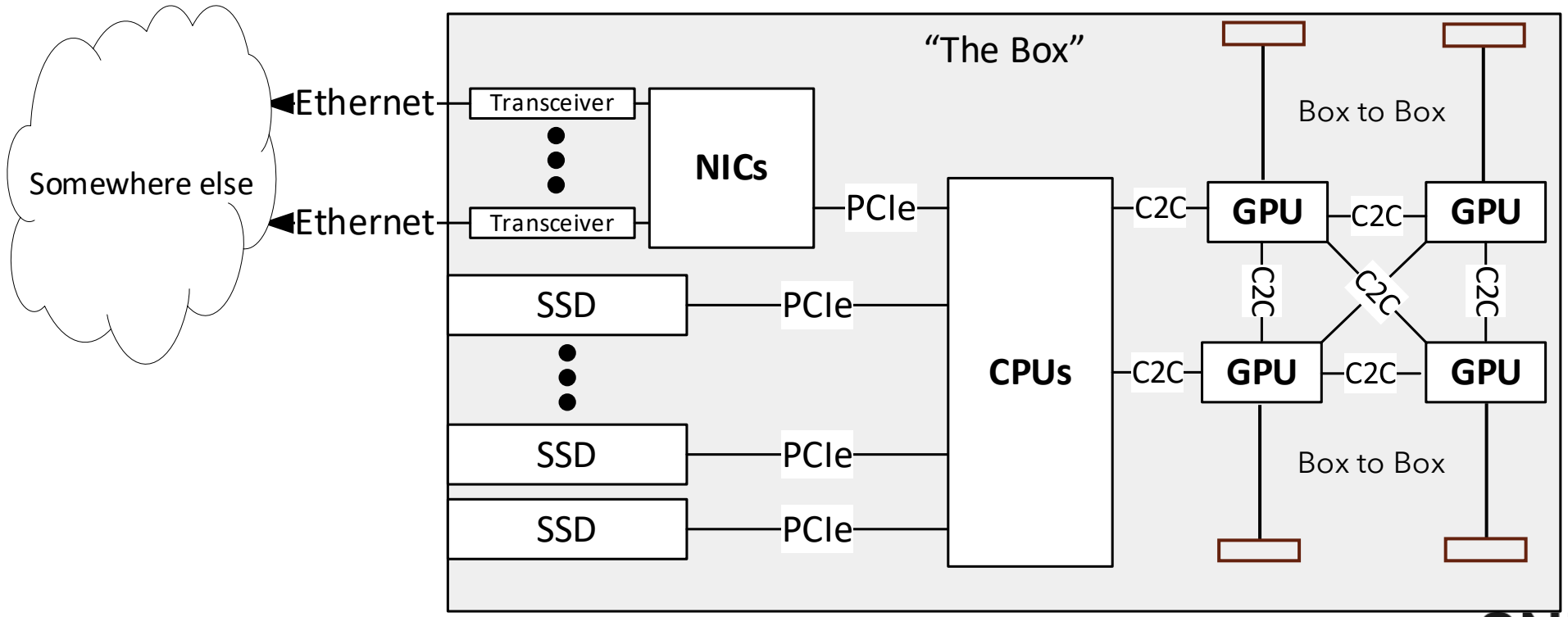


First off, why faster?

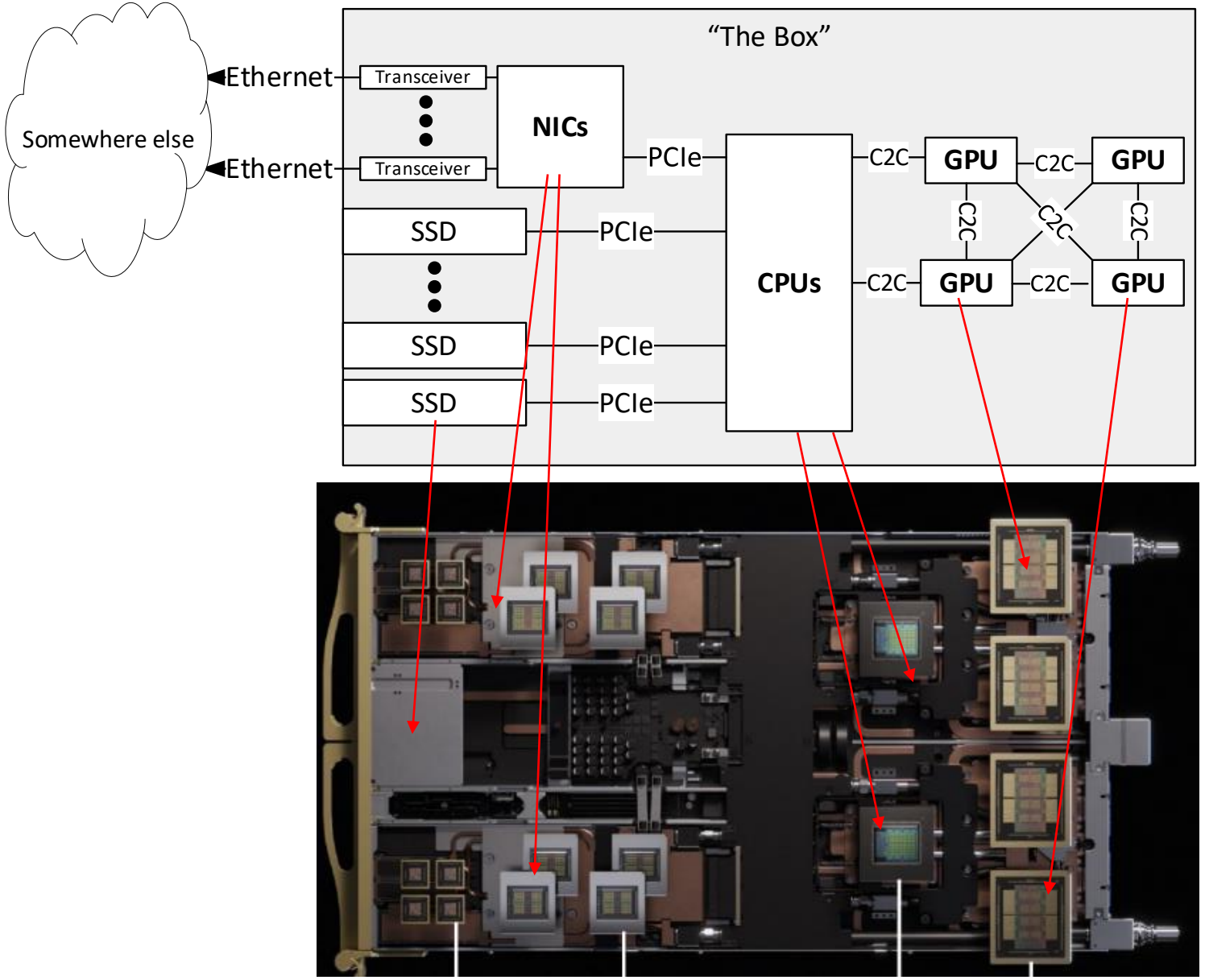
- Observing two major trends in storage for AI:
  - Storage provides supplemental memory capacity
  - A lot of cold data is warmer
- What this means:
  - Lots of data needs to be moved rapidly outside the box
  - Lots of data needs to be moved rapidly within the box
  - Storage needs to keep up with this data movement
- The interconnects need to keep up
  - Not keeping up means data is not getting acted upon

What is this Box?

- The Overly Simplified AI Box (example)
  - Out of box: Ethernet
  - In the box: PCIe and Chip to Chip (C2C) interconnect



Box in the Wild



SNIA | DSN A SNIA Community

Copper connections and challenges of 400G

Copper connections and the challenges of AI

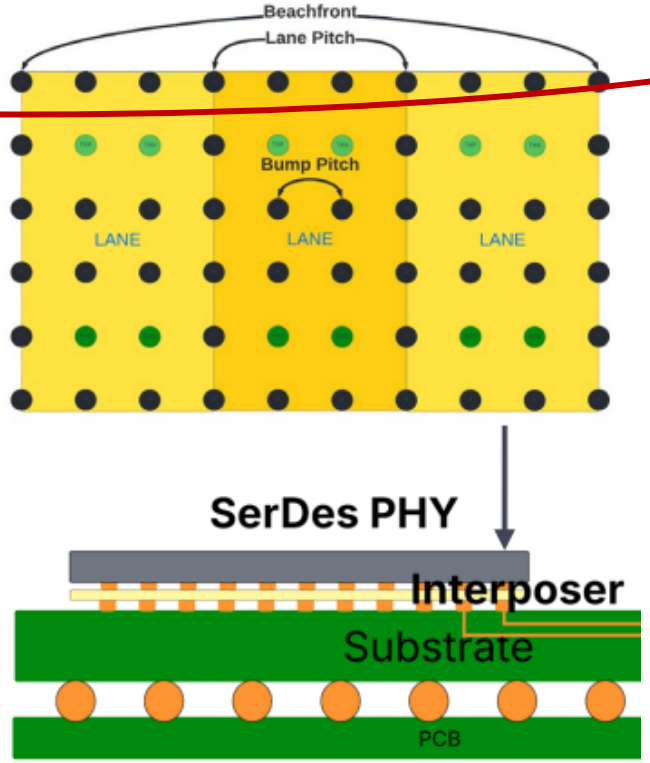
- Why 400G?
- The basics
- It’s all about the channel
- Channel breakdown
- Models and measurements
- COM decisions
- Future work

- Note: Opinions are my own

Why 400G?

- Higher bandwidth, lower latency of AI systems

Beachfront is the major limitation on scaling the Network IO BW

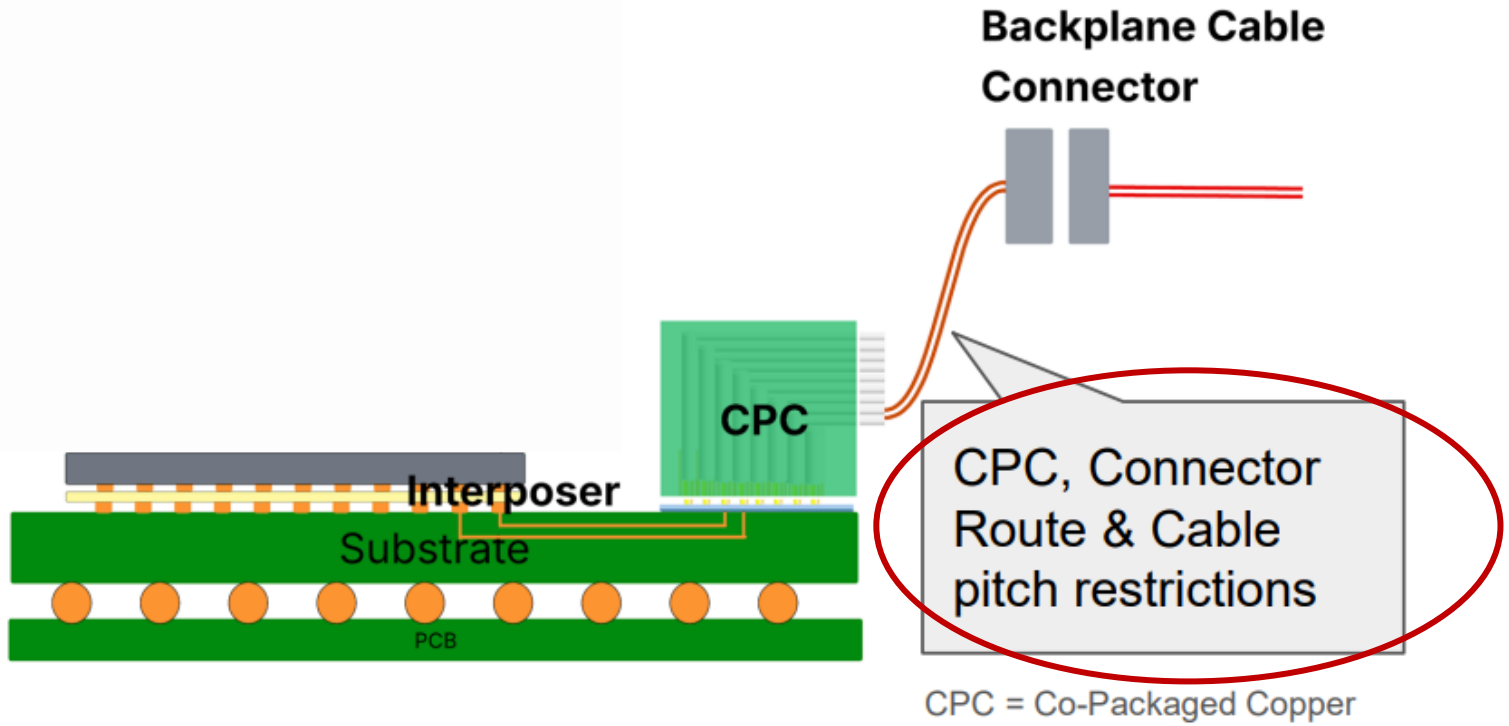


Halil Cirit (Meta) EA TEF 2026

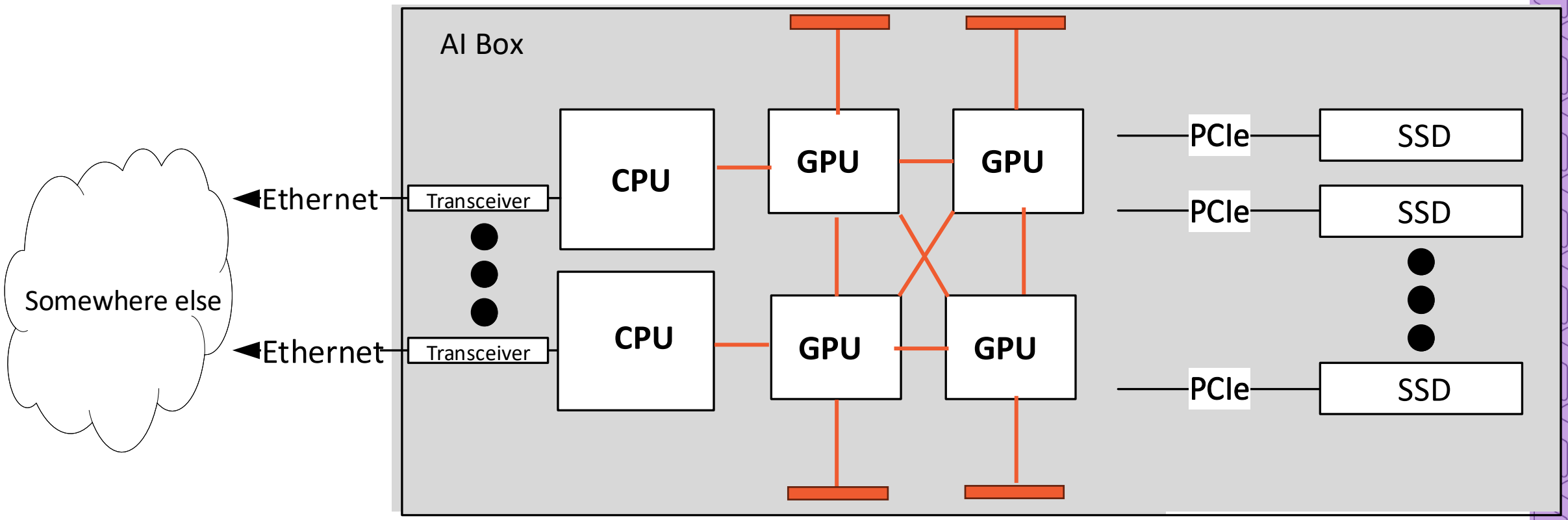
Why 400G?

- Higher bandwidth, lower latency of AI systems

Halil Cirit (Meta) EA TEF 2026



AI design



The basics

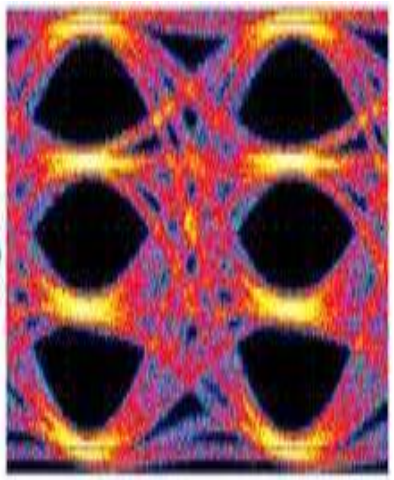
- How to design a high speed copper interconnect
  - Modulation/DSP performance
  - Package
  - Breakout region
  - Trace routing
  - Connector
  - Cables

Modulation/DSP performance

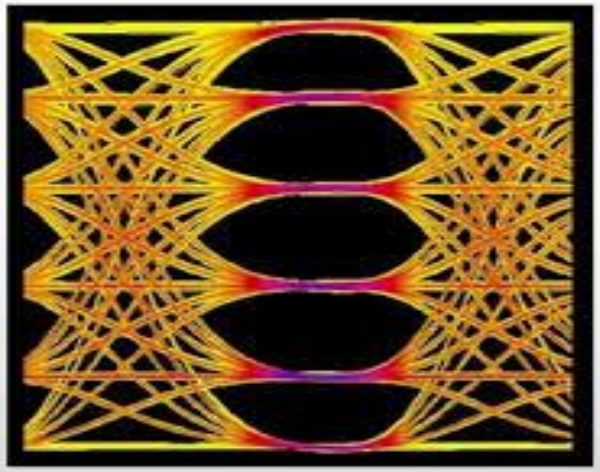
Data Rate	448G			
Signal levels (PAM)	8	6	4	
Bits per symbol	3	2.5	2	
Signaling rate	150	180	224	GBd
Nyquist Frequency	75	90	120	GHz
SNR reduction	6.3	3.7	0	dB

My opinion: DSP can do any PAM modulation for 400G

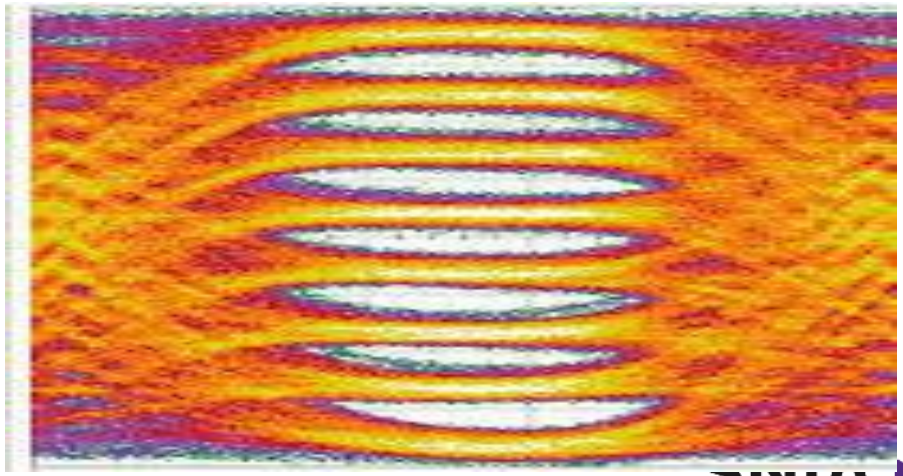
PAM4



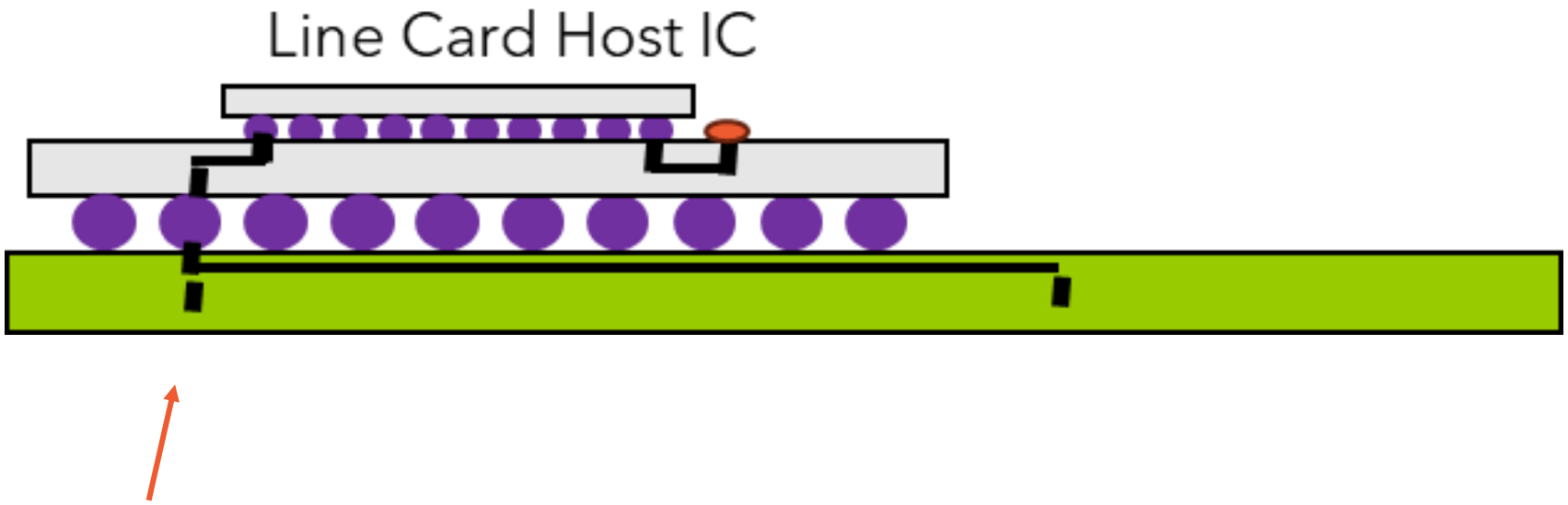
PAM6



PAM8



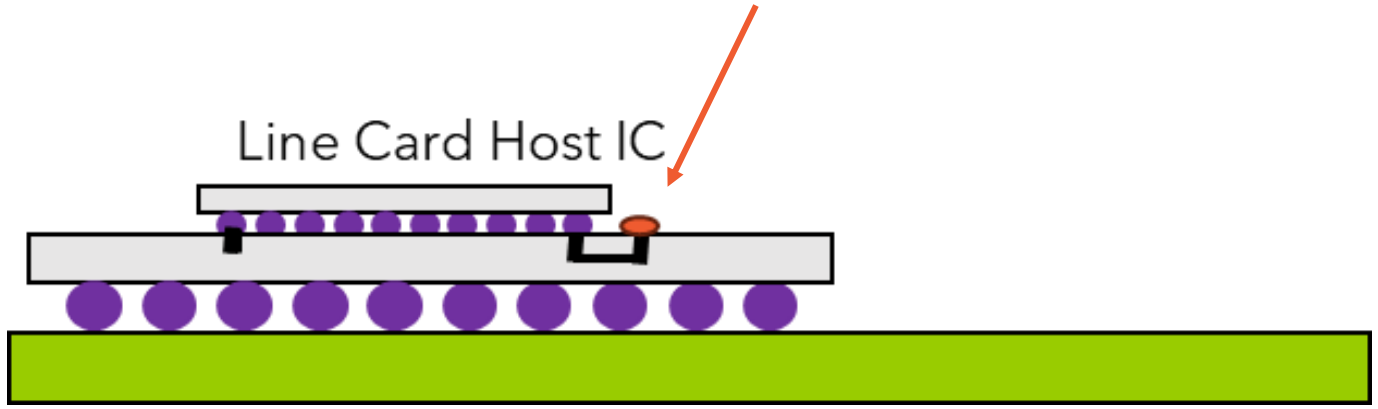
Package and Breakout region



Standard Package traces and PCB breakout create SI challenges

Package and Breakout region

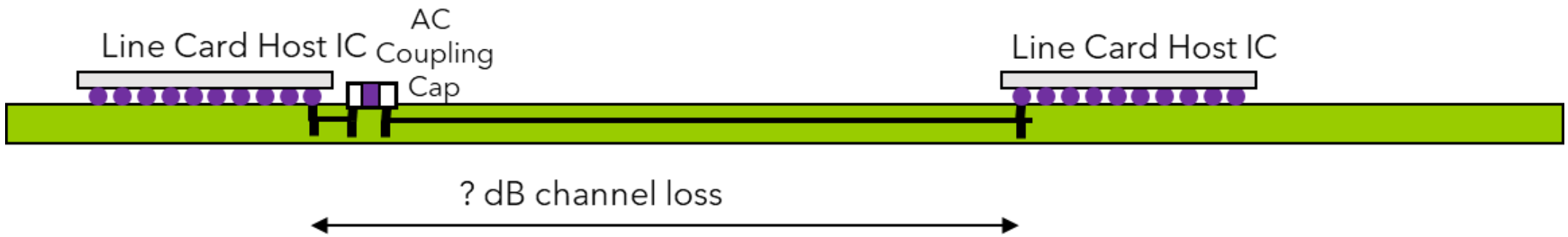
Co-Packaged Copper (CPC) eliminates PCB breakout region giving reduced IL and better SI.



My opinion: 400G will require better package materials and CPC routing

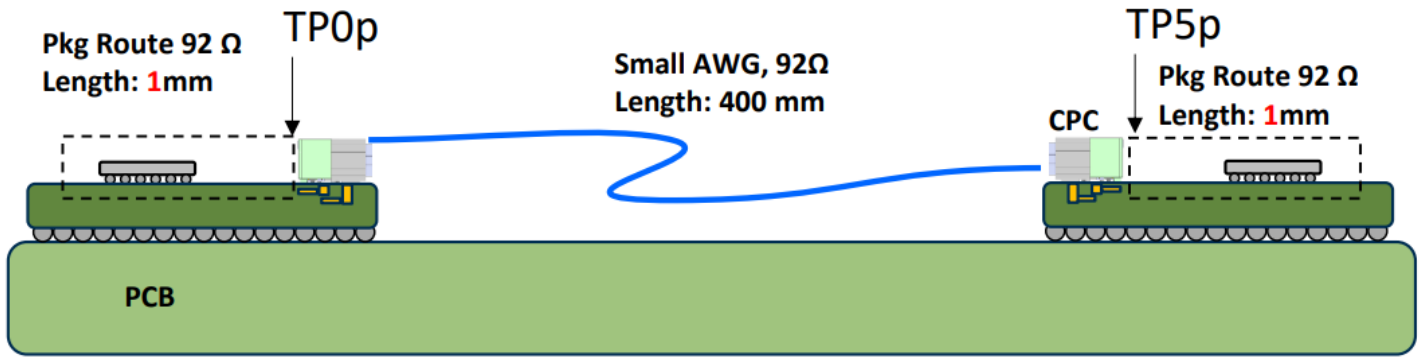
Trace routing

- Traditional specs are IL (trace width, surface roughness) and Via stubs (Reflections)



Trace routing

- Flyover cables are a much better solution
  - Lower loss
  - Better SI



My opinion: Traditional PCB routing will not work at 400G, skew will be important

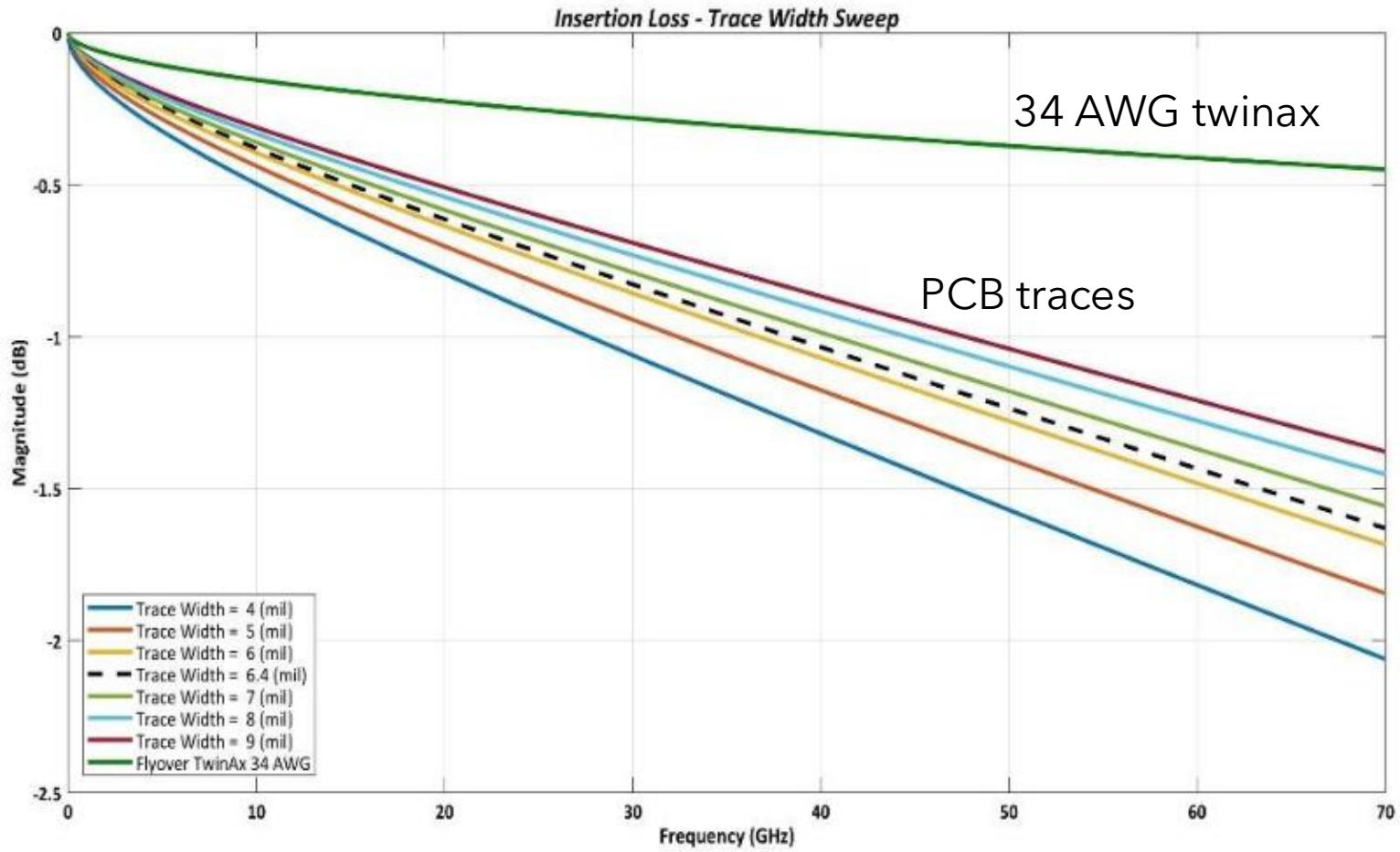
Connectors

- Connector BW is key and appears to be the biggest issue with determining the optimum modulation.
  - IL, Crosstalk

No connector S parameter data that I can share

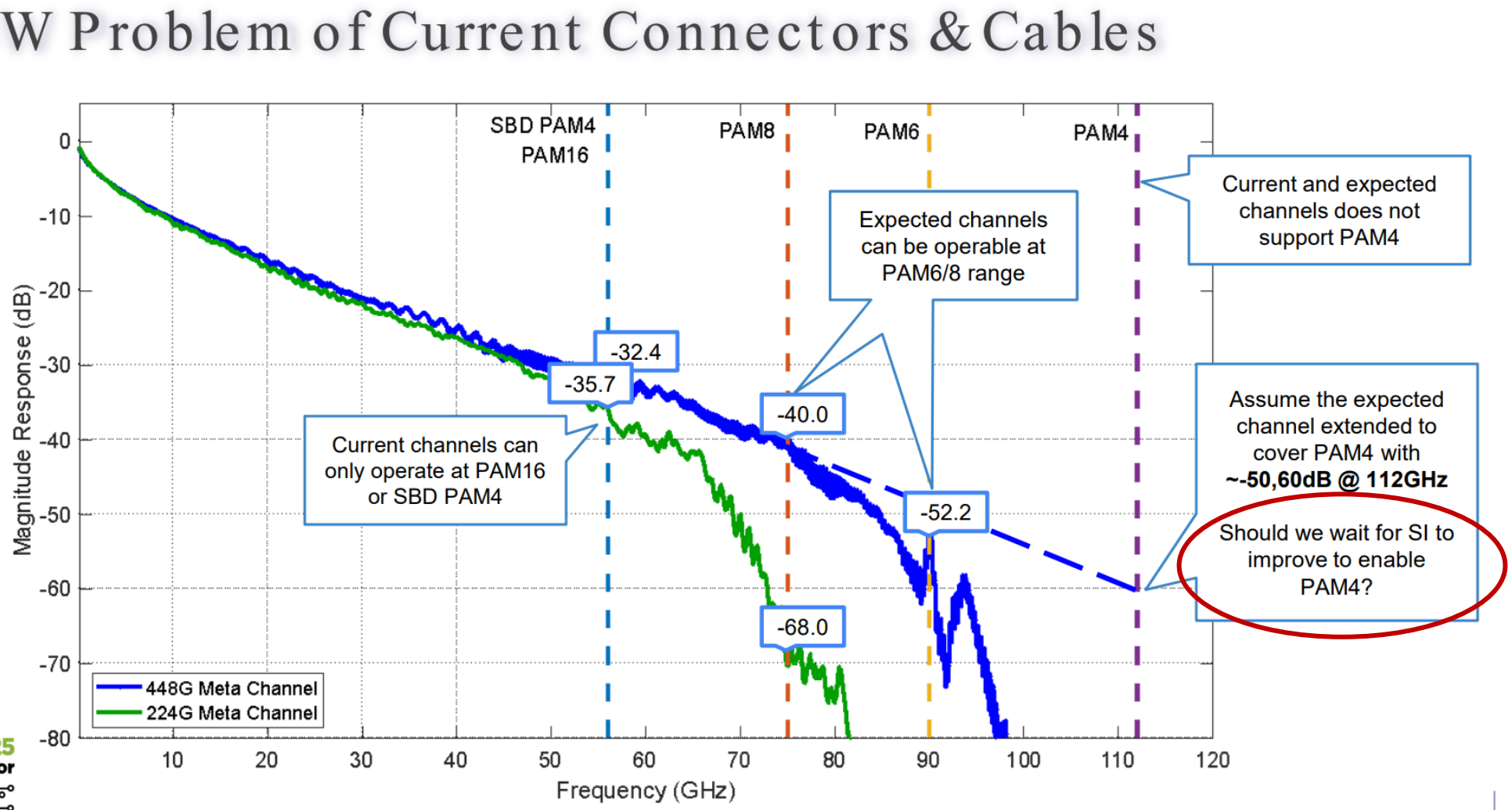
Cables

- Cable loss is highly dependent on wire gauge
- Cable loss is generally linear with good Signal Integrity



Its all about the channel (Halil Cirit (Meta) EA TEF 2026)

- My opinion: We cannot wait for improved channels

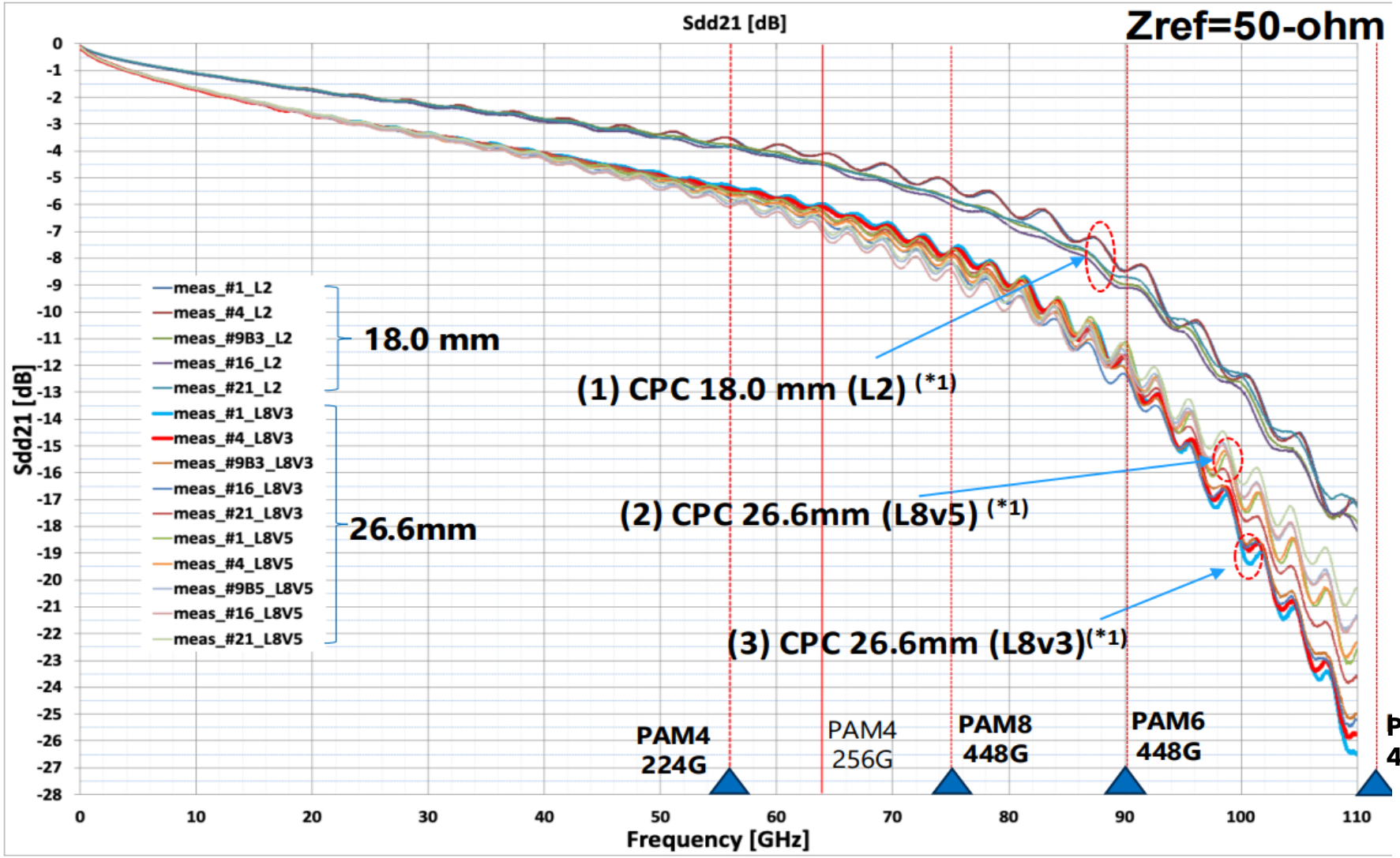


400G Channel breakdown

- Die, Package, CPC, internal cable, connector, external cable
  - Die: Internal AC coupling, low parasitics, 92 ohm terminations
  - Package: Improved materials, Trace design, CPC pads
  - Flyover cables with CPC: IL, CPC pads and cable contacts
  - Backplane/front panel connector: IL and crosstalk (Tradition front panel modules using PCB pads will not work)

Package

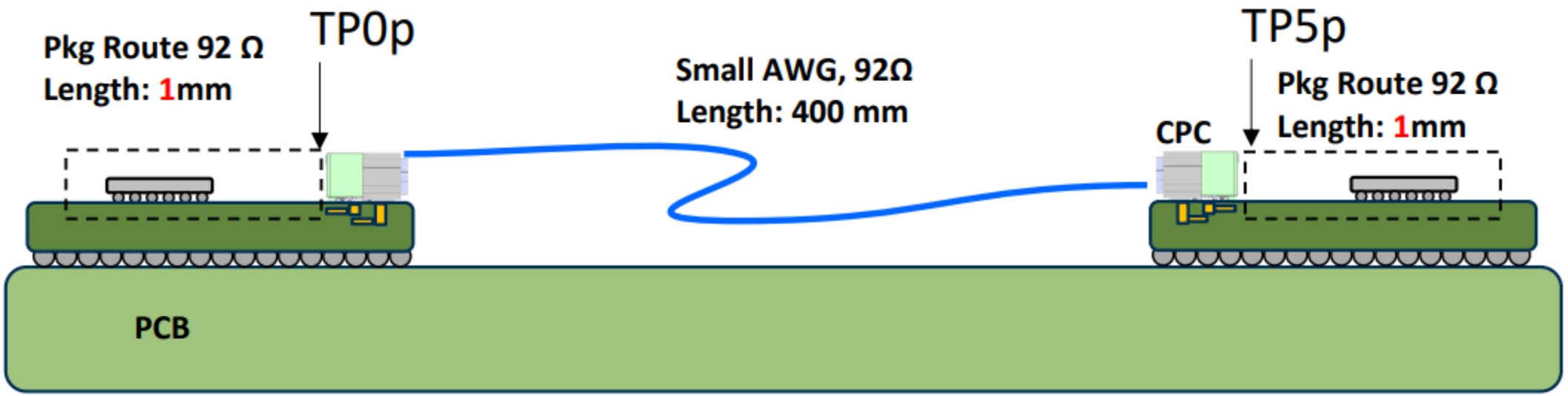
4.1 CPC: Insertion loss Sdd21



Package loss with existing materials can work for some channels but longer reach copper interconnects will benefit from lower loss packages

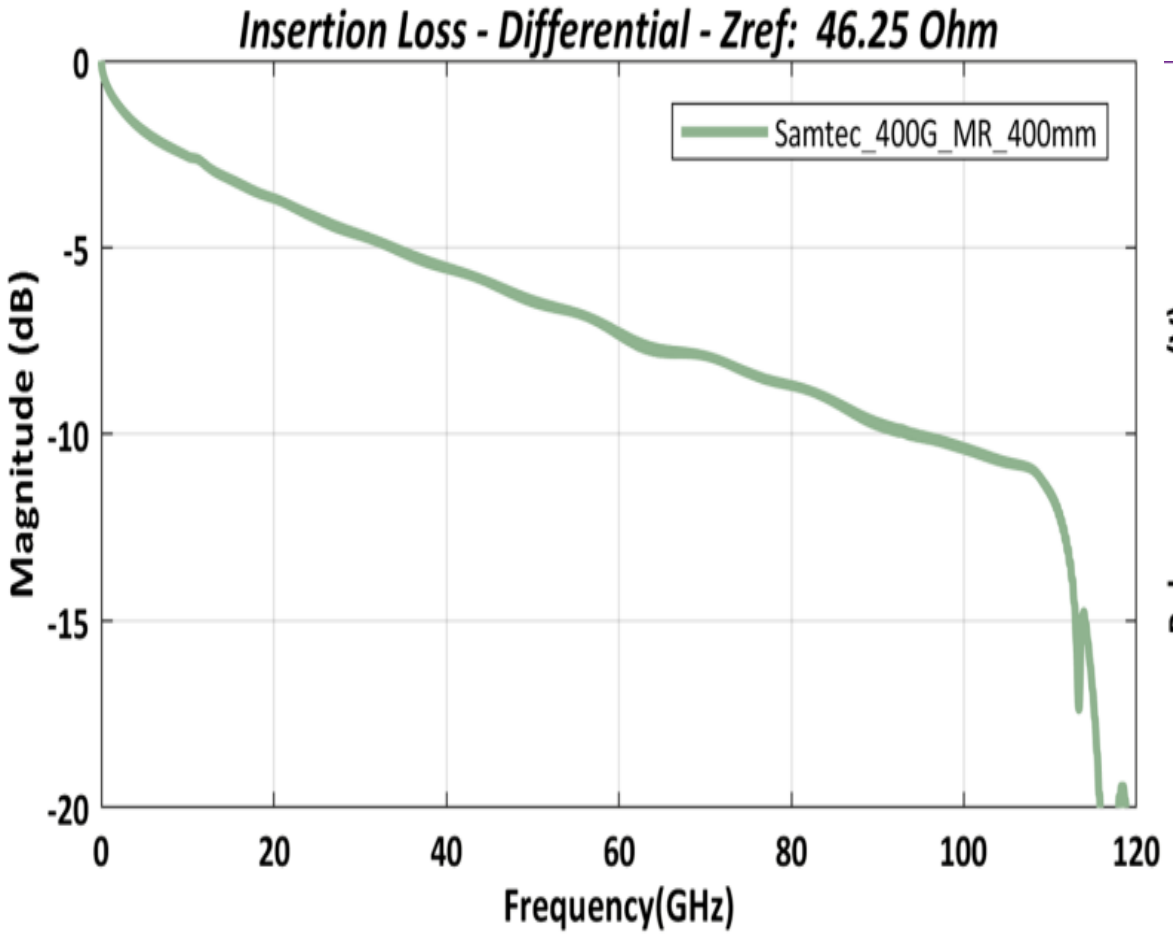
CPC and flyover design

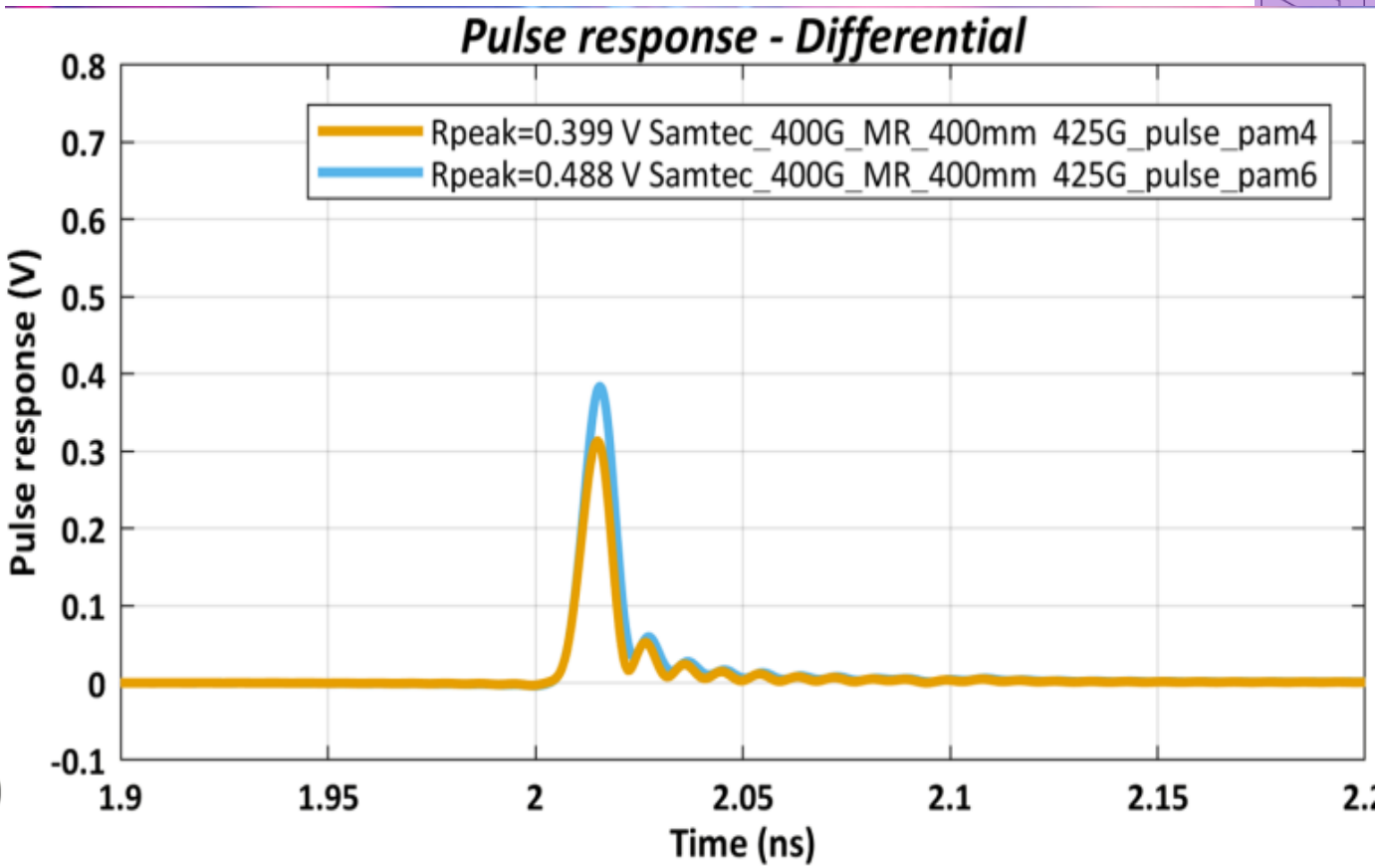
- Best SI possible



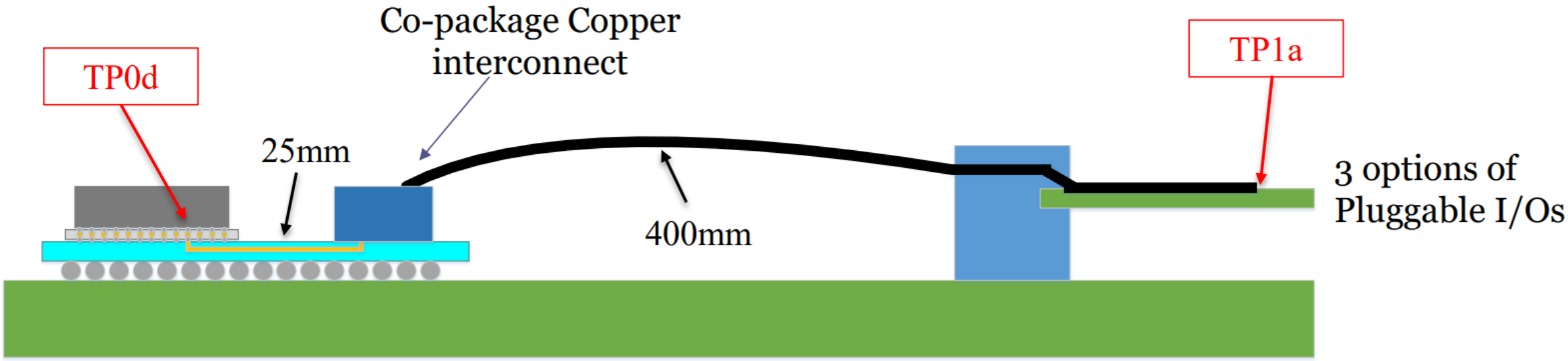
Insertion Loss and pulse response of Chip to Chip channel

- My opinion: PAM6 is the best modulation for 400G Chip to Chip channels





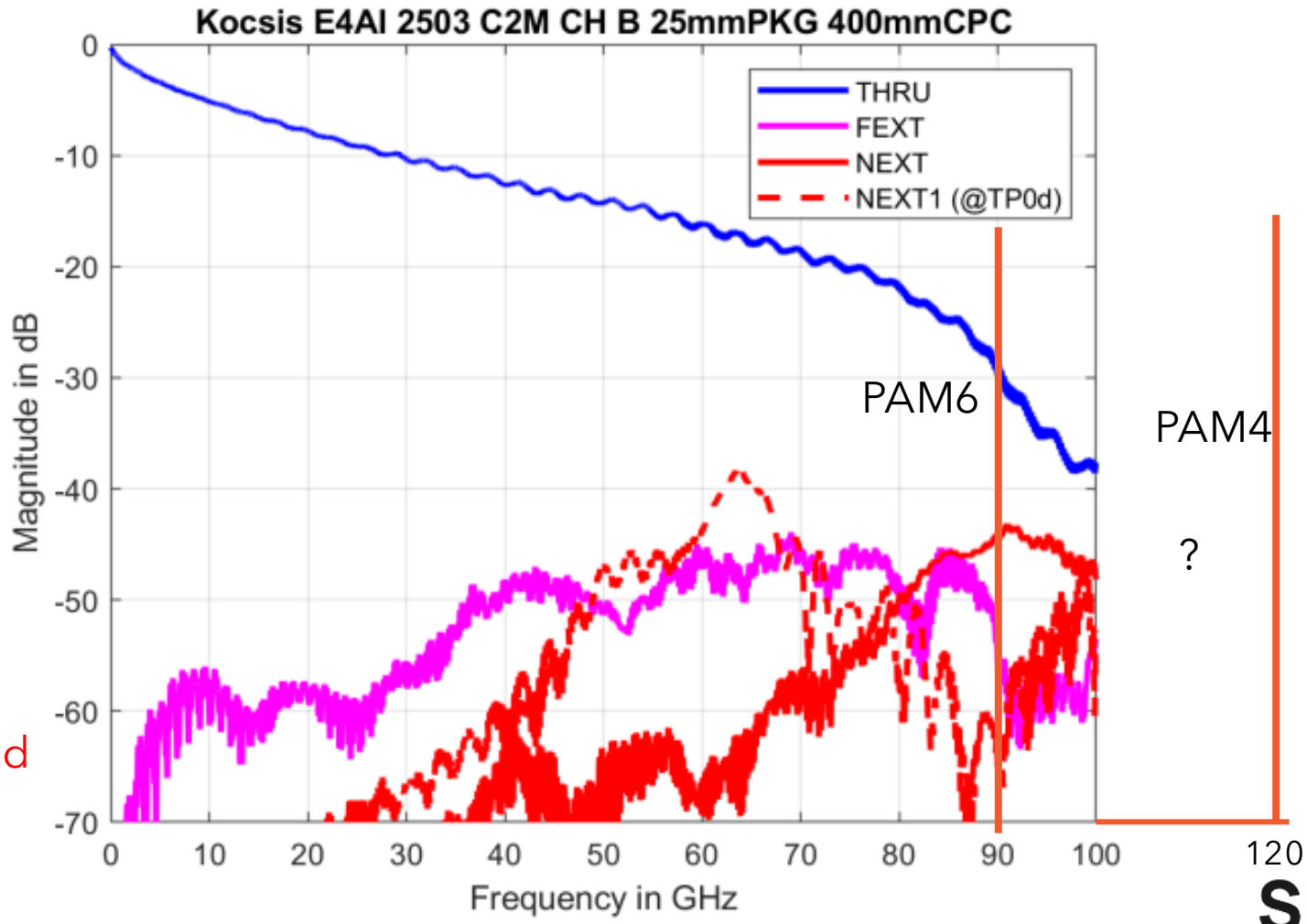
Front Panel Pluggable



Courtesy Amphenol

Backplane channel performance

- My opinion: PAM6 is the best modulation for 400G Chip to Module channels

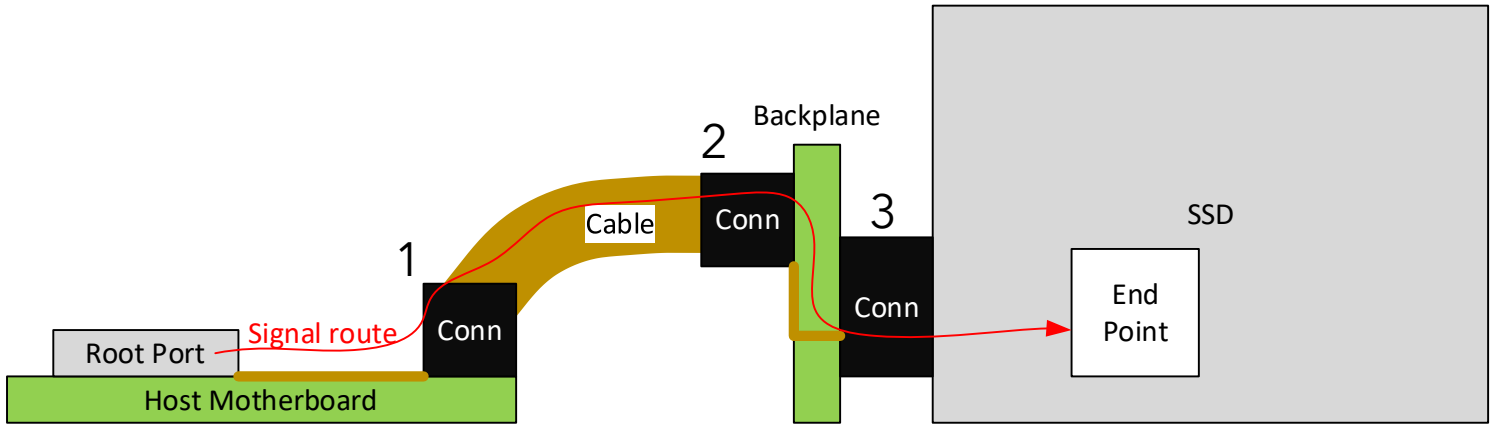


Courtesy Amphenol
PAM4/6 Red lines added

PCIe Connectors and Cabling to SSDs up to PCIe 8.0

In Box Connectors and Cabling

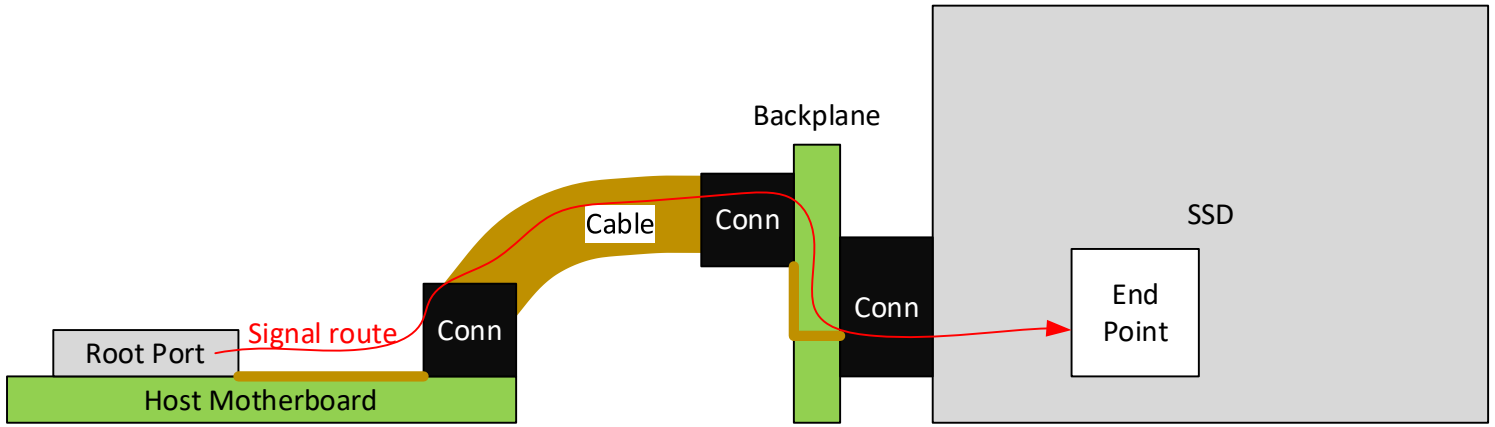
- SSDs are generally not close to the host silicon (PCIe Root Complex)
- Traditionally, SSDs were connected to the host through a “3 connector” topology.



Signaling Budget: Insertion Loss

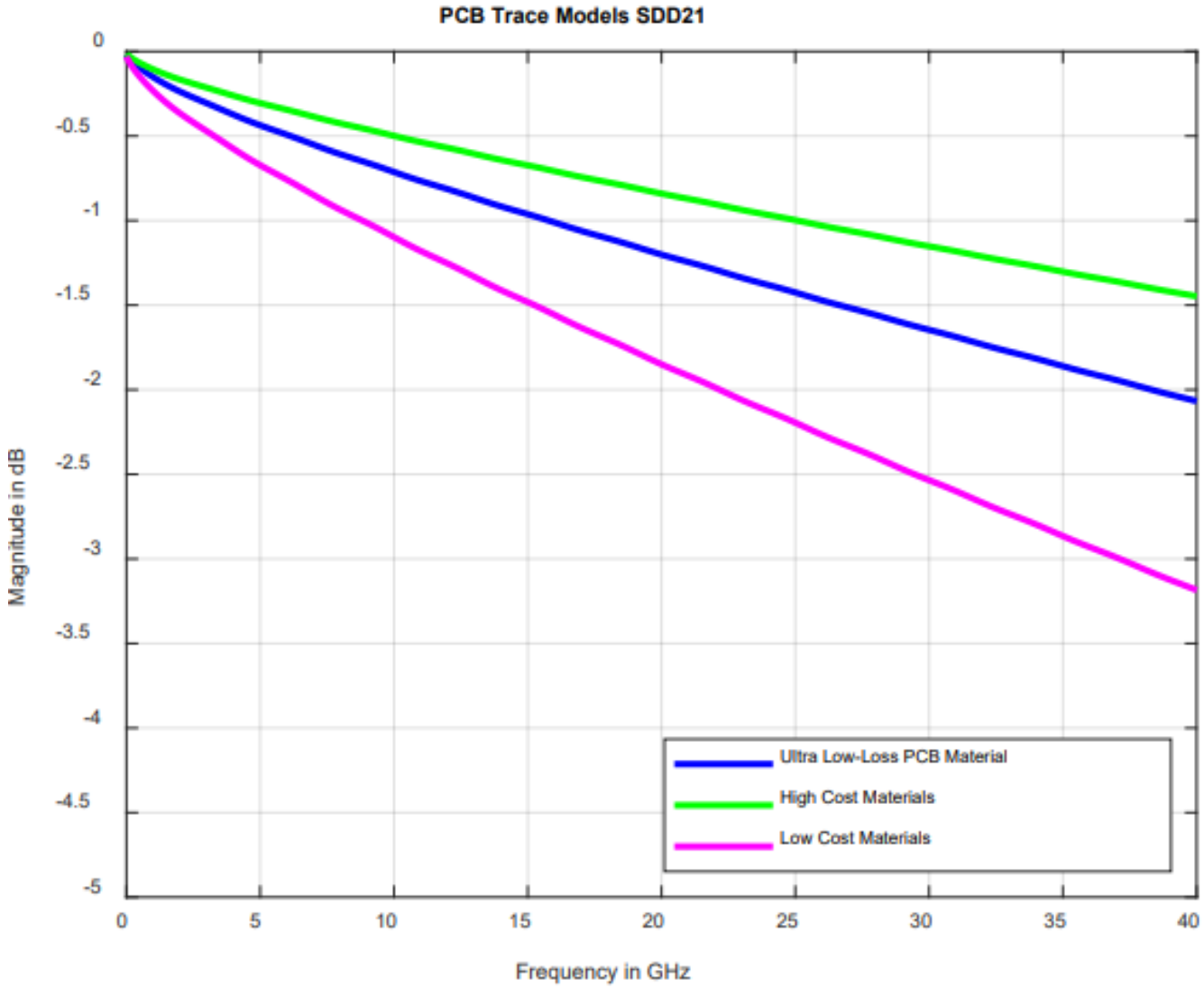
- A way to approximate how far you can connect between Root Port and end point is insertion loss.
  - If the channel loss is greater than the target, then one or more mitigations:
    - Reduce channel length (if possible)
    - Active components needed (e.g., retimer)
    - Lower loss PCB materials
  - For PCB's at 16 GHz:
    - Cheap: ~1.5 dB/in
    - Less cheap: ~1 dB/in
    - Not cheap: ~0.7 dB/in

Spec	Insertion Loss Budget			
Pcie revision	5.0 (32GT/s)	6.0 (64 GT/s)	7.0 (128 GT/s)	8.0 (256 GT/s)
Nyquist	16 GHz	16 GHz	32 GHz	64 GHz
Insertion Loss target	**36 dB**	**32 dB**	**36 dB**	TBD
Root Port	9 dB	8 dB	9 dB	
SSD	7 dB	6 dB	7 dB (est)	
Everything Else	**20 dB**	**18 dB**	**20 dB**	
1M mated cable	7.5 dB	7 dB	Tbd	
SSD connector	1 dB	0.75 dB	Tbd	
PCB routing budget	**11.5 dB**	**10.75 dB**	**Tbd**	



Challenge: Insertion loss at 32 GHz+

- At 32 Ghz, PCB routing insertion loss is ~70% worse
  - 0.7dB/in->1.2dB/in,
  - 1dB/in->~1.7dB/in
- At 64 GHz, this becomes much larger
- Note: Copper cable insertion loss is about 5-10x better but also degrades over frequency

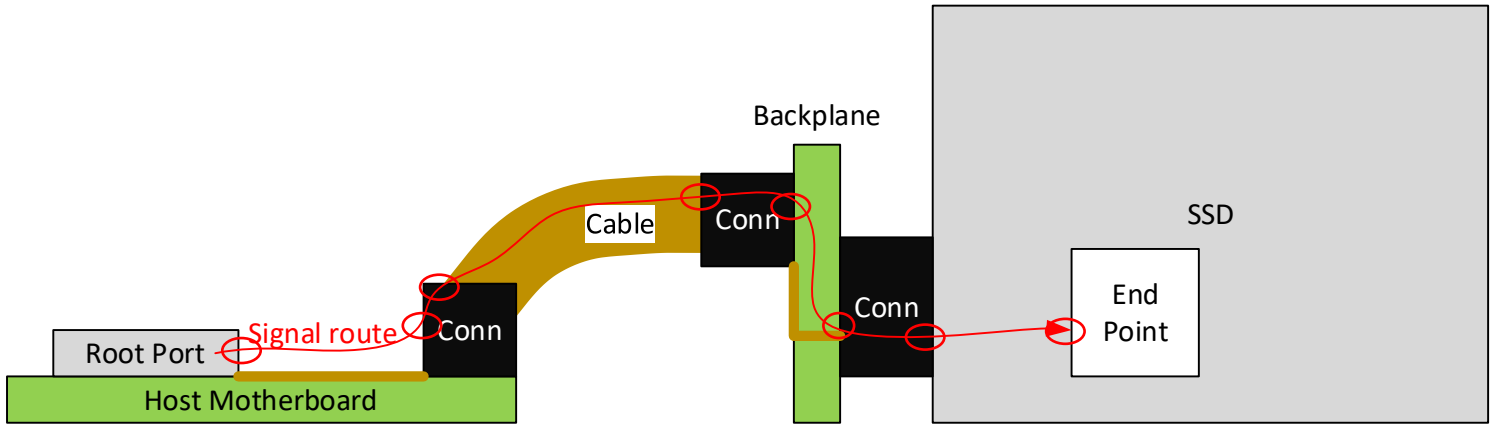


Credit: Mohiuddin Mazumder
PCIe Cabling Solutions for Next Generation AI, HPC and Storage Applications.

SNIA DSN A SNIA Community

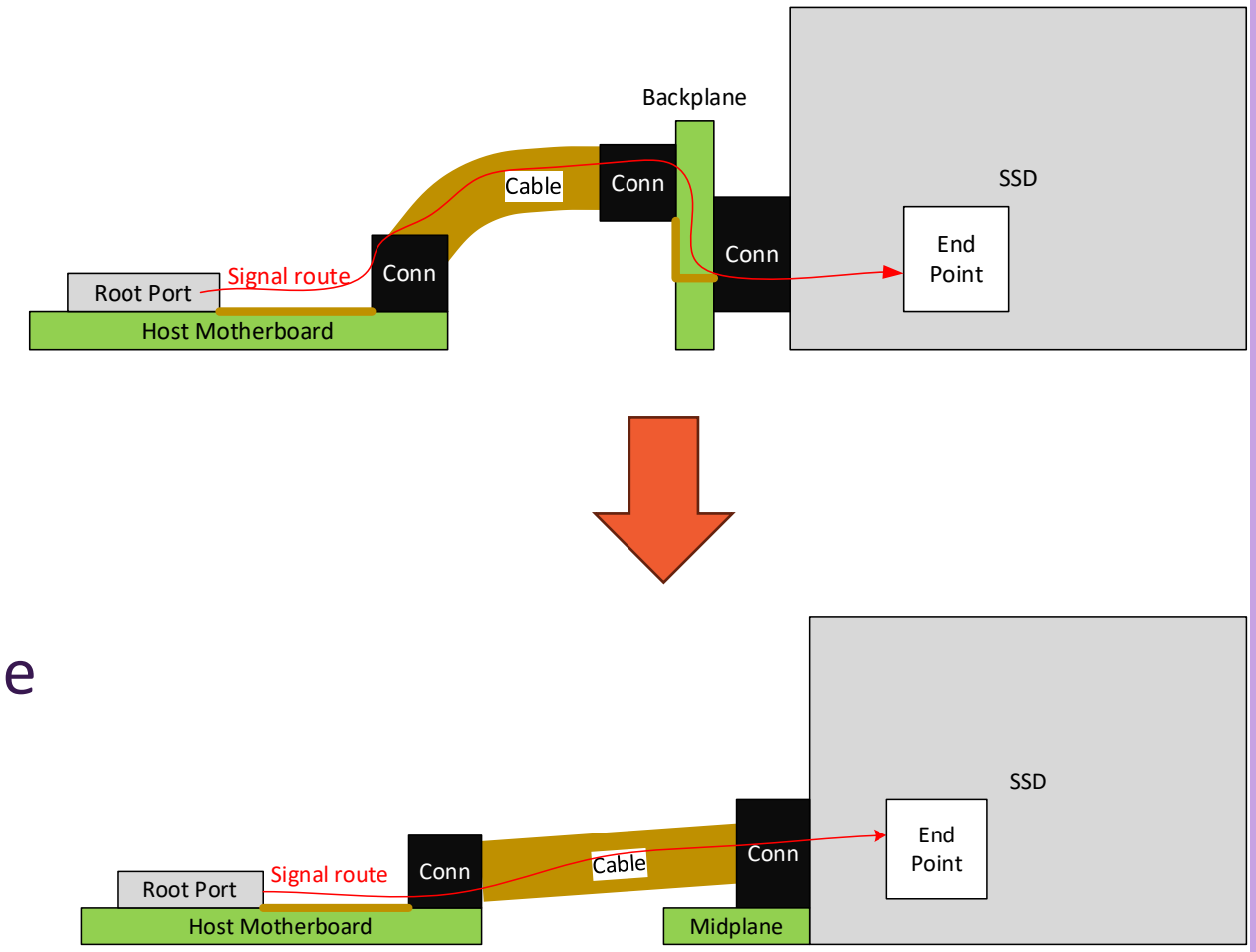
Additional Challenge: Return Loss

- Return loss is the reflected power that did not make it to its destination
- Every impedance discontinuity is a point of loss.
  - Via transitions
  - Coupling caps
  - Connectors
  - Etc.,
- This also becomes a larger problem at higher speeds



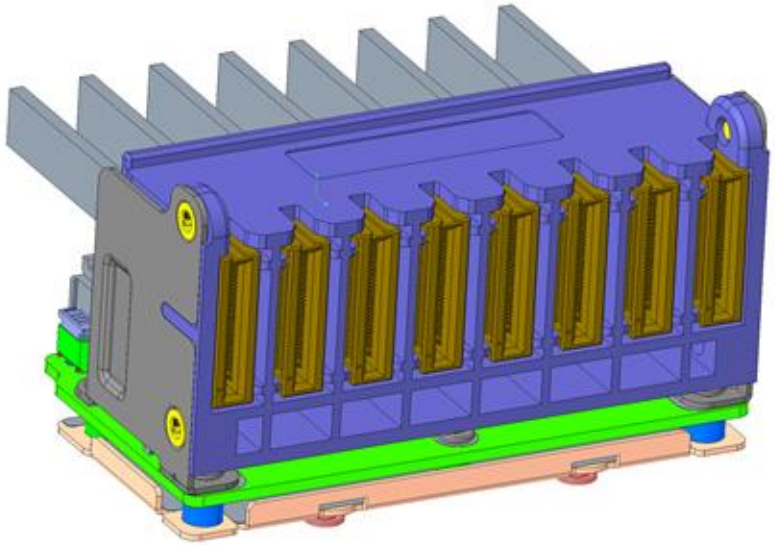
Resolving these Challenges: 3->2 connector topology

- Insertion Loss reduction
  - Backplane PCB routing eliminated
  - 1 connector eliminated
  - Lower profile connectors help with reduce routing length.
- Return loss reduction
  - 1 connector eliminated (2+ reflection points)
  - Backplane eliminated (no transition via, 1-2 reflection points)
- This is already implemented for some PCIe 5.0 and PCIe 6.0 systems
  - Hybrid cables

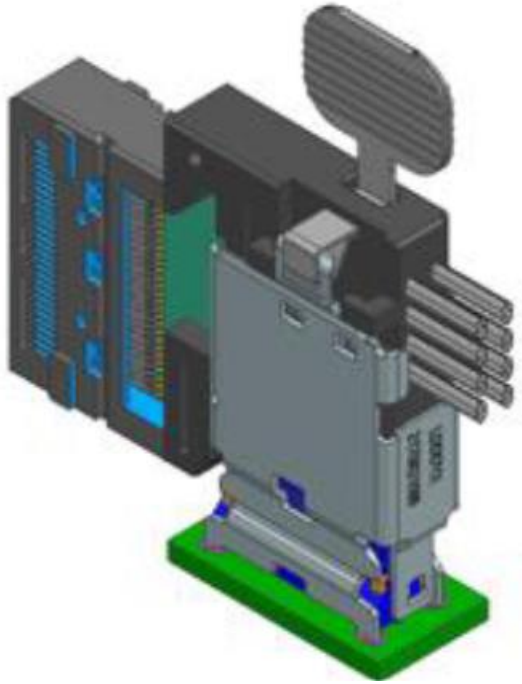


Hybrid Cables

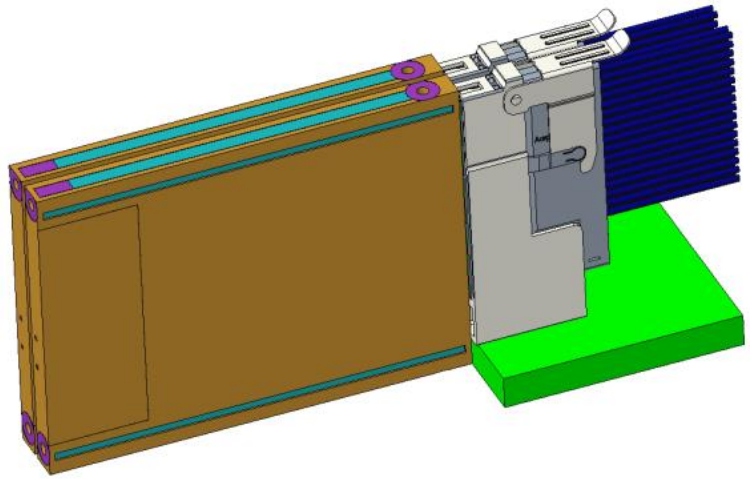
SFF-TA-1016



SFF-TA-1035



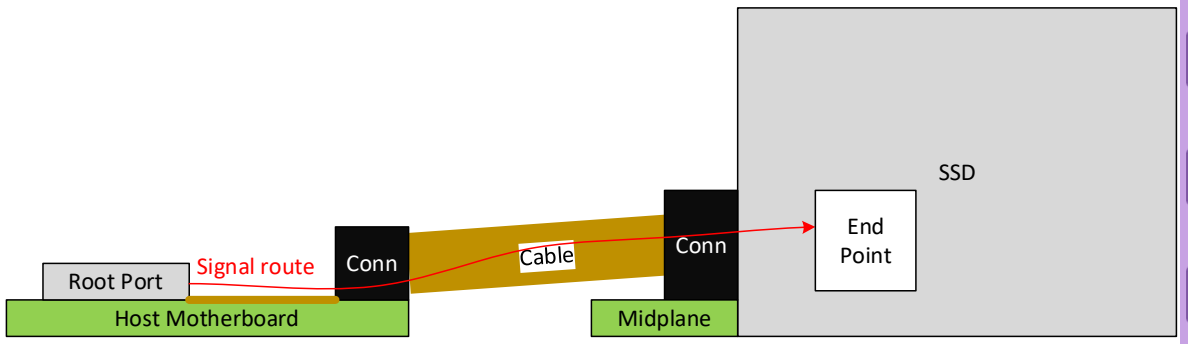
SFF-TA-1044
(under development)



So, all problems solved?

- For PCIe 7.0, yes and no
  - With “not cheap” materials yes.
  - A retimer may still be needed in some cases.
- For PCIe 8.0, active components required.
- Some enablers to explore:
  - Active copper cables
    - Retimers at one or both ends of cable
  - Optical cables (copper interconnect)
    - Direct Drive ECN (PCI-SIG)
  - “Not not cheap” materials

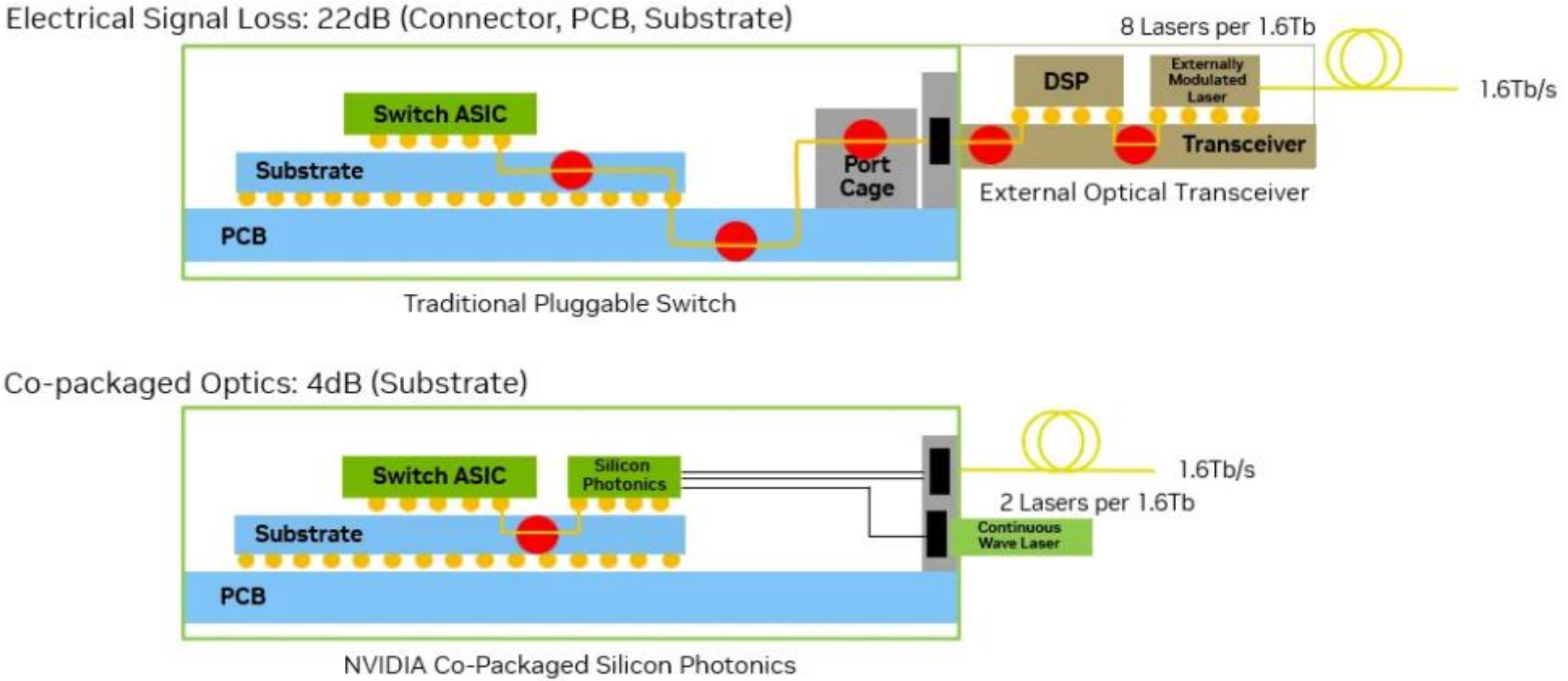
Spec	Insertion Loss Budget			
Pcie revision	5.0 (32GT/s)	6.0 (64 GT/s)	7.0 (128 GT/s)	8.0 (256 GT/s)
Nyquist	16 GHz	16 GHz	32 GHz	64 GHz
Insertion Loss target	**36 dB**	**32 dB**	**36 dB**	TBD
Root Port	9 dB	8 dB	9 dB	
SSD	7 dB	6 dB	7 dB (est)	
Everything Else	**20 dB**	**18 dB**	**20 dB**	
1M mated cable	7.5 dB	7 dB	Tbd	
SSD connector	1 dB	0.75 dB	Tbd	
PCB routing budget	**11.5 dB**	**10.75 dB**	**Tbd**	



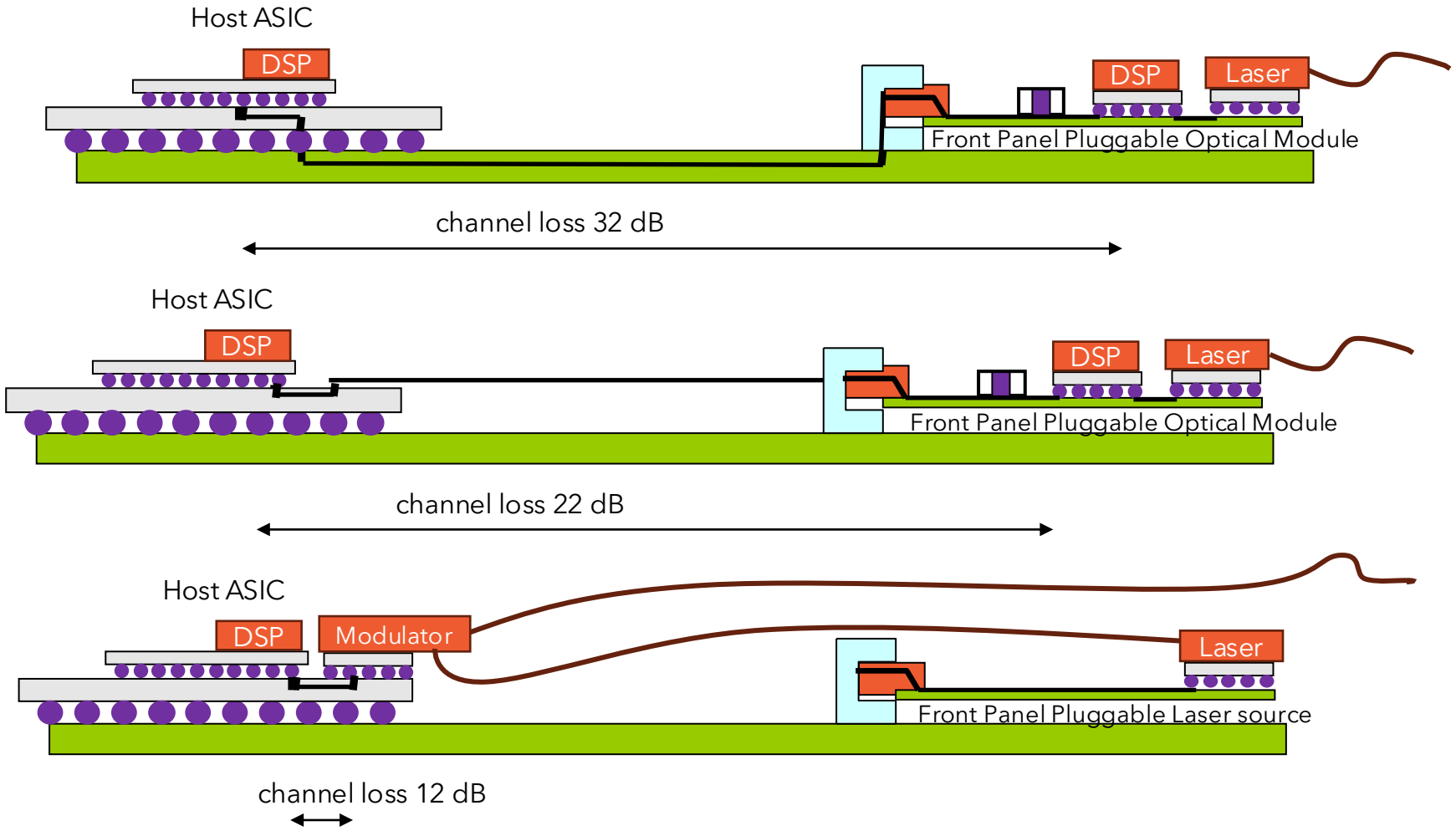
SNIA DSN A SNIA Community

The inflection point of Copper and Optical

200G AI optical links

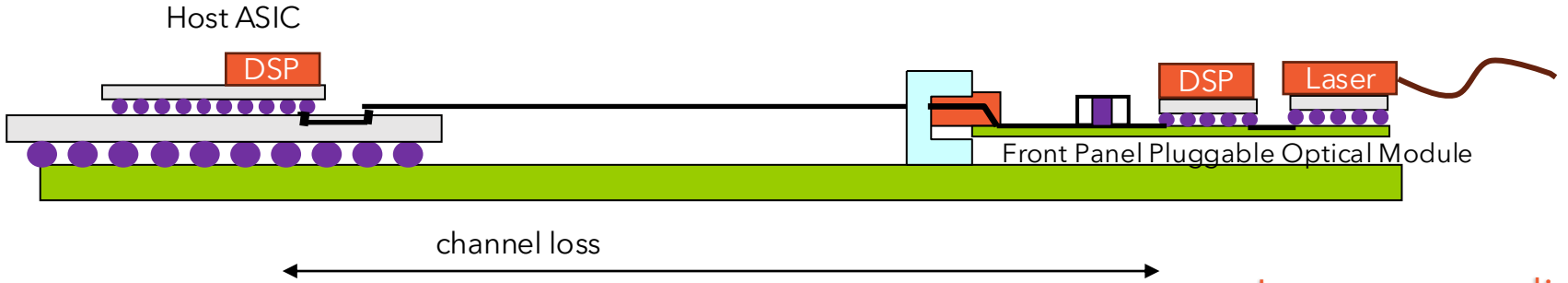


200G AI optical



Large scale adoption of AI optical

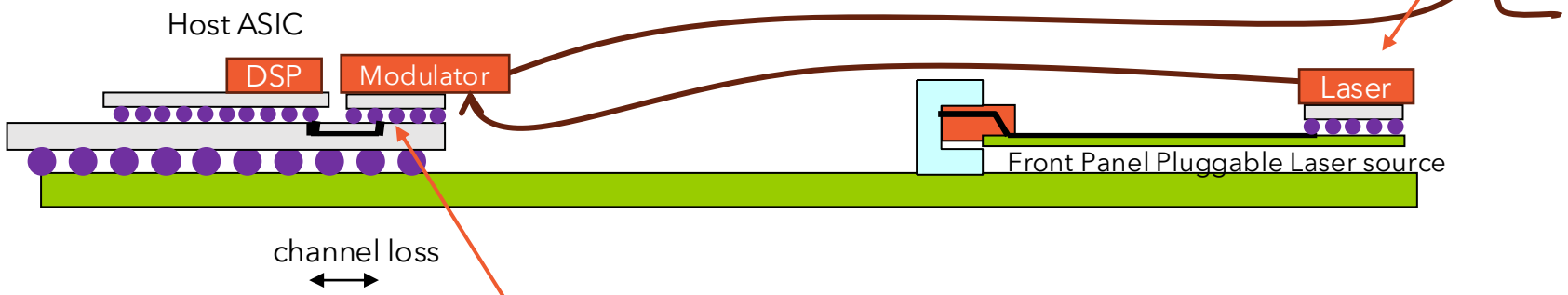
Support for front panel pluggable modules (100m-10/40km)



30W module*
$$$$$$

Laser power distribution must be improved
(support for more CPO ports per laser)

Support for low cost, low power, short reach (100m) interconnects



5W solution*
$

CPC and CPO must share common pads (power distribution to CPO needs to be solved

*Note: power and cost are based on 200G/lane 1.6Tb/s modules

SNIA DSN A SNIA Community

AI needs of SSDs

Look back at the 1st Slide

- Storage provides supplemental memory capacity
  - What this means: Higher Bandwidth and IOPS are needed.
  - This also means more power
  - This can be called an AI SSD
- A lot of cold data is warmer

The AI SSD

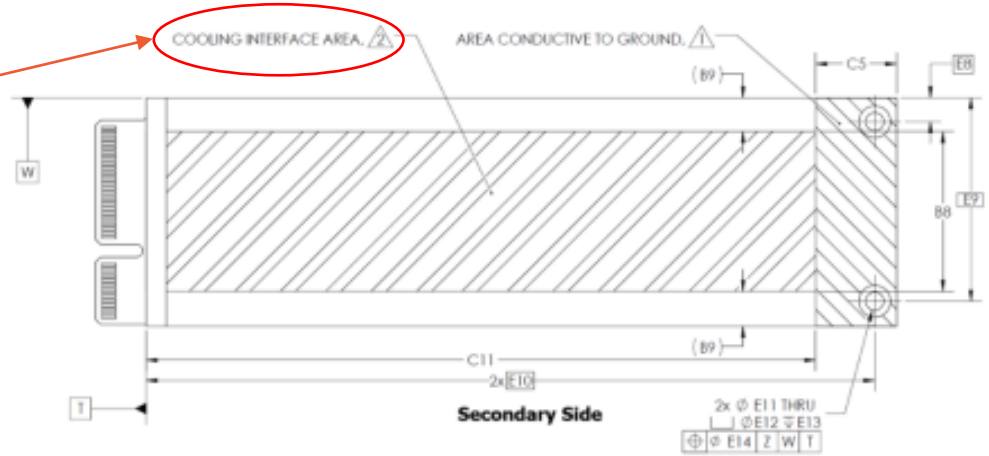
- These devices are driving storage bandwidth
- They will also consume more power
  - 2x the bandwidth on PCIe
  - Higher backend (ONFI) Bandwidth
  - Higher performance ASIC
- Power not scaling lower as fast
  - Will not stay the same but also will not double
  - Greater than 50% increase from PCIe 6->7 and 7->8
- Power is a problem we need to solve.

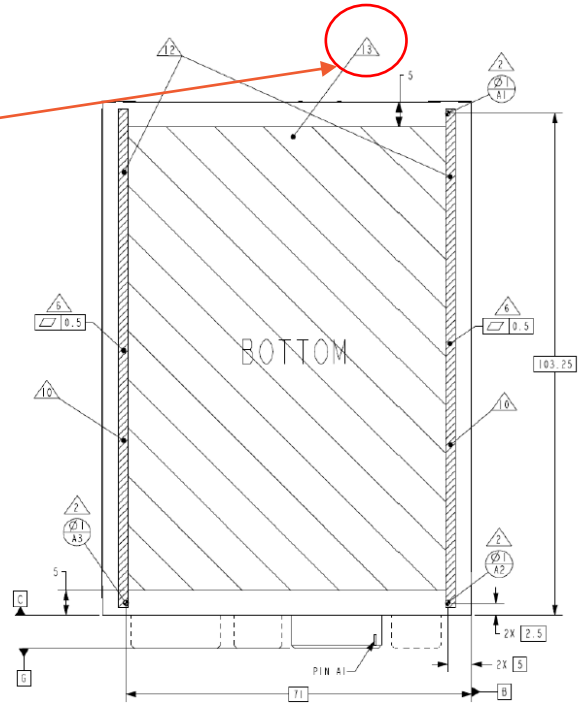
EDSFF Power Limiters in the SSD

- Connector: Allows up to ~79 W RMS
  - >79W requires connector modifications
- Form factor: Many dependencies
  - It largely comes down to maintaining enclosure temp
  - Air cooling (convection) is less efficient than a conductive cooling surface (e.g., cold plate)
  - For this reason, direct liquid cooling is starting to appear.

Work already happening here:

- SFF-TA-1006 (E1.S) Rev 2.0 defined a cooling interface area
  - Contact point for cold plate or other cooling component
- SFF-TA-1008 (E3) Draft 2.1.2 defines a similar thermal contact zone.
- These two examples show a single side of the SSD contacting with a cooling device (e.g., cold plate).





SNIA DSN A SNIA Community

Cooling Coverage

- It's ultimately dependent on SSD dimensions, flow rate (of air or liquid), and temperature that is trying to be maintained at the SSD
- Current Estimations:
  - Air should mostly be able to cool up through PCIe 6 but may be challenged to support PCIe 7 and beyond.
  - An SSD with 1 large contact surface to a cold plate will cool up through PCIe 7 but may be challenged to support PCIe 8 and beyond.
  - An SSD contact surfaces on both sides of the SSD should provide sufficient cooling benefit.

Cooling method	10W	20W	30W	40W	50W	60W	70W	80W
Air								
DLC, 1 large contact surface on SSD								
DLC, 2 large contact surface on SSD (Opposite Sides)								

Look back at the 1st Slide

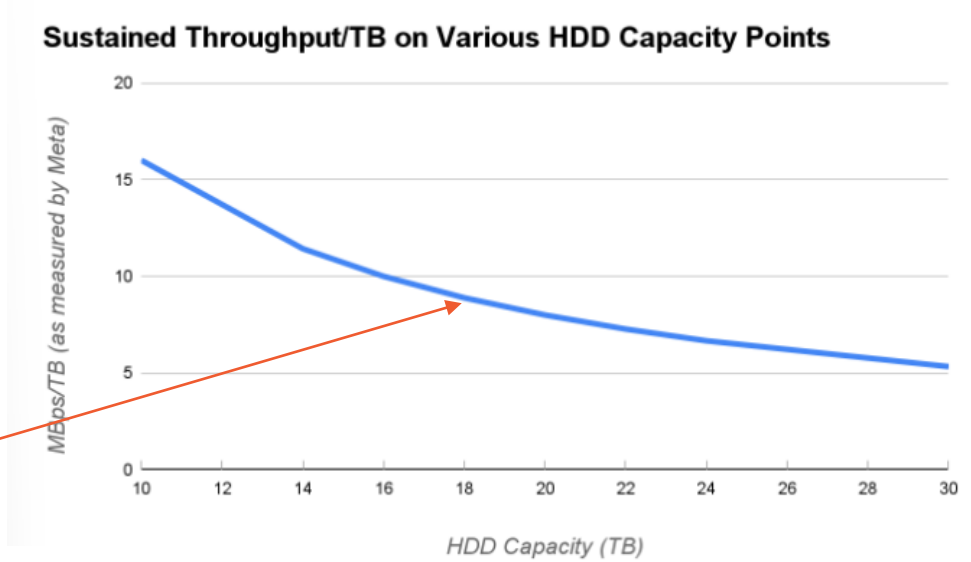
- Storage provides supplemental memory capacity

- A lot of cold data is warmer
  - What this means: As capacity goes higher, bandwidth/TB needs to stay the same
  - For an HDD, this is very difficult
  - For an SSD, this can be mitigated
  - This can be called a Near Line SSD

Near Line SSD

- QLC, NAND scaling continue lowering costs
- HDDs have difficulty scaling performance/TB
- QLC performance is lower than TLC but better than HDDs
- Hyperscalars have targets

Source: Meta blog: A case for QLC SSDs in the data center
https://engineering.fb.com/2025/03/04/data-center-engineering/a-case-for-qlc-ssds-in-the-data-center/



	HDD (Bulk Storage)	QLC SSD (Capacity Tier)	TLC SSD (Performance Tier)
Capacity (TB)	20-30	64-150	8-16
Acquisition Cost ($/TB)	Low	Med	High
Performance (BW/TB)	Low	Med	High
Power (W/TB)	High	Low	High

Source: OCP Storage Tech Talk (May 14, 2025)
https://www.youtube.com/watch?v=ppPGAngXX7c

More Context

- HDD Scaling
  - To continue their TCO curve HDDs need to keep growing capacity
  - Physics limits their maximum IOPs
  - IOPs per TB wall
- Data Lake Temperatures
  - A portion of the HDD tier is warming up
  - Guess is that AI is playing a role
  - Customers are paying to store the data and want a return on that
- TCO
  - Encompasses a lot more that just acquisition cost
    - Power, cooling, life span, replacement rate, etc.
  - Does not need to reach parity to start the ball rolling

Need to move to some form of HDD Displacement

Requirements

- Minimum of 64 NAND placements for high capacity at low die stacks
- Fit vertically in a 2U chassis for node density
- Large board area for low-cost componentry
- Symmetrical connector registration for component placement flexibility
- E1 LED placement for simplicity and cost
- Moderate performance (8-10 MBs/TB?) for low TCO (QLC)
- EDSFF connector and electricals (future proof for G7+)
- NVMe protocol (who would pick anything else in this day and age?)

Near Line SSD

- Focus on minimizing this cost
  - Increase the nominal capacity of the SSD
  - Reduces overall rack cost.
    - Less non-NAND components per rack
    - Less SSDs to hit same rack capacity
    - Less control nodes per rack
- Increasing nominal capacity means more NAND dies.
  - E2 is the most cost-effective path to get here

	HDD (Bulk Storage)	QLC SSD (Capacity Tier)	TLC SSD (Performance Tier)
Capacity (TB)	20-30	64-150	8-16
Acquisition Cost ($/TB)	Low	Med	High
Performance (BW/TB)	Low	Med	High
Power (W/TB)	High	Low	High

SSD	Performance SSD	Capacity SSD	Near Line SSD
NAND Dies	16-128	128-512	1024-2048+

Option	Pros/Cons
Option 1: Use existing form factors and double NAND dies/package to stay at ~32 packages	+Use existing form factors -Higher cost -Thermals
Option 2: Use existing form factors and make the form factor thicker to support 64+ packages(multiple PCBs)	+Use existing form factors -Front panel area -Higher cost -Thermals
Option 3: New form factor supporting 64+ packages	+Cheaper +Thermals -New Form Factor

Summary on AI Needs of SSDs

- An AI SSD will consume more power to meet the higher performance
- Direct Liquid Cooling will help solve cooling the AI SSD
- Near Line SSD will need to maintain a balance of good enough performance vs. capacity vs. cost.
- To get to lower overall system costs, a higher capacity SSD like E2 is needed.

One last thing: SFF TWG Participation

- We are solving problems around higher-speed Ethernet and PCIe interconnects to solve AI bottleneck problems while improving existing interconnects and form factors.
- Our members include participants involved in ASICs/CPUs, Data centers, interconnects, networking, research, server systems, storage devices, test equipment, and transceivers.
  - We develop specifications to support a broad range of usages.
- Benefits:
  - Participation in development of SFF specifications, information documents, and reference guides
  - Ability to open new projects
  - Access to all presentations, all drafts, prior publications, and supplemental material relevant to all SFF projects
- Resources:
  - Public Site: https://www.snia.org/sff
  - Specifications: https://www.snia.org/sff/specifications
  - Additional questions? Please send mail to sff-chair@snia.org

SNIA | DSN
A SNIA Community

Q&A

After this Webinar

- Please rate this webinar and provide us with your feedback
- This webinar and a copy of the slides are available at the SNIA Educational Library snia.org/educational-library
- A Q&A from this webinar, including answers to questions we couldn't get to today, will be posted on our blog at sniablog.org
- Follow us on LinkedIn and X @SNIA

SNIA | DSN
A SNIA Community

Thank You