



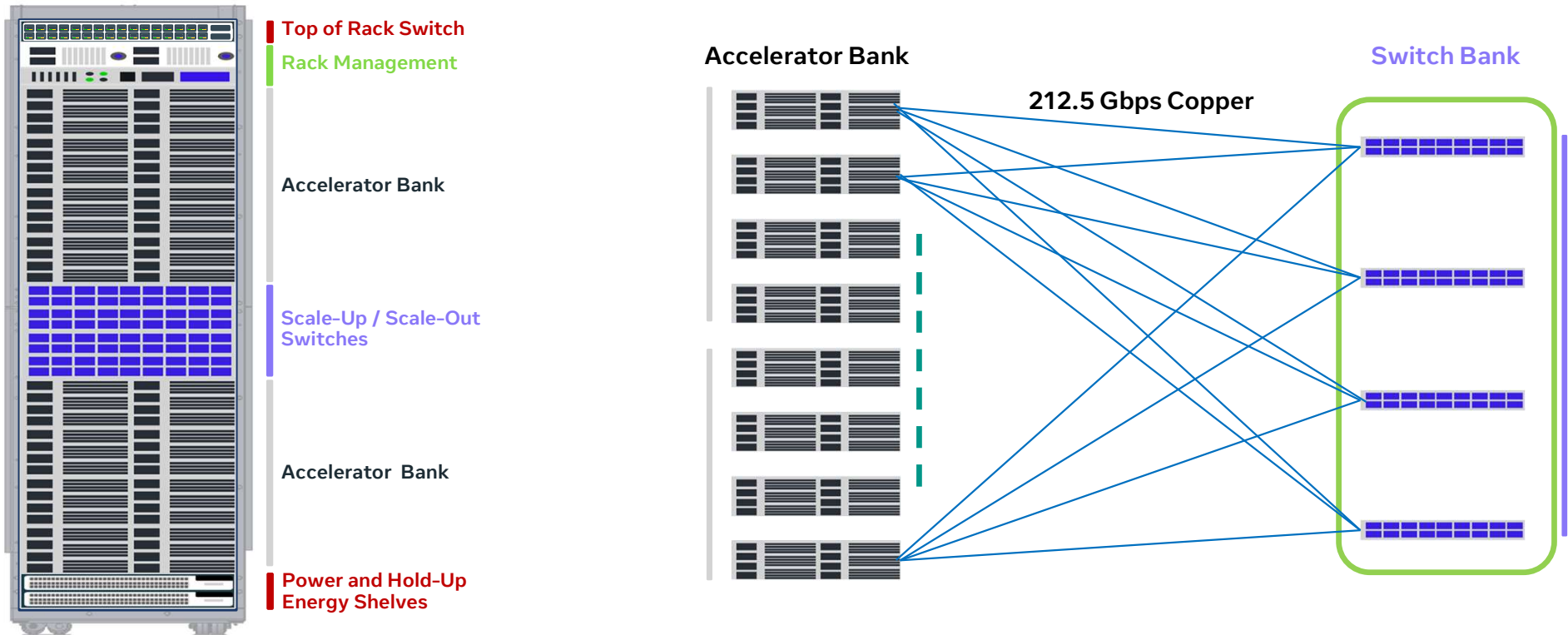
# AI Systems and Interconnects: LR Channel for Scale-up

Srinivas V., Xu Wang  
04-15-2025

# Overview

- Consideration for long reach (LR) type channels to support single hop high bandwidth connectivity between accelerators
  - Systems form factors are the same as current generation making this challenging
- Contribution covers an initial assessment for an LR channel for scale-up connectivity
  - Assumes use of Co-packaged Copper Interconnect (CPC) on both ends
  - Evaluates feasibility with Baseband PAM modulation

# Current Generation Scaleup Connectivity

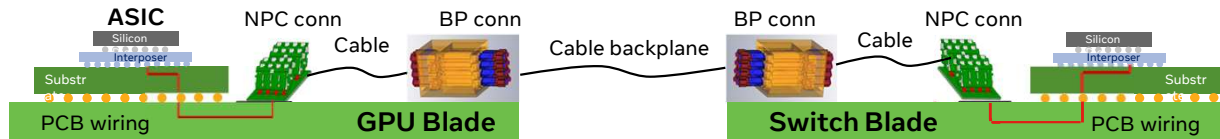


- All to all connectivity between accelerators and switches
- 1 to N GPUs per tray
- Configurable no of network trays
- Long reach (LR) links for scale-up

# Current and Next Gen Scaleup Connectivity

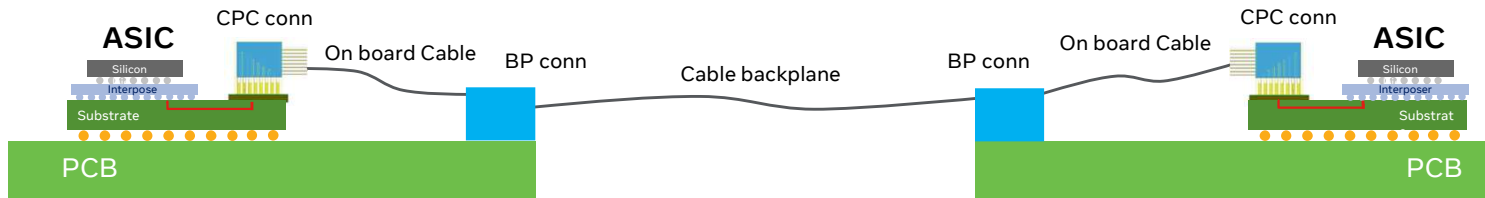
- 212G Scale up

- Near package connector + flyover cables through a dedicated cable backplane cartridge
- Near package connector + flyover cables through an orthogonal direct connector attach
- Worst case cable lengths end to end can be close to ~2 meters
- Packages can consume 25-30% of the overall channel budget
- Manage impairments with package resonances and package/PCB footprint crosstalk

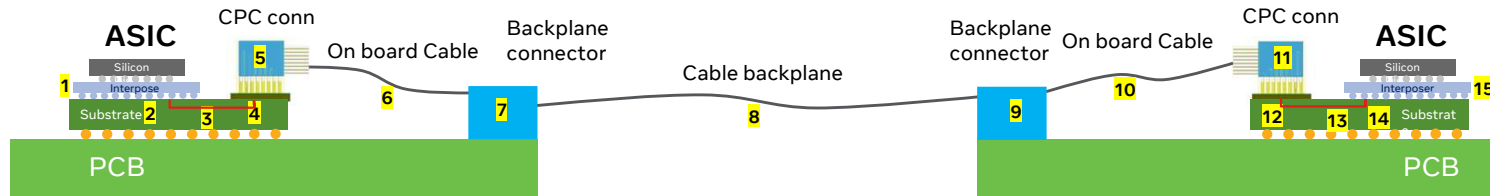


- Next Generation Scale-up

- Support similar cable backplane connectivity and physical reach as previous generation ~ 2 meters
- CPC connector attach to the same first level package as CoWoS interposer
  - Eliminates package resonances and PCB crosstalk impairments
- Limited cable lengths to 1.6 m at this stage for early analysis
- On board cable lengths can be eventually stretched using gauge adapters to go from AWG 32 on CPC side to AWG 26 towards the backplane side to make up the difference



# 448 Gbps LR Channel with CPC connectors



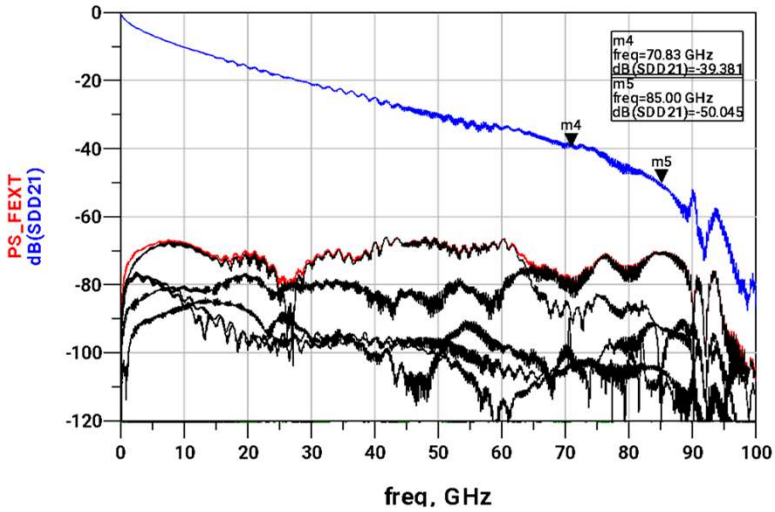
	Interconnect (Bump to Bump)	Notes
1	CoWoS-L interposer	1, 2 are extracted together
2	Organic substrate c4 escape routing, 10 mm, 16/35/16 um	80 C, nomSR, GL107 (current generation material) <sup>1,2</sup>
3	Organic substrate wiring, 30.4 mm, 86 ohms, 40/64/40 um	80 C, nomSR, GL107, 0.165 dB/mm at 75 GHz (current generation material) <sup>1</sup>
4	CPC connector footprint	
5	Next gen CPC connector	Diff pair count / footprint optimized
6	On board cable 300 mm, 92 ohms	AWG 32, 0.368 dB/inch at 75 GHz (current generation cable)
7	Next gen Backplane IO connector	Cable in-out M-F
8	Cable Backplane, 1.0 meter, 90 ohms	AWG26, 0.205 dB/inch at 75 GHz (current generation cable)
9	Next gen Backplane IO connector	Cable in-out M-F
10	On board cable 300 mm, 92 ohms	AWG 32, 0.368 dB/inch at 75 GHz (current generation cable)
11	Next gen CPC connector	Diff pair count / footprint optimized
12	CPC connector footprint	
13	Organic substrate wiring, 23 mm, 86 ohms, 40/64/40 um	80 C, nomSR, GL107, 0.165 dB/mm at 75 GHz (current generation material)
14	Organic substrate c4 escape routing, 10 mm, 16/35/16 um	80 C, nomSR, GL107, (current generation material)
15	CoWoS-L Interposer	14, 15 are extracted together

**Note 1:** Scope for improvement with next gen build-up, skip layer routing and next gen cable interconnect

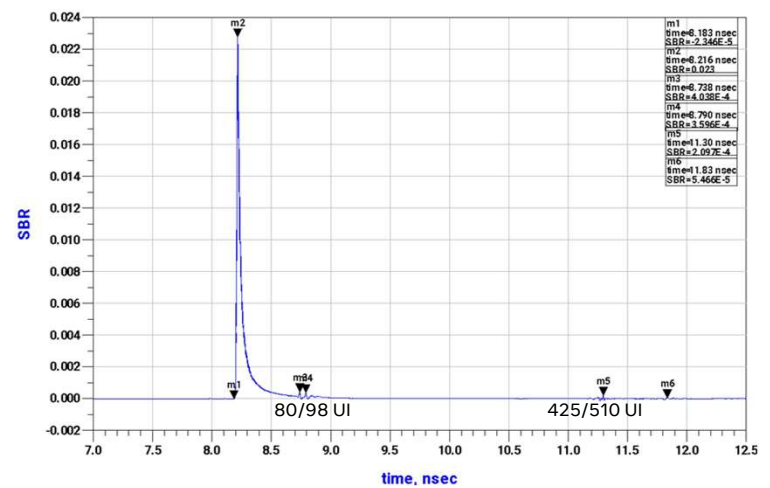
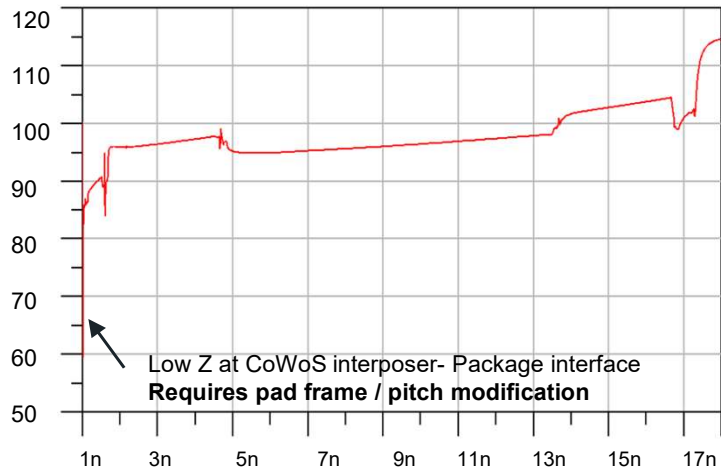
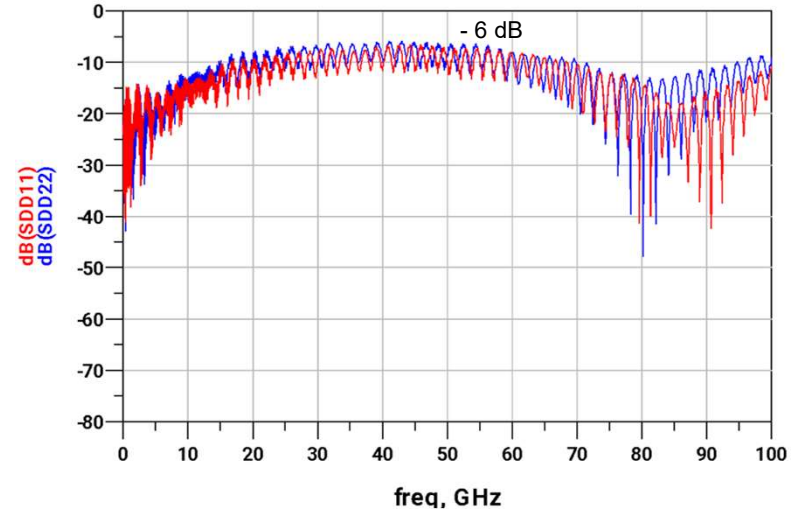
**Note 2:** Nominal surface roughness

# CPC Based LR Channel Data

Loss and crosstalk



Return loss



# COM Analysis

- Computation based on COM 4.8, with following modifications to 802.3dj config<sup>[1]</sup> parameters
  - Data rate: 425 Gbps
  - Device parasitics ( $C_d$ ,  $L_s$ ) scaled by ~30%
  - DER\_0: 1E-4
  - FFE: 7+1+40 fixed; floating tap span to 100 UI
  - $A_v$ : 0.385 V;  $N_{qb}$ : 6.5;  $\eta_0$ : 1E-09 V<sup>2</sup>/GHz
  - Crosstalk files included in the simulations

	<b>PAM-8</b>	<b>PAM-8<sup>[2]</sup></b>	<b>PAM-6</b>	<b>PAM-6<sup>[2]</sup></b>
COM margin	-1.55	-0.13	-1.93	-0.91
DER_DFE	2.70E-03	8.60E-04	4.60E-03	2.20E-03
DER_MLSE <sup>[3]</sup>	9.95E-04	1.25E-04	1.50E-03	4.21E-04

- COM results do not show sufficient margin, needs more investigation
- CPC connector launch and the backplane connector mating interface contribute to late reflections (~ 2x over the tray cable delay)

**Note 1:** Baseline 448G config spreadsheet from Rich Mellitz

**Note 2:** Increased floating tap groups to 4 and span to 550 UI

**Note 3:** Turned off additional impairment penalty applied before MLSE calculation (modified code from Hossein Shakiba)

## Serdes IP Vendor Analysis

	Data rate	Channel	DSP	Conclusions
Marvell	<ul style="list-style-type: none"> <li>425Gbps</li> </ul>	CPC/long reach	FFE+DFE+MLSD	<ul style="list-style-type: none"> <li>PAM6 modulation - yields 1dB SNR margin with respect to KP FEC limit (<math>&lt; 5E-5</math>)</li> <li>PAM8 modulation - yields 2.4 dB SNR margin with respect to KP FEC (<math>&lt; 4e-6</math>)</li> <li>Additional note: <i>Adding optionally Inner FEC is going to improve the SNR margin</i></li> </ul>
Maxlinear	<ul style="list-style-type: none"> <li>430 Gbps</li> <li>7.5% FEC overhead</li> </ul>	CPC/long reach	FFE+DFE+MLSD	<ul style="list-style-type: none"> <li>Tx with 1.2 V swing</li> <li>&gt; 8+ dB Tx pre-emphasis (5 tap FFE)</li> <li>BER <math>&lt; 1E-5</math> (both PAM-8, 6)</li> </ul>
Broadcom	425 Gb/s not including inner code overhead <ul style="list-style-type: none"> <li>170 Gbaud PAM-6</li> <li>142.5 Gbaud PAM-8</li> </ul>	CPC/long reach	Includes MLSD	<ul style="list-style-type: none"> <li>PAM-6 not analyzed</li> <li>PAM-8 needs inner code with 2.4 dB net coding gain for 3 dB margin [1]</li> <li>Significant reflection(s) with very high round-trip delay</li> </ul>
		Lower loss CPC/long reach <ul style="list-style-type: none"> <li>Package lengths reduced to 25 mm on each substrate</li> <li>~4.5 dB reduction at 70.8 GHz</li> </ul>	Includes MLSD	<ul style="list-style-type: none"> <li>PAM-6 needs inner code with 1.3 dB net coding gain for 3 dB margin</li> <li>PAM-8 needs inner code with 1.4 dB net coding gain for 3 dB margin</li> <li>Significant reflection(s) with very high round-trip delay</li> </ul>

[1] Net coding gain is coding gain adjusted for higher signaling rate due to encoding. Margin is relative to the IEEE P802.3dj maximum BER allowance ( $2.76e-4$ )



## Summary

- Copper interconnect with LR reach is required for high bandwidth connectivity between accelerators
  - Support similar physical reach as today's system form factors
  - Retimer less configurations preferred
- CPC attach directly to ASIC substrate might be the only viable option to support this interface
  - Requires development of low loss, high bandwidth wire gauge transition adapters
  - Enable small CPC footprint on package and large ~ AWG26 type gauge for bulk of the interconnect
- Analysis with Meta LR channel
  - Poor convergence with COM (needs deeper investigation)
  - Some channel improvements possible to improve (loss/reflections)
  - More analysis needed to converge on an optimal solution with necessary DSP + FEC, considering trade-offs with IO power/design complexity
  - Measured data on channels and with test chips