Google

Technical Feasibility of Passive Copper Cables at 200Gbps/lane

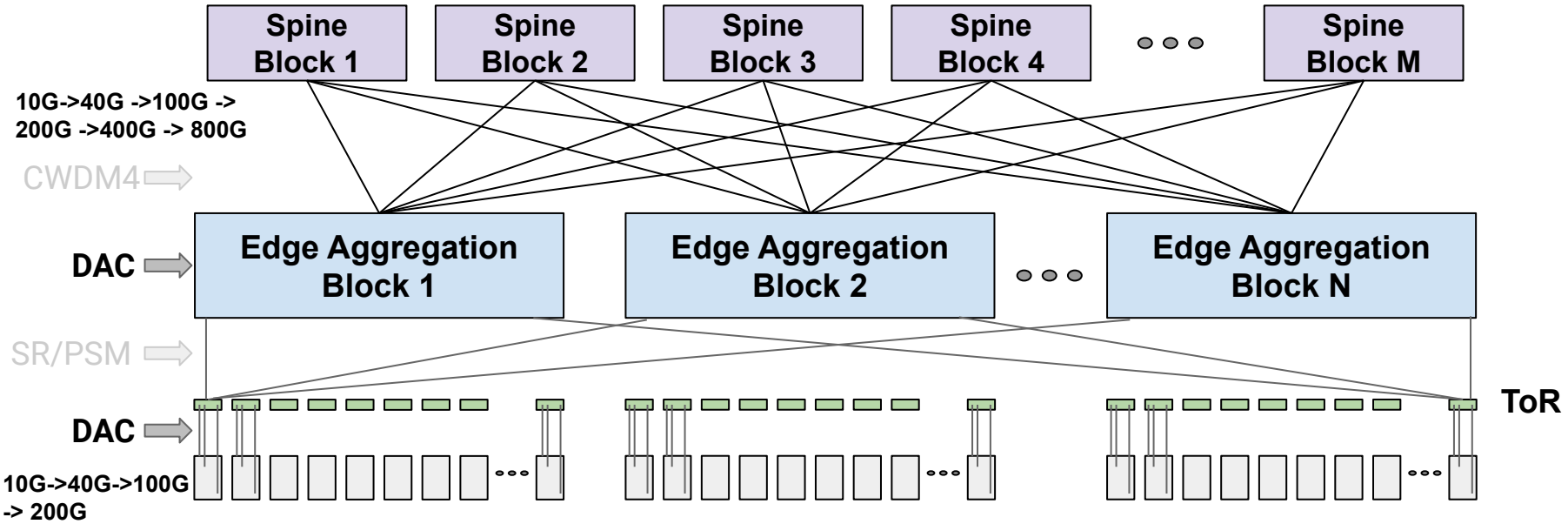
Leesa Noujeim, Sara Zebian, and Hong Liu

IEEE 802.3 Beyond 400G Ethernet Study Group, May 17 2021 (r2)

Outline

- Motivations for 200Gbps per lane
- 200Gbps/lane electrical components readiness
- Link budgets for 200Gbps/lane

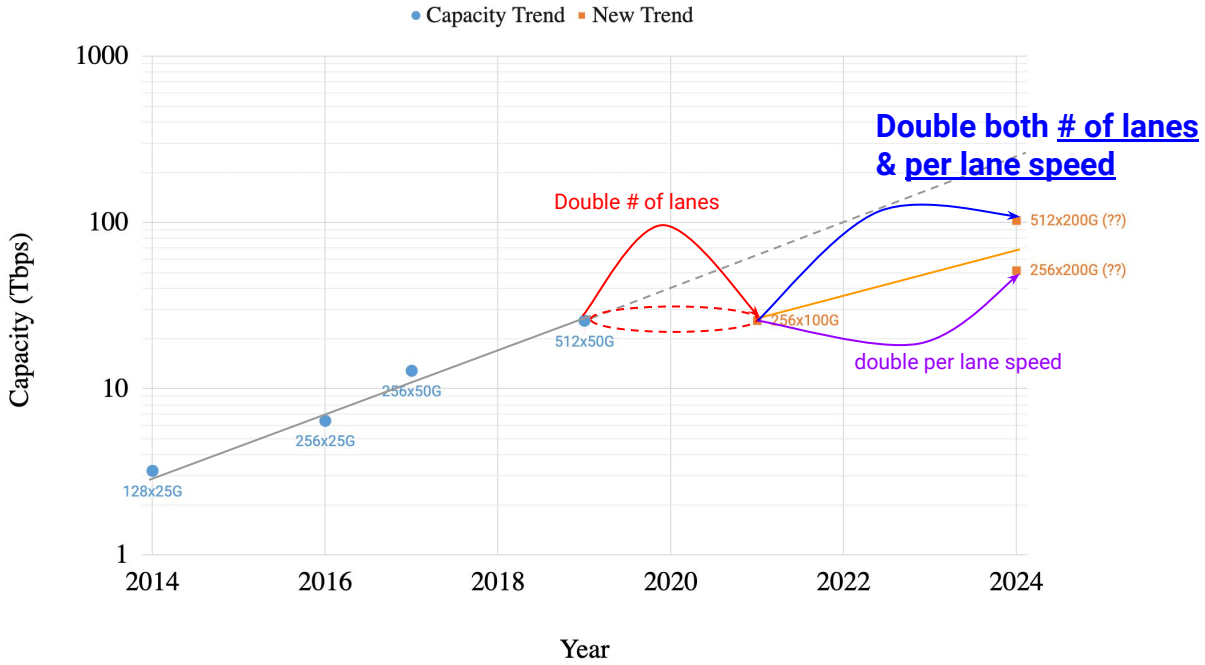
>400G will be needed in Datacenter Network fabrics



- DAC (Direct Attach Copper) is the most efficient interconnect for edge block and TOR

https://www.ieee802.org/3/B400G/public/21_03/lam_b400g_01a_210329.pdf

Motivations for 200Gbps/lane



- Switch capacity growth slowed down due to limitation of power, SI and # of packaging pins
- 200Gbps/lane is critical to scale the switch bandwidth and efficiency of networking fabric.

https://www.ieee802.org/3/B400G/public/21_03/lam_b400g_01a_210329.pdf

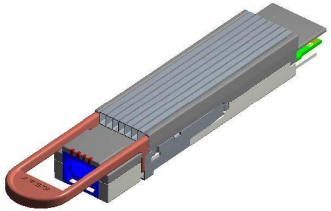
Pluggables for 200 Gbps per lane

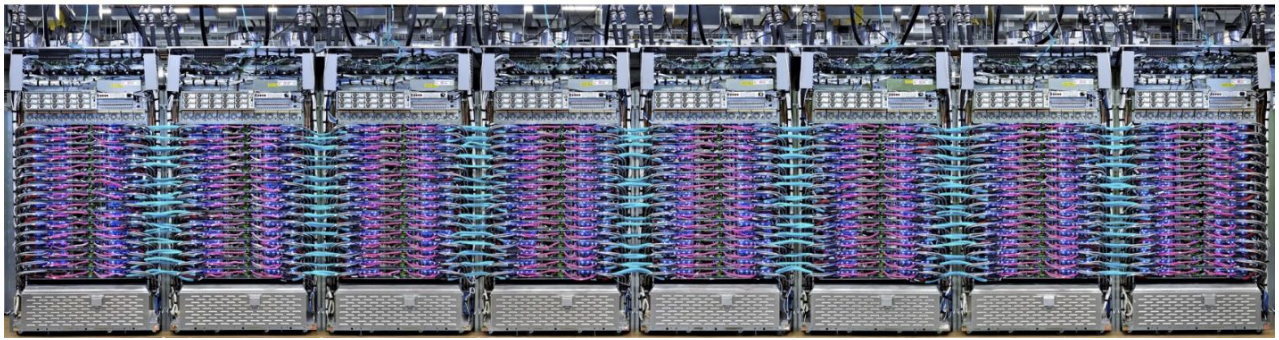
Pluggable front panel ports provide:

- Evolutionary path for optics to 1.6Tbps/module (2x800G)
- Flexibility to optimize for various reaches: CR, SR, LR ...

Given front panel pluggable ports for optics:

=> next step is to explore feasibility copper cables in these ports



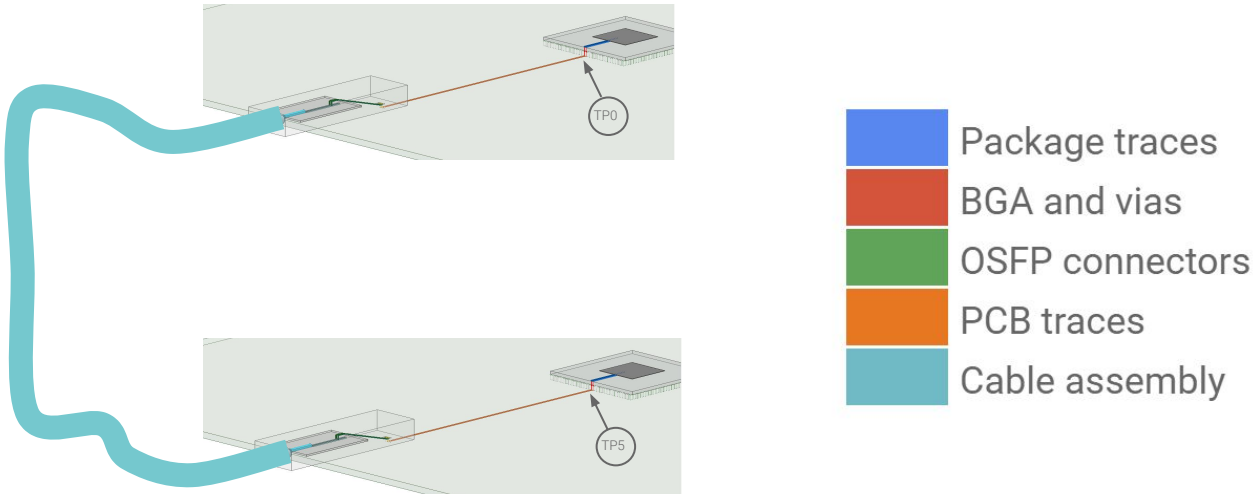


http://cloud.google.com/tpu/

Channel Loss Budget: Key Components

Channel is traditionally specified between TP0 and TP5

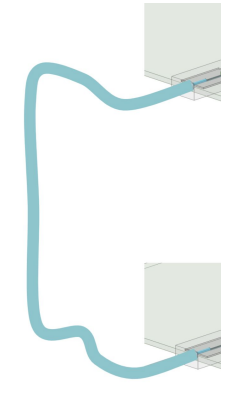
- BGA/vias and packages are included in analysis



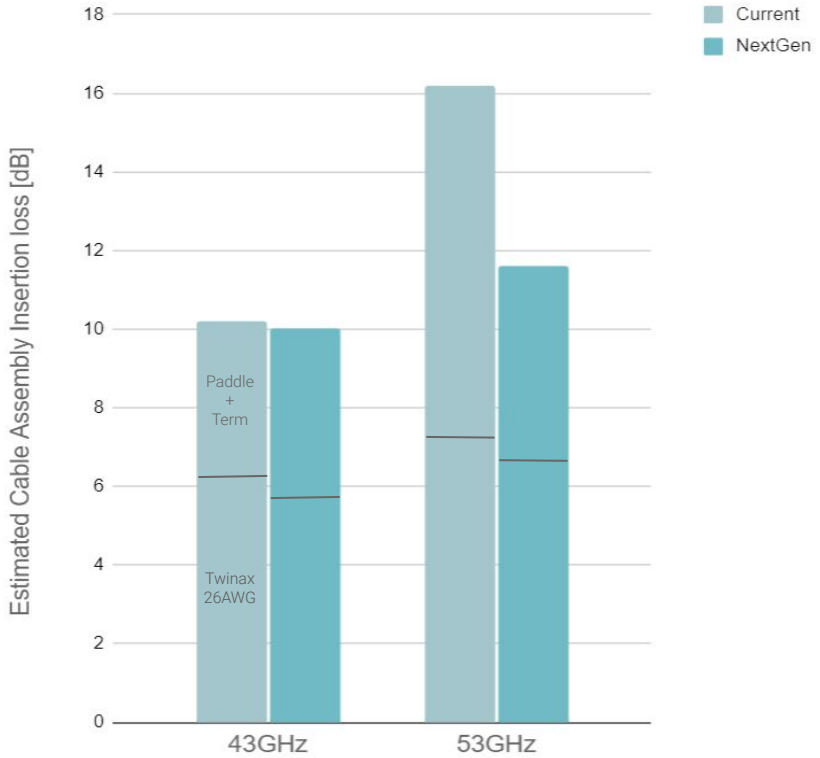
Key Component: cable assembly*

Engineered rackifications are key to managing cable lengths

- enforce long-to-short port combinations
- Key applications can be enabled by 1.0m cables



*excluding connector



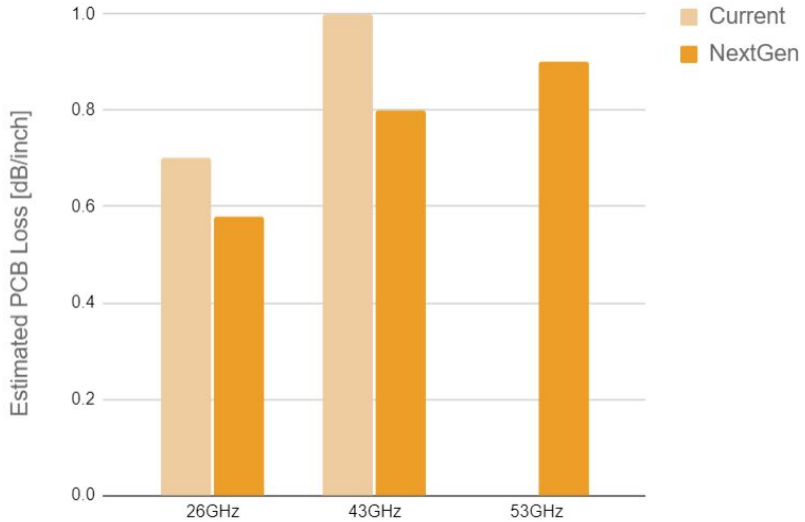
Key Component: PCB trace loss

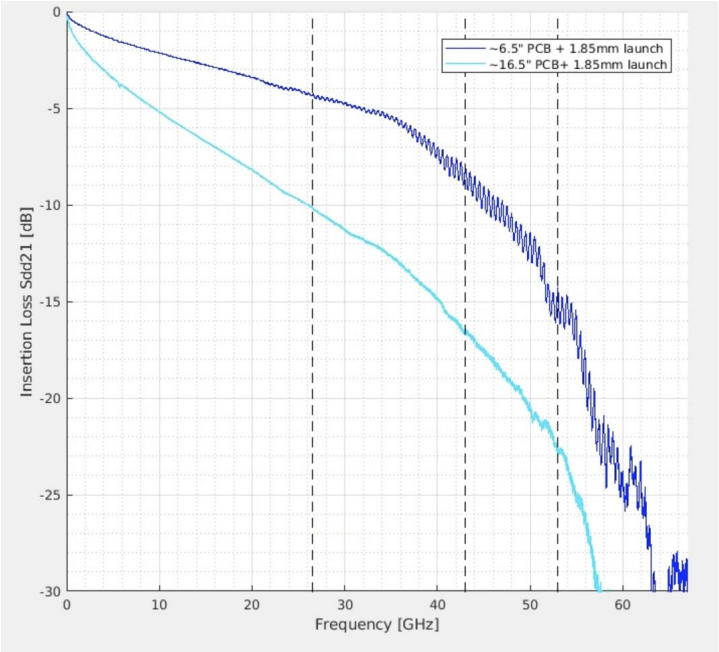
PCB differential pair loss improvements: evolutionary (no exotic materials)

- Wider trace with improved surface roughness
- Lower dielectric loss

Length reduction: port partitioning short/long

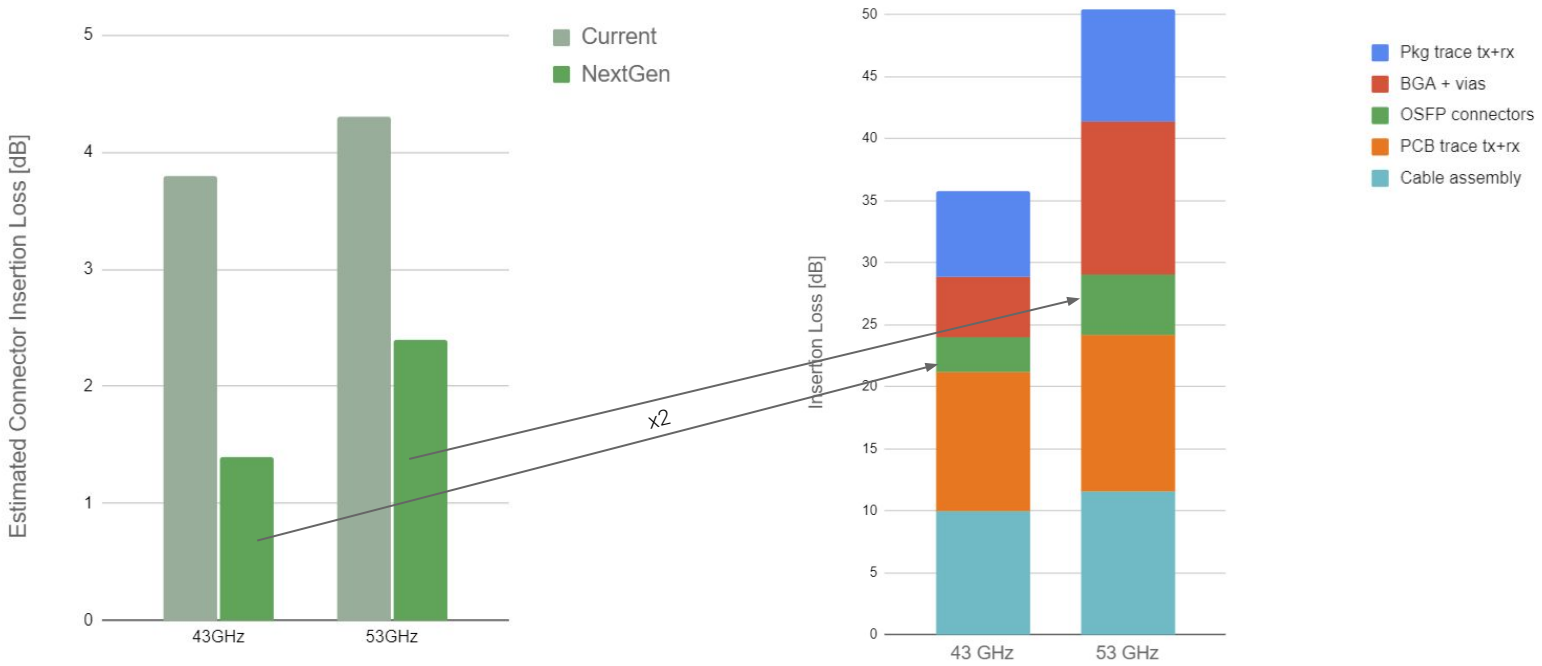
- Classic 32-port switch: ~9" long [1]; ~75% < 7"; ~60% < 5" (short) [2]
- Engineered rackifications: disallow long-long [3]





Key Component: connector

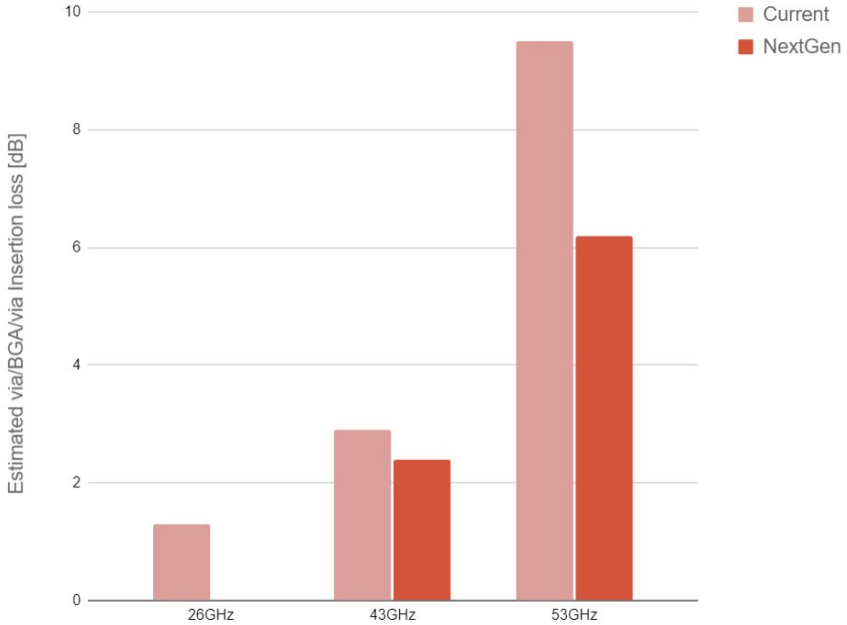
OSFP200 MSA includes contributions on feasibility of connector
Improvements focus on mating interface (stub, tolerance) as well as SMT interface

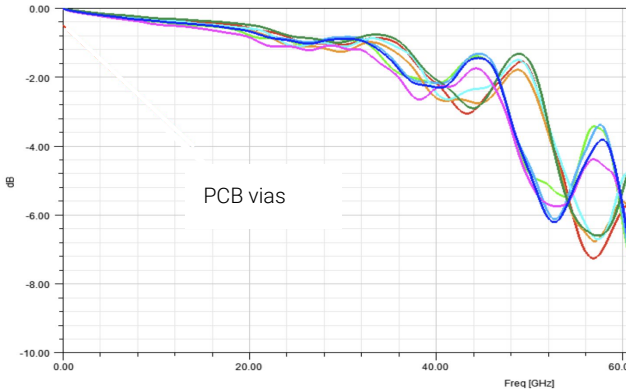


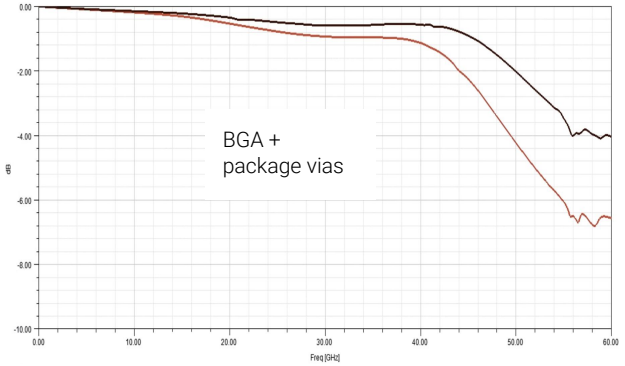
Key Component: PCB vias, BGA, package vias

CALL TO ACTION: need cross-functional assessment

- Behavior strongly influenced by manufacturing capabilities
- Cross-sectional dimensions approach lambda/4 => over-moding





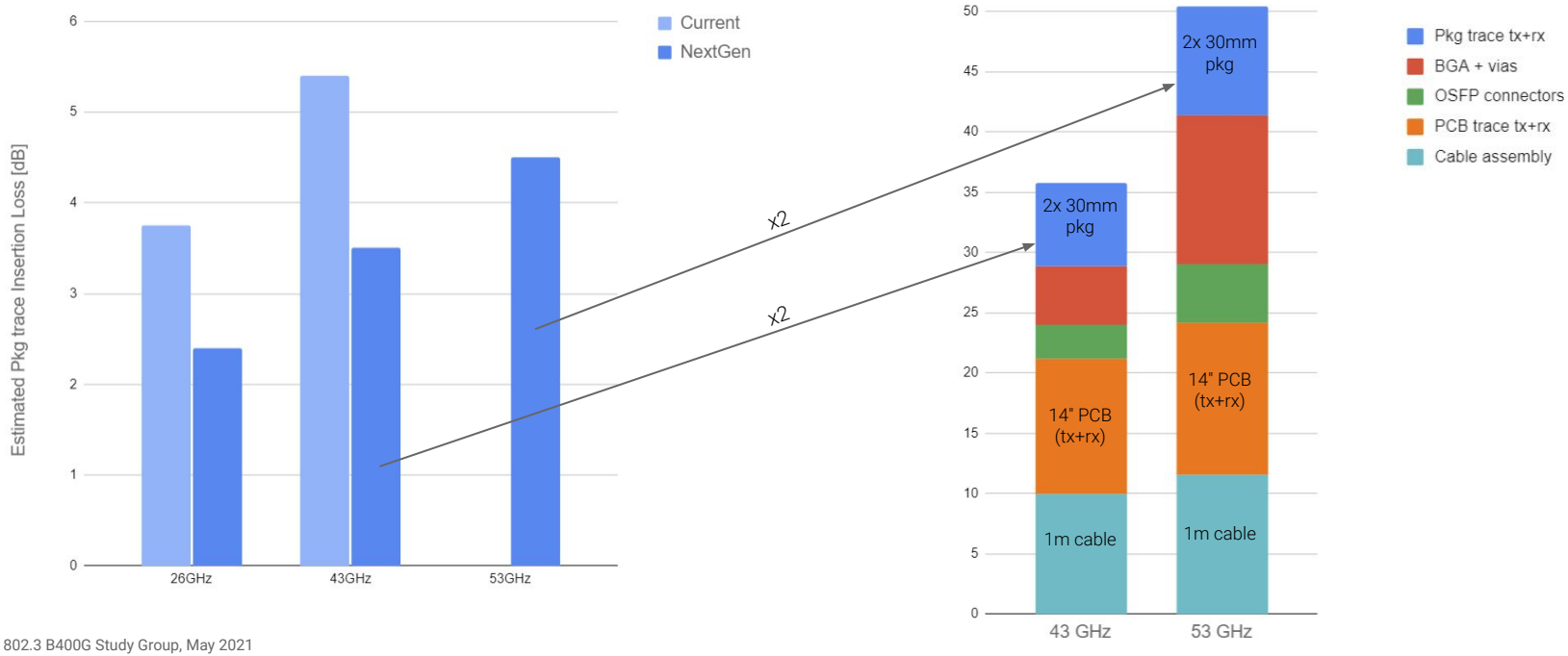


Simulations courtesy Leo Li, Google

Key Component: package differential pair

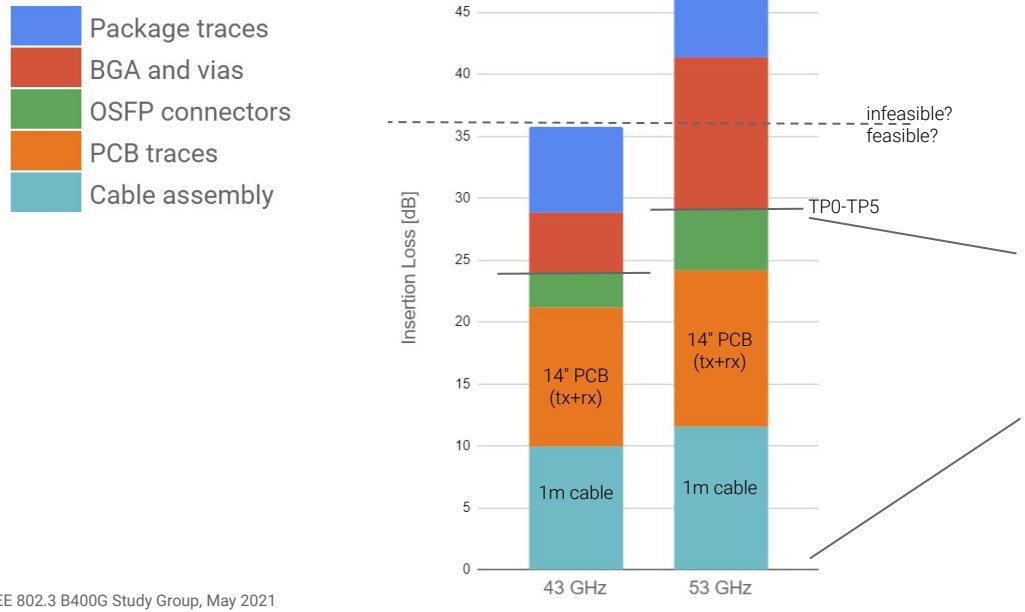
Package loss increasing disproportionately compared with prior generations

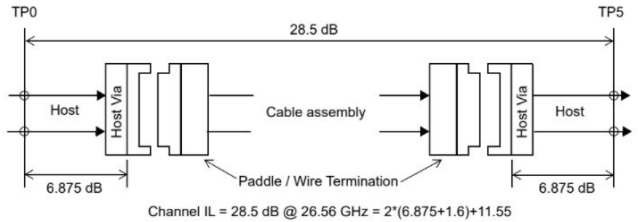
- based on large switch: 30mm package trace [4]
- asymmetric links (eg TOR-NIC) have shorter (eg 15mm) package trace



Link Budget

- <36dB bump2bump @ 43GHz: loss budget is promising
- >50dB @ 53GHz much more challenging; may be viable with serdes and BGA/via improvements
- Next steps: consider insertion loss deviation and crosstalk





Channel IL = 28.5 dB @ 26.56 GHz = 2*(6.875+1.6)+11.55

NOTE—Channel IL derived from cable assembly host, and mated test fixture

Figure 162A–4—Channel insertion loss at 26.56 GHz

Summary

- Passive copper cable at 200Gbps/lane seems feasible: ~36dB bump-to-bump for PAM6
- Need further technical analysis of channel, particularly for BGA/via impairments beyond 45GHz
- Serdes capabilities, FEC, and power efficiency not covered - need full link analysis with channels informed by this feasibility study

Call for action: refine analysis of passive copper cable interface at 200Gbps/lane.

THANK YOU!

References:

[1] https://www.ieee802.org/3/100GEL/public/18_03/stone_100GEL_01_0318.pdf
[2] https://ieee802.org/3/ck/public/18_05/stone_3ck_01a_0518.pdf
[3] https://www.ieee802.org/3/ck/public/adhoc/apr28_21/dawe_3ck_adhoc_01_042821.pdf
[4] https://www.ieee802.org/3/100GEL/public/18_03/lim_100GEL_01b_0318.pdf

Contributors:

Kevin Zhang, Leo Li, Cedric Lam,