# The Case for 36 dB C2M Insertion Loss at 200G

Pirooz Tooyserkani, Upen Reddy Kareti, Yi Tang

November 12-16, 2023

# List of Supporters

- Samuel Kocsis, Amphenol
- Valery Kugel, Juniper Networks
- Brian Welch, Cisco Systems
- Nathan Tracy, TE Connectivity
- Megha Shanbhag, TE Connectivity
- Liav Ben-Artsi, Marvell
- Phil Sun, Credo
- Kapil Shrikhande, Marvell

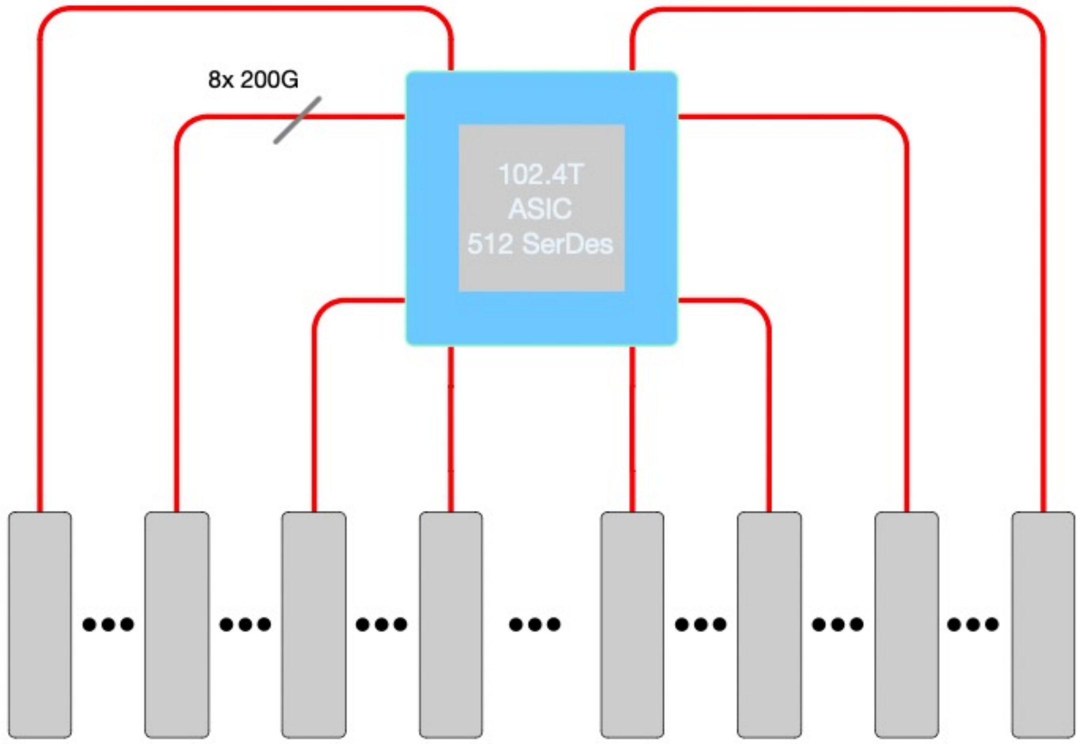
# Higher Bandwidth System Requirement

- AI/ML is the driving force behind the push for higher bandwidth Ethernet Switch systems, hence next step is 200G/lane
- Advances in Silicon is now hitting 102.4T bandwidth doubling the current 51.2T switch capacity
  - A 51.2T system is comprised of 512 SerDes with 64x 800GE front panel ports
  - At 102.4T the radix stays the same but the speed doubles to 64x 1.6T per port
  - Rack density for AI/ML cluster is a strong market priority and maintaining 2RU is a strong market need

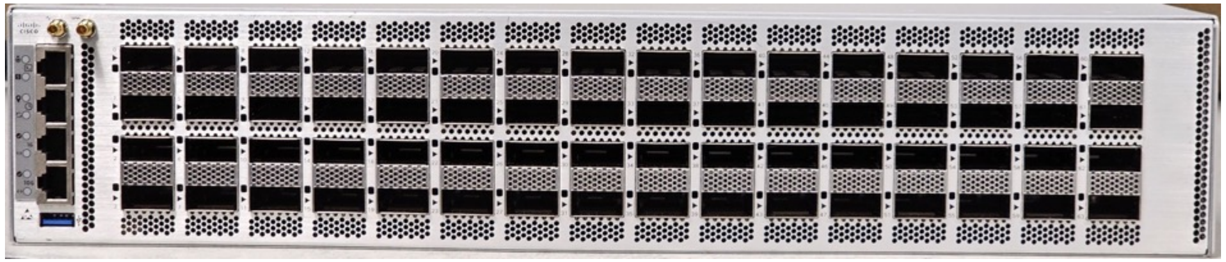
2RU 64-port switch: A crucial application requiring optimal support

# Building a 64 port 2RU Switch - options

- Single ASIC in a fixed box fanning out 512 SerDes to the front panel ports
- Option 1:
  - Fully PCB routed (preferred due to cost) requires 36 dB C2M IL for 200G
- Option 2:
  - Fully PCB routed up to 32 dB channels, cabled interfaces to remaining ports
- VLC is not considered here due to rack density requirements (4RU minimum)



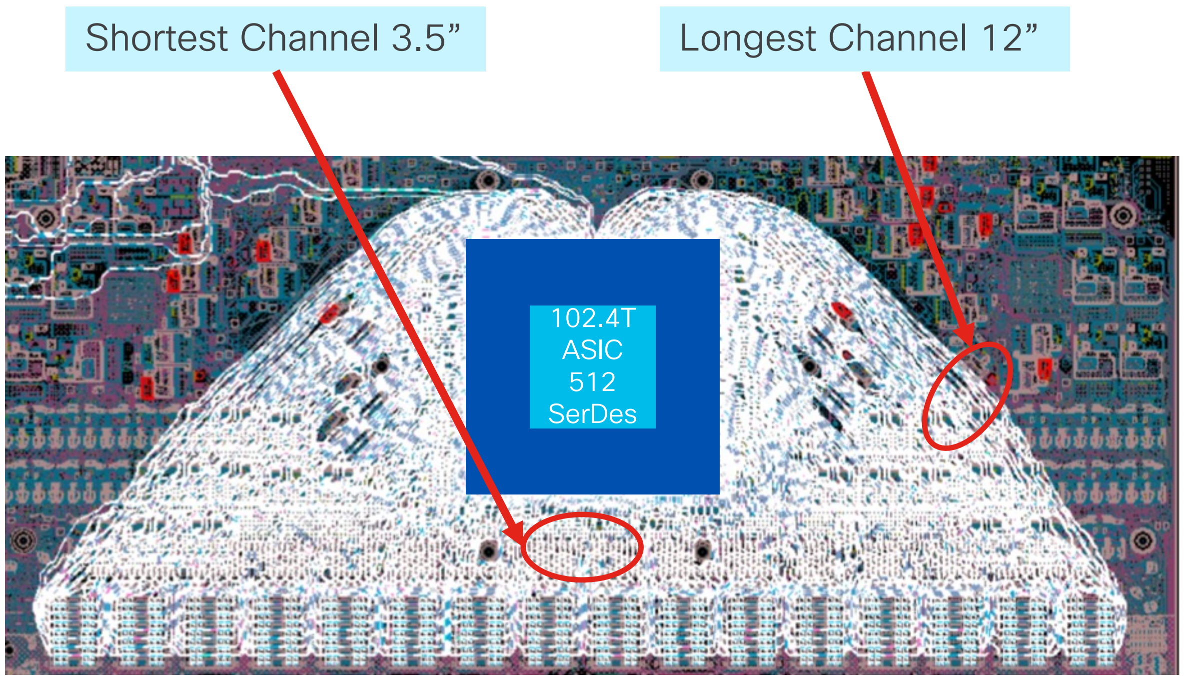

64x1600G cages (e.g. QSFP-DD1600)



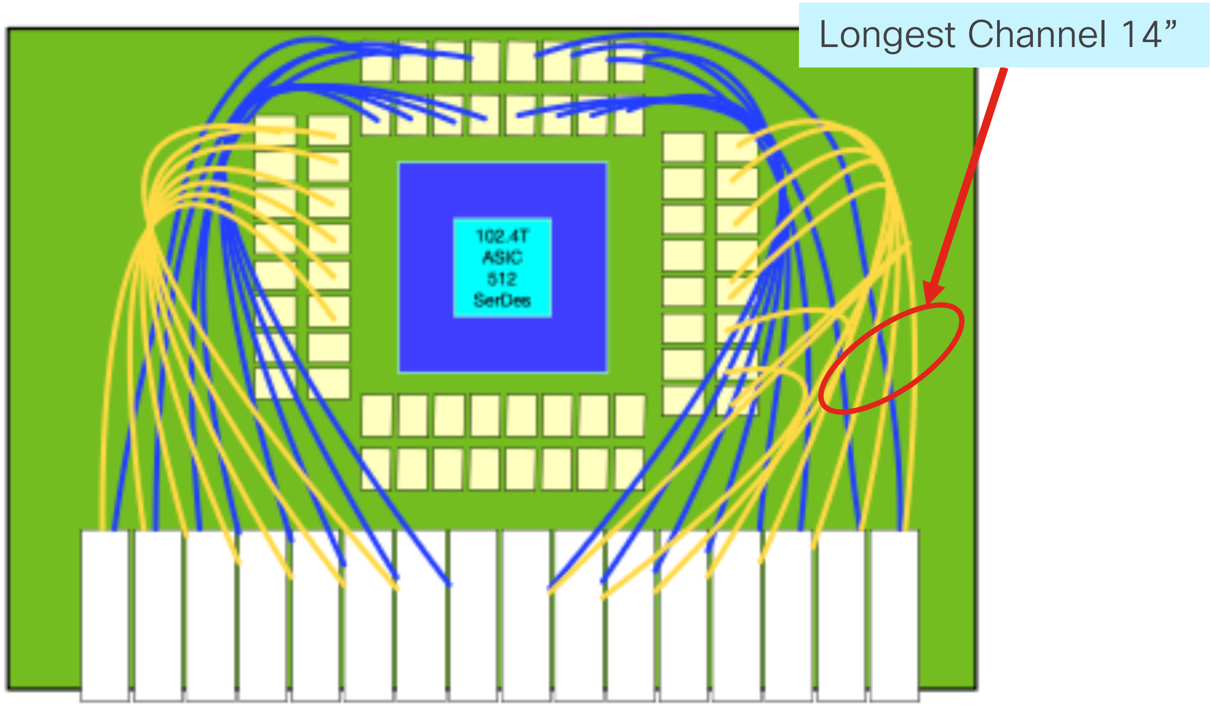
# Retimers

- 102.4T Systems must be significantly better than 51.2T system from thermal and power consumption perspective
  - End-user overall power consumption limitations must be taken care of due to facility limitations and green initiatives
  - Doubling the data rate limits the reaches using reasonably low-power SerDes architecture
  - Thermal aspect of the design is becoming very difficult with respect to system cooling solutions and is also a burden on overall system power consumption - higher fan speed
- Use of retimers -
  - Total system power consumption
  - System thermals is negatively affected by addition of retimers
  - Real estate - retimers require POLs, power filtering, heat sink and mounting holes which limits the PCB area usability

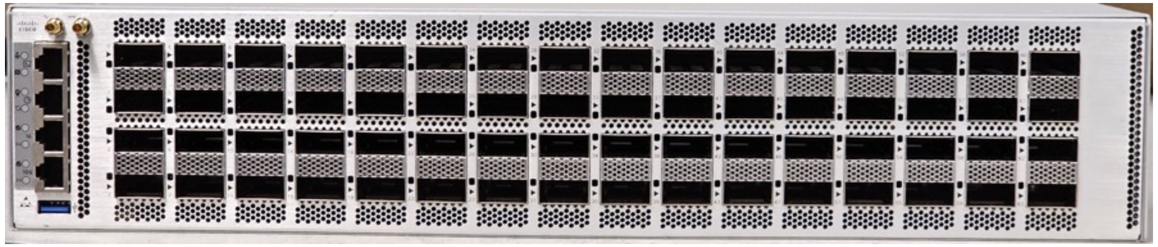
# 102.4T – Implementation design study




PCB Routed



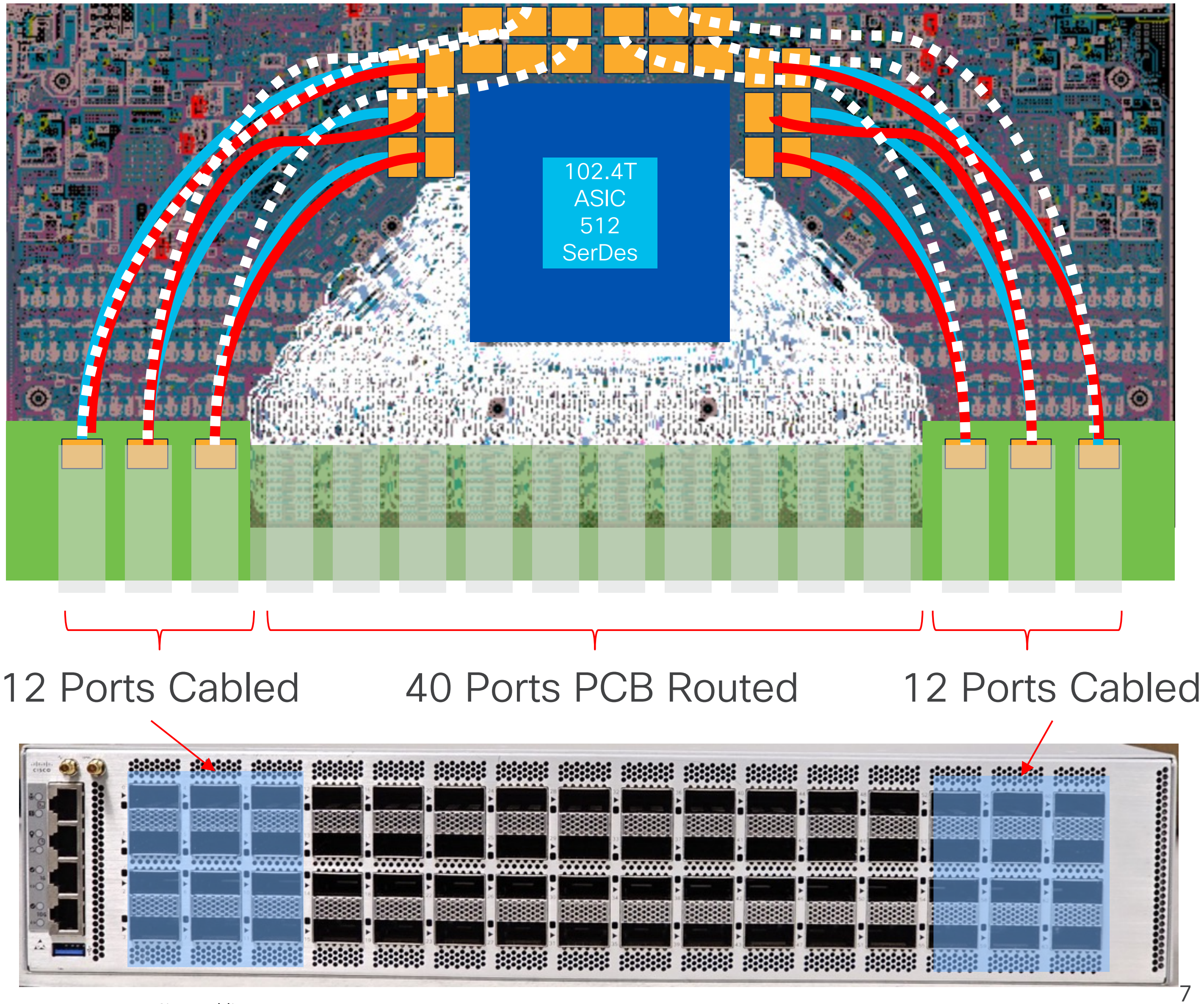

Cabled Host



ASIC to connector = 3.5”
Cable to Port = 14”

# Hybrid PCB and Cabled Solution

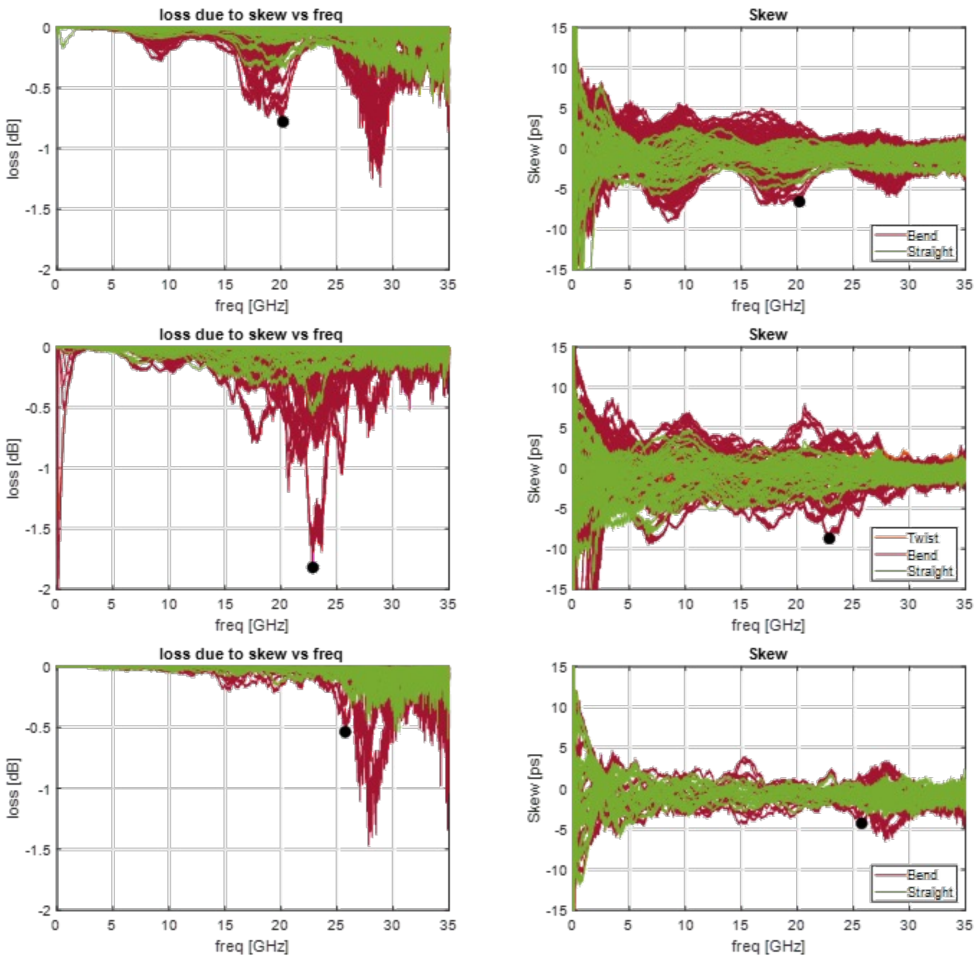
- Supporting 32 dB channel loss requires a hybrid solution
- This means 2/3 or 42 can be PCB routed and 1/3 or 22 ports require cable
- However, the hybrid solution does come with the following issues that needs to be addressed:
  - PCB Skew of P/N is fixed and does not vary much (temperature) after fabrication, but cable skew may vary from assembly to assembly due to bend/twist and temperature
  - Assembly complexity is a disadvantage of cabled solutions



# Challenges with Cabled Host

- At 100G with cabled ports, we see evidence that P/N skew is causing problems. At 200G we expect it to be much more severe
- Cable skew can occur during manufacturing (tolerance) or during installation of the cable (cable twist/bend) – It is important to limit cable skew to a reasonable value, including temperature impact and it is Important to limit the variability of cable skew for systems once deployed
- Skew leads to pulse shrinkage, increases insertion loss in the channel, and degrades the SNR
- The negative effects of P/N skew are currently observed in cabled hosts operating at 100G per lane
- While the impact can be mitigated at 100G, it is expected to worsen with the transition to 200G (UI of 9.4ps)
- A more detailed explanation of the impact of skew will be provided in a future presentation

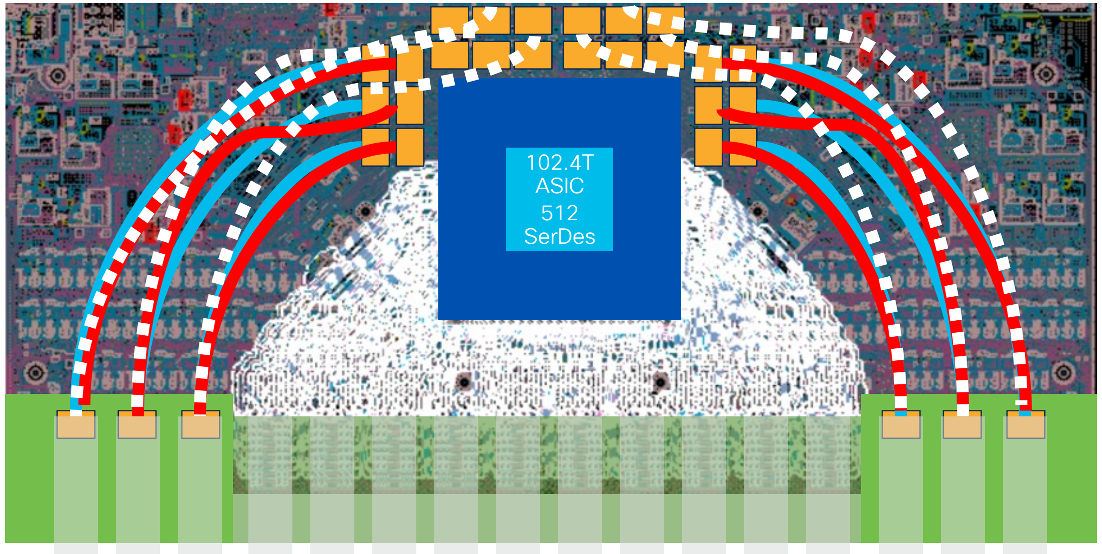
# Cable Frequency Domain Measurements




- Performance of multiple cables was measured in the frequency domain under various realistic conditions
- Significant skews were detected in straight cables provided by multiple cable suppliers
- Skew intensifies when the cables are twisted or bent
- Increased insertion loss as a consequence of skew

Note: the data shown is for 100G cable, 200G cable data – TBD

# Find Paths to Both Implementations


PCB Implementation: max **36 dB IL**
(manageable PCB routing skew)

Cabled/Hybrid Implementation:
max 32 dB IL (with **Undeterministic cable skew**)

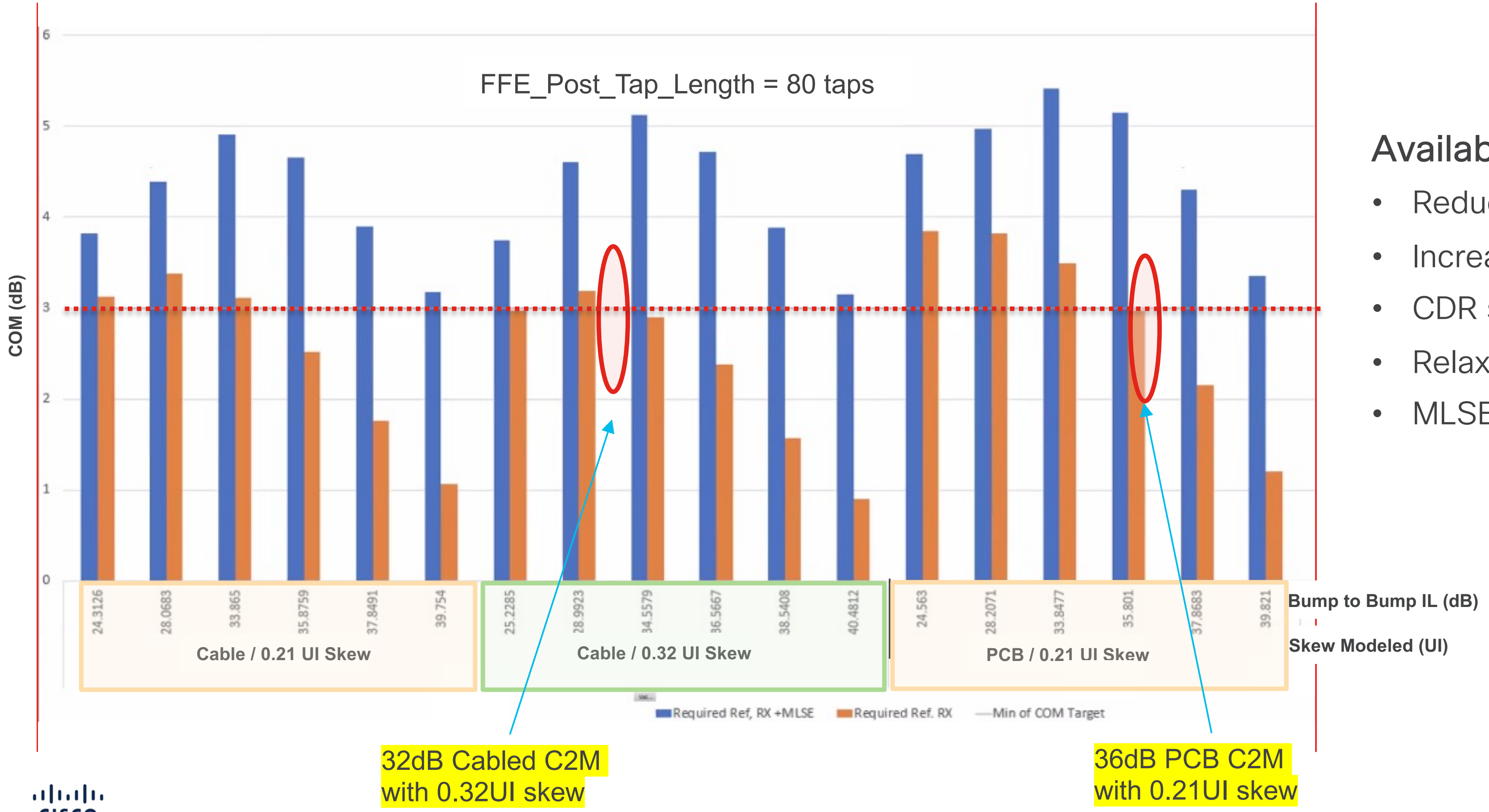
- Must support both PCB routed design, and cabled host systems
- 2x1 stacked connector design is needed for both cases
- Cabled ports are needed for flexibility in system design and implementation

# 36 dB Routed vs. 32 dB Cabled

- Complexity Comparison:
  - Despite the challenge posed by a 36 dB C2M equalizer, its complexity maybe comparable to a 32 dB equalizer designed to handle high cable skew
- Additional Loss from Cable Skew:
  - A 32 dB channel with cable skew incurs an effective additional insertion loss, necessitating the use of a 32+ dB equalizer to compensate for the loss and ensure the preservation of signal integrity
- Manageable Skew in PCB:
  - Skew in PCB can be effectively managed, providing more predictable performance. In contrast, cabled solutions have more complicated skew mechanisms, which needs to be addressed
- Advantages of PCB Routed Front Panel:
  - A PCB routed front panel offers several benefits, including lower cost, improved thermal performance, and reduced assembly complexity, when compared to a hybrid solution
- Advantages of Cabled Host:
  - Cabled host provides lower loss compared to PCB routed board and reduced in route density (skew managed)

# Find Paths to Both Implementations

The RX equalization capabilities required for both 36dB PCB and 32dB Cabled C2M channels are similar




Available Tools & Ongoing Studies:

- Reduction of noise (eta_0)
- Increase of taps and equalizer reach (UI)
- CDR sampling point optimization
- Relaxation of tap limits
- MLSE

cisco

# Conclusions and next step

- A crucial application: 2RU 64 ports switch
  - Ensuring consistent performance and supporting high-volume manufacturing is critical
  - Economic, power and performance considerations require PCB routed solution
  - Standardization shall support both PCB and cabled host implementations

- Call to action:
  - Collective efforts to enable both PCB and cabled implementations
    - Insertion loss: a minimum of 36dB C2M, must be analyzed in tandem with power optimization
    - Skew: study and establish the skew limit
    - 2x1 stacked connector: meeting SI performance requirements

| Table 93A-1 parameters | | | |
|---|---|---|---|
| **Parameter** | **Setting** | **Units** | **Information** |
| f_b | 106.25 | GBd | |
| f_min | 0.05 | GHz | |
| Delta_f | 0.01 | GHz | |
| C_d | [0.4e-4 0.9e-4 1.1e-4 ;0.4e-4 0.9e-4 1.1e-4 ] | nF | [TX RX] |
| L_s | [0.13 0.15 0.14; 0.13 0.15 0.14 ] | nH | [TX RX] |
| C_b | [0.3e-4 0.3e-4 ] | nF | [TX RX] |
| z_p select | [12] | | [test cases to run] |
| z_p (TX) | 1 1 1 1 1 1 1 1 1; 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ; | mm | [test cases] |
| z_p (NEXT) | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0; 0 0 0 0 0 0 0 0 0 0 0 0 | mm | [test cases] |
| z_p (FEXT) | 1 1 1 1 1 1 1 1 1; 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ; | mm | [test cases] |
| z_p (RX) | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0; 0 0 0 0 0 0 0 0 0 0 0 0 | mm | [test cases] |
| PKG_Tx_FFE_preset | 0 | | |
| C_p | [0.5e-4 0.5e-4] | nF | [TX RX] |
| R_0 | 50 | Ohm | |
| R_d | [50 50] | Ohm | [TX RX] |
| A_v | 0.5 | V | vp/vf= |
| A_fe | 0.5 | V | vp/vf= |
| A_ne | 0.5 | V | |
| L | 4 | | |
| M | 32 | | |
| **filter and Eq** | | | |
| f_r | 0.75 | *fb | |
| c(0) | 0.54 | | min |
| c(-1) | [-0.4:0.02:-0.3] | [-0.4:0.02:0] | [min:step:max] |
| c(-2) | [0:0.02:0.04] | [0:0.02:0.2] | [min:step:max] |
| c(-3) | [-0.04:0.02:0] | [-0.04:0.02:0] | [min:step:max] |
| c(-4) | [-0.02:0.02:0.04] | [0.02:0.02:0.02] | [min:step:max] |
| c(1) | [-0.04:0.02:0.04] | -0.12:0.02:0.04 | [min:step:max] |
| N_b | 1 | UI | |
| b_max(1) | 1 | | As/dffe1 |
| b_max(2...N-b) | [0.3 0.2*ones(1,22)] | | As/dffe2...N_b |
| b_min(1) | 0 | | As/dffe1 |
| b_min(2...N-b) | [-0.2 - 0.2*ones(1,22)] | | As/dffe2...N_b |
| g_DC | [-20:1:0] | dB | [min:step:max] |
| f_z | 42.5 | GHz | |
| f_p1 | 42.5 | GHz | |
| f_p2 | 106.25 | GHz | |
| g_DC_HP | [-6:1:0] | | [min:step:max] |
| f_HP_PZ | 1.328125 | GHz | |
| Butterworth | 1 | logical | include in fr |
| Raised_Cosine | 0 | logical | include in fr |
| RC_Start | 6.70E+10 | Hz | start freq for RCos |
| RC_end | 7.97E+10 | Hz | end freq for RCos |
| | | | |
| ffe_pre_tap_len | 4 | UI | |
| ffe_post_tap_len | 80 | UI | |
| ffe_tap_step_size | 0 | | |
| ffe_main_cursor_min | 0 | | |
| ffe_pre_tap1_max | 0.7 | | |
| ffe_post_tap1_max | 0.7 | | |
| ffe_tapn_max | 0.7 | | |
| ffe_backoff | 0 | | |
| | | | |
| Sample adjustment | [0 0] | phase | |
| ts_anchor | 0 | | |

| I/O control | | |
|---|---|---|
| DIAGNOSTICS | 1 | logical |
| DISPLAY_WINDOW | 1 | logical |
| CSV_REPORT | 1 | logical |
| RESULT_DIR | .\results\C2M_B_{date}\ | |
| SAVE_FIGURES | 0 | logical |
| Port Order | **[1 3 2 4]** | |
| RUNTAG | C2M_B | |
| COM_CONTRIBUTION | 0 | logical |
| **Operational** | | |
| ERL Pass threshold | 9.7 | dB |
| COM Pass threshold | 3 | db |
| VEC Pass threshold | 10.69073041 | db |
| DER_0 | 2.67E-05 | |
| T_r | 4.00E-03 | ns |
| FORCE_TR | 1 | logical |
| PMD_type | C2Mcom | |
| EW | 1 | |
| **TDR and ERL options** | | |
| TDR | 1 | logical |
| ERL | 1 | logical |
| ERL_ONLY | 0 | logical |
| TR_TDR | 0.01 | ns |
| N | 2000 | |
| TDR_Butterworth | 1 | logical |
| beta_x | 0 | |
| rho_x | 0.618 | |
| TDR_W_TXPKG | 0 | |
| N_bx | 0 | UI |
| fixture delay time | [ 0 0 ] | |
| Tukey_Window | 1 | |
| **Noise, jitter** | | |
| sigma_RJ | 0.01 | UI |
| A_DD | 0.02 | UI |
| eta_0 | 6.00E-09 | V^2/GHz |
| SNR_TX | 33 | dB |
| R_LM | 0.95 | |
| 11-2022 BenArtsi pkg | oif2022.065.02 | |
| highlighted are under re-consideration | | |
| | | |
| MLSE | 1 | |
| | | |
| AC_CM_RMS | 0 | |
| | | |
| 1.25E-08 | | |
| | | |
| **Batch control options** | | |
| BATCH_RUN | 1 | logical |
| CHANNEL_DIR | ..\Channels\All\ | |
| | | |
| | | |

| Table 93A–3 parameters | | |
|---|---|---|
| **Parameter** | **Setting** | **Units** |
| package_tl_gamma0_a1_a2 | [0 0.0008455 0.000340225] | |
| package_tl_tau | 0.00644805 | ns/mm |
| package_Z_c | [92 92 ; 70 70; 80 80; 100 100] | Ohm |
| | | |
| **Parameter** | **Setting** | |
| board_tl_gamma0_a1_a2 | *[0 6.44084e-4 3.6036e-05]* | 1.5 db/in @ 56G |
| board_tl_tau | *5.790E-03* | ns/mm |
| board_Z_c | 100 | Ohm |
| z_bp (TX) | *125* | mm |
| z_bp (NEXT) | *0* | mm |
| z_bp (FEXT) | *125* | mm |
| z_bp (RX) | *0* | mm |
| C_0 | [0.2e-4 0] | nF |
| C_1 | [0.2e-4 0] | nF |
| **Include PCB** | **0** | **logical** |
| | | |
| **Selections (rectangle, gaussian,dual_rayleigh,triangle** | | |
| Histogram_Window_Weight | gaussian | selection |
| Qr | *0.02* | UI |
| | | |
| **ICN parameters** | | |
| f_v | 0.594 | Fb |
| f_f | 0.594 | Fb |
| f_n | 0.594 | Fb |
| f_2 | 79.688 | GHz |
| A_ft | 0.450 | V |
| **A_nt** | 0.450 | V |
| | | |
| **Floating Tap Control** | | |
| N_bg | 0 | 0 1 2 or 3 groups |
| N_bf | 3 | taps per group |
| N_f | 120 | UI span for floating taps |
| bmaxg | 0.2 | max DFE value for floating taps |
| B_float_RSS_MAX | 0.1 | rss tail tap limit |
| N_tail_start | 61 | (UI) start of tail taps limit |
| | | |
| **Receiver testing** | | |
| RX_CALIBRATION | 0 | logical |
| Sigma BBN step | 5.00E-03 | V |

# COM 4.1

Values of parameters in the highlighted fields are being assessed.

cisco

# Thank you!