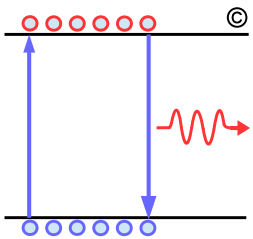


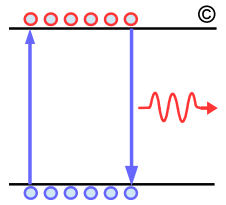
Evolution of Serial Ethernet Bitrate to 100 Gb/s

Ali Ghiasi
Ghiasi Quantum LLC

TEF 2016: The Road to Ethernet 2026

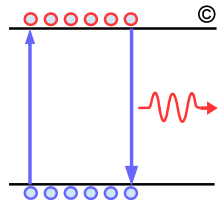
Sept 29, 2016

Overview



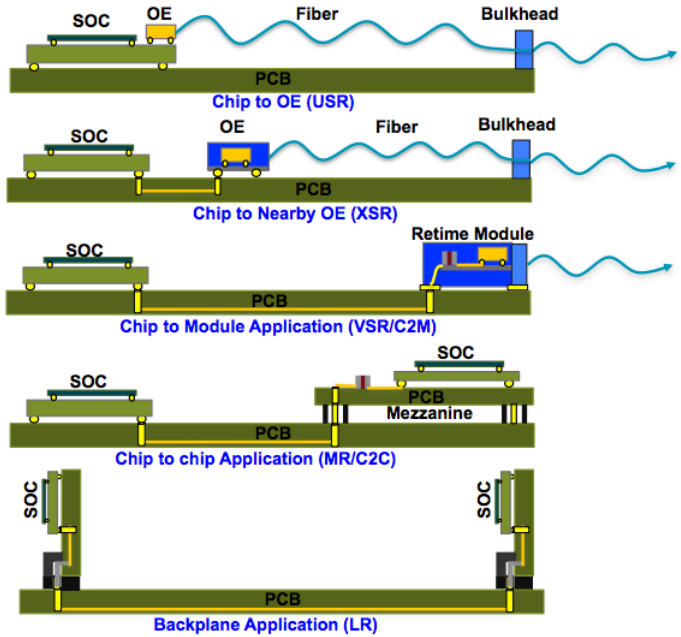
- **Current 50 Gb/s/lane and future 100 Gb/s/lane eco-system**
- **Evolution of signaling and front panel BW**
- **Pluggable at 50 Gb/s/lane and 100 Gb/s/lane**
- **Cu DAC support at 100 Gb/s/lane**
- **Backplane support at 100 Gb/s/lane**
- **What about on-board-optics "OBO" ?**
- **Analysis assume PAM4 for both 50 Gb/s and 100 Gb/s signaling**
  - Higher order PAM modulation not considered due to need for eco-cancellation, stronger FEC, larger latency, and higher power.

The 50G/lane Interconnect Ecosystems

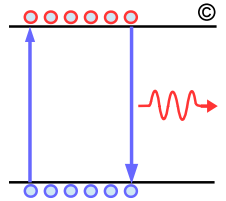


- The 25G/lane Ethernet eco-system in near future will be upgraded With 50G/lane PAM4 combined with RS(544,514) FEC
- IEEE 802.3bs are defining C2M, C2C, backplane, and Cu cabling based on PAM4
- OIF has defined both NRZ and PAM4 for MR, VSR, XSR; NRZ for XSR; and PAM4 for LR.

Application	Standard	Modulation	Reach	Coupling	Loss
Chip-to-OE (MCM)	OIF-56G-USR	NRZ	< 2 cm	DC	2 dB@28 GHz
Chip-to-nearby OE (no connector)	OIF-56G-XSR/ OIF-56G-XSR	NRZ/ PAM4	<10 cm <10 cm	DC DC	8 dB@28 GHz 4.2 dB@14 GHz
Chip-to-module (one connector)	OIF-56G-VSR/ IEEE CDAUI-8 OIF-56G-VSR	NRZ/ PAM4	<25 cm <25 cm	AC AC	18 dB@28 GHz 10 dB@13.3 GHz
Chip-to-chip (one connector)	OIF-56G-MR/ IEEE CDAUI-8 OIF-56G-MR	NRZ/ PAM4	< 50 cm < 50 cm	AC AC	35.8 dB@28 GHz 20 dB@14 GHz
Backplane (two connectors)	OIF-56-LR IEEE 50G-KR	PAM4 PAM4	<100 cm <100 cm	AC AC	27.5dB@14 GHz 30dB@13.3 GHz



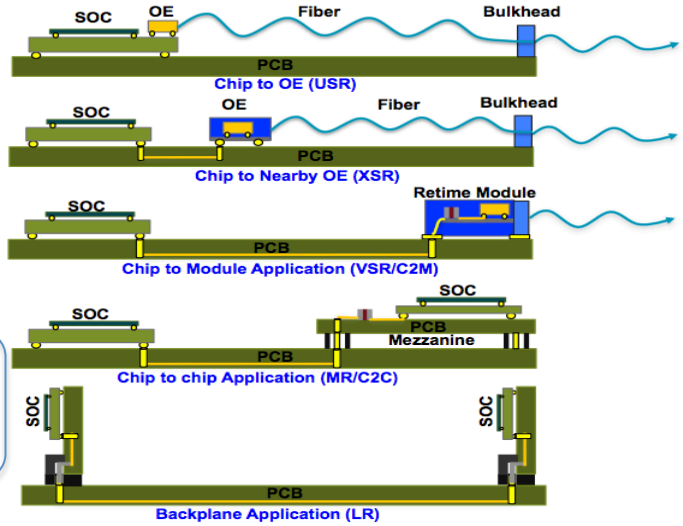
The 100G/lane Interconnect Ecosystems



- **OIF kicked off 100G/lane activity by organizing a serial 100G workshop co-located with OFC 2016**
- **OIF started 112G-VSR project during Q3-2016 meeting – OIF project typically takes ~ 3 years**
- **Facing major obstacle as we migrate to 100G/lane:**
  - Lack of technically viable passive Cu directly driven from switch ASIC
    - Will drive the market to higher cost Active DAC and AOC
  - Conventional backplane need to be replaced with higher cost cabled backplane
  - Chassis system with backplane will continue to declining in favor of 1 RU/ 2RU boxes
  - On board optics "OBO" coupled with modular boxes will enable more scalable lower cost network.

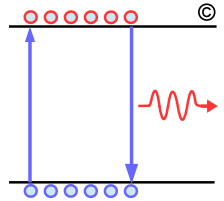
Application	Standard	Modulation	Reach	Coupling	Loss*
Chip-to-OE (MCM)	TBD	PAM4	< 2cm	DC	~2 dB@28 GHz
Chip-to-nearby OE (no connector)	TBD	PAM4	<10 cm	DC	~8 dB@28 GHz
Chip-to-module (one connector)	OIF-112G-VSR	PAM4	< 25 cm	AC	~18 dB@28 GHz
Chip-to-chip (one connector)	TBD	PAM4	< 50 cm	AC	~30 dB@28 GHz
Cabled Backplane (two connectors)	TBD	PAM4	<100 cm	AC	~30dB@28 GHz

Chip to chip loss is identical to cabled backplane loss

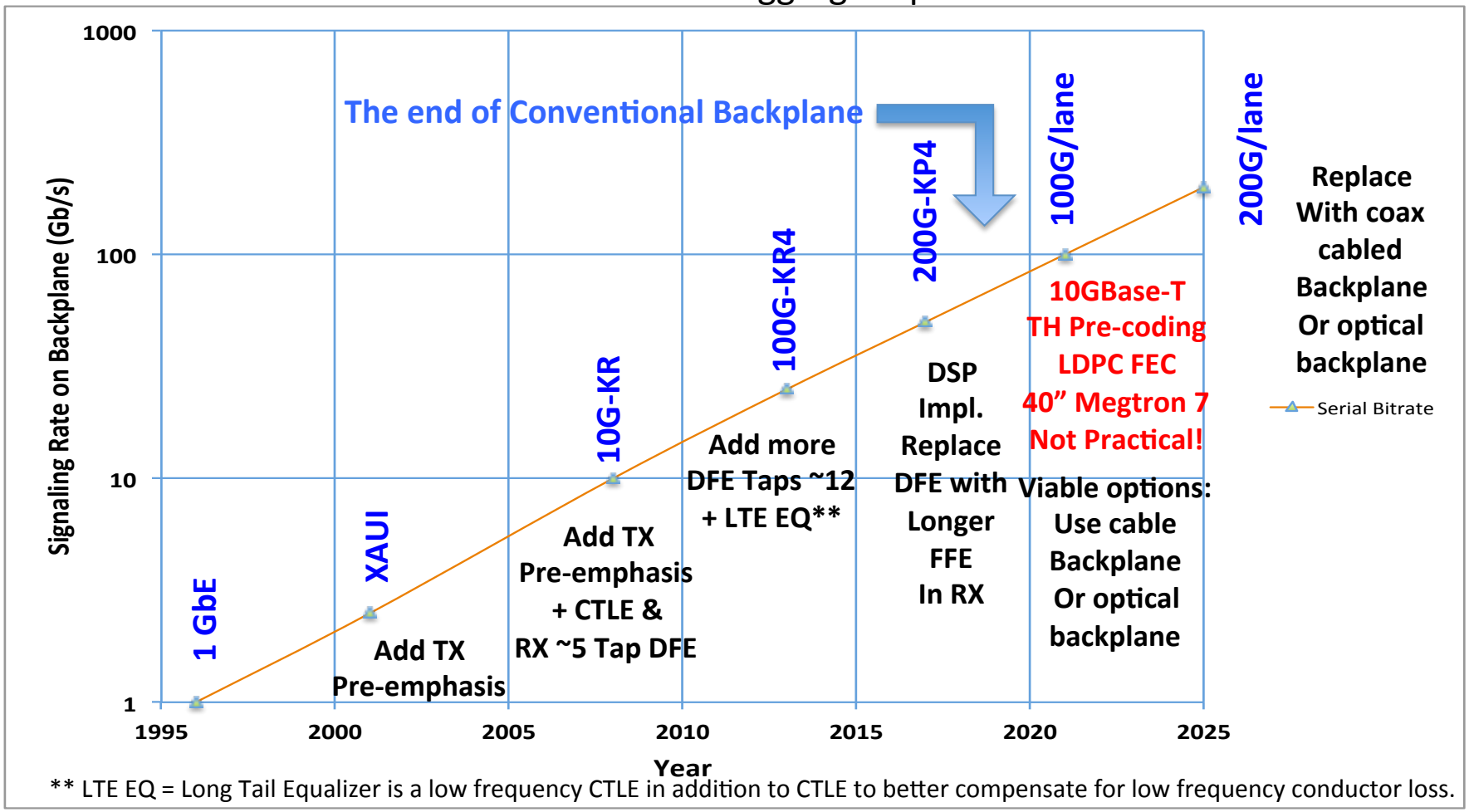


* Losses are the author best guess.

Evolution of Serial Bit Rate

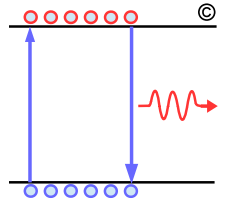


- **Since 2008 common SerDes have been used for backplane and Cu DAC**
  - Serial bit for x4 interfaces are ¼ of aggregate port bit rate

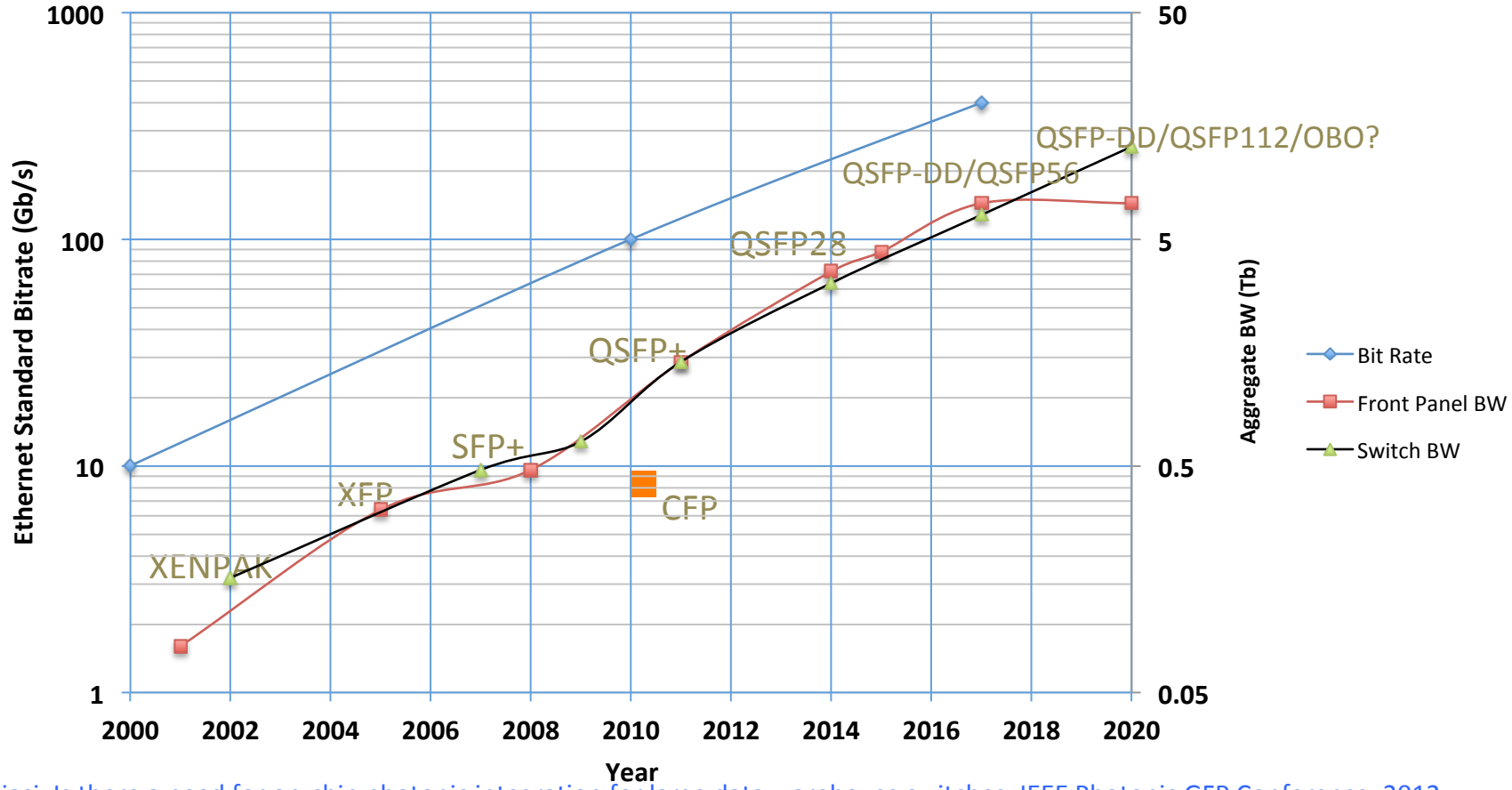


** LTE EQ = Long Tail Equalizer is a low frequency CTLE in addition to CTLE to better compensate for low frequency conductor loss.

Evolution of the Front Panel Bandwidth

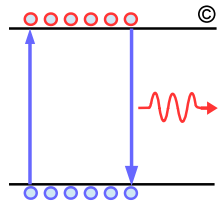


- **When will OBO be necessary?**
  - In 2012 at GFP* conference I showed front panel BW may limited to 3.6 Tb
  - With emergence of QSFP-DD and future QSFP112 there is path to 14.4 Tb front panel BW
  - OBO is somewhat interesting at 100G/lane by simplifying electrical equalization to CTLE
    - But OBO does not enable removal of power hungry CDR/Mux in the module!



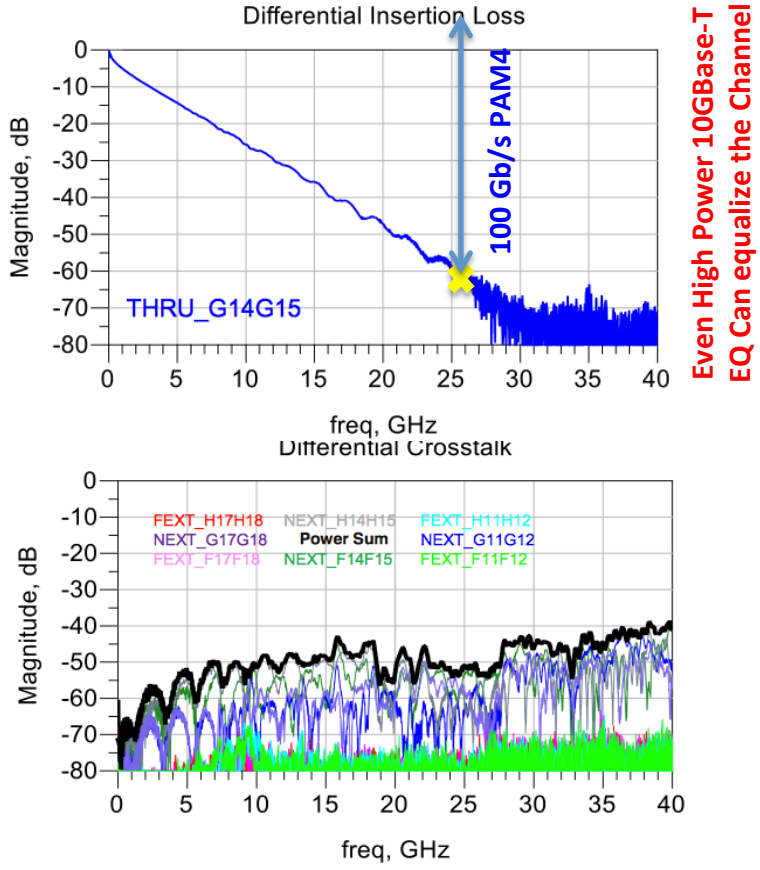
*Ali Ghiasi, Is there a need for on-chip photonic integration for large data warehouse switches, IEEE Photonic GFP Conference, 2012 .

Conventional Backplane no Longer Feasible at 100 Gb/s!

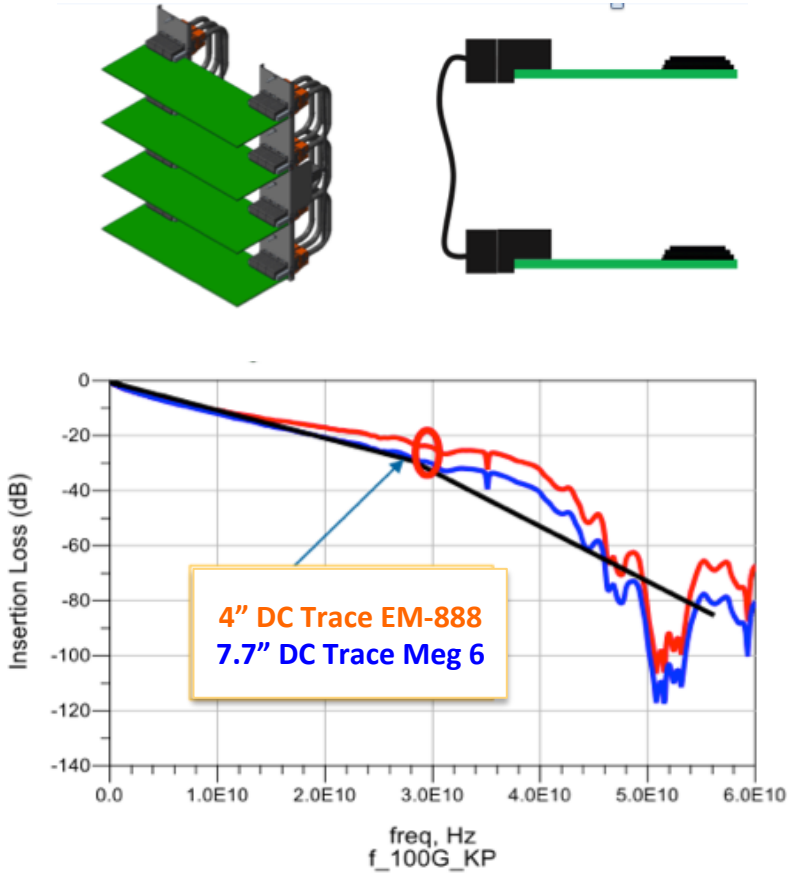


- **TE Whisper 40” conventional backplane at 100 Gb/s PAM4 Nyquist has a loss of ~65 dB ***
- **1 m cabled backplane is viable with short daughter-card, in effect every lane needs a retimers!**

TE Whisper Conventional Backplane 40” with Meg 6 HVLP *



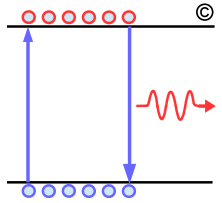
TE Whisper 1 m Cabled Backplane **



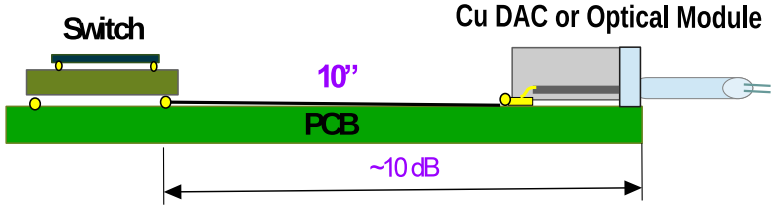
* TE Whisper channel, http://www.ieee802.org/3/cd/public/channel/Reference_document_for_TE_Connectivity_Backplane_S-Parameter_Channels_07_28_16.pdf

** Achieving 100 Gb/s Channels, David Hester TE Connectivity, OIF 2016 100 Gb/s Workshop.

Evolution of Pluggable Modules

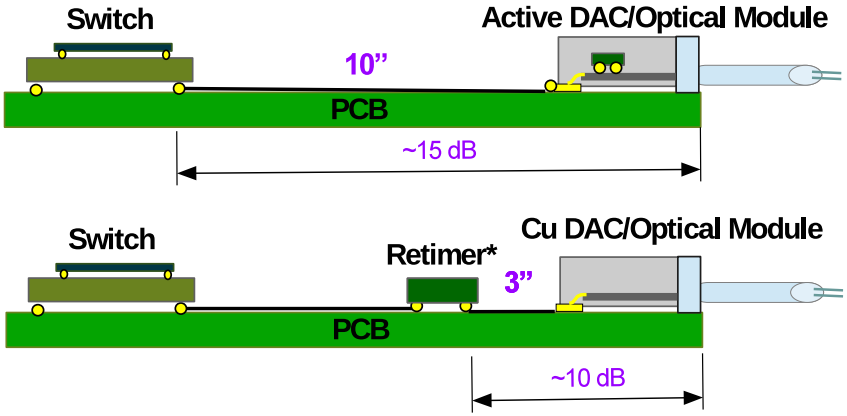


Pluggable at 25 Gb/s and 50 Gb/s



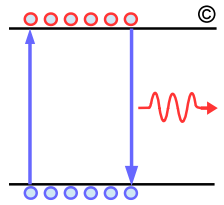
- PHY less design – what we are used to
  - Supports passive Cu DAC
  - Switch directly drives optical modules
  - Switch directly drives 3 m of Cu DAC
  - Offers optimum power and cost.

Pluggable at 100 Gb/s

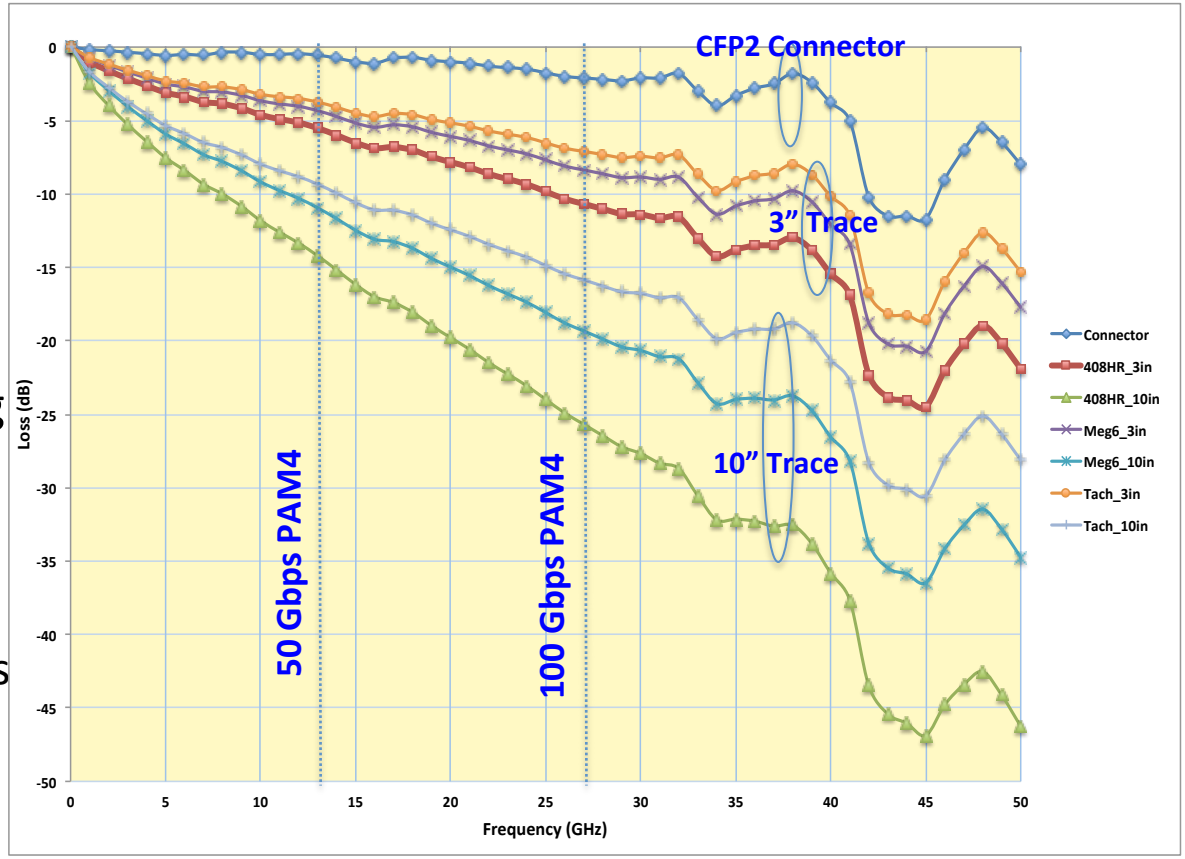


- Option I – PHY less design
  - Doesn’t support passive Cu DAC
  - Switch directly drives pluggable module, active Cu DAC, or AOC
  - Support 10” of Megtron 7/Tachyon PCB
  - Offers improve power and cost
- Option II – Require PHY close to every module
  - Supports passive Cu DAC, active DAC, and AOC Support 3” of Megtron 7/Tachyon PCB
  - Supports Active Cu DAC and optical modules
  - Retimer adds significant cost and power.

Extending Chip-to-Module (VSR) loss from 50 to 100 Gbps

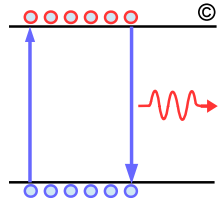


- **Connector assumed to be Yamaichi CFP2 capable of 50 Gbaud (100 Gbps with PAM4) operation***
  - C2M channel loss investigated with following material 408HR, Megtron 6 HVLP, Tachyon HVLP for 5.5 mil ½ oz stripline
  - Host ASIC directly driving the front panel 10" trace
    - Current 50 Gbps C2M loss is 10 dB
    - At 100 Gbps C2M loss is ~18 dB
  - A retimer driving the front panel 3".

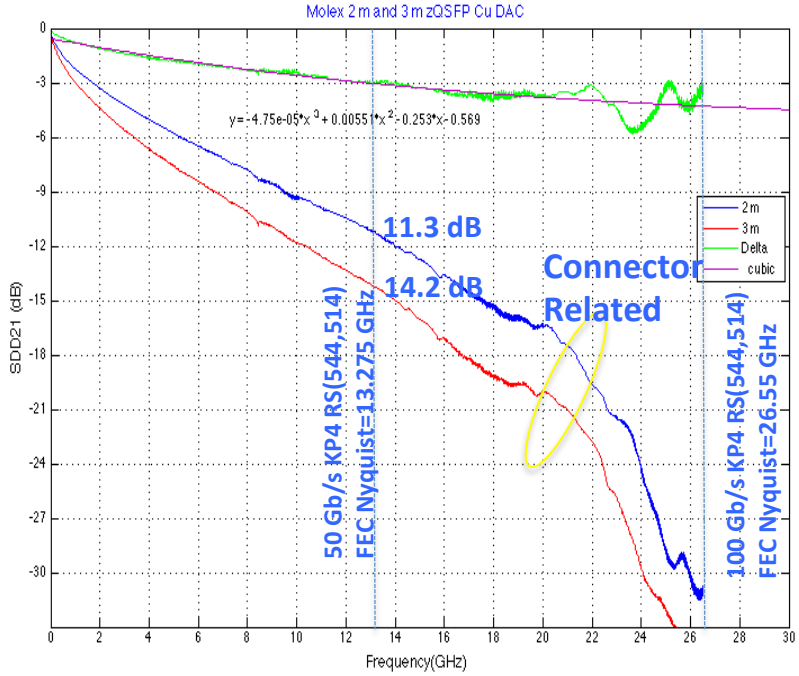


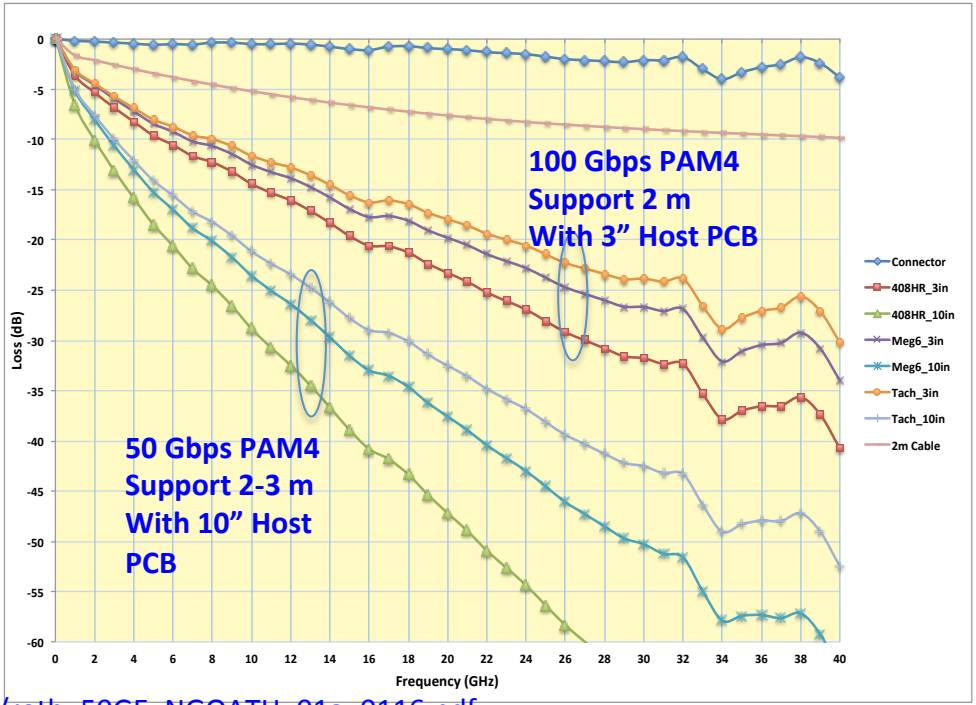
* CFP2 connector, http://www.ieee802.org/3/400GSG/public/13_05/nishimura_400_01a_0513.pdf

Extending Cu DAC Operation from 50 to 100 Gbps



- **Construction of the hypothetical 100 Gb/s Cu DAC**
  - De-embed Molex zQSFP cable response [2]
  - Use two of Yamaichi CFP2 connector response to create hypothetical Cu [3]
  - Hypothetical 2 m Cu DAC with 10” trace has loss of ~50 dB instead 3” host could be supported with end-end loss of ~ 30 dB
  - The 3” host PCB requires adding 32 retimers will double the line card cost and power
  - A better solution is to go with 10” PCB (PHY-less) and instead go with active DAC or AOC.

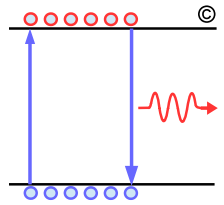




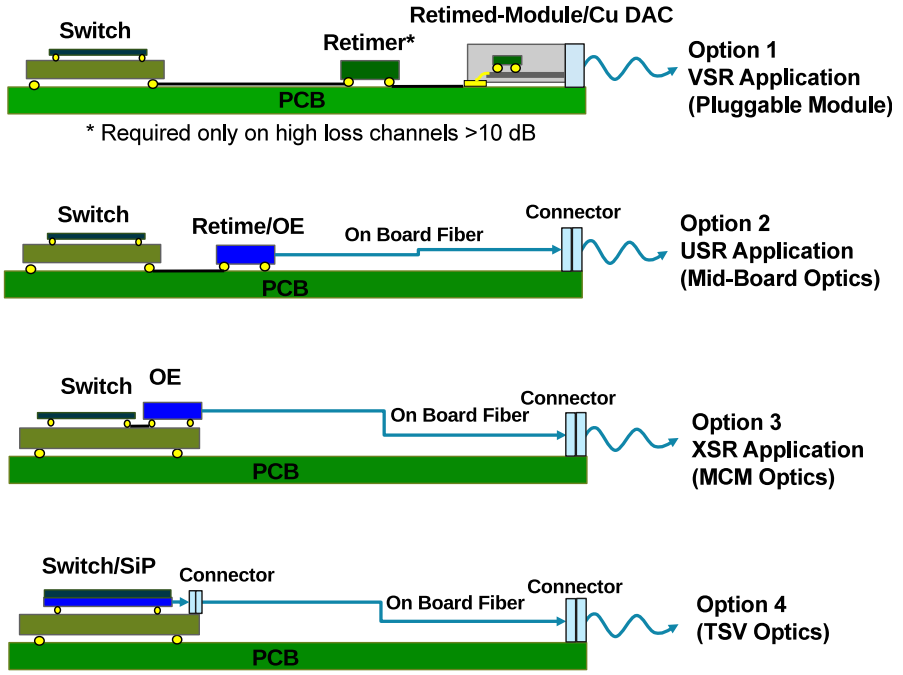
*zQSFP cable data, http://www.ieee802.org/3/50G/public/Jan16/roth_50GE_NGOATH_01a_0116.pdf

**CFP2 connector, http://www.ieee802.org/3/400GSG/public/13_05/nishimura_400_01a_0513.pdf

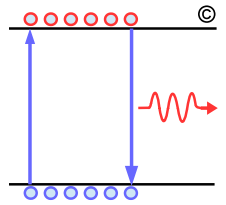
Toward the Holy Grail



- **When are we going to see manufacturable OBO or co-packaged where there is cost, power, and density advantages?**
  - During DOT.COM era OBO were very popular till you start manufacturing!
  - The SFP's and QSFP's have delivered the required front panel BW offering flexibility and low cost Cu DAC
  - Mid-board optics at 50G/lane may not provide sufficient advantage over QSFP56/QSFP-DD with negligible power saving and eliminating low cost Cu DAC!
  - Mid-board optics at 100 Gb/s main advantage will be simpler equalizer assuming 10 dB vs 18+ dB channel
  - Eliminating the retimer in the optics deliver the biggest power saving but require the an MCM or stack-die implementation!



Summary



- **With 802.3bs and 802.3cd defining C2M, C2C, backplane, and Cu cabling expect to see 50 Gb/s PAM4 eco-system to follow the foot step of the 25G**
- **OIF has recently started the CEI-112G-VSR (C2M) project**
- **Unless we discover the Holy Grail of cheap WDM integrated optics sooner than later we will need 100G/lane IO**
- **The transition to serial 100G/lane will not be smooth like 50G/lane transition**
  - Even with material like Megtron 7 or Tachyon 100G C2M the 10" C2M (VSR) loss will be ~18 dB
  - At 18 dB loss the retimers in the module equalizer need to be comparable to C2C equalizer and not just a simple low power CTLE
  - Passive DAC likely will not be supported as it require retimers close to each cage
  - Active DAC or AOC will fill the passive DAC void as the system gets optimized for cost and power
  - PHY-less host design with 10" trace coupled with QSFP-dd or QSFP112 still is very compelling solution and delivers 14.4 Tb/s capacity
- **To achieve any significant power saving over pluggable, require an MCM or a TSV implementation where the retimers are eliminated**
  - OBO at 100 Gb/s/lane with ~10 dB PCB loss offer incremental power advantage by allowing to stay with simple CTLE style of equalizer.