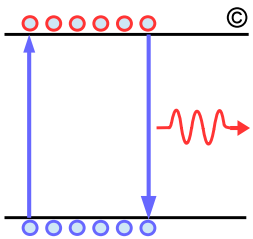


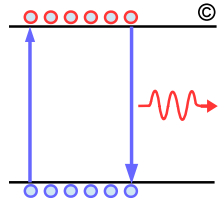
50 GbE, 100 GbE and 200 GbE PMD Requirements

Ali Ghiasi
Ghiasi Quantum LLC

NGOATH Meeting
Atlanta

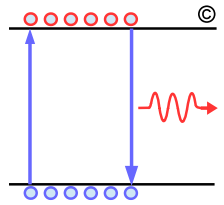
January 20, 2015

Observation on 50GbE, 200 GbE, and NG 100GbE PMDs



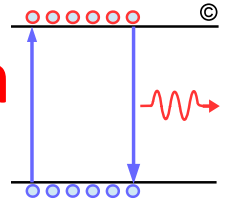
- **50 GbE and 200 GbE are complimentary set of standards just as we observed in the market place the complimentary nature of 25GbE/100 GbE**
  - Current generation of switch ASIC offer 4x25 GbE breakout for small incremental cost
  - Next generation switch ASIC will offer 4x50GbE breakout for same economics
  - Complete eco-system require backplane, Cu cable, 100 m MMF, possibly 500 m PSM4, 2000 m, and 10,000 m PMDs and should follow 25/100GbE PMDs
- **NG 100 GbE PMDs attributes and requirements**
  - Currently with the increase in volume the market is enjoying significant cost reduction for 100 GbE PMDs such as 100GBase-SR4, PSM4, and CWDM4/CLR4
  - Cost may not be the main driver to define NG 100 GbE PMDs with exception of CAUI-2
  - Currently defined 100 GbE PMDs will require inverse-mux with introduction of 50G ASIC IO
    - A PMA-PMA device could address any I/O mismatch
    - Simplest form of PMA/PMD implementation occurs for the case when # of electrical lanes = # of optical lanes/λ
  - Do we need with every generation of electrical I/O 25G, 50G, 100G introduce new 100 GbE PMDs which are optimized for given generation of ASIC but not optically backward compatible
  - The decision to define new optical PMD should not be taken lightly to save a PMA-PMA mux!

Today's Ethernet Market Isn't Just about Enterprise

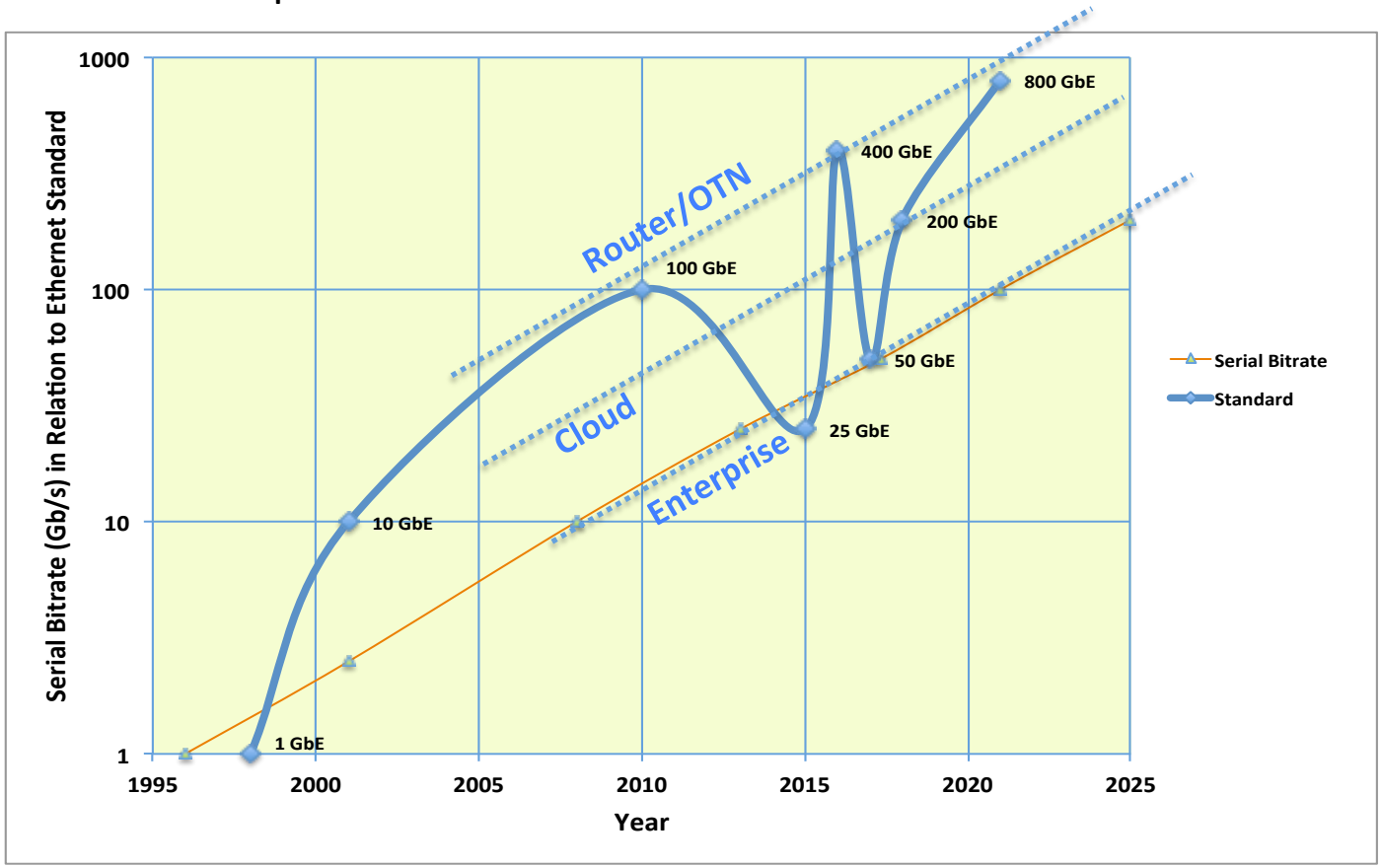


- **Router/OTN**
  - Leads the deployment with fastest network interface
  - Drives bleeding edge technology at higher cost and lower density
- **Cloud data centers**
  - Clos fabrics typically operate at lower port speed speed to achieve switch ASIC radix of 32 or 64
  - Drives cost-power-density to enable massive data center built out
  - Forklift upgrade doubling capacity every ~2.5 years with doubling of switch ASIC capacity
- **Enterprise**
  - Enjoys the volume-cost benefit of deploying previous generation of Cloud Data Centers technology
  - More corporate IT services are now hosted by the Cloud operator
  - According to Goldman Sachs research from 2013-2018 Cloud will grow at rate of 30% CAGR compare to 5% for Enterprise IT
    - http://www.forbes.com/sites/louiscolumbus/2015/01/24/roundup-of-cloud-computing-forecasts-and-market-estimates-2015/.

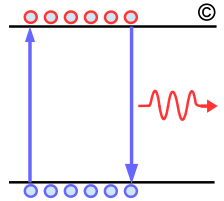
Ethernet Serial Bitrate and Port Speed Evolution



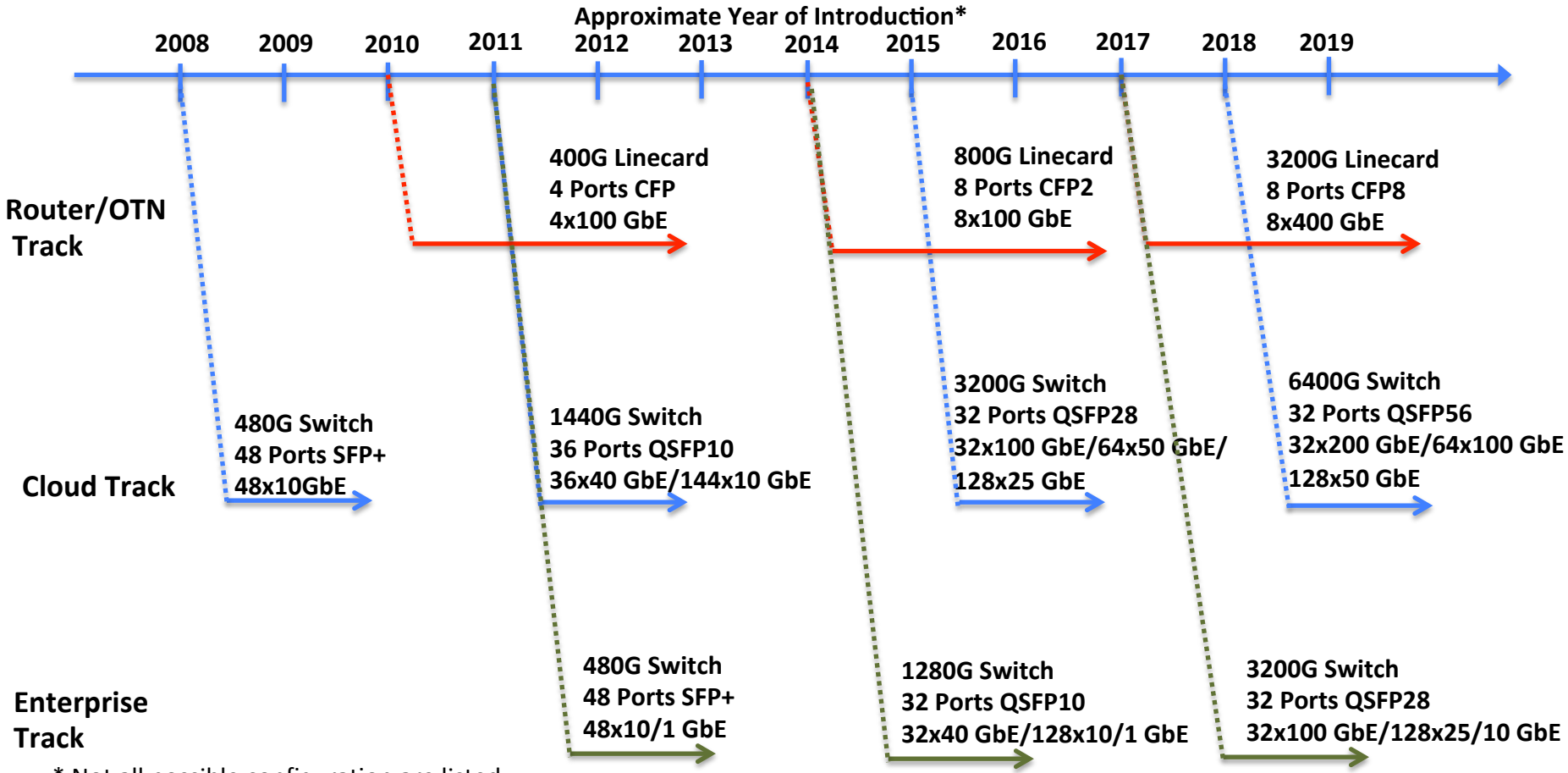
- **Router/OTN, Cloud, vs Enterprise applications**
  - NGOATH project addresses next generation Cloud and Enterprise
  - 50 GbE is not only an interface on Cloud server but also a replacement for 40 GbE in the Enterprise.



Evolution of Ethernet Speed-Feed



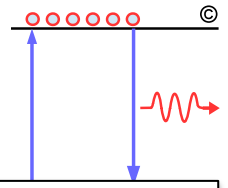
- **NGOATH project is addressing the need for next generation Cloud track**
  - P802.3.bs addressing the need for next generation Router/OTN track
  - 25G SMF project addressing the need of next generation Enterprise/campus

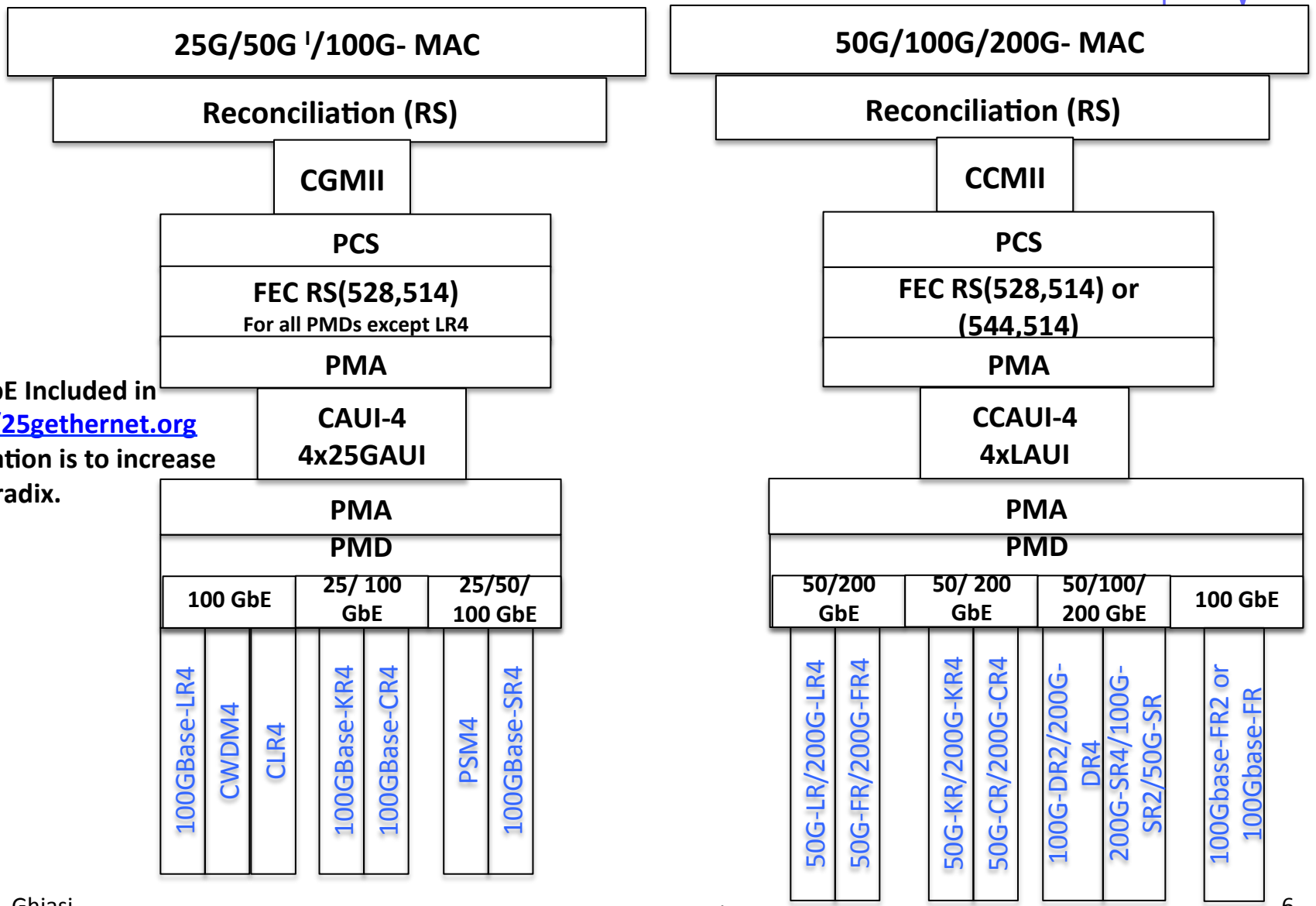


* Not all possible configuration are listed.

A. Ghiasi

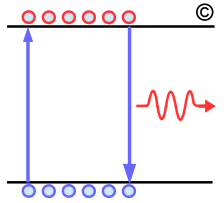
Current and Next NGOATH PMDs



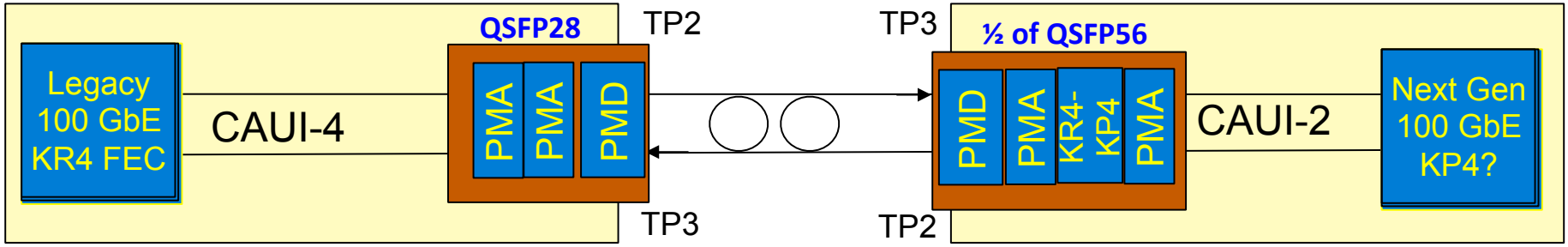


I.50 GbE Included in http://25gethernet.org application is to increase fabric radix.

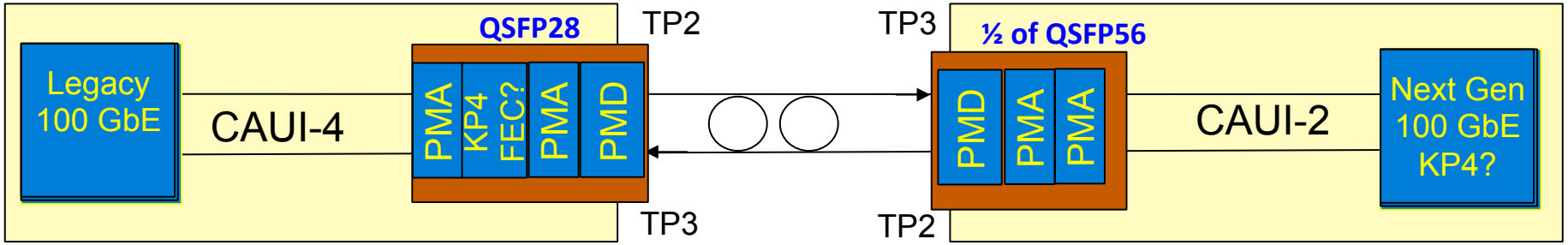
The Challenge with 100 GbE Next Gen PMDs



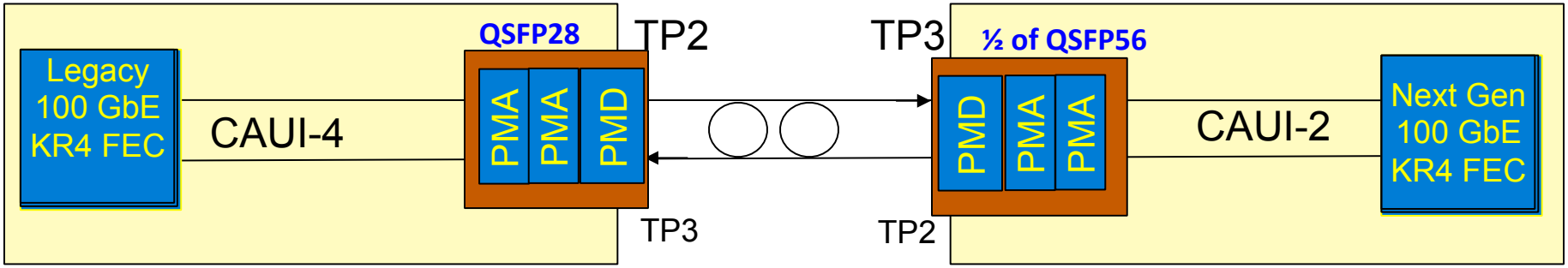
- **Approach support existing 100 GbE PMD**



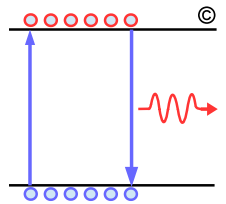
- **Approach to support new 100 GbE PMDs**



- **The simplest approach if feasible is to define new 100 GbE PMDs based on KR4 FEC.**

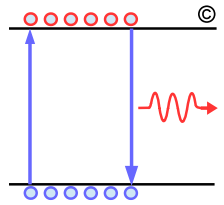


Observation: 50/200 GbE are the enablers while 100 GbE is a nice addition



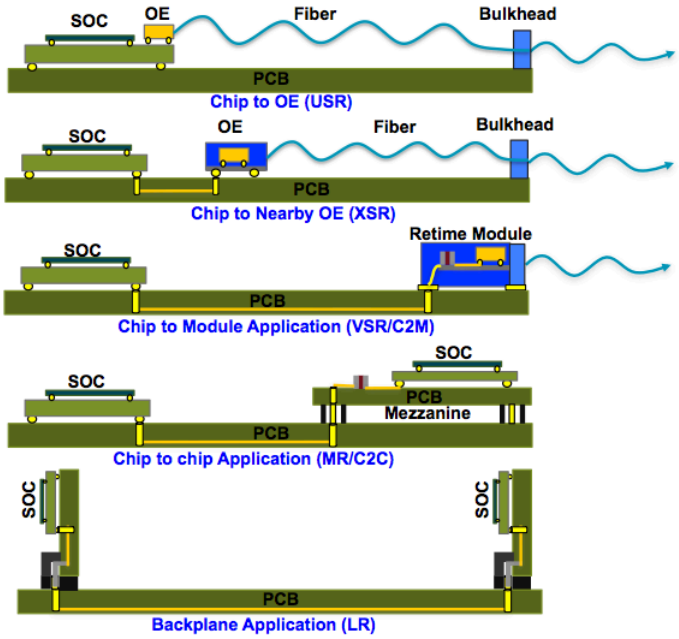
- **200GbE PMDs (Application next generation Cloud Data Center)**
  - 200Gbase-LR4 based on CWDM4 enables next generation uncooled low cost PMD
    - Does 200Gbase-FR4 offers significantly lower cost solution to define separate PMD?
  - 200Gbase-DR4 offers 200 GbE as well as 50/100 GbE breakout
  - 100Gbase-SR4 offers 200 GbE as well as 50/100 GbE breakout
  - 200Gbase-KR4 with 30+ dB required to meet 1 m backplane
    - Backplane loss will determine exact cable reach of 3 to 5 m
- **100GbE PMDs (Application doubling radix in Cloud and could be a better match with next generation 50G ASICs IO)**
  - 100Gbase-LR2 to enables next generation uncooled low cost PMD
    - Does 100Gbase-FR4 offers significantly lower cost solution to define separate PMD?
  - 100Gbase-DR2 use ½ of the 200Gbase-DR4
  - 100Gbase-SR2 options: use ½ of the 200Gbase-SR4 or define a dual-λ duplex
  - Too early to define serial 100 Gb/s and no need to define 2 lanes Cu KR2/CR2
- **50GbE PMDs (Next generation servers and next Gen Enterprise/campus)**
  - 50Gbase-LR required for the campus and access application
  - Do we need to define both 50Gbase-FR and 50GBase-DR?
  - 50Gbase-SR
  - 200Gbase-KR4 with 30+ dB required to meet 1 m backplane
    - Backplane loss should determine the exact cable reach 3 to 5 m.

50 Gb/s/lane Interconnect Space

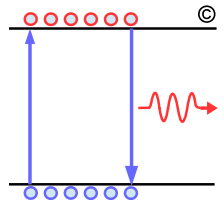


- **OIF has been defining USR, XSR, VSR, MR, and LR**
  - OIF-56G-LR is good starting point but does not support practical 1 m backplane implementation but 27.5 dB is insufficient to build practical backplanes!

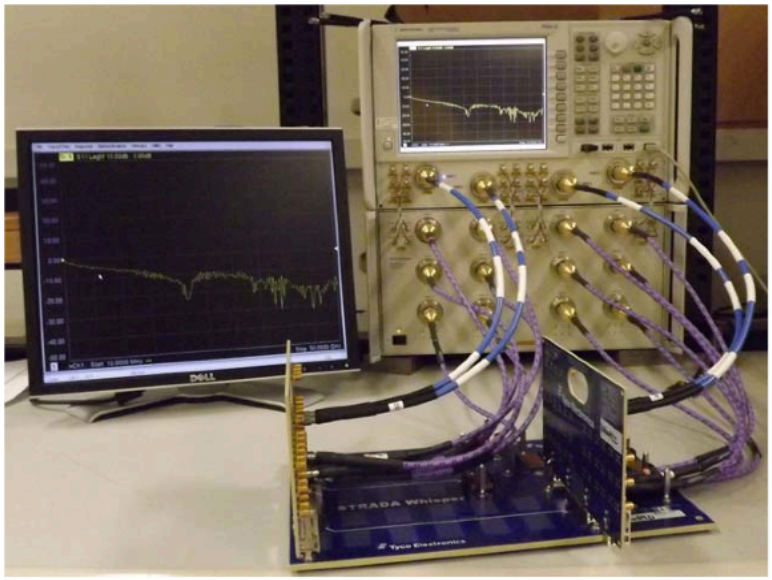
Application	Standard	Modulation	Reach	Coupling	Loss
Chip-to-OE (MCM)	OIF-56G-USR	NRZ	< 1cm	DC	2 dB@28 GHz
Chip-to-nearby OE (no connector)	OIF-56G-XSR	NRZ/ PAM4	<5 cm	DC	8 dB@28 GHz 4.2 dB@14 GHz
Chip-to-module (one connector)	OIF-56G-VSR	NRZ/PAM4	< 10cm	AC	18 dB@28 GHz 10 dB@14 GHz
	IEEE CDAUI-8	PAM4	< 10 cm	AC	10 dB@13.3 GHz
Chip-to-chip (one connector)	OIF-56G-MR	NRZ/PAM4	< 50 cm	AC	35.8 dB@28 GHz 20 dB@14 GHz
	IEEE CDAUI-8	PAM4	< 50 cm	AC	14 dB@13.3 GHz
Backplane (two connectors)	OIF-56-LR IEEE	PAM4	<100 cm 100 cm	AC AC	27.5dB@14 GHz x dB@13.275



TE Whisper 40” Backplane
“The Gold Standard”



See: http://www.ieee802.org/3/bj/public/jul13/tracy_3bj_01_0713.pdf



H11-H12	H14-H15	H17-H18
G11-G12	G14-G15	G17-G18
F11-F12	F14-F15	F17-F18

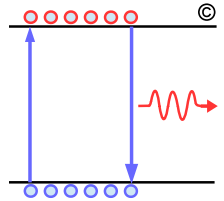
- All data is measured and includes 2.4mm test points
- Measurements are pair G14-G15 centric .s4p files
- 4 Near-End and 4 Far-End measurements
- Data is from 0-30GHz in 10MHz steps

DAUGHTER CARD
- Board Material = Megtron6 VLP
- Trace length = 5”
- Trace geometry = Stripline
- Trace width = 6 mils
- Differential trace spacing = 9 mils
- PCB thickness = 110mils, 14 layers
- Counterbored vias, up to 6mil stub
- Test Points = 2.4mm (included in data)

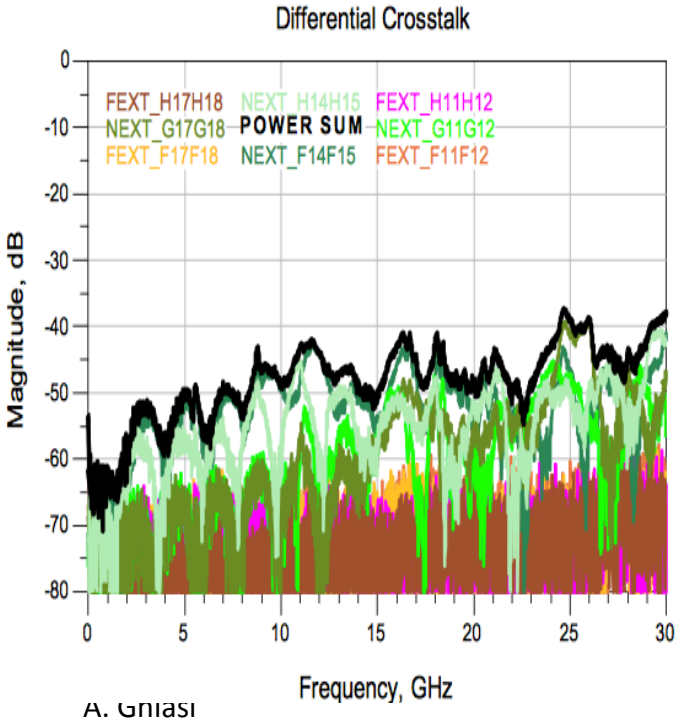
BACKPLANE
- Board Material = Megtron6 HVLP
- Trace length = 30”
- Trace geometry = Stripline
- Trace width = 6 mils
- Differential trace spacing = 9 mils
- PCB thickness = 200 mils, 20 layers
- Counterbored vias, up to 6mil stub

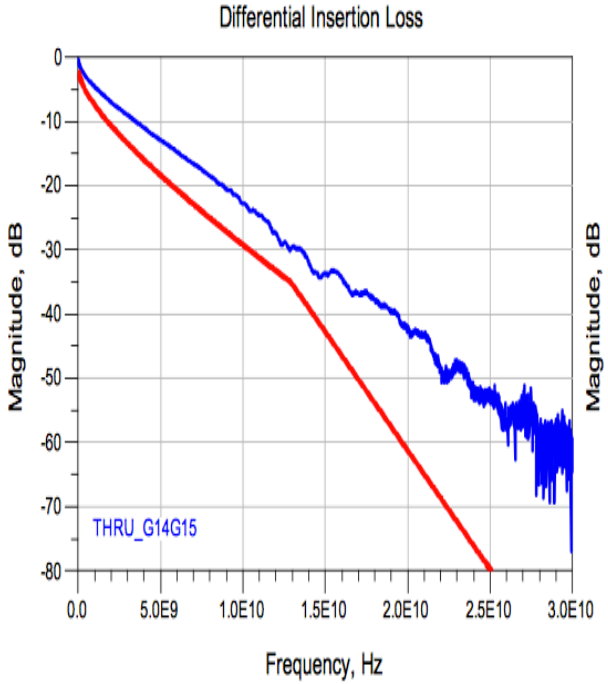
CONNECTORS
- Dataset 1 includes
  - Mated standard STRADA Whisper connector at each end
- Dataset 2 includes
  - Mated Embedded Capacitor STRADA Whisper connector at one end and,
  - Mated standard STRADA Whisper connector at other end

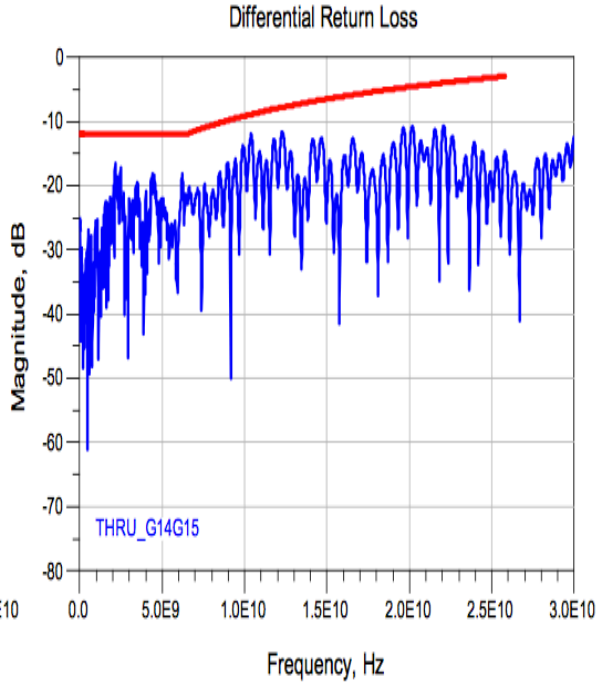
Response of 40” TE Whisper Backplane with Megtron 6



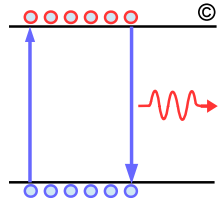
- 30” backplane Megtron 6 HVLP with 6 mils traces
  - For dense application more practical trace width will be 4.5-5 mils
- Daughter cards 5” each Megtron 6 VLP with 6 mils traces
- The loss is ~30 dB at 12.87
- Actual implementation may need to use narrower traces like 4-5 mils increasing the loss further
- With backplane not shrinking 30-32 dB loss is required for practical line cards.







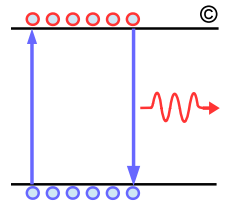
25G/50G Channel Summary Results for TE Whisper 1 m Backplane



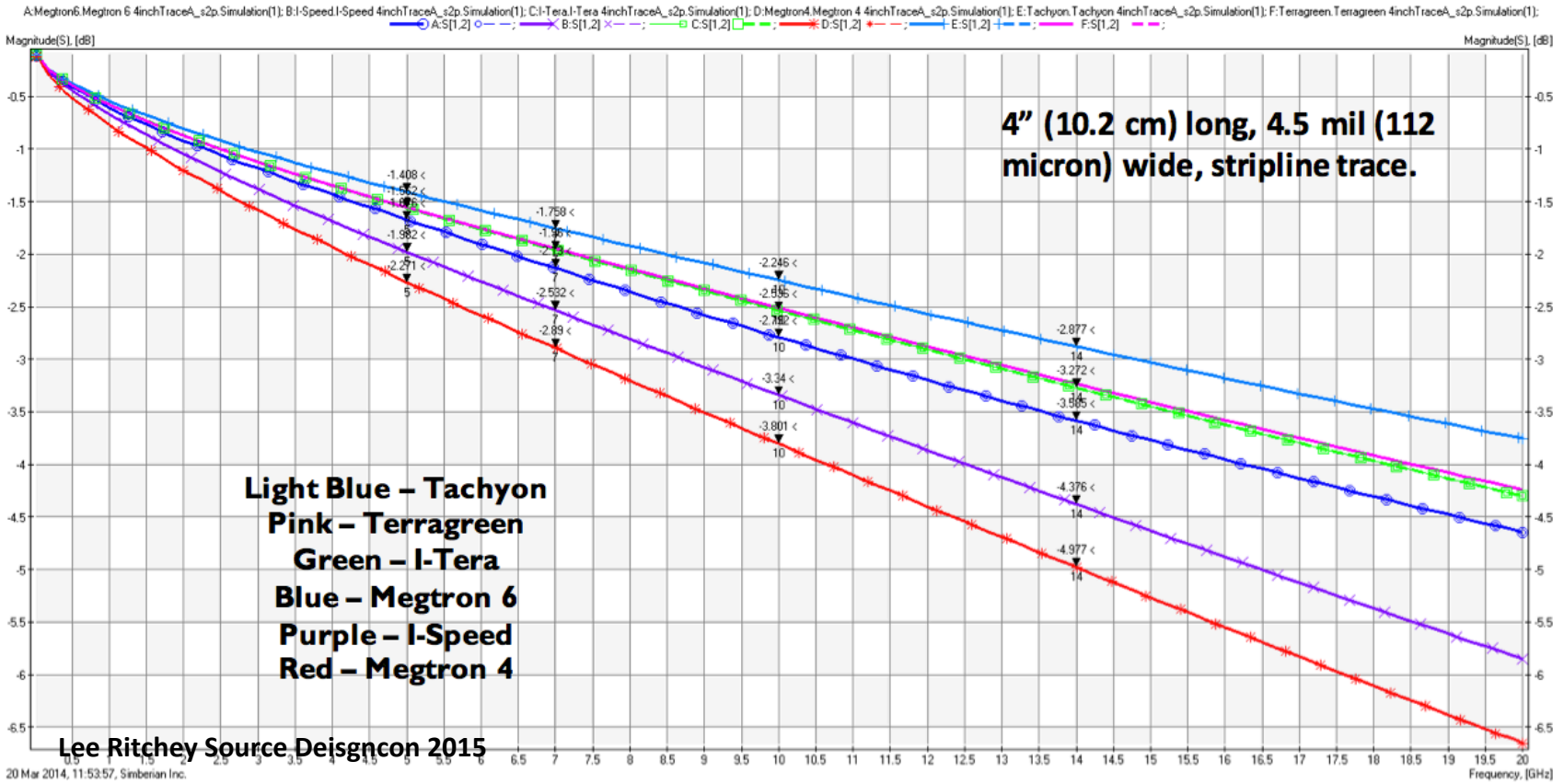
- **Closing the link budget on 30 dB channel with 2 dB COM margin is not trivial**
  - A loss reduction also not an option given TE backplane trace width rather wide of 6 mils where typical linecard trace would be in 4.5-5 mils!

Test Cases	Channel IL (dB)	Channel + PKG IL (dB)	ILD	ICN (mV)	PSXT (mV)	COM (dB)
25G NRZ With IEEE 12 mm Package	28.4	30.4	0.37	1.60	4.0	5.5
25G NRZ With IEEE 30 mm Package	28.4	32.5	0.37	1.63	3.3	4.8
25G PAM4 With IEEE 12 mm Package	16.4	17.1	0.05	0.98	2.0	5.7
25G PAM4 With IEEE 30 mm Package	16.4	18.1	0.05	0.98	1.8	5.7
50G PAM4 With IEEE 12 mm Package	29.7	34.7	0.41	1.65	3.1	
50G PAM4 With IEEE 30 mm Package	29.7	36.7	0.41	1.64	2.66	

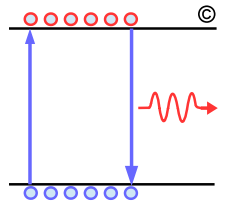
More Advance PCB Material Only Modestly Improves the Backplane Loss



- **Even moving from Megtron 6 DF~0.005 to Tachyon DF~0.0021 the loss only improves by ~20%**
  - With DF≤ 0.005 loss is now dominated by conductor size and roughness.



Summary



- **In 802.3 today we need address 3 markets**
  - Router/OTN bleeding edge technology and speed
  - Cloud driven by fork lift upgrade as result of switch BW doubling every ~2.5 years
  - Enterprise leveraging last generation cloud technology
- **50/200 GbE offer optimum solution set for next generation cloud with 50 GbE for servers and 200 GbE for fabrics**
  - In current data center build out 50 GbE (25G MSA) is deployed to double radix and fabric capacity
    - In next generation data centers high density 100 GbE likely will be deployed to build ultra scale fabric
- **Next Gen 100 GbE PMDs can be based on 200 GbE PCS/FEC or it can be defined to be backward compatible using Clause 82 PCS and KR4 FEC**
  - The advantage of using common FEC for 100 GbE and 200 GbE is to achieve identical performance for a PMD operating in full rate or break out mode
  - Considering the investment made in current 100 GbE PMDs backward compatibility should an important consideration
- **To enable next generation 6.4 Tb line card, the backplane based on improved FR4 material must operate at 50 Gb/s/lane**
  - A minimum loss of 30 dB is required for construction of 1 m conventional backplane
- **The 802.3 need to balance cloud applications driven by fork lift upgrade as well as synergy and compatibility across Ethernet eco-system.**