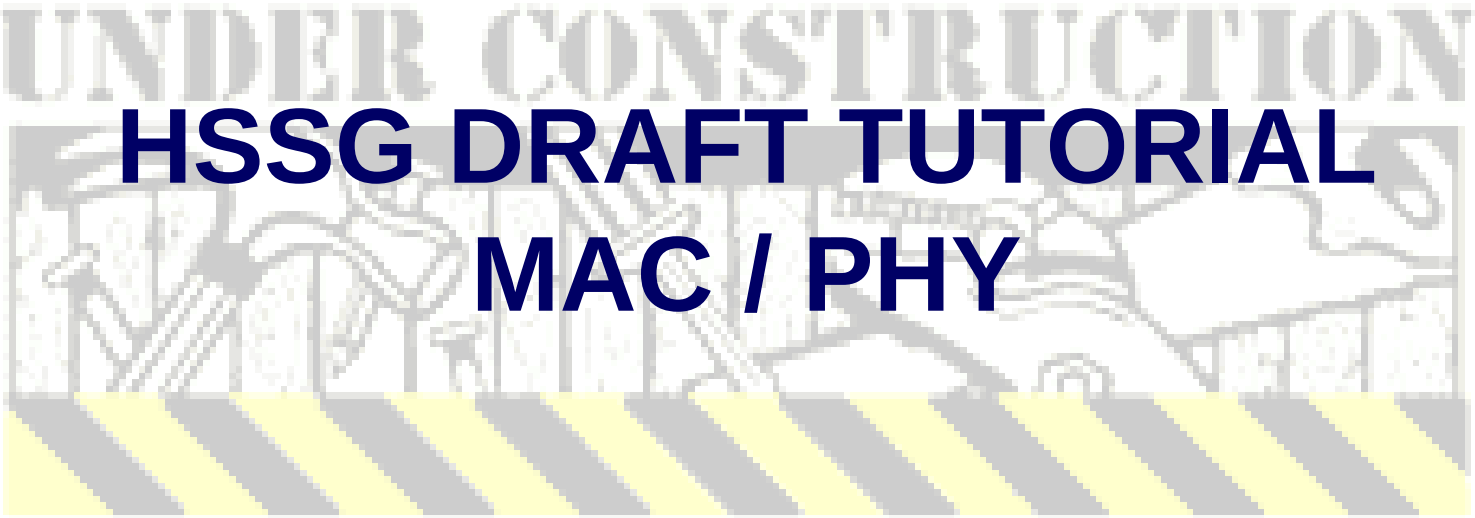
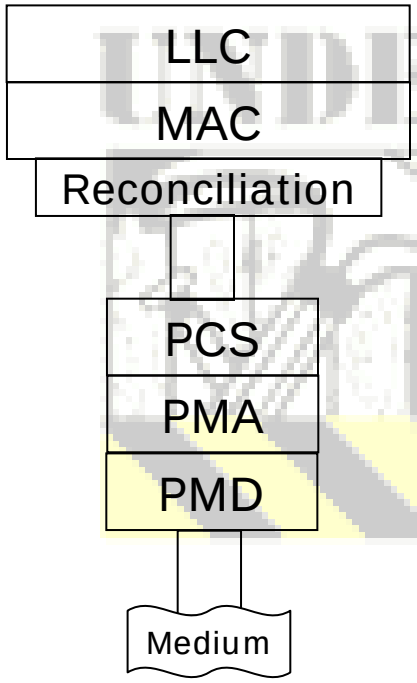


HSSG DRAFT TUTORIAL
MAC / PHY

John Jaeger, Infinera

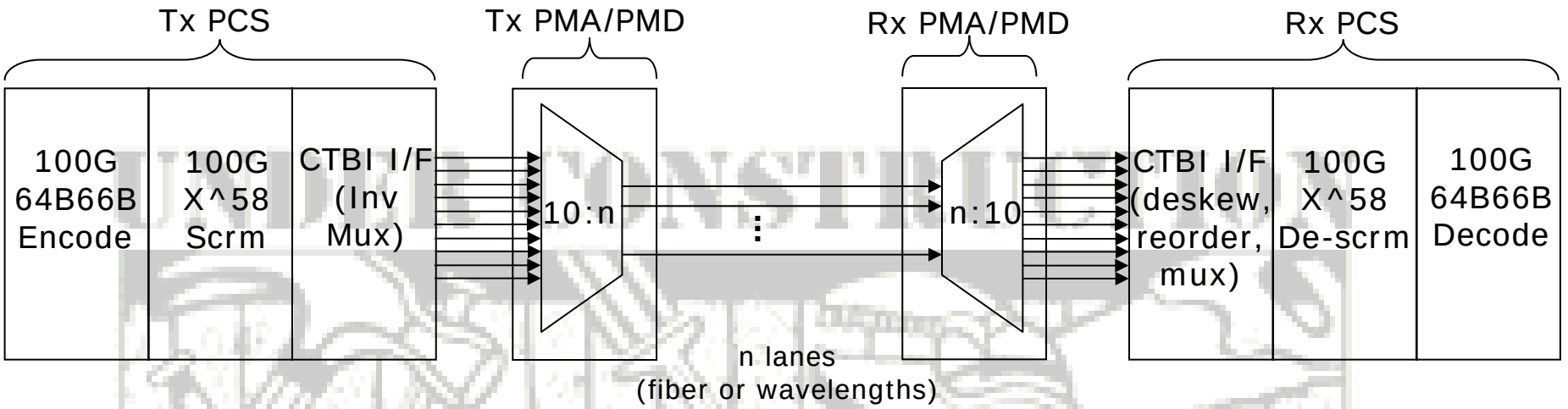
Overview – 100G MAC & PHY



Generalized LAN CSMA/CD Layers

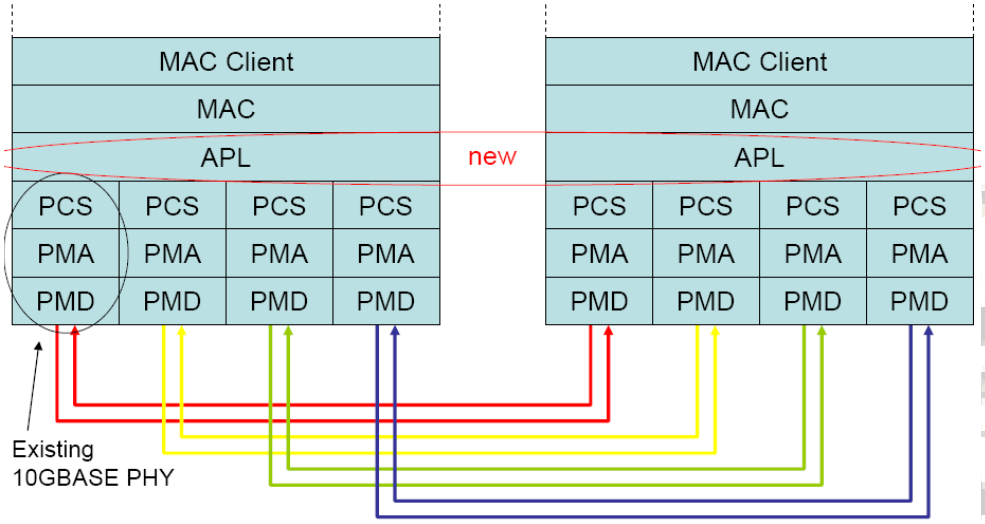
- Consistent with previous Ethernet rates, extension to 40,000 & 100,000 Mb/s data rates
  - Frame format; Services; Management attributes
- MAC
  - Proposing no changes to the MAC operation
  - Technical feasibility material reviewed:
    - CRC checker, general MAC functions, gear-boxes, host interfaces – roll-ups at .13um & 90nm data points
    - No issues or concerns identified
- PCS
  - Specified PCS(s) need to accommodate:
    - 40G backplane PHY; 40 & 100G copper cable PHYs; 40 & 100G MMF PHYs and 100G SMF PHYs
  - Commonality, leveraging existing 10G technology, and working with the specified multi-channel/cable/fiber/wavelength PMDs will be examined
  - Two example approaches reviewed (next slides)

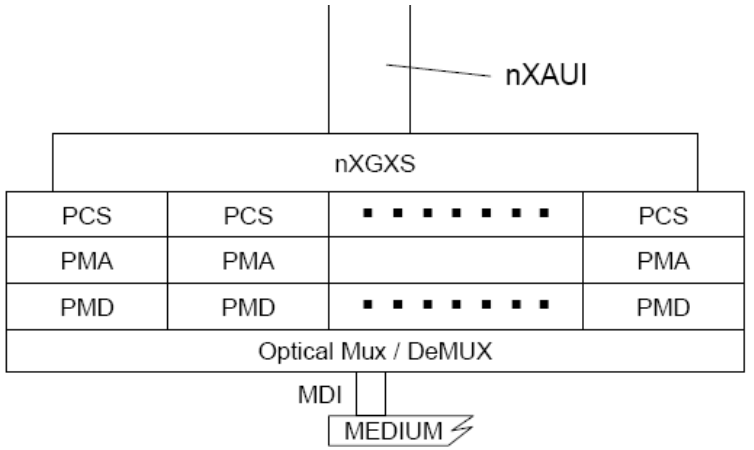
CTBI Technical Overview



- Standard 64B/66B PCS (running up to 10x faster)
- 100G Example: 10 Lane Electrical PCS to PMA/PMD Interface (CTBI)
  - 64B/66B aggregate is inverse mux'ed to Virtual Lanes
  - A unique identifier is added to each Virtual Lane on a periodic basis
  - Virtual lanes are bit mapped to/from the 10 CTBI electrical lanes
  - Virtual lane alignment and skew compensation is done in Rx PCS only
- PMA maps 10 lane CTBI to n lane PMD
  - PMA is simple bit level muxing and demuxing
  - no realignment in PMA (for either electrical or optical skew)
- PCS and Virtual Lane overhead is very low, and independent of frame size

APL Technical Overview





Multi-wavelength PHY Example (nxXAUI)

- The PME aggregation concept from 802.3ah can be used with existing 10GBASE PHYs – Aggregation at the Physical Layer (APL)
- Variety of fragment format considerations discussed (header, size, fragment CRC…)
- APL:
  - Assumes equal speed links, point-to-point, & full duplex links
  - Resilient and scalable
  - Ensures ordered delivery and detects lost or corrupted fragments
  - Minimal added latency
  - Fits well with multi-port (quad/octal) PHYs
- An APL control protocol would be specified

Physical Layer Specifications to be Defined

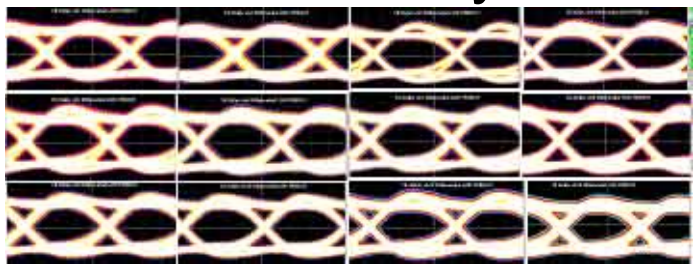
	40G	100G
At least 1m backplane	√	
At least 10m cu cable	√	√
At least 100m OM3 MMF	√	√
At least 10km SMF		√
At least 40km SMF		√

Draft HSSG Tutorial
MMF Section

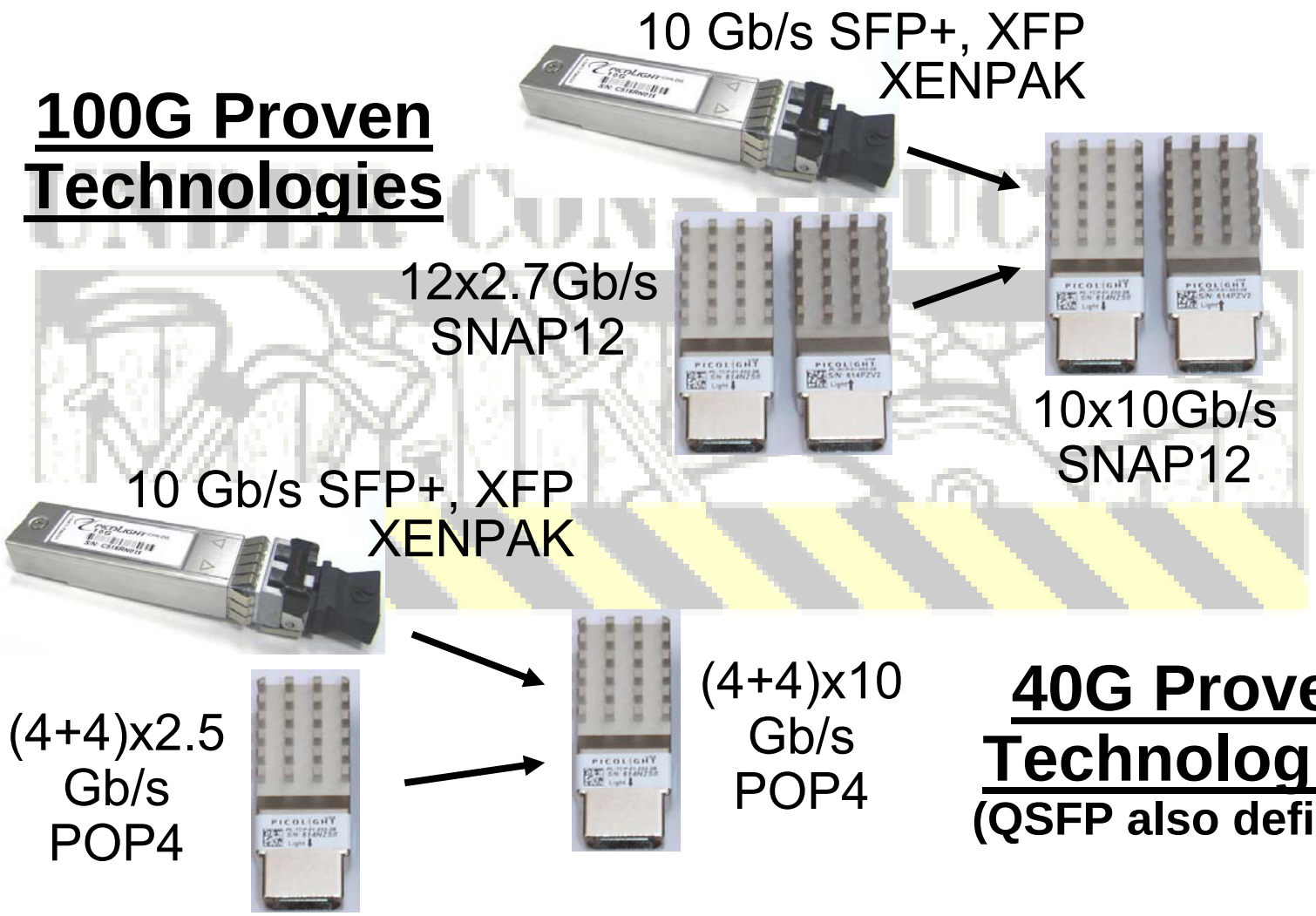
Jack Jewell – JDSU

MMF PMD Review (40G and 100G)

- Objective: At least 100m on OM3 MMF (40G and 100G)
- Platform: Parallel fibers; ~10Gbps/ch; 850nm VCSEL arrays
  - Combines existing 10GbE serial and 12x2.7G (or (4+4)x2.5G) parallel product technologies - both are technically sound and economical
  - Considering 10x10.3Gbaud (64/66), 12x10Gbaud (8B/10B), others
  - Considering WDM (e.g. 2 λ's) to reduce fiber count/cost
- Advantages:
  - Low power for ~100m reach - initially ~3W (~1.5W) for 100G (40G)
  - Small footprint for high-density interconnects
  - Low cost
- Early Demo: 12x10GbE; >300m by IBM/Picolight; OFC 2003



100G and 40G Form Factors

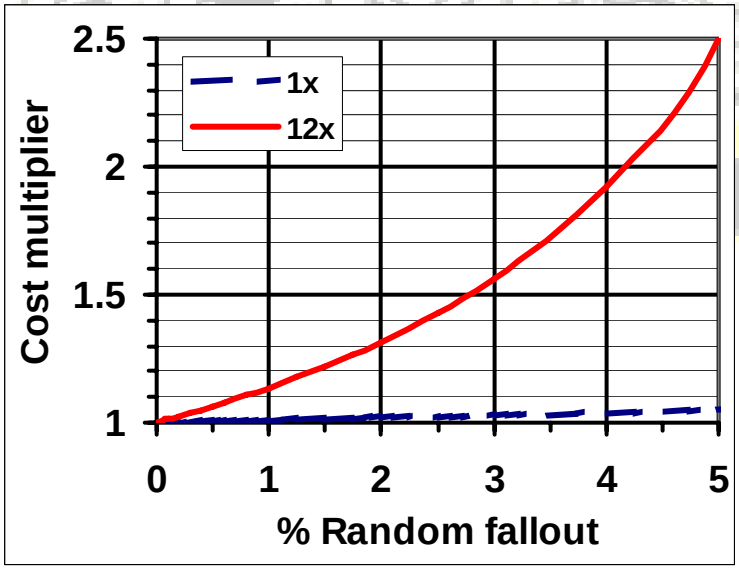


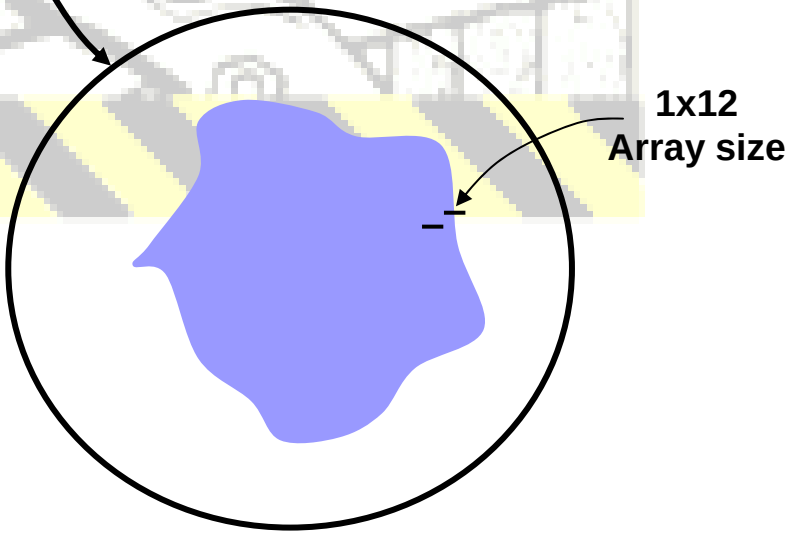
Economic Feasibility - VCSEL Yield

10x10G VCSEL yield is necessary for cost feasibility

- **VCSEL array is only a small portion of a 10x10G overall cost (same applies to 1x10G and 12x2.7G)**
- **Random microscale fallout:**
  - **12x a small number is still a small number - not significant factor**
- **Areal-dependent performance fallout:**
  - **Affects 1x and 12x similarly (slight penalty for 12x, less for 4x)**

1x4 yield slightly higher than 1x12





Array yield will not be a significant cost factor

Technical Feasibility - VCSEL Reliability

Wearout Time (Depends on uniformity)

- Perfect uniformity → 12x same as 1x
- Exp. reports: 12x wearout time ~1/2 as for 1x

Random Failures

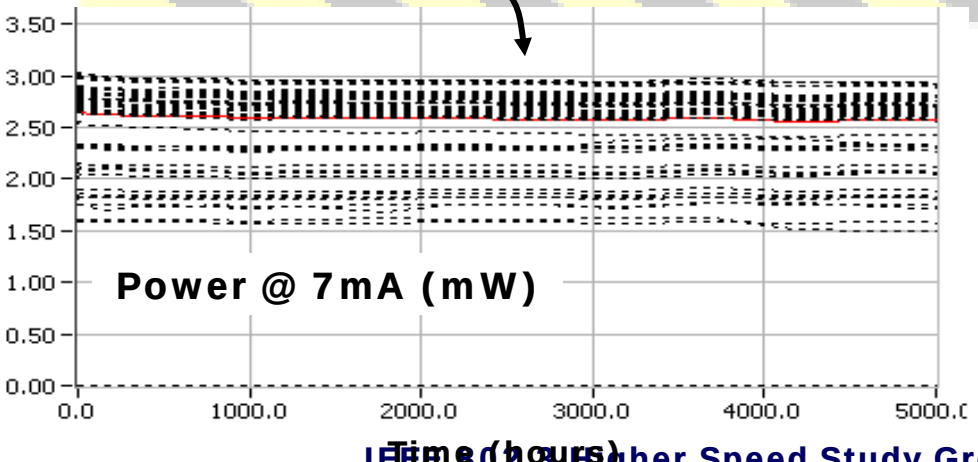
- 12x array failure rate nearly 12 times the 1x failure rate
- Virtually all "random" failures are eliminated through burn-in

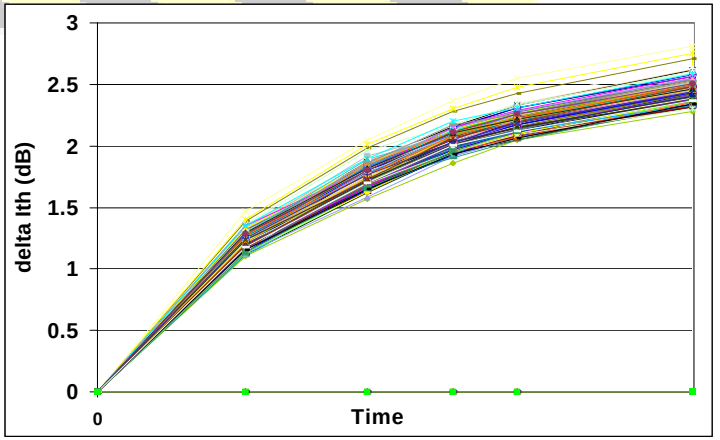
ESD-Related Failures

- Above-threshold ESD events damage 1x and 12x about equally

Non-Hermetic Packaging

- 12x3G VCSELs robust to harsh environments





MMF Data, Conclusions

Burn-in Fallout

- 12x3G only slightly higher than 1x3G
- 1x10G only slightly higher than 1x3G
- ∴ 10x10G fallout expected to be only slightly higher than 12x3G

Field Data

- For 1x10G and 12x3G products (10^4-10^6 each), JDSU has not experienced any failures due to VCSEL manufacture/technology

Expectation

- 10G VCSEL technology will mature to the level that 3G is today

Conclusions

- The MMF PMDs will be there for 100Gig and 40Gig Ethernet
- The MMF PMDs will be cost effective
- The MMF PMDs will be reliable

Draft HSSG Tutorial
SMF Section

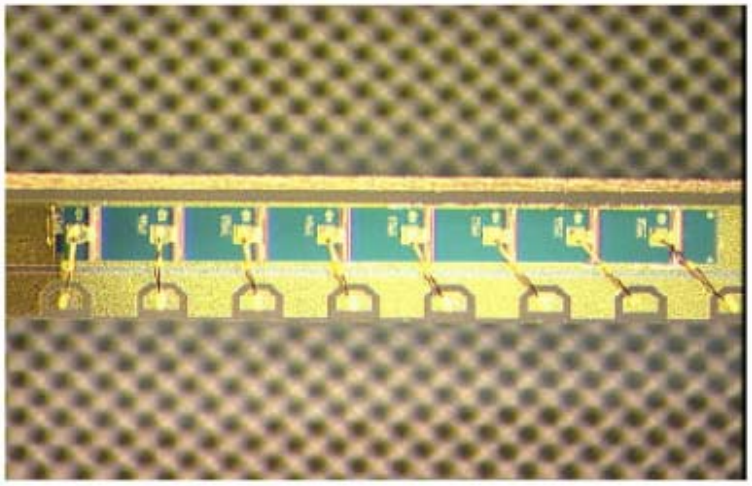
Chris Cole – Finisar

SMF PMD Technical Alternatives

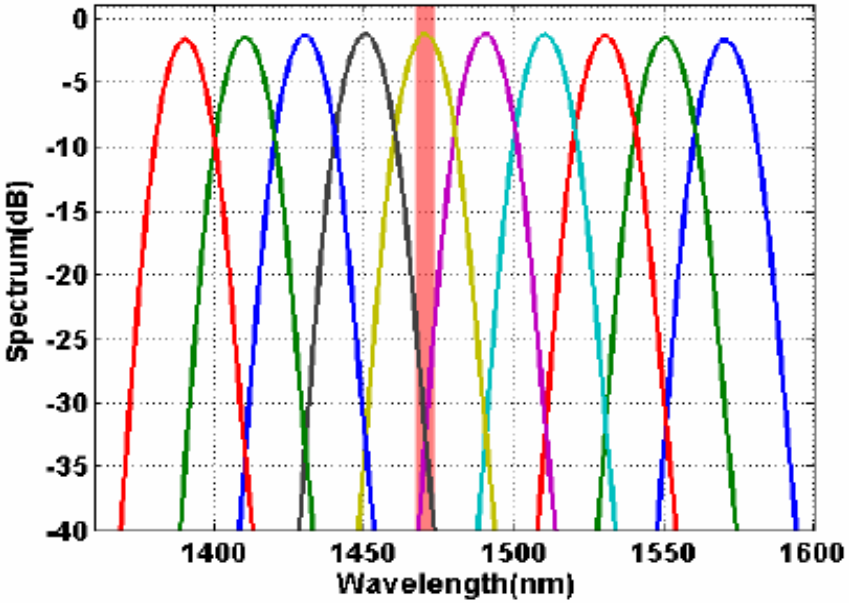
SMF Technology	10km 1310nm	40km 1310nm	10km 1550nm	40km 1550nm
10x10G DML		**OA**	clairardin_01_0107 hartman_01_0107	**OA** schrans_01_0107 tsumura_foah_1206
10x10G ML		**OA**	martin_01_0307	**OA** jaeger_01_0107
5x20G / 4x25G DML	cole_01_1106 traverso_02_0307	**OA**		**OA + DC**
5x20G / 4x25G ML	cole_01_0307 jiang_01_0407 traverso_02_0307	**OA** cole_01_0507 jiang_01_0507 traverso_01_0307		**OA + DC**
2x50G DQPSK ML		**OA + DC**	**+ DC**	**OA + DC** duelk_01_0107 takeda_01_0107
1x100G Serial ML		**OA + DC**	**+ DC**	**OA + DC** fisher_01_0107

- OA = Optical Amplification (or APD), DC = Dispersion Compensation
- Green shading designates alternatives under consideration by HSSG contributors.

10x10G 1550nm CWDM DML Transceiver



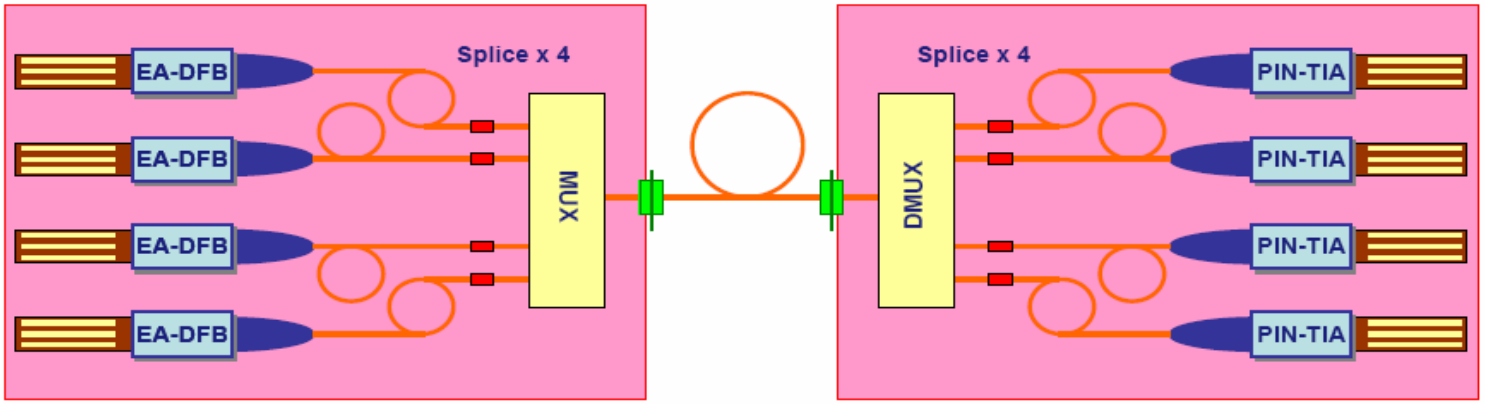
- *8 x 10Gb/s CWDM DFB array*
- *1470 – 1610nm, 20nm spacing*
- *Open, clean 10G back-to-back eyes*

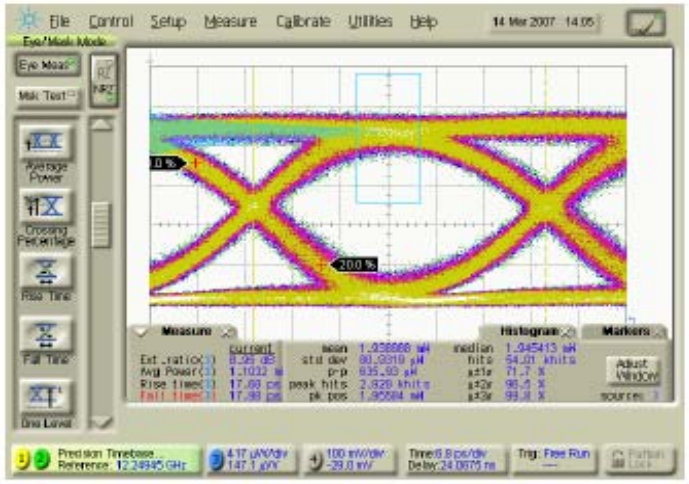


TX Mux designed with broad Gaussian 6.4nm Passband allowing 4nm for DFB registration.
5pm/°C $\Delta\lambda/\Delta T$ difference between Si and InP over Temp Range provides for 20°C for temperature differential across chip.
High thermal conductivity of Si provides good thermal management

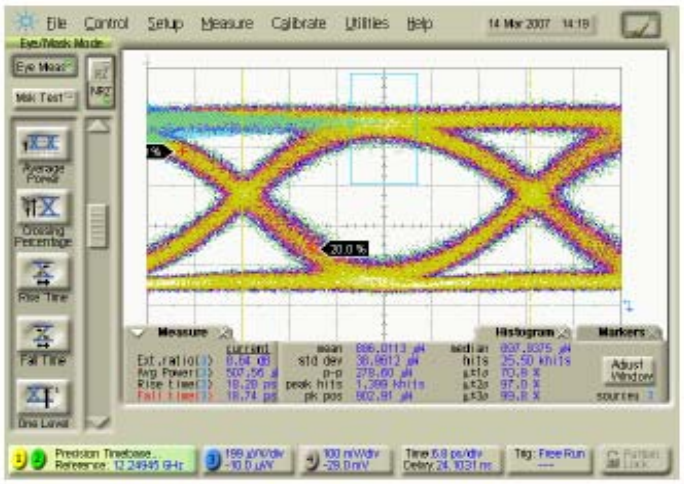
- 10G 1550nm DFB array component discussion; clairardin_01_0107
- CyOptics (DFB array,) Kotura (Mux)

4km/10km 4x25G 1310nm EML Transceiver





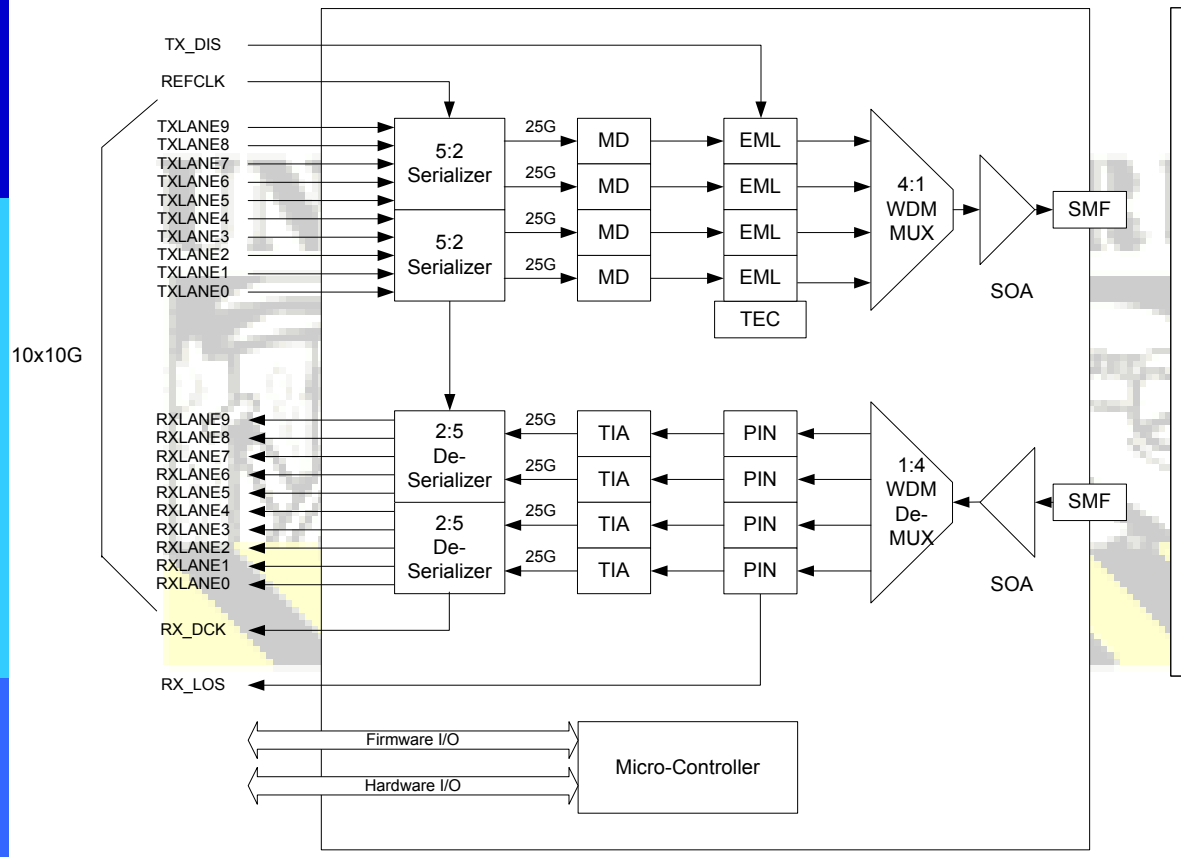
1m

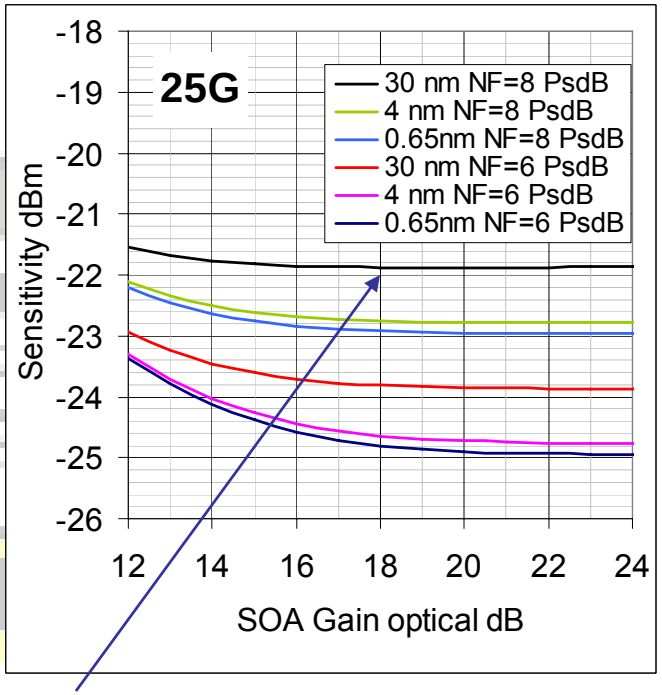


10km

- OpNext 4x25G EML Transceiver discussion; traverso_02_0307
- JDSU 25G EML components discussion; jiang_01_0407

40km 4x25G 1310nm EML OA Transceiver

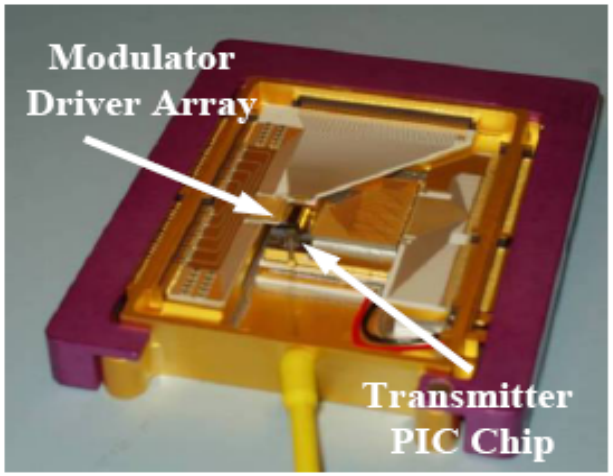




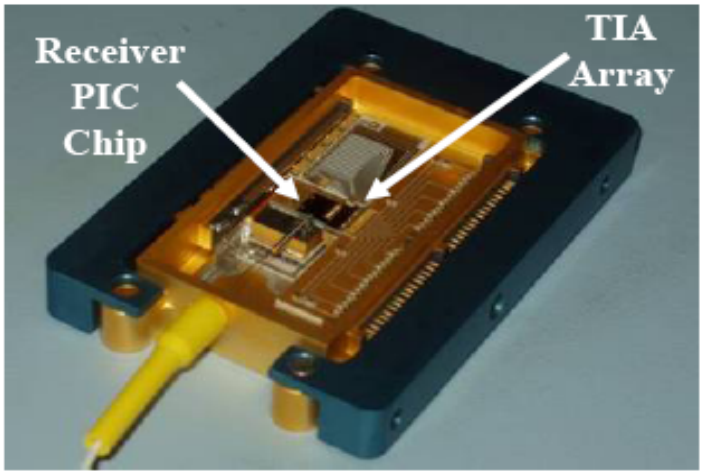
Required -22dBm Receiver Sensitivity is achievable with SOA technology.

- Finisar 1310nm Optical Amplification discussion; cole_01_0507

10x10G 1550nm DWDM EML Transceiver

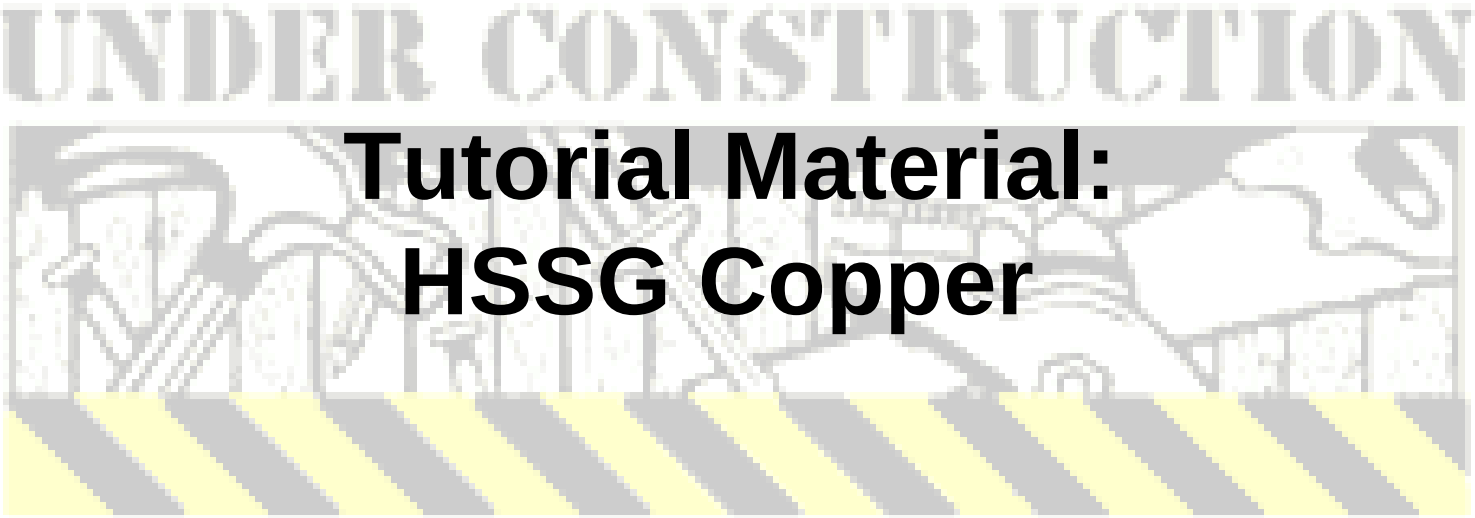


100Gb/s Tx Module



100 Gb/s Rx Module

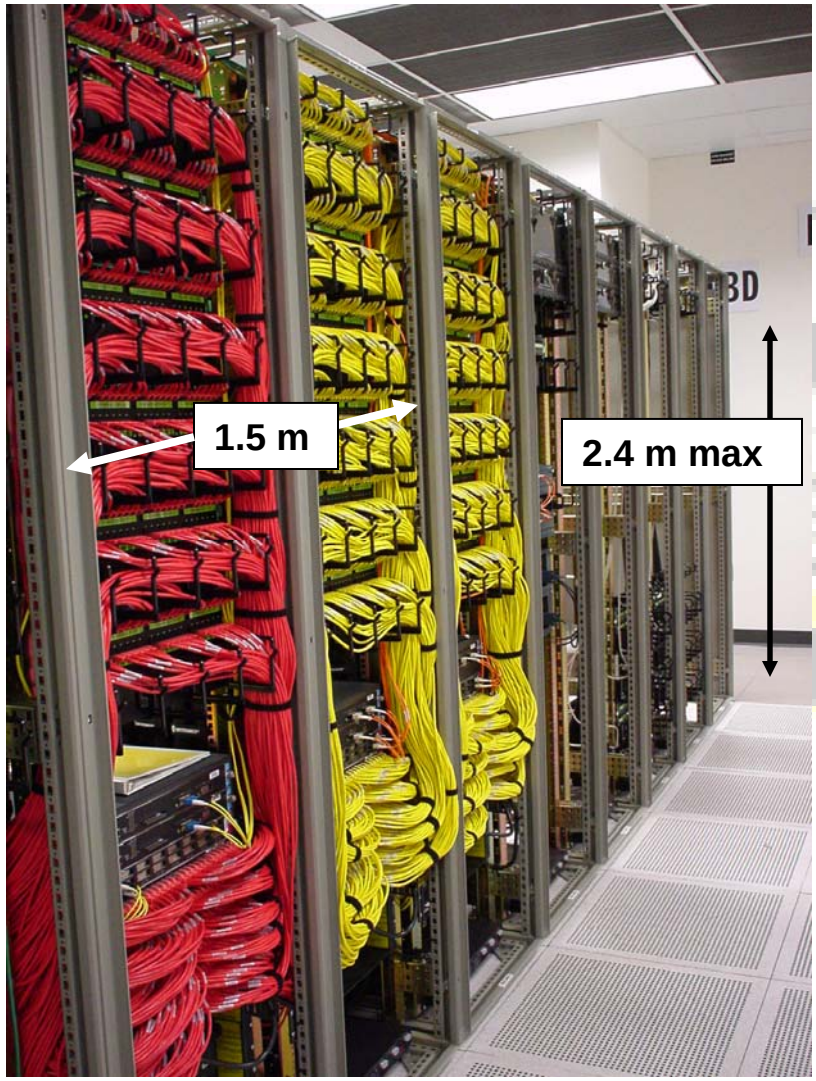
- Infinera 100Gbps 1550nm PIC components; jaeger_01_0107.
- Example of future cost reduction approach for high volume applications.



Tutorial Material: HSSG Copper

Chris Di Minico

High Speed Copper Interconnect: Applications



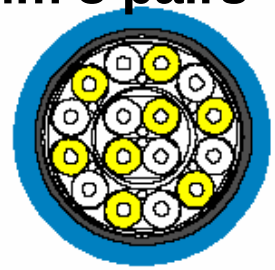
Intra/Inter rack/cabinet applications and High Performance Computing Interconnect

TIA-942 - Cabinet and rack height

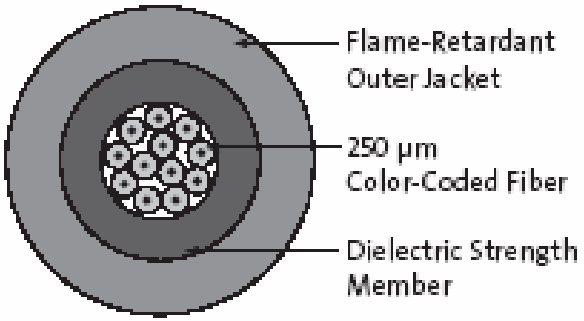
- **The maximum rack and cabinet height shall be 2.4 m (8 ft).**
- **Preferably no taller than 2.1 m (7 ft) for easier access to the equipment or connecting hardware installed at the top.**

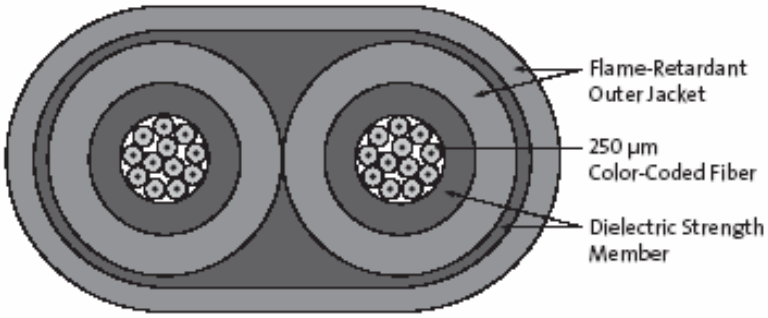
Twinaxial copper cable assembly

- 100 ohm 8 pairs – 16 conductors



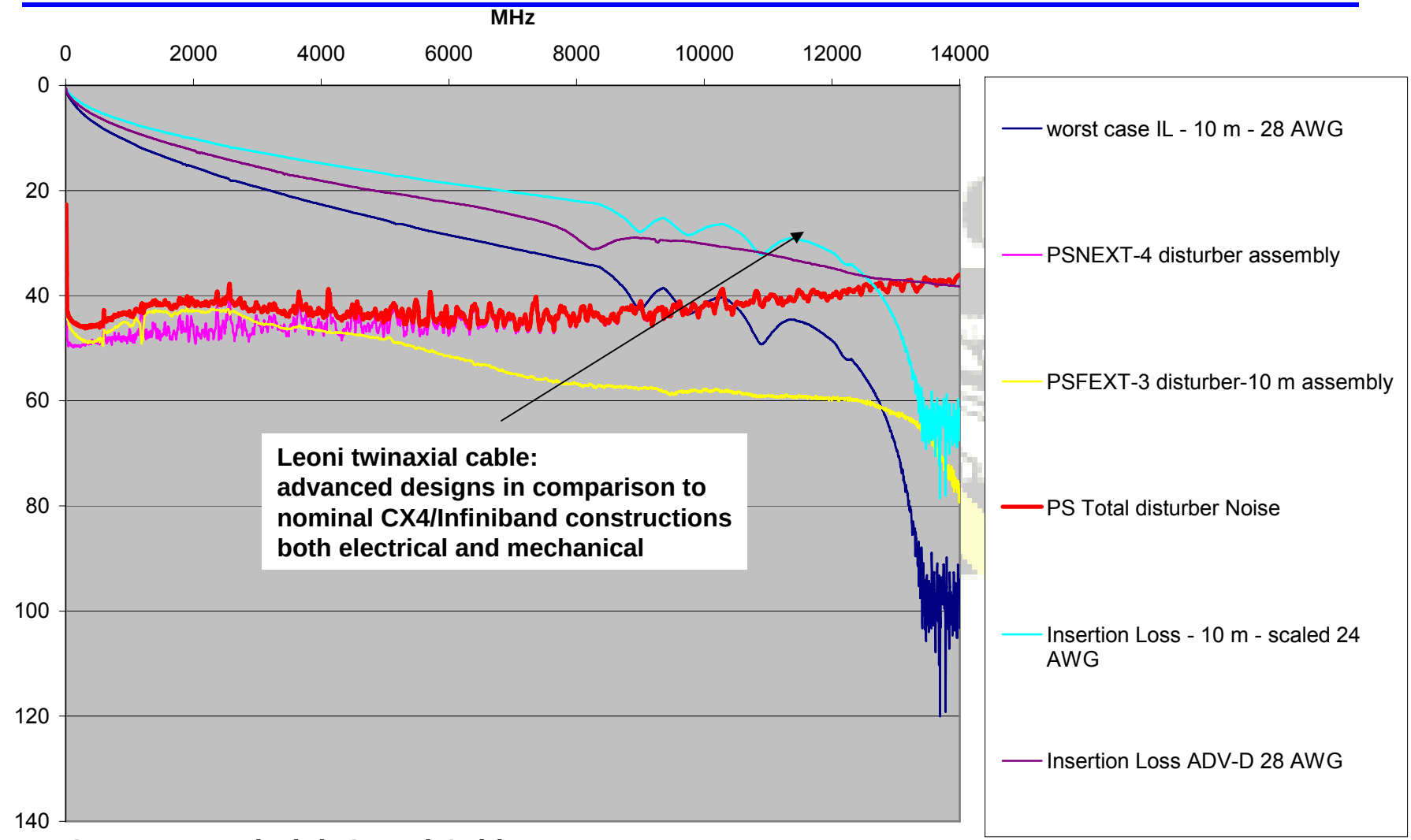
- 28 AWG - 5.6 mm-7.2 mm (0.220 in – 0.281 in) – Leoni twinaxial designs
- 250 µm - 12-fibers groups for use with multifiber connectors





- 12-Fiber OFNP Fiber Optic Cable 4.4 mm (0.17 in)
- 24-Fiber OFNP Fiber Optic Cable 8.3 mm (0.33 in)

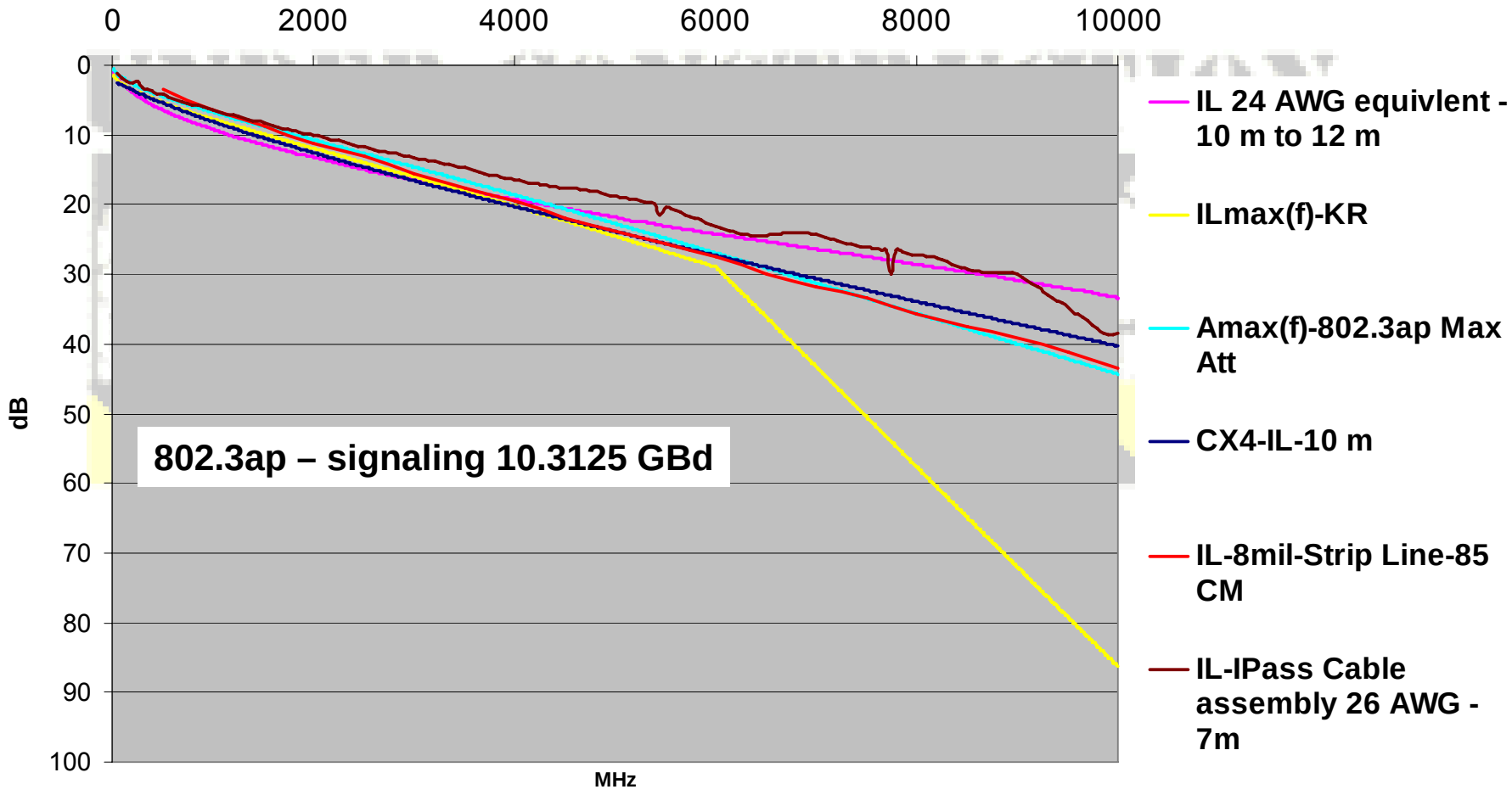
Cable assemblies - 28 AWG/24 AWG/28 AWG ADV-D 10 meter



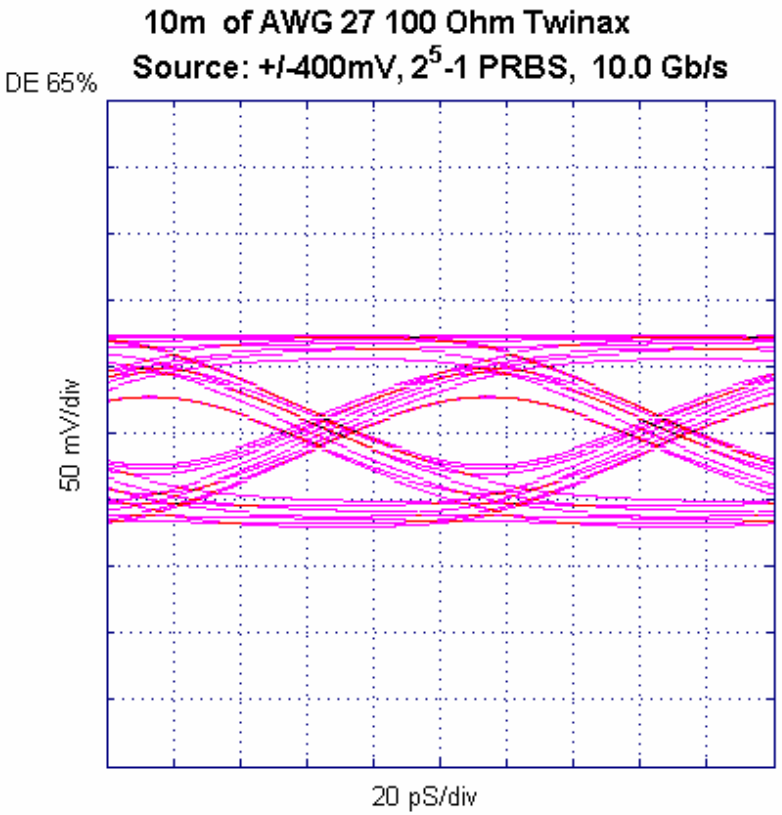
Source: Leoni High Speed Cables

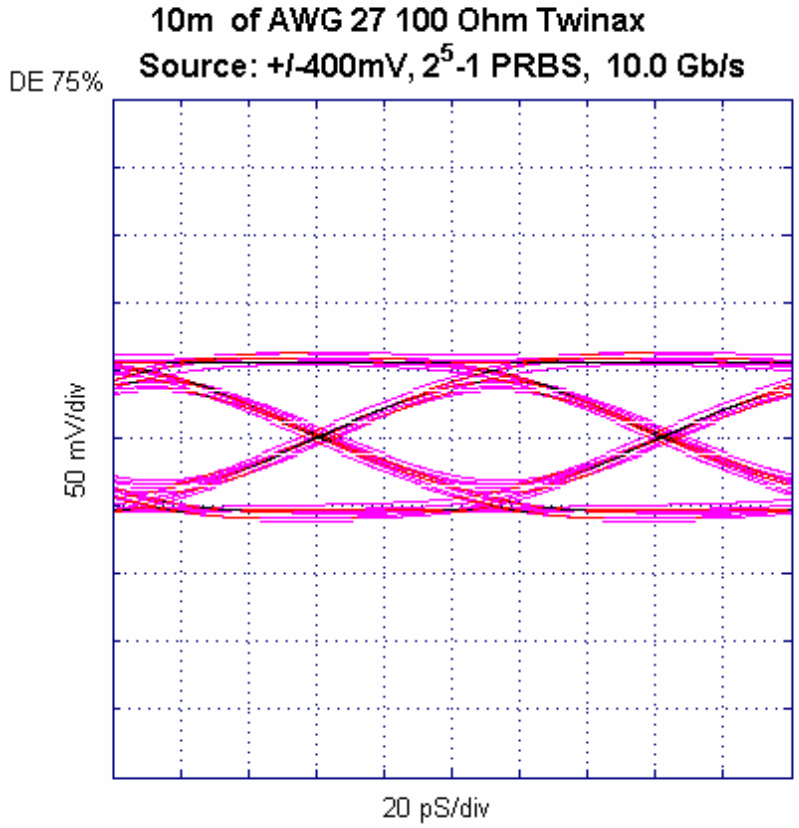
802.3ap - Ethernet Operation over Electrical Backplanes

802.3 ap Insertion Loss vs twinaxial cable assembly IL



Eye Patterns for 24 AWG cable -10 to 12 meters





Various levels de-emphasis without equalization

Source: Herb Van Deusen, W.L. Gore

Lane Rate, Signaling rate, channel bandwidth

10 m cable + connectors @ 6 dB Margin

maximum achievable lane rate for each coding gain

Maximum Lane rate	Maximum signaling rate	Info bits/baud/ dim	Modulated Info bits/baud/dim (2)	Channel bandwidth (1)	Copper Gauge	Code gain	Length	Modulation
Mb/s	Mbaud			MHz	AWG	dB	meters	
10889.28	8180.00	1.33	1.33	4090.00	28	0	10	PAM-3
13984.52	9476.64	1.38	1.48	4738.32	28	2	10	PAM-4
17555.24	10097.78	1.55	1.74	5048.89	28	4	10	PAM-4
24950.75	17290.00	1.44	1.44	8645.00	24	0	10	PAM-4
30727.91	16261.68	1.77	1.89	8130.84	24	2	10	PAM-4
36785.11	16888.89	1.94	2.18	8444.44	24	4	10	PAM-5
17212.31	13210.00	1.30	1.30	6605.00	28 ADVD	0	10	PAM-3
22044.24	13626.17	1.51	1.62	6813.08	28 ADVD	2	10	PAM-4
27147.01	13742.22	1.76	1.98	6871.11	28 ADVD	4	10	PAM-4

(1)Channel bandwidth=.5*(Maximum signaling rate) (2)Infor/bits/baud/dim adjusted for coding gain

maximum achievable lane rate for each coding gain that yields 1 bit/baud

Maximum Lane rate	Maximum signaling rate	Info bits/baud/ dim	Modulated Info bits/baud/dim	Channel bandwidth	Copper Gauge	Code gain	Length	Signaling
Mb/s	Mbaud			MHz	AWG	dB	meters	
10439.00	10439.00	1.00	1.00	5219.50	28	0	10	PAM-2/NRZ
12630.13	13580.00	0.93	1.00	6790.00	28	2	10	PAM-2/NRZ
14857.53	16690.00	0.89	1.00	8345.00	28	4	10	PAM-2/NRZ
21639.98	21639.98	1.00	1.00	10819.99	24	0	10	PAM-2/NRZ
23439.51	25170.00	0.93	1.00	12585.00	24	2	10	PAM-2/NRZ
24885.04	27950.00	0.89	1.00	13975.00	24	4	10	PAM-2/NRZ
16224.36	16224.36	1.00	1.00	8112.18	28 ADVD	0	10	PAM-2/NRZ
18748.43	20159.60	0.93	1.00	10079.80	28 ADVD	2	10	PAM-2/NRZ
21194.44	23813.98	0.89	1.00	11906.99	28 ADVD	4	10	PAM-2/NRZ

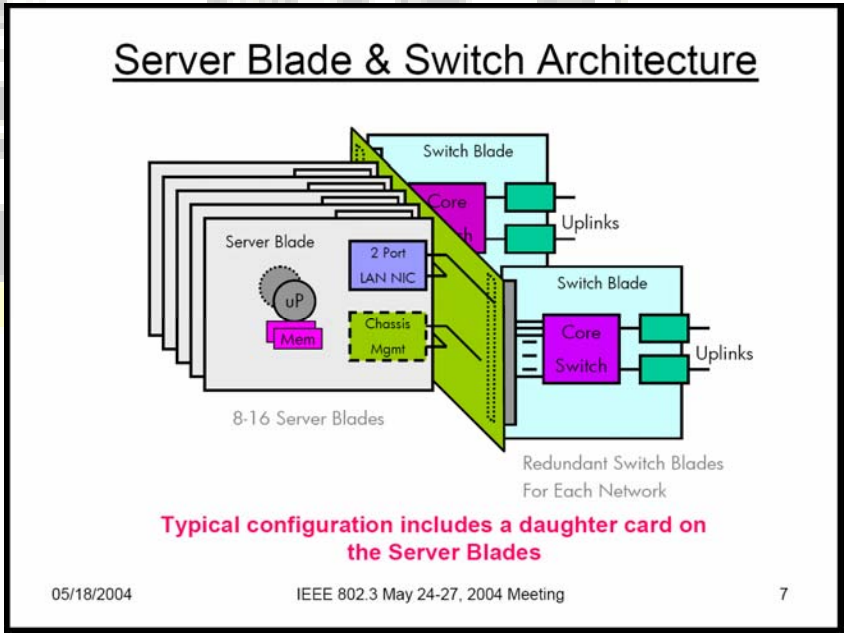
Source: George Zimmerman, Solarflare Communications, Chris DiMinico, MC Communications

Conclusions

- Technical feasibility, economic feasibility, and market potential for a Higher Speed copper interconnect demonstrated.
- Up to 10 meter reach consistent with intra/inter rack application and HPC cluster distances.
- High speed study group copper interconnect objective of at least 10 meters to address intra/inter rack applications and high performance computing (HPC) interconnects.

40 Gigabit Backplane Ethernet

- IEEE 802.3ap™-2007 defines 1 and 10 Gigabit Ethernet operation over a modular platform backplane (1 m objective)
  - 1000BASE-KX (Gigabit Ethernet)
  - 10GBASE-KX4 (10 Gigabit Ethernet, 4 x 3.125 GBd)
  - 10GBASE-KR (serial 10 Gigabit Ethernet)
- Blade servers: 2nd generation backplane
  - Based on 10GBASE-KX4 architecture...
  - ...but satisfy 10GBASE-KR channel requirements
  - 40 Gb/s capability inherent
- "40GBASE-KR4" leveraged from 10GBASE-KR standard



Source: Koenen, "Channel Model Requirements for Ethernet Backplanes in Blade Servers", May 2004.

Summary

- Bandwidth requirements are growing for all applcations in the Ethernet EcoSystem
- The future bandwidth needs of Networking and Computing / Server are diverging
- The project targets the next generation of Ethernet with two rates