100G-Optics Next Gen Possible PMD Set

Ali Ghiasi

IEEE 802.3 100GNGOTX Study Group

Sept 15 2011

Chicago

Overview

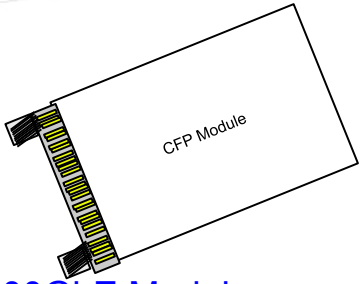
- Historical Evolution of 10GbE
- Module roadmap
- I/O Trend
- What is driving force to 4x25G unretimed “cPPI”
- Mega data centers
- 100Gbase-SR4
- 100Gbase-FR4

100GbE Current PMD Set and What is Missing?

- **Optical interface parameters for 100GBase-LR4 (10 km) and 100GBase-ER4 (40 km) defined in the 802.3ba**
  - In addition 802.3ba defined 10x10G 850 nm MMF and Cu cable which will phase out with introduction of native 25G I/O
- **100GCU project will define 4x25G Cu cabling**
  - Target reach is 5 m
- **100GNGOPTX need to study and possibly define**
  - CAUI-4 chip to chip retimed interfaced (OIF VSR as starting point but possibly with reach of OIF-SR) (please see ghiasi_02_0911)– low hanging fruit
  - cPPI-4 chip to module unretimed interface (please see ghiasi_03_0911) – med hanging fruit
  - 4x25G 850 nm MMF parallel optics link – low hanging fruit
  - 4x25G duplex SMF targeted application mega data centers– major development

Evolution of 10 Gig Ethernet Module Form Factor

- **10 GbE started with XSBI (16 lanes) and over last 8 years migrated to single lane unetimed SFI!**
  - CFP is 100 GbE equivalent Gen 2
  - CFP2 is 100 GbE equivalent Gen 2
  - CFP4/QSFP2 100 GbE equivalent Gen 3



1st Gen 40/100GbE Module

Gen 0
300 Pins
XSBI-16 Lanes

Gen 1

Gen 1+
X2
XAUI-4 Lanes

Gen 2

Gen 3

Xenpak
XAUI-4 Lanes

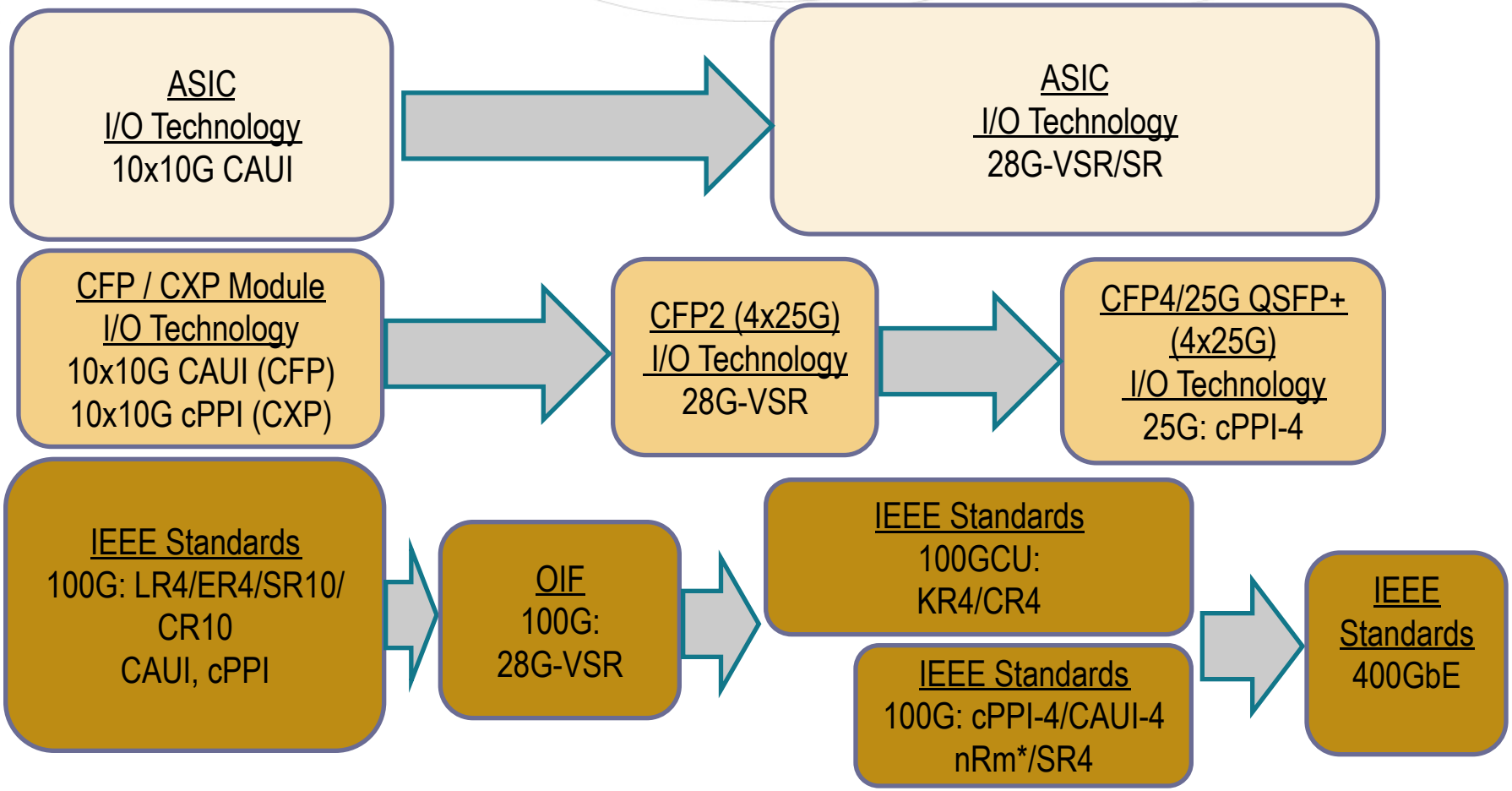
XFP
XFI-1 lane

SFP+
SFI-1 Lane

Gearbox/Mux/De-mux → **Retimed** → **Unretimed**

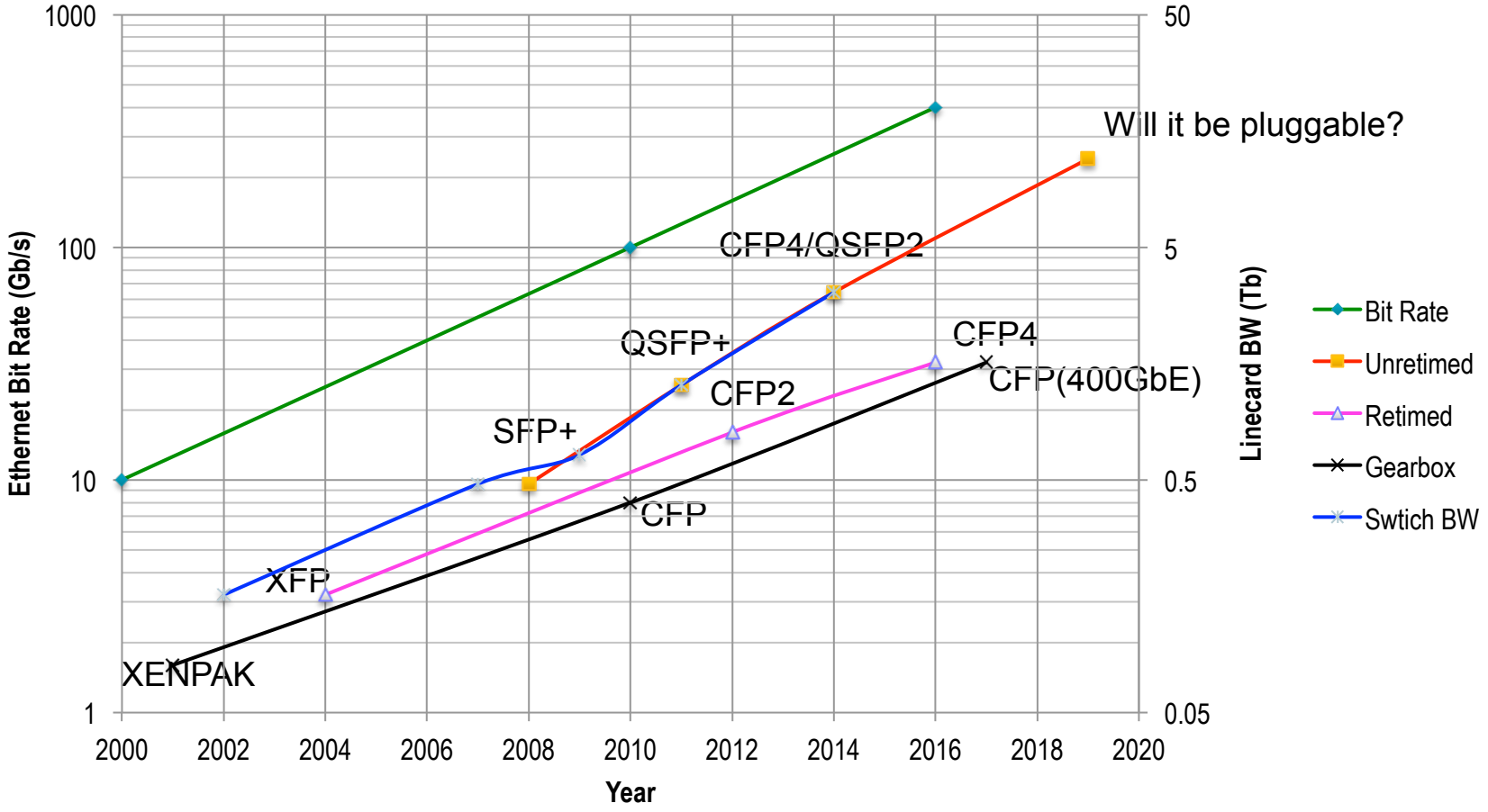
2001 2002 2003 2004 2005 2006 2007 2008 2009

100GbE I/O Trends



Ethernet Bit Rate and Projected Line Card I/O Density

- In 2008 with availability of switch/ASICs the SFP+ enabled 48 ports designs
- Unretimed modules based on cPPI-4 are needed to support projected switch/ ASIC BW growth



Largest US Data Centers*

* Dana Hull, San Jose Mercury News, April 22 2011

** Status as of April 2011

*** Longest Run Estimated to be = 2*√(area)=603 m

Tulsa Times May 3 2007 states 1.4M Sq-ft plant assume data center is 70% of plant

Various media reported likely size to be ~500K Sq-ft

Median size of data centers in US 190000 sq feet

Typical run assuming 190000 sq feet area =√(area)=132.9 m

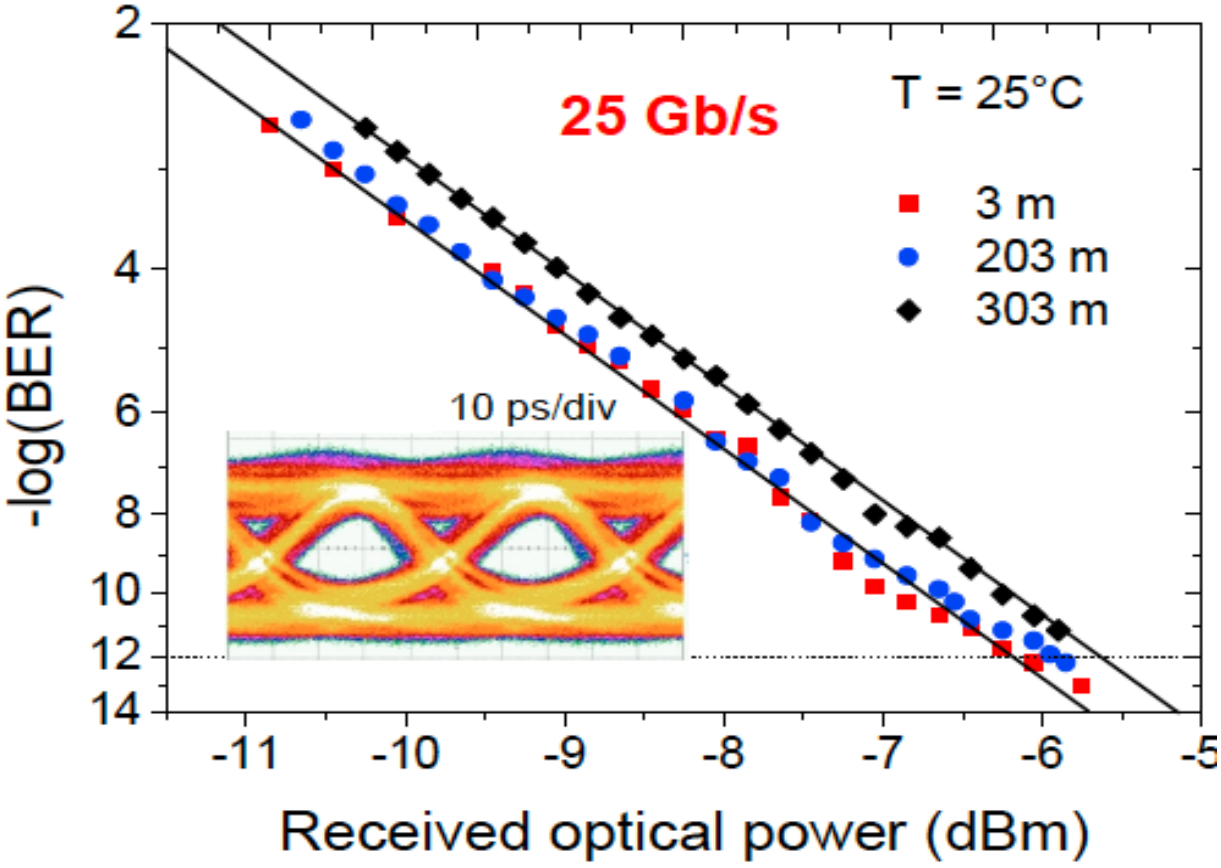
Data Center Location	Status	Square footage	Power (MW)	Longest Run(m)*
Amazon				
Boardman, OR	Not operational	100000	5	192.8
Mcnary, OR	Under Construction	120000	6	211.2
Manasas, VA	Operational	110000	5	202.2
Ashburn, VA	Operational	180000	9	258.6
Stering, VA	Operational	125000	6	215.5
Apple				
Maden, NC	Nearly Completed	500000	100	431.1
Newark	Operational	100000	15	192.8
Facebook				
Princeville, OR	Nearly Completed	307000	40	337.8
Forest City, OR	Under Construction	300000	40	333.9
San Jose, CA	Lease	25000	5	96.4
Santa Clara, CA	Lease	86000	15	178.8
Santa Clara, CA	Lease	50000	8	136.3
Ashburn, VA	Lease	49000	8	134.9
Ashburn, VA	Lease	45000	8	129.3
Google				
Berkely County, SC	Under Construction	200000	76	272.6
Council Bluff, IA	Under Construction	200000	76	272.6
Dalles, OR	Operational	200000	70	272.6
Lenoir, NC	Phase 2	337000	76	353.9
Mayers County, OK #	Likely Operational	980000	100.8	603.5
HP				
Atlanta, GA	Operational	200000	20	272.6
Atlanta, GA	Operational	200000	20	272.6
Austin, TX	Operational	100000	20	192.8
Colorado Spring, CO	Operational	250000	10	304.8
Tusla, OK	Operational	250000	20	304.8
IBM				
Boulder, CO	Operational	300000	60	333.9
RTP, NC	Operational	100000	30	192.8
Microsoft				
Boydton, VA ##	Likely Operational	500000	NA	431.1
Chicago, IL	Operational	700000	60	510.0
Quincy, WA	Operational	500000	27	431.1
San Antonio, TX	Operational	477000	27	421.0
Des Moines, IA	Under Construction	NA	NA	
Twitter				
Salt Lake City, UT	Lease	15000	2	74.7
Sacramento, CA	Lease	NA	2	
Yahoo				
Ashburn, VA	Operational	112000	10	204.0
Lockport, NY	Operational	NA	18	
Omaha, NB	Operational	180000	20	258.6
Quincy, WA	Operational	180000	26	258.6
Quincy, WA	Under Construction	NA	72	

Some Thought on 100GBase-SR4

- **Natural extension of the 40GBase-SR4 to 4x25G operation and a key objective of this project**
- **Starting reach objective should be the same as reach defined in 802.3ba**
  - 100 m on OM3
  - 150 m on OM4
- **What are the key challenges for operation of VCSEL link at 25G**
  - VCSEL rise/fall time 15-20 ps where the device can't be made much smaller
  - Operation at 80 C°
  - Photodiode capacitance 100-150 ff where the device can't be made much smaller and still collect the light from 50 um core
- **A portion of VCSEL/PD impairments at 25G could be equalized as previously it was shown the use of 4G VCSELs for operation at 10G**
  - Ghiasi & Dudek, "Benefits of EDC and linear receivers for short reach 40/100GE", IEEE Communication Magazine, Feb 2008
- **FEC could be utilized to improve the power budget**
  - see http://www.ieee802.org/3/100GCU/public/sept11/bhoja_01_0911

25G VCSEL State of Art Results

- **As presented by James Lott in T11 11- 263v1**
  - Significant amount of linear distortion is visible in the eye diagram where an EDC could help



Some Thought on Med Reach Duplex SMF PMD "100GBase-nR4"

- **The reach for 100GBase-SR4 likely will not be much longer than 100 m**
  - In large data center any reach beyond 100 m one has to use 100GBase-LR4 which is overkill and does not meet the power-cost-density requirements
  - Based on public data from all major US data center it seems that 600 m reach will satisfy all of the Mega data centers in US based on assumption that max length ≤ 2*√(area)
- **Having already defined duplex SMF PMD what hurdle must be met to define a 2nd SMF PMD**
  - Cost 3x 10Gbase-LR
  - Power and size with in the envelope of 25G QSFP+
- **What are the possible candidates for 100GBase-nRm**
  - The most perceivable PMD is 4λ CWDM but can it deliver cost-power-density
  - Multi-core SMF would require new fiber and connectors
  - Single λ PAM-16 at 25GBd possibly the most promising but can we implement PAM-16 in analog
  - Dual λ PAM-4 analog implementation of PAM-4 on unimpaired channel is trivial
  - Single λ PAM-4 at 50GBd will eliminate direct mod DFB

Summary

- **The 100GNGOPTX project will define the 2nd generation set of PMDs for 100 GbE and likely the last round to define additional PMDs for the same application space**
  - Specially in the case of 100Gbase-nRm if we can't come up with compelling solution now it would be better not to define it
- **There is very clear application space for 100Gbase-SR4, CAUI-4 retimed interface, and cPPI-4 unretimed interface**
- **The combination of 100G CR4, CAUI-4, cPPI-4, and SR4 at least will enable the high density 100GbE applications in 25G QSFP+/CFP4 form factors**
- **100 GbE PMD set will be incomplete till duplex SMF PMD with reach of 500 m in 25G QSFP+/CFP4 form factor becomes reality**
- **Implication of not having high density – low cost – low power duplex SMF**
  - The market will be fragmented and multiple line card have to be designed
  - Will add more pressure on 100Gbase-SR4 to go longer!
- **Strongly encourages major data center operator to share their cable reach distribution so we can better tailor these new PMDs.**

Thank You