100GE/40GE skew budget

IEEE 802.3ba TF
Dallas November 2008

Mark Gustlin – Cisco
Pete Anslow – Nortel
Dimitrios Giannakopoulos - AMCC

Supporters

- Gary Nicholl – Cisco
- Farhad Shafai – Sarance Technologies
- Francesco Caggioni - AMCC
- Brad Booth – AMCC
- Chris Cole – Finisar
- Norbert Folkens – JDSU
- Faisal Dada - JDSU
- Thananya Baldwin – Ixia
- Jerry Pepper – Ixia
- Zhi Wong – Altera
- Mike Peng Li – Altera
- Divya Vijayaraghavan – Altera
- David Ofelt – Juniper
- Magesh Valliappan – Broadcom
- Steve Trowbridge - Alcatel-Lucent
- Anthony Torza – Xilinx
- Petar Pepeljugoski – IBM
- Andy Weitzner – Marvell

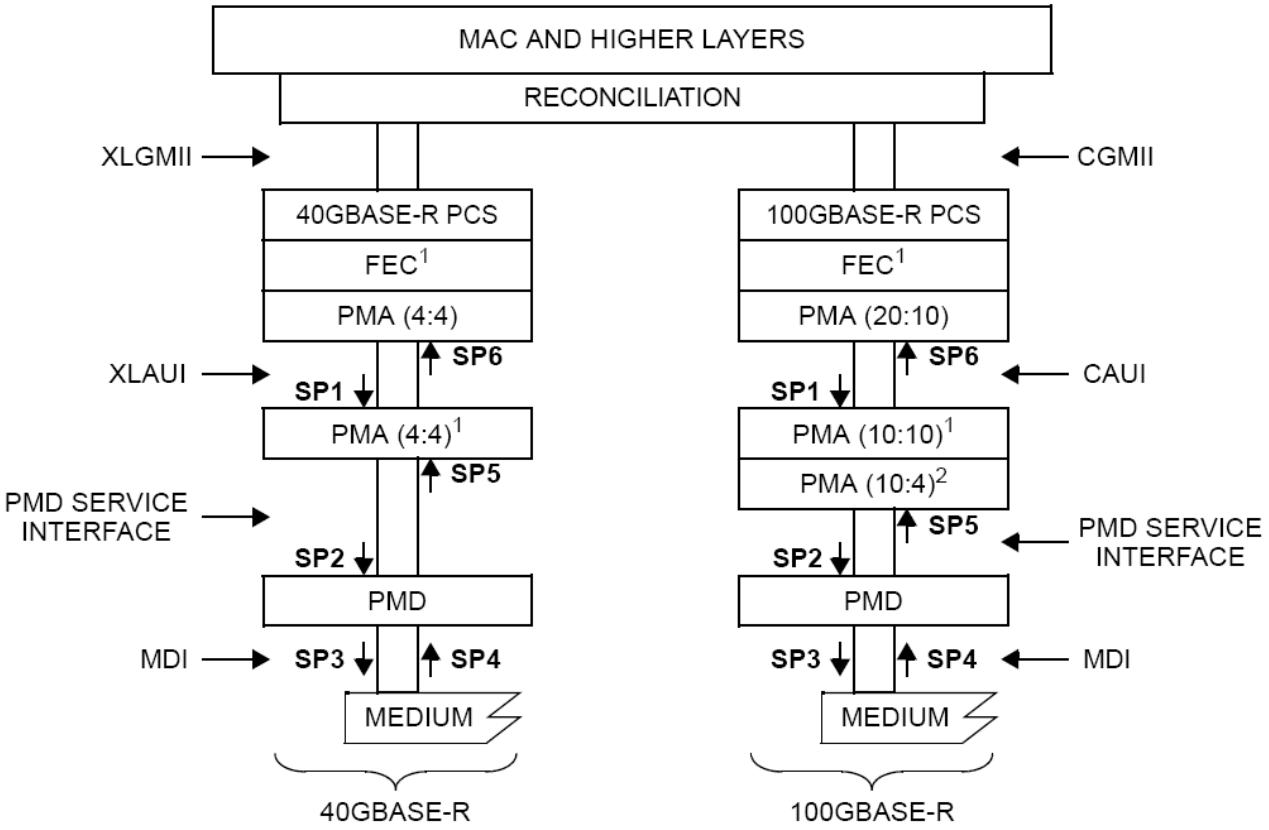
Skew Definition

- **In 40GE and 100GE, information will be transmitted in parallel links (lambdas, fibers, copper cables), typically not serially**
- **Since different paths can have different delays, skew will be introduced between them**
- **Source information needs to be reconstructed at the remote end, therefore de-skewing is needed at appropriate points**
- **Need to identify skew contributors and where we must compensate for skew**
- **Skew considered in this presentation is lane-to-lane skew, not the skew between the positive and negative parts of a differential pair**

High Level Skew View Point

- **What are the causes of maximum skew**
  - **Fixed path length differences**
    - **Copper traces**
    - **Cables**
    - **Fibers etc.**
  - **Parallel path FIFOs not synchronized**
  - **Propagation differences between media**
    - **Caused by wavelength differences**
    - **Stress in fibers, etc.**
- **What are the causes of dynamic skew (a subset of max skew)**
  - **Group delay: variable due to laser wavelength shift with temperature and wavelength drift over time**
  - **Fiber stress variation**
  - **DMD: variable due to launch & coupling variation**
  - **Electrical functions**
    - **Temperature variation causing variable gate delay**

Skew Points with integrated or no FEC

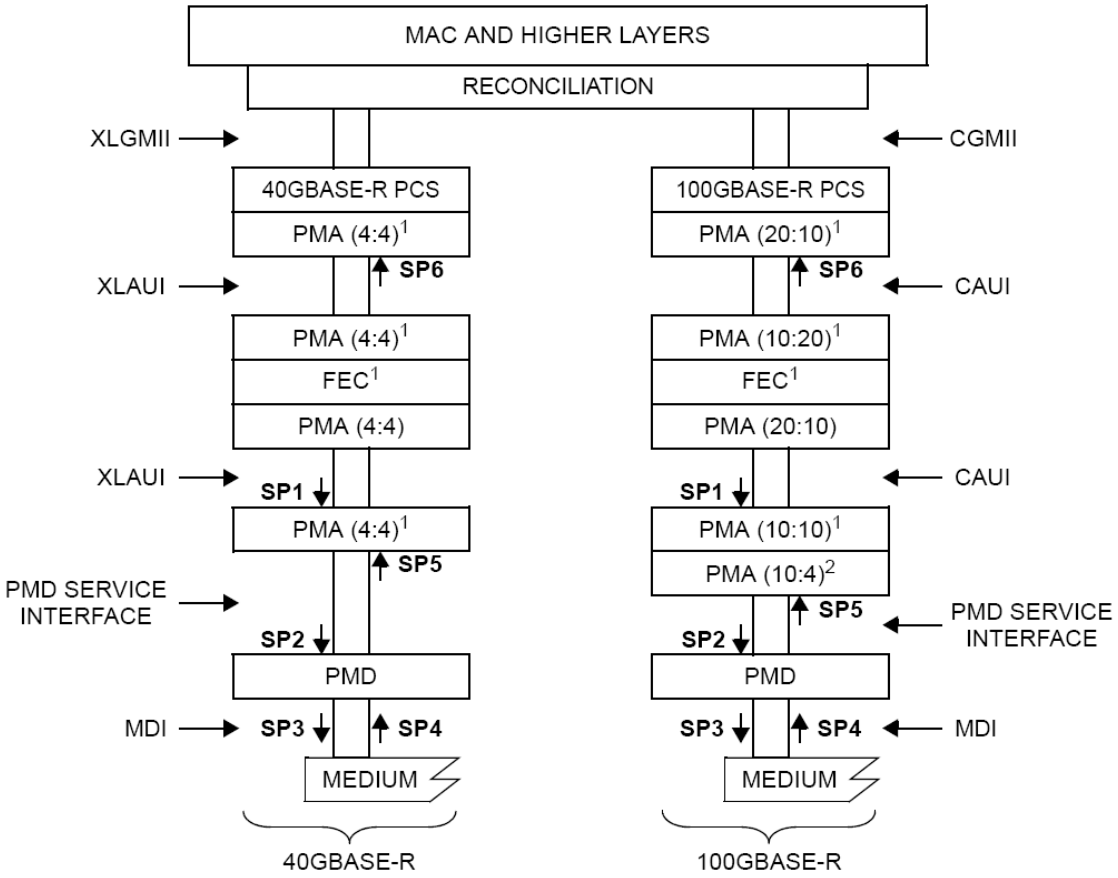


- **If XLAUI or CAUI is implemented , limits at SP1 and SP6 apply**
- **If PMD Service Interface is implemented, limits at SP2 and SP5 apply**
- **At MDI, limits at SP3 and SP4 apply**

Note 1 : optional layers

Note 2 : Conditional based on PMD type

Skew Points with separate FEC



- **XLAUI or CAUI is implemented , limits at SP1 and SP6 apply**
- **If PMD Service Interface is implemented, limits at SP2 and SP5 apply**
- **At MDI, limits at SP3 and SP4 apply**

Note 1 : optional layers

Note 2 : Conditional based on PMD type

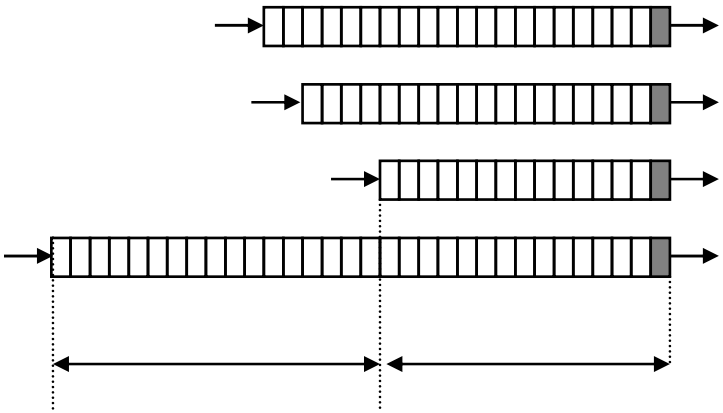
Skew budget definition in the PCS

- **PCS distributes data (with Lane Markers) into 20 (100 GE) or 4 (40 GE) virtual lanes using 66b block distribution**
- **PMAs distribute data by bit multiplexing (when needed)**
- **Skew will be presented as:**
  - **The maximum skew**
  - **The dynamic skew**
    - **Skew change over time due to environmental and/or other conditions**
    - **A subset of the maximum skew**
- **Maximum skew will be compensated for in the Rx PCS**
  - **Determines minimum FIFO size required at PCS sink**
- **Dynamic skew needs to be tolerated at each appropriate sink point, examples are:**
  - **PCS Rx sink (SP6)**
  - **Tx PMA (SP1)**
  - **Rx PMA (SP5)**

De-skew buffer

Example – 4 x 10G receiver PCS

When link is established

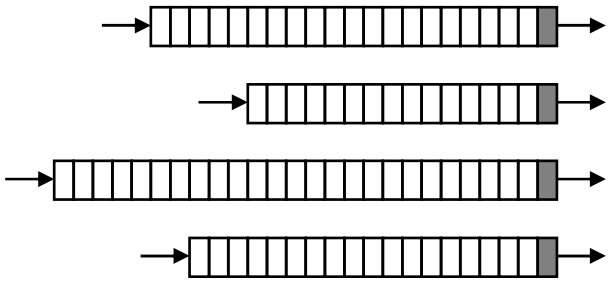


Initial skew

All buffers must contain at least as much as the max dynamic skew

Lane markers read out of each lane at the same time

Some time later



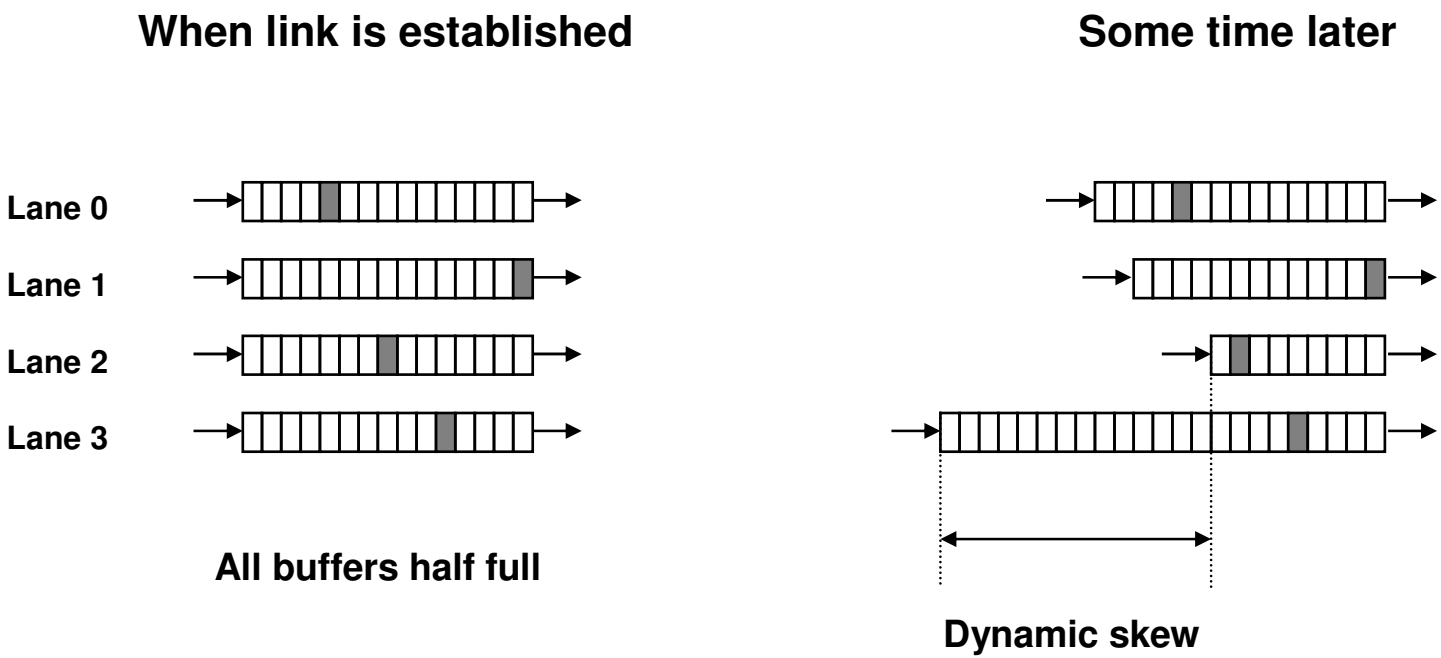
Dynamic skew has changed the fill level of the buffers

The maximum skew is the largest difference in the fill level of the buffers at any time

= first bit of the lane markers

Dynamic skew buffer (m != n)

Example – 4 x 25G receiver gearbox



Takes no account of the lane markers

= first bit of the lane markers

PCS maximum skew (TX and RX)

- **This slide analyzes the skew of the PCS layer as well as that of its attached PMA (4:4 or 20:10) and (optional) FEC, as shown in the system architecture diagram in slide 5. Related skew points are SP1, SP6**
- **Skew can be introduced due to 10G Tx SerDes FIFOs not being aligned, differences in FIFO fill levels translates into skew**
- **Another contributor can be the high speed serializer or deserializer stage in the SerDes (PMA)**
- **2 case studies: ASIC or FPGA solution**

  ASIC case: TX= 2 ns, RX= 2 ns

  FPGA case: TX = 25.5 ns, RX = 14.3 ns
- **If FEC is external, PCS budget includes the FEC contribution**

- **More detailed analysis in giannakopoulos_01_0508**

CAUI/XLAUI maximum skew (TX and RX)

- **For a chip to chip interconnect, assume a total distance of 8-12” on the host board, and there would be 1” or so on the module**

- **Propose a generous 4” of trace length difference allowance, equates to 0.88 ns per direction (TX/RX)**

PMD/PMA max skew (TX and RX)

- **This contributor relates to the PMA associated with SP2, SP5 and SP6 points in the system architecture diagram (4:4, 10:10, 10:4)**
- **In a 10:4 (4x25G, 100GE) PMA scenario, a 64-bit internal bus implementation can introduce 128 bits of skew per direction**

  6.4 ns (64 bits) due to FIFO fill difference

  6.4 ns (64 bits) due to serializer/deserializer function
- **PMA to PMD connection**

  Traces should in any case be carefully laid out

  Propose 1” (per direction), which is 0.22 ns (TX/RX)
- **Total PMA = 6.4 ns + 6.4 ns + 0.22 ns ~ 13 ns per direction**
- **PMD contribution should be low, assume PMD skew by itself < 1ns**

Maximum Transmission skew of PMDs

PMD	Description	Max Skew Budget ns (UI@10G)	Notes
100GBASE-ER4	100GE 40 km	1.3ns (13UI)	
100GBASE-LR4	100GE 10 km	0.3ns (3UI)	
100GBASE-SR10	100GE 100m	4.5ns (45UI)	
100GBASE-SR10+	100GE 300m	13.6ns (136UI)	Speculative
100GBASE-CR10	100GE Copper 10m	0.5ns (5UI)	
100GBASE-CR10+	100GE Copper 30m	1.5ns (15UI)	Speculative
40GBASE-KR4	40GE backplane	??	
40GBASE-LR4	40GE 10 km	1.7ns (17UI)	
40GBASE-SR4	40GE 100m	4.5ns (45UI)	
40GBASE-SR4+	40GE 300m	13.6ns (136UI)	Speculative
40GBASE-CR4	40GE Copper 10m	0.5ns (5UI)	
40GBASE-CR4+	40GE Copper 30m	1.5ns (15UI)	Speculative

Note: This is a mix of PMDs that we have objectives for and various speculative PMDs

Recommended maximum skew contributions

Contributor	Maximum	Proposed (ns)	Proposed (UI 10G VL)	Proposed (UI 5G VL)
PCS TX, FEC, PMA (at CAUI/XLAUI)	25.5ns	28ns	~289UI	~144UI
Electrical CAUI/XLAUI i/f TX	.88ns	1ns	~10UI	~5UI
PMA TX	13ns	13ns	~134UI	~67UI
Electrical PMD service i/f	0.22ns	1ns	~10UI	~5UI
PMD TX	<1ns	1ns	~10UI	~5UI
Transmission	13.6ns	100ns	~103UI	~206UI
PMD RX	<1ns	1ns	~10UI	~5UI
Electrical PMD service i/f	0.22ns	1ns	~10UI	~5UI
PMA RX	13ns	13ns	~134UI	~67UI
Electrical CAUI/XLAUI i/f RX	.88ns	1ns	~10UI	~5UI
PCS RX, FEC, PMA	14.3ns	20ns	~206UI	~103UI

Proposed Max Skew Point values

Skew Point	Proposed Standard	UI for 10G VL	UI for 5G VL	Normative/ Informative
SP1	29ns	~299 UI	~150 UI	Normative if exposed
SP2	43ns	~443 UI	~222 UI	Normative if exposed
SP3	44ns	~454 UI	~227 UI	Normative
SP4	144ns	~1484 UI	~742 UI	Normative
SP5	146ns	~1505 UI	~753 UI	Normative if exposed
SP6	160ns	~1649 UI	~824 UI	Normative if exposed
TOTAL (at PCS RX)	180ns	~1856 UI	~928 UI	

For all the SP (Skew Points) that are measurable (exposed), the recommendation is to make them Normative
For Skew points with associated PCB track, the spec would be of the form xx ns including 1 ns allowance for track

Maximum dynamic skew of PMDs

PMD	'Standard'	Transmission medium	Notes
100GBASE-ER4	100GE 40 km	373ps (9.6UI)	
100GBASE-LR4	100GE 10 km	93ps (2.4UI)	
100GBASE-SR10	100GE 100 m	676ps (7UI)	
100GBASE-SR10+	100GE 300 m	2.0ns (21UI)	Speculative
100GBASE-CR10	100GE Copper 10m	50ps (0.5UI)	
100GBASE-CR10+	100GE Copper 30m	150ps (1.5UI)	Speculative
40GBASE-KR4	40GE Backplane	???	
40GBASE-LR4	40GE 10 km	766ps (8UI)	
40GBASE-SR4	40GE 100 m	676ps (7UI)	
40GBASE-SR4+	40GE 300 m	2.0ns (21UI)	Speculative
40GBASE-CR4	40GE Copper 10m	50ps (0.5UI)	
40GBASE-CR4+	40GE Copper 30m	150ps (1.5UI)	Speculative

Note: This is a mix of PMDs that we have objectives for and various speculative PMDs

Recommended dynamic skew contributions

Contributor	Maximum	Proposed
PCS TX, FEC, PMA (at CAUI/XLAUI)	194ps (2UI)	200ps (~2UI)
Electrical CAUI/XLAUI i/f TX	0	0
PMA TX	194ps (2UI)	200ps (~2UI)
Electrical PMD service i/f	0	0
PMD TX	<100ps (<1UI)	200ps (~2UI)
Transmission	2.0ns (21UI)	2.8ns (~29UI)
PMD RX	<100ps (<1UI)	200ps (~2UI)
Electrical PMD service i/f	0	0
PMA RX	194ps (2UI)	200ps (~2UI)
Electrical CAUI/XLAUI i/f RX	0	0
PCS RX, FEC, PMA	194ps (2UI)	200ps (~2UI)

Note: All UI are @10G

Proposed Dynamic Skew Point values

Skew Point	Proposed Standard	UI at 10G	UI at 25G	Normative/ Informative
SP1	200ps	~2UI		Normative if exposed
SP2	400ps	~4UI	(~10UI)	Normative if exposed
SP3	600ps	~6UI	~15UI	Normative
SP4	3.4ns	~35UI	~88UI	Normative
SP5	3.6ns	~37UI	(~93UI)	Normative if exposed
SP6	3.8ns	~39UI		Normative if exposed
At PMA/PCS RX	4ns	~41UI		

For all the SP (Skew Points) that are measurable (exposed),
the recommendation is to make them Normative

More on Dynamic Skew

- **Ok, I have the dynamic skew numbers, now what?**
- **For designs with a PMA gearbox (m != n), the gearbox has a dynamic skew buffer per input lane**

  Size is 2x the max dynamic skew for that corresponding path

  Start reading out of the wander buffers when they are half full

- **For designs without a PMA gearbox (m=n), the maximum skew already includes the dynamic skew numbers, your receive PCS input FIFOs/buffers need to be able to track the dynamic skew**

  Note: An increase in maximum skew capability does not impact latency, only buffer depth. An increase in dynamic skew capability does increase latency because you must wait to start reading out data from the receive FIFOs until there is enough data in the least filled FIFO to allow for the maximum dynamic skew variation that we expect for a worst case interface.

- **In addition, depending on the PMA design it might need to track dynamic skew**

  For instance, if you clock all outputs with a common clock

Q & A

Thank you !

Backup - Fiber characteristics tool

- **Fiber characteristics tools (spreadsheets) officially adopted by IEEE and used by P. Anslow to calculate transmission skews are in:**

http://www.ieee802.org/3/ba/public/tools/Fibre_characteristics_V_3_0.xls

http://www.ieee802.org/3/ba/public/may08/kolesar_02_0508.xls