Optimum 40GBase-CR4 Lane Assignment

Jan 24, 2010

Ali Ghiasi
Broadcom Corporation
aghiasi@broadcom.com

Problem Statement

- Response to comment 872
- Classic QSFP pin out carried forward in 802.3ba is not optimized if lane orders must be maintained
  - Clause 85 requires maintaining connecting lanes to MDI contact per definition of table 85-12 because of training requirements
  - On other hand Clause 86 allow any lane to MDI contact connection
- With single routing layer one could connect host ASIC to 40GBase-SR4 MDI but the requirement of CL85 require to use 3 precious routing layers
  - This will force use of more expensive board with blind vias
- The conflicting requirements of CL85 and 86 could also result in incompatible systems

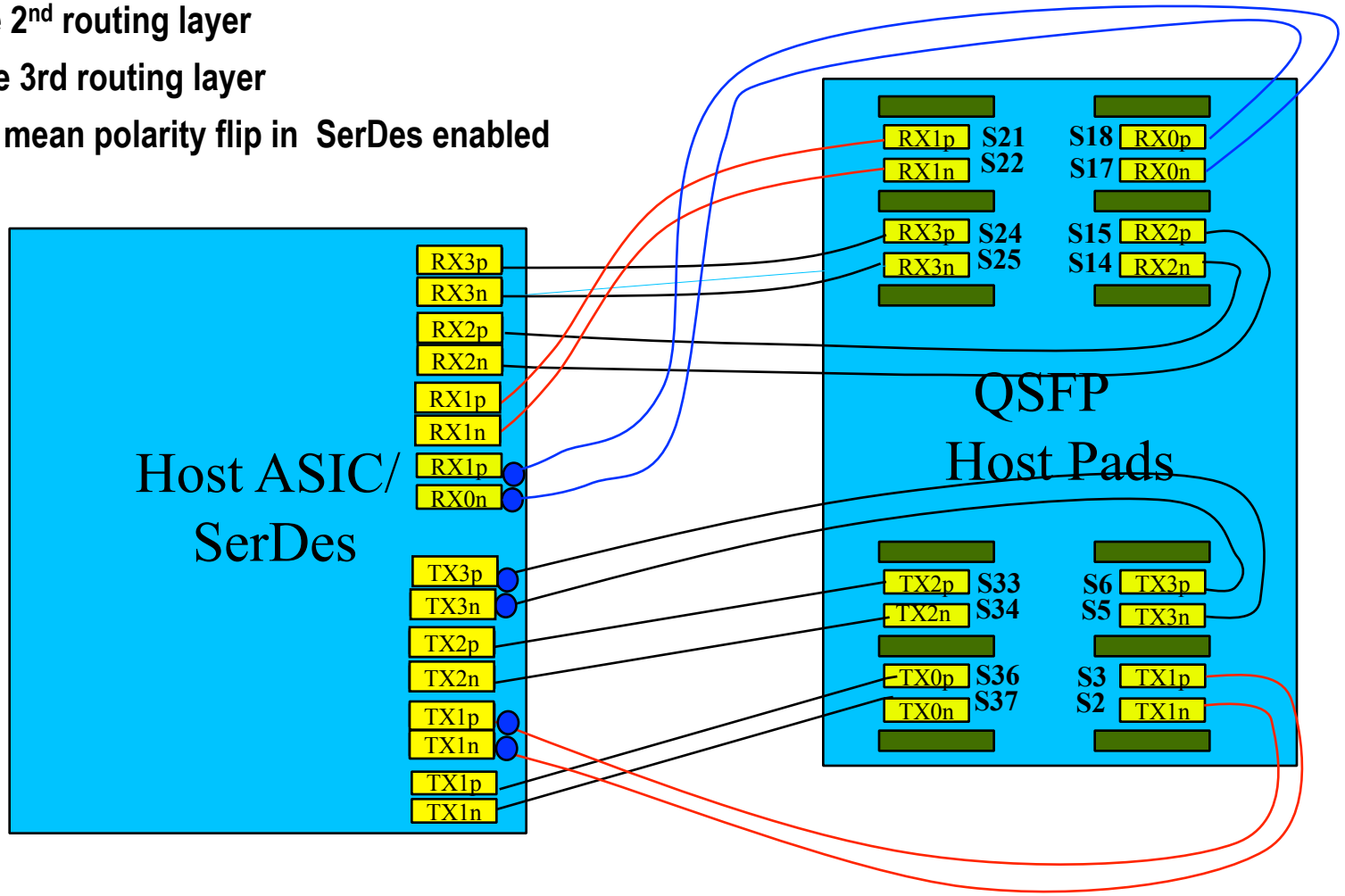
Current 10Gbase-CR4 lane to MDI Contact Mapping

Table 85–12—Style-1 40GBASE-CR4 lane to MDI connector contact mapping

Tx lane	MDI Connector contact	Rx lane	MDI Connector contact
signal gnd	S1	signal gnd	S13
SL1<n>	S2	DL2<p>	S14
SL1<p>	S3	DL2<n>	S15
signal gnd	S4	signal gnd	S16
SL3<n>	S5	DL0<p>	S17
SL3<p>	S6	DL0<n>	S18
signal gnd	S7	signal gnd	S19
SL2<p>	S33	DL1<n>	S21
SL2<n>	S34	DL1<p>	S22
signal gnd	S35	signal gnd	S23
SL0<p>	S36	DL3<n>	S24
SL0<n>	S37	DL3<p>	S25
signal gnd	S38	signal gnd	S26

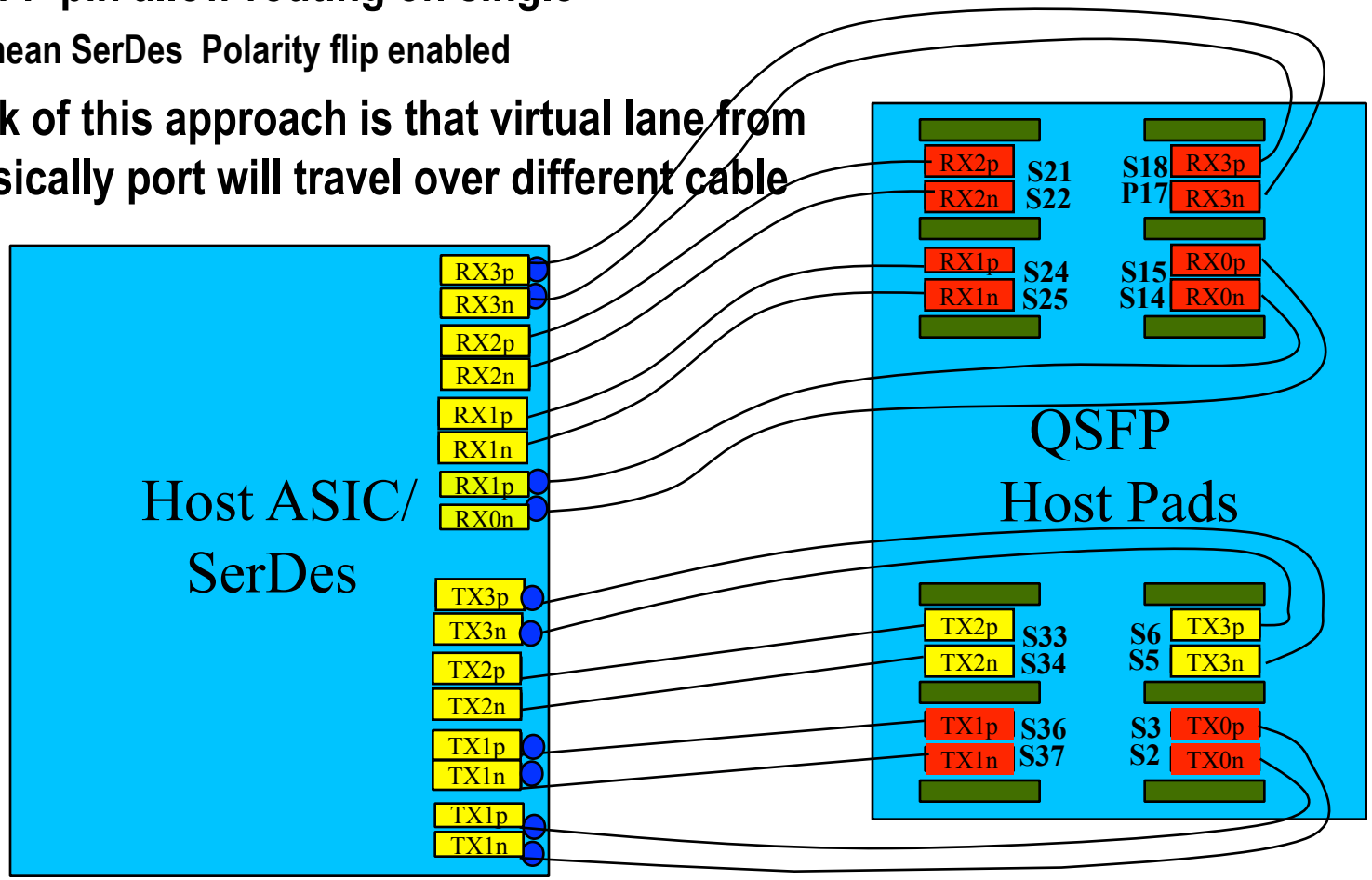
Current QSFP Pinout

- QSFP pin out is not optimized and require 3 routing channel
  - Red is the 2nd routing layer
  - Blue is the 3rd routing layer
  - Blue dots mean polarity flip in SerDes enabled



Optimized QSFP Pinout

- Optimized QSFP pin allow routing on single
  - Blue dots mean SerDes Polarity flip enabled
- One draw back of this approach is that virtual lane from the same physically port will travel over different cable pairs



* Red pads on QSFP means contact was reassigned

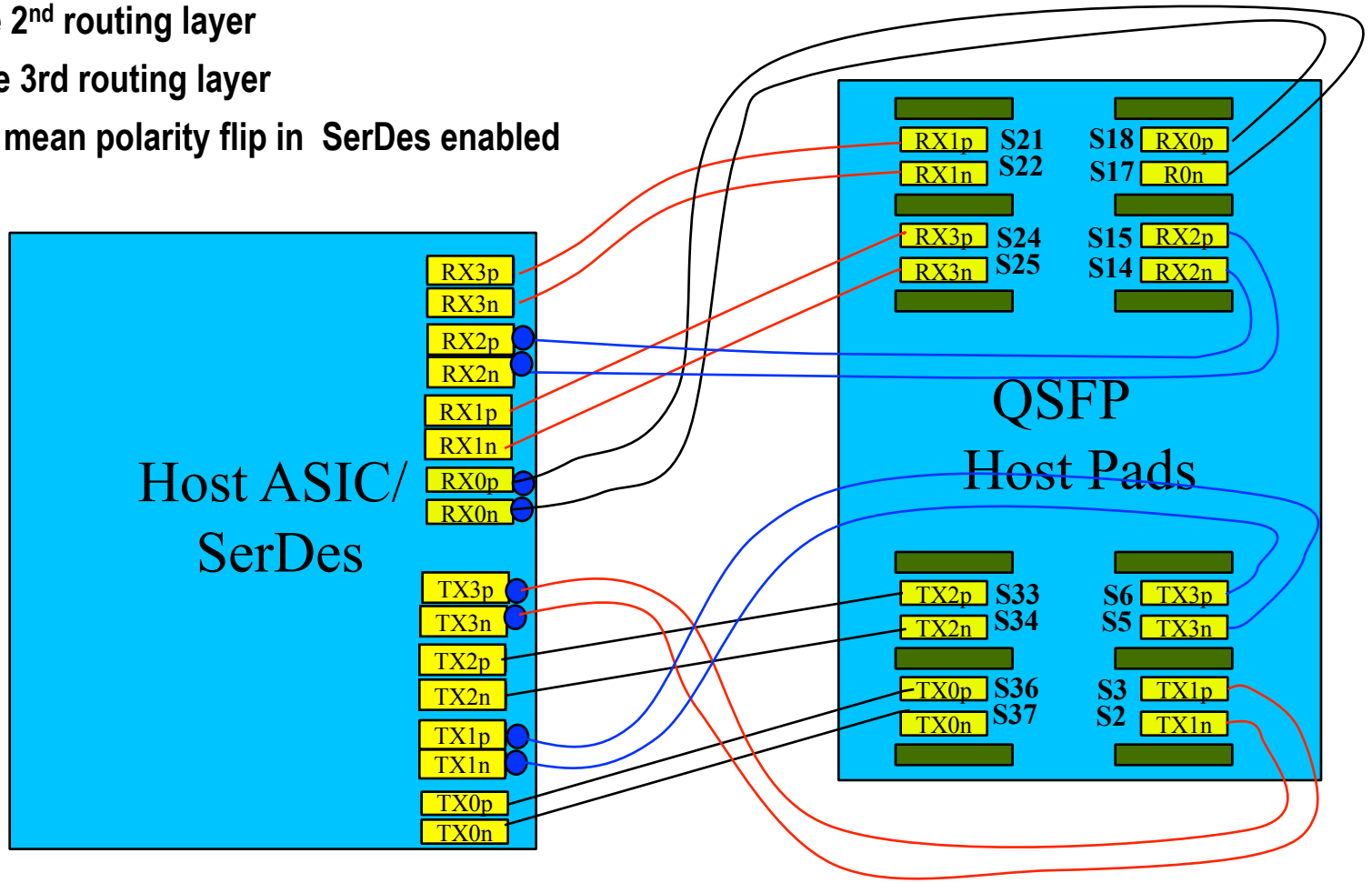
Optimum 10Gbase-CR4 lane to MDI Contact Mapping

Table 85–12—Style-1 40GBASE-CR4 lane to MDI connector contact mapping

Tx lane	MDI Connector contact	Rx lane	MDI Connector contact
signal gnd	S1	signal gnd	S13
SL0 ~~SL1<n>~~	S2	SL0 ~~DL2<p>~~	S14
SL0 ~~SL1<p>~~	S3	SL0 ~~DL2<n>~~	S15
signal gnd	S4	signal gnd	S16
SL3<n>	S5	DL3 ~~DL0<p>~~	S17
SL3<p>	S6	DL3 ~~DL0<n>~~	S18
signal gnd	S7	signal gnd	S19
SL2<p>	S33	DL2 ~~DL1<n>~~	S21
SL2<n>	S34	DL2 ~~DL1<p>~~	S22
signal gnd	S35	signal gnd	S23
SL1 ~~SL0<p>~~	S36	DL1 ~~DL3<n>~~	S24
SL1 ~~SL0<n>~~	S37	DL1 ~~DL3<p>~~	S25
signal gnd	S38	signal gnd	S26

Example Host Levering MLD Flexibility

- In this example TX2/RX2<n,p> where swapped with TX4/RX4<n,p>
  - Red is the 2nd routing layer
  - Blue is the 3rd routing layer
  - Blue dots mean polarity flip in SerDes enabled



PMD to MDI Mapping 40GBase-CR4 and 100Gbase-CR10

- Change MDI contact name form SLx to TXx and change DLx to RXx in table 85-12, 85-13, and 85-14
- Create a new table showing mapping of PMD lanes to MDI contacts as shown below
  - Configuration below guarantees CR4 auto-negotiation and make sure transmitter or receiver are not crosswired

Transmit PMD Lane	Transmit MDI Contact	Receive PMD Lane	Receive MDI contact
SL0<p>	TX0<p>	DL0<p>	RX0<p>
SL0<n>	TX0<n>	DL0<n>	DX0<n>
xLi<p>	TXi<p>	xLi<p>	RXi<p>
xLi<n>	TXi<n>	xLi<n>	RXi<n>
for 40Gbase-CR4 i=1, 2, 3 and for 100Gbase-CR10 i=1,2,..,9. For a given lane identifier value_i, x can take value of S or D respectively identifying either transmit PMD or receive PMD lanes.			

Summary

- The best solution would be to redefine the MDI contact definition for optimum host PCB layout
  - Since where we are in the project we would have to leave the cable definition unchanged
  - However this approach will carry given TX/RX lane traffic on different cable pairs which could complicate the testing
- Choosing an optimum MDI contact assignment can reduce the PCB routing layers from 3 to 1 and not choosing connector definition results in unnecessary pain we all have to carry!
- Allowing CL85 to take advantage of MLD lane reordering could provide some relief for the CR4/CR10 hosts
  - This is the best option for where we are in the project even in the example shown here to advantage was seen.