



40 and 100 Gigabit Ethernet  
PCS and PMA Overview



**Mark Gustlin**

**Ethernet Summit**

**February 2010**

**San Jose**

# Agenda

- PCS and PMA requirements
- PCS Overview
- 64B/66B encoding
- Data Distribution
- PMA Multiplexing
- Optional FEC
- Summary

# Requirement for the PCS and PMA

- PCS = Physical Coding Sublayer, PMA = Physical Medium Attachment
- The PCS performs the following functions:
  - Delineates Ethernet frames
  - Supports the transport of fault information
  - Provides the data transitions which are needed for clock recovery on SerDes and optical interfaces
  - Bonds multiple lanes together through a striping/distribution mechanism
  - Supports data reassembly in the receive PCS even in the face of significant parallel skew and with multiple multiplexing locations
- The PMA performs the following functions:
  - Bit level multiplexing from M lanes to N lanes
  - Clock recovery, clock generation and data drivers
  - Loopbacks and test pattern generation and detection

# 100/40GE PCS Overview

- 10GBASE-R 64B/66B based PCS (10 Gb/s serial PCS)
  - Run at 100 Gb/s or 40 Gb/s serial rate instead of 10 Gb/s
  - Includes 66 bit block encoding and scrambling
- Multi-Lane Distribution
  - Data is distributed across n PCS lanes 66 bit blocks at a time
  - Round robin distribution
  - Periodically, unique alignment marker blocks are added to each virtual lane to allow deskew in the receive PCS
- PMA maps n lanes to m lanes
  - PMA performs simple bit level multiplexing
  - Does not know or care about PCS coding
- Alignment and static skew compensation is done in the Rx PCS only

# 64B/66B Encoding Details

- 10GBASE-R 64B/66B based PCS just run faster
- Adds a two bit sync header to each 64 bit block of data, legal values are '01' for data, and '10' for control blocks

The fact that '00' and 11' are not used allows the receiver to sync up to the blocks

Block lock state machine looks for 64 '01' or '10' patterns 66 bits apart to declare lock (no instances of '00' and 11')

Block lock state machine looks for 16 instances of '00' and 11' within 64 sync headers to declare out of lock

- Delineates frames and control information
- Scrambles data with a self synchronous scrambler

Limits baseline wander and provides transitions for clock recovery

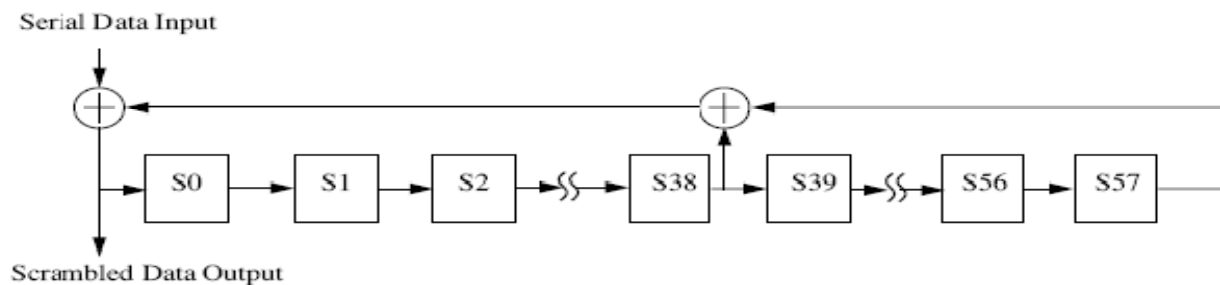


Figure 49-8— Scrambler

# 64b/66b Block Formats

- Data Block Format



- Control Block Formats



- Idle Block Example



- Start Block Format



- 7Byte Terminate Block Format



# The Need for Data Striping

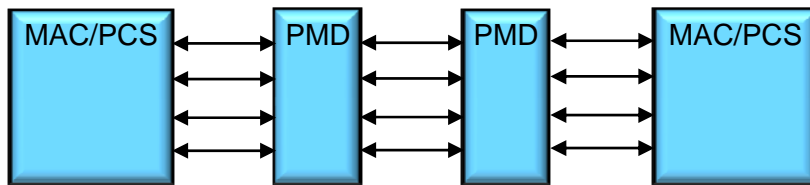
- All PMDs for 100 Gb/s and 40 Gb/s Ethernet have multiple lanes

Either multiple fibers, coax cables, wavelengths or backplane traces

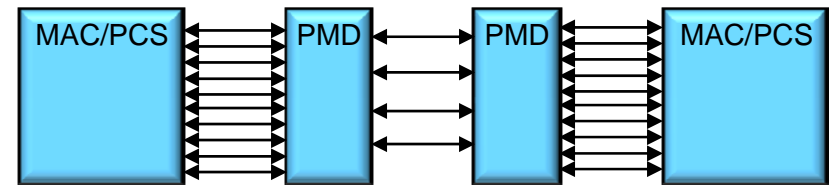
Individual bit rates of 10.3125 Gb/s or 25.78125 Gb/s (new PMD will have a rate of 41.25 Gb/s)

Module interfaces are also multiple lanes, not always the same number of lanes as the PMD interface

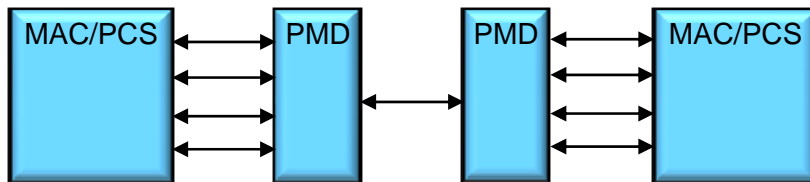
- Therefore the PCS must support a mechanism to distribute data to multiple lanes on the transmit side, and then reassemble the data in the face of skew on the receive side



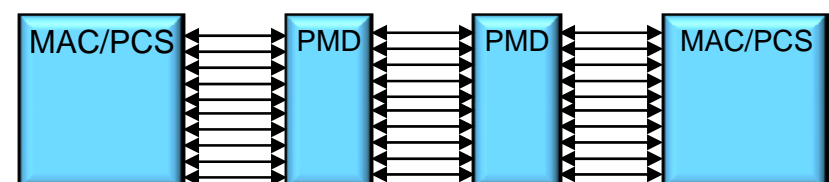
All Current Variants of 40 Gb/s Ethernet



Single-mode 100 Gb/s Ethernet



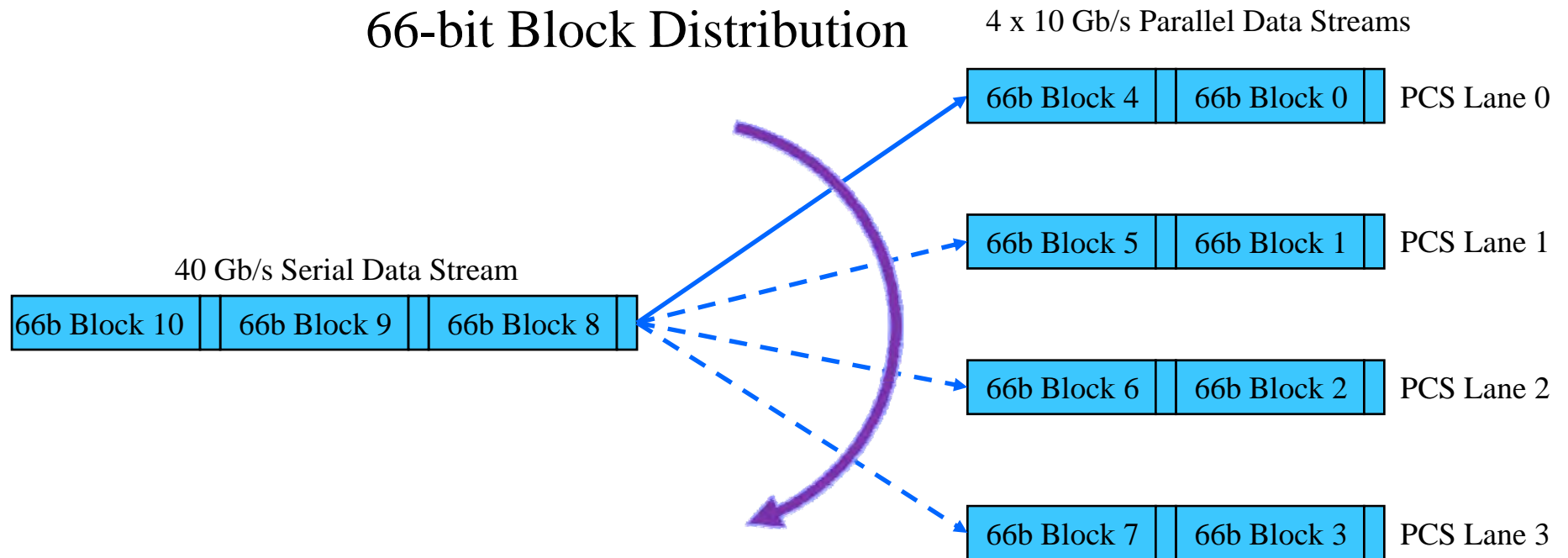
40Gb/s Ethernet Single-mode Fibre PMD Study Group



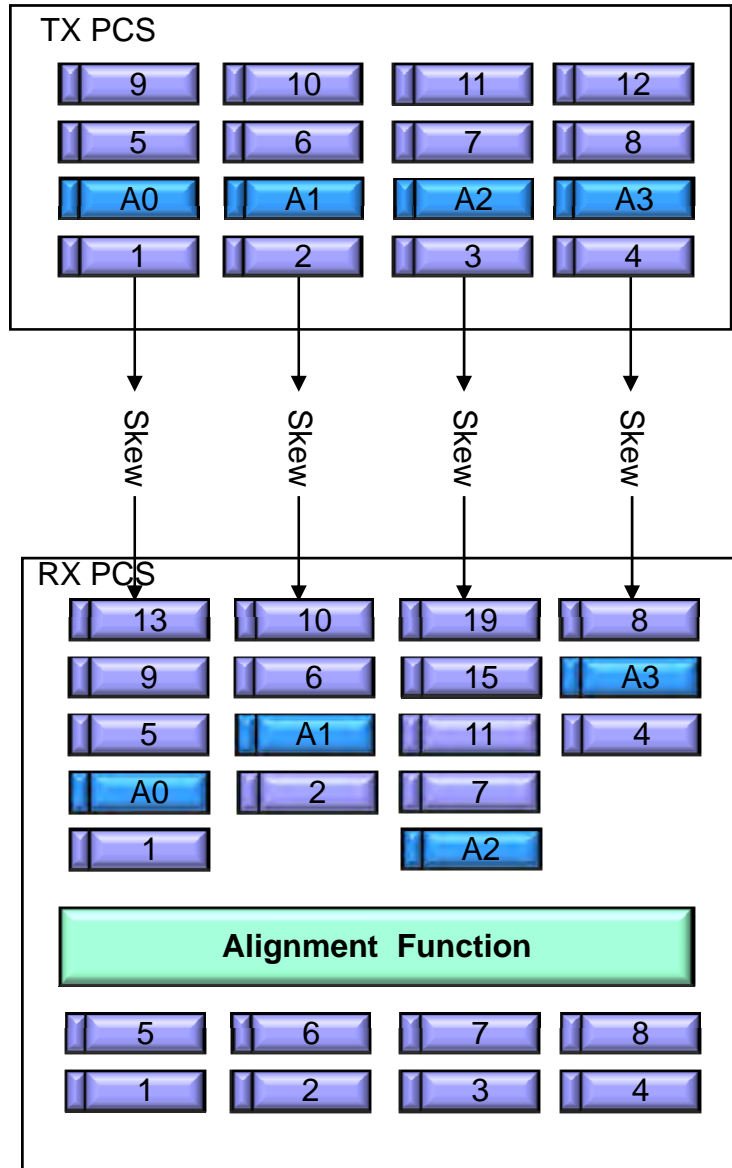
Multi Mode 100 Gb/s Ethernet

# Transmit Data Striping – 40 Gb/s

## Round Robin 66-bit Block Distribution



# Alignment Mechanism – 40 Gb/s Example



## TX PCS Functions:

Encode data into blocks

Scramble data

Add alignment markers periodically

Every 16k blocks on each lane

## RX PCS Functions:

Re-Align 66 bit blocks

Remove the Alignment blocks

Then descramble and decode

Alignment markers are unique 66b blocks, each lane has a unique marker, markers are not scrambled

# Alignment Marker Block Format

- **Alignment marker format:**



- The alignment markers are added after the data stream has been encoded, scrambled and distributed to multiple lanes
- The alignment markers are not scrambled, this allows the receiver to find the markers before descrambling
  - Which is necessary to deskew the streams
- The 24 bit Marker x field is populated with a fixed value per PCS lane
  - Looks like a scrambled pattern, but it remains constant on a given lane
  - 20 unique markers for 100 Gb/s, 4 markers for 40 Gb/s
- Alignment markers are inserted every 16k blocks on each lane at the same time, interrupting data
- Each alignment marker includes an 8 bit Bit Interleaved Parity (BIP) calculation for BER determination
  - Covers all bits since the last alignment marker was sent
- The Marker field and the BIP field are both inverted in the 2<sup>nd</sup> half of the alignment marker to provide a DC balanced block

# Changing Widths

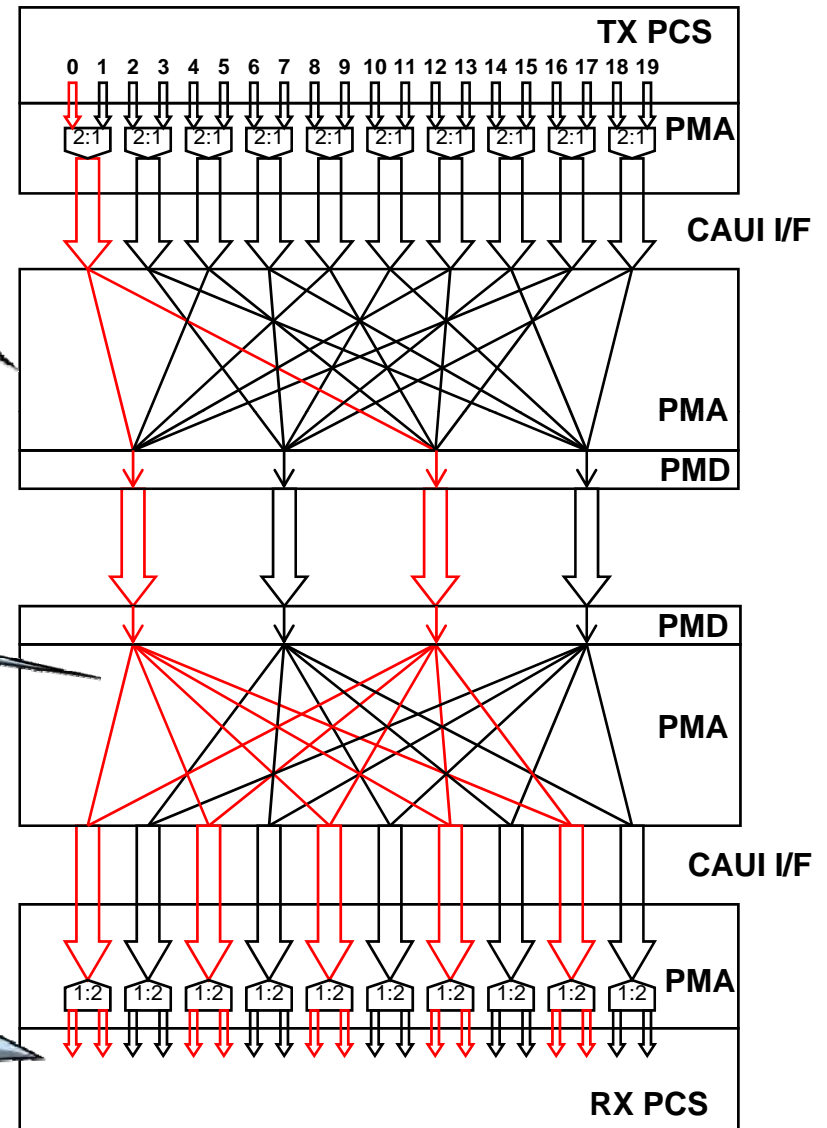
- For the 100 Gb/s Ethernet single mode PMD, we have a 10 lane electrical module interface and then a 4 lane (wavelength) PMD
- The standard is defined so that all that must be done to switch widths is simple bit multiplexing (and de-multiplexing on the other end)
- No need to deskew lanes before changing widths in the optical module
- 40 Gb/s Ethernet has four PCS Lanes, this can support 4, 2 or 1 lanes
  - Initial standard will only use 4 lanes, new PMD standard likely to use 1 lane
- 100 Gb/s Ethernet has twenty PCS Lanes, this can support 20, 10, 5, 4, 2 or 1 lanes
  - Initial standard will use 10 or 4 lanes or a combination of the two

# Possible Paths Through the Link

Skew on input electrical interface determines which optical lane PCS Lane 0 passes through

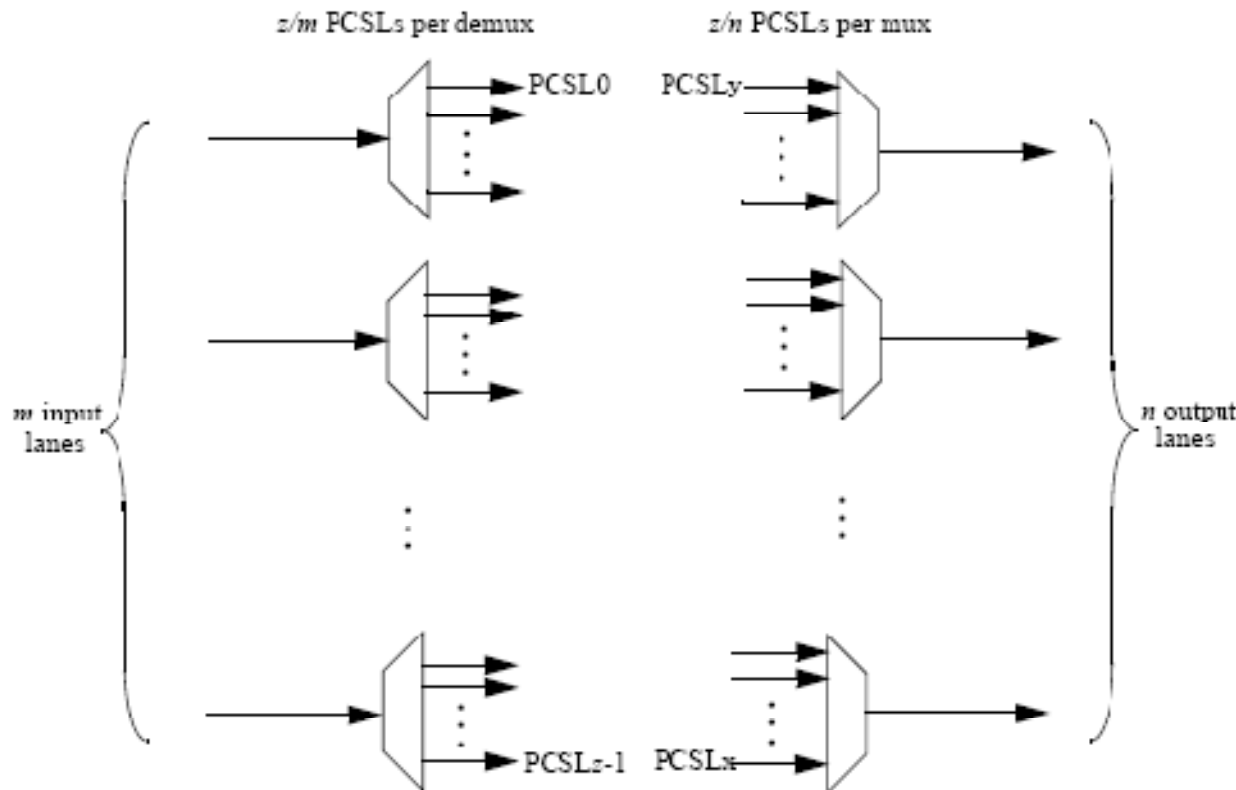
Skew on optical interface and previous electrical interface determines which output electrical lane PCS Lane 0 passes through

PCS Lane 0 can appear on any of the red outputs depending on the skew of the electrical and optical interfaces. PCS receivers are designed to accept any PCS Lane on any physical lane.



# PMA Functions

- A primary function of the PMA is to multiplex M input lanes to N output lanes where needed
  - Bit level multiplexing only
- The PMA also performs clock recovery, clock conversion, test pattern generation and detection and loopbacks where applicable



# Forward Error Correction

- The copper interfaces, cable and backplane, can support optional FEC to improve Bit Error Rate performance
- Re-uses the Clause 74 FEC from backplane Ethernet (10 Gb/s)
- If used, an independent FEC encoding stream is run on each PCS Lane  
A total of 4 or 20 FEC streams
- Utilizes the redundancy that is inherent in the 64b/66b sync field

Table 74-1—FEC block format

T <sub>0</sub>	64 bit payload Word 0	T <sub>1</sub>	64 bit payload Word 1	T <sub>2</sub>	64 bit payload Word 2	T <sub>3</sub>	64 bit payload Word 3
T <sub>4</sub>	64 bit payload Word 4	T <sub>5</sub>	64 bit payload Word 5	T <sub>6</sub>	64 bit payload Word 6	T <sub>7</sub>	64 bit payload Word 7
T <sub>8</sub>	64 bit payload Word 8	T <sub>9</sub>	64 bit payload Word 9	T <sub>10</sub>	64 bit payload Word 10	T <sub>11</sub>	64 bit payload Word 11
T <sub>12</sub>	64 bit payload Word 12	T <sub>13</sub>	64 bit payload Word 13	T <sub>14</sub>	64 bit payload Word 14	T <sub>15</sub>	64 bit payload Word 15
T <sub>16</sub>	64 bit payload Word 16	T <sub>17</sub>	64 bit payload Word 17	T <sub>18</sub>	64 bit payload Word 18	T <sub>19</sub>	64 bit payload Word 19
T <sub>20</sub>	64 bit payload Word 20	T <sub>21</sub>	64 bit payload Word 21	T <sub>22</sub>	64 bit payload Word 22	T <sub>23</sub>	64 bit payload Word 23
T <sub>24</sub>	64 bit payload Word 24	T <sub>25</sub>	64 bit payload Word 25	T <sub>26</sub>	64 bit payload Word 26	T <sub>27</sub>	64 bit payload Word 27
T <sub>28</sub>	64 bit payload Word 28	T <sub>29</sub>	64 bit payload Word 29	T <sub>30</sub>	64 bit payload Word 30	T <sub>31</sub>	64 bit payload Word 31
32 parity bits							

# Summary

- Simple 10GBASE-R based PCS
- Distribution layer to support multiple physical lanes/lambdas
- Complexity in the optical module is low
  - Simple bit muxing even when  $m \neq n$
- Based on proven 64B/66B framing and scrambling
- Supports an evolution of optics and electrical interfaces
  - New Single-mode PMD will not need a change to the PCS