100GE Layering Model Overview

Trey Malpass
Ye Min

IEEE 802.3 Higher Speed Study Group 10-13 Sept 2007

HUAWEI TECHNOLOGIES Co., Ltd.

Supporters

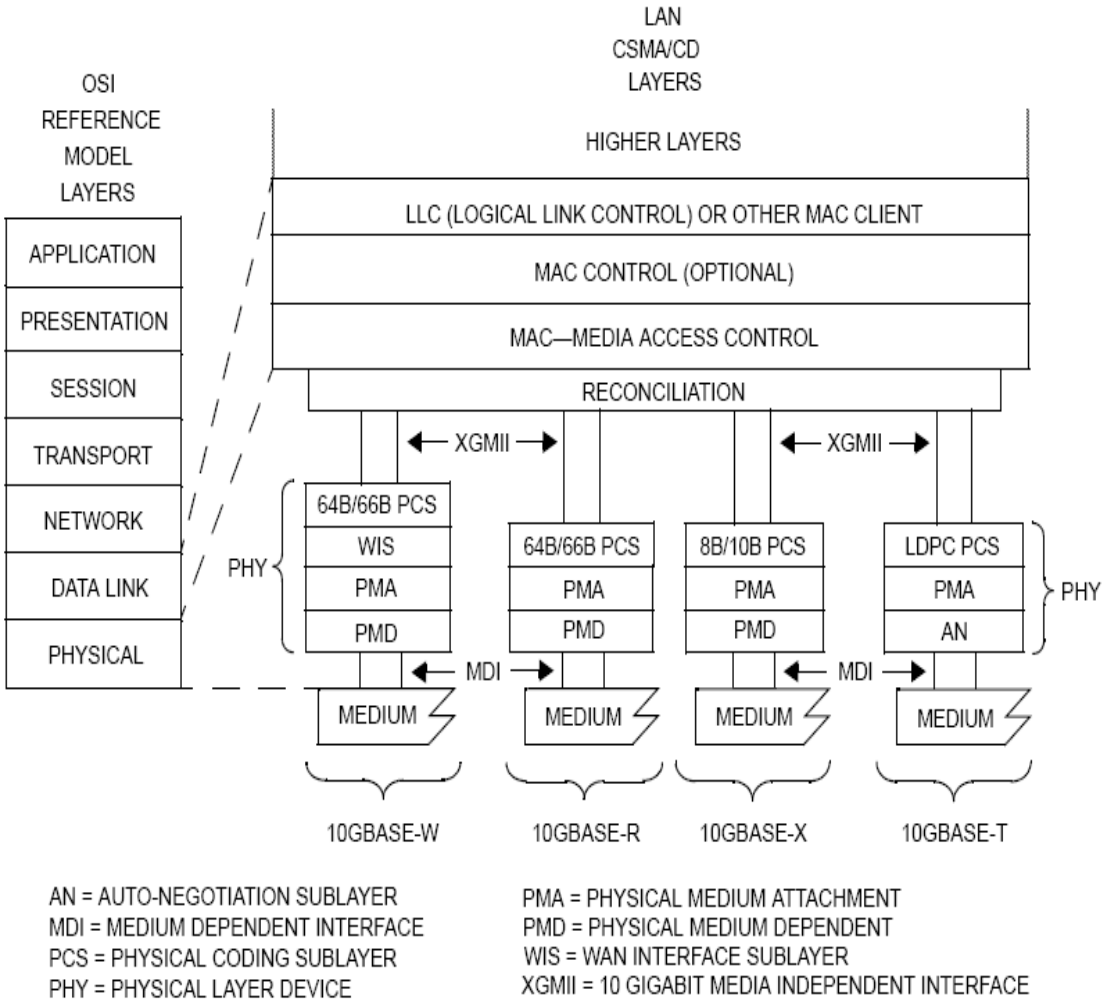
- Frank Chang – Vitesse
- Dan Dove – HP Procurve

Contents

- Ethernet Layer Model
- 100GE Layer Model
- 100GE Multi-Lane Model
- PCS Function in Multi-Lane

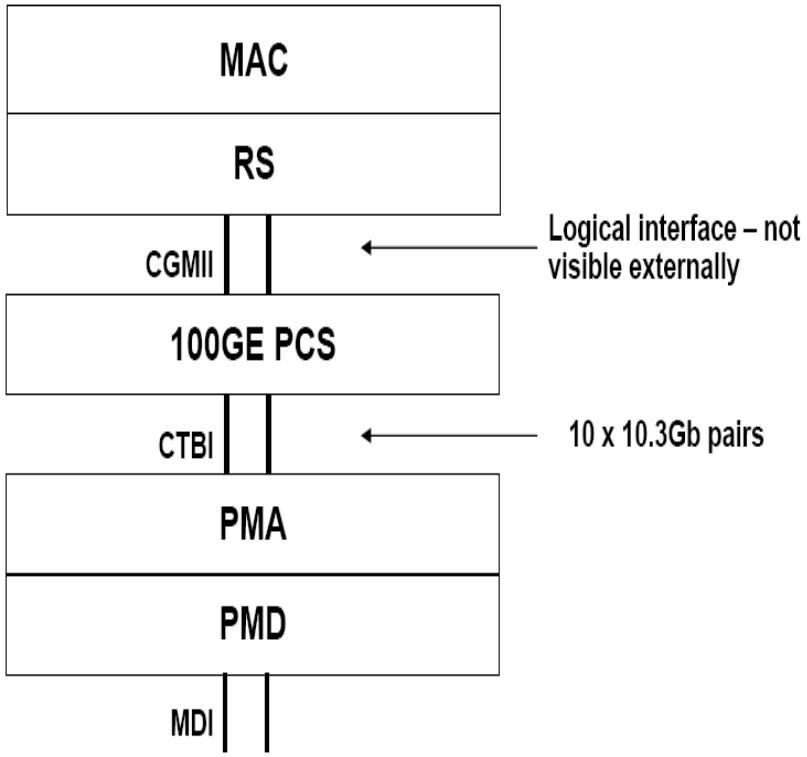
Ethernet Layer Model

- OSI reference model layers
- 10GBASE-R Physical layer
- Includes:
  - RS
  - PCS
  - PMA
  - PMD



100GE Layer Model

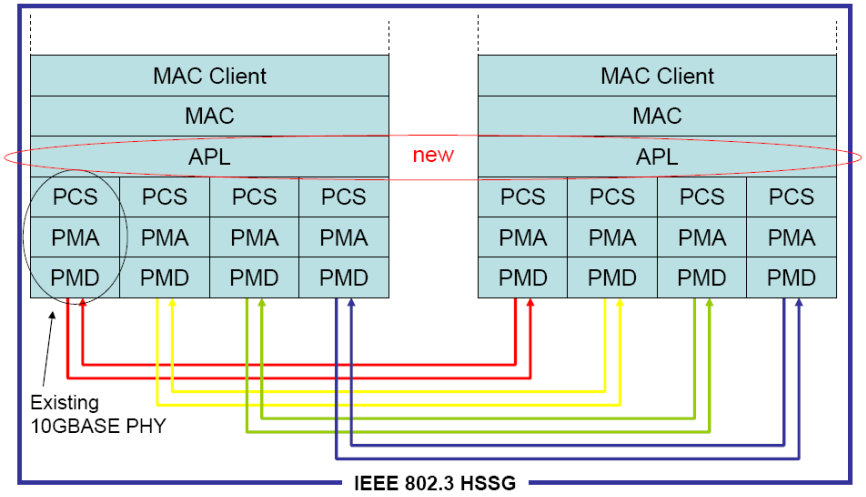
- One option
  - Single PCS that is tolerant to different PMDs
  - CTBI is a high-speed bit interleaved interface, virtual lanes enable deskewing of parallel bus.
  - CTBI could also be used to interconnect backplane or chip-to-chip.

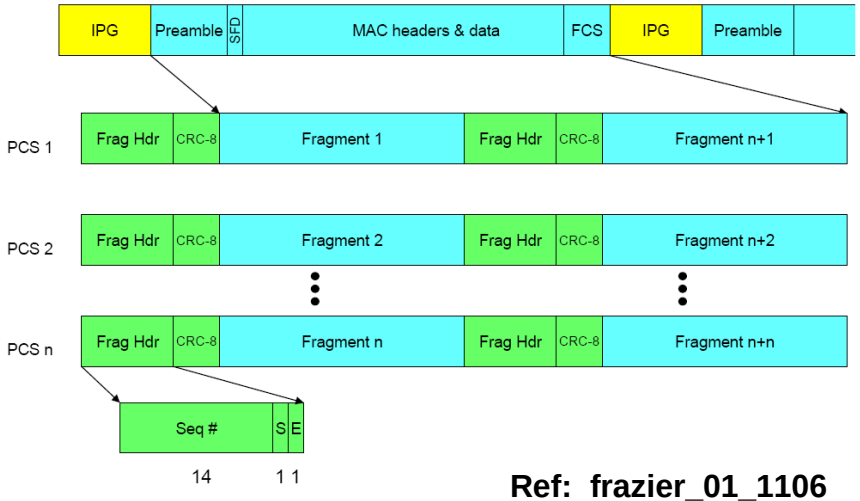


Ref: Gustlin_01_0107

100GE Layer Model

- Another option
  - Individual PCS for each lane of a multi-lane PMD
  - APL can implement packet fragmentation, distribution, collection and reassembly.
  - Fragmentation headers added so that receiver can reassemble fragments correctly

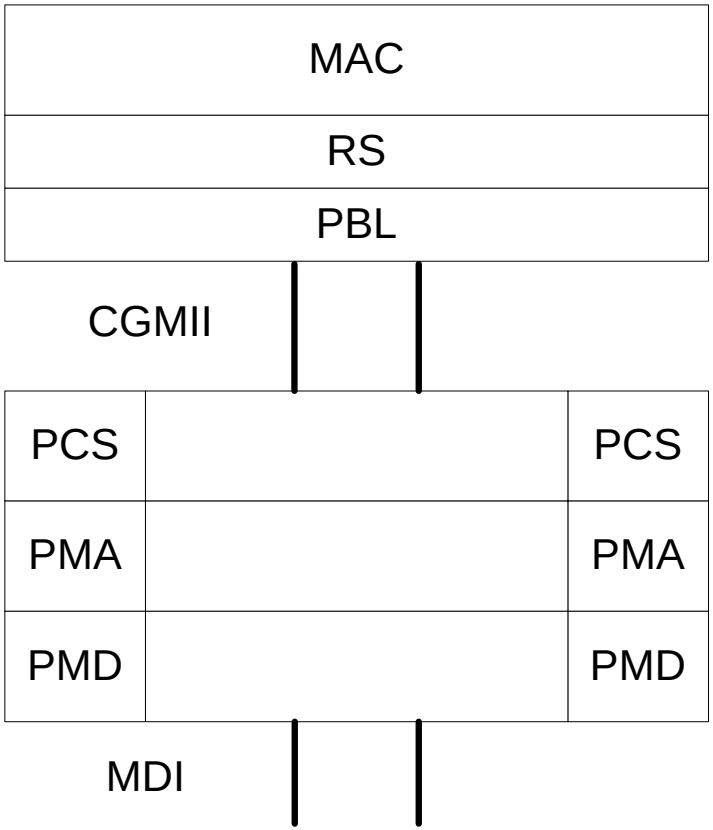




Ref: frazier_01_1106

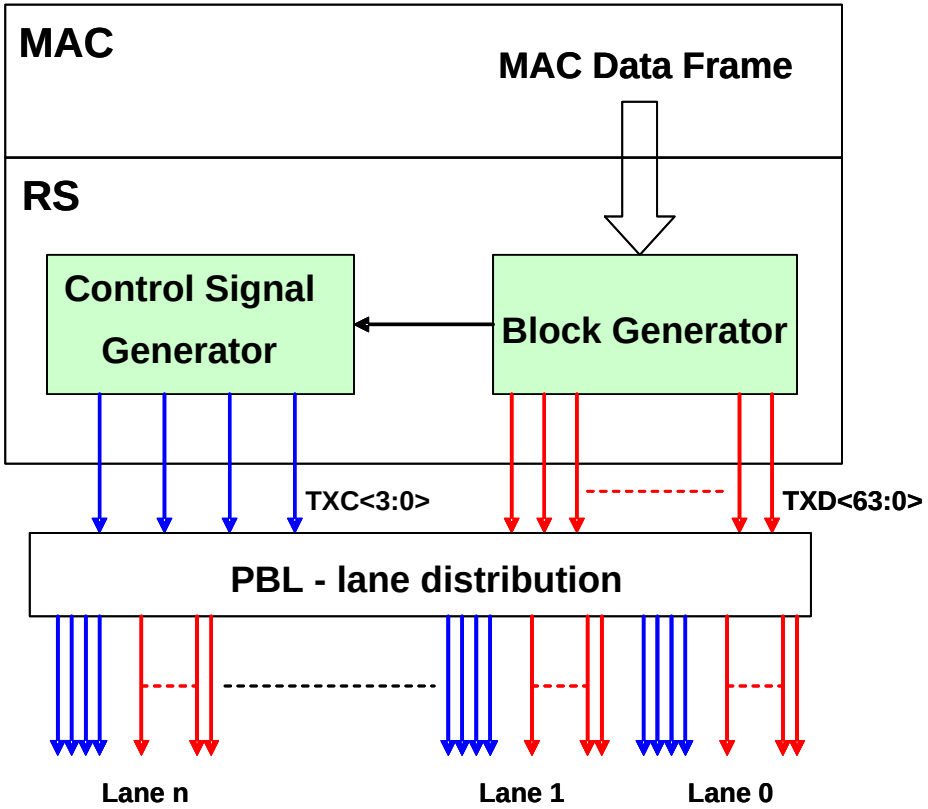
100GE Multi-Lane Model

- Multi-Lane Model
  - PBL (Physical Bundling Layer), which is similar to the APL, bonds the physical lanes
  - PBL distributes the data frame based on 64-bit blocks, supports PCS lanes based on 64B/66B
  - PBL performs the block distribution and reassembly
  - Optionally supports different types/speeds of serial lanes (10*10G, 4*25G)
  - Provides a scalable, resilient multi-lane model



RS (Reconciliation Sublayer)

- **RS will now include a block generator and control generator; data grouped into 64-bit blocks and 4-bit control (similar to 64B/66B) for use by PBL.**
- **Based on the type of block, RS generates the 4-bit control signal.**
- **PBL distributes the 64-bit blocks to each lane. PBL is the block distribution/reassembly function.**



Why use 64-bit blocks?

- **In XGMII, the Interface includes 8bits data (TXD<7:0>) and 1bit control signal (TXC) per lane, TXC indicates data/control state of the byte. For CGMII, the larger data width would require many more TXC bits**

- **In the 100GE multi-lane model, each PCS lane could still use 64B/66B coding. If PBL distributes bytes to different lanes, 64B/66B would need new block types. So,**
  - Distribute data to each lane based on 64-bit blocks for ease of 64B/66B coding

- **Distribute based on 32 bits?**
  - In belhadj_01_1106, if striping based on 32 bits (Column) , many more new block types are needed.

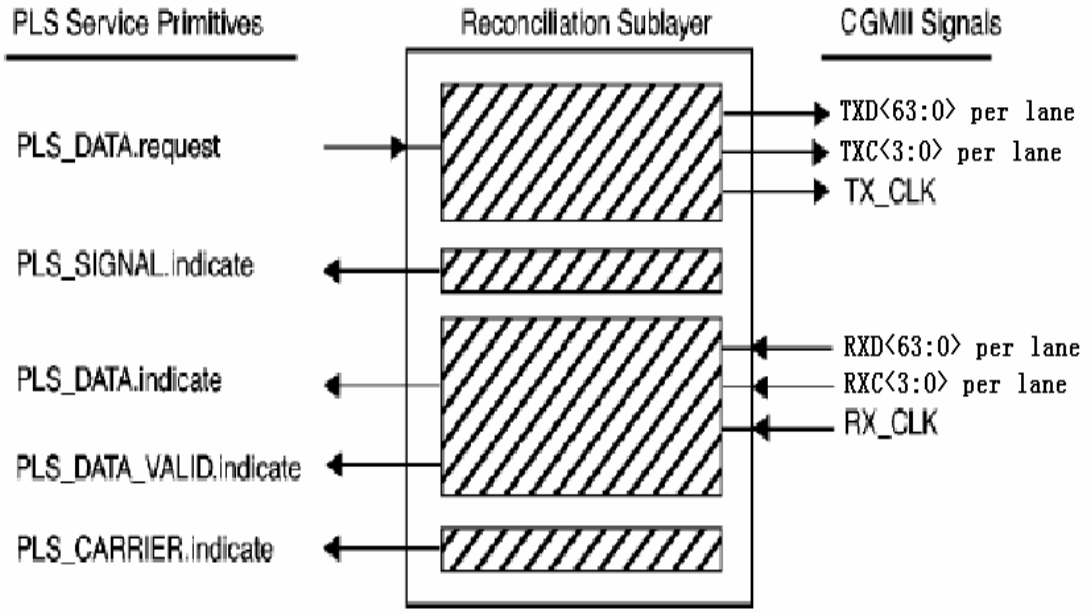
Block and Block Type

- The existing 64B/66B encode uses 16 block types
- In 100GE Multi-Lane model, new type blocks are added for
  - Alignment
  - Error
  - Null
- The Order_set service is for the XGMII or XAUI to monitor link. These codes can be developed into the PHY OAM for 100GE Multi-lane

TXD/RXD								TXC/RXC	Description
D	D	D	D	D	D	D	D	0000	Data
S	D	D	D	D	D	D	D	1001	Start
C	C	C	C	S	D	D	D	1010	Start
T	C	C	C	C	C	C	C	1000	Terminate
D	T	C	C	C	C	C	C	0111	Terminate
D	D	T	C	C	C	C	C	0110	Terminate
D	D	D	T	C	C	C	C	0101	Terminate
D	D	D	D	T	C	C	C	0100	Terminate
D	D	D	D	D	T	C	C	0011	Terminate
D	D	D	D	D	D	T	C	0010	Terminate
D	D	D	D	D	D	D	T	0001	Terminate
C	C	C	C	C	C	C	C	1111	Control
A	A	A	A	A	A	A	A	New	Alignment
E	E	E	E	E	E	E	E	New	Error
N	N	N	N	N	N	N	N	New	Null
								Reserved	Order_set

CGMII Interface

- **CGMII interface**
  - **N= number of lanes in PMD**
  - **N*64data bits (N*TXD<63:0>), N*4 control bits (N*TXC<3:0>) , one clock for transmit**
  - **N*64data bits (N*RXD<63:0>), N*4 control bits (N*RXC<3:0>) , one clock for receive**
  - **4 control bits indicate the block type**



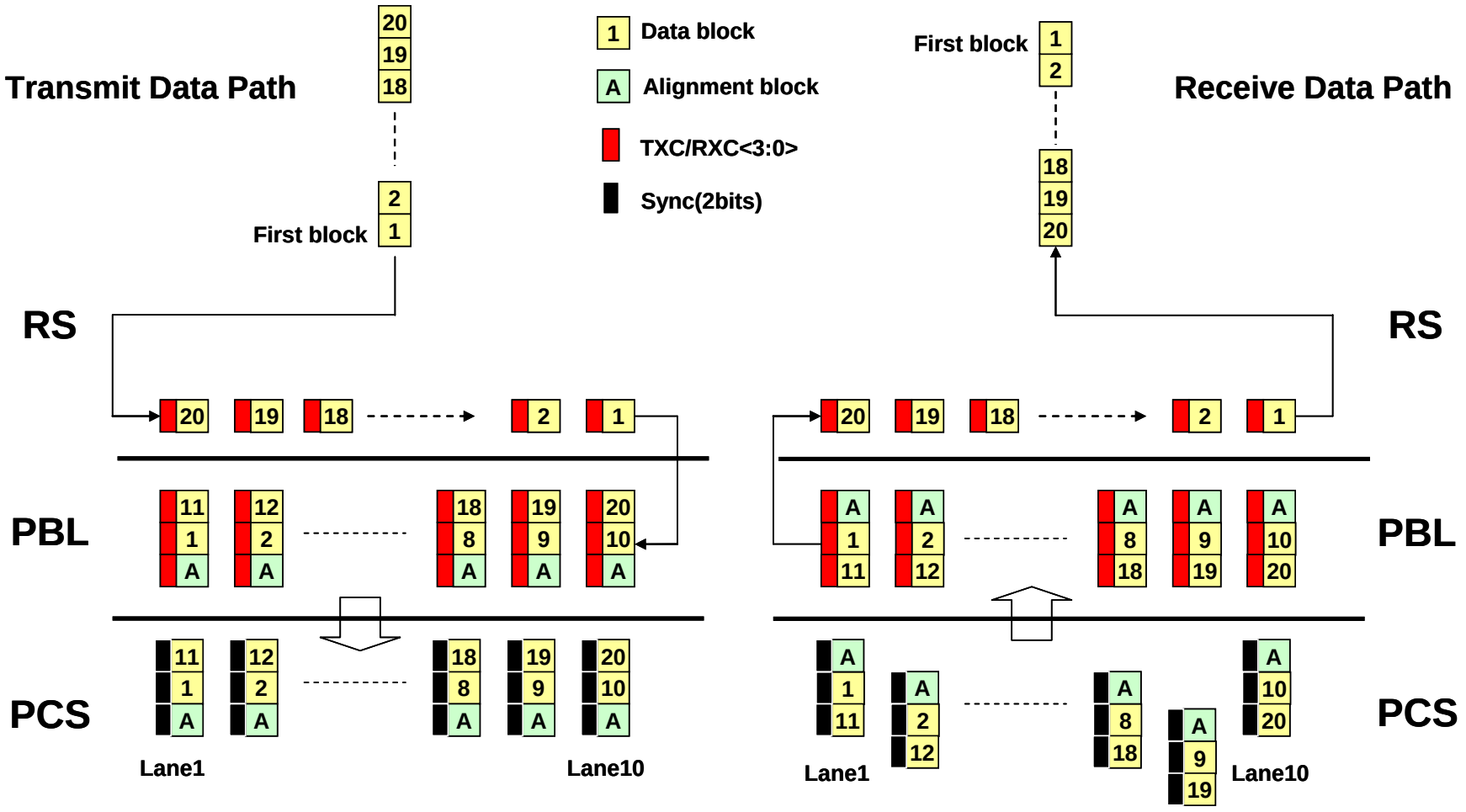
PBL – Physical Bundling Layer

- **Bundle the multi-lanes in the PBL**
  - **Optionally support 1~10 lanes for different bandwidths**
    - **4 lanes = 40GE, 10 lanes = 100GE**
  - **Could support 4x25G lanes for future PMDs with clock-rate change**

- **PBL performs the blocks distribution and reassembly**

- **Alignment mechanism is used to compensate for multi-lane differential delay**

PBL distribution and combination



How to support different numbers of PMD lanes

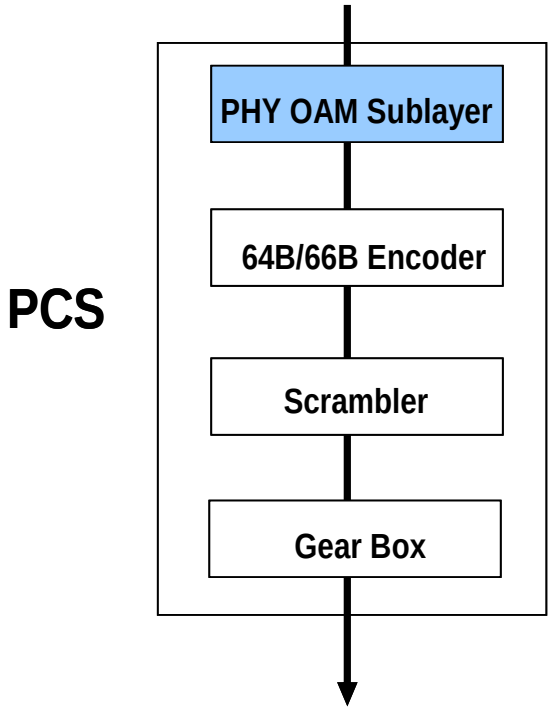
- One option
  - PBL distributes 64-bit data blocks to the PCS by Round Robin across a logical CGMII interface of different data widths and clock speed
    - For 4x10G, CGMII is 256 bits and 156.25 MHz clock (single edge)
    - For 10x10G, CGMII is 640 bits and 156.25 MHz clock (single edge)
    - For 4x25G, CGMII is 256 bits and 195.3125 MHz clock (dual edge)
- Another option
  - Virtual lane (ref: Gustlin_01_0107). Define 20 virtual lanes on which data is interleaved based on 64-bit blocks
  - Will need alignment mechanism on virtual lanes, which will be an additional alignment to the PBL alignment
  - For 4x25G PMDs, blocks are interleaved to 20 virtual lanes, and then 10:4 gearbox is used to distribute blocks from 20 virtual lanes to 4 PMDs

PCS

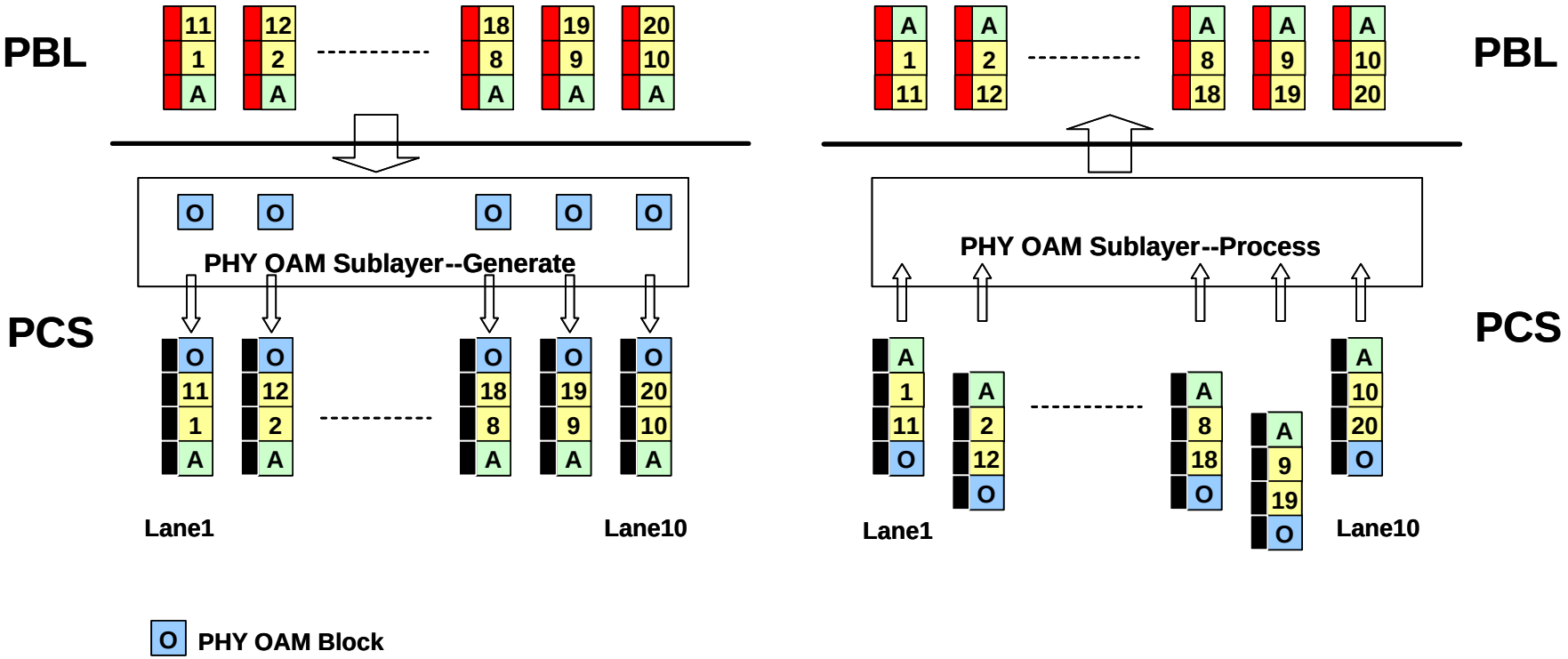
- **Encode and Decode the 64bit block based on the 64B/66B code rule (using pre-code information in the 4-bit control words)**
  - Add 2-bit synchronization to each block on each lane to determine the block boundaries.
- **Scramble and descramble the block**
- **Clock rate compensation**
- **Link status report for each lane (PHY OAM&P)**

PCS function – PHY OAM

- Define a new 64-bit block type for PHY OAM
- PCS layer generates and processes the PHY OAM block which is combined into the data flow
- PHY OAM function—(for each lane)
  - Lane fault reported per lane---(LLF/RLF)
  - BER monitor
  - Lane status monitor and notification
    - Lane delay monitor
    - Lane status notification and negotiation

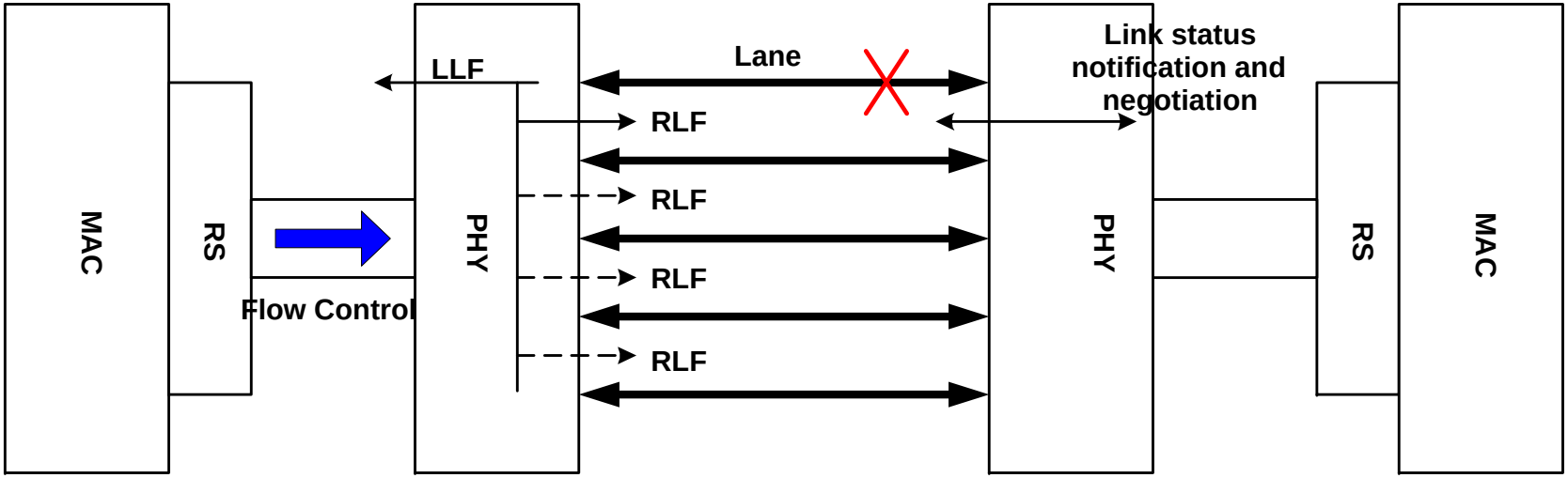


PCS function – PHY OAM



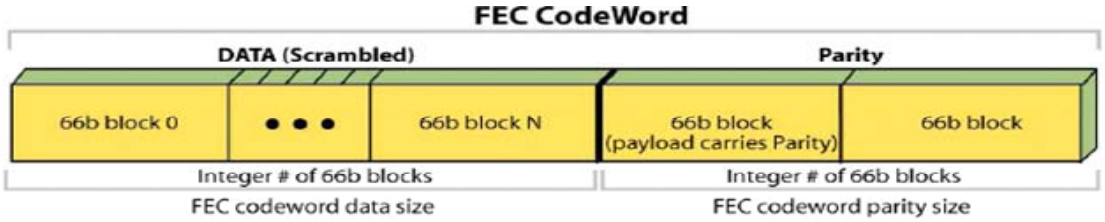
PCS Function – Resiliency

- **Partial Fault Protection**
  - **If there is a Partial Lane failure, the data flow could still be transmitted on the remaining available lanes**
  - **PCS layer monitors for lane failure (LOS, BER, etc.)**
  - **Provide a mechanism to add/delete lanes (negotiation mechanism)**
  - **After the failed lanes recovered, they would be reused in the ETH link**
  - **Flow control in the MAC layer (backpressure, flow control, QoS, etc.) may need consideration**
- **More detailed discussion in malpass_02_0907**

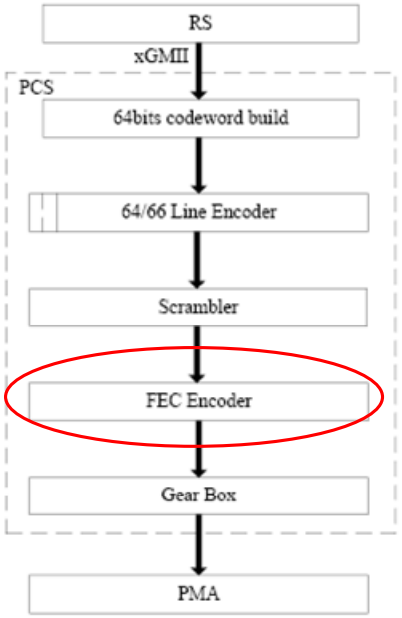


Other PCS Functions

- **other functions in PCS (Open issues)**
  - **For higher speed lanes (25G+), FEC should be supported**
  - **FEC discussions in 10G EPON Task Force**
    - **FEC framing is based on the block**



FEC on 10GEPON



Summary

- **The multi-lane model can support many PMD styles (including serial)**
- **PBL performs distribution and reassembly based on 64-bit blocks**
  - **CGMII interface is based on Nx 64-bit blocks**
  - **Alignment mechanism needs to be added**
- **More PCS functions per lane can be considered**
  - **PHY OAM provides lane status monitor**
  - **Partial fault protection improves the ETH link reliability**

Thank You