Joint ITU-T/IEEE Workshop on Carrier-class Ethernet

Proposed 100GE PCS and Electrical Interface

Gary Nicholl,
Senior Technical Leader
Cisco Systems
The following slides present a proposal for a 100GE PCS and PMA/PMD interface, with the goal of enabling relatively simple optics modules.

Topics include:
- Ethernet Architecture Overview
- Motivation and goals for the PCS
- Overview of the proposed PCS
- Virtual Lane Concept
- PMA (Serdes) Complexity
- 100GE and OTN/FEC
- Summary
IEEE 802.3 Architecture

MAC: Media Access Control
PCS: Physical Coding Sublayer
PMA: Physical Media Attachment
PMD: Physical Media Dependent
MII: Media Independent I/F

Serial Optics Module (i.e. Xenpak)

Geneva, 31 May – 01 June 2007

** Note XFI is not defined by IEEE
Traditional 100G Functional Partitioning

- Same functional split as 10GE
- Electrical interface could be scaled XAUI
- PCS is specific to optics
- Complicated optical module (includes electrical interface deskew and PCS)

**MAC ASSP**
- MAC
- RS
- CGMII
- CGXS

**Optical Module**
- CGXS
- CGMII
- 100GE PCS
- PMA
- PMD
- MDI
- CAUI: 100G AUI
- CGMII: 100G MII
- CGXS: CGMII Extender
Alternative 100G Functional Partitioning

- PCS is generic
- PCS provides coding for both electrical and optical interfaces
- Electrical interface is high speed parallel serdes
- Relatively simple optical module (no PCS)

**MAC ASSP**
- MAC
- RS
- CGMII
- 100GE PCS

**Optical Module**
- PMA
- PMD

CGMII: 100G MII
CTBI: 100G Ten Bit Interface
Goals & Motivations for 100G PCS

- Provide line coding for both the electrical (module) interface as well as the optical interface
- Support a high speed, parallel, serdes-based electrical interface, with in-band deskew
- Allow for a simple optical module
  - No framing, alignment or deskew in module
  - Simple bit muxing serdes (PMA)
- Ensure that the PCS can scale with technology
  - Not optimized for a particular near term PMD
  - Single PCS for all current and many future PMDs (means handling different optical lane combinations, from parallel to serial)
  - Single PCS for current and future module interfaces (means handling different electrical interface bus widths, narrower and faster over time)
Proposed 100GE PCS Overview

- 64B/66B based PCS (same as 10GE, but 10x faster)
- 10 Lane MAC/PCS to PMA/PMD Electrical Interface
  - 100G Ten Bit Interface (CTBI)
  - Each lane runs at 10.3125G
  - Data is inverse-multiplexed across the 10 CTBI lanes
  - In-band deskew mechanism
- Support 1-12 PMD lanes using same electrical Interface and PCS
- PMA maps 10 lane CTBI to n lane PMD
  - PMA is simple bit level muxing
  - Does not need to know or care about PCS coding
- Alignment and skew compensation in Rx PCS only
- Assumes same line coding can be used for the CTBI and PMD
100GE PCS Signal Flow

**CTBI Interface**
- Electrical, Serdes based
- 10 x 10.3125G
- Common for all PMDs
- Supports simple bit mux PMA

**Optics Interface**
- 1 to n lanes
- Lanes can be fibers or wavelengths
- Current IEEE Objectives:
  - 10km on SMF (4 x λ @ 25G)
  - 40km on SMF (4 x λ @ 25G)
  - 100m on MMF (10 x fibers @ 10G)
CTBI – Virtual Lane Concept

- Data from the MAC is first encoded into a continuous stream of 64B66B blocks and scrambled (100G aggregate stream).
- The 100G aggregate stream is inverse-multiplexed into a number of ‘virtual lanes’, based on 64B66B blocks.
- A unique marker (66B word) is added to each virtual lane, at the same time and on a periodic basis.
- The Rx PCS uses these virtual lane markers to identify, deskew and reorder the virtual lanes, and reassembles the 100G aggregate stream.
- Conceptually similar to SONET/SDH and OTN virtual concatenation .. but with some differences.
How Many Virtual Lanes are Needed?

The number of virtual lanes generated is scaled to the Least Common Multiple (LCM) between the 10 lane CTBI interface and the n lane PMD. This ensures that:

- A given virtual lane is always transmitted over the same CTBI and PMD lane combination (follows the same path).
- There is no skew introduced within a virtual lane.
- BUT … different virtual lanes do take different paths, and therefore arrived skewed and out-of-order at the Rx.

<table>
<thead>
<tr>
<th>Number of Electrical Lanes (CTBI)</th>
<th>Supportable PMDs</th>
<th>Virtual Lanes Needed (LCM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1, 2, 3, 4, 5, 6, 8, 10, 12</td>
<td>120</td>
</tr>
<tr>
<td>10</td>
<td>1, 2, 3, 4, 5, 10</td>
<td>60</td>
</tr>
<tr>
<td>10</td>
<td>1, 2, 4, 5, 10</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>1, 2, 5, 10</td>
<td>10</td>
</tr>
</tbody>
</table>
How are Virtual Lanes Generated?

100G Aggregate Stream (of 64B/66B words)

Add 66 bit Marker words periodically (used to identify and realign virtual lanes at Rx)

= 66 bit word
How are VLs mapped to CTBI lanes?

Virtual lanes are assigned to CTBI lanes in pairs, *and bit interleaved.*

CTBI #1 is a simple bit interleave of virtual lane #1 and #2, etc.

*Geneva, 31 May – 01 June 2007*
How are VLs mapped to PMD lanes?

Each CTBI lane contains a bit interleave of 2 x virtual lanes

Each PMD lane contains a bit interleave of 5 x virtual lanes

Every 5\textsuperscript{th} bit on a PMD lane is from the same virtual lane
Complexity of PMA is Very Low

- PMA is a simple 10:4 gearbox for a 4 Lambda Optical I/F
- What is the gearbox?
  - **Tx:** 10 x 1:2 demuxes + 4 x 5:1 muxes
  - **Rx:** 4 x 1:5 demuxes + 10 x 2:1 muxes
Can CTBI be used for 100G OTN?

LAN Application

MAC/PCS

10

CTBI
10x10.312G

WDM Optics (4x25G)

10:4

4x26G

Parallel Optics (10x10.312G)

10x10G

10:10

MAC/PCS

100G OTN/FEC

10

4x28G

DWDM Optics (DQPSK, DPQPSK, etc)

10

CTBI
10x11.095G

Ribbon Fiber
100m on OM3 MMF

4xλ, 10/40km on SMF

Ribbon Fiber
100m on OM3 MMF

1xλ, 80-2000km

Same MAC/PCS

Same 10:4 Gearbox

Geneva, 31 May – 01 June 2007
Summary

- The proposed PCS supports all PMDs (serial and parallel)
- Complexity is reasonable within the PCS
- Complexity in the optical module is very low
- Based on proven 64B/66B framing and scrambling
- Uses a common 10 lane electrical interface for all PMDs
- Electrical interface is very feasible at 10 x 10.3125G
  - XFI and SFI interfaces have proven the serdes
- Allows for a MAC rate of 100.000G
- PCS and alignment overhead very low (~ 3%) and independent of packet size
- CTBI ‘concept’ could be extended to an 100G OTN interface
Backup Material
Alignment Proposal

- Send alignment on a fixed time basis
- Alignment word also identifies virtual lanes
- Sent every 16k 66bit blocks on each virtual lane
  - approx once every 210 us for a 20 virtual lanes
  - Limits max skew, could also make it programmable
- Alignment word interrupts packets
- Takes only 0.006% (61PPM) of the Bandwidth
- Rate Adjust FIFO will delete enough IPG so that the MAC still runs at 100.000G with the CTBI at 10.3125G
  - Looks like a clock difference of 261PPM instead on 200PPM maximum
## Alignment Proposal

### Generic 10GBASE-R Control Block

<table>
<thead>
<tr>
<th>10</th>
<th>Block Type = 0x1e</th>
<th>C₀</th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
<th>C₅</th>
<th>C₆</th>
<th>C₇</th>
</tr>
</thead>
</table>

### Proposed Alignment Word

<table>
<thead>
<tr>
<th>10</th>
<th>Block Type = 0x1e</th>
<th>0x4c</th>
<th>0x55</th>
<th>0x4c</th>
<th>0x55</th>
<th>0x4c</th>
<th>0x55</th>
<th>~VL#</th>
<th>VL#</th>
</tr>
</thead>
</table>

VL# = Virtual Lane Number (0-n)
Open Issues and Next Steps

- Investigate the ‘randomness’ of the aggregate scrambled data after muxing onto both the CTBI and PMD lanes.
- Investigate error conditions
  - Scrambler error multiplication
  - Alignment errors etc
- Investigate the impact of multiple alignment words aligning
  - Can cause long runs of 1’s or 0’s
- Can one PCS really be used for all PMDs?
  - Long haul interfaces might require FEC?
  - If a particular PMD (i.e DWDM) requires a different PCS then the CTBI becomes more like XAUI for that case (an extender layer)
- Investigate complexity of deskew logic in Rx PCS
More Details

Look at:
For additional details.

Thank You!