CEI-28G:
Paving the Way for 100 Gigabit

June 2010

Authors:

John D'Ambrosia, Force10 Networks

David Stauffer, IBM Microelectronics

Chris Cole, Finisar

_____

1. This work represents the opinions of the authors and does not necessarily represent the views of their affiliated organizations or companies.

Table of Contents

1.0 Executive Summary

While the first generation of devices supporting 100 Gigabit per second (Gb/s) will utilize a 10 lane by 10 Gb/s signaling rate, future power and size reduction requirements will drive the need for a faster, narrower interface.   Noting these future needs, the Optical Internetworking Forum has initiated development of standards for the next generation of electrical signaling for long reach and short reach applications.

This paper provides an overview of the CEI-28G project, which defines electrical specifications for 28 Gbaud/s signaling that will be used for next generation chip-to-chip and chip-to-module applications that support transmission of 100 Gb/s data rates, such as and 100 Gigabit Ethernet and OUT-4.

2.0 Introduction

The mission of the Optical Internetworking Forum (OIF) is to foster the development and deployment of interoperable products and services for data switching and routing using optical networking technologies. This requires addressing multiple issues related to optical internetworking at the carrier, system vendor, and component vendor levels. This integrated approach strengthens the OIF's ability to fulfill its mission.

The Physical and Link Layer (PLL) Working Group of the OIF specifies implementation agreements (IA) that not only enhance the interoperability of communication links at the optical level but also at the electrical interface, which includes the electrical properties and protocol, between the different components in a typical optical networking application.

Recognizing the industry's need for the entire component level infrastructure to be upgraded to support higher system capacity, the OIF authorized the PLL Working Group to begin the Common Electrical I/O 25 Gb/s (CEI-25) project. This project defines electrical specifications for up to 28 Gbaud/s signaling for chip-to-chip applications and 25 Gbaud/s signaling for backplane applications. These signaling rates would allow the development of narrower interfaces for 100 Gb/s, such as 100 Gigabit Ethernet. Such interfaces would enable smaller package sizes, lower pin count components, connectors and optical modules, lower power dissipation, and clockless interfaces.

3.0 OIF CEI-28G-SR Electrical Specification Overview

The CEI-25 project builds on the legacy of the CEI effort for the 10-11 Gbaud/s generation, which provided solutions for Short Reach (SR) applications (chip-to-chip and chip-to-module), and Long Reach (LR) applications (backplane). The short reach chip-to-module interface being developed targets a maximum baud rate of 28 Gbaud/s and is designated CEI-28G-SR. This baud rate results from distributing ~112Gb/s OTU-4 rate over four lanes. OTU-4 overhead is added and decoded in the Physical Layer chip, and is not sent across inboard interfaces to link layer devices. Therefore the ~26 Gbaud/s data rate for CEI-25G-LR is sufficient to support backplane interfaces.

Development of the CEI-28G-SR and CEI-25G-LR electrical interfaces has been based on evolutionary improvements in both signaling technology and channel technology. As electrical data rates move from 10 Gbaud/s to 25 Gbaud/s, the increase in power dissipation must be limited to 1.5x per link. While these higher data rates could be achieved entirely through more advanced signal processing in the Serdes device, such a solution would not meet market needs for power and levels of integration. Therefore improvements to signal processing must be evolutionary, not revolutionary, and must be accompanied by evolutionary improvements to circuit board, connector, and backplane technology.

The signaling solution selected by OIF for CEI-28G-SR is "equalized NRZ". Non-Return-to-Zero (NRZ) signaling uses two signal levels to encode '0' and '1' bit values. Equalized NRZ signaling additionally uses equalizer circuits in both the transmitter and receiver to correct for various signal impairments introduced by the channel. A Feed Forward Equalizer (FFE) in typically included in the transmitter, and a Decision Feedback Equalizer (DFE) is included in the receiver. The output of a transmitter containing an FFE is in effect a multi-level signaling code where the transmitted signal amplitude is not only based on the current bit value, but also values of adjacent bits. The coefficient settings for the FFE are optimized for the channel, thereby producing better performance than traditional multi-level signaling schemes using fixed levels.

In addition, the receiver definition typically includes a DFE. This circuit varies the slicing threshold of the receiver based on the history of prior bits received. Work presented in OIF shows good performance at data rates of interest is achieved by cascading a Continuous Time Equalizer (CTE) with the DFE in the receiver. This evolution of the receiver reference model, combined with evolution in channel requirements, is expected to be the basis for both the CEI-28G-SR and CEI-25G-LR electrical interfaces.

4.0 Overview of Applications

Electrical interface standards service two application spaces as shown in Figure 1: the chip-to-module electrical connection between the optics module and the Physical Layer device, and the chip-to-chip connections between link layer Network Processing Elements (NPE) and Physical Layer devices or other NPEs. The CEI 28 Gbaud/s standard targets chip-to-module interfaces and chip-to-chip interfaces on the same circuit board. The CEI 25 Gbaud/s standard targets chip-to-chip interfaces across a backplane.
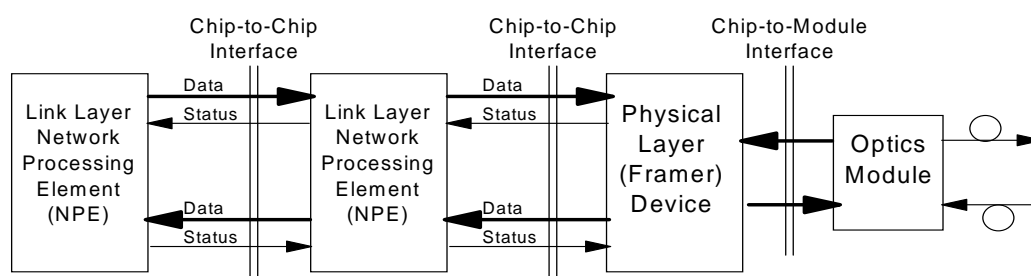
Figure 1 – Chip-to-Module and Chip-to-Chip Applications within a Network System

While several standards do exist for chip-to-chip interfaces, produced both by OIF and by IEEE, this application space is primarily dominated by proprietary interfaces and by de facto standards driven by system vendors. The value proposition for a higher speed interface in this application space is therefore based on the generic assertion that any solution which uses fewer package pins and reduces cost and power is useful. Reducing the number of differential pairs that must be routed in a system reduces the cost of backplanes, connectors, and circuit boards. As throughput density requirements increase for systems, such reduction is necessary avoid having to substantially increase the number of differential pairs being routed. Narrower, higher speed interfaces also allow reduction of the number of chip package pins. As electrical data rates move from 10 Gbaud/s to 25 Gbaud/s, a power dissipation increase of 1.5x per link represents the limit of what the market has historically found to be acceptable. Assuming this target, an

interface constructed with 25 Gbaud/s Serdes devices is more cost effective and has better power utilization per Gigabit than the equivalent interface constructed with 10 Gbaud/s Serdes devices.

Conversely, the chip-to-optics application space requires closer examination. Electrical interfaces between the optics module and the Physical Layer device tend to be shorter and do not cross a backplane. Cost and power of the optics module tends to be the dominant concern for this interface. Evaluation of these metrics is dependent on the optics architecture being targeted, and therefore the value proposition for narrower, higher-speed interfaces must be examined in this context.

Two optics architectures are considered in the following sections. The 4 channel optics architectures being developed in the IEEE 802.3ba 40 Gb/s and 100 Gb/s Ethernet amendment are likely to become the dominant standard for 100 Gb/s networks. Similar 4 channel architectures are also of interest in the transport of 100 Gb/s in telecom networks For example, the OIF 100G DWDM transmission project is also discussing 4 channel architectures. The objective of this project is to aid the industry in the development of transceiver technology for transport of 100G signals in long distance backbone networks.

5.0 100 Gigabit Ethernet[i]

The IEEE 802.3ba 40Gb/s and 100Gb/s Ethernet amendment has defined a number of physical layer specifications. The 100GBASE-LR4 physical layer specification is a WDM solution that targets a distance of at least 10km over single-mode fiber. The 100GBASE-ER4 physical layer specification is a WDM solution that targets a distance of at least 40 km over single-mode fiber. Both of these physical layer specifications are based on four wavelengths at 25 Gbaud/s per wavelength across one single mode fiber in each direction. It is anticipated that this optical solution will also be the basis for the ODU-4 specification for telecom networks that is being explored by the ITU-T Study Group 15 Question 6.

Figure 2 illustrates an architectural layer diagram for the first practical implementation of 100GBASE-LR4. In this scenario the Media Access Control (MAC), Physical Coding Sub-layer (PCS), and a Physical Medium Attachment 20:10 Sub-layer (PMA) reside in one component. This component connects to an optical module containing a PMA 10:4 Sub-layer and the Physical Medium Dependent Sub-layer (PMD).

The physical interconnection between the two PMA sub-layers is provided via the "CAUI" (pronounced "COW-EEE"). The "CAUI" is an optional interface between any two PMA sub-layers, which makes it useful for architectural partitioning in an implementation. It is a retimed interface that is based on 10 lanes. The 64B/66B encoding provided in the PCS sub-layer above the PMA results in an actual data rate per lane of 10.3125 Gbaud/s, which translates to an effective data rate of 10 Gb/s.
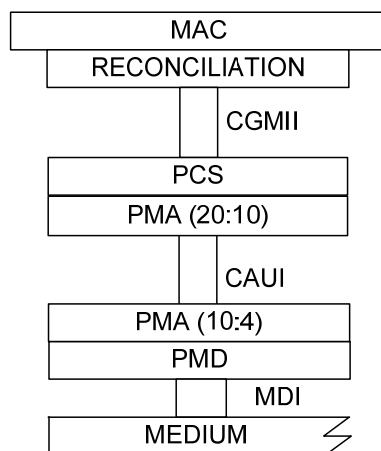
Figure 2 – Potential Implementation for 100GBASE-LR

In the implementation shown in Figure 2, the PMA sub-layer above the CAUI translates the 20 PCS lanes (PCSL) associated with the 100GBASE-R PCS to the 10 physical lanes associated with CAUI. The lower PMA sub-layer then serializes / deserializes the 10 lanes of 10.3125 Gbaud/s associated with the CAUI to the 4 lanes of 25.78125 Gbaud/s (25 Gb/s 64B/66B encoded) associated with the PMD sublayer. Retiming is provided by each PMA sub-layer in the receive path. Figure 3 illustrates a first generation implementation concept of a module intended to support 100GBASE-LR4.
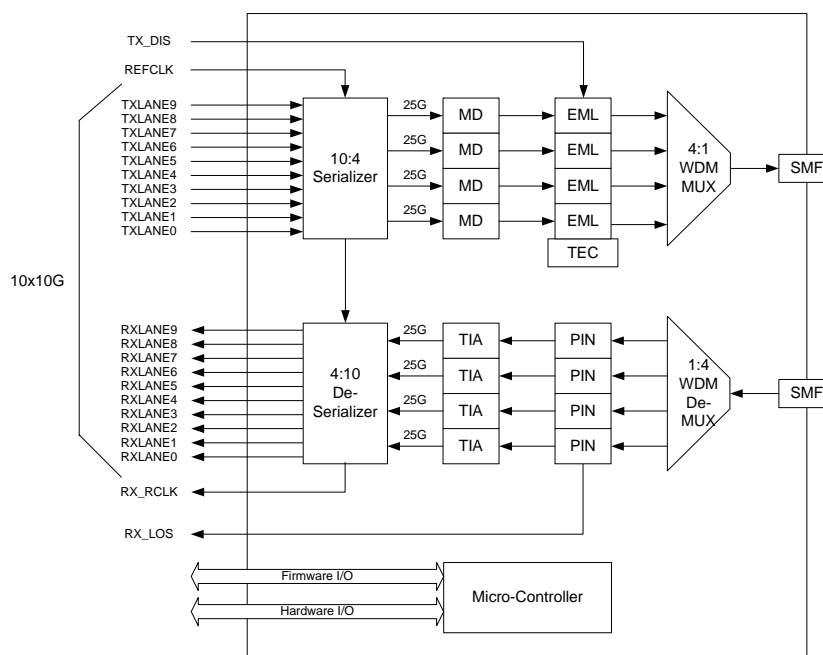
Figure 3- Generation 1 100GBASE-LR4 Implementation Concept

The module utilizes a CAUI interface, which incorporates the 10:4 PMA sub-layer noted above. This translates to an internal 10:4 serializer on the transmit path and a 4:10 deserializer in the receive path of the module. The non-integer ratio of the multiplex stage in the serializer / de-serializer has higher power consumption because the internal circuits would be operating at the 25 Gbaud/s output rate.[ii] The

combination of the higher power consumption and the requirement for a 40-pin interface (20 differential pairs) will result in a larger module footprint.

Figure 4 illustrates a next generation implementation concept of a module intended to support 100GBASE-LR4.[iii] The use of a 25 Gbaud/s based interface would reduce the width and the elimination of the serializer / deserializer in the module will lead to lower power consumption, as well as a smaller module size.
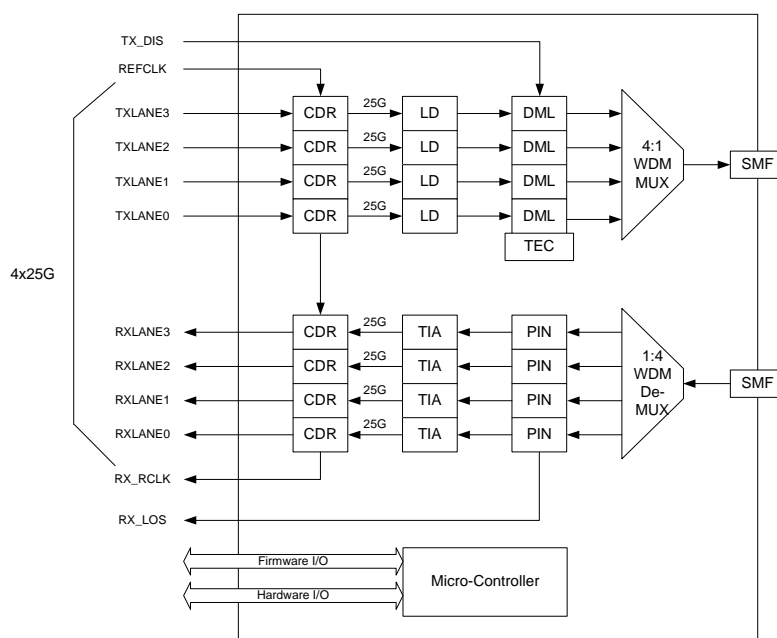

Figure 4- Generation 2 100GBASE-LR4 Implementation Concept

The second generation of 100GBASE-LR4 modules, shown in Figure 5, is already being envisioned. In this module architecture the mux/demux associated with the front stage in the first generation is moved outside of the module to a separate 10:4 SERDES, which resides on the host. These changes will lead to lower cost and power next generation form factor.

The role of CEI-28G-SR in future generations of 100 Gigabit Ethernet is shown in Figure 5. In the second generation of 100GBASE-LR4 modules, an interface, based on CEI-28G, will be used to connect the discrete SERDES to the module, which has a 4:4 PMA sublayer in its front stage to provide retiming. In the third generation the CEI-28G interface has been integrated into the host ASIC, and connects directly to the module with the 4:4 PMA sublayer in its front stage to provide retiming.
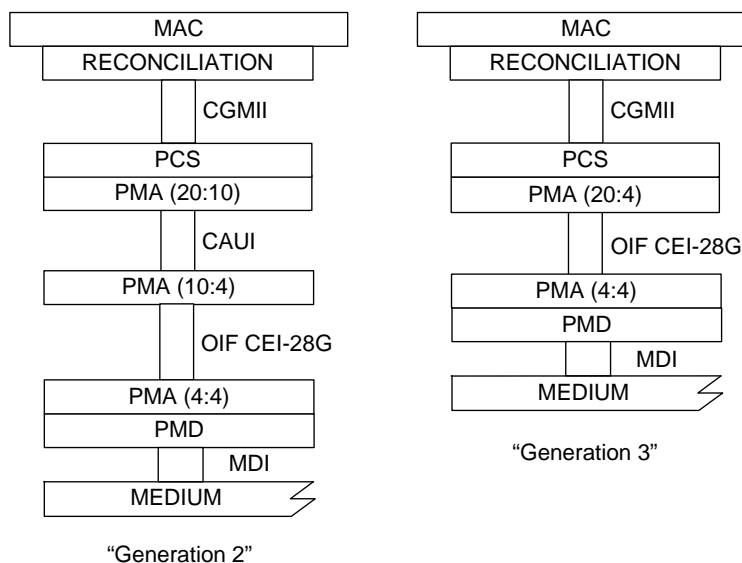
Figure 5 – Potential Next Generation Implementations for 100GBASE-LR

6.0 OIF 100G DWDM Transmission Project

OIF is developing optical standards for 100 Gb/s networks under the 100G DWDM transmission project. This project is developing transceiver technology for long distance networks where a single fiber may contain multiple 100 Gb/s signals. The target optical channel spacing of 50 GHz dictates that the 100 Gb/s signal must be limited to utilizing a single wavelength. OIF has chosen to specify Dual Polarization Quadrature Phase Shift Keying (DP-QPSK) as a modulation method to achieve the requirements of this application space.
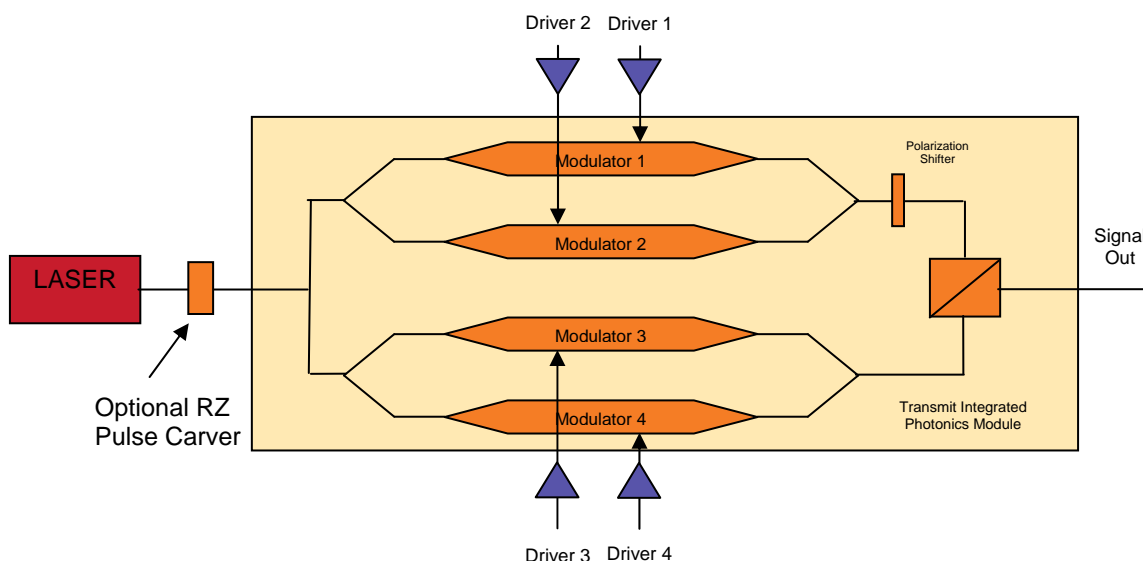

Figure 6 – Block Diagram of DP-QPSK Transmitter Module

First, it should be noted that the DP-QPSK modulation scheme incorporates two polarization lanes, each of which uses two bits to code each QPSK symbol. A block diagram of the transmitter module is shown in Figure 6. The natural chip-to-module interface for this modulation scheme is four bits wide, with one bit driving each modulator. Early module implementations using 10 lane by ~11 Gbaud/s interfaces based on CAUI will require 10:4 multiplexor stages in the optics module transmission path, and 4:10 demultiplexor stages in the receive path. As was the case with Ethernet, it will be desirable to eliminate this logic from the optics module in later generations.

A second concern is that 100G DWDM is targeting long haul applications where the maximum regeneration spacing is in the range of 1000 to 1500 km (with up to 6 ROADMs). To achieve transmission over this distance without increasing the number of ROADMs required, it is necessary to consider FEC coding which is stronger than the standard 7% FEC defined in G.709. Approximately 20% FEC bandwidth overhead is required to achieve performance envisioned for some 100G DWDM systems. Strong FEC codes are not expected to be standardized, and are expected to be implemented in the optics module.

This requirement for FEC in the optics module creates an even greater burden on implementations using CAUI. In an Ethernet system, the CAUI coding can be multiplexed and demultiplexed in the PMA sub-layer without needing to decode the CAUI protocol and align bits between lanes. However, in a 100G DWDM module, any misalignment of bits at the input of the FEC encoder will produce incorrect FEC codewords. Therefore, the CAUI protocol must be decoded and bits must be aligned prior to the 10:4 multiplexor in the optics module. This places an even higher power and packaging burden on the optics module than was the case for Ethernet, and a greater impetus to move to a four lane interface.

Following a similar evolutionary path to Ethernet applications, it is expected that initial 100G DWDM modules will use a 10 lane by ~11 Gbaud/s interface. This interface will likely use a protocol similar to CAUI (but at a higher data rate) in order to maintain commonality with modules designed for Ethernet and ITU-T standards. However, as was the case with Ethernet, this solution is less than optimal and drives power and packaging concerns. Therefore, given the characteristics of the modulation scheme chosen, the natural chip-to-module interface for this modulation scheme is four bits wide.

7.0 Conclusion

The migration of networks to 100 Gb/s will initially be implemented by electrical interfaces that utilize a 10 lane by 10 Gbaud/s signaling rate. Ultimately, power and size limitations associated with such a wide interface will drive the need for a faster, narrower interface. Given the characteristics of the optical solutions being developed by various industry bodies, it is anticipated that the logical width for a narrower electrical interface would be 4 bits wide in each direction. Recognizing this need, the OIF's The Physical and Link Layer (PLL) Working Group has been working on the Common Electrical I/O 25 Gb/s (CEI-25) project, which includes electrical specifications for 28 Gbaud/s signaling for chip-to-chip applications. This work will enable narrower interfaces for 100 Gb/s applications, such as 100 Gigabit Ethernet, which will enable smaller package sizes, lower pin count components, connectors and optical modules, lower power dissipation, and clockless interfaces.

---

[i] This paper is not intended as an overview of the IEEE P802.3ba architecture. Readers wishing to read such an overview are directed to: D'Ambrosia, Law, Nowell, "40 Gigabit Ethernet and 100 Gigabit Ethernet Technology   Overview,   November   2008" http://www.ethernetalliance.org/images/40G_100G_Tech_overview(2).pdf, Ethernet Alliance White Paper, 2008.

[ii] Cole, Allouche, Flens, Huebner, Nguyen, "100GBE Optical LAN Technologies," IEEE Applications & Practice, December 2007.

[iii] Cole, Allouche, Flens, Huebner, Nguyen, "100GBE Optical LAN Technologies," IEEE Applications & Practice, December 2007.