Optical interconnects for computing applications

Mitchell Fields
mitch.h.fields@avagotech.com
Avago Technologies, 350 W. Trimble Road, San Jose, CA 95131

Introduction

Interconnect requirements for computing systems are application dependent. The metrics may include cost, data rate, reach, reliability, temperature range of operation, BER, immunity to EMI, and weight. Compared to optical signaling, a typical advantage of electrical signaling is cost. However, optical technologies can enable more flexibility in system design and architecture, especially when optics are tightly coupled to the host ASIC. Here, two specific technologies that enable tight integration of optics into computing systems are described: MicroPOD™ and LightPeak™ optical components. The base technologies are reviewed as well as the metrics and features of these components that make them appropriate for particular applications. As next-generation computing systems are developed, optical technologies must focus on the appropriate metrics to ensure adoption.

LightPeak and MicroPOD optical modules

LightPeak modules, shown in Figure 1, are dual-lane 10G optical transceivers (a single module includes two independent transmitters and receivers). MicroPOD modules, shown in Figure 2, are twelve-lane 10G optical transmitters and receivers (a single module is either a 12-lane transmitter or 12-lane receiver). Figure 1 and Figure 2 also illustrate typical application examples for LightPeak and MicroPOD, respectively. Lightpeak is applicable to "thin" fabrics for which 10 – 40Gb/s of connectivity is sufficient for the application [1]. MicroPOD can be used for a wide variety of applications but is targeted for those which require 100's or 1000's of Gb/s of connectivity per network element. MicroPOD is most notably employed in IBM's POWER7-IH high-performance computing system, of which the 10-PF BlueWaters system at NCSA has been detailed [2, 3]. In these IBM systems, all off-drawer I/O is optical via eight "hub" modules. Each hub module provides ~6 Tb/s of aggregate optical I/O and is populated with up to 28 MicroPOD Tx/Rx pairs (the BlueWaters system uses ~220,000 transmitter/receiver pairs).

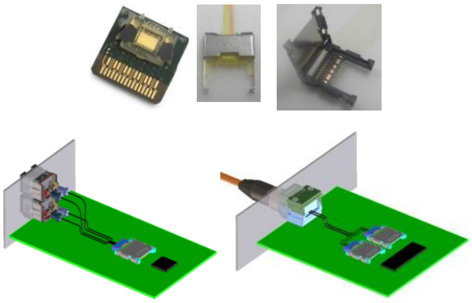

Figure 1: Top row - 2x10G Lightpeak module, optical connector, and socket photographs. Bottom row - example applications showing 20G and 40G aggregate connectivity.

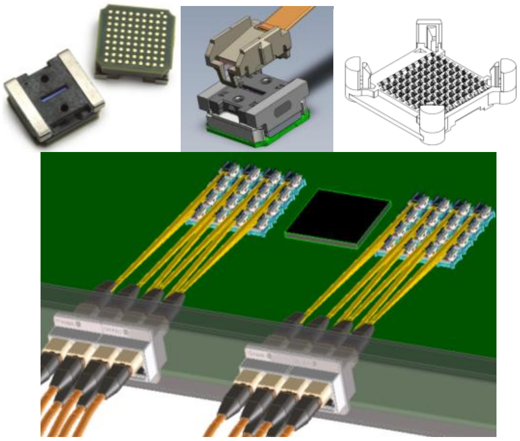

Figure 2: Top row - MicroPOD optical and electrical interface photographs, optical connector, and LGA socket. Bottom - densely tiled configuration highlighting tight ASIC coupling.

A comparison of the technologies is made in Table 1. LightPeak is designed for low cost, power, and profile. MicroPOD is also designed for low cost, but targets systems requiring dense interconnects. The top-attach Prizm connector was specifically designed to enable dense tiling of the MicroPOD near the host ASIC.

Table 1: Details of LightPeak and MicroPOD

Parameter	LightPeak	MicroPOD
Format	2-lane transceiver	12-lane TX and RX
IC technology	CMOS	BiCMOS
Laser source	2xVCSEL array	12xVCSEL array
Receiver	2xPiN array	12xPiN array
Electrical socket	LightPeak	μLGA
Optical connector	LightPeak	Prizm
Footprint (with and w/o socket and optical connector)	12x12x2 16x16x2.3	7.8x8.2x3.9 10.7x11.1x7.1
Power (mW/Gbps)	14	25
Nominal reach (m)	30	100
Diagnostic capability	ID and basic alarms	Full diagnostics
Signal integrity features on electrical interface	None	Equalization and de-emphasis

Performance and capabilities

Many computing applications require performance over a wide temperature range. Robust receiver performance is also mandatory to ensure link integrity. Other important features include capabilities to mitigate jitter in the ASIC/optics electrical channel and module diagnostics.

Figure 3 highlights sample data from LightPeak modules demonstrating transmitter stability over a wide case temperature range as well as good receiver sensitivity. The CMOS IC and low-threshold VCSELs result in low power dissipation per 10 G link (<14 mW/Gbps worst case).

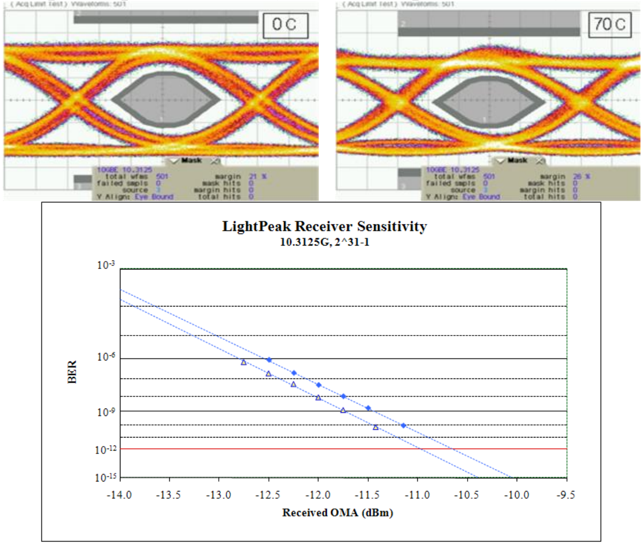

Figure 3: Sample LightPeak module transmitter eye diagrams at 0C and 70C and receiver sensitivity plots.

Figure 4 shows characteristic MicroPOD performance. The MicroPOD ICs contain circuitry specifically designed to target stable performance over temperature. The BiCMOS ICs enable robust transmitter and receiver performance that allow for enhanced link margin. The worst-case power dissipation is 25 mW/Gbps (maximum settings applied to all signal-integrity features). Thermal management is possible via water or air cooling with appropriate heat sinking to the copper side bars of the MicroPOD.

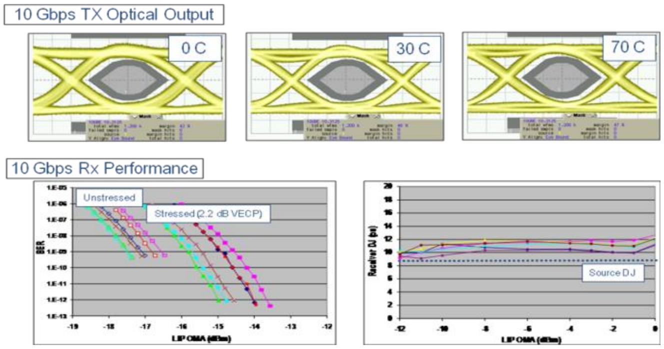

Figure 4: MicroPOD transmitter and receiver performance.

In dense embedded applications with >10G signaling, mitigation of jitter due to skin-effect frequency dependent losses is critical to ensure link margin. The optical components should provide programmable equalization blocks on the electrical inputs of the transmitter modules and programmable de-emphasis blocks on the electrical outputs of the receiver modules. Figures 5 and 6 highlight these signal-integrity features of MicroPOD (receiver output voltage swing is also programmable).

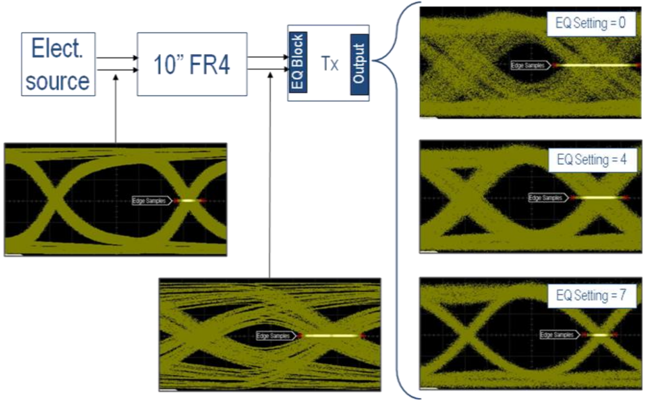

Figure 5: MicroPOD transmitter input equalization capability.

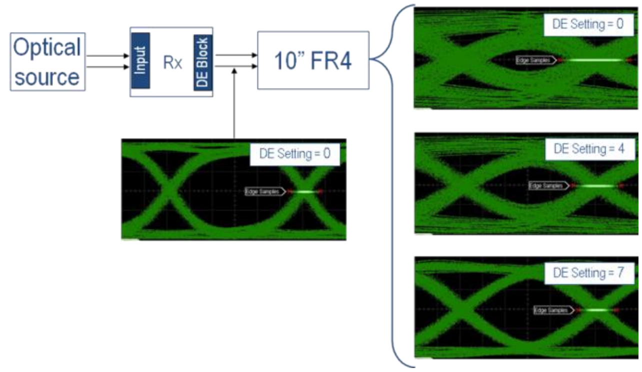

Figure 6: MicroPOD receiver output de-emphasis capability.

Diagnostic capabilities of optical modules in embedded applications are important for real-time monitoring and fault isolation. For example, transmitter power, laser bias current, receiver power, module temperature, voltages, and various alarms based upon these diagnostics are all available in the MicroPOD modules. LightPeak does not provide these diagnostics as the main target application (consumer) does not mandate these features nor could it bear the burden of the cost.

Summary

Electrical signaling will be challenged to meet the needs of next-generation computational systems. However, for optics to successfully displace electronics, the optical components must achieve the requisite metrics.

References

[1] S. Addagatla, M. Shaw, S. Sinha, P. Chandra, A. Varde, and M. Grinkrug, *Direct Network Prototype Leveraging Light Peak Technology*, 18th IEEE Symposium on High Performance Interconnects, 2010.

[2] A. Benner, D. Kuchta, P. Pepeljugoski, R. Budd, G. Hougham, B. Fasano, K. Marston, H. Bagheri, E. Seminaro, H. Xu, D. Meadowcroft, M. Fields, L. McColloch, M. Robinson, F. Miller, R. Kaneshiro, R. Granger, D. Childers, and E. Childers, *Optics for High-Performance Servers and Supercomputers*, Proceedings of OFC, 2010.

[3] M. Fields, J. Foley, R. Kaneshiro, L. McColloch, D. Meadowcroft, F. Miller, S. Nassar, M. Robinson, and H. Xu, *Transceivers and Optical Engines for Computer and Datacenter Interconnects*, Proceedings of OFC, 2010.