100 Gigabit Ethernet Drivers and Technology

White Paper

The world is currently experiencing a dramatic escalation in the demands placed upon its global data communications infrastructure. The increase is driven predominately by a shift from text only traffic with modest bandwidth requirements to full multimedia content with bandwidth requirements orders of magnitude more. To accommodate the ever increasing demand for bandwidth, communication links have evolved from copper based channels running in the MB/s range to fiber optic based links running in excess of 10 Gb/s. In modern data centers for example, 10 Gigabit Ethernet (GBE) ports are commonly used as uplinks for 24- and 48-port 1 GBE switches. The demand for 10 GBE is only expected to increase as switch vendors begin to roll out their next generation platforms with individual port capacities of 10 Gb/s. For these types of systems, it seems logical that any link to an upstream network element requires a throughput much greater than 10 Gb/s. Emerging 100 GBE technologies will provide a path to realizing these high capacity uplinks.

Ethernet is a communications protocol based on a series of IEEE 802.3 standards embedded in software and hardware devices, and is the dominant cabling and low level data delivery technology used in local area networks. First developed in the 1970s, it was published as an open standard by DEC, Intel and Xerox and later described as a formal standard by the IEEE. Ethernet has become the networking communications technology of choice for so many applications largely because of wide availability of reliable and cost effective products. Strict adherence by the vendors to IEEE 802.3 standards has resulted in a very robust protocol, leading to an ever increasing desire to deploy Ethernet further into the end-to-end network.

Ethernet is also beginning to make inroads into the Storage Area Networking (SAN) space. Benefiting from broader data access, better reliability, lower maintenance costs, and improved disk utilization, networked storage is replacing local hard drives traditionally found in servers. Disk I/O is one of the primary bandwidth consumers in servers, and moving the disks to a remote network drive increases the demand for network I/O bandwidth. Internet SCSI (iSCSI) using Gigabit Ethernet networking is well-proven and capable of supporting the performance, reliability and availability needs for critical applications for mid-tier enterprises. The emerging Fiber Channel over Ethernet (FCoE) standard will enable the consolidation of both SAN and Ethernet traffic onto one common network adaptor reducing the ever growing number of adaptors required. The goal of FCoE is to create a converged network where servers would use the same interface for the SAN and LAN, and both networks would use one set of cables and switches.

In order to address the requirements for the next evolution of Ethernet speed, the IEEE 802.3 Higher Speed Study Group (HSSG) convened to discuss possible solutions. In November 2006, the group concluded that the divergence in bandwidth requirements between the networking and computing industries necessitated a bifurcation in base data rates. To service the next generation of Internet backbone and network aggregation points 100 Gb/s Ethernet seemed like the logical choice, while 40 Gb/s Ethernet appeared better suited to match the bandwidth requirements required by computing technologies such as multi-core processing. For the 100 Gb/s Ethernet standard, the group agreed to support reaches of at least 100 m on OM3 (optical multimode 3, laser optimized) multimode fiber and at least 10 km on single-mode fiber. In November 2007, the IEEE 802.3 working group authorized the HSSG to become the IEEE P802.3ba task force whose purpose is to extend the 802.3 protocol to operating speeds of 100 Gb/s and 40 Gb/s while maintaining maximum compatibility with the installed base of 802.3 interfaces. The tentative timeline for this task force calls for delivery of a final set of specifications by 2010. In the meantime, the global demand for bandwidth will be fueled in large part by the following:

- Video sharing, streaming and video on demand
- High Definition video transmission
- Online media content delivery and storage
- Enterprise traffic
- Data exchange between corporate, academic and medical research facilities, and high performance computing installations

Video Streaming / Download

One of the most widely recognized providers of streaming video on the Internet today is YouTube. Best described as a video file sharing provider, YouTube had over 238 million visits in December 2007. Even though YouTube streams its videos using a modest resolution and bit rate (Flash 7 at 250 kb/s), the potential instantaneous aggregate bit rate for YouTube and other sharing providers like MSN, Yahoo, AOL and MySpace is staggering. In addition, many major television and cable networks have begun to offer their programming on an on-demand basis over the Internet in both standard and High Definition formats. In January 2007, Netflix joined this group by introducing a feature which allows customers to view movies using real time playback technology based upon a browser applet. This particular piece of technology requires a minimum download capacity of 1Mb/s to the user, and with a 3 Mb/s connection speed DVD quality resolution is possible. Adding to this, in May of 2008, Netflix announced the availability of a set top box which allows users to watch streamed movies without the need of a PC.

According to ComScore, a Reston Virginia Internet marketing firm, US based companies delivered more than seven billion video streams over the Internet to 126 million unique viewers in March 2007. YouTube alone was responsible for 1.1 billion streams to 55 million unique viewers. Using three minutes as an average video length, and 4 megabytes per minute for low resolution video (just over 500 kb/s) gives an average file size of 12 MB. Delivering 1.1 billion video streams would require about 12.5 PB (Petabytes) of bandwidth for the month. This corresponds to a stack of standard definition DVDs over one mile high.

Other communications service providers are expanding the availability of video on demand, "triple play" (high speed internet, television, and telephone over a single broadband connection), and virtual private networks (VPNs). As demand for these services is only expected to grow over time, content providers are vigorously deploying 10 GBE technologies in an effort to upgrade and expand their data centers.

Media File Storage and Delivery

While similar to the streaming video providers, media file storage and delivery providers differ in that the content they supply ends up residing with the end user. That is to say, the end user receives a copy of the data of interest. Perhaps the most well known of this type of provider is Apple iTunes™. As of August 1, 2007, Apple announced it had sold more then three billion songs through the iTunes online store. In addition to music downloads, iTunes now offers television programs, movies, audio books, and games.

The explosive growth of digital photography has led to the rise of photo storage and sharing websites. In November, 2007 one such site, flickr®, indicated it had received it's two billionth image upload, and that it continues to receive three to five million new photos per day. The bandwidth required to service all of these uploads and downloads will undoubtedly continue to increase as consumers move to ever higher resolution digital cameras.

Enterprise and High Performance Computing

In order to increase performance and reduce costs, data center managers are increasingly turning to server virtualization and clustering technologies. High performance computing (HPC) clusters, which consist of many servers networked together to execute large computing tasks through parallel processing, are steadily gaining in speed due to advancements in server processor technology. As the processing power increases, so too does the need for faster networking I/O within the cluster.

Within the data center, 10 GBE technologies are being deployed at an ever increasing rate. This deployment is being enabled, in part, by the availability of low cost pluggable fiber optic transceiver modules. As 10 GBE grows within the data center, there is a greater need for higher speed switch uplinks for network aggregation. It is widely believed by many of the world's largest data center operators that the switch to switch aggregation capacity of their networks must be increased by at least an order of magnitude to cope with the increasing demand for bandwidth. For servicing this demand, 100 GBE provides a path forward. It is expected that 100 GBE will be deployed not only in switch uplinks within the data center but also in inter-building, inter-campus, and MAN/WAN connectors for enterprise networks.

100 GBE in the Physical Layer

The 100 GBE physical media devices (PMDs) can be constructed utilizing a variety of different technologies. Some of these technologies are commercially available today while others are still in the early development phase. The exact implementation of a100 GBE physical media device will be driven by design constraints such as required link length, power dissipation, form factor, and perhaps most importantly, cost.

Historically, single mode fiber based transceivers have been used to address applications requiring link distances of greater than a couple of hundred meters, while 850 nm based multimode devices addressed the short reach (< 500 m). Technically speaking, single mode solutions are able to satisfy short reach applications, but there is often a substantial difference in cost between single mode and multimode options. For example, in today's market, the purchase price of a 1310 nm single mode 10GBASE-LR XFP transceiver is about two times that of a 850 nm multimode 10GBASE-SR XFP. Given the technologies needed for implementing 100 GBE, the large disparity between single mode and multimode pricing will likely remain.

The most commonly considered way of achieving 100 GB/s optical transmission with today's technology is by breaking the data stream up into several lower speed streams. Over single-mode fiber, this is accomplished by utilizing wavelength division multiplexing (WDM). For transmission over multi-mode fiber, the data is sent in parallel over individual fiber links which are bundled together within a single cable assembly. WDM systems allow for multiple data streams to be carried over the same fiber by using sources of varying wavelength. The sources, usually Fabry Perot (FP) or Distributed Feedback (DFB) lasers are combined using an optical multiplexer element and launched into a single mode fiber. On the receive side, the multi-colored data stream is separated according to wavelength using an optical De-MUX element and routed to an array of discrete receiver blocks.

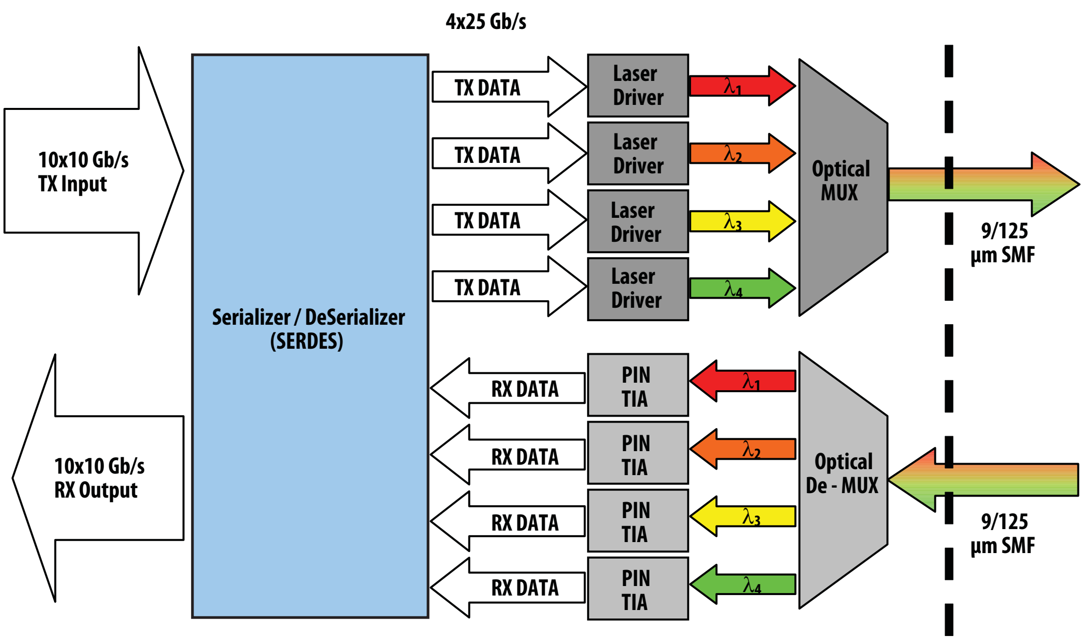


Figure 1. Wavelength Division Multiplexing

Communication ASICS available today are routinely able to operate up to 10 Gb/s. This suggests that in the near term, the electrical interface of a 100 GBE PMD will be parallel in nature with ten lanes running at 10 Gb/s per lane. While it is certainly possible to create a 10-wide WDM solution, a PMD with a lower optical channel count is more desirable when one considers power dissipation, reliability, footprint, and most importantly, cost. With this in mind, the consensus seems to be that a 4x25 Gb/s based WDM transceiver will initially emerge as the preferred SMF solution for long reach applications. The partitioning of the 100 Gb/s data stream into four optical channels of 10 Gb/s is similar to the partitioning called out in the 10GBASE-LX4 standard. The exact wavelength spacing is yet to be finalized but should be sufficiently wide enough to allow for un-cooled lasers.

To be sure, there are still many technical hurdles that need to be overcome before we see large scale deployment of 4x25 Gb/s WDM transceivers. One major piece missing today is the commercial availability of 10:4 serializer / deserializer ICs (SerDes). These components are necessary to translate the ten lanes of 10 Gb/s data into four lanes running at 25 Gb/s. In August of 2008, Avago Technologies announced it had demonstrated operation of a 20 Gb/s SerDes. Unfortunately, commercially available products operating at 25 Gb/s are not yet available. Moreover, while laser sources operating up to 40 Gb/s are available today, they are not packaged in a form suitable for integration into a low cost transceiver. These lasers are typically housed in butterfly type packages with fiber pigtails. What is needed is a direct modulated (DM) or externally modulated laser (EML) capable of being assembled into a compact transmitter optical subassembly (TOSA). To date, the fastest commercially available compact TOSA are specified to around 10 Gb/s.

In contrast to the WDM option, the technology required to achieve cost effective 100 GBE links via multimode parallel optics is available today. In fact, 100 Gb/s operation of parallel optics modules in a "real world" setting has already been publicly demonstrated. Utilizing prototype 12-channel parallel optic transmitter and receiver pairs supplied by Avago Technologies, Ixia and Infinera Corporation demonstrated a complete 100G Ethernet fiber optic link. In the demonstration, a prototype Ixia 100 GbE test set generated a pre-standard 100 GbE signal, which was handed off to a prototype 100 GbE interface on the Infinera DTN DWDM transport system via Avago's 100 GbE multimode parallel optic technology. The 100 GbE signal was then switched and transported by the DTN through XO Communications' long-haul DWDM network to Los Angeles, where it was looped back and transported back to the Infinera booth in Las Vegas.

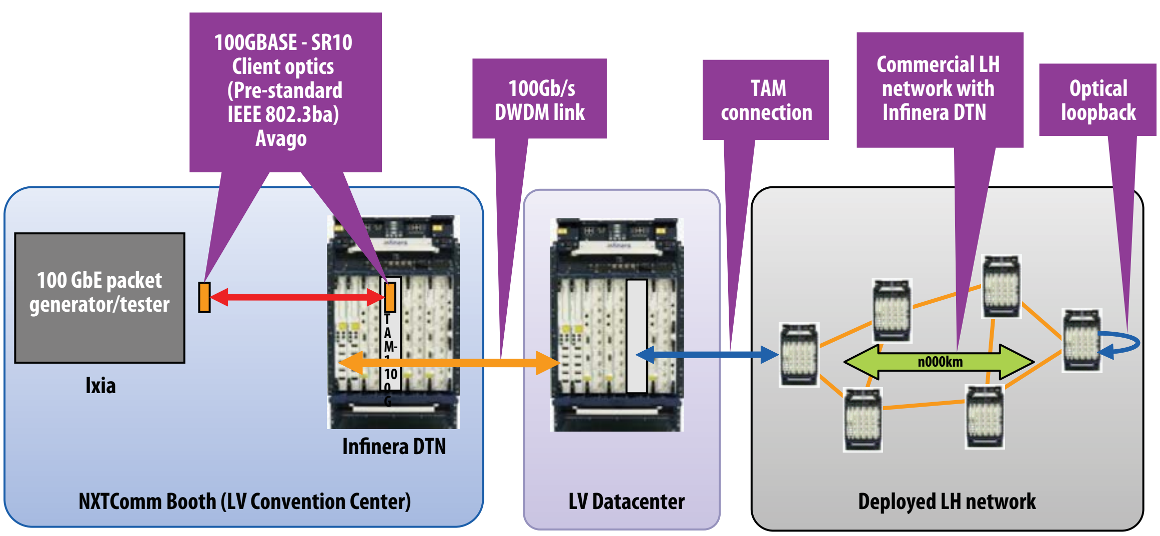


Figure 2. 100 GBE Demonstration

This demonstration was not completely optimized for 100 Gb/s operation, but the Avago Technologies' prototype 12-channel parallel optic modules show encouraging performance at 10 Gb/s per channel, able to deliver a 0.4 UI eye opening at the RX output after 50 m of OM3 fiber.

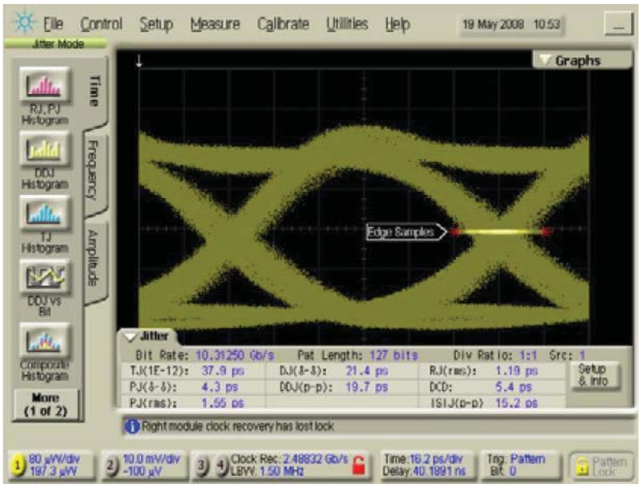


Figure 3. Typical parallel transmitter output at 10.3125 Gb/s

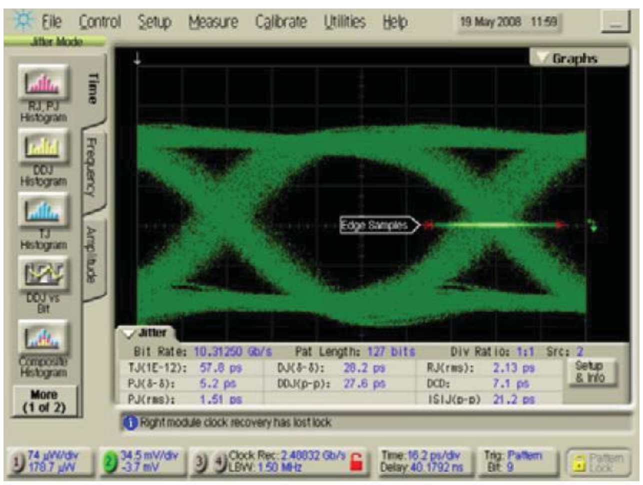
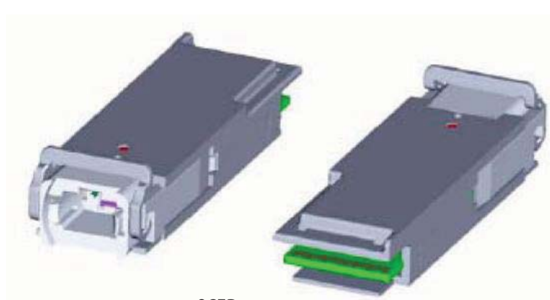


Figure 4. Typical parallel receiver output at 10.3125 Gb/s

Parallel fiber modules, hereafter referred to as parallel optics, use fiber ribbon cables which commonly contain either 12 or 24 individual fibers terminated by an MTP (or MPO) connector. Today, parallel optics modules are available in variety of configurations and form factors which are specified in a number of multi-source agreements (MSA), such as SNAP-12, POP4, and PPOD and QSFP. SNAP-12 specifies separate 12-channel transmit and receive Z-axis pluggable modules, while POP4 describes a 4-channel Z-pluggable transceiver. PPOD is similar to the SNAP-12, but specifies a higher data rate, typically 5-6 Gb/s. Like POP4, QSFP specifies a pluggable transceiver but with a different electrical connector.

Avago Technologies offers a complete family of multimode parallel optic solutions based upon internally developed 850 nm based VCSELs, PINs and driver ICs with a maximum aggregate bandwidth of up to 72 Gb/s (12 channels @ 6 Gb/s). Extending this bandwidth to 100 Gb/s should be relatively straightforward.

SNAP-12, POP4, PPOD



QSFP

Figure 5. Parallel optic module form factors

The primary building blocks for a parallel optics based 100 GBE solution are as follows: multi-channel laser driver IC, multi-channel transimpedance amplifier / post amplifier IC, PIN diode array, and vertical cavity surface emitting laser array (VCSEL). For a 10x10 Gb/s module, these building blocks will be based upon 10G VCSEL, PIN, and IC technologies which have a proven track record of performance and reliability in single channel applications. The exact form factor a 100 GBE short reach solution will eventually adopt is yet to be determined. However, given the desire of equipment vendors to provide a "pay as you go" approach to bandwidth expansion, it will most likely be some type of pluggable module (or modules if separate TX and RX modules are considered). Front panel pluggability offers the advantage of ease of insertion and extraction, but usually at the expense of signal integrity, for example the fiber optic module must sit farther from the host ASIC. Mid-board mount solutions, such as SNAP-12 / PPOD, allow the optics to be placed closer to the ASIC but then cable management from the module to an external bulkhead connector becomes an issue.

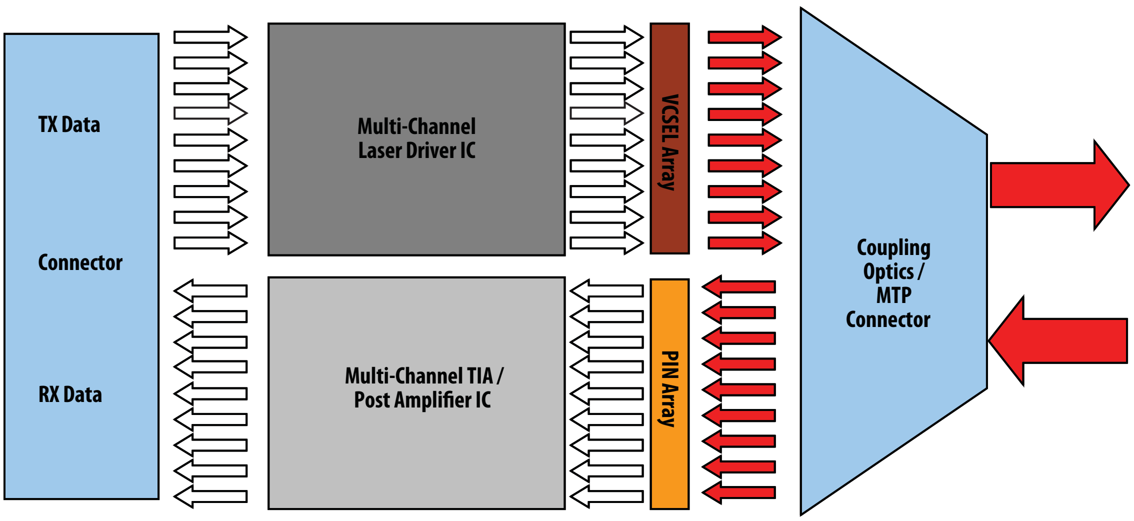


Figure 6. Parallel optics transceiver

There is no doubt that the exploding demand for bandwidth will drive the development of faster and lower priced PMDs. In the short term, and looking forward, parallel optic solutions will be instrumental in realizing low cost, high capacity, short reach 100 GBE links.

For product information and a complete list of distributors, please go to our web site: **www.avagotech.com**

Avago, Avago Technologies, and the A logo are trademarks of Avago Technologies in the United States and other countries.
Data subject to change. Copyright © 2005-2010 Avago Technologies. All rights reserved.
AV02-1660EN - March 25, 2010