On-Board Optical Interconnection
CTR III TWG Report #3
MIT Microphotonics Center, April 2013

Digest

Executive Summary

Optical interconnects have become the established rack interconnection in NEBS and ETSI communications systems, in carrier grade data centers, high performance computing, and high speed packet and circuit switching applications. They are also emerging as the favored solution for board-board communications within equipment racks as pervasive demand for bandwidth increases. The next application evolution for optical interconnects will be on-board, inter-module communications. As bandwidth demands continue to increase, the challenges of scaling cost, density, power efficiency, and thermal control will require optimization with both component and system integration. As yet, no dominant designs for embedded or pluggable, on-board optical I/O have emerged despite advances in optical cables for consumer electronics, peripheral interconnects and embedded optics for high performance computing systems.

The purpose of this report is i) to assess the viability and maturity of the key elements required to implement on-board embedded optical interconnects, and ii) to identify gaps and opportunities to be addressed by the R&D community to achieve on-board optical solutions. We target a notional “Optical PCI Express 3.0 bus” for this assessment to represent a leading high performance board-level interconnect bus presently used in personal computers and servers systems. (PCIe is a registered trademark of PCI-SIG.) A scalable, on-board optical solution has the potential to revolutionize server design. Two features make PCIe 3.0 compelling: i) it represents a possible solution for the pin-out limitations emerging in the fixed point-to-point links in high performance CMOS ASICs, and ii) the PCIe architecture has the flexibility needed to address pluggable peripheral functions such as accelerated processors, InfiniBand host channel adapters, SSD arrays and other demanding I/O. As such, we believe the Optical PCI Express 3.0 **O-PCle** bus provides an appropriate target for understanding how board level optical interconnects can scale in computing applications. An effective implementation of O-PCle can provide enormous system benefits for designers and users of computing platforms based upon its architectural flexibility, enhanced modularity and scalability of bit rate and bus width and distance.
**Key Points**

- The requirements for on-board optical I/O are significantly different than for longer reach board-to-board and rack-rack applications. These attributes include compactness, modularity, reduced power consumption and effective component and manufacturing integration. The expected higher component count, as optical interconnection penetrates closer to the processor and switch, implies strong cost constraints that must compete with incumbent cost/connection metrics at each level of the interconnection hierarchy.

- Parallelism-driven architectural changes from tiered hierarchies to non-blocking, high connectivity networks that put the communication bottleneck at network capacity rather than processor access are driving both port count and bandwidth requirements.

- Packaging and test dominate the cost of short reach optical interconnection. Integration and high bandwidth density partially compensate for cost, but aggressive packaging and chip-level test cost reduction initiatives are needed in the industry.

- High Performance Computing has been a premium entry point for backplane and inter-board optical interconnection, but the extreme demands of this implementation has not anticipated commodity adoption at the board level. Bandwidth partitioning, port modularity, protocol flexibility and connector implementations must all be re-engineered to meet more aggressive commodity cost points.

- The baseline O-PCIe performance is i) parallel lanes of no less than 10 Gb/s and more likely 25 Gb/s each, and ii) distances from 0.1 meters to 3 meters. Parallelism scaling form 2 to 128 lanes is expected with data rate scaling from 8 to 25 Gbps user rates. Scaling to 16 Gbps is expected in the next generation.

- Most of the base (discrete) optical and terminal electronic building blocks are available and sufficiently mature, but they have not been assembled into forms suitable for scalable on-board interconnects. This status is true for both VCSEL and silicon microphotonic based solutions. We anticipate significant synergies for electronic-photonic integration.

- Silicon microphotonic devices for the modulation, detection and guiding of light meet or exceed currently envisioned needs for integrated solutions. Integration of the light source into this platform will significantly reduce packaging complexity and cost.

- The candidate solutions tend to group compatible technologies together into sets. For example, multimode, VCSEL-based solutions tend to include large core diameter multimode fibers/waveguides and large area photodetectors. Silicon microphotonic solutions tend to include small-core, step-index, single-
mode waveguides with tight coupling tolerances to small waveguide-integrated photodetectors. Electronics are optimized for low power and high sensitivity. Both sets may embrace coarse WDM (CWDM) and multilevel signaling formats in the future, but only the single mode solutions can realistically embrace dense WDM and ultra-high spectral efficiency.

- While single channel, multimode VCSEL and multimode fiber connections meet cost and energy efficiency factors, WDM connections will soon be required to meet port count and bandwidth density scaling. Critical, strategic decisions for packaging R&D and standardization will follow investment in solutions compatible with both single channel and WDM interconnects.

- Multi-core fiber shows promising performance in loss and reduced crosstalk and compact coupling with photonic integrated circuits with high optical port density. Low manufacturing cost and a connector technology will be needed for competitiveness with ribbon-based fiber arrays.

- High port count architecture drives attention to protocol transparency and seamless optical interconnection through all levels of the traditional interconnection hierarchy. Interfaces and link power management must be driven by specifications at the system level.

- Key component challenge: system level specifications that define re-matable interfaces, modular compact optical connector configurations and appropriate cable designs. These targets are worthy of ongoing R&D investment.

- Key integration challenge: packaging and assembly designs that i) are low cost and ii) minimally impact CMOS chip-cooling and electrical I/O requirements, while offering iii) low power and flexible optical access. Single mode solutions currently suffer from cost challenges due to tight alignment tolerances.
Objectives

The objective of this assessment is to evaluate the relative maturity and completeness of a variety of solutions for a strawman optical PCIe 3.0 (O-PCIe) bus. PCIe is a standard used to interconnect CPU chips, controller chipsets, APUs, GPUs, and high speed peripheral I/O used in most modern server and computing systems. It is enormously capable, robust and mature. O-PCIe is adopted here as a proxy for a suite of similar buses (see Figure 1 below), including Hypertransport, Rapid I/O, derivatives such as HMC AMC and CompactPCI, Thunderbolt and others. As a proxy, we add memory interconnect as a possible application for our O-PCIe. By being concrete, we can identify firm requirements and then assess various optical solutions against those requirements. We focus on the underlying data-carrying performance of the solutions, and we, thus, de-emphasize the constraints imposed by protocols and higher level logic and functional compatibility.

![Generic Compute System Diagram](image)

Figure 1: On-Board interconnects for a generic computation system.

The main O-PCIe performance requirements are the following.

- **User data rate per lane:** 8-16 Gbps (PCIe3.0 and PCIe4.0)
- **Number of bidirectional lanes:** 1 to 128
- **Distance:** 0.1 to 1 meters
- **Bandwidth Density:** >2 Gbps/mm
- **Manufacturability:** Technology is available for manufacturing.

This analysis does not directly address field-pluggable PCI card interfaces. Rather the
focus is on those applications with static embedded buses among high speed memory, switches, controllers and CPU chips within servers and communication systems. A key reason for this focus is that PCIe implements point-to-point links that are easy to construct in a static connectivity mode for both electrical and optical implementations, but are currently very difficult to construct for swappable-card slot, optical implementations. There is no technical solution for a low loss, low cost pluggable multidrop optical bus as yet, although HP’s hollow-waveguide approach is solid progress in that direction.
Motivation

**Power:** The Cloud Computing market is expected to grow from $16.5 Billion in 2009 at a CAGR of 27% through 2014. Cost and energy consumption are limiting in-house data center operations. In the time frame from 2010-2014 the number of servers deployed in cloud applications is expected to triple to 1.35 million.

Energy is the single largest part of the data center cost-of-ownership equation. Data center capacity is no longer limited by floor space. Instead, servers consume more power than the data center can supply, or generate more heat than data centers can cool.

![Evolution of Optical interconnects](image)

**Figure 2:** From M. Taubenblatt, IBM, July, 2012.

**Pin count:** Today’s state of the art server and PC chipset packages are limited to about 4000 pins, operating at near the internal clock speed of the CMOS chips. As Scaling forward, multicore processing within the package will attain greater system-on-chip functionality, and the associated increase in pin count demand will be critically constrained by surface area, heat dissipation, RF interference and data rate. Optical I/O may alleviate several of these constraints and enable additional generations of scaling. Transmission line capacitance increases the fraction of system power dissipation of off-chip electrical drivers. New approaches are needed to continue scaling bandwidth and maintain or increase distance.

Ultra-short reach optical interconnects are expected to deliver enhanced capability over electrical-only solutions. These attributes include:
• Energy efficiency
  o Energy per bit at a given distance
  o Lower distortion and crosstalk
• Lower cost per Gb/s
  o Lower cost-of-ownership
  o Retrofits, upgrades and infrastructure reuse
  o Integrate-ability/Backwards Compatibility
  o Fewer parts and interfaces
• Architectural flexibility
  o Interface simplification
  o Layout flexibility
  o Modularity
• Form Factor
  o Pinout density
  o Plug density
  o Stackability

Microphotonic technology has recently achieved milestones that now enable it to be considered as a complete solution for the delivery of the above attributes. Electronic interconnect scaling has benefited enormously from advances in adaptive compensation and regeneration to mitigate impairments arising from impedance discontinuities and crosstalk at package boundaries and along parallel transmission lines.

Once a printed circuit board (PCB) has adopted optical I/O at the chip package level, there are compelling reasons to extend it across the board. Because optical link performance and costs are largely independent of distance, optical buses within PCBs will provide architectural freedom for designers to place components and functions where other constraints are more easily managed. Multichannel, shared memory, for example, can be placed where it can be more easily accessed by multiple sockets and remote processors.
Figures of Merit

The solutions identified will be evaluated against the following Figures of Merit. These presume that the basic performance and functional requirements discussed above are met.

- **Bandwidth Density** measured either in Gbps/mm or Gbps/mm$^2$, depending on the configuration.
- **End-to-end energy** per user bit, measured in pJ/bit or its equivalent Gbps/mW.
- **Cost** is difficult to quantify, because it depends on system level benefits. It can be assessed in terms of how well it uses low cost high speed automated manufacturing, simple robust athermal designs with minimal parts, standard equipment and design tools. This ‘gateway analysis’ is particularly important at the packaging and test levels, where costs have traditionally been high.
- **Latency**: The point to point latency must be minimized in this clock-cycle limited I/O environment. This minimization means that additional time division multiplexing or demultiplexing and complex error correction, especially FEC, will be unavailable. Latency places heightened expectations on signal fidelity and bit rates.
- **Integrate-ability**: The historical benefits of integration favorably affect the yield, reliability, cost, energy and bandwidth density metric performance.
- **Scaling**:
  - **Time** domain (bit rate)
  - **Space** domain (scaling lane count to 128 lanes and higher)
- **Spectral** domain (number of wavelengths, spatial modes) to increase bandwidth density
- Potential to support longer distance

- **Functionality**: The ability of the technology to perform functions with reduced energy and footprint and to scale to future needs such as broadcast and switching.
Key Technologies

Silicon Nanowire Waveguides and Devices
A silicon nanowire waveguide typically has a height of 220 nm set by SOI wafers and a width of about 480 nm depending on desired single-mode operation, scattering loss and birefringence. TE polarization is often chosen to carry data information in the waveguide. The fundamental advantages of silicon nanowire for waveguiding and other on-chip functions are (i) high index contrast and (ii) compatibility with CMOS processes. The high index contrast is critical for minimizing the device size and increasing birefringence in the waveguide to avoid polarization conversion.

A high index contrast in silicon nanowire waveguides allows for submicron light confinement and small bending radius on the order of 2 µm. This makes it possible to achieve a free spectral range (FSR) of tens of nanometers using a silicon nanowire microring resonator, with reasonably small bending loss.

Another benefit of having a high index contrast is to produce a large birefringence in the nanowire waveguides by making the waveguide height and width slightly different. This can effectively reduce the conversion of polarization states to each other during propagation over the waveguide. Polarization diversity circuit can be used to handle and process data signals carried by two polarization states simultaneously.

On the other hand, a high index contrast may in turn cause some problems or, more strictly speaking, make some problems more severe, which have to be carefully addressed in device design and fabrication.

One problem is that the propagation loss is sensitive to sidewall roughness that is introduced in device fabrication. Currently, a loss of 1.7 dB/cm is achievable, which is dominated by scattering loss. The loss can be reduced by producing smoother sidewalls of the nanowire waveguides with a smaller standard deviation of roughness.

Another potential issue is that the high index contrast can make the effective refractive index of nanowire waveguides more sensitive to the dimension inaccuracies caused by fabrication imperfection. Post-processing would be required to trim the waveguide and correct the effective index to the designed value, which otherwise may cause a significant wavelength drift of the resonance wavelength when a ring resonator is formed. Demonstrated trimming schemes include (i) a local heater placed very close to the resonator to thermo-optically induce an effective index change and (ii) a so-called adiabatic ring with a low doping level for thermal tuning. These trimming schemes cause additional power consumption to the system, while the second trimming technology is demonstrated to be much more efficient than the first one.

Note that the effective index offset is a fundamental issue to not only microresonators but also any interference-based devices such as Mach-Zehnder structures and arrayed waveguide gratings (AWGs). Nevertheless, a resonator may suffer from it more, since light travels many times around the cavity. A general method to mitigate the issue is to use materials with a smaller material refractive index, such as silicon nitride.
In summary, silicon submicron nanowire waveguides with a high index contrast is good for the realization of miniaturized optical circuits integrated with electronics in pace with, high yield requirements, large volume manufacturing, and low costs. However, this technology is sensitive to material and fabrication parameters such as index, waveguide widths, layer thickness, roughness, etc. and these limit performances.

**Waveguide-integrated Photodetectors**

**Role:** Waveguide-integrated germanium photodetectors are making in increasing impact in commercial photonic devices. Fifteen years of R&D have established reliable processes to deposit high quality Ge directly on Si. CMOS-integrated Ge photodetectors are performing as well or better than discrete III-V based photodetector receiver and transceiver systems. Due to the ease of integration, Ge photodetectors are widely used in receivers, in transceivers and to monitor laser power.

**Status:** In a CMOS process flow, Ge is directly grown on Si using either a one or two step growth process. The bottom Si substrate that is used as a single crystal template for Ge is generally patterned as a waveguide and utilized to evanescently couple the light into the Ge detector. Alternatively, the light can be butt-coupled into the Ge detector by using a multilevel waveguide configuration. High yield and reliable performance of 20 Gb/s and responsivities reaching 1 A/W can be achieved. Recently, 40 Gb/s data rates have been reached with little impact on yield and responsivity.

![Figure 4: Bandwidth efficiency products for high performance waveguide integrated Ge detectors. J.F. Liu, Handbook of Si Photonics](image)

Figure 4 shows the bandwidth-efficiency product vs. device area for some high performance Ge detectors. Due to the ease of integration with trans-impedance amplifiers, waveguide-integrated Ge detectors are being used in a growing number of applications. Foundry services for photonic devices are already offering Ge detector modules with guaranteed bandwidth performance of 20 Gbps in their design kit (e.g., OpSys).

Currently the germanium for all Ge detectors is epitaxially grown at the transistor level due the temperature requirements for the fabrication of the detectors. In the
long term, it is desirable to limit the thermal budget to ~450°C in order to facilitate device fabrication for back end, monolithic optical interconnect layers. The limitation of the thermal budget puts severe restrictions on fabrication. First, crystalline Si for Ge epitaxy is not available in the back end metal interconnect stack. Furthermore, doped poly-Si is not available in the back end, and electrical contacts to the Ge devices become a challenge. Several research groups have started to develop back-end processes for Ge detectors. Physical vapor deposition of Ge has been used to demonstrate deposition of mono-crystalline epitaxial layers of Ge on Si at 300°C. Single crystal Ge on an amorphous Si (a-Si) seed layer, using a CVD process has also been demonstrated. The challenge of reliable electrical contacts at these low processing temperatures has not been addressed widely enough to provide a commercially viable solution. The performance of MSM detectors, for instance, suffers from high dark current. Further research is needed to provide high performance Ge detectors for back-end processes.

**Silicon OEICs**

Silicon microphotonics is perhaps the fastest growing technology segment in short reach interconnects. Several industrial foundries are issuing process development kits which will enable a second generation of device designers and transceiver makers to participate in the technology. Silicon optical modulators and receivers are potentially well suited to CMOS integration especially in silicon-carrier-chip and system-on-chip platforms. Today, silicon optoelectronic integrated circuits (Si-OEICs) are used exclusively in active optical cables. This application is not substantively differentiated from VCSEL-based optical links except at distances beyond 300 meters where the single mode performance of Si-OEICs on SMF are clear advantages. Si-OEICs promise to use less electrical power than VCSELs, because their small size and integration with silicon CMOS electronics introduces much lower parasitic capacitance. Si-OEICs also promise to leverage WDM and higher order modulation formats to vastly increase bandwidth density. However, no monolithically integrated light source is available in current technology, so the OEIC must couple to a direct bandgap semiconductor for operation as an optical link. Several solutions have emerged (wafer bonding, hybrid assembly and external coupling of III-V laser sources) but none have established a dominant position.

**Status:** Silicon OEICs using free carrier plasma Mach-Zehnder (MZ) and Franz-Keldysh effect electro-absorption (EA) modulation have emerged from the laboratory and are in active optical cable (AOC) products. Today, the dominant design employs four parallel lanes of modulation using a single shared light source. Data rates of 10 Gb/s per lane are in production with 4x28 Gb/s products announced. Arrays of 12 are certainly feasible for an optical PCI application. WDM with four channels has been demonstrated, and advanced modulation formats beyond simple on-off keying have achieved spectral efficiencies of 8b/Hz. Very intense R&D is occurring worldwide on this system, and new results are announced almost monthly.

**Challenges and Future:** Because of emerging standards for 100 Gb/s Ethernet, the first large scale use of Si-OEICs is expected to serve that market in direct competition with VCSELs using 4x28 Gb/s technology. Si-OEICs will also enable the transition of data center networks from specialized Multimode structured cabling to a longer-lived single
mode infrastructure. Because SiOEICs operate well both at extremely short distances (<1m) and beyond 300 meters, we anticipate that Si-OEICs will be ubiquitous in applications with a wide range of distance needs. WDM will enable performance scaling, and the technology is well suited to leverage the spatial multiplexing benefits of emerging multicore optical fiber. The linearity of the modulators suggests that advanced modulation formats such PM-DQPSK with coherent detection may offer yet higher bandwidth density and power efficiency.

**VCSELS**

**Role:** For more than a decade, VCSELS have been the dominant modulated laser source for short reach interconnects. They are directly modulated at low current levels and deliver optical signals on laser-optimized fiber over short distances with minimal distortion. They are easy to assemble with conventional multimode fiber, and they have a well developed ecosystem of cables, connectors, detectors and drive electronics. The basic modulation rates needed for board level interconnection are commonly available.

**Status:** VCSELS are available in a range of wavelengths from the dominant 850nm range to the high performance 980-1060nm range and in the long-reach, single-mode 1310 and 1550 range. VCSEL arrays of 12 units operating at 10-14 Gb/s are commercially available, and arrays of 4 units at 28 Gbps will soon be in the marketplace. Above 32 Gbps VCSELs exhibit lifetime limitations due to the high operating current density. Today, VCSELs have significant package and test cost advantages. VCSELs can be assembled with passive alignment technique, and they are tested in wafer form without need to cleave or polish a crystal edge. For enhanced performance, they can be engineered with matched pre-distortion circuitry and with dispersion compensation. These electrical enhancements enable more robust performance and improved manufacturing yield. Over 10 million VCSELs are sold every year, so an amortized manufacturing infrastructure gives them a cost advantage for short distance applications.

**Challenges for the Future:** Directly modulated VCSELs have proven remarkably resilient to displacement by competing technologies. Engineering improved material properties has enabled increased reliability and modulation rates. However, it has been difficult to achieve the very highest modulation speeds in direct modulation modes, prompting work on external modulated approaches for speeds at and above 50 Gbps. To minimize current density, VCSELs operate in multiple spatial modes and thus require MMF for transport. An additional limitation is their broad spectral width under direct modulation which couples with fiber chromatic and modal dispersion to limit useful link distance, especially in data-center-scale networks. MMF has a limited bandwidth-distance figure of merit, but this concern has not yet become a limitation for board level interconnects.

**Polymer Waveguides**

Multi-mode polymer waveguide technology is the leading technology candidate for commodity optical interconnection applications at the board level. Manufacturing compatibility is inherent for self-supporting flexible waveguides containing films that can be configured-on or associated-with substrates for short reach interconnectivity.
and functionality at the board level. Critical attributes provide a conceptual basis for widely implementing polymer waveguides as optical links and functional components. Optical building block configurations, connectivity, and functional components impact system performance and exploratory design capabilities, such as single mode waveguides. Generic solutions that addressed in the text are:

- **Polymer Waveguide Fabrication**: Generic review of the methods for waveguide formation; creating ridges and/or trenches; using monomer diffusion.

- **Application Embodiments**: Connectivity for waveguide arrays covering MT style connections, I/O mirrors, chip and lens coupling, and resulting configurability. Projections of system performance based on: materials, side walls and ROC, I/O mirrors, and fiber coupling.

- **Configurability for Practical Application Functions**: Diverse design options for practical solutions are reviewed: density and mode fill, and functionality such as splitters and combiners.

- **Performance Related Attributes**: solder reflow, aging, temperature cycling, moisture/temperature, radiation.

**Optical fiber**

**Role**: Silica optical fiber has been the established solution for optical transport media since the invention of low loss silica fiber in the early 1970s. Silica fiber dominates because: i) it has low optical loss (below 0.5 dB/km for multimode (MM) fiber and less than 0.15 dB for single mode fiber (SMF); ii) it has high tensile strength; iii) it has a small cross section; and iv) it has an industrial ecosystem of splicers, connectors, cables and transceivers, especially on the 125 micron outside diameter (OD) form factor. These attributes compare favorably against hollow and polymer waveguides.

**Status**: Present silica fibers have improved reliability and lower loss at tight bends, because a smaller OD (100 and 80 microns) lowers bend stress and increases index contrast to minimize light leakage in the bend. Losses in single mode systems are negligible, but single mode fiber still has proven difficult to align in high unit volumes at low cost. By trading off low alignment loss with alignment speed, SMF can become a good choice for board level interconnects. Polarization maintaining fiber is also an excellent choice for polarization sensitive, board level silicon photonics, but costly PM-connectors would be required.

**Challenges for the future**: 1) For continued improvement in power efficiency, lower coupling loss and high speed automated passive alignment will be needed. 2) Being able to terminate the silica fiber end easily and consistently, especially under field conditions is also important. 3) Emerging technologies include photonic bandgap fiber with negligible bend loss, and multiple core fibers with either single mode or multimode waveguide cores. Despite their promise, these emerging technologies will require development of new connectors, cabling infrastructure and standards to qualify for commodity, board level implementation.
The Chip-to-Chip Coupling Challenge

To fulfill future aggregate off-chip bandwidth demands, optical interconnects are a promising solution to replace copper interconnects for chip-to-chip communication. A typical chip-to-chip coupling system consists generally of four components: light source, photo detector, coupling elements, and optical waveguides. In chip-to-chip optical link schemes, the transmission distance is very short (less than one to several centimetres), so the waveguide loss and dispersion are less important than board-to-board optical interconnect applications. In the absence of waveguide propagation losses, the dominant coupling loss in chip-to-chip optical links is due to the off chip coupling. The ideal optical I/O schemes should have the following features: low cost; high density; low coupling loss; immune to position misalignments; easy packaging, etc.

Current chip-to-chip optical interconnect systems deploy VCSELs as light sources, polymer waveguides on PCB boards as the waveguides between chips, and 45-degree mirrors as deflection couplers. The coupling loss stems from many factors such as angle deviation, waveguide index contrast, axis tilting, lateral misalignment between VCSEL/PD and waveguides. Direct coupling is the simplest option but can only be used when the light divergence angle is small or when the separation gap between VCSEL/PD and waveguide is short (<50 µm). Waveguide coupling (Optical pillar) or micro-lens assisted coupling is required for large separation gaps (>50 µm). A micro-lens is helpful to reduce the mode size and improve the lateral and vertical misalignment tolerance. However, by using micro-lenses, the spatial size of the optical field is large so high-density integration is difficult. Flexible pillars satisfy the requirement of high-density, large misalignment tolerance. Coupling schemes with micro-lenses or vertical pillars show better coupling efficiency compared to free space coupling for large longitudinal displacement. For applications in which mode size reduction is required, micro-lenses are currently indispensable. On the other hand, micro-lenses occupy a larger spatial area compared to vertical pillars. Consequently, vertical pillars are preferred in applications with high-density optical devices. Considering the fact that one side of the chip is reserved for electrical bonding and wiring, and the other side is widely used for heat dissipation, edge coupling is recommended. Coupling loss mainly comes from mode size mismatch, lateral misalignment, and gaps between chips. The key challenges of an edge coupling approach include mode size converters to reduce the coupling loss from mode size mismatch, small footprint optical connectors with large lateral position misalignment tolerance, and easy packaging.

The main difficulties and suggestions of current chip-to-chip optical interconnect systems are summarized as below:

- **Light Source:** the primary light sources for optical interconnects in current ultra-short distance optical interconnect are VCSELs. VCSELs are limited due to a dedicated bonding process (cost), large mode size (low density), multimode (no WDM), etc. On the other hand, in-plane lasers provide many merits compared to VCSELs. For instance, in-plane lasers can easily be aligned with and coupled to waveguides. Furthermore, the modulation frequency of in-plane lasers can be increased significantly by fabrication of multiply segmented...
contacts along the active layer. Moreover, the coupling components such as mirrors, pillars and lens can be avoided, which greatly increases the stability and misalignment tolerance of the coupling systems.

- Waveguides: waveguides used in current ultra-short distance optical interconnects include multi-mode polymer waveguides on the PCB board, multi-mode fiber arrays, optical pillars for vertical coupling, and free space coupling. All of the above are not suitable for WDM schemes due to their multi-mode nature in which inter-modal dispersion will affect the modulation significantly. Single mode fiber array is not a proper choice due to the connector complexity and cost. A single-mode waveguide with high misalignment tolerance, low cost, will be required in chip level future optical interconnects.
System Perspective

Manufacturing Considerations Including Cost
A basic assumption for this manufacturing discussion is that the high end chips that PCI Express 3.0 will support will require multi-Tb/s I/O that is provided by optical technology using 100’s of 10Gb/s channels. In addition, we target a cost of $0.10 per Gb/s for the PCI Express 3.0 application (vs the ~ $1.00/ Gb/s for Active Optical Cables).

To illustrate the advantage optical I/O offers consider that a conventional 125 microns diameter (~1/64 of a mm²) optical fiber can transmit Tb/s data rates many meters with virtually no loss of light. An electrical differential pair operating at 25Gb/s requires about 1 mm² of area and has a range of ~1 meter.

A PCI Express 3.0 Optical Solution requires manufacturing Optical Engines (integrated, multi-channel transceivers) and the related optical interconnect structures that perform these functions:

• **E-O Function**: Electronic to Optical conversion at the data source

• **Optical Interconnect Elements**: integrated with/into electric circuit boards to provide:
  - optical waveguides to carry light
  - interfaces to sources and detectors
  - multiplexing and demultiplexing
  - optical connectors

• **O-E Function**: Optical to Electronic conversion at the data destination

The technology can be implemented using optical engines that are combined in a structure (package) with CMOS semiconductor chips. The CMOS processor chips that form the system communicate large amounts of data with one another using the optical engines interconnected through optical media. The processor chips drive the optical engines with a stream of data that the engines utilize to modulate light. The light carries the data through an optical media, such as a fiber or waveguide, to the desired destination where another optical engine converts the light back to an electrical signal.

Optical media are embedded in conventional circuit boards to add optical interconnect/transmission capability to conventional circuit boards. The optical engines are tightly coupled with the CMOS semiconductor chips in a series of structures described below.
The hardware to perform the Optical engine functions (E-O and O-E) can be built using two different approaches:

- Assembly of discrete components which are available today.
- Building components on a wafer using integration technology that is emerging.

Six basic structures, or packages, were evaluated with two variations (light source in-package vs. light source off-package) of four, for ten structures in total. The first two assumed building with discrete components; the last four structures utilize a CMOS chip combined with an optical interposer. The interposer is built utilizing semiconductor wafer integration. The CMOS chip contained the electronics; the interposer contained the optical functions. The CMOS chip and optical interposer are combined utilizing a flip-chip like approach to yield an optically-enabled I/O CMOS chip. These optically-enabled chips are then electrically and optically interconnected utilizing a circuit board that has optical media built into it to create a system with a ultra-high I/O data rates.

![Costs per unit](image)

Figure 5: Cost analyses for ten different optical interconnect designs.

The relative cost per unit is illustrated in Figure 5. The A versions have the light sources built into the package; the B version utilizes an off-package light source that adds cost. Versions 2 and 3 are more expensive because hundreds of VCSELs, photodetectors and lenses must be purchased. In versions 4, 5, 6 and 7 all of these parts are built simultaneously (monolithically) on wafers located in their final positions and interconnected with metallization. The wafer is diced to make individual interposers, and the CMOS chip is mounted directly on the interposer utilizing a package-on-package (POP) approach.
Figure 6: Schematic representation of design 4A: Chip-on-Interposer with on-package light source.

Figure 6 shows one example of the concepts described here: the 4A design of an electronic chip package on optical interposer. This version incorporates an optical interposer, and Figure 6 includes a sub sketch that illustrates the key features of the daughter card used to interconnect these optically enabled CMOS packages.

With the integrated approach, either multimode or single mode devices can be built on the optical interposer. In principal, integration is equally able to produce both. However, as the technologies to produce any version of the optical interposer are just emerging, and they may, in fact, be driven by PCI Express needs.

Optically-enabled circuit boards to provide the package-to-package interconnection are further along than the optically enable chips. Optically-enabled circuit boards are currently available from several vendors. The cost in volume is unclear, but the cost of adding the optical capability is likely to be less than that of incorporating optical engines into the chip structures and package. Implementation of the concepts outlined here will require careful selection and further development of the technologies described in other sections of this report.
Technology Transition to Short Reach Optical Interconnection

As shown in Figure 7 below, optical interconnects have been introduced commercially into ever shorter distance links to meet system needs of higher data rates. First introduced in box-box or rack-rack communications, they have penetrated to within the racks, within a shelf and between pluggable circuit packs over pluggable and fixed backplanes. The next frontier is within a PCB. The requirements change as the distance shrinks. These changes include higher density and higher link or lane count, and the communicating devices get closer to the IC. By one estimate, the number of links rises by close to an order of magnitude for each level of packaging closer to the IC. Because of the desire to consume energy only in doing the designed function, power dissipation wasted on time division multiplexing and de-multiplexing, regeneration, level shifting, error correction, pre-distortion should be minimized. To meet this constraint, optical implementation should be as tightly integrated and electrically compatible with the native CMOS circuitry as possible.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Storage &amp; Network I/O</td>
<td>Inter-Rack (node to node)</td>
<td>Intra-Rack &amp; Backplane (node to node)</td>
<td>Card &amp; Proc-memory BW</td>
<td>Optical switch (ckt)</td>
</tr>
<tr>
<td>BOX TO IO DEVICES</td>
<td>BOX TO BOX</td>
<td>INSIDE THE BOX</td>
<td>INTRA CARD</td>
<td>OPTICS BASED ARCH</td>
<td></td>
</tr>
<tr>
<td>Initial Enabling Technology</td>
<td>200 Mbps LEDs, MM fiber optics</td>
<td>2.5 Gbps vcsel arrays, Parallel fiber optics</td>
<td>10Gbps integrated highly parallel fiber optics,</td>
<td>10-25-40Gbps on module, highly parallel, dense, + WDM? Multi-core?</td>
<td>High functionality Photonics switch IC's</td>
</tr>
<tr>
<td></td>
<td>1Gbps Ethernet</td>
<td>5Gbps (2008)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/O bottleneck solved</td>
<td>BW at 100m</td>
<td>BW at &gt;10m Off-rack</td>
<td>BW at &gt;1m Off-module</td>
<td>BW at &lt;1m Off-module, DIMM</td>
<td>Switch BW Network</td>
</tr>
<tr>
<td>System Introduction</td>
<td>IBM S390</td>
<td>IBM Federation Switch/ASC1 Purple</td>
<td>IBM P775</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Figure 7: Penetration of optical interconnect technology to ever shorter distance with time. Courtesy M. Taubenblatt, IBM (July 2012)

Global IP traffic has increased eightfold over the past 5 years and will increase threefold over the next 5 years reaching 1.3 zetabytes (ZB) per year by the end of 2016. Data available for storage is predicted to reach 3.3 ZB in 2015. This growth in data usage and storage is driven by critical business applications, high speed internet, multimedia networking applications, and “Big Data” causing applications to become more data-intensive and thus requiring higher data throughput, availability, processing speeds and storage capacities. Interconnect speeds in data storage systems based on
the Serial Attached SCSI (SAS) point-to-point bus protocol will increase to 24 Gbps by 2016\(^\text{iii}\), while the InfiniBand protocol used predominantly for rack-to-rack communication in enterprise Data Centers and High Performance Computers (HPC) will provide 25 Gbps per link by 2013\(^\text{iv}\).

The resulting increase in capacity, processing power and bandwidth density will severely impact design, cost and performance of future data center and HPC systems, bringing increased demands in terms of signal integrity and power consumption. As communication bandwidths increase, the conversion point from electrical to optical interconnects migrates ever closer to the on-board processing complexes, whether these are CPUs on a server blade, SAS controllers, expanders, PCIe or InfiniBand switches.

The performance of state-of-the-art HPC processors is scaled to future generations with the assumption that the chip size (2.5cm x 2.5cm) do not increase substantially. An IBM Power7, 45 nm multi-core processor has a chip I/O bandwidth of ~0.25TB/s and consumes ~200W/chip. If scaling the CMOS critical dimensions down to the 9 nm node provides a ~25X increase in the number of computing nodes and ~2X increase in clock speed, then the commensurate chip I/O will increase to ~12.5TBps, or ~10Tbps/cm bandwidth density at the chip edge. Assuming the on-chip global interconnects require ~5X of the chip I/O\(^\text{v}\),\(^\text{vi}\), an on-chip bandwidth density of 80Tbps/cm\(^2\) is projected. If the on-chip global and off-chip optical interconnects are allocated 40% of the total chip power budget (80 W), a chip-scale interconnect energy consumption lower than 133 fJ/bit is needed, assuming the same photonic techniques are used for the on-chip global and off-chip fabric.

Electrical links are loss limited at high data rates, and photonic links support higher data rate and longer reach. As photonic links displace electrical at shorter reach, photonic interconnects will trend to a higher fraction of the interconnect count within the overall communications ecosystem. While the external requirements of real-time video transmission are forecast to grow to two-thirds (2/3) of global demand by 2017—an increase of 16x — the near term product driver capable of enabling such demand will be the transceiver to support optical retrofit with the ‘active optical cable’. The long term driver will result from new functional architectures that are capable of partitioning imaging, logic and communication in order to more efficiently scale information capacity at constant energy and footprint.

The significant consequences of short reach optical interconnection will be:

- increases in total addressable market (TAM)
- decreases in margins
- increases in common manufacturing infrastructure and research focus
- revolutionary changes in network architectures, standards and chip-level integration within the information hardware industry

High unit volumes at short reach will revolutionize photonic economics. High level photonic integration will drive a new manufacturing paradigm that will determine cost and, consequently, design. Scaling performance at constant cost will enable transition
from the traditional 10-15 year ‘network build’ interval to 3 year product cycles. This scaling-induced obsolescence will be self-supporting, as the revenues will fund more research with a common industry focus. Standards and industry consolidation will help to decrease cost and price. The disruptive solution to the photonics industry problem, however, will not be performance scaling, but shifting out of the demand curve to a wide range of cross-market applications.

The penetration of optical signaling in a system interconnection hierarchy is dictated by the traditional communication figure of merit: the distance x bandwidth product. The cost, energy and bandwidth density of copper lanes compete less favorably with optical interconnects at higher data rates. These tradeoffs were described in detail in the 2011 CTR III TWG release “Copper Scalability”. Optical signaling is pervasive for all data rates in telecommunication networks for distances greater than 10km. A transition to optical lanes with 10Gb/s data rates for rack-to-rack communication in local area networks (LANs), data centers and high performance computing (HPC) installations is now in progress. The penetration of optical lanes to board level interconnects, the next transition, is the subject of this study.

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>B</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-rack</td>
<td>100 m</td>
<td>1–10 Gb/s</td>
<td>$5/Gb/s (AOC)</td>
</tr>
<tr>
<td>Backplane</td>
<td>1 m</td>
<td>100 Gb/s – 1Tb/s</td>
<td>$0.1/Gb/s (*Cu)</td>
</tr>
<tr>
<td>Linecard</td>
<td>30 cm</td>
<td>300 Gb/s – 3Tb/s</td>
<td>$0.01/Gb/s (*Cu)</td>
</tr>
<tr>
<td>Chip</td>
<td>3 cm</td>
<td>3Tb/s – 30 Tb/s</td>
<td>$0.001/Gb/s (*Cu)</td>
</tr>
</tbody>
</table>

*Copper connectors
Backplane: $0.1/Gb/s ($1.50 for 2 differential pairs @ 10Gbps)
Linecard: $0.01/Gb/s ($0.20 for mated differential pair @ 20Gbps)
Chip: ITRS estimates for BGA

Table 1: Aggregate Data Rate, Distance and Price Thresholds for Implementation of Optical Interconnection

Table 1 shows the price expectations for optically-enabled interconnection scaled to constant cost/Gb/s. If there were no legacy of electrical interconnection, a hierarchical architecture might not be the favored design. Hierarchy implies interfaces that consume space, time and energy as new signaling protocols are introduced. Copper wires require hierarchy and interfaces to achieve signal integrity for longer distance transmission. Typically, copper lines are designed for 20dB loss within each hierarchical level. The fundamental design rule for communication chip and board architecture is i) to use the highest data rate consistent with 20dB loss at the required link length; ii) to add repeater stations along the link to maintain signal integrity; and iii) to add additional single channel lanes to meet the link bandwidth requirements. This low risk path is consistent with fault intolerant system performance, but it necessitates passage through the full technology transition ritual for the replacement of electronic lanes by photonic lanes.
The photonic interconnection legacy of high-performance, high-margin discrete network components poses a barrier to new technology as well. The three key areas to be addressed are packaging, test and design.

- Short reach optical interconnection technology requires a pluggable-once component without embedded fiber attach.
- Transceivers will be the near term focus for photonic integration. Monolithic, chip level integration will require low cost, single mode I/O. This transition will be driven by bandwidth density requirements, but adoption is critically dependent of low cost package and test solutions.
- Design is the primary constraint to scaling; the photonic reach and bandwidth advantages must be executed with highly leveraged, architectural synergy.

### Electronic/Photonic Convergence in Silicon

**MIT Short Reach Communication Technology Roadmap**

<table>
<thead>
<tr>
<th>Component</th>
<th>Fiber Lasers Detectors</th>
<th>MUX EDFA</th>
<th>100Gb/s Ethernet &gt;1Gb/s I/O Tbps Transceiver</th>
<th>Microphotonics: ICs and ASICs</th>
<th>Pervasive Microphoton Ics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>ETDM</td>
<td>DWDM</td>
<td>Inter-rack Chip-Chip</td>
<td>Backplane PWB</td>
<td>Pervasive edge functionality</td>
</tr>
<tr>
<td>Target Application</td>
<td>WAN</td>
<td>WAN</td>
<td>Access HPC Data Center Automotive Consumer Devices</td>
<td>1Gb/s Access 10Tbps WAN Optical Router</td>
<td>Transparent Network</td>
</tr>
<tr>
<td>Package</td>
<td>Fiber</td>
<td>Fiber pigtail</td>
<td>AOC Optical Engine</td>
<td>E-P Chip Stack</td>
<td>Optical I/O Pins</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Technology</th>
<th>IC: Al/SiO₂</th>
<th>IC: Cu/SiO₂</th>
<th>Hybrid Flip Chip Hybrid Bonded</th>
<th>CMOS Optical Plane</th>
<th>Monolithic E-P processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>DRAM, SRAM μProc</td>
<td>DSP μProc</td>
<td>Parallel Processing</td>
<td>Spectral Efficiency</td>
<td>Global Optical Control</td>
</tr>
<tr>
<td>Constraint</td>
<td>Yield</td>
<td>Shrink</td>
<td>E-P Packaging E-P Test Reliability</td>
<td>E-P CAD</td>
<td>Analog/Digital Functionality</td>
</tr>
</tbody>
</table>

Figure 8: Roadmap for the electronic-photonic convergence that is driving short reach optical interconnection.

---