

# Optical Interconnects in Systems

A. F. J. Levi

**Abstract** — Future enhancement of system performance will decreasingly rely on reduction in transistor dimensions. Rather, performance gains will increasingly come from improved hardware and software architectures and emerging technologies such as optical interconnects which provide a new design-space for system designers. It is possible that hybrid opto-electronic interconnects may revolutionize system implementation in the next decade.

**Index Terms** — Optical interconnects, parallel fiber-optics, free-space optical interconnect, IO bottleneck, end-system, quality-of-service

## I. INTRODUCTION

There are various ways to rationalize the use of optics in otherwise electronic information processing systems. In his article in this Special Issue, David Miller places the emphasis on basic physical phenomena. From this starting point he argues that optical components provide solutions to physical problems that limit the use of electrical interconnects in systems. By way of contrast and to provide an alternative perspective, in this article the focus is to evaluate the impact and benefit of advanced optical interconnect technologies in systems assembled from a large number of electronic components.

Over the last few decades advances in complexity and performance of circuits has been carefully managed by the microelectronics industry to follow what has become known as “Moore’s Law”. This rule-of-thumb states that the performance of computers and the associated silicon integrated circuits increase by a factor of two every eighteen months. The Semiconductor Industry Association has institutionalized Moore’s Law via the “SIA Roadmap” which tracks and projects needed advances in most of the electronics industry’s technology [1]. Never-the-less, the impossibility of sustaining a continued reduction of transistor device dimensions is well illustrated by the fact that Moore’s law predicts DRAM cell size will be less than that of an atom by the year 2020. Well before this endpoint is reached, quantum effects dominate device performance and conventional electronic circuits fail to function.

It is likely that future system performance gains from simple scaling of transistor device dimensions will not contribute as much as they have in the past. Performance improvements will increasingly come from new architectures, better operating systems, and introduction of

new technologies such as photonics. From this perspective, systems will continue to improve their performance, not by relying heavily on incremental reduction in transistor size (scaling), but rather by creating new more efficient architectures which exploit, among other things, photonics. In this picture, photonics becomes a key element in an expanded design space for a system architect. Photonics technology will be used in systems to extend and enhance performance thereby ensuring Moore’s Law continues well beyond the time when continued transistor scaling is no longer practical. Increasingly, photonics will be an option, often the only option, for designers trying to deliver systems which track Moore’s metric.

The challenge of stretching Moore’s Law well beyond the year 2020 can be met by designing systems more intelligently with new, efficient, architectures and by exploiting non-electronic technologies such as photonics.

In fact, relatively simple photonic components have already been successfully incorporated into a large number of systems. Telephone and campus networks use fiber-optic links to connect switches, data-base storage devices are connected to servers via fiber-optic links, and compact disks are read using laser-diode based sensors. However, the intimate integration of high-functionality photonic components into the heart of otherwise electronic systems is a significant task requiring a high level of cooperation between component manufacturers and system integrators.

In the next section, the Input / Output (IO) bottleneck in electrical systems is introduced. The very high edge-connection bandwidth density, low power, and electrical isolation offered by modern optical interconnect solutions is discussed. Section III describes opportunities for optics in emerging convergence technologies and section IV discusses the hardware and software origins of the end-system bottleneck. Section V briefly introduces switched architectures. Sections VI and VII describe perceived barriers to inserting opto-electronics into CMOS based systems along with possible future directions for opto-electronic component research.

## II. LARGE SYSTEM INTERCONNECTS

Since large systems are, by their very nature, constructed from a number of smaller components, the development of cost-effective approaches to interconnection and packaging is a fundamental issue faced by any system designer. Often, the competitive advantage of a given system is determined by the integration and interconnection strategy of the system designer.

As a specific example, consider how the number of electric IO signal lines determines packaging choices in the design

---

A.F.J. Levi is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-1111, USA.

of a large system such as a telephone switch. Fig. 1 illustrates, as a function of (a) packaging choice and (b) interconnection length-scale, dominant interconnection technologies in part of a large high-performance system [2] (in this case a 1024 x 1024 switch matrix in a 5ESS telephone system [3]).

Due to economies of scale and the need to implement different functions, large systems consisting of greater than one billion transistors are created from a number of individual Integrated Circuits (ICs). Fig. 1(a) shows electrical IO and packaging option with number of gates per IC. To help quantify the situation, packaging solutions are assumed to be driven by Rent's rule (where  $(IO/k)^{1.8} =$  number of gates). The high IO bandwidth density created by efficient packaging using High Performance Multi-Chip Modules (HPMCMs) and Printed Circuit Boards (PCBs) results in an edge-connection IO bottleneck at the board-to-board level. Electrical interconnects simply fail to provide the needed edge-connection *bandwidth density* (measured in units of Gb/s/cm).

Taking a closer look at Fig. 1(a) one sees that the number of gates per IC is indicated on the horizontal scale and the number of IO terminals required is on the vertical scale. The lower curve (labeled IC) indicates the number of IO required per IC as a function of the number of gates. In this example, Rent's rule shows that a *single* IC containing 100 million gates (~ billion transistors) and an IO of 14,000 signals would satisfy the system need. Unfortunately single-chip solutions do not provide flexibility in system design and do not provide economics of scale that are inherent to designs that use multiple generic ICs. Flexible, cost-effective systems tend to use multiple ICs each with about 0.1 to 1 million gates and corresponding IO in the range 300 to 1,100. The system designer must now identify packaging solutions for multiple ICs. In Fig. 1(a), the curve above the one labeled IC is the IO created when ICs are packaged in efficient HPMCMs which are assumed to contain up to a maximum of ten ICs. The HPMCM are themselves mounted on PCBs. IO flows off the edge of PCBs. Groups of up to ten PCBs are packed into a shelf. Up to six shelves are packed together to form a frame. IO from PCBs flows via the frame backplane and / or via point-to-point links. Today's existing electrical solutions for the edge-connection IO bandwidth density for PCBs are inadequate and have become a system bottleneck.

Another perspective on the same bandwidth density issue may be achieved by examining interconnect hierarchy as a function of interconnect length. Fig. 1(b) illustrates system interconnect as a function of length scale and packaging option. The approximate data rate above which controlled impedance lines are required for electrical interconnects is also shown. Significantly, requiring controlled impedance for wires longer than 3 mm which support Gb/s signaling results in increased power consumption [4]. This increase in power dissipation has an impact on electrical interconnect approaches and provides an opportunity for efficient, high-speed, low-power optical solutions [4 - 5]. On length scales greater than several tens of meters, standardized conventional serial fiber-optic Local Area Network (LAN) technologies such as Fibre Channel [6],

ATM (OC12, OC48), Gigabit Ethernet, and FDDI [7] have been successfully implemented. Viewed as a system-level packaging problem, networking, such as occurs in a LAN, is merely a way to interconnect components of a distributed system.

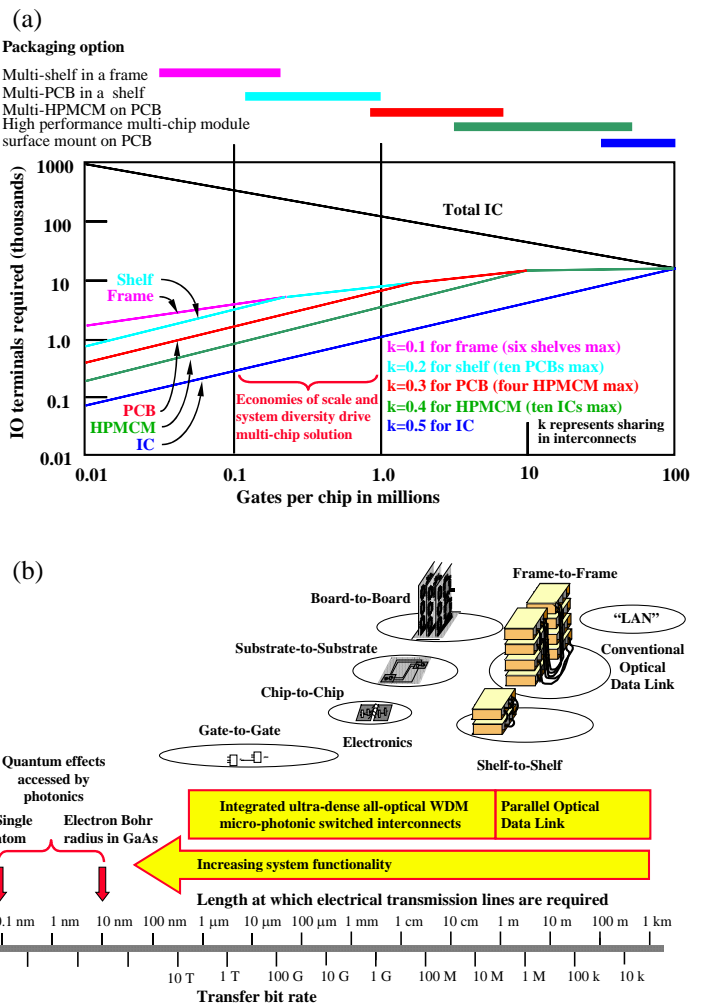


Fig. 1. (a) Large-system IO as a function of gates per IC and the indicated packaging options which are assumed to be determined by Rent's rule. The system contains 100 million gates. If constructed from one hundred 1-million gate ICs each IC would have an IO of about 1,100 and so the corresponding point on the Total IC IO curve is 100 x 1,100 = 110,000. (b) Illustration of system interconnect hierarchy as a function of length scale and packaging option. Also shown is the approximate data rate above which controlled impedance lines are required for electrical interconnects.

As shown in Fig. 1, there exists an opportunity for advanced parallel fiber-optic technologies to impact high-performance system design on length scales from 1m to a few 100 m. Parallel fiber-optic modules are a potentially low-cost solution to a specific IO bottleneck that has developed in large telecommunication and supercomputer systems [2, 8]. This bottleneck, which exists in electrical systems, arises due to the high-density of relatively slow-speed electrical connections needed at the edge of cards, shelves and between frames. Parallel optics with its high form-factor and high-speed gives a high edge-connection

bandwidth density and hence offers an excellent technical solution to this bottleneck. The cost of future system implementation has been driven down to an attractive level by advances in components such as Vertical Cavity Surface Emitting Lasers (VCSELs) [9 - 10] plastic waveguides, plastic array connectors [11], and low-skew fiber-optic ribbon [12].

Parallel fiber-optic interconnects is emerging as an immediate solution for the edge-connection bandwidth density bottleneck. Parallel fiber-optic interconnects will be used for multi-GB/s interconnects over length scales of 1 m to 1 km. Excellent progress in bringing this technology to market has been made in recent years. The article by Takashi Yoshikawa in this Special Issue is an example of an effort that has made a significant contribution.

The introduction of new optical interconnect technologies is an opportunity to re-evaluate system architectures by providing designers with a new cost / performance trade-off. This expansion of the design space has potential to revolutionize system implementation in the next few decades. Even today, optics could be used to improve system design by replacing electrical interconnects at the edge of printed circuit boards with optical interconnects.

At present, electrical interconnects are the solution of choice for intra- and inter-chip connections on length scales up to approximately 1m. The reason for this is simply that no viable cost-effective alternative has emerged. If new technologies such as free-space optical interconnects are successfully developed, this could provide the system designer with a dramatically enhanced set of choices and new approaches to system optimization. In addition, there is the possibility that advanced optical technologies could provide new types of system functionality. For these and other reasons, there has been a great deal of research in this subject in recent years.

The challenge of providing *useful* optical interconnection on length scales shorter than 1 m will most likely *require* functions significantly more advanced than a simple point-to-point link. Examples of increased function include all-optical switching, all-optical media access control, optical memory access, some level of optical computation, and controlling quantum effects using photonics.

In recent years this research challenge has begun to be addressed by a number of groups. Initially, most approaches have made use of VCSELs due to their small size, low power consumption, high efficiency, and the ease with which area arrays may be fabricated. When combined with optical receivers and CMOS electronics, a few thousand VCSELs, each signaling at Gb/s rates, can be used to provide an opto-electronic switched backplane with Tb/s interconnect bandwidth. Such Tb/s backplanes have interconnection lengths in the cm range. The FAST-Net project described by Mike Haney in this Special Issue is an example of current research in this subject. Masatoshi Ishikawa and David Plant also discuss programs with approximately similar objectives.

### III. CONVERGENCE TECHNOLOGIES

At present there are two convergence technologies which provide special opportunities for optics. The two markets are in the area of telecommunications and the networked professional work environment.

In telecommunications the trend is to merge subscriber services to provide a single service for voice, email, web, TV, and entertainment. Switching and server systems capable of delivering such services to the consumer will use optical interconnects to eliminate the IO bottleneck illustrated in Fig. 1.

In the professional work environment, a *productivity convergence* is taking place involving the merging of information processing and network resources. This convergence of computing and communications in the *professional campus* environment has created a new class of applications with significant bandwidth requirements and more stringent delivery-time constraints [13]. In the case of high-performance workgroup environments in small campus networks, several bandwidth intensive multimedia applications such as multicast videoconferencing, remote graphics visualization, and distributed computing have emerged. Conventional network interfaces and network designs are severely strained to meet the Quality-of-Service (QoS) demands of these applications at an affordable cost.

Meanwhile, inexpensive Personal Computers (PCs) have continued to improve in performance and commercialization of processors with clock rates in excess of 1 GHz is fast approaching [14]. The use of such machines in a high-bandwidth networked cluster environment can provide significant re-configurable computing resources at moderate cost.

Additional technological support comes from high-performance CMOS-based ICs which can be designed to interface between a host computer and network [15] providing bandwidth capabilities that were only available previously using expensive technologies such as bipolar Emitter-Coupled Logic (ECL). CMOS-based IC design is attractive because it leverages the cost benefits of an inherently simpler process compared to silicon bipolar and leverages the infrastructure of high-volume commodity ICs for high-performance circuit applications. It also offers a higher level of integration and the potential to replace older multi-chip circuitry with single-chip CMOS-based solutions.

The network physical medium needed to meet the requirements of a professional high-performance workgroup environment could, in principle, be either electrical or optical. Current small area networks are constructed using electrical links. Copper cables are however bulky and restricted in density, bandwidth and interconnect distance capabilities. For example, the electrical link described by Walker et al. [16] provides a 10 Gb/s serial link over distances less than 20 m, using relatively thick coaxial cable. In contrast, fiber-optic based links offer thin, flexible cable, essentially unlimited bandwidth, immunity from electromagnetic interference, and a significantly higher edge-connection density bandwidth. A parallel fiber-ribbon constructed using multi-

mode fibers is a natural and low-cost solution to provide the needed total interconnect-bandwidth while interfacing to the wide data buses of CMOS-based systems in a less complex and power-efficient way compared to conventional serial link approaches.

An interconnect standard which recognized these benefits early and specifically included parallel optical interconnects is the High Performance Parallel Interface, HIPPI-6400 [17]. Emerging standards such as NGIO and FutureIO will also incorporate specification for parallel optical interconnects [18 - 19].

As discussed previously in section II, the edge-connector bandwidth density (Gb/s/cm) is a known barrier to efficient system integration. One solution exploits the inherently high-bandwidth density of ribbon-fiber based parallel optical interconnects. Such an approach leverages the industry standard MT fiber connector which uses an approximately 6 mm wide ferrule holding 12 multi-mode glass fibers on a 250 mm pitch. In this case, an opto-electronic transceiver port with 12 transmit and 12 receive fibers is best accommodated using a sub-12 mm wide (< 2 x 6 mm) electronic interface.

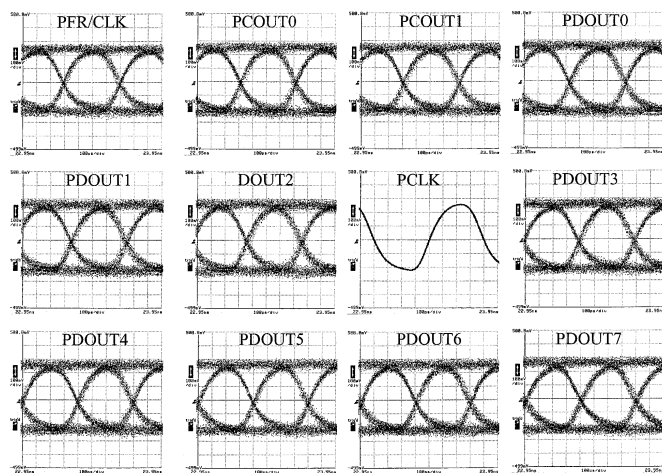


Fig. 2. Measured eye-diagrams and photograph of CMOS interface IC. Low-cost 0.5  $\mu\text{m}$  CMOS is used to bridge between a relatively slow parallel electrical interface and a very high-speed parallel optical interface [20]. Each one of the 11 high-speed Tx and 11 high-speed Rx data lines signal at 2.5 Gb/s. There is one high-speed Tx clock and one high-speed Rx clock. The 1 cm-wide integrated 1:2 / 2:1 mux / demux circuit has an internal and external bisection data bandwidth of 55 Gb/s = 11 x 2 x 2.5 Gb/s. The high-speed IO is source and load terminated and LVDS compliant. The IC has a 2.7 ns Tx / Rx mux / demux end-to-end latency.

The challenge for CMOS is to deliver the needed bandwidth density to the opto-electronic interface while at the same time effectively bridging to the lower bandwidth density of conventional electronics. In fact, low-cost 0.5  $\mu\text{m}$  CMOS has been shown to be capable of providing a

multiplexed 55 Gb/s/cm interface [20]. This edge-connector bandwidth density is a factor of 10 greater than the competing state-of-the-art HIPPI-6400 parallel *electrical* interconnect. Fig. 2 shows an example of such a low-cost 0.5  $\mu\text{m}$  CMOS used to bridge between a slow parallel electrical interface and a very high-speed parallel optical interface. Each one of the 11 high-speed transmit (Tx) and 11 high-speed receive (Rx) data lines signal at 2.5 Gb/s.

To gain perspective on the potential impact this type of bandwidth density, consider Fig. 3. On the left is a photograph of the Agilent-PONI parallel fiber-optic module which can support greater than 50 Gb/s per linear cm. The module uses 12-wide fiber ribbon with 2.5 Gb/s signaling per fiber. The high bandwidth density is achieved using a standard plastic MTP optical connector and BGA surface mount to the PCB for electrical connection. By way of comparison, the IBM 8263 ATM switch shown on the right hand side has a backplane which is approximately one meter in size and has a total capacity of 12.8 Gb/s or about one quarter that available from the PONI technology.

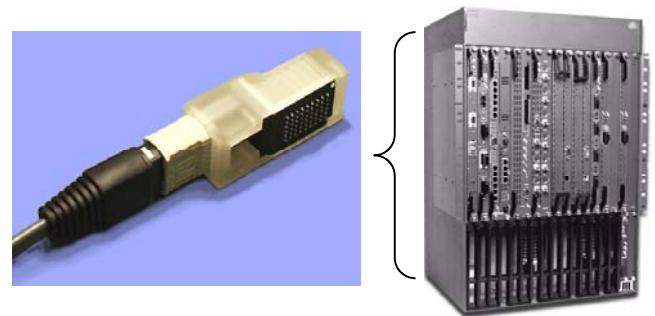


Fig. 3. The Agilent-PONI parallel fiber-optic module (left) can support greater than 50 Gb/s per linear cm. The module uses 12-wide fiber ribbon with 2.5 Gb/s signaling per fiber. The high bandwidth density is achieved using a standard MTP optical connector and BGA surface mount to the PCB for electrical connection. By way of comparison, an IBM 8263 ATM switch (right) has a backplane which is approximately one meter in size and has a total capacity of 12.8 Gb/s.

#### IV. THE END-SYSTEM BOTTLENECK

From a system perspective, the relative merits of individual components such as optical links should be assessed in terms of overall system benefit. This includes both cost, reliability, availability, and performance. For commodity consumer products such as PC end-systems, cost and availability introduce significant constraints. Traditional long-distance telecommunication and switching equipment does not have quite such stringent cost constraints and tends to emphasize performance.

To illustrate the complexity and broad nature of system-level issues, consider the factors contributing to the so called “end-system bottleneck” for a PC connected to a high-speed network. The PC user wants unimpeded access to local and networked resources which provide video, graphics, text, audio, and other services. From the user perspective the QoS can be parameterized in terms of measurable quantities such as sustained application data



though-put, latency, and jitter. The difference between raw performance at the individual hardware / software component level and the application level can be dramatic. Fig. 4(a) illustrates a typical Intel Pentium PC architecture with LAN attached to a Peripheral Component Interface (PCI) IO bus. If the user requests an uncompressed two-way video link via the network then a minimum sustained data rate of  $640 \times 480 \times 24 \times 30 = 221 \text{ Mb/s}$  must be transmitted in each direction. This estimate assumes minimal NTSC video with 24 bit color. In a television studio production and editing environment, broadcast quality video such as D1 CCIR-601 [21] has data rate requirements in excess of 250 Mb/s. In practice, the use of general-purpose platforms such as PCs results in a significant degradation in end-system QoS. This occurs due to a combination of hardware and software data bottlenecks.

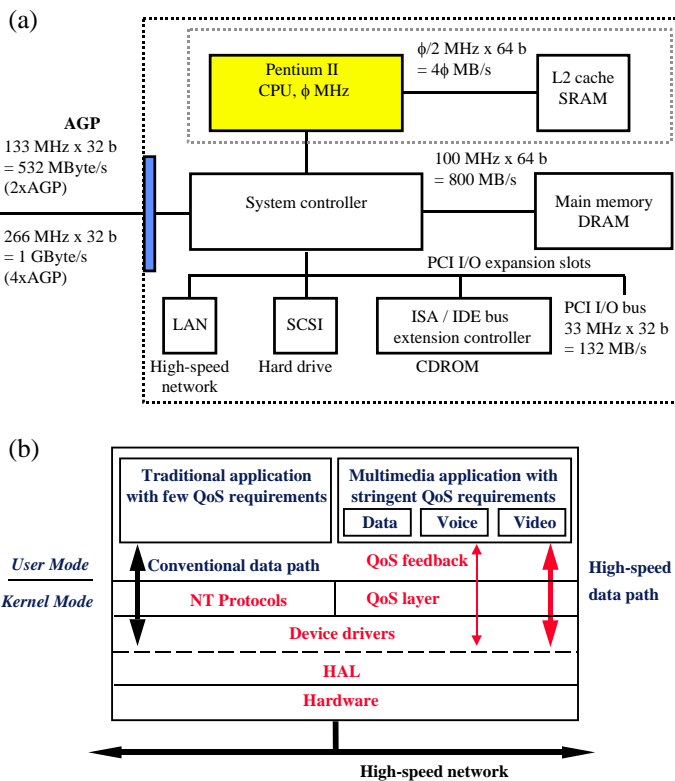


Fig. 4. (a) Schematic of the Intel Pentium II bus-based architecture. The CPU has a clock operating at  $\phi$  MHz. A typical numerical value of the clock rate  $\phi$  is approximately 500 MHz. High-speed LAN connection is made via a PCI adapter card. Data is physically moved from the adapter card to DRAM main memory via the system bus controller. After processing data is then moved from main memory to the display via the system bus controller and AGP. Because the system bus is a *shared* resource, it is not possible to sustain the peak though-put of individual physical components throughout the system data-path. (b) Protocol stack illustrating layers of Kernel and User Mode software between high-speed network physical layer and the user application. Device drivers must be modified to accommodate high sustained through-put networked uncompressed video applications with QoS requirements.

As illustrated in Fig. 4, data from the high-speed LAN enters the system via a PCI adapter card. Data is then physically moved to DRAM main memory via the system

bus controller. Typically, the CPU performs operations on the data such as copying it to new locations in memory before moving it to the display via the system controller and AGP. Such copying introduces unnecessary loading, jitter, and latency. Implementation of single or zero copy protocols and shared memory implementations reduces, but does not eliminate, the QoS impact.

In addition to reducing the copying of data in memory, an important technique which has been shown to reduce message latency is the concept of active messages in which a process at the receiving node can start before the complete message has arrived [22]. This is achieved by including in the message the address of the function which is to be invoked upon arrival at the destination. Another approach used by Hamlyn [23] creates a high-performance network interface by implementing sender-based memory management. Ultimately, changes to the operating system have to be made so that these methods of improving through-put and reducing latency are both invisible to the user and safe to use in a multiprogramming environment.

Because data is physically moved via *shared* resources such as memory, the system controller, and PCI bus, there are often significant physical limits imposed on achievable sustained through-put,  $R_{tot}$  for a given system implementation. Simple rate equation analysis reveals that  $1/R_{tot} = \sum 1/R_i$  where  $R_i$  is the sustainable rate of a given individual process such as memory copy. Experiments show that typical applications capture less than 20% of the available network bandwidth [24], i.e. less than 200 Mb/s for a Gigabit Ethernet LAN. The *shared* physical resources and system architecture of a general-purpose PC fails to deliver even a single stream of studio-quality uncompressed NTSC video to the end-user.

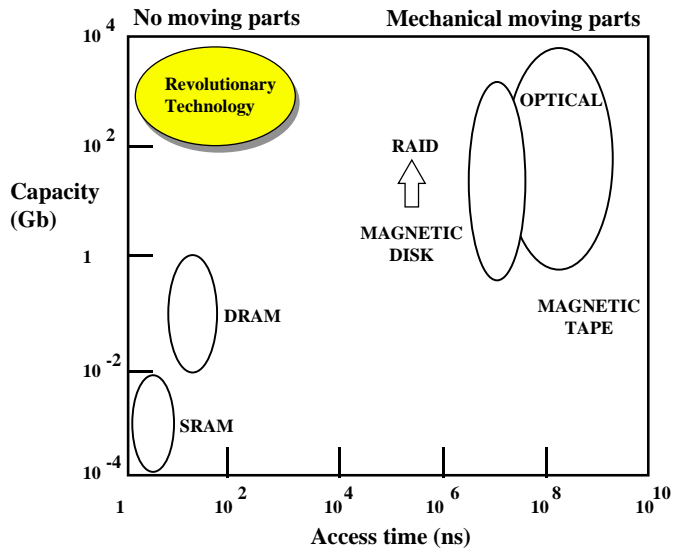


Fig. 5. Illustration of the memory access bottleneck. Ideally, many terabits of stored memory can be accessed at random with ns latency. The realization of such a memory which is both practical and can be integrated into a system requires introduction of revolutionary technology.

Another constraint that has long been known about is memory access. While processor clock rate doubles every 18 months, main memory data transfer speeds increase 10%

every 18 months. This widening performance gap is well documented [25] and illustrates the failure of conventional architectures and interconnect implementations to deliver bandwidth and latency that matches improvement in processors. Fig. 5 illustrates one aspect of this memory access bottleneck. Conventional SRAM and DRAM have relatively small capacity and relatively long access times so that processors have to wait many clock cycles to retrieve data from main memory in the event of a cache miss. The existence of components, possibly photonic, capable of storing terra-bits accessed at random with ns latency would mitigate the memory access bottleneck. The realization of such a memory which is both practical and can be integrated into a system requires introduction of revolutionary technology.

In addition to these physical constraints, conventional device drivers, operating systems such as Microsoft NT, and typical applications impose further rate-limiting operations on the data. In fact, the Kernel / User mode architecture itself implies fundamentally unnecessary copying of data and hence reduction in achievable sustained throughput. The operating system and the existence of shared resources introduces further degradation in QoS by imposing extra latency and jitter. Normally the situation becomes severely degraded when more than one high-bandwidth application is requested by the end-user. Unfortunately, typical device drivers and operating systems such as NT are not equipped to provide the needed QoS management. Performance could be enhanced by implementing a high-speed data path and QoS management software that blurs the Kernel / User mode boundary as illustrated in Fig. 4(b). However, in addition to a host of software difficulties, such an approach is always limited by the constraints of shared hardware resources in bus-based architectures.

The above discussion of end-system bottlenecks is only one very specific example of a system optimization problem. Although different systems such as servers have different needs, the common theme of shared resource allocation and management nearly always emerges as key to system-level optimization. An important lesson from such studies is that there are significant benefits if both hardware and software can be optimized *simultaneously*.

## V. SWITCHED ARCHITECTURES

At this point it should be clear that a change in architecture might be of benefit. In particular, it would be helpful if the network, processor, frame buffer, main memory, etc., were served more democratically. By this one means that the network, processor, display, and main memory all have essentially equal access to what is traditionally the processor bus. In this situation processor performance might not be optimized but the overall functionality and potential utility of the networked system could be dramatically enhanced.

Initial changes to system architecture which attempt to solve some of these difficulties have already been implemented by a number of vendors. These new

organizations are designed to accommodate concurrent access or transfer at high-sustained data rates between the processor, memory, and IO. An example of such an implementation is the SGI Origin switch-based system architecture [26]. This architecture interconnects two processors (2P) per node via a hub to local main memory. The hub implements a full four-way cross-bar between processors and provides switched access to a multi-port IO cross-bar switch. This IO switch is the building block of a high-bandwidth, low-latency scalable shared-memory multi-processing architecture which can support up to 512 nodes. Fig. 6 indicates memory bandwidth and latency performance targets for different processor configurations. The interconnect used in this system conforms to the HIPPI-6400 standard and so, in principle, could directly benefit from the use of parallel optical interconnects. Beyond this application of opto-electronic interconnects, it is conceivable that in the future all-optical cross-bar switches will replace the CMOS-based devices used in systems today.

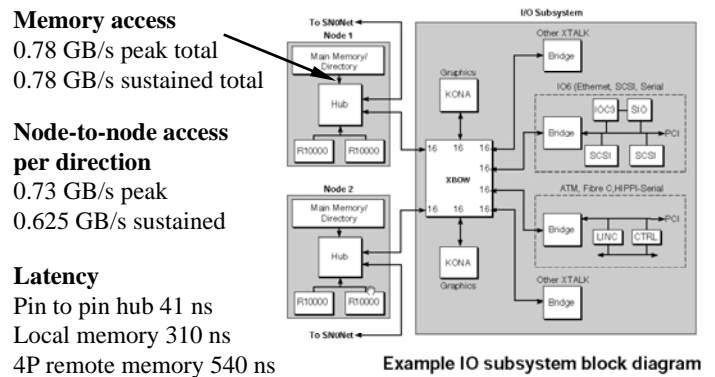


Fig. 6. Schematic of the SGI Origin switched-based system architecture. The IO subsystem is designed to connect multiple processor boards. Each processor board has two processors (2P) and memory connected via a hub. Electrical interconnects with 44 signal pins per direction over distances of up to 5 m are used to connect nodes. From [26]. Local memory access time is 310 ns, 4P remote memory access time is 540 ns, 8P average remote memory 707 ns, 16P average remote memory 726 ns, 32P average remote memory 773 ns, 64P average remote memory 867 ns, and 128P average remote memory 945 ns. Node-to-node 16 kB page block transfer takes less than 30  $\mu$ s.

## VI. BARRIERS TO INSERTING OPTO-ELECTRONICS IN CMOS-BASED SYSTEMS

Opto-electronic interconnects can provide new system design optimization and functionality such as an essentially length independent link and reduced power consumption compared to copper-based interconnects of a given bandwidth. Distributed systems with reduced heat density are enabled by optical interconnects. It is also possible to reduce IO count because optics can provide higher signaling rates. In addition, opto-electronic interconnects offer the benefit of electrical isolation and low electromagnetic interference.

Unfortunately, despite these and other technical advantages, there are a number of important barriers to inserting new

photonic technology into systems. Placing undue emphasis on metrics such as component cost, reliability, availability, and performance is certainly over-simplifying the problem. A system designer or, more often, a design team, might be offered immediate availability of reliable high-performance photonic components at no cost and still reject their use in a system! There are some very compelling reasons for doing so.

The mere fact that systems with twice the previous performance are delivered every 18 months puts a great deal of stress on designers and limits approaches to system design. The necessity of meeting project deadlines forces the system design team to reuse and leverage as much as possible from previous designs. The end result is an inherent conservatism in design. Even if the design team overcame this conservatism and decided to take the risk of incorporating advanced photonic technology, they might be forced to reconsider. One reason is that system risk is shared by more than the design team and its part of the corporation. Unfamiliarity with the technology can impact the ability of subcontractors, customers, marketing, and service personnel to support the new system. In this way, risk is propagated through every aspect of the corporation's business.

One approach to overcome these difficulties is for component manufacturers to work cooperatively among themselves and with system integration companies to provide comprehensive system-level support. Inserting opto-electronics in CMOS systems would be easier if component manufacturers worked to remove barriers to adoption by system designers. In addition to keeping component cost low, proving reliability as good or better than copper, and ensuring availability via multi-sourcing and standardization, the list of support activities might include:

- One-stop *technology shopping* and *design support* for:
  - Standard packaging and board-level integration
  - Standard CMOS library cells
  - Standard optical sub-assembly (OSA) footprint
- Standards that *help* the designer including:
  - Standard evaluation boards
  - Standardized mechanical design
  - Standard system testing
  - Standard system software

There is no doubt that such an effort is a substantial undertaking requiring a level of cooperation and unity at least comparable to that achieved by the SIA.

## VII. FUTURE DIRECTIONS AND CONCLUSION

The preceding has touched on some of the potential impact and benefits of today's advanced optical interconnect technology. In the future, it is likely that photonic components with greatly enhanced capabilities will become available. For example, micro- and nano-photonic circuits may be used to perform Media Access Control (MAC) in the all-optical domain. Such circuits could provide ps node latency, deadlock protection, and adaptive routing in mesh-based System Area Networks (SANs). The integration of

these components might involve direct die attach to CMOS circuitry and the use of PCBs with copper and optical waveguides. A highly speculative schematic cross-section of such a heterogeneously integrated scheme is shown in Fig. 7.

All-optical micro-photonic routing based on Wavelength Division Multiplexing (WDM) may have an inherent system advantage by providing a natural means to eliminate deadlock conditions that occur in switched-based SANs such as mesh routers. One approach is to have virtual WDM sub-networks that the system can use to extract itself from the deadlock [27]. Micro-photonic WDM components have potential to provide the key enabling technology for deadlock-free all-optical switched architectures.

Exactly how much functionality should be incorporated into optical components is not obvious at present. However, as indicated in Fig. 1(b), it seems reasonable to suggest that the level of functionality or intelligence built into future micro-photonic designs will likely depend on how far optics penetrates into the system.

Another factor which influences adoption of advanced technologies is the system power budget. Air-cooled systems can be characterized by a maximum average power dissipation density. As an example, consider a 20-slot 6U-card VME sub-rack chassis. Each card is approximately 6" x 9" and can dissipate up to 50 W with 500 cubic feet per minute of air flow. The cards are placed in the chassis on a 0.8" pitch. The *external* dimensions of a typical chassis are 10.5" x 19" x 13" giving a maximum average system power density of slightly less than 0.4 Watts per cubic inch. This is essentially identical to the maximum average system power density of high-performance workstations such as the HP J7000. From a thermal point of view, the difference between the VME system and the HP workstation is that heat dissipation can be more uniformly distributed throughout the volume of the VME chassis. Any reduction in power dissipation that adoption of efficient micro-photonic components and subsystems can have allows allocation of additional power to typically power-starved resources such as the CPU.

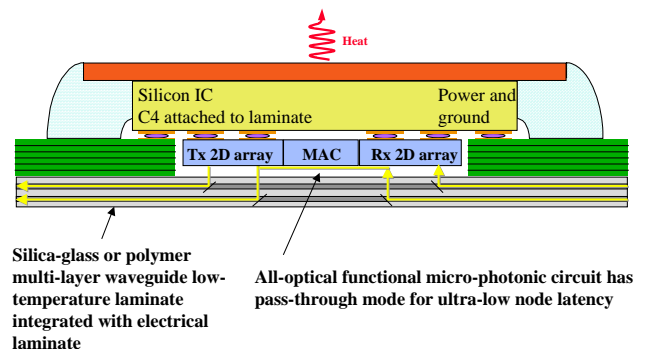


Fig. 7. Schematic cross-section illustrating one possible approach to integration of CMOS with all-optical MAC for a SAN.

While component research indicates that optics can provide the massive bandwidth density and enhanced functionality to make systems more efficient than the best efforts of conventional pure electronic systems, system architects

require a wide range of additional criteria to be met. In the past, these criteria may have been used as a barrier (and sometimes as a mere pretext) to prevent the wide-spread use of optics in systems. However, as performance gains from conventional scaling of transistor dimensions begin to diminish, so will some of the reluctance to explore the insertion of advanced opto-electronic interconnects in systems. Photonic component manufacturers and system integrators will have to cooperate to push Moore's Law beyond the year 2020.

### VIII. ACKNOWLEDGEMENT

This work is supported by DARPA under agreement #MDA972-97-3-0008.

### IX. REFERENCES

- [1] Semiconductor Industry Association *National Technology Roadmap for Semiconductors 1997* (Sematech Inc., Austin Texas 1997); <http://www.sematech.org>
- [2] R. A. Nordin, A. F. J. Levi, R. N. Nottenburg, J. O'Gorman, T. Tanbun-Ek, and R. A. Logan, "A System Perspective on Digital Interconnection Technology," *J. Lightwave Tech.*, vol. 10, pp. 811-827, 1992.
- [3] W. S. Hayward, "The 5ESS switching system," *AT&T Tech. J.*, vol. 64, 1985.
- [4] B. Madhavan and A. F. J. Levi, "Low-power 2.5 Gbps VCSEL driver in 0.5  $\mu\text{m}$  CMOS technology," *Electron. Lett.* vol. 34, pp. 178-179, 1998.
- [5] O. Kibar, D. A. van Blerkom, Chi Fan, and S. C. Esener, "Power minimization and technology comparisons for digital free-space optoelectronic interconnections," *J. Lightwave Tech.* vol. 17, pp. 546-555, 1999.
- [6] ANSI X3.230:199x, "Fibre Channel - Physical and signaling interface (FC-PH)," *American National Standards Institute*, August 1994.
- [7] ANSI X3.148-1988, "Fiber Distributed Data Interface (FDDI) - Token Ring Physical Layer," *American National Standards Institute*, Nov. 1988.
- [8] R. A. Nordin, W. R. Holland, M. A. Shahid, "Advanced Optical Interconnection Technology in Switching Equipment," *J. Lightwave Tech.*, vol. 13, pp. 987-994, 1995.
- [9] G. M. Yang, M. H. MacDougall, and P. D. Dapkus, "Ultra-low threshold vertical cavity surface emitting lasers obtained with selective oxidation," *Electron. Lett.*, vol. 31, pp. 886-888, 1995.
- [10] D. G. Deppe, D. L. Huffaker, H. Y. Deng, Q. Deng, and T. H. Oh, "Ultra-low threshold current vertical cavity surface emitting lasers for photonic integrated circuits," *IEICE Transactions on Electronics*, vol. E80-C, pp. 664-674, 1997.
- [11] T. Satake, T. Arikawa, P. W. Blubaugh, C. Parsons, and T. K. Uchida, "MT Multifiber Connectors and New Applications," *44th Electronics Components and Technology Conference*, pp. 994 - 999, May 1994 (IEEE cat# 94CH3241-7).
- [12] A. P. Kanjamala and A. F. J. Levi, "Subpicosecond skew in multimode fibre ribbon for synchronous data transmission," *Electron. Lett.*, vol. 31, pp. 1376-1377, 1995.
- [13] B. J. Sano and A. F. J. Levi, "Networks for the professional campus environment," *Multimedia Technology for Applications*, Eds. B. Sheu and M. Ismail, chapter 13, pp. 413-427. IEEE Press, Piscataway, New Jersey, 1998.
- [14] J. Silberman, N. Aoki, D. Boerstler, J. Burns, S. Dhong, A. Essbaum, U. Ghoshal, D. Heidel, P. Hofstee, K. Lee, D. Meltzer, H. Ngo, K. Nowka, S. Posluszny, O. Takahashi, I. Vo, and B. Zoric, "A 1.0 GHz single-issue 64b PowerPC integer processor," *IEEE International Solid-State Circuits Conf.*, pp. 230-231, February 1998 (IEEE cat# 98CH36156).
- [15] B. Raghavan, Y.-G. Kim, T.-Y. Chuang, B. Madhavan, and A. F. J. Levi, "A Gbyte/s Parallel Fiber-Optic Network Interface for Multimedia Applications," *IEEE Network* vol. 13, pp. 20-28, 1999.
- [16] R. C. Walker, K.-C. Hsieh, T. A. Knotts, and C.-S. Yen, "A 10 Gb/s Si-Bipolar TX/RX Chipset for Computer Data Transmission," *IEEE International Solid-State Circuits Conf.*, pp. 302-303, February 1998 (IEEE cat# 98CH36156).
- [17] HIPPI-6400-OPT Working Draft T11-1, Project 1249-D, <http://www.cic-5.lanl.gov>, Rev 0.7, *American National Standards Institute*, August 1998.
- [18] NGIO Specification 1.0, <http://www.gioforum.org>
- [19] Future IO Developers Forum, <http://www.futureio.com>
- [20] B. Madhavan and A. F. J. Levi, "A 55.0 Gb/s/cm data bandwidth density interface in 0.5  $\mu\text{m}$  CMOS for advanced parallel optical interconnects," *Electron. Lett.* vol. 34, 1846-1847, 1998.
- [21] CCIR recommendation 601-2, "Encoding Parameters of Digital Television for Studios," vol. XI — part 1, *International Telecommunications Union*, Geneva, 1990.
- [22] C. Chang, G. Czajkowski, and T. von Eicken, "MRPC: a high performance RPC system for MPMD parallel computing," *Software - Practice and Experience*, vol. 29, pp. 43-66, 1999.
- [23] G. Buzzard, D. Jacobson, M. Mackey, S. Marovich, and J. Wilkes, "An implementation of the Hamlyn sender-managed interface architecture", *Oper. Syst. Rev.*, vol. 30, pp. 245-259, 1997.
- [24] S. Zeadally, G. Gheorghiu, and A. F. J. Levi, "Improving end system performance for multimedia applications over high bandwidth networks," *Multimedia Tools and Applications* vol. 5, pp. 307-322, 1997.
- [25] D. A. Patterson and J. L. Hennesy, *Computer architecture: a quantitative approach* San Francisco, CA: Morgan Kaufmann, 1996.
- [26] <http://www.sgi.com/origin/numa.html>
- [27] T. Pinkston, <http://www.usc.edu/dept/ceng/pinkston/>