Architectural approach to the role of optics in monoprocessor and multiprocessor machines

Jacques Henri Collet, Daniel Litaize, Jan Van Campenhout, Chris Jesshope, Marc Desmulliez, Hugo Thienpont, James Goodman, and Ahmed Louri

The relevance of introducing optical interconnects (OI's) in monoprocessors and multiprocessors is studied from an architectural point of view. We show that perhaps the major explanation for why optical technologies have nearly been unable to penetrate *into* computers is that OI's generally do not shorten the memory-access time, which is the most critical issue for today's stored-program machines. In monoprocessors the memory-access time is dominated by the electronic latency of the memory itself. Thus implementing OI's inside the memory hierarchy without changing the memory architecture cannot dramatically improve the global performance. In strongly coupled multiprocessors the node-bypass latency dominates. Therefore the higher the connectivity (possibly with optics), the shorter the path to another node, but the more expensive the network and the more complex the structure of electronic nodes. This relation leaves the choice of the best network open in terms of simplicity and latency reduction. The bottlenecks resulting from and the benefits of implementing OI's are discussed with respect to symmetric multiprocessors, rings, and distributed shared-memory supercomputers.

OCIS codes: 200.4650, 200.2610.

1. Introduction

Although numerous studies are in progress worldwide for developing short-distance optical interconnects (OI's), it clearly emerges from the literature that most of them are based on technological arguments and that global operation of the targeted architecture has not been fully analyzed. The proceedings of the conferences that constitute Refs. 1 and 2 provide a nonexhaustive presentation of the current state of the art in this field. Many studies attempt to improve some part of the machines but offer no assurance that this progress improves the global operation of the whole system.

In this study we follow a different approach, which consists of analyzing the role of short-range OI's in monoprocessor and multiprocessor machines from an architectural point of view. Most of the discussion throughout the paper focuses on the relevance of OI's in reducing memory-access latency (MAL), which is the most critical and permanent issue in stored-program computer architectures. The consideration of optics leads to the following paradox: On the one hand, OI's extend the communication bandwidth but generally do not *directly* change (or address) the access latency to the memory, which is dominated by electronic processing times in both monoprocessors and tightly bound multiprocessors. On the other hand, OI's may increase the interprocessor network connectivity (thus reducing the path that separates remote nodes) at the expense of shifting most of the latency problem to the electronic domain and, in particular, to the design of efficient node circuits. The

J. H. Collet (collet@laas.fr) is with the Laboratoire d'Analyse et d'Architecture des Systèmes du Centre National de la Recherche Scientifique, 7 avenue du colonel Roche, F-31077 Toulouse, France. D. Litaize (litaize@irit.fr) is with the Institut de Recherche en Informatique, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse, France. J. Van Campenhout (jan.vancampenhout@rug.ac.be) is with the Vakgroep Universteit Gent, St. Pieternieuwstraat 41, B-9000 Gent, Belgium. C. Jesshope (c.r.jesshope@massey.ac.nz) is with the Institute of Information Sciences and Technology, Massey University, New Zealand. M. Desmulliez (m.desmulliez@hw.ac.uk) is with the Department of Computing and Electrical Engineering, Heriot-Watt University, Riccarton EH14 4AS, UK. H. Thienpont (hthienpo@vub.ac.be) is with the Laboratory for Photonics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. J. Goodman (goodman@cs.wisc.edu) is with the Department of Computer Sciences, University of Wisconsin, Madison, Wisconsin 53706. A. Louri (louri@ece.arizona.edu) is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, Arizona 85721.

Received 26 May 1999; revised manuscript received 8 September 1999.

0003-6935/00/00671-12$15.00/0

choice of the best network is therefore an open issue in terms of implementation simplicity and latency reduction. These difficulties may explain in part why OI's today are not deployed in general-purpose machines and are considered for use potentially in dedicated processors (that do not execute instructions or fetch data stored in a memory).

The contents of this paper are as follows: The current state of the application of optical technologies to communication and interconnection networks is reviewed in Section 2. The memory issue and its influence on the architecture of today's machines are reviewed in Section 3. In Section 4, we discuss the relevance of implementing OI's in monoprocessors. No dramatic increase in global performance is expected in such systems, as the intrinsic memory latency is the dominant latency. The potentially stronger impact of OI's in multiprocessor machines is discussed in Section 5. Simplicity of implementation is often preferred for small multiprocessor machines, making the introduction of OI's particularly attractive in symmetric multiprocessors (SMP's) and ring architectures for connecting approximately 100 nodes. In these architectures OI's can provide a huge bandwidth that can minimize contention latency (related to the traffic saturation), as is also explained in Section 4, while they maintain the simplicity of the node structure. In supercomputers, providing a global shared view on a physically distributed memory places a heavy burden on the interconnection network and, in particular, on the development of low-latency high-connectivity electronic nodes. The introduction of OI's in new reconfigurable architectures and in dedicated processors is discussed briefly in Sections 6 and 7, respectively.

2. Brief Review of the Role of Optics in Communications Networks

The current state of the competition between optics and electronics for the processing and the transmission of information is reviewed here as a function of the communication distance.

A. Telecommunications Networks

Optical communications have won the battle for long-distance transmission in wide-area networks (WAN's) and metropolitan-area networks (MAN's). There are at least three reasons for this: (1) The bandwidth limitation of OI's is much less pronounced than that of electrical transmission,[3] losses are much lower, and in future systems the effects of nonlinear dispersion can be countered by use of solitons. (2) Parallel transmissions are not usable over long distances because skew makes the synchronization of the different reception channels particularly complex. (3) Multiwavelength (optical) transmissions make possible the extension of the transmission bandwidth at almost no cost to the network infrastructure. Thus one permanent objective of long-distance communications consists of increasing the transmission bandwidth through a single monomode fiber.

B. Local-Area Networks

Local-area networks were first designed for data transmission between computers. We can distinguish between company networks and industrial networks that operate in a hostile environment with real-time constraints. Each computer (PC, workstation, etc.) in a company network is connected to a hub through a few tens of meters of links and operates with an ethernet protocol. Hubs themselves are interconnected by means of high-throughput links that operate under various protocols such as the ethernet, the fiber-distributed data interface, and ATM. Links from computers to hubs generally are implemented with the preinstalled metallic cables of the building network, whereas serial optical (i.e., fiber-distributed data interface) links are used mostly for connections between hubs. Industrial networks also exhibit a hierarchical structure. Each level may use a specific field bus, such as the Profibus,[4] the Fieldbus,[5] the controller area network (developed by Bosch for cars[6]), aviation industry standards like the Aeronautical Radio, Inc., Model 429 or Model 629 (developed by the Airline Electronic Engineering Committee), the manufacturing automotive protocol (developed by General Motors), or the interbus S.[7]

C. Short-Distance Communications and Interconnects

The transition from serial telecommunications to the computer world (dominated by parallel interconnects) occurs at short distances that extend over a few meters. A large bandwidth is needed for computer clusters and multiprocessor interconnects, as, for instance, in the CRAY Model T3E,[8] the IBM Model SP2,[9] the Intel Model Paragon,[10] and the Silicon Graphics Model Origin systems.[11] Electronic parallel interconnects dominate because they allow, across a few meters, the extension of the global bandwidth without increasing the operation frequency. These interconnects are generally a cheap solution. However, several networks that initially were implemented with paralleled electrical links now offer serial (or parallel) optical alternatives for increasing the bandwidth and the transmission distances. Some of these networks are HIPPI (high-performance parallel interface) at 6.4 Gbits/s,[12] SCI (scalable coherent interface) at 1.6 Gbits/s,[13] and Myrinet at 1.28 Gbits/s.[14] These networks make possible the building of multicomputers for supporting cluster computing, which is an area in which there is much experimentation at the moment. Cluster computing is motivated partly by the preoccupation of developing of more modular, low-cost hardware that would simplify maintenance and compatibility issues for manufacturers. However, it must be stressed that cluster computing is suitable for some but not all applications because the latency of internode communications becomes extremely long (with respect to the processor cycle it is currently 1 ns) when the intercomputer distance attains a few meters (1 m translates to 5 ns). Therefore a distributed system will execute numerous applications much more slowly

(especially applications with a distributed memory and those that require numerous internode exchanges) than does a tightly bound multiprocessor enclosed in a single cabinet.

D. Ultrashort-Distance Interconnects

At present ultrashort-distance interconnects ranging from a few centimeters to a few tens of centimeters are in the electronic domain. The machines under consideration here are monoprocessors (PC's and workstations) or SMP's such as the Silicon Graphics Model Power Challenge[11] and the Sun Model Enterprise.[15] In these machines the communication latency is never controlled by the propagation (internode or interunit) but by electronic terms (memory latency, routing time, etc.). This control constitutes an essential difference from the distributed systems described in Subsection 2.C. The bandwidth extension has been achieved to date by the increasing of transmission parallelism and by the replacement of shared buses (which are multipoint electrical lines) with dedicated point-to-point parallel interconnections. For instance, in the Pentium architecture data are transmitted between the memory controller and the chip set through a 64-bit-wide bus at 100 MHz and possibly will be transmitted in the future with 128 bits to attain 12.8 Gbits/s.[16]

E. Intrachip and Multichip-Module Interconnects

By far most of the communications at this level are in the electronic domain. However, an optical clock distribution exists at the interboard level in the CRAY Model T-3D (Ref. 17), and research is being carried out to extend this technology to the intraboard level for the CRAY Model T-90.[18] On the research side studies are in progress for the construction of optical backplanes[19,20] and on micro-optical components for interchip and intrachip communications, for example, at Vrije Universiteit Brussel.[21,22]

F. Optics in Logic-Level Processing

Most of the all-optical computing studies launched in the middle of the 1980's have been reoriented toward special-purpose systems because (1) the dramatic increase in electronic processing power has progressively eliminated many arguments that favored optical binary processors (switching time, switching energy) and (2) the performance of optical numerical demonstrators comparatively has stagnated. Today, interest in optical processing seems to be limited to dedicated processors.

Parallel optical I/O can be exploited for information transport in smart pixels, early vision processing, artificial retina applications,[23] and database and symbolic applications.[24] The low-level processing in the fields of pattern and image recognition might take advantage of the potential of optics for the extremely fast implementation of simple Boolean functions, as has been demonstrated by recent experiments that were carried out in the field of ultrafast optics.[25]

3. Optical Technologies and the Memory Issue

A. General Remarks

In Section 2, we showed that OI's have ruled the telecommunications world by the replacement of electrical links with optical serial links without affecting the communication protocols or the network architectures. They will likely invade local-area networks similarly. The insertion of OI's *into* computers is much more problematic despite numerous advantages that favor optical technologies.[26] Some of these arguments are listed below:

- Optical technology can provide extremely high bandwidth almost independently of the interconnect length at the length scales considered.
- OI's have advantages in terms of weight and volume. The interconnect density is potentially higher because of (1) the much lower interference of optical signals, either in free space (electron beams versus light beams) or when guided (mutually interfering conductors versus optical fibers), and (2) the fact that each optical communication channel requires a cross section of the order of the wavelength of the light used. Hence per given total cross section and interconnect length a larger number of interconnects is possible optically. Connection of high-density OI's is much less bulky than that of electrical interconnects.
- Broadcasting is feasible because of the capability of high fan-out. However, high fan-out requires high power, which introduces some limitations, such as an increase in the latency.
- Wavelength-division multiplexing can be used to increase bandwidth and achieve special advantages. Although it requires tunable light sources, there is no interference between wavelengths; in terms of networking add–drop capabilities and wavelength routing are possible.
- Optics is at least competitive in terms of power- and control-supply requirements: The speed–power product of advanced centimeter-range OI's is becoming better than that of electrical ones.[27–29]
- Optics may be capable of the rapid reconfiguration of static interconnection patterns. Light polarization also makes possible switching and fast configuration routing. Apart from power losses the means needed for optical reconfiguration do not impair signal quality or even latency, as is often the case with electrical switching or reconfiguration.
- OI's provide galvanic isolation between interconnected subsystems. This isolation leads to improved noise immunity and security for applications that monitor high-voltage systems.

With all the above arguments, how does one explain the paradox that optical technologies have so far hardly gained application in computers? We feel that the underlying reason is that optics generally does not directly shorten the access time to the memory, which is the most crucial issue for stored-program machines. Therefore it is extremely

difficult to translate the numerous technological arguments that claim optical advantages into an optical architecture (or a demonstrator) that might effectively overcome electronic computers (except for some dedicated applications that are related to vision or image processing). We briefly review the memory problem in Subsection 3.B.

B. Memory Issues and the Architectural Evolution of Machines

The evolution of stored-program computers has repeatedly been influenced by the fact that the performance of microprocessors has increased much more rapidly than has that of memory systems. In early microprocessor systems (i.e., in the 1970's), processors operated at approximately the same rate that memory could be cycled, and the processor was connected directly to the memory system [dynamic RAM[30] (DRAM)]. This is no longer the case. Whereas processor speeds have been doubling every few years, dynamic memory has increased in speed only marginally over the past two decades, although its size has also doubled every few years. This lack of improvement in speed means that the time needed by the processor to fetch instructions or data from the main memory has increased permanently compared with the processor-cycle time, making direct exchanges with the main memory more and more penalizing for the *global* performance of the computer.

Latency is the key parameter of memory–processor interactions (much more so than the bandwidth, as in telecommunications) because the processor exchanges very short bursts of information (usually at least one word, i.e., 4 bytes, and more often a cache block at a time, i.e., 32 or 64 bytes). The processor never establishes a steady communication stream with the memory. The major limitation to the speed of DRAM is in the circuits used for detecting the stored charge on a memory cell. There is a trade-off between the size of the memory and the rate at which this tiny stored charge can be sensed. Static RAM (SRAM), on the other hand, is optimized to be significantly faster than DRAM, although SRAM speed is achieved at the cost of a larger memory cell and therefore a significantly reduced memory size (or an increased memory cost). For this reason most desktop and server systems use DRAM memories to maintain large memory size at modest cost. The architecture of chips and machines has evolved permanently to maximize global performance despite the degradation of the MAL. A typical example is shown in Fig. 1 for the PC architecture. The evolution of PC's and multiprocessor machines has consisted primarily of *hiding* the MAL with hardware or software solutions because it has been impossible to change the memory technology. Therefore

- Modern computer architectures use a complex hierarchy of memories. Two on-chip level 1 (L1) caches are provided in pipelined architectures—one for data and one for instructions—because data and instructions must be read concurrently. These caches typically are 32–64 kbytes in size (128 kbytes are expected in the next processor, Model K7, from AMD). L1 caches can then be connected to a level 2 (L2) cache, also on chip, or to a much larger off-chip cache. The latter typically is approximately 1 Mbyte in size. If there is a L2 on-chip cache, there may also be a level 3 (L3) cache off chip. The off-chip cache will then be connected to the main memory. Caches work by exploitation of the locations of references in access to data, either spatially when data adjacent to those recently used are likely to be reused or temporally when data recently used are likely to be used again. The aim of a cache is to make the memory system appear to be as large as the largest component and to appear as fast as the fastest component. Unfortunately, when the cache system does not work well through a lack of locality the slowdown is severe because the DRAM memory-access time is at least an order of magnitude larger than the processor's cycle time.
- Current microprocessors are designed to *hide* the latency associated with a memory fetch. Techniques used to tolerate high-latency memory include speculative execution in which the results of branches and even data values are predicted. When a misprediction occurs, data generated along wrong branch paths or based on mispredicted values must be cleaned up and the original state restored. Another technique used is out-of-order execution in which instructions start and even terminate before previous instructions in the instruction stream. Operands of instructions that have been completed out of order must be held in renaming buffers prior to being retired. Thus operands to dependent instruction must then be retrieved from these registers and not the registers indicated by the instruction. This process requires tables for register-renaming results that have not been retired as well as for memory for data-flow matching to determine which instructions

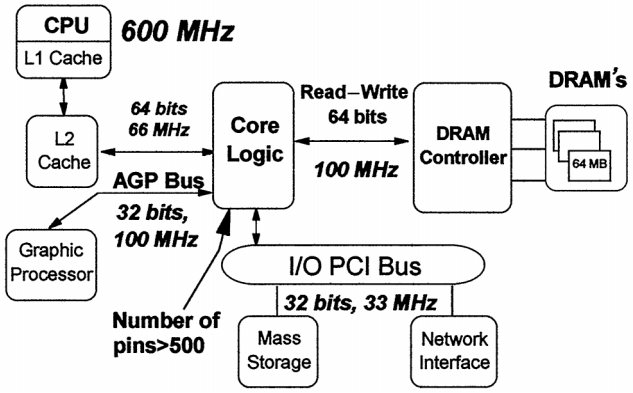

Fig. 1. Current PC architecture: Note the two levels of caches (L1 and L2) and the absence of the shared-memory bus, which is replaced with dedicated point-to-point connections between the logic core, the graphic processor, the DRAM controller, and the L2 cache. Note also that the number of pins in the core logic is greater than 500! AGP, accelerated graphics port; PCI, peripheral-component interface; MB, megabytes.

can be executed. The prediction and the clean-up logic and the additional registers and tables used in out-of-order execution mean that modern microprocessors are very complex. Less complex mechanisms exist for tolerating high latency into main memory, such as microthreading or multithreading.[31,32]

In addition to these latency-tolerant techniques most processors attempt to issue more than one instruction in a single cycle by use of multiple-execution units. This process is meant to increase throughput for a given clock cycle, again at the expense of complexity. Most recent microprocessors have at least four-way issue but seldom achieve an effective instruction per cycle count of greater than two. These general considerations hold for any computer with specific constraints, depending on the number of processors and on the communication network of the machine.

C. Evolution or Revolution of the Architecture?

Two approaches prevail with respect to the evolution or revolution of architectures (with some possible intermediate points of view):

1. The first approach, which we call the evolutionary approach, consists of trying to integrate optical communication systems in forthcoming machines. This approach requires the analysis of communication bottlenecks in existing or future computers and the capability of optical communications to solve these problems (i) with much more effectiveness than electronic solutions and (ii) with a good chance to reach a cost-effective mass production. Thus this approach tries to predict the role of optical communications in the next 5–10 years, starting with the present state.
2. The second approach, which we tentatively call the mutational approach, considers that optics might induce new computer architectures with outperforming specifications that will justify abandoning (or at least dramatically modifying) existing electronic solutions. This is a more speculative approach about the possible long-term evolution of computer architectures. Note that it does not release designers from having to know quite well the state of the art of existing electronic architectures that cannot be reduced to the pure communication aspects if they are to propose and demonstrate the advantages of the new optical solutions.

In the rest of the paper, we follow the evolutionary approach, as it is much less risky than the mutational approach and can be used as a reference for appreciating the relevance of more advanced proposals. Architectures are analyzed in ascending order of the number N of connectable processors. Although hundreds of network topologies have been proposed, there are only a handful of commercial implementations, which reduce to mostly buses, rings, meshes, tori, and central switches. Thus we begin with the monoprocessor. Then we consider SMP's (say, typically $2 < N < 32$–64), rings (say, potentially $10 < N < 100$), and supercomputers that could connect up to several thousands of processors.

4. Optical Interconnects in Monoprocessors

What could be the role of optical interconnects in monoprocessor machines? The registers are linked to L1 caches, the L1 to the L2 cache, the L2 to the memory or in some machines to the L3 cache, and the L3 cache to the memory. The L1 and the L2 caches, which are often integrated in the processor chip, are built with SRAM's and can operate at the processor frequency. Inserting OI's at this level (i.e., between the register and L1 or between L1 and L2) increases the transfer latency (because of the optoelectronic conversion time) and degrades processor performance.[33]

Perhaps the integration of optical communications ought to be considered for the longest distances in the memory hierarchy, namely, between the last cache level (considered to be L2 in the following) and the main memory. Two terms contribute to the MAL, namely,

- The *intrinsic* MAL, which is the leading term to the MAL and depends on the internal architecture and on the technology of the DRAM's.
- The communication latency between L2 and the memory.

The need for memory bandwidth in future machines will grow dramatically owing to the increase of processor-operation frequency and to foreseeable architectural evolutions such as the extension of instruction-level parallelism of processors and the use of higher-order nonblocking caches. Today, processors run at nearly 700 MHz and issue as many as 4 instructions/cycle. In the years 2005–2010 operation at nearly 4 GHz is expected with the execution of 32 instructions per cycle, corresponding to a sole instruction bandwidth B_0 (between the processor and L1) of the order of 400 Gbytes/s. The two levels of cache will substantially reduce the main-memory access to a few percent of B_0, but, all in all, a bandwidth in the range of 10–20 Gbytes/s is expected between L2 and the memory! Introducing OI's (operating at approximately 1–2 GHz) might exhibit some advantages here (see the arguments listed in Subsection 3.A), but ultimately the role of OI's will depend on the evolution of chip packaging and motherboard technology. The point is that no major problems are envisioned by electronics engineers in moving out to 1-GHz processors with electronic signaling in monoprocessor and tightly coupled architectures, as explained below.

Neither bus frequency of the order of 500 MHz nor bus width in the 500-pin range is seen as an insurmountable obstacle. The memory bus at this end of the market is not a significant issue, as its cycle time is limited by the DRAM speed. The effects of high latency can be mitigated by the provision of very wide

electrical buses that carry as many as 512 bits (plus error-correction bits) of data simultaneously. Possibly this broad bus might be split into several independent narrower buses (say, 64 bits wide) to make access parallel to different memory banks. These solutions might require new chips with several thousands of pins, but current mass-produced devices including from 2500 to 10,000 pins are already available in research[34] with possibly a role for optical interconnects.

In summary, the possible introduction of OI's (between the registers and L1, L1 and L2, or L2 and the main memory) seems hypothetical because (1) the memory-access time is dominated by the intrinsic memory latency and optical communications (whatever their bandwidth is) do not change this issue, and (2) the bandwidth challenge between L2 and the memory, which is in the range of 10 Gbytes/s, seems accessible to the forthcoming electronic packaging and to the motherboard technology. Electronic solutions will likely suffice for building cheap and efficient monoprocessor machines within the next 10 years.

5. Optical Interconnects in Multiprocessor Architectures

A. General Remarks

The logical way to reach a performance level not accessible with a monoprocessor consists of connecting several processors through an interconnection network, thus building a multiprocessor system. As was stressed in Section 4, the MAL is a critical parameter for the performance of the machine. In multiprocessors it can increase drastically and can be as much as 3 orders of magnitude larger than the processor cycle time in the worst case, in particular, when the memory is distributed in a number of processors (or clusters of processors) and the execution of the application necessitates numerous internode transfers. It is possible to distinguish (at least) five contributions to the latency, namely,

(1) The *intrinsic* memory latency already mentioned in Section 4.

(2) The *software* latency (communication overhead) associated with formatting, sending, and receiving messages. It is not clear at this time how optics might reduce software latency. A comprehensive study of the effect of the high communication bandwidth capability on the overall communication latency has not yet been done. Are there possible architectural innovations in interprocessor communications with OI's that will eliminate or largely diminish the effect of software latency?

(3) The (network) *propagation* latency, which depends on the network topology and the processing overhead for routing and solving contention problems. This latency is discussed extensively below.

(4) The (network) *contention* latency, which critically depends on network saturation. The network latency is the sum of the propagation and the contention latencies.

(5) The *coherence* latency in which maintaining the coherence of the caches in tightly bound machines requires broadcasting (or multicasting) coherence messages through the communication network. This coherence maintenance also contributes to slowing down the memory access. This factor strongly depends on the network topology. Snooping protocols, usually implemented with buses, are much simpler and faster than directory-based protocols that have to be implemented in distributed networks.[30]

Which types of OI's (topologies, technologies, packaging schemes, etc.) are the most suitable in the short term as well as in the long term? First, it is clear that network latency is the dominant term in existing multiprocessors (in a monoprocessor this is the intrinsic memory latency). Second, the network latency in strongly coupled multiprocessors is dominated by the bypass time of electronic nodes and *not* by the internode-propagation time (IPT). This is a key difference from the telecommunications networks because of the short internode distance in the systems under consideration [i.e., a few tens of centimeters (see Subsection 2.D)] with an IPT in the range of a few nanoseconds. Optics can provide high connectivity, but increasing the connectivity generates new issues and must be used parsimoniously, as discussed below:

- When the connectivity is low [1 for the unidirectional ring shown in Fig. 2(a)] node processing is simple (only the add–drop of information in the network, possibly error detection and correction, and priority treatment) but is still longer than the IPT. The drawback of unidirectional rings is that the number of nodes a message must pass through in its round trip to remote memory (RTRM) equals the total number of nodes.
- Increasing the connectivity is therefore particularly attractive {by use of meshes, tori, hypercubes, etc. [see Fig. 2(b)]} because the average internode distance decreases. Unfortunately, two new issues arise:

(1) The average internode distance generally decreases sublinearly versus the connectivity, whereas the network complexity increases linearly. Thus increasing the connectivity sooner or later becomes prohibitively expensive. Let us consider an example for clarity, namely, the connection of $N = 256$ nodes with k-dimensional bidirectional tori. The average internode distance is $D(k) \approx (k/4)N^{1/k}$, and the number of connections is $N_C(k) = Nk$. For $k = 2$ [a two-dimensional (2-D) mesh], $k = 3$ [a three-dimensional (3-D) torus], and $k = 4$, we get $D(2) \approx 8$, $D(3) \approx 4.8$, and $D(4) \approx 4$, respectively. Therefore, from $k = 2$ to $k = 3$, D is divided by 1.66, and from $k = 3$ to $k = 4$ by 1.2. But simultaneously the number of connections is multiplied by 1.5 and 1.33, respectively,

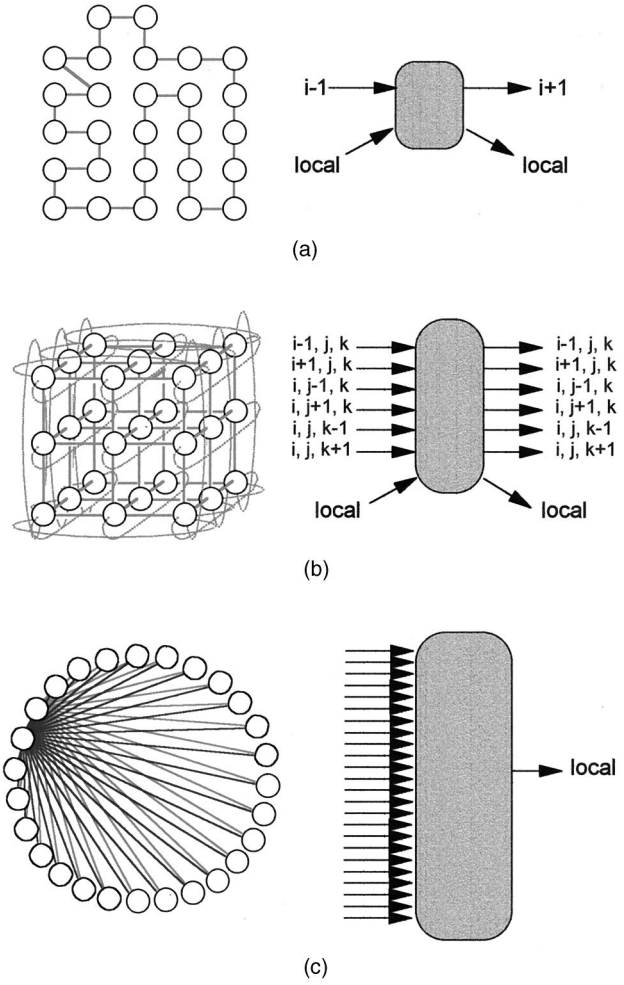

Fig. 2. Three networks for connecting $N = 27$ nodes: (a) A unidirectional ring that requires that all the nodes (27) be bypassed in a RTRM. The node is a 2×2 switch. (b) A 3-D torus with an average RTRM of approximately 4. This second network is approximately 7 times faster in terms of the hop number than that shown in (a), but the node becomes a 7×7 switch. The increase in the node-bypass time depends critically on the switch design. (c) A fully interconnected network. For clarity, the connections of only two nodes are drawn. The RTRM reduces to 1 hop, but the input structure of each node becomes an N-to-1 multiplexer that must operate in the asynchronous mode to reduce the MAL.

showing that increasing the connectivity becomes more and more costly.

(2) The node-bypass latency $T(k)$ increases with the connectivity because of the increase in node complexity. The increase of $T(k)$ depends critically on how sophisticated the implementation of the node switch can be, depending on the operation frequency, the parallelism of transmissions, the communication protocol, the routing algorithm, the admissible cost, etc. A simple solution consists of decomposing a k-dimensional switch in k successive one-dimensional (1-D) switches optimized for straight traffic. In that case the average internode latency $L(k)$ of bidirectional tori scales as $L(k) \approx D(k) + (k - 1) = (k/4)N^{1/k} + (k - 1)$, where $D(k)$ is the average number of bypassed nodes. The second term accounts for the increase of the note latency. Again, with $N = 256$ and $k = 2, 3, 4$, we deduce that $L(2) \approx 9, L(3) \approx 6.8$, and $L(4) \approx 7$, respectively, showing that increasing the connectivity from three to four does not shorten the average internode latency. The conclusion here is that the topological arguments on the benefits of increasing the connectivity in strongly coupled multiprocessor networks cannot be separated from analyzing the complexity of the node-routing implementation.

- If the network is fully interconnected [Fig. 2(c)] the internode distances are minimal. However, this network topology transfers most of the latency-reduction challenge to the design of the node-input circuits that must multiplex N asynchronous inputs and serialize access to node memory. Preserving the memory-ordering constraints also seems to be a particularly complicated issue for a fully connected shared-memory machine. Perhaps the input block of *each* node of a fully connected network might be seen as a common bus shared by the N inputs with a snooping protocol to maintain the coherence.

Therefore finding the connectivity that provides the lowest network latency (i.e., the propagation latency in our classification at the beginning of this subsection) for a given number of N nodes is an open and a complicated problem.[35] When N is very large (say, several thousand, as in a supercomputer) the hypercube topology may be attractive because the connectivity scales as $\log_2 N$ (see Refs. 36 and 37 for comparisons with other topologies). But this solution is obviously expensive. For instance, with $N = 1024$, twelve links per node chip are necessary, i.e., $\log_2 P + 2$ links/chip, requiring approximately 800 pins for a transmission parallelism of 64. For smaller machines, say, N less than 64 or 128, simplicity may be favored by the use of shared buses or rings. The latencies of the memory or the network are purely electronic. OI's can contribute to eliminating the contention latency by the provision of a huge bandwidth. This can be decisive for machines based on 1-D interconnection networks such as SMP's or rings (see Subsections 5.B and 5.C).

The scalability of network parameters is hardly predictable in the long term. It is clear that the operation frequency of electronic nodes will increase, whereas the IPT is incompressible. Thus the IPT might become the leading contributor to remote-access latency, with a corresponding increase in memory-access time (in terms of the electronics cycle) with strictly no hope of reduction. This long-term evolution would transform any strongly coupled multiprocessor into a weakly coupled machine. However, this situation is not inexorable, as it is based on the sole scalability of the operation frequencies of processors and nodes. It is likely that it will be accompanied by a reduction of dimensions (a possible metric being, for instance, the processor power divided by the processor volume) so that the processor cycle and the internode distance would diminish si-

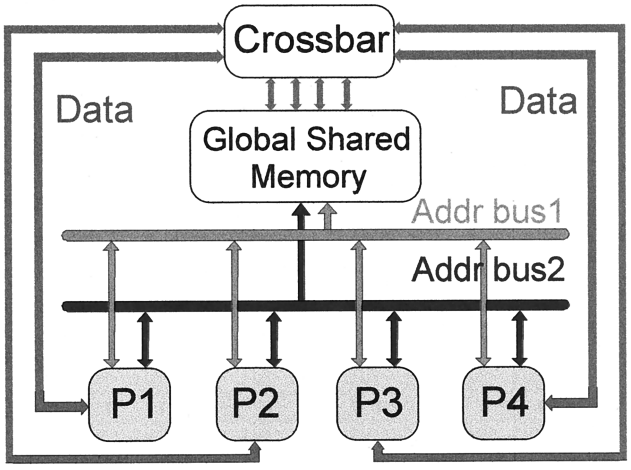

Fig. 3. Modern SMP architecture: Two address buses (Addr bus1 and Addr bus2) access the memory while preserving the coherence of the caches.

multaneously, maintaining the preeminence of the node-bypass latency over the IPT.

B. Symmetric Multiprocessors

A SMP is a physically shared-memory machine with a uniform memory-access time. Early machines comprised a small number of processors, e.g., not exceeding 32, connected to a memory system through a single multiplexed shared bus.[38] The architecture has evolved, and the number of nodes is larger today than was expected just a few years ago, with as many as 64 processors for the Sun Microsystems Model Enterprise 10,000.[39]

Interconnect links for data and addresses have been separated in modern machines (see Fig. 3). Today the solution for increasing the *data* bandwidth consists of connecting each processor by a private link to a crossbar and then connecting the crossbar to the memory. But the necessity of preserving the coherence of the cache makes this method unsuitable for addresses. Therefore the shared-address bus has become a critical communication bottleneck of the SMP architecture because it serializes accesses to the memory and adds an important contention latency to the MAL. The greater the number of processors, the longer the contention latency. The solution, which would consist of increasing the bus-operation frequency, is particularly complicated because the bus is a multipoint electronic line.

The only palliative solution has so far consisted of duplicating the number of address buses to reach the needed bandwidth, each bus being attached to an address-memory range. In addition, bandwidth can be increased by use of the notion that a logical bus can be implemented with a tree structure. This approach seems to be impractical for large SMP's (i.e., for SMP's with 64, 128, or more processors) and makes optical solutions attractive. The simplest technological change might consist of replacing each electrical *address* bus with an optical bus that connects the processors to several interleaved memory banks. This approach is attractive for three major reasons:

- Bus operation to as high as 1 GHz (or higher) would become possible (by replacement of the electrical bus operating at nearly a few tens of megahertz) because the transmission of optical pulses in guides is not penalized by capacity effects and critical-load adaptations that are encountered for electrical transmissions in a multipoint line. As a result the SMP architecture (i.e., the processor, the bus, and the memory) would become more scalable.
- Parallel transmission through optical lines is almost skew free in the gigahertz domain for transmission over a few tens of centimeters. This simplifies data recovery in the case of parallel transmission.
- The introduction of such an optical connection basically would not change the bus operation, which would always rely on the access serialization and a snooping protocol to maintain the coherence of caches.

However, several severe limitations cannot be ignored, namely,

- The top transaction efficiency of a shared bus is close to 1 transaction/bus cycle with pipelined arbitration. This limit becomes a bottleneck for large SMP's with 32 or more processors.[40] Speeding up the bus will surely improve the machines' operation but will not solve all the contention problems of large multiprocessors, as it seems unrealistic to assume that the bus might operate faster than the processor. Large SMP's with several optical buses seem inevitable.
- The bus-operation frequency cannot be increased without ensuring that each cache controller is able to check and update its directory at the bus-operation frequency.
- Speeding up the bus will sooner or later generate an integration issue because the bus length ought to be limited to make sure that the optical signals can be stationary within a single bus period. The light-propagation velocity ($c = 20$ cm/ns) requires the bus to be shorter than 10 cm at 2 GHz, shorter than 5 cm at 4 GHz, etc. This size constraint disappears if more than one transaction can be in progress simultaneously in the bus. In that case the bus architecture is akin to that of rings, as described in Subsection 5.C.

C. Rings and Hierarchical Rings

Ring multiprocessors are distributed-memory machines with nonuniform memory-access time.[41] The ring is a *multiaccess* interconnection topology that is attractive for the following reasons:

(1) It permits the use of simple interfaces because the ring connects to a given node by means of only one input and one output port. The node–ring interface is basically a 2×2 switch. This simplicity reflects itself in a relatively low requirement for the number

of connecting wires, which often corresponds directly to the number of pins on physical connectors. The number of connections is considerably smaller than in 2-D or 3-D network topologies (torus, mesh, hypercube) in which more than one direction exists for incoming and outgoing signals, inducing a larger overhead for processing routing.

(2) It provides a natural broadcasting and multicasting mechanism. This feature can be exploited in the implementation of producer-driven prefetching of data, which can improve performance significantly. Unlike with the bus, it is not possible to order parallel messages between different pairs of nodes, and for most implementations flow control can violate the ordering constraints. Thus the ring structure preserves a *partial* ordering of transmitted packets that can be exploited for implementation of a cache coherence scheme.[42]

(3) There are point-to-point connections between successive nodes that do not suffer from the undesirable effects such as loading and signal reflections from multiple connectors that plague electrical bus-based schemes and effectively reduce their feasibility to small sizes (see Subsection 5.B). Therefore signals can be transmitted on such links at high clock rates. Operation at 3–4 GHz with a parallelism of 246 will provide a huge bandwidth close to the terabit-per-second range.

The effective bandwidth is determined essentially by the transfer rates attainable at individual nodes, and it can be improved by an increase in the clock frequency or the width of the transfer path. The bandwidth of a multiprocessor interconnection network can also be increased by means of a hierarchical structure whereby a number of localized transfers can take place concurrently on several rings. For example, if several local rings are connected by means of a central ring the number of concurrent transfers that can be supported is much higher if the transfers are between only sources and destinations on the same local ring. Transfers that pass through the central ring take more time than local ring transfers, but they are, in general, shorter than they would be if all nodes were connected to a single long ring.

D. Supercomputers and Distributed Shared-Memory Systems

Top-of-the-range supercomputers use more than 1000 processors. Although the memory may be distributed, each processor can access all the memory in the system. The distributed-shared memory (DSM) architecture attempts to provide a single addressing space for the distributed memory to enable the user to gain transparent access to computational resources in scalable systems. One achieves this by hiding the remote-communication mechanisms from the application writer, thus preserving the programming ease and the portability of shared-memory systems. Additionally, the scalability and the cost effectiveness of underlying distributed-memory systems are also inherited.

However, local memory is accessed much faster than is remote memory in which messages have to be exchanged across the network to fetch data. The nonuniformity is due not only to the network topology but also to the packaging technologies and will be degraded substantially by heavy traffic loads and congestion. Additionally, the reliance on locality and memory allocation requires heavy caching of memory to reduce remote references. Unfortunately, caching shared data introduces the problem of cache coherence, the solution of which relies significantly on the efficiency of the interconnection network. As was stressed in Section 5, designing low-latency nodes is also a critical issue. The problem will get worse with advances in the speed of current microprocessors in which the price–performance advantage of microprocessors is increasing. To build a scalable DSM multiprocessor that utilizes state-of-the-art off-the-shelf microprocessors with gigabyte-per-second connections to local memory, one must utilize technology that supports interprocessor connections at least in the gigabyte-per-second range and average access times to shared data in the nanosecond range. OI's could be the only cost-effective technology for internode communications.

6. Optical Communications in Reconfigurable Architectures

In nonconventional architectures (such as custom computing) in the reconfigurable-computing domain increasing use is being made of arrays of field-programmable gate arrays (FPGA's). High interconnect density is critical because of the difficulty in finding (or the nonexistence of) natural, weakly interconnected partitioning of the functions. Furthermore, the on-chip interconnect facilities are relatively slow owing to their configurability. This slowness is not a passing phase that depends on integration level but will persist as the density of chips grows continuously. OI's may have a role in FPGA arrays, essentially to speed them up, by the introduction of a new routing layer. The density of optical chip-to-chip I/O may be applicable here even for very short distances such as adjacent chips. If OI's can be justified for this purpose, they may even be feasible for replacing the relatively slow electrical communications within a single chip. Implementing wide buses from the FPGA chip to the reconfiguration memory to achieve rapid reconfiguration could be of benefit in nonconventional architectures.

7. Dedicated Optoelectronic Processors

Dedicated processors are very different from general-purpose monoprocessors or multiprocessors because they generally do not execute stored programs. They are designed for a specific task, which is often related to the processing of optical information. The MAL extensively discussed in Sections 3–6 is no longer a problem (as there is no memory or almost no exchange with the memory). Dedicated optoelectronic processors traditionally have an optoelectronic front-end and back-end interface. The optical data

streams impinge on photodetectors that convert the light intensity of beams into electronic signals that are amplified and processed electronically. The resulting processed data can be converted back into optical signals for further processing if necessary. The number of optical channels can range from 10 to nearly 10,000 over a 1 cm $\times$ 1 cm chip area.

The communication bandwidth becomes a critical issue. For example, in a vision machine a 1024 $\times$ 1024 correlation on a 1024 $\times$ 1024 image requires approximately 170 $\times$ 10^6 multiply-and-accumulate operations, which corresponds to 10^{10} operations/s at a video frame rate of 30 frames/s. In the same way, a matrix multiplication of 1024 $\times$ 1024 requires approximately 10^9 multiply-and-accumulate operations, which corresponds to 6 $\times$ 10^{10} operations/s for the same frame rate. General-purpose electronic machines cannot cope with the input–output needs of such computationally intensive applications. Optically interconnected electronic chips have been shown to be the only technology to date that is capable of providing a match between computationally intense applications.[43] There is a case therefore for dedicated optically connected electronic information-processing systems for applications that require a high data bandwidth capability. Such applications range from image-processing-primitive operations (Fourier transformation, 2-D convolution and correlation, and dot-product and dot-matrix multiplications) to switching fabrics in telecommunications.

Demonstrators based on optically interconnected electronics deal with such functions. See, for example, the fast-Fourier-transform machine built at the University of North Carolina, Charlotte, North Carolina, which has an I/O bandwidth of 29 Gbytes/s and calculates a 1024 fast Fourier transform in a few microseconds,[44] and the bitonic sorter at Herriot-Watt University, Riccarton, UK, which sorts 1024 15-bit-deep words within 10 μs.[45]

8. Conclusion

The most critical issue in computer architectures (from the monoprocessor to large multiprocessor systems) is the access time to the main memory. This is the key problem that architectures must live with, regardless of whether they are optoelectronic. Although memory chips have become much denser (and therefore much larger), they have not become significantly faster. Furthermore, the techniques that currently are used to realize large memory systems suitable for multiprocessors create coherency problems for which no simple, well-scaling electrical solutions are known. This situation further aggravates the latency properties of complex memory systems. It also makes the processor architecture complex, as many techniques in the processor are specifically targeted at the problems of memory latency and bandwidth limitations as well as the unpredictability of hierarchical memory systems. Thus designing new low-latency memory chips is a critical challenge. The changes are open, possibly at the physical level (with the introduction of new materials, for instance, superconductors), at the architectural level regarding the organization of memory (for instance, with the design of multiported memories), or alternatively by the cooling down of current memory chips to approximately the liquid-nitrogen temperature.

With respect to future monoprocessor machines, the possible introduction of OI's in the memory hierarchy seems hypothetical, as it is likely that the evolution of the electronic packaging and the motherboard technology will suffice to build efficient machines within the next 10 years.

With respect to multiprocessor machines, the situation is more favorable, but a major problem is the economic risk factor of introducing a new technology. This is a strategic rather than a technological issue, but unless optics can solve a major problem and provide a significantly better solution at a cheaper cost no one is going to take the risk of an untried technology.

Introducing optics is conceivable in several types of multiprocessor machines. For instance, the supercomputer line has traditionally been the first to experiment with novel techniques because the economic risk may be acceptable in the manufacture of a supercomputer for which performance is the primary issue. But one might also consider developing systems that exploit affordable technology with the basic idea that the commercial success of multiprocessors will depend not only on their computational capability but also on their cost–performance ratio. Successful products might be those that would allow configuration of a feasible entry-level machine at a correspondingly low cost that could then be expanded into a larger system merely by the acquisition of additional hardware modules of essentially the same kind. The multiprocessors related to the different strategies are reviewed below.

A. Symmetric Multiprocessor Machines

Perhaps the most significant area in which optical technology might be applicable is in the upper limits of SMP's. Rings and buses are 1-D networks whose performance depends critically on the traffic bandwidth of the communication system. The huge bandwidth aids in reducing the traffic latency caused by contention access. In SMP, the number of nodes is much larger today than was expected only a few years ago, and there would be great benefit in increasing it further, although the techniques required for bus-based implementation are already heroic. If optical technology could help extend SMP, even by a factor of 2 beyond the current maximum (64 processors for the Sun Microsystems Model Enterprise 10,000), it would readily find application.

B. Rings

Optical implementation of a ring-structured backplane makes it possible to achieve highly parallel links with very large bandwidth (of the order of terabits per second). Complete parallel transmissions (requiring the implementation of a parallelism as

high as 600 channels to insert simultaneously several transactions into the ring) would enable the realization of the huge bandwidth needed in forthcoming processors.[40] A massively parallel optical ring (i.e., several thousands of channels) could be divided into several concentric subrings to increase the bandwidth further. However, it must be stressed that the keys to success will be (1) the efficiency of the optical–electronic interface, (2) the capacity to carry out node operations (add, drop, bypass, and possibly on-the-fly error correction) in one or two ring cycles to compress as much as possible the different node latencies, and (3) the capacity to maintain the coherence of the local cache at the ring-operation frequency. To be successful, it will be necessary for the OI to show considerably enhanced performance in comparison with its electrical counterpart.

C. Supercomputers

One might consider a demonstrator of a highly parallel computer connected by a hypercube interconnection network that uses wide, fast buses. The one area in which optics can have a major advantage over electronics is in the interconnection buses in a network-based parallel computer. Here a high pin-out and a high data rate are both required to minimize the latency of network-based memory access. One could imagine a router chip with 10 buses of 500 bits, which would be impossible for electronic communication. This chip would require the collaboration of a parallel-computer manufacturer who is experienced in the use of commodity microprocessor parts. The major issue would be to obtain a microprocessor die in which the pin-out could be taken to flip-chip–bonding sites for emitters and for which optical inputs were also provided.

D. Reconfigurable-Network Machines

Architectures capable of exploiting the ability of optics to support static networks that can be reconfigured very quickly (in one cycle?) may find optics attractive. This reconfiguration rate is largely incompatible with modern shared-memory systems, which require much asynchronous, variable-sized packet switching. One application suggested is the use of cache-consistency protocols that use write–update, with special multicasting configurations to exploit knowledge about sharing patterns. However, write–update protocols aggravate the memory-ordering problems, and this application may depend on future directions in this as-yet unsolved problem.

E. Dedicated Optoelectronic Processors

There are still issues concerning the design, fabrication, and testing of such systems as those described above. On the design front the smartness of the pixel (i.e., the degree of complexity in relation to the number of electronic gates) that each optically interconnected element should possess to maximize the overall aggregate data-throughput rate is still under study. Useful figures of merit can include the power–weight product of the overall system, the power-consumption density per gigabit per second, and the throughput rate itself. The performance of such demonstrators is limited by the available laser-source power, the electronic chip area, or the heat-removal capability of the system. The interfacing technology between the optoelectronic components and the VLSI circuitry is still immature with several options still under consideration: flip-chip bonding, epoxy glue, anisotropic bonding, monolithic integration, etc. The optical hardware needs to be further miniaturized and to be compliant industrially, especially in terms of alignment accuracy and connections to the outside world. The yield and the testing of such systems have not been studied in great detail and deserve more fundamental study.

This study partly summarizes the conclusions of the Workshop on Optical Communications and Computer Sciences (WOCCS) that was held in Toulouse, France, in March 1999. This workshop was sponsored by the European Commission in the context of the MicroElectronic Advanced Research Initiative (Mel ARI). It is our pleasure to thank Pierpaolo Malinverni for his permanent support. We also wish to thank the participants, namely, Howard L. Davidson, Alfred Forchel, Gilles Jacquemond, Graham Jenkin, John D. Lambkin, Dominique Lavernier, Henk Neefs, Rami Melhem, Matthias Pez, Roger Vounckx, and Zvonko Vranesic, of the WOCCS for numerous fruitful and informal discussions. The position we have taken in this paper does not necessarily reflect the opinions of all the members who attended the workshop.

References and Notes

1. P. Chavel, D. A. B. Miller, and H. Thienpont, eds., *Optics in Computing '98*, Proc. SPIE **3490** (1998).
2. IEEE Computer Society, *Proceedings of the Fourth International Conference on Massively Parallel Processing Using Optical Interconnects*, Montreal, Quebec, Canada, 22–24 June (IEEE Computer Society, Los Alamitos, Calif., 1997).
3. D. A. B. Miller and H. M. Ozaktas, "Limit to the bit-rate capacity of electrical interconnects from the aspect ratio of system architecture," in the Special Issue on Parallel Computing with Optical Interconnects, J. Parallel Distribut. Comput. **41,** 42–52 (1997).
4. See the URL http://www.profibus.com.
5. See the URL http://www.fieldbus.org.
6. See the URL http://www.can-cia.de.
7. See the URL http://www.interbus.com.
8. S. Scott, "Synchronization and communication in the T3E multiprocessor," in *Proceedings of the Seventh International Conference on Architectural Support for Programming Languages and Operating Systems* (Association for Computing Machinery, New York, 1996), pp. 26–36.
9. C. B. Stunkel, D. G. Shea, D. G. Grice, P. H. Hochschild, and M. Tsao, "The SP-1 high performance switch," in *Proceedings of the Conference on Scalable High Performance Computing* (IEEE Computer Society, Los Alamitos, Calif., 1995), pp. 150–157.
10. See the URL http://www.ssd.intel.com.
11. See the URL http://www.sgi.com/origin.
12. For a complete document on HIPPI 6400, see the URL's http://www.noc.lanl.gov/~jamesh/hippi64; http://www.scizzl.com; and http://www1.cem.ch/HSI/sci/sci.html.

13. S. Scott, M. Vernon, and J. R. Goodman, "Performance of the SCI ring," in *Proceedings of the Nineteenth International Symposium on Computer Architecture* (Association for Computing Machinery, New York, 1992), pp. 403–414.
14. N. Boden, D. Cohen, R. E. Felderman, A. E. Kulawik, C. L. Seitz, J. N. Seizovic, and W.-K. Su, "Myrinet: a gigabit-per-second local area network," IEEE Micro. **15,** 29–38 (1995).
15. See the URL http://www.sun.com/servers/midrange/e6500/e6500.spec.html.
16. See the URL http://www.rambus.com/html/documentation.html.
17. S. Tang, T. Li, F. Li, L. Wu, M. Dubinovski, R. Wickman, and R. T. Chen, "A 1-GHz clock signal distribution for multiprocessor supercomputers," in *Proceedings of the International Conference on Massively Parallel Processing Using Optical Interconnects* (*MMPOI96*) (IEEE Computer Society, Los Alamitos, Calif., 1996), pp. 186–191.
18. R. T. Chen, L. Wu, F. Li, S. Tang, M. Dubinovski, J. Qi, C. L. Schow, J. C. Campbell, R. Wickman, B. Picor, M. Hibbs-Brenner, J. Bristow, Y.-L. Liu, S. Rattan, and C. Nodding, "Si CMOS process compatible guided-wave multi-Gbit/s optical clock signal distribution system for the Cray T-90 supercomputer," in *Proceedings of the International Conference on Massively Parallel Processing Using Optical Interconnects* (*MMPOI97*) (IEEE Computer Society, Los Alamitos, Calif., 1997), pp. 10–24.
19. T. Szymanski and H. Scott, "Design of a terabit free-space photonic backplane for parallel computing," in *Proceedings of the Conference on Massively Parallel Processing Using Optical Interconnects* (*MMPOI95*) (IEEE Computer Society, Los Alamitos, Calif., 1995), pp. 16–27.
20. Y.-S. Liu, B. Robertson, G. C. Boisset, M. H. Ayliffe, R. Iyer, and D. V. Plant, "Design, implementation, and characterization of a hybrid optical interconnect for a four-stage free-space optical backplane demonstrator," Appl. Opt. **37,** 2895–2914 (1998).
21. G. Verschaffelt, R. Buczynski, P. Tuteleers, P. Vynck, V. Baukens, H. Ottevaere, C. Debaes, S. Kufner, M. Kufner, A. Hermanne, J. Genoe, D. Coppée, R. Vounckx, S. Borghs, I. Veretennicoff, and H. Thienpont, "Demonstration of a monolithic multichannel module for multi-Gb/s intra-MCM optical interconnects," Photon. Technol. Lett. **10,** 1629–1631 (1998).
22. V. Baukens, G. Verschaffelt, P. Tuteleers, P. Vynck, H. Ottevaere, M. Kufner, S. Kufner, I. Veretennicoff, R. Bockstaele, A. Van Hove, B. Dhoedt, R. Baets, and H. Thienpont, "Performance of optical multi-chip-module interconnects: comparing guided-wave and free-space pathways," J. Eur. Opt. Soc. A **1,** 255–261 (1999).
23. J. C. Rodier, P. Chavel, A. Dupret, E. Belhaire, P. Garda, D. Prevost, and P. Lalanne, "Video-rate simulated annealing for stochastic artificial retinas," Opt. Commun. **132,** 427–431 (1996).
24. J. Tanida and Y. Ichioka, "Programming of optical array logic: image data processing," Appl. Opt. **27,** 2926–2939 (1988).
25. A. M. Weiner, "Femtosecond optical pulse shaping and processing," Prog. Quantum Electron. **19,** 161–237 (1995).
26. D. A. B. Miller, "Physical reasons for optical interconnection," Int. J. Optoelectron. **11,** 155–168 (1997).
27. G. Yayla, P. Marchand, and S. Esener, "Energy and speed analysis of digital electrical and free-space optical interconnections," in *Optical Interconnections and Parallel Processing: The Interface*, A. Ferreira and P. Berthome, eds. (Kluwer Academic, Dordrecht, The Netherlands, 1997), Chap. 3.
28. H. S. Hinton, *An Introduction to Photonic Switching Fabrics* (Plenum, New York, 1993).
29. O. Kibar, D. A. Van Blerkom, C. Fan, and S. Esener, "Power minimization and technology comparisons for digital free-space optoelectronic interconnections," J. Lightwave Technol. **17,** 546–555 (1999).
30. J. L. Hennesy and D. A. Patterson, "Buses connecting I/O devices to the CPU/memory," in *Computer Architecture, a Quantitative Approach*, 2nd ed. (Morgan Kauffmann, Los Altos, Calif., 1996), Sec. 6.3.
31. A. Bolychevsky, C. R. Jesshope, and V. B. Muchnick, "Dynamic scheduling in RISC architectures," IEEE Proc. Comput. Digital Technol. **143,** 309–317 (1996).
32. A. Iannucci, *Multithreaded Computer Architecture—A Summary of the State of the Art* (Kluwer Academic, Dordrecht, The Netherlands, 1994).
33. H. Neefs, P. Van Heuven, and J. Van Campenhout, "Latency requirements of optical interconnects at different memory hierarchy levels of a computer system," in *Optics in Computing '98*, P. Chavel, D. A. B. Miller, and H. Thienpont, eds., Proc. SPIE **3490,** 552–555 (1998).
34. H. Davidson, Sun Microsystems, 901 San Antonio Road, Palo Alto, Calif. 94303 (private communication, 11 March 1999).
35. J. H. Collet and L. Fesquet, "Comparison of the latency for an optical bus and several 2-D electronic topologies," in *CD-ROM of the Proceedings of the Eleventh International Parallel Processing Symposium* (*IPPS*) (IEEE Computer Society, Los Alamitos, Calif., 1997), CD addresses X:\workshps\wocs\collet.pdf; X:\workshps\wocs\collet.ps.
36. K. H. Wang, *Advanced Computer Architecture: Parallelism, Scalability, Programmability* (McGraw-Hill, New York, 1993).
37. A. Louri, B. Weech, and C. Neocleous, "A spanning multichannel linked hypercube: a gradually scalable optical interconnection network for massively parallel processing," IEEE Trans. Parallel Distribut. Sys. **9,** 497–512 (1998).
38. P. Sindhu, J. M. Frailong, J. Gastinel, M. Cekleov, L. Yuan, B. Gunning, and D. Curry, "XDBus: a high-performance, consistent, packet-switched VLSI bus," in *Technical Digest of the Spring '93 Computer Conferences* (*CompCon*) (IEEE Computer Society, Los Alamitos, Calif., 1993), pp. 338–344.
39. A. Charlesworth, "Starfire: extending the SMP envelope," IEEE Micro. **1,** 39–49 (1998).
40. W. Hlayel, D. Litaize, L. Fesquet, and J. H. Collet, "Optical versus electronic bus for address-transactions in future SMP architectures," in *Proceedings of the Conference on Parallel Architecture and Compilation Techniques* (*PACT*) (IEEE Computer Society, Los Alamitos, Calif., 1998), pp. 22–29.
41. Z. G. Vranesic, M. Stumm, D. M. Lewis, and R. White, "Hector: a hierarchically structured shared-memory multiprocessor," Computer **24,** 72–79 (1991).
42. L. A. Barroso and M. Dubois, "The performance of cache coherent ring-based multiprocessors," Tech. Rep. CENG-92-19 (Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, Calif., 1992).
43. A. V. Krishnamoorthy and D. A. B. Miller, "Scaling optoelectronic-VLSI circuits into the 21st century: a technology roadmap," IEEE J. Select. Top. Quantum Electron. **2,** 55–76 (1996).
44. R. G. Rozier, F. E. Kiamilev, and A. V. Krishnamoorthy, "Design and evaluation of a photonic FFT processor," J. Parallel Distribut. Comput. **41,** 131–136 (1997).
45. M. P. Y. Desmulliez, F. A. P. Tooley, J. A. B. Dines, N. L. Grant, D. J. Goodwill, D. Baillie, B. S. Wherrett, P. M. Foulk, S. Ashcroft, and P. Black, "Perfect-shuffle interconnected bitonic sorter: optoelectronic design," Appl. Opt. **34,** 5077–5090 (1996).