Performance scaling comparison for free-space optical and electrical interconnection approaches

Michael W. Haney and Marc P. Christensen

Projected performance metrics of free-space optical and electrical interconnections are estimated and compared in terms of smart-pixel input–output bandwidth density and practical geometric packaging constraints. The results suggest that three-dimensional optical interconnects based on smart pixels provide the highest volume, latency, and power-consumption benefits for applications in which globally interconnected networks are required to implement links across many integrated-circuit chips. It is further shown that interconnection approaches based on macro-optical elements achieve better scaling than those based on micro-optical elements. The scaling limits of micro-optical-based architectures stem from the need for repeaters to overcome diffraction losses in multichip architectures with high bisection bandwidth. The overall results provide guidance in determining whether and how strongly a free-space optical interconnection approach can be applied to a given multiprocessor problem. © 1998 Optical Society of America

OCIS code: 200.4650.

1. Introduction and Motivation

Smart-pixel throughput capabilities are projected to exceed 1 $(\text{Tbit/s})/\text{cm}^2$.[1] The hope is that this capacity will enable free-space optical interconnects (FSOI's) to provide significant throughput, size, and power-consumption advantages over all-electronic interconnection technologies. For accomplishing this goal, new architectures for interconnection-limited problems must be devised that exploit the ability of smart pixels to combine parallel high-density input–output (I/O) with local electronic logic.

Clearly optics provides the highest potential payoff for those problems that must dedicate a large portion of resources to interconnecting multiple processors in a dense, compact environment that challenges conventional electrical interconnect packaging approaches. In particular, three-dimensional (3-D) free-space optics may offer the ability to overcome the throughput and global interconnection limitations of conventional two-dimensional (2-D) metallic interconnection technology by exploitation of the additional spatial dimension. Our purpose in this paper

is to explore and compare the geometric scaling rules for 2-D metallic and 3-D free-space optical interconnection topologies. Such scaling relations will be useful for quantifying the benefits of optical interconnection approaches in given problem domains.

The focus of this paper is on those multiprocessor applications that require global high-density interconnections characterized by high minimum bisection bandwidth (BB)—a widely accepted measure of the degree of interconnection difficulty in networks. The BB of a network is defined as the bandwidth (BW) that crosses a boundary that cuts the network in half: It is a measure of wiring difficulty.[2] In architecture design there is a direct trade-off between minimum BB and latency in a network. It is therefore generally desirable to implement networks with the largest minimum BB that can be achieved practically to solve a given problem. The ability of optical elements to interconnect large arrays in space-variant patterns, without cross talk in the medium, suggests that FSOI techniques are particularly promising for problems with high BB. In particular, optical space-variant approaches to performing high-BB perfect-shuffle[3] and related patterns have been studied for some time.[4–9]

Chip-area requirements for high-density, interconnection-limited integrated circuits (IC's) were found to be proportional to BB^2.[10] In this paper circuit-area analysis is extended to problems for which the IC interconnection area is not sufficient to achieve the desired multiprocessor links. The total

interconnection area must therefore be determined for interconnection packaging technologies lower in the interconnection hierarchy, i.e., for multichip modules (MCM's) and, for the most highly interconnected problems, for printed circuit boards (PCB's). The total area requirement is used as a basis for estimating performance costs, such as volume, latency, and power consumption.

In general, the total circuit area will be the sum of the interconnection area and the processor area. The focus of this paper is on those problems for which the area dedicated to interprocessor interconnection dominates. It follows that the volume, latency (path length), and power-consumption performance metrics will then also be limited by the interconnection requirements. Path-length requirements for mesh-interconnected 3-D topologies have been analyzed based on an extension of Rent's rule.[11] This paper approaches path length, volume, and power consumption on the basis of global (i.e., shuffle-based) interconnection capability of free-space optical interconnections.

In Section 2 of this paper basic VLSI electrical interconnection-area scaling requirements are extended to MCM's and PCB's and then extended further to latency, power, and volume scaling rules. These parameters are derived as a function of the BB of the architecture. Section 3 is a derivation of the same parameters for interconnections based on optoelectronic technology. The emphasis is on globally interconnected systems in which a multichip data interchange dominates the interconnection requirements. In the discussion of Section 4 the derived scaling laws for different interconnection technologies (electrical, macro-optical, micro-optical) are compared to define those problem domains in which each technology has the greatest benefit. The conclusion, Section 5, summarizes the key results and relates the analysis to recent experimental developments.

2. Electrical Interconnection Requirements

A. Network Partitioning

The starting point for performance scaling analysis is the BW density capacity of the interconnection technologies. For the electronic packaging hierarchy the linear BW density is different in each level of the packaging hierarchy (IC, MCM, PCB). The linear BW [measured in (terabits per second) per centimeter] density stipulates the maximum BW that can cross any boundary as a function of the length of the boundary. Two types of boundary readily lend themselves to this analysis: internal bisection boundaries within partitions and external boundaries between partitions. Figure 1 depicts these two types of BW-limited boundary for the case of a single IC partition placed on a MCM.

To relate linear BW density to area requirements it is necessary to partition the architecture repeatedly into smaller equally sized sets of nodes. The requirements of every subpartition are calculated on the basis of the linear BW density of the interconnec-
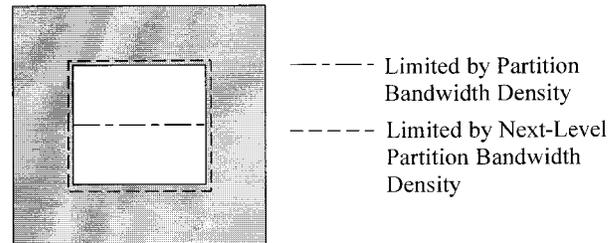

Fig. 1. Internal and external interconnection BW partition boundaries for metallic interconnection technologies.

tion technology. Often the optimum partitioning of the system, in the least-area sense, is the minimum bisection that separates the network into two equal groups and cuts the fewest wires. However, in general partitioning into any prime number of groups should be considered. For example, it is possible that the optimum partition—one that minimizes the BW between partitioned subsets—of a group of nodes might be a trisection (three equal-sized groups of nodes with fewer wires cut than a bisection of the nodes). To simplify the discussion we assume bisection partitions in this paper.

Figure 2 depicts an example interconnection architecture with 16 nodes. The I/O requirement for the entire system is 8B, where B is the BW of a single wire and there are four inputs and four outputs. Figure 3 depicts the minimum BB partitioning of the system: the cut wires are denoted by dashed lines. The internal minimum BB of the architecture is seen to be 6B. These cut wires are now part of the external BW requirements and are therefore added to I/O BW requirements. Figure 4 depicts the next level of minimum bisection partitioning. In general, this partitioning is repeated until each partition contains only one node. Figure 5 shows a tree depicting the resulting partitions for the example network shown in Figs. 2–4. Each node of the tree is labeled with the partition, the BB requirements of the partition,
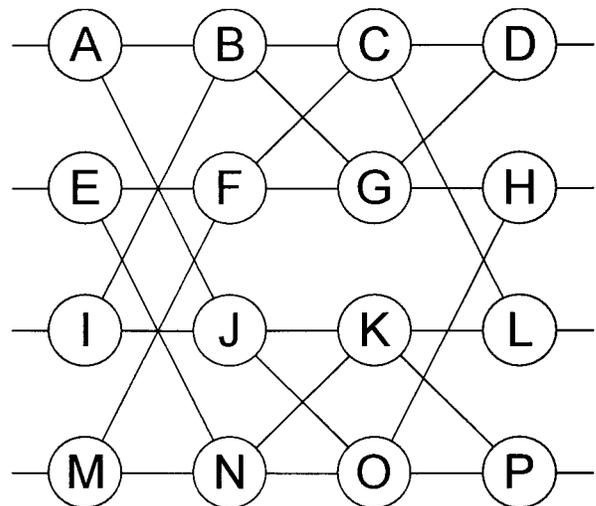

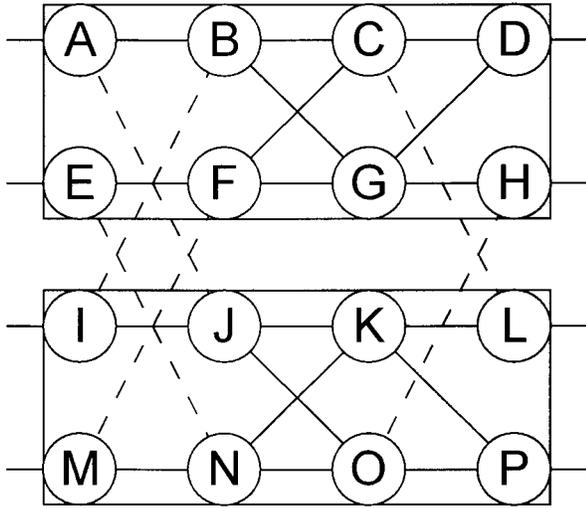Fig. 2. Example interconnection network of 16 nodes.

Fig. 3. Top-level minimum bisection partitioning of the nodes depicted in Fig. 2.

and the I/O requirements of the partition. Network partition trees are useful for determining the requirements for the interconnected architecture in different technologies.

For relating the BB and the I/O requirements of a node of the tree to area, the maximum capacities of the different levels of the packaging hierarchy must be determined. This is driven by the maximum practical or realizable size of each level. If one assumes a maximum size of a square package ($A^{1/2}$) with uniformly distributed nodes and a linear BW density (D_{layer}) for that layer, then Eqs. (1) and (2)
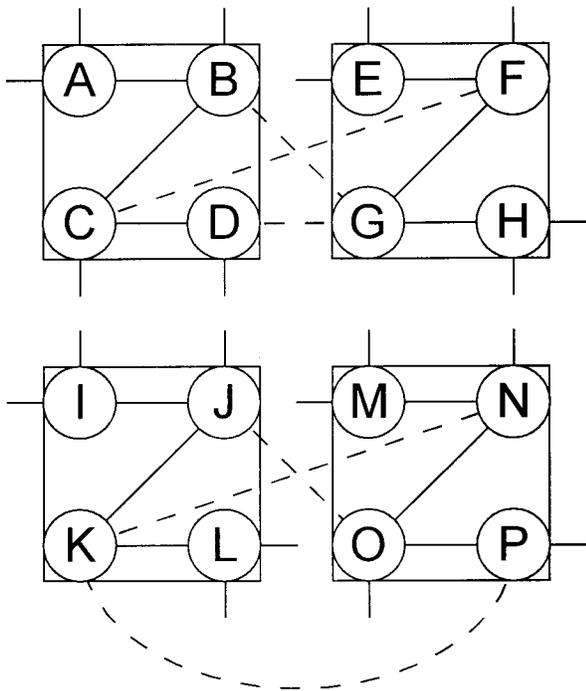

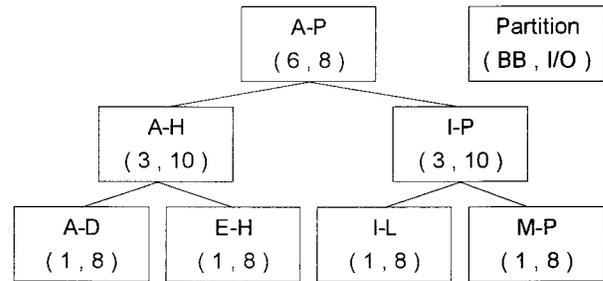Fig. 4. Subpartitioning of Fig. 3 into second-level bisections.

Fig. 5. Bisection tree for depicting the two levels of bisection of Fig. 2. Each node is labeled with the partition name, the internal BB of the partition, and its I/O BW requirement. This tree can be completed to four levels of bisection, thereby creating partitions of one node each.

dictate the maximum partition BB and I/O of that packaging layer:

$$\text{BB}_{\text{max}} = \sqrt{A_{\text{max}}}\, D_{\text{layer}}, \qquad (1)$$

$$\text{IO}_{\text{max}} = 4\sqrt{A_{\text{max}}}\, D_{\text{layer}}. \qquad (2)$$

Note that, when the partition boundary coincides with a technology boundary, e.g., the partition is an entire chip placed on a MCM, the I/O D_{layer} is determined by the lower hierarchical layer. As illustrated in Fig. 1, all data lines that leave a chip must cross the chip-package perimeter in the MCM layer, no matter how dense the connections between the chip and the MCM.

When the maximum capacities of each layer for BB and I/O BW are known, the tree of Fig. 5 can be traversed to calculate the required substrate area. Beginning with any node in the bottom row, determine first if that node can be realized within a single IC, and if so, then determine what size is required. If it can be realized, then traverse up the tree to its parent node. Now it must be determined if the parent node is realizable, while simultaneously realizing both of its daughter nodes in half an IC. The tree is climbed in this fashion until a given partition cannot be realized in the IC layer. From this point the process continues with the lower packaging layers (e.g., the MCM followed by the PCB) until the root node is reached. When this node is reached the total interconnection substrate area is estimated by the calculation of the maximum total area required across all three layers of the hierarchy. Note that the area specified by the bisection tree is the area required for only interconnection. It is possible that the total chip area, and not the lowest hierarchical layer, i.e., the topmost bisection, drives the area requirements. For example, when this analysis is applied to an architecture in which the first BB (topmost node) is extremely low but the subsequent partitions are characterized by large BB, the interconnection-area requirement of the topmost node (e.g., the PCB) will not result in an area large enough to mount the resultant MCM's and IC's. In this case the higher packaging layers clearly drive the interconnection-area requirements. This system would be charac-

terized by having many dense IC's interconnected on MCM's but with little interconnection between the MCM's in the PCB layer. In this case the maximum of the MCM or the IC area would determine the overall architecture area requirement.

When a network requires a regular global interconnection pattern, defined as an architecture for which each level of the bisection tree results in half the BB requirements of the previous level, the topmost partition determines the overall area. Butterflies and shuffles are examples of regular global interconnection patterns. In this case the above analysis is a direct extension of the VLSI area complexity analysis[10] to lower levels of the hierarchy. When the architecture is not a regular global interconnection pattern, the first bisection does not necessarily drive the area requirements. In this case, the bisection tree provides a mechanism to identify and quantify the area that drives the interconnection bottlenecks of the architecture. When this analysis is applied to nonglobal networks, the effect is to reduce the network to its most problematic (and therefore most highly interconnected) subnetworks, which are, within themselves, global in nature. The aggregate area requirement for a general problem must account for the subelement-area requirements at each bisection. Subsection 2.B extends this analysis to equations for area, power, latency, and volume.

B. Geometric Scaling Rules for Planar Metallic Interconnections

Subsection 2.A identified the optimum partitioning of a network and determined the BB requirements for each partition. From this, the area required for each partition is given by

$$A_{i,j} = \text{MAX}\left[\left(\frac{\text{BB}}{D_{\text{layer}}}\right)^2, A_{i-1,2j} + A_{i-1,2j-1}\right], \quad (3)$$

where i is the layer of the tree (numbered from bottom to top in Fig. 5) and j is the node within that layer (numbered from left to right). Equation (3) states that the interconnection-area requirement of a node is the maximum of its own BB requirements and the sum of its two daughter nodes' BB requirements.

The substrate-area interconnection requirement can be used to determine other important performance parameters. For example, interprocessor signal latency may be an issue when synchronous operation of the multiple processors is desired. In planar metallic technology the worst-case maximum path length L_{max} between processors will be the diagonal distance across the interconnection substrate:

$$L_{\text{max}} = \sqrt{2A} = \sqrt{2}\,\frac{\text{BB}}{D_{\text{layer}}}, \quad (4)$$

where A is the area requirement. To the extent that latency is proportional to the maximum distance between processors, L_{max} is a measure of latency in the network.

The total packaging volume for the interconnection can be bounded if it is assumed that each layer of the metallic interconnection hierarchy has a finite height H_{layer} that is the required clearance for the enclosure of the circuit, as determined by practical packaging constraints. For example, possible enclosure heights for the three levels of metallic packaging might be 0.1, 0.5, and 1 cm for H_{IC}, H_{MCM}, and H_{PCB}, respectively. The volume required for a given metallic interconnection package is therefore

$$V_{\text{layer}} = H_{\text{layer}}A = H_{\text{layer}}\left(\frac{\text{BB}}{D_{\text{layer}}}\right)^2. \quad (5)$$

The interconnection network's power-consumption requirement is also related to the geometric constraints of the planar interconnection hierarchy. Although the exact scaling rules for power consumption will depend on the details of the metallic technology used and other operational characteristics, it is useful to bound the power-requirement scaling rules for later comparison with optical interconnection scaling rules. If the electrical interconnections within a level are viewed as lumped capacitive loads (as, for example, in the short interconnections on IC's), then the power of each line will scale as the length of that line. In this domain the power requirements are bounded by

$$P_{\text{capacitive}} = P_c\sqrt{A}\,D_{\text{layer}}\sqrt{A} = P_c D_{\text{layer}}A = P_c\frac{\text{BB}^2}{D_{\text{layer}}}, \quad (6)$$

where P_c is the power required per unit length per unit BW (in watts per centimeter per terahertz), so the product of P_c and D_{layer} has units of watts per square centimeter. This represents an upper bound on the power requirements. A lower bound is derived under the assumptions of lossless transmission lines for the propagation of data. In this case the power is bounded from below by

$$P_{\text{lossless}} = P_l\text{BB}, \quad (7)$$

where P_l is the power required (in watts per terahertz) for driving the lossless transmission lines. Equations (6) and (7) provide bounds on the trends of power scaling as a function of the BB in the metallic packaging hierarchy. An actual implementation will therefore likely scale somewhere between the lower bound, which scales as BB, and the upper bound, which scales as BB^2. These bounds are presented here to facilitate a later comparison with optical interconnection requirements.

3. Optical Interconnection Requirements

A. Representations of Free-Space Optical Interconnections

Systems based on FSOI's can be categorized by the ratio of lenses to optical I/O. This is a measure of the degree of space variance in the optical system. Figure 6 is a depiction of the range of FSOI approaches. In general, planes of optical I/O's may be interconnected to each other. For simplicity, Fig. 6
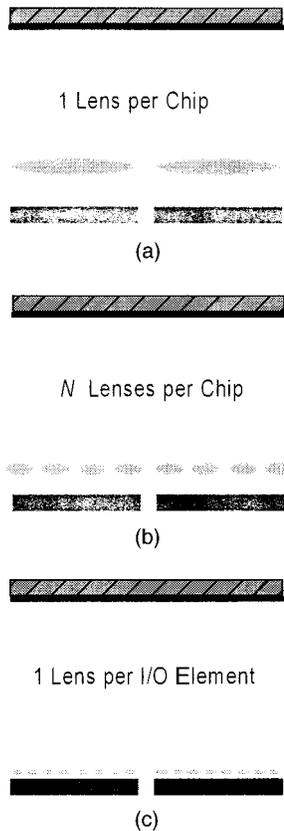
Fig. 6.   Three basic optical interconnection approaches, shown for a reflective architecture.

depicts single-plane reflective architectures in which all the smart-pixel resources are distributed on a common plane.   Figure 6(a) depicts a one-chip-per-one-lens scheme termed a macro-optical interconnection because the lenses are approximately the size of the smart-pixel chips: several millimeters or larger.[12]   In this case many optical I/O's are located beneath each lens.   Figure 6(c) depicts a micro-optical approach with one lens for each optical I/O.   In this case the beam-steering elements have diameters equal to the pitch of the high-density optoelectronic I/O, of the order of hundreds of micrometers.   Figure 6(b) depicts an intermediary approach with several lenses per chip and several I/O's per lens.   In principle, a single lens per I/O provides maximum flexibility, since an arbitrary interconnection pattern can be implemented with appropriate lens elements. Achieving arbitrary interconnection patterns in an approach like that of Figs. 6(a) and 6(b) requires local electronic interconnections and multiple passes through the optical system.[13]   The shape of the modules depicted in Fig. 6 is dependent on the f-number ($f_\#$) of the optics utilized.   As the $f_\#$ approaches unity, the reflective module approximates a cube in form.[14,15]

As depicted in Fig. 6, the interconnection architecture consists of an array of point-to-point links.   In principle, scaling to larger arrays with a larger BB simply requires larger multichip smart-pixel arrays

with the interconnection volume scaled appropriately to maintain the approximately cubic aspect ratio. Such scaling entails longer link lengths.   Under the assumption that smart-pixel-based interconnections require optoelectronic densities of $\sim 1000/\mathrm{cm}^2$, diffraction losses limit the lengths of these links. Macro-optics, with lens sizes of millimeters or more, scale well into free-space volumes with sizes of thousands of cubic centimeters.   However, the combination of a high I/O density and long link paths leads to diffraction limits in the micro-optical approach and thereby affects the scaling properties.

Determining the performance scaling of micro-optics requires a determination of the maximum allowable throw distance between optical elements above emitters and detectors.   The optical element size is set by the pitch of the smart-pixel I/O, limiting it to hundreds of micrometers or fewer.   Under the assumption of vertical-cavity surface-emitting lasers (VCSEL's) for the optical emitters, the propagation of Gaussian beams can be applied.[16–19]   The loss and cross-talk tolerances of the design and the type of beam forming that is implemented determine the actual throw distance.   For example, the microelements can be configured to achieve minimum divergence or a minimum beam waist.   Both yield similar throw-distance results.   The following example is a minimum divergence-angle estimate.   The loss criteria are set as follows: The input lens should capture at least 99.9% of the VCSEL light (to allow a close approximation to Gaussian beam propagation between the two lenses), and the throw distance should be constrained by the requirement that the receiving lens capture 86% of the light (i.e., matched to the beam waist).   Given a micro-optical element with diameter d, focal length f, and VCSEL beam waist w_0, the beam waist at the transmitting element aperture is given by

$$w_1 = w_0 \left[1 + \left(\frac{f\lambda}{\pi w_0^2} \right)^2 \right]^{1/2} = \frac{d}{2k}, \qquad (8)$$

where λ is the wavelength and k is chosen as 2.12 to maintain the Gaussian approximation to collect 99.9% of the light at the transmitting aperture.[17] The beam waist at the receiving element is therefore given by

$$w_2 = w_1 \left[1 + \left(\frac{z_{\max}\lambda}{\pi w_1^2} \right)^2 \right]^{1/2} = \frac{d}{2}. \qquad (9)$$

Equations (8) and (9) can be solved to determine the maximum throw distance $z_{\max}$:

$$z_{\max} = (k^2 - 1)^{1/2} \left(\frac{\pi d^2}{4\lambda k^2} \right). \qquad (10)$$

As an example, with $d = 200$ μm, $\lambda = 0.85$ μm, and $k = 2.12$, the maximum throw distance equals ~ 1.5 cm.   This first-order approximation assumes that the beams propagate perpendicularly to the optical elements.   The throw distance will actually be re-

duced for optical beams that propagate at steep angles because of the cosine projection of the beam waist. From geometric constraints, z_{max} and the $f_\#$ of the optics determine the mirror height h, as given by

$$h = \frac{f_\# z_{max}}{(1 + 4f_\#^2)^{1/2}}. \tag{11}$$

Diffraction effects on micro-optical architectures dictate that large-BB systems, characterized by large substrate areas, do not retain the cubic form of macro-optics. The short distances of micro-optics dictate a low aspect ratio for the interconnection volume. Furthermore, this short throw distance limits the lateral displacement of any given link, thereby requiring repeaters for connecting globally distributed nodes. This need for repeaters has a great effect on the scaling of micro-optical architectures, as detailed below.

B. Geometric Scaling Rules for Three-Dimensional Smart-Pixel-Based Architectures

Since FSOI's are not confined to planar links, the interconnection density limitations stem from the *area* I/O density capabilities of smart-pixel technology and the ability of optical elements to perform the interchip data interchange functions. FSOI concepts based on interleaved imaging of subarrays, as depicted in Figs. 6(a) and 6(b), are able to link arrays of smart-pixel I/O's with a resolution well beyond that required for achieving the anticipated terabit per (second per square centimeter) I/O densities of smart-pixel arrays. The area of the smart-pixel surface and the density $D_{I/O}$ [in terabits per (second per square centimeter)] of the optical I/O therefore determine the maximum BW that crosses external boundaries for the FSOI. If the interconnection pattern is global in that every IC communicates with every other IC, then half of the total IC area contains optical I/O's that cross any bisection boundary. The BB capacity of the FSOI is thus given by half the total smart-pixel I/O. For example, if $D_{I/O} = 1$ (Tbit/s)/cm², then the BB capacity of the FSOI is $D_{I/O}A_c/2$, where A_c is the total smart-pixel chip area employed. When this is inverted the area required for macro-optical interconnections is

$$A_{macro} = \frac{2BB}{D_{I/O}}. \tag{12}$$

As discussed above, the micro-optic's requirement for repeaters changes this for multichip architectures by reducing the effective density of the I/O, including only those emitters originating and not repeating data. The equation for micro-optics in terms of this effective density D_{eff} is

$$A_{micro} = \frac{2BB}{D_{eff}}, \tag{13}$$

where D_{eff} is given by

$$D_{eff} = \frac{D_{I/O}h}{f_\#(A_{micro})^{1/2}}, \tag{14}$$

where $h/f_\#$ is a normalized lateral throw distance and A_{micro} is the new area requirement. Solving for A_{micro} in Eqs. (13) and (14) yields

$$A_{micro} = \frac{4BB^2 f_\#^2}{D_{I/O}^2 h^2}. \tag{15}$$

Volume, latency, and power requirements may be derived directly from the above area analysis. Since there is no packaging hierarchy in free-space optics, only one area BW density is required. The volume required for macro-optical systems is approximated by

$$V_{macro} = A_{macro}^{3/2} f_\#, \tag{16}$$

whereas the fixed throw distance of micro-optics results in a volume of

$$V_{micro} = h A_{micro}. \tag{17}$$

From geometry, the maximum path length for both macro- and micro-optical architectures is

$$L_{max} = [A(1 + 2f_\#^2)]^{1/2}. \tag{18}$$

However, the areas for micro- and macro-optical systems scale differently, resulting in a different overall scaling in maximum path length. Similarly, the power requirements of optically interconnected modules are derived directly from area requirements. This power-requirement relation is described by

$$P = ANP_{link}, \tag{19}$$

where A is the area populated by the I/O, N is the total number of I/O's per square centimeter, and P_{link} is the power per I/O link.

Table 1. Example Planar Interconnect Parameters

Parameter	IC	MCM	PCB
A_{max}	6.25 cm²	225 cm²	N/A
D_{layer}	0.5 Tbits/cm	0.2 Tbit/cm	0.1 Tbit/cm
h_{layer}	0.1 cm	0.5 cm	1 cm
$P_c D_{layer}$	5 W/cm²	5 W/cm²	5 W/cm²
$P_{lossless}$	N/A	20 mW(Gbit/s)	40 mW/Gbit

Table 2. Example Optical Interconnect Parameters

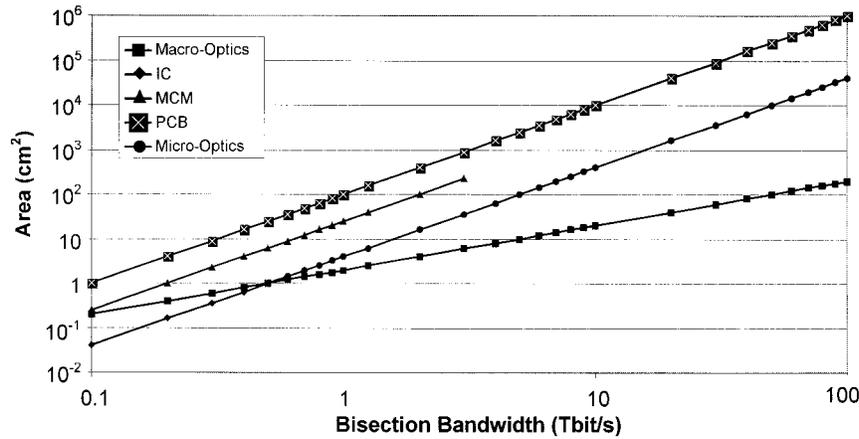| Parameter	Macro-Optical Interconnects	Micro-Optical Interconnects
λ	850 nm	850 nm
h_{max}	N/A	1 cm
$f_\#$	1	1
$D_{I/O}$	1 (Tbit/s)/cm²	1 (Tbit/s)/cm²
P_{link}	5 mW/link	5 mW/link
N	1000 links/cm²	1000 links/cm²

Fig. 7. Area-scaling graph for macro-optical, micro-optical, and metallic planar interconnections.

4. Discussion

Tables 1 and 2 contain sample parameters for making comparisons among planar metallic, micro-optical, and macro-optical interconnections. Although the actual values may vary, the slopes of the scaling equations are fixed. Figure 7 plots the area scaling of planar interconnects, micro-optical interconnects, and macro-optical interconnects based on the sample parameters. Figure 7 shows that the FSOI area requirement grows in direct proportion to the BB.[20] However, this scaling argument applies to only the macro-optical architecture. The macro-optical and the micro-optical architectures scale identically until the micro-optical architecture hits its diffraction-limited throw distance (at <1 Tbit/s). At higher BB's the micro-optical architectures scale at the same rate as the metallic architectures ($A \sim BB^2$) because of their need for repeaters.

Figure 8 depicts the interconnection-volume scaling requirements for the discussed technologies. Note that, while micro-optics has a much larger area than macro-optics, the difference in volume is not as extreme. This is due to the forms of the micro-optical and the macro-optical architectures. Micro-optical architectures are broad and flat, whereas macro-optical architectures are cubic in nature. However, after the diffraction-limited throw distance is exceeded, the micro-optical volume requirements scale as BB^2, as does electronics, whereas the macro-optical volume scales only as $BB^{3/2}$. For the selected parameters macro-optical architectures have a greater than 2 orders of magnitude advantage over metallic approaches for BB approaching 10 Tbits/s. It is noteworthy that the apparent wasted empty volume in cubic-shaped optical interconnection architectures actually provides this significant advantage.

Figure 9 depicts the maximum path-length scaling requirements for the discussed technologies. Clearly, for a low BB (<1 Tbit/s), IC technology is superior. However, for greater BB's, macro-optical path lengths scale as $BB^{1/2}$, whereas micro-optical and electronic path lengths all scale linearly with the BB. As discussed above, the maximum path length relates directly to the latency and the skew in the synchronization of multiprocessor systems. The data show that macro-optical systems will have a
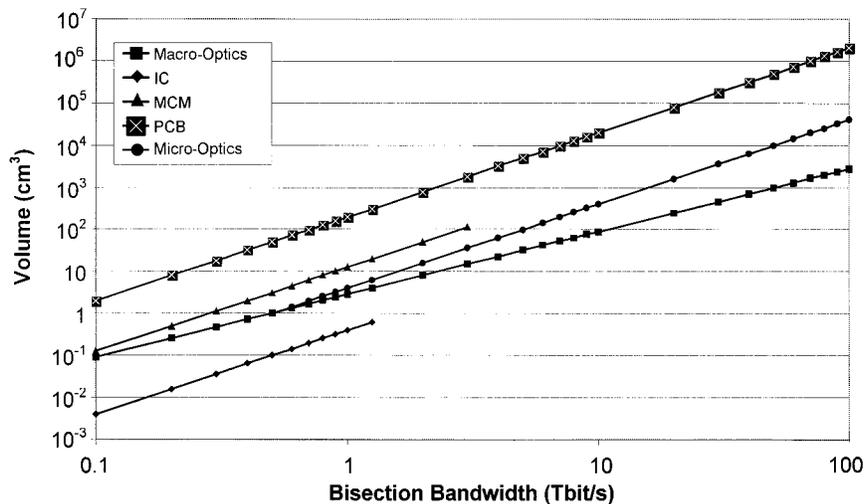

Fig. 8. Volume-scaling graph for macro-optical, micro-optical, and metallic planar interconnections.
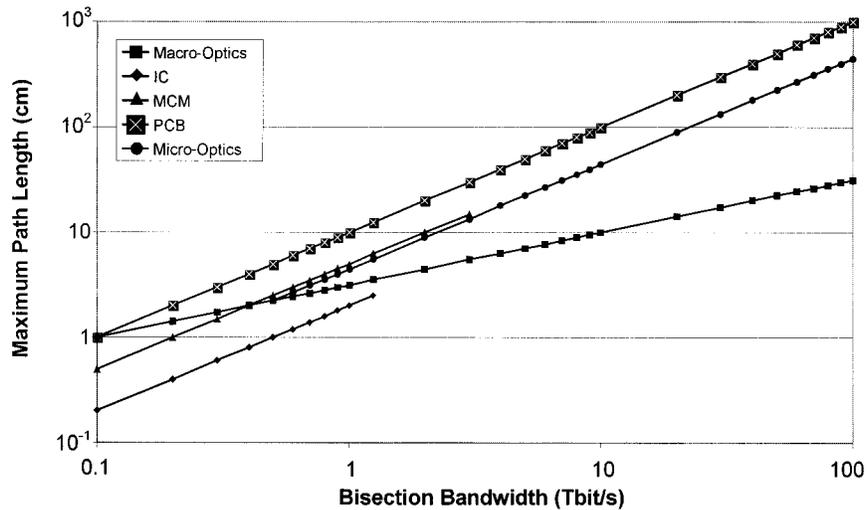
Fig. 9.   Maximum path-length scaling graph for macro-optical, micro-optical, and metallic planar interconnections.

significant advantage in latency in the ~10-Tbit/s BB regime.

Finally, Fig. 10 depicts trends for the interconnection power-consumption requirements for the relevant technologies. The electronic packaging layers are bounded on the graph by lossless transmission line analysis and lumped capacitive loading. Note that there are two lines each for the MCM and the PCB layers that represent these bounds. Macro-optics again achieves the best scaling (~BB) and matches the best possible electronic-scaling slope. Micro-optics, however, scales as poorly as the worst-case electronic power requirements (~BB^2).

Although all the performance metrics described above are derived from substrate-area considerations, user-defined metrics can combine them. For example, the product of power consumption and volume may be a critical figure of merit for some applications. As can be seen from Figs. 9 and 10, the advantages of macro-optical FSOI architectures are amplified when such measures are combined.

5. Conclusion

To realize the potential of the rapid advances being made in high-throughput smart-pixel technology, it is necessary to develop architectures based on macro-optical interconnection modules. Figure 11 shows a photograph of a prototypical macro-optical reflective multichip interconnection module. This system links four smart-pixel chips[21] with a 2 × 2 array of miniature projection lenses. This approach has achieved accuracies of ~10 μm across MCM substrates of 10 cm in extent for lens arrays as large as 4 × 4.[12]

The fundamental conclusion of this paper is that FSOI approaches have the most favorable scaling advantages when multiple IC's are globally interconnected, i.e., when multiple chips communicate
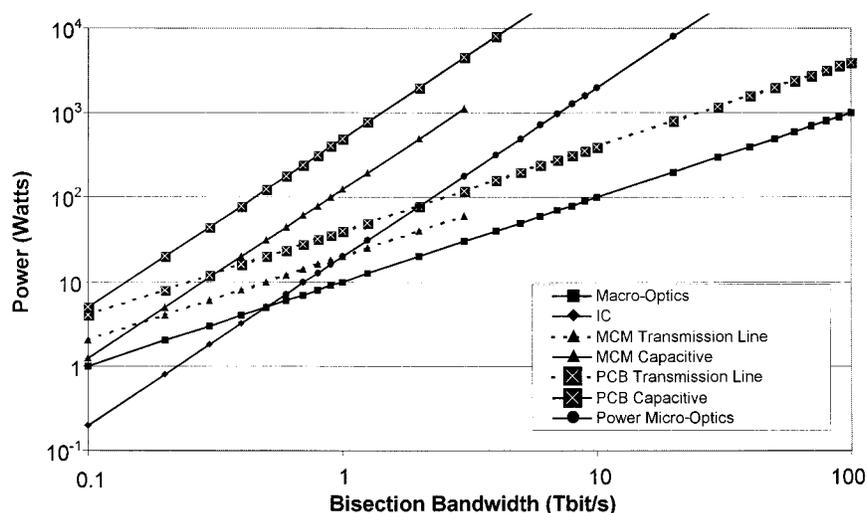

Fig. 10.   Power-scaling graph for macro-optical, micro-optical, and metallic planar interconnections.
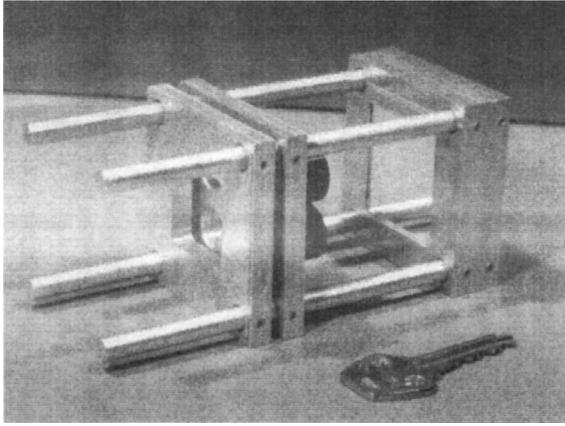
Fig. 11. Photograph of a prototypical macro-optical reflective multichip interconnection module.

simultaneously with multiple chips. This case typifies the multiterabit-per-second BB regime in which the FSOI has the greatest pay-off. The fundamental advantage of a macro-optical FSOI over metallic interconnections in terms of substrate-area-based metrics does not rely on the actual BW densities of the routing layers. It stems directly from the *reduction* in density in metallic interconnections as the BW is placed in lower layers. The only technological improvement that could overcome this fundamental advantage would be to raise the lowest routing level (PCB) densities to the densities of optical interconnections. This is not projected to happen, as density increases tend to trickle down from increased chip densities to increased MCM densities to increased PCB densities. As long as the metallic packaging hierarchy remains, the advantage of the FSOI will hold true. In other words, although electronic interconnection technology will continue to improve in density (as, we hope, will smart-pixel-based FSOI technology), the height and the placement of jumps between the metallic interconnect curves will change somewhat. However, the basic and fundamental advantage of the FSOI, as embodied in the lower slope and the lack of partition boundaries (i.e., *no* interconnect packaging hierarchy), for optics will remain.

References

1. T. Nakahara, S. Matsuo, S. Fukushima, and T. Kurokawa, "Performance comparison between multiple-quantum-well modulator-based and vertical-cavity-surface-emitting laser-based smart pixels," Appl. Opt. **35,** 860–871 (1996).
2. F. T. Leighton, *Introduction to Parallel Algorithms and Architectures; Arrays, Trees, Hypercubes* (Morgan-Kaufmann, San Mateo, Calif., 1992).
3. H. S. Stone, "Parallel processing with the perfect shuffle," IEEE Trans. Comput. **C-20,** 81–89 (1971).
4. A. W. Lohmann, "What classical optics can do for the digital optical computer," Appl. Opt. **25,** 1543–1549 (1986).
5. G. Eichmann and Y. Li, "Compact optical generalized perfect shuffle," Appl. Opt. **26,** 1167–1169 (1987).
6. S.-H. Lin, T. F. Krile, and J. F. Walkup, "2-D optical multistage interconnection networks," in *Digital Optical Computing*, R. Arrathoon, ed., Proc. SPIE **752,** 209–216 (1987).
7. K.-H. Brenner and A. Huang, "Optical implementations of the perfect shuffle interconnection," Appl. Opt. **27,** 135–137 (1988).
8. C. W. Stirk, R. A. Athale, and M. W. Haney, "Folded perfect shuffle optical processor," Appl. Opt. **27,** 202–203 (1988).
9. A. A. Sawchuk and I. Glaser, "Geometries for optical implementations of the perfect shuffle," in *Optical Computing '88*, P. Chaval, J. W. Goodman, and G. Roblin, eds., Proc. SPIE **963,** 270–282 (1988).
10. C. D. Thompson, "Area–time complexity for VLSI," in *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, 1979), pp. 81–88.
11. H. Van Marck and J. Van Campenhout, "Three-dimensional optoelectronic architectures for massively parallel processing systems," in *Proceedings of the Fourth International Conference on MPPOI*, J. Goodman, S. Hinton, T. Pinkston, and E. Schenfeld, eds. (IEEE Computer Society, New York, 1997), pp. 178–182.
12. R. R. Michael, M. P. Christensen, and M. W. Haney, "Experimental evaluation of the 3-D optical shuffle interconnection module of the sliding Banyan network," J. Lightwave Technol. **14,** 1970–1978 (1996).
13. M. P. Christensen and M. W. Haney, "Two-bounce free-space arbitrary interconnection architecture," in *Proceedings of the Fourth International Conference on MPPOI*, J. Goodman, S. Hinton, T. Pinkston, and E. Schenfeld, eds. (IEEE Computer Society, New York, 1997), pp. 61–67.
14. M. W. Haney and M. P. Christensen, "Fundamental geometric advantages of free-space optical interconnects," in *Proceedings of the Third International Conference on MPPOI*, A Gottlieb, Y. Li, and E. Schenfeld, eds. (IEEE Computer Society, New York, 1996), pp. 16–23.
15. M. W. Haney, M. P. Christensen, K. Raj, and P. Milojkovic, "Packaging advantages of macro-optical free-space interconnections over micro-optical and electrical interconnections," in *Advances in Electronic Packaging 1997*, Vol. 19-1 of EEP Series, E. Suhir, M. Shiratori, and Y. C. Lee, eds. (American Society of Mechanical Engineers, New York, 1997), pp. 811–817.
16. R. G. Schell and G. Tyras, "Irradiance from an aperture with a truncated-Gaussian field distribution," J. Opt. Soc. Am. **61,** 31–35 (1971).
17. P. Bellard and J. P. Crenn, "Changes in the characteristics of a Gaussian beam weakly diffracted by a circular aperture," Appl. Opt. **21,** 522–527 (1982).
18. L. D. Dickson, "Characteristics of a propagating Gaussian beam," Appl. Opt. **9,** 1854–1861 (1970).
19. F. B. McCormick, F. A. P. Tooley, T. J. Cloonan, J. M. Sasian, and H. S. Hinton, "Microbeam optical interconnections using microlens arrays," in *Photonic Switching*, H. S. Hinton and J. W. Goodman, eds., Vol. 8 of OSA Proceedings Series (Optical Society of America, Washington, D.C., 1991), pp. 48–51.
20. M. R. Feldman, C. C. Guest, T. J. Drabik, and S. C. Esener, "Comparison between electrical and free space optical interconnects for fine grain processor arrays based on interconnect density capabilities," Appl. Opt. **28,** 3820–3829 (1989).
21. M. W. Haney and M. P. Christensen, "Smart pixel based Viterbi decoder," in *Optical Computing*, Vol. 10 of 1995 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1995), pp. 99–101.