

Optical Interconnections for VLSI Systems

JOSEPH W. GOODMAN, FELLOW, IEEE, FREDERICK J. LEONBERGER, SENIOR MEMBER, IEEE, SUN-YUAN KUNG, SENIOR MEMBER, IEEE, AND RAVINDRA A. ATHALE

Invited Paper

The combination of decreasing feature sizes and increasing chip sizes is leading to a communication crisis in the area of VLSI circuits and systems. It is anticipated that the speeds of MOS circuits will soon be limited by interconnection delays, rather than gate delays. This paper investigates the possibility of applying optical and electrooptical technologies to such interconnection problems. The origins of the communication crisis are discussed. Those aspects of electrooptic technology that are applicable to the generation, routing, and detection of light at the level of chips and boards are reviewed. Algorithmic implications of interconnections are discussed, with emphasis on the definition of a hierarchy of interconnection problems from the signal-processing area having an increasing level of complexity. One potential application of optical interconnections is to the problem of clock distribution, for which a single signal must be routed to many parts of a chip or board. More complex is the problem of supplying data interconnections via optical technology. Areas in need of future research are identified.

I. INTRODUCTION

There are several roles that optics can play in the field of computation. Best known are the well-demonstrated applications of optics to analog computation. Examples include acousto-optic spectrum analyzers, convolvers, and correlators [1], [2], as well as systems for forming images from synthetic-aperture radar data [3], [4]. Such analog approaches offer very high processing speeds, but low accuracy and limited flexibility in terms of the types of operations that can be performed. These shortcomings have led to a search for applications of optics to digital [5], [6] or other types of numerical computation [7], [8].

A digital computer or computational unit consists primarily of nonlinear devices (logic gates) in which input signals interact to produce output signals, and interconnections between such devices or groups of devices of various sizes and complexity. The nonlinear interactions required of individual computational elements are realized in optics

only with considerable difficulty. Various kinds of optical light valves have been utilized to realize a multitude of parallel nonlinear elements [9], [10], but the speeds at which such devices can operate are exceedingly slow by comparison with equivalent electronic elements. Recent discoveries in the area of optical bistability have generated renewed interest in the possibility of constructing optical logic gates that are even faster than their electronic counterparts, but currently the efficiency of such devices is low and the device concepts are too little explored to allow a full assessment of their potential. It seems safe to say that the construction of optical logic gates with speeds, densities, and efficiencies equaling or exceeding those of electronic gates remains problematical, although future progress is certainly possible.

While optics lags behind electronics in the realization of the needed nonlinear elements, nonetheless the horizon for electronics is not without clouds. It is generally realized that the exponential growth of semiconductor chip capabilities cannot continue indefinitely, and that indeed important limits are beginning to be felt already. These limits arise not from difficulties associated with the further reduction of gate areas and delays, but rather from the difficulties associated with interconnections as dimensions are further scaled downward and chip area continues to increase [11]–[13].

Given the above facts, it is natural to inquire as to whether optics might offer important capabilities in overcoming the interconnect problems associated with microelectronic circuits or systems. Encouragement is offered by the observation that the very property of optics that causes difficulty in realizing nonlinear elements (it is difficult to make two streams of photons interact) is precisely the property desired of an interconnect technology. Further encouragement is offered by the noted successes of fiber optics in satisfying modern communication needs on a more macroscopic scale.

The purpose of this paper is to explore the possible means by which optics might contribute to the solution of interconnect and communication problems in integrated circuits (ICs) and systems. Attention is by no means limited to fiber optics, but rather places emphasis on integrated optics and free-space interconnection techniques as well. The ideas are admittedly speculative to some degree, but an attempt is made to introduce realistic numbers wherever

Manuscript received February 3, 1984.

J. W. Goodman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA.

F. J. Leonberger is with Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02173, USA.

S.-Y. Kung is with the Department of Electrical Engineering-Systems, University of Southern California, University Park, Los Angeles, CA 90080, USA.

R. A. Athale is with the Naval Research Laboratory, Code 6530, Washington, DC 20375, USA.

possible. To our knowledge there has been no previous attempt to systematically explore the area of optical interconnections in a comprehensive way.

Section II provides an overview of the current state of IC technology, with emphasis on the limitations posed by interconnect requirements. Section III provides further motivation by considering the implications of signal/image processing algorithms with respect to interconnections. Some algorithms require only local interconnects, while others fundamentally benefit from global or more flexible interconnect capabilities. Section IV reviews the current state of the art of optical technology relevant to the interconnect problem. Section V introduces a number of specific optical approaches to interconnections, at various levels, including intrachip, interchip, and interboard (machine-to-machine communication is excluded from consideration, due to the large attention it has already received by others). Finally, Section VI identifies the outstanding problems that must be solved for optical interconnections to find real application in the microelectronics field, and suggests future directions for research.

II. OVERVIEW OF INTEGRATED CIRCUIT TECHNOLOGY AND INTERCONNECT LIMITATIONS

The purpose of this section is to give a brief overview of the current state of metal-oxide-semiconductor (MOS) IC devices and systems, to outline the limitations posed by the interconnect problem, and to very briefly discuss some VLSI architecture issues. The treatment is far from exhaustive, but rather focuses on aspects that are relevant to the central question addressed in this paper; namely, what role might optics play in helping to solve the interconnect problem. Emphasis here is on silicon (Si) MOS circuits, since they are the base for the vast majority of VLSI-based computational power. Comments on the use of optical interconnections in hybrid GaAs-Si circuits as well as high-speed Si bipolar and GaAs ICs are found in Sections V and VI.

A. MOS Circuits [14]

MOS circuits are typically constructed from n-channel enhancement and depletion transistors (NMOS), or n-channel and p-channel enhancement transistors (CMOS). An n-channel transistor is made in a p-type substrate, whereas a p-channel transistor is made in an n-type substrate. In an n-channel transistor, the drain and source regions are created by n-type diffusions. The gate is made of a conductor (polysilicon) over a thin oxide covering the region between the drain and source diffusions. When the voltage of the gate is raised with respect to the drain, source, and substrate voltages, electrons are attracted to the surface of the substrate. Above a certain threshold, the number of electrons is so large that they form a conducting channel between the source and the drain.

The basic MOS module is the inverter circuit. Fig. 1 shows basic NMOS and CMOS inverter circuits. Usually, the transistor connected to ground is called the pulldown transistor, while the transistor connected to V_{dd} is called the pullup transistor. The pullup transistor of the NMOS inverter is a depletion-mode transistor, i.e., it is always on. The pullup transistor of the CMOS inverter is a p-channel enhancement-type; it will be on only when the voltage at

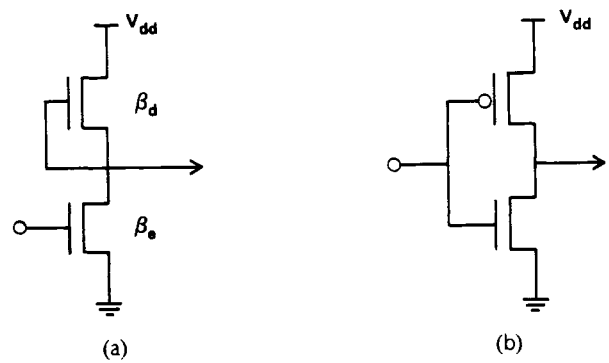


Fig. 1. Basic NMOS and CMOS inverter circuits. (a) NMOS-depletion load static inverter. (b) CMOS static inverter.

the gate is low with respect to the voltages at the drain and source. When the inverters are properly designed, their outputs can be used to drive the input of the next inverter stage.

A major difference between NMOS and CMOS is the power dissipated. A CMOS inverter draws power only in a transient condition, due to the fact that normally only one of the transistors is on. An NMOS inverter, on the other hand, draws power whenever the pulldown transistor is conducting. One way of alleviating this problem is to use dynamic logic, which is however more difficult to design and requires more area. Consequently, CMOS is in general preferable to NMOS.

B. Integrating Circuits on a Chip

A strong point of VLSI is the availability in the near future of a *hierarchical* and *multilevel* design method and the associated software packages. Such approaches are imperative due to the extremely large ($> 100,000$) numbers of MOS gates per chip in current technology. Usually four description levels are considered: 1) architectural, 2) register transfer, 3) logic/circuit, and 4) layout. An upper level description should be an elegant and powerful abstraction of the more detailed implementation at the lower level. For instance a control unit can be simply a box at the architectural level, one or more finite-state machines (FSM) at the register transfer level, an MOS programmable logic array (PLA) circuit description at the circuit level, and a collection of rectangles at the layout level. At each of the description levels the cells are hierarchically specified to decrease the complexity of the description. In order to reduce the design costs, a modular design approach is used; it is often less expensive to implement a general module that can be used in a number of different places than to implement a specific module that can be used only once.

The actual process of designing the layout of a VLSI chip is fairly well supported at the moment. Existing interactive layout editors and design rule checkers relieve the designer from most of the tedious work of specifying and checking the layout. An important aid in specifying the layout of a VLSI circuit is the so-called "stick diagram" [14]. A stick diagram specifies the topology of the circuit, i.e., the relative positions of the transistors and their interconnections. In a stick diagram the transistors are symbolically depicted as the crossing of polysilicon and diffusion lines. A stick diagram adequately models the functional behavior (i.e.,

the logic gates and their interconnections) of the circuit. However, it does not allow the specification of certain capacitive effects, such as bootstrapping.

At the circuit level, we decompose the circuit into three major types of building blocks. The basic *memory* module is the one bit register cell; the basic *logic* module is the AND-OR-INVERT gate, and the basic *arithmetic* module is the full adder. Fundamentally, a VLSI circuit consists of these three types of modules. A somewhat special (but widely used) logic module is the PLA. A PLA can be used to implement any set of Boolean equations, and if combined with a state register, can even implement a complete FSM. A PLA can be generated directly from a register transfer level specification. In general, with the increasing use of high-level design aids, we see more and more programs that are able to synthesize large portions of a VLSI circuit from a high-level (register transfer) specification.

The building block approach, described above, when combined with high-level tools such as silicon compilers/assemblers, gives the VLSI designer the flexibility and modularity needed to cope with the ever increasing complexity of VLSI design.

C. Effects of Scaling on Device and Interconnection Delays

The exponential growth of IC complexity and capabilities experienced since the birth of the industry has been caused by a combination of scaling down of the minimum feature size achievable, and a scaling up of the maximum chip size, both subject to the constraint of reasonable yield. The scaling process has many beneficial effects, but also eventually causes difficulties if combined with "stuffing," i.e., the addition of circuitry in order to realize more complex circuits in the same area of silicon that was used before scaling [15]. Here we wish to briefly discuss the good and bad effects of scaling.

We will assume that all the dimensions, as well as the voltages and currents on the chip, are scaled down by a factor α (an α greater than one implies that sizes or levels are shrinking). Consider first the effects of device scaling. Obviously, when scaling down the linear dimensions of a transistor by α , the number of transistors that can be placed on a chip of given size scales up by α^2 . In addition, the power dissipation per transistor decreases by a factor α [16], due to the fact that both the threshold voltage and the supply voltage are scaled down by α . Finally, we note that the switching delay of a transistor is scaled down by α , due to the fact that the channel length is decreased by a factor α .

Scaling also affects the interconnections between devices. Fig. 2 depicts the effect of scaling down a conductor

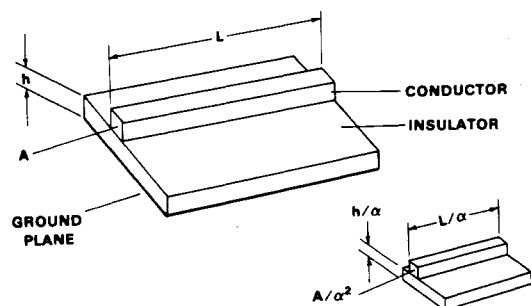


Fig. 2. Scaling of a conductor (scaling factor α).

by a factor α . Since the cross-sectional area of the conductor is decreased by a factor α^2 , the resistance per unit length will increase by a similar factor. If the length of the conductor is scaled by α (as simple scaling implies), then the net increase of resistance is in proportion to α . At the same time, scaling implies changes of the capacitance of the interconnection. Regarding the conductor as one plate of a parallel-plate capacitor, scaling down of both linear dimensions of the plate by α implies a decrease of the capacitance by α^2 . However, scaling down also implies a decrease by α of the thickness of the oxide insulating layer separating the plates of the capacitor. Hence the capacitance of a fixed interconnection scales down by α . We see the scaling up of resistance and down of capacitance exactly cancel, leaving the RC time constant and the interconnect delay unchanged.

It is well-recognized in the IC literature that the scaling laws outlined above will eventually pose serious problems for the VLSI industry [11]–[13], [17]–[19]. First, it is clear that since gate delays decrease with scaling while interconnect delays remain constant with scaling, eventually the speed at which a circuit can operate is dominated by interconnect delays rather than device delays. However, the situation is actually somewhat worse than the above considerations imply, due to the fact that as scaling and stuffing occur, the lengths of the interconnects required do not scale down with the inverse of α , as was assumed. Rather, as the complexity of the circuit being realized increases, the distances over which interconnections must be maintained on a chip of fixed area may stay roughly constant. It has been argued from statistical considerations [20] that a good approximation to the maximum length $L_{\max}$ of interconnection required is given by

$$L_{\max} = \frac{A^{1/2}}{2} \quad (1)$$

where A represents the area of the chip. Note that if in addition to scaling, chip size is increased, the interconnect problem becomes further exacerbated. As a consequence of these considerations, it has been estimated that by the late 1980s, chip speeds will be limited primarily by interconnect delays [11]. The tantalizing possibilities for bringing optics to bear on this interconnection bottleneck are prime motivating factors for the considerations of this paper.

D. Effects of Scaling on Numbers of Interconnections Required

In accord with the hierarchical nature of design, a complicated VLSI chip can be regarded as consisting of a multitude of subunits of circuitry, called "blocks," connected to form larger circuit units, called "super-blocks" [11]. As scaling proceeds, the complexity of the blocks can be made greater, and the number of blocks that can be realized in a single super-block also grows. As the number of elements in a block increases, the number of interconnections required from that block to other blocks also increases. If the assumption is made that interconnections to or from a given block must be supplied around the perimeter of that block, then the limitation imposed by the amount of available perimeter implies that the number of interconnections that can be supplied grows as the square root of the area of the block, or equivalently as the square root of the number of devices contained within the block. However, there is a

well-known empirical relation, known as *Rent's rule*, which specifies that the number of interconnections M required for a block consisting of N devices grows as approximately the $2/3$ power of N , i.e.,

$$M \approx N^{2/3}. \quad (2)$$

At the chip level, the disparity between the number of interconnections that can be realized and the number required becomes more and more severe as the number of devices within the chip grows larger through scaling and stuffing. For example, a circuit consisting of 100 000 gates requires about 2000 interconnections. For a 10 mm $\times$ 10 mm chip works out to a connection pad every 20 μ m. It should be noted that Rent's rule applies only to circuits consisting of logic elements. Memory cells require fewer interconnections. In addition, it is required that each circuit be a small "random" subset of the entire logic system [13].

At the chip level, some of the limitations implied by Rent's rule can be overcome by use of metal bump technology for making interconnections possible from the interior of a chip, rather than just from the edges. Optical techniques may ultimately provide an alternate and more flexible means for providing interconnections directly to the interior of a chip.

As the number of devices realizable on a single chip grows, the assumptions underlying Rent's rule may become invalid, and the exponent associated with that rule may fall to less than 0.5. Nonetheless, from a practical point of view, connections to and from a chip are likely to remain a problem area, due to the bandwidths and driving powers required, as well as the continuing need to distribute signals to and from the interior of the chip.

E. Effects of Scaling on Electromigration

There is one further negative effect of feature size scaling that should be mentioned, namely *electromigration*. While current is scaled down inversely as α , the cross-sectional area through which that current must flow is scaled down inversely with α^2 . The net result is that current density increases in proportion to α . Such an increase leads to greater electromigration effects, by which is meant the movement of conductor atoms under the influence of electron bombardments, resulting ultimately in the breaking of conductor lines. The potential of optical interconnections as a means for reducing the electromigration problem is of considerable interest here.

F. System Considerations

Arrays of identical VLSI processing elements can be organized to perform a variety of signal processing functions, as will be reviewed in the next section. Interconnections in such VLSI systems are customarily implemented in a two-dimensional circuit layout with a few crossover layers. With such technology, communication is ideally restricted to localized interconnections, since communication is very expensive in terms of area, power, and speed [14]. Dynamic interconnections are desirable for general-purpose computing and in certain signal processing applications.

The timing framework is a very critical issue in system design. There exist two different timing schemes, *globally synchronous* and *locally synchronous* approaches. In the globally synchronous scheme, a global clock network must distribute the clocking signals throughout the entire system. Clock skew associated with global clock distribution

has become an increasingly important factor in limiting achievable clock speeds. Under such circumstances the locally synchronous approach has some advantages, in that there need be no global clock, and information transfer is by mutual convenience and agreement between each processing element and its immediate neighbors. The performance of such a scheme is less affected by the scaling of technology than is the globally synchronous scheme, and it can be implemented with a simple handshaking protocol [21]. Optics may provide an alternative method for solving these synchronization problems, as discussed in greater detail in Section V.

III. ALGORITHMIC IMPLICATIONS OF INTERCONNECTS

Further motivation for considering optical interconnections is provided by algorithmic considerations. The capabilities and limitations of the interconnect technology utilized in realizing a computational or signal processing unit play a substantial role in determining the speed and flexibility of the operations that can be achieved by that unit. Different algorithms require different degrees of interconnect globality, and it is the implications of such considerations, particularly as they pertain to electronic and optical interconnections, that are the subject of this section.

To effectively exploit the special features of optical and electronic technologies, the mathematical operations needed in signal processing and computational operations must be cast in suitable algorithms. Optical signals can flow through three-dimensional space to achieve the required interconnect pattern between elements of a two-dimensional data array before executing the desired operation between them. Current VLSI-based electronic systems, on the other hand, are inherently two-dimensional in nature. While work on three-dimensional VLSI is in progress, success in this endeavor will primarily increase the density of computational elements, rather than alleviating the constraints imposed by interconnect limitations. For current VLSI technology, the interconnect paths as well as the processing elements have to share what is essentially a common plane. These topological considerations, as well as cross-talk and interconnect delay limitations, impose a restriction to nearest neighbor interconnections as being highly desirable in VLSI parallel-processing systems. The algorithmic mapping of the same mathematical operation with optical interconnects and electronic interconnects may be quite different in order to conform to the different constraints of the interconnect patterns. In subsequent discussion, we will consider four different classes of signal processing operations that are defined by the type of interconnect patterns that are needed to implement those operations on a parallel array of processors.

A. Point Operations

In this category of operations, each point in the one- or two-dimensional data array is processed completely independently. If the one-dimensional input is a time sequence, then these operations are referred to as "memoryless" operations. All the points in the input array may be processed in the same way or each could have its own independent instruction set. In either case, it is clear that once the input data array is loaded into the array of processing elements, then each element can carry out its own predetermined processing task completely independently of the rest of the

elements in the array. The interconnectivity required by these operations is therefore minimum and will be apparent only while loading the data into or unloading the data from the processor array. These operations can therefore be carried out in parallel, whether interconnections are provided by optical or electronic means.

Optical interconnections have the advantage of being able to input the entire two-dimensional data array in parallel using the third dimension for data propagation. On the other hand, in an electronic parallel processor, such as the Goodyear Massively Parallel Processor (MPP) [22], the data can be input and output only along the edges of the two-dimensional array, one row/column at a time.

In those cases where the operations needed are complex, the overhead associated with the data input and output may be small compared to the computational load of the main operation, thus minimizing the need for a completely parallel input/output link between the processing element and the outside world. Examples of such operations can be found especially in the field of image processing. The value of a picture element can be transformed according to a prescribed nonlinear function in order to modify its contrast. Another example would involve correcting for spatially varying sensitivity of a two-dimensional sensor array by suitable post-processing. Addition/subtraction of matrices is yet another example of this type of operation.

B. Matrix Operations

A large number of signal and image processing algorithms can be expressed in terms of matrix operations. The multiplication of two matrices is one of the most basic operations in matrix algebra (a vector-matrix multiplication can be considered as a special case of this more general operation). Such a multiplication is described mathematically as follows:

$$C_{ij} = \sum_{k=1}^N A_{ik} B_{kj}, \quad i, j = 1, \dots, N \quad (3)$$

where A , B , and C are assumed to be $N \times N$ square matrices, for the sake of convenience. Alternatively, the output matrix C can be defined as a sum of N outer product matrices formed by multiplying column vectors of A by corresponding row vectors of B

$$C = \sum_{k=0}^N C^{(k)} \\ C_{ij}^{(k)} = A_{ik} B_{kj}, \quad i, j = 1, \dots, N \quad (4)$$

where the second line defines the outer product between the k th column of A and the k th row of B . It is evident from the above equations that this operation involves a high degree of interconnectivity between the elements of the input matrices and the output matrix. Thus all the elements in a given row of A and a given column of B will contribute to one element of the output matrix C . Conversely, one element of matrix A (or B) contributes to all elements of the corresponding row (or column) of matrix C .

Taking advantage of these properties, the global interconnect capabilities of optics can be exploited to build high-speed, high-throughput parallel optical processors to perform matrix multiplication [23]. On the other hand, the

regular nature of the interconnectivity suggested in (3) and (4) implies that these operations can also be carried out via recursive and locally interconnected algorithms implemented with systolic architectures in VLSI [21], [24]. In such algorithms, all processors perform nearly identical tasks and each processor repeats a fixed set of tasks on sequentially available data. A recursive algorithm is said to be locally interconnected if the space indices of the data elements input to the same processor in successive recursions are separated by no more than a given limit [21]. In matrix operations, these indices are found to differ by 1, thus indicating nearest neighbor type of interconnectivity.

It can be seen from (3) and (4) that the computation involved in matrix multiplication grows as $O(N^3)$, where N is the dimension of the matrices. The use of a two-dimensional array of processing elements that perform multiplication and addition of two numbers along with a suitable nearest neighbor interconnection network can be shown to carry out the matrix multiplication in $O(N)$ time. Thus the global interconnection capability offered by optics does not provide any significant computational advantage over systems using only nearest neighbor interconnections when dealing with simple matrix multiplications.

It is worth noting that some algorithms for matrix operations, more complicated than the simple product discussed above, have been proposed that require nearest neighbor connectivity on a three-dimensional surface (e.g., a torus) [25]. Clearly, the communication problem posed by cutting such a surface for compatibility with a planar processor geometry offers opportunities for contributions by optical interconnections. Another potential role for optical interconnections is in the problem of clock distribution in a globally synchronous systolic processor (see Section VI for more detailed consideration of the clock distribution problem).

C. Fourier Transforms and Sorting

Fourier transformation and sorting are two important signal processing operations that entail global interconnections between all the elements of the input array. In other words, every element of the output array is affected by all elements of the input array, and conversely, each element of the input array affects all elements of the output array. The computations involved in the Fourier transform are complex multiplication and addition, whereas in sorting it is the comparison operation.

The discrete Fourier transform (DFT) of a one-dimensional sequence is defined by

$$X(k) = \sum_{i=0}^{N-1} x(i) W_N^{ik}, \quad k = 0, \dots, (N-1) \quad (5)$$

where

$$W_N^{ik} = \exp[-j2\pi(ik)/N].$$

It can be seen that if implemented in a straightforward fashion, this operation involves computation that grows as $O(N^2)$. But the regular structure of the problem as well as the periodic nature of the W_N^{ik} suggest a more efficient algorithm, in which the computation grows as $O(N \log N)$ [26]. However, this computational savings comes at the expense of a global and more complicated interconnection pattern between the input elements than that implied by

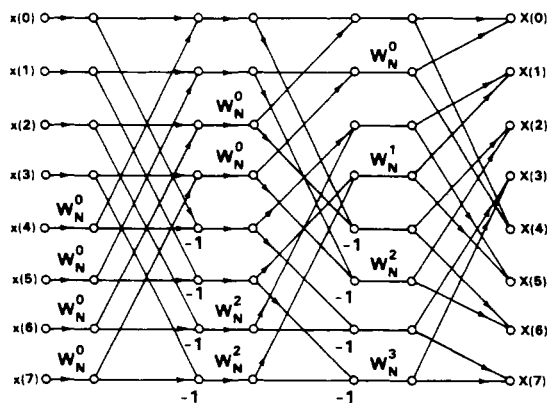


Fig. 3. Flowgraph for the FFT.

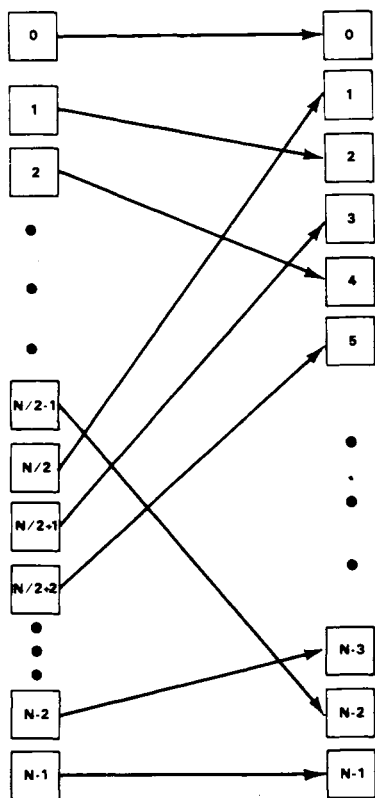


Fig. 4. Perfect shuffle interconnections.

(6). Fig. 3 shows the familiar butterfly diagram of a fast Fourier transform (FFT), which indicates that the interconnections change at different stages of the computation. This requirement for dynamic interconnections can be avoided by resorting to a fixed but global interconnect pattern known as the "perfect shuffle" and shown in Fig. 4. The perfect shuffle can be applied repeatedly at each stage of the FFT to produce the interconnect pattern required for that stage [27], presumably at a cost of extra time required to complete the interconnections.

The operation of sorting a sequence of numbers involves elementary operations of comparing two numbers and arranging them in descending order at the output. One algorithm for efficient sorting is Batcher's bitonic sort algorithm [28]. The basic principle behind this algorithm is the "divide-and-conquer" method of breaking a large problem

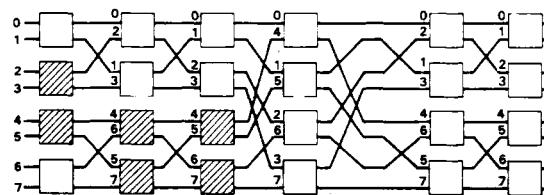


Fig. 5. Sorting network for eight items based on Batcher's bitonic sort algorithm.

into several smaller problems, and then combining their solutions at the output to generate a solution to the more general problem. The network for implementing Batcher's sort algorithm is shown in Fig. 5. Again it is seen that the interconnections between elements are global and dynamic. Since the FFT algorithm is also based on the divide-and-conquer principle, it should come as no surprise that the interconnections required for both operations are identical and hence can be generated by repeated application of the perfect shuffle [27].

Thus both of the operations discussed above belong to the class of operations requiring global and dynamic interconnection between the different elements of the input array. The regularity and structure of the different interconnect patterns leads to a simpler realization of the dynamic interconnects via repeated use of a global but fixed interconnect network (the perfect shuffle network in the case discussed). If either the fixed or the dynamic interconnect networks can be implemented easily and efficiently using optics, then a speedup of the throughput without a corresponding increase in the hardware can be achieved.

D. Space and Time Variant Operations

In image restoration and pattern recognition, one is interested in studying structures of various sizes in images, from a single point and its nearest neighbors to an image covering the entire available field. The interconnectivity between each input point and all object points is determined by the degree of globality of the operation under consideration. When the size and nature of the interconnectivity of the operation to be performed is invariant over the image space, as well as in time, then the problem is considerably simpler, and fixed optical or electronic (depending on the degree of globality required) interconnection networks can be used.

For problems encountered in restoration of images degraded by atmospheric turbulence and in processing signals obtained from a dynamic sensor array, the interconnectivity varies in space and time. Furthermore, the interconnect patterns could be data-dependent (as often is assumed in modeling the human visual system), making it impossible to foresee the interconnect requirements at different stages of processing without having foreknowledge of the input.

The computational throughput of a parallel processor implementing these types of operations will be critically affected by the availability of a dynamic and global interconnect network. Without such a capability, the processors could be idle for a significant number of cycles while the data are routed to the correct processors. A much higher degree of supervision on the part of the controller would be required, and there would be many more input/output operations to and from memory. All these

considerations decrease the computational efficiency of the parallel processors and point to the importance of flexible interconnects.

E. Overview

The four classes of signal/image processing operations discussed above require varying amount of complexity on the part of the interconnect network that routes signals to processors. As the complexity increases, so does the impact of a flexible interconnect technology on the throughput of a parallel processor, whether optical or electronic. In the case of point operations, the motivation for a flexible and global interconnect capability is weak; interconnects are most important in this case when many input/output operations to the processor array are required in practice. In the second category, namely, operations involving matrix algebra, the global but fixed interconnects inherent in the operations can be converted to nearest neighbor interconnects via the recursive formulation. The computational throughput can be high when pipelining is used. Therefore, in this type of application, global interconnections will be beneficial only when the nature of the application produces frequent breaks in the pipelining of the processor. The third category of operations involves global and dynamic interconnects, which can be reduced to repeated application of a global but fixed interconnect network. If these fixed interconnects are implemented in a fast and efficient way, perhaps with the help of optics, then a high computational throughput can be achieved without excessive hardware and without a dynamic interconnect structure. The last category considered contains operations with a minimum amount of regularity and structure, along with possible data and time dependency in the interconnect requirements. A global and programmable interconnect network will be vital to achieving high throughput and high efficiency with parallel processors implementing these operations.

IV. OVERVIEW OF ELECTROOPTIC TECHNOLOGY AS IT PERTAINS TO THE INTERCONNECT PROBLEM

The rapid growth of the fiber telecommunications industry has been in part due to the development of near-infrared optical sources, modulators, and detectors. These components are particularly attractive for solving the interconnect problem. In this section the state of the art of these components, as well as of the transmission media for directing light to specific locations, will be reviewed.

A. Light Sources, Modulators, and Detectors

The alternatives for optical sources are diode lasers and light-emitting diodes (LEDs) [29]. While much of the telecommunications effort is now directed at InP-based devices whose emission wavelengths match the 1.3–1.5- μm optimum fiber band, the 0.85- μm range of GaAs/GaAlAs emitters is of most interest here because of the wavelength sensitivity compatibility with on-chip Si detectors. Typical efficient low-threshold GaAs lasers are of the index-guided type and emit 3–15 mW with threshold currents of approximately 30 mA and differential efficiencies in the 20–30-percent range. The spectral bandwidth is of the order of 20 Å (about 0.2 percent). Typical active laser area is

500 × 2 μm , but packaged chips are generally wider for bonding. Such lasers are edge emitters with cleaved mirrors and emission patterns typically 10 × 35 degrees in width. They have projected lifetimes in the $\geq 10^5$ -h range. In the laboratory, a number of techniques have been proposed for noncleaved mirrors for more compact devices. Although surface-emitting GaAs lasers have seen limited development, both distributed feedback [30] and heterointerface-mirror [31] types have been reported. These devices are to date relatively inefficient and in the latter case require cryogenic cooling.

LEDs are well developed and can be designed for edge or surface emission. In contrast to diode lasers, they are inefficient (100-mA drive) with a low power output (of the order of 1 mW) that has a broad emission pattern (Lambertian for a surface emitter) and a wide spectral range (≈ 450 Å). However, they are more stable and, in particular, less temperature sensitive than diode lasers.

To modulate these sources, either direct current modulation or external modulation can be used. Diode lasers can be large-signal modulated to rates of approximately 2 GHz with minimal pattern effects, and LEDs are generally limited to < 100-MHz modulation but have recently been reported to operate at rates up to 500 MHz [32]. These sources present dynamic resistive loads of approximately 10 Ω . For situations where driving a capacitive load is preferred, one may use external modulators. Here full on–off modulation requires approximately 4 V in a waveguide modulator for a bandwidth to about 3 GHz with a capacitance of approximately 3 pF. [33] The use of LiNbO₃ waveguide modulators to monitor interchip signals in VHSIC circuits has recently been reported [34].

For modulating an optical wavefront propagating through free space, a large variety of electrically addressed light modulators can be used. Recently, several devices have been proposed with drive requirements compatible with Si integrated circuitry, and hence are particularly relevant to this study. These include a LiNbO₃ phase modulator [35], a cantilevered beam-deflector made in Si [36], and a deformable mirror device [37]. The last two approaches, being mechanical in nature, are somewhat limited in their speeds (≈ 10 kHz) while the phase modulator needs a complicated optical system for conversion of phase modulation to intensity modulation. A recent proposal for using the electroabsorption effect in GaAs with guided or unguided optical waves has the advantage of potentially high-speed operation (≈ 1 GHz) and of direct intensity modulation [38].

In the detector area, avalanche photodiodes, p-i-n photodiodes, and photoconductors are candidate devices [29]. Avalanche devices require large biases (≈ 50 V) and are relatively temperature sensitive, and photoconductors have a relatively low impedance. Thus p-i-n photodiodes are the most attractive for interconnects. These diodes require only a few volts bias and have quantum efficiencies of approximately 70 percent. High-performance devices have been made in GaAs as well as Si. Self-scanned Si detector arrays are also well developed with scanning rates up to 40 MHz available commercially.

B. Circuits

For circuit interconnects, it is quite relevant to ask what overall transduction efficiency is obtainable using opto-

electronics. Laser efficiencies for devices with uncoated mirrors can approach approximately 30 percent, but minimum power dissipation will remain in the 1-mW range, due to the 1-V diode turn-on and 1-mA current fundamentally needed for lasing [39]. Optical signals in the 0.5-mW range could produce signals from a photodiode–transimpedance preamplifier of approximately 100 mV. To achieve this signal level (at frequencies typically ≤ 100 MHz) requires at least two transistors; subsequent amplification to volt level logic states will require a few additional transistors. Thus it is not unreasonable to consider integrating these detection circuits on silicon ICs if the photodiode fabrication process can be made compatible with logic circuit fabrication. This could be more of a difficulty with MOS than with bipolar technology.

An alternative method for changing logic states is by direct optical injection into a gate. Initial results utilizing this technique have been reported for both Si [40] and GaAs [41] circuits. At present the peak optical power levels required exceed those available from the low-threshold diode lasers discussed here, but more progress in this area is anticipated.

C. Interconnect Elements

Optics is attractive for interconnects because of the inherent noninteraction of multiple photon beams passing through or near one another. Here, media to be considered include free space, optical fibers, and integrated optical waveguides. An attractive means for exploiting optical beam noninteraction is to use free-space propagation with either focused (e.g., holographic) or unfocused techniques. Holographic optical elements are fairly well developed. They may be written with visible light in dichromated gelatin or silver halide emulsions. Reflective elements with efficiencies limited only by surface reflections (i.e., with efficiencies in the high 90 percentiles) can be realized in dichromated gelatin when imaging with visible light [42]. In bleached silver halide materials, high diffraction efficiencies are hard to achieve in reflective elements due to the limited spatial frequency response of even high-resolution materials, but efficiencies of the order of 70 percent are readily achieved on transmission [43]. Comparable performance should be possible in imaging diode laser or LED emission in the near infrared. Computer-generated holograms can also be constructed, in some cases with the help of electron-beam lithography for writing the hologram. In any of these cases, the equivalent of $F/1$ optical systems can be achieved.

Due to fiber-optic systems, multimode fibers and associated components such as microoptic lenses, star couplers, and wavelength multiplexing modules are under development with some commercial availability. These components could lend themselves to signal distribution. For example, an N -port fiber star coupler can distribute a signal at one of N input ports equally to N output channels. Recently, a coupler with $N = 100$ was reported with total channel loss of 5 ± 0.05 dB over that expected for equal power division [44].

Another candidate for signal distribution is integrated optics. While most work has been reported for guides in electrooptic materials such as LiNbO₃ and GaAs [45], numerous workers have fabricated guides in glass or in

oxide films on Si substrates. These latter guides can have quite low propagation loss (< 0.1 dB/cm) [46] and passive components such as couplers and splitters can be formed. Here the work on multimode as well as single-mode devices is relevant.

The interconnection can be made changeable by inclusion of an active element in these schemes. With holographic optical elements, a tantalizing (though long-range) possibility is the incorporation of dynamic holographic materials, such as those now being studied for four-wave mixing applications. A shorter term solution might entail a bank of holographic mapping elements in conjunction with a real-time mask which selects the appropriate interconnection pattern. Some candidates for such a real-time mask are matrix-addressed liquid-crystal devices, and the matrix-addressed magnetooptic spatial light modulator [47]. Another approach for a dynamic interconnect could be the implementation of an optical crossbar switch as a special case of an optical matrix–vector multiplier [48]. Dynamic interconnects can also be obtained in the domain of guided-wave optics using integrated directional couplers in LiNbO₃ [49]. Cascading several such two-port switches can yield an arbitrary interconnection with high speed capabilities.

D. Hybrid and Monolithic Approaches

To discuss the use of optics in electronic interconnections raises immediately questions of materials technology, since optical sources cannot be formed from silicon. Hybrid and monolithic approaches are conceivable. In the former area, hybrid laser–amplifier circuits are available from a number of suppliers. For detection, hybrids containing a Si diode and preamplifiers are available. These hybridization techniques have to date followed conventional technological approaches to achieve several hundred-megahertz bandwidth. Although monolithic approaches can minimize parasitic capacitance for increased performance, there has been only limited work on monolithic Si photoreceivers, due to process compatibility issues and limited needs for bandwidth.

In the monolithic area, there have been limited reports of GaAs edge-emitting lasers monolithically integrated with electronics. In particular, FET drivers with lasers have been reported, including modulation up to 1 GHz [50]. Recently, a 4:1 multiplexer and laser were reported with speeds to 150 MHz [51]. Monolithic GaAs p-i-n diode–FET amplifiers have also been reported by several groups, and response times compatible with 500 Mbits/s have been achieved [52]. In each case, difficulties associated with process incompatibility between optoelectronic and electronic device requirements had to be overcome.

Similar efforts on both monolithic and hybrid InP-based circuits are being pursued because of the important telecommunications applications. However, these sources typically emit at $1.3 \mu\text{m}$, a wavelength not readily detected by Si diodes. Thus this work will have most relevance to wide-band fiber interconnects between machines which may be remotely located, a topic which, as previously mentioned, will not be discussed here.

An exciting topic in monolithic integration is the emerging area of heteroepitaxy. Here one attempts to grow crystalline films of one semiconductor on a dissimilar (and thus not lattice matched) crystalline substrate. The most recent

relevant example for the interconnect application is the growth of GaAs on Si and the subsequent fabrication of devices in both materials on the same wafer. Initial reported growth results utilized an intervening layer of Ge to relieve part of the lattice mismatch [53], but dislocation densities were too high (10^7 cm^{-2}) to fabricate practical devices. Recently an overgrowth technique has been reported which has reduced the dislocation density to 10^3 – 10^4 cm^{-2} [54], so that optoelectronic device fabrication in the GaAs layers looks promising.

V. POSSIBLE APPLICATIONS OF OPTICAL INTERCONNECTS

A. Introduction

Some of the difficulties expected in the further extension of integrated electronic technology to circuits and systems with ever greater complexity have been mentioned in Section II. In what respects does optics offer a potential solution to some of these problems?

Propagation of signals on metallic interconnections is governed by the basic laws of circuits containing distributed elements. The exact character of the distributed elements depends on whether interconnections are considered at the chip level or the board level. At the chip level, interconnections can be viewed as distributed and unterminated *RC* transmission lines [11], [18]. At the board level, the transmission lines may contain significant distributed inductance and may be terminated or unterminated. The velocity of signal propagation on such transmission lines depends on the capacitance per unit length. Thus as more and more devices having capacitive components of admittance are attached to an interconnection, the velocity of propagation decreases, and the time required for charging the line to a predetermined voltage level increases.

By way of contrast, propagation of optical signals, whether confined to waveguides or through free space, takes place at a speed that is independent of the number of components that receive those signals, namely, at the speed of light in the medium of concern. Thus we identify the first of several advantages of optical interconnects, namely the *freedom from capacitive loading effects*. Such freedom is responsible for the greater flexibility of optical interconnections with respect to fan-in and fan-out, vis-a-vis electrical interconnections.

A second advantage of optical interconnections over their electronic counterparts is their superior *immunity to mutual interference effects*. The stray capacitances that exist between proximate electrical paths introduce cross-coupling of information to a degree that increases with the bandwidth of the signals of interest. In contrast, optical interconnects suffer no such effects, although care must be exercised to assure that light scattering does not introduce a similar result (of different origin). In any case, there is no electrical coupling between the high-frequency modulations of two proximate beams of light.

A third potential advantage of optical interconnect technology lies in *freedom from planar or quasi-planar constraints*. Waveguides can pass through waveguides without significant cross-coupling (provided the angle of intersection exceeds about 10°), and free-space light beams can pass through free-space light beams without significant

interaction. Such flexibility can simplify the problems associated with routing signals on a complex chip or board.

A fourth advantage to be considered lies in the potential for certain types of optical interconnections (namely, free-space focused interconnects, to be discussed in more detail shortly) to achieve *reprogramming by means of a dynamic optical interconnect component*. In principle, the interconnection pattern associated with a chip can be changed at will, by writing appropriate information to the interconnect element. Such a capability would be rather difficult to realize using conventional electronic interconnect technology.

Finally, a fifth reason for considering optical interconnections lies in the possibility (discussed in Section IV-B) of *direct injection of optical signals into electronic logic devices*. Such intimate coupling of optical signals into electronic devices, with the bypassing of separate detectors for optical-to-electronic conversion, could greatly simplify the interface between the optical interconnect and electronic device technologies, as well as offer significant speed advantages vis-a-vis purely electronic connections to the logic, provided suitable low-power approaches can be found.

For optoelectronics to be useful for VLSI interconnections, the size, efficiency, and power requirements must be compatible with IC environments. This is in fact the case, based on the previously given review of component capabilities. For example, detector areas of about $100\text{-}\mu\text{m}$ diameter with about 1 mW of incident optical power should have acceptable power dissipation and signal levels. The overall efficiency of laser drive current to detector output current can approach 20 percent and only a modest number of transistors should be needed after the detector to produce logic-level signals.

B. Classification of Optical Interconnect Problems

In order to discuss the optical interconnect issue, it is useful to describe a hierarchy of connections with some perspective on size considerations. The fundamental chip we take for example is a $10 \times 10\text{-mm}$ MOS chip, having $\geq 300\,000$ transistors. This level of complexity is less than that of recently reported NMOS circuits, but in general, at this level of complexity the chip can be wiring dominated and the "pin-out" problem can be severe. Such chips have typically $100\text{-}\mu\text{m}$ center-to-center spacing for bond pads at the periphery and are mounted on a chip carrier with about 50-mil spacing between contacts at its outer edge.

Levels of interconnects to be considered include within a single chip, at the wafer level between undiced chips, between packaged chips on a common board, and between boards. Here it is assumed that a board will be about 1×1 ft and will contain about 100 chips. As previously mentioned, the highest level interconnect, between machines, will not be considered here.

Within this hierarchy of interconnections it is appropriate to consider several types of signal distribution. The most straightforward case is clock distribution. Here an identical signal is distributed to a large number of nodes. Since only one-way fixed communication channels are involved, the source can be a separate discrete circuit. The more general case is the distribution of data and control signals. Here more specialized and two-way links would be desirable. In addition, since various degrees of flexibility in the intercon-

nects are desirable, both static and dynamic interconnects need to be explored. In the following sections, both clock and data distribution will be considered. For each case, the hierarchy of interconnections will be matched to the available optical technologies summarized previously.

C. The Problem of Clock Distribution

A problem that appears amenable to immediate attack using optical technology is that of clock distribution at the chip, wafer, or board level. Most (but not all) computing architectures require synchronous operation of a multitude of devices, circuits, and subsystems. Synchronism is maintained by distributing to all parts of the system a timing signal, called the clock. One of the chief difficulties encountered in designing circuits and systems for high-speed operation is the phenomenon known as "clock skew," a term which refers to the fact that different parts of the circuit or system receive the same state of the clock signal at different times. In this section we consider several possible approaches to using optics for distribution of the clock, with the aim of minimizing or eliminating clock skew. Attention is first focused on the problem of distributing the clock within a single chip. Consideration is then given to the problem at the wafer and board levels.

1. Intra-Chip Clock Distribution: The interconnections responsible for clock distribution are characterized by the facts that they must convey signals to all parts of the chip and to many different devices. These requirements imply long interconnect paths and high capacitive loading. Hence the propagation delays are large and depend on the particular configuration of devices on the chip. Here we consider methods for using optics to send the clock to various parts of the chip. It is assumed that optics is used in conjunction with electronic interconnects, in the sense that optical signals might be used to carry the clock to various major sites on the chip, from which the signals would be further distributed, on a local basis, by a conventional electronic interconnection system.

The clocks used in MOS technology are generally two-phase [14]. Presumably only one of these phases will be distributed optically, the other being generated on the chip after the detection of the optical timing signal.

A variety of optical techniques can be envisioned for accomplishing the task at hand, and therefore we devote some time to delineating these various approaches and specifying their strengths and weaknesses. The main distinction between these approaches occurs in the method used to convey light to the desired locations on the chip.

Index-guided optical interconnections: The first major category of optical interconnect techniques we refer to as "index guided." Light is assumed to be carried from some single source generating an optical signal modulated by the clock to many other sites by means of waveguides. The waveguides could be of either of two types. One type could use optical fibers for carrying the optical signals. The second type could use optical waveguides integrated on a suitable substrate.

If fibers are chosen as the interconnect technology, then the following approach, illustrated in Fig. 6, might be used. A bundle of fibers is fused together at one end, yielding a core into which light from the modulated optical source (probably a laser diode) must be coupled. Light coupled in

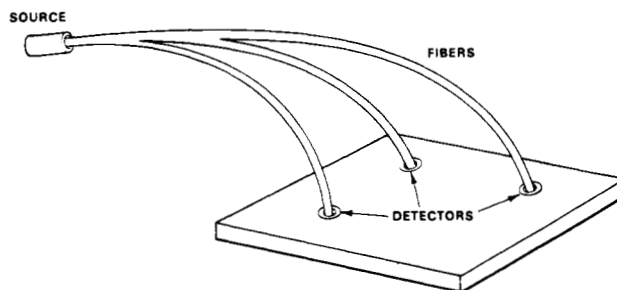


Fig. 6. Distribution of the clock by means of fibers.

at the fused end is split as the cores separate, and transmitted to the ends of each of the fibers in the bundle. Each fiber end must now be carefully located over an optical detector that will convert the optical clock to an electrical one. Alignment of the fiber and the detector might be accomplished with the help of micropositioners (analogous to a wire bonding machine), and UV-hardened epoxy could be used to hold the fiber in its proper place permanently. The difficulties associated with the fiber-optic approach stem from the alignment requirements for the fibers and detectors, and from the uniformity requirements for the fused-fiber splitter. It should also be noted that the fibers cannot be allowed to bend too much, for bends will cause radiation losses that may become severe. Lastly, we should mention that the use of fibers, and the requirements regarding allowable degrees of bending, imply that this interconnect technology will occupy a three-dimensional volume, rather than being purely planar, and this property could be a disadvantage in some applications.

If integrated optical waveguides are chosen as the interconnect technology, then the geometry might be that shown in Fig. 7. The waveguides might be formed by sputtering of

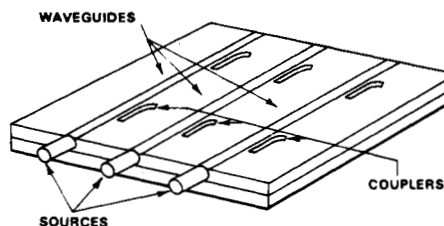


Fig. 7. Distribution of the clock by means of integrated optical waveguides.

glass onto a silicon dioxide film on the Si substrate. These guides are shown as straight in the figure, a configuration chosen again because of the large losses anticipated if this type of light guide is bent at a large angle. Optical signals must be coupled into each of the separate guides. Such signals might be generated by a single laser diode and carried to the waveguides by fibers, or separate sources might drive each of the guides, with the clock distributed to the different sources electrically. Presumably light must be coupled out of each of the straight waveguides at several sites along its length, with a detector converting the optical signal to electronic form at each such site. The difficulties associated with the waveguide approach to the problem, neglecting the bending problem which has been intentionally avoided, stem primarily from the requirement

to efficiently couple into and out of the guides. Careful alignment of the sources or fibers with the integrated waveguides is required, and couplers with short lengths are desired to remove the light from the guides and place it onto the appropriate detectors. Present waveguide technology requires distributed couplers with rather large dimensions ($5\text{ }\mu\text{m} \times 1\text{ mm}$) compared with the feature sizes normally thought of in electronic IC technology. A major advantage of the integrated optics distribution system lies in its planar character and the small excess volume it requires. A disadvantage is the comparatively inflexible geometry dictated by the necessity to avoid large bends of the waveguides.

Free-space optical interconnections: A second major category of optical interconnects can be referred to as "free-space" techniques. For such interconnects, the light is not guided to its destination by refractive index discontinuities, but rather by the laws that govern the propagation of light in free space. It is helpful to distinguish between two types of free-space interconnect techniques, "unfocused" and "focused."

Unfocused interconnections are established simply by broadcasting the optical signals carrying the clock to the entire electronic chip. One such approach is shown in Fig. 8. A modulated optical source is situated at a focal point of

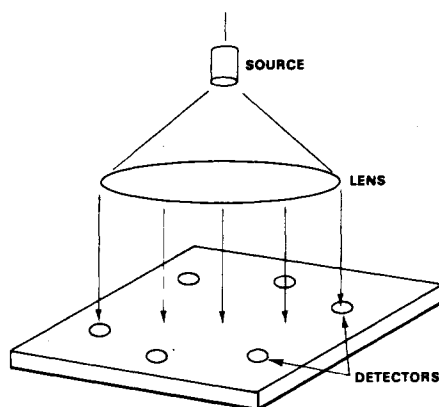


Fig. 8. Unfocused broadcast of the clock to the chip.

a lens that resides above the chip. The signal transmitted by that source is collimated by the lens, and illuminates the entire chip at normal incidence. Detectors integrated in the chip receive the optical signals with identical delays, due to the particular location of the source at the focal point of the lens. Hence in principle there is no clock skew whatever associated with such a broadcast system. However, the system is very inefficient, for only a small fraction of the optical energy falls on the photosensitive areas of the detectors, and the rest is wasted. Inefficient use of optical energy may result in requirements for the provision of extra amplification of the detected clock signals on the chip, and a concomitant loss of area for realizing the other electronic circuitry required for the functioning of the chip. Moreover, the optical energy falling on areas of the chip where it is not wanted may induce stray electronic signals that interfere with the proper operation of the chip. Therefore, it is likely that an opaque dielectric blocking layer would be needed on the chip to prevent coupling of optical signals at

places where they are not wanted. Openings in this blocking layer would be provided to allow the optical signals to reach the detectors. Alternate unfocused interconnection techniques could be imagined that use diffusers rather than a lens. Note that all such techniques require a three-dimensional volume in order to transport the signals to the desired locations.

The last category of optical interconnections is free-space "focused" interconnections, which can also be called "imaging" interconnections. For such interconnections, the optical source is actually imaged by an optical element onto a multitude of detection sites simultaneously. As indicated in Fig. 9, the required optical element can be realized by

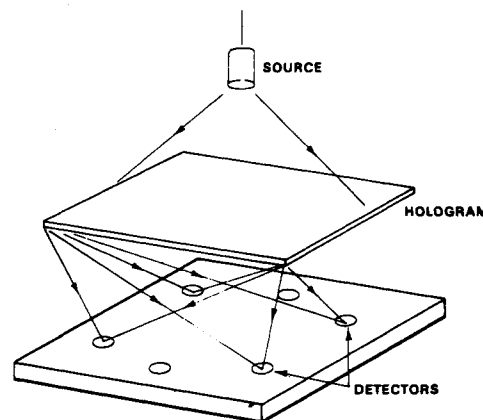


Fig. 9. Focused optical distribution of the clock using a holographic optical element.

means of a hologram, which acts as a complex grating and lens to generate focused grating components at the desired locations. The efficiency of such a scheme can obviously exceed that of the unfocused case, provided the holographic optical elements have suitable efficiency. Using dichromated gelatin as a recording material, efficiencies in excess of 99 percent can be achieved for a simple sine wave grating. When a multitude of focused spots are to be produced, the efficiency will presumably be lower, but should be well in excess of 50 percent. The flexibility of the method is great, for nearly any desired configuration of connections can be realized.

The chief disadvantage of the focused interconnect technique is the very high degree of alignment precision that must be established and maintained to assure that the focused spots are striking the appropriate places on the chip. Of course, the spots might be intentionally defocused, decreasing the efficiency of the system, but easing the alignment requirements. Thus there exists a continuum of compromises between efficiency and alignment difficulty. Fig. 10 illustrates a possible configuration that retains high efficiency but minimizes alignment problems. The imaging operation is provided by two two-element lenses, in the form of a block with a gap between the elements. A Fourier hologram can be inserted between the lenses, and it establishes the desired set of focused spots. The hologram itself consists of a series of simple sinusoidal gratings, and as such the position of the diffracted spots is invariant under simple translations of the hologram. The source is permanently fixed on the top of the upper lens block after it has been aligned with a detector at the edge of the chip,

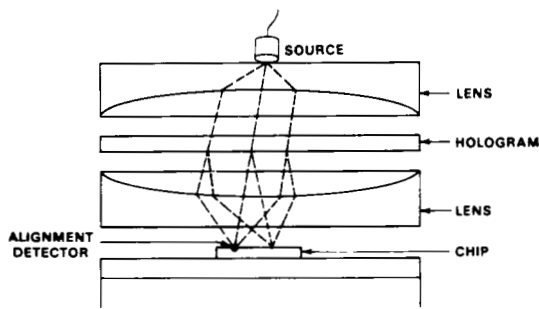


Fig. 10. Configuration for focused clock distribution that minimizes alignment problems.

thereby establishing a fixed optical axis. The only alignment required for the hologram is rotation. The position of the image spots is determined by the spatial frequencies of the gratings in the hologram, which could be established very precisely if the hologram were written, for example, by electron-beam lithography.

Focused interconnect systems, like the unfocused ones, require a three-dimensional volume above the chip. If holographic elements are used, thought must be given to the effects of using a comparatively nonmonochromatic source such as an LED. A spread of the spectrum of the source results in a spread of the focused energy, so the primary effect is to reduce the efficiency with which light can be delivered to the desired detector locations.

2. *Clock Distribution at the Wafer and Board Levels:* Clock distribution at the wafer and board levels is also a strong candidate as an application for optical interconnections. The primary differences in the optical requirements in the two cases derive from the different physical sizes of chips, wafers, and boards, as discussed above. At the chip and wafer levels, the discussion of the previous sections carries over essentially without change. At the board level, the physical areas are sufficiently large that the free-space and integrated optics approaches could at best be used for coverage of only a portion of an entire board. The preferred approach seems to lie with the use of optical fibers for conducting signals to remote parts of a single board, between which the greatest clock skews would otherwise be anticipated. Hierarchical schemes can also be envisioned, in which a network of fibers distributes the clock to a series of widely separated sites, from which either integrated optical waveguides or free-space connects are used to distribute these signals more locally to detectors, where the optical signals are converted to electronic form and distributed on an even more local scale to the various devices that require the clock signal.

D. Data Interconnects

The clock distribution problem discussed above is a particular case where long propagation distances are encountered. The more general case is data exchange between different components of a system. It was seen earlier that, if the process of scaling down feature size and increasing chip size is continued, then at a certain point the speed of the chip will be limited by the delay time associated with the interconnections between different components of the circuit, rather than the switching times of the components themselves. Therefore implementing the longer intercon-

nections optically could potentially enhance the performance of the chip. Communications between different chips are often limited by the number of pins available on the chips for communication with other chips. In addition, as interconnection lengths increase, the size and power requirements of the drivers on the chip also increase. Thus it becomes difficult to communicate with chips that are widely separated, and to have a large fan-out.

In addition to these hardware considerations, several important applications in signal and image processing demand much more flexible interconnections between processing elements of a VLSI-based parallel processor than electronic interconnect technology can provide. It was seen in Section III that some important classes of algorithms require global and dynamic interconnections between the elements of the processor array. Optical interconnections could make valuable contributions to these problems for all levels of integration, from one processor per chip to an entire array of processors on a single wafer.

The optical data interconnects utilize some components (the detectors and the interconnect elements) that are common with the problem of optical clock distribution. However, the conversion of signals from electronic to optical form poses much greater demands in the case of data interconnects. The different technologies available for direct modulation of sources, as well as external modulators (as reviewed in Section IV) will be particularly relevant here. As with the case of clock distribution, we discuss the problem of data interconnects at the two levels of intrachip and interchip communication.

1. *Intrachip Data Communication:* As a first scheme to be considered, a GaAs chip with optical sources is connected in a hybrid fashion (with conventional wire bond techniques) to a Si chip such that light is generated only along the edges of the Si chip (see Fig. 11). The sources

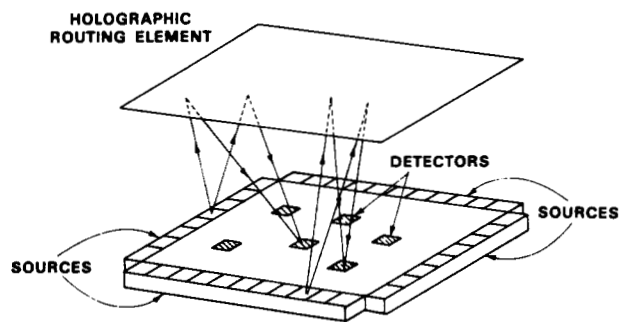


Fig. 11. Hybrid GaAs/Si approach to data communication.

could be of the edge-emitting or surface-emitting type. The optical signals are routed to the appropriate locations on the Si chip using conventional and/or holographic optical elements. The Si chip will contain detectors to receive the optical data streams generated by the sources. Since the detector-amplifier combinations can be fabricated in Si, every computational component on the Si chip could be capable of receiving data. Free-space propagation of optical signals allows a large fan-out as well as the possibility of changing the routing via a programmable two-dimensional mask, such as a reflective spatial light modulator.

In the second scheme to be considered, index-guided

structures are used for routing signals instead of free-space propagation. This scheme can be well illustrated by a specific example involving a one-dimensional parallel processor array which is interconnected by a perfect shuffle network (see Section III). The schematic diagram of the system for the case of 4 processing elements is shown in Fig. 12. The Si chip contains 4 processing elements which

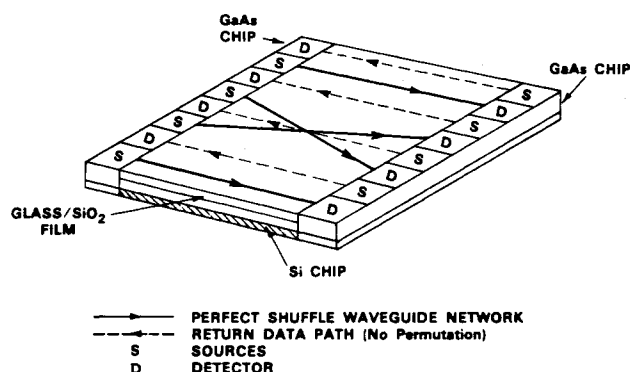


Fig. 12. Optical perfect shuffle network.

output data to the left and accept data from the right. Each processing element is associated with source-detector pairs in the GaAs chips that are mounted on both sides of the chip. Thus there exists a bidirectional optical data path for each processing element. The routing of optical data is performed by a crossing network of channel waveguides that can be formed in glass/SiO₂ layers grown on top of the Si chip. The waveguide network connecting the sources on the output side of the processor array (i.e., the left side) with the detectors on the input (right) side perform a perfect shuffle while the reverse optical paths correspond to a processing element connected to itself (no permutation). Using this scheme it is possible to shuffle data as many times as necessary to obtain the desired interconnect pattern before processing by the electronic part of the system. Since the sources and detectors can operate at several hundred megahertz rates (or more), and the optical interconnection delays are negligible, the permutation operation could be performed extremely rapidly. Such a processor design would be very efficient in implementing the FFT or Batcher's sorting algorithm.

An alternative approach to the same problem would use waveguides formed in a LiNbO₃ substrate, with active switches incorporated in the permutation network. Such a network would be programmable. A further modification might allow detection to be distributed throughout the Si chip, coupling light from LiNbO₃ waveguides to detectors via tapping mechanisms, such as grating couplers, evanescent wave coupling, etc.

In the third scheme to be discussed, electrical-to-optical conversion is carried out via external modulation of a uniform optical wavefront. In this scheme, the routing of optical signals is carried out via conventional and holographic optical elements and free-space propagation. The different modulation mechanisms were mentioned earlier along with their speed limitations. Fig. 13 shows the schematic diagram of this approach to optical interconnects, where the modulators are operated in a reflection mode.

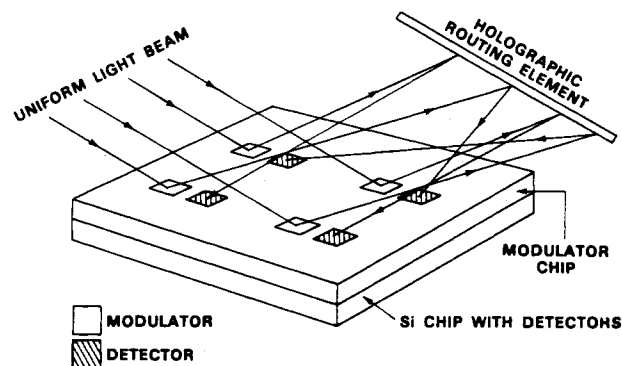


Fig. 13. Optical interconnections using external modulators operating in reflection mode.

The optical elements that perform the routing after the uniform light signals have been modulated could include a dynamic mask, thus introducing programmability in the interconnect patterns. The main advantage of using external modulators is their low power consumption and the relative ease with which components in the interior of the chip can be accessed. It should be noted that the problems encountered here in realizing the electrical-to-optical converters are somewhat similar to those encountered in the development of spatial light modulators. There are significant differences, however, in that the modulation mechanisms required for interconnect schemes need much faster response times (several megahertz versus tens of kilohertz), but need only binary (rather than analog) output (1 bit versus 10 bits).

An alternative approach to the same concept can achieve high-speed operation with low voltage levels via integrated optical modulators. As previously described, such modulators are highly asymmetric, being typically a few millimeters long and only tens of micrometers in width. Due to this shape, such an approach would make most sense when the communication distances involved are more than a few millimeters. The small widths can be exploited to fabricate a number of such modulators on a chip in order to provide one modulator for each source of data on the chip. Thus the two-dimensional arrangement of the processing elements in an array will be converted into a linear light distribution. This linear distribution can then be mapped into another linear distribution in an arbitrary fashion (including one-to-many mappings, if needed) via an optical crossbar switch, which can be realized using an optical matrix-vector multiplier, such as has been reported in the literature [48]. When the matrix mask that encodes the interconnect pattern is changeable in real time, a programmable interconnect network of the processing elements in an array is obtained. Fig. 14 shows the schematic diagram of such a system. The main limitation of this scheme will be the total number of components that can be connected this way within a chip. The theoretical maximum will be given by the number of integrated optical modulators that can be accommodated within the linear dimension of the chip. Since the routing is still performed by free-space propagation and focusing/imaging optics, the entire system will be three-dimensional, requiring large volume and a highly stable optical system. A hybrid version can also be envisioned, where the final delivery of the permuted optical data streams

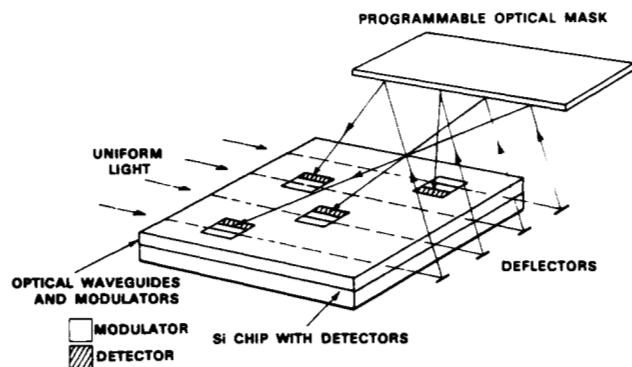


Fig. 14. Optical interconnections using integrated optical modulators.

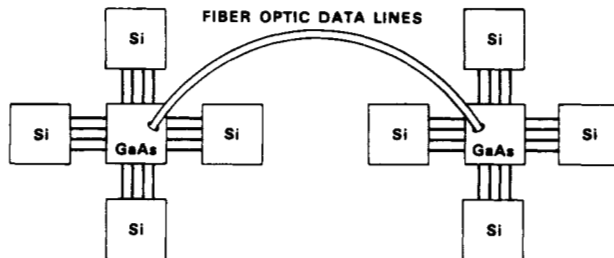


Fig. 15. Configuration with clusters of Si chips around GaAs chips, and with communication between GaAs chips via optical fibers.

to the detectors at appropriate locations on the chip can be performed with permanently bonded optical fibers.

A fourth and final approach to the data-routing problem is a monolithic one that presupposes the successful development of heteroepitaxial growth of GaAs on Si and subsequent construction of devices in both materials. The sources in the GaAs might be of the surface-emitting type. The optical routing could then be performed via free-space propagation and imaging components. The main advantage to such an approach is the optical access it provides to data sources that may be interior to the Si chip, rather than requiring that data be routed electrically to the edge of the chip before being converted into optical form.

2. Interchip Data Communication: Unlike the case of electrical interconnections, an increase in the length of optical data paths does not reduce the bandwidth of the link (all the paths of concern here are so short that optical attenuation is assumed to be insignificant). Therefore, the schemes discussed previously for intrachip communication using free-space propagation or fibers are equally applicable to interchip communication. The arguments are somewhat similar to those made for wafer- and board-level clock distribution with optics. Some advantages of this type of interchip data communication (above and beyond the potential for higher bandwidth and lower mutual interference afforded by optics) are the ability to access interior components of a chip directly and the potential for realizing programmable interconnects using real-time masks. On the other hand, optical data routing using imaging optics and free-space propagation requires careful alignment and uses the third spatial dimension, requiring a larger volume than would otherwise be the case. Therefore, optical fibers may be preferred as the communication medium at the board level.

A novel way of exploiting the large information carrying capacity of fiber optic systems would entail the use of time multiplexing. Thus several electrical connections could be replaced by a single high-speed fiber-optic line. Additional multiplexing and demultiplexing circuitry would be required on a GaAs chip. Conceptually, one can imagine a cluster of Si chips connected to a GaAs chip with short electrical connections. The GaAs chip performs the multiplexing and electrical-to-optical conversion before sending the data over a fiber-optic line to another cluster. The GaAs chip will also have detectors and demultiplexing circuits. The idea is described pictorially in Fig. 15.

VI. FUTURE DIRECTIONS

In this section our goal is to identify the most likely and profitable directions in which research on optical interconnections might move in the next few years. The view is based on the considerations discussed in previous sections, but inevitably has an element of subjectivity. These opinions should not be construed as ruling out the possibility of a "breakthrough," conceptual or practical, not foreseen in the previous discussions.

A. Clock Distribution

In many respects the problem of optical clock distribution seems to be the easiest to attack with current technology. The problem is a real and serious one. A single off-chip source is needed, and the distribution system can be as simple as global broadcast, with no focusing of light. However, presumably the extra efficiency of a focused distribution system will result in higher detected current and voltage levels, thus minimizing the need for extra amplification on the chip. The tradeoffs between optical efficiency and amplification needed on chip are interesting ones and worthy of future study.

Some experience with optically supplied clocks is needed before the benefits and drawbacks of the idea can be fully understood. Probably a combination of a macroscopic optical clock distribution feeding a collection of more local electrical distribution systems on the chip will be optimal. If unfocused broadcast is to be used, opaque overcoating layers will be needed to prevent introduction of unwanted signals at locations other than detectors. Experimentation with these ideas is badly needed.

If focused distribution systems are to be used, methods for relieving alignment problems, such as that illustrated in Fig. 10, must be developed. Further optical studies of the efficiencies of holographic optical elements in the infrared would be helpful.

B. Hybridization

Many of the proposed interconnect techniques rely on the mating of Si and GaAs chips. In some cases the GaAs provides only the optical source, but in others a more complex optoelectronic chip is needed. Initial efforts could focus on previously developed techniques for chip placement and bonding of the two dissimilar types of chip. However, to maximize the efficiency of relieving the interconnect and pin-out problems in large-scale Si circuits, the development of techniques that minimize the length of

interchip interconnect lines, and in effect eliminate the bonding pads, are needed. The pads are considerably larger than the device lines, and pins on the conventional ceramic package are even larger, as described previously. Two examples of approaches that could be pursued are a metal-bump technique and a close-packed planar hybrid. In the former technique, the electrical and optoelectronic chips would be vertically mated by a series of metal posts formed on one of the circuits. This technique has been used, for example, to connect infrared detector arrays to Si chips [55]. In the latter approach, the chips could be butted up to one another and interchip metallization formed on an intervening planarizing dielectric [56]. Manufacturable versions of these technologies could allow both more interconnects to the Si and the efficient utilization of GaAs chips.

C. Monolithic Approaches

The integrating on GaAs of both optical and electrical devices is an exciting area currently in early development, and progress was summarized in Section IV-B. Progress in integrated optoelectronics will be paced by the development of GaAs ICs. If ICs with medium-scale integration or higher complexity can be fabricated with high yield, there may be strong motivation to use optics for interconnects in all-GaAs high-speed or radiation-tolerant processors.

A long-term and conceptually appealing technique for realizing optical interconnects is to grow single-crystal GaAs on Si and then form the appropriate devices in each material, thereby eliminating the need for hybridization. In addition to optoelectronic devices, GaAs electronic devices could also be formed (e.g., high-speed digital circuits for multiplexing signals to Si circuits). This technology could lend itself to many circuit configurations, including the formation of strategically located GaAs islands on a large Si wafer. To date, initial efforts have focused on materials issues, as discussed in Section IV-D. The demonstration of heteroepitaxial circuits, of course, requires the careful determination of process-compatible approaches to Si circuit formation, GaAs layer growth, and GaAs fabrication. For example, at present the GaAs growth occurs at temperatures of $\approx 680^\circ\text{C}$, suggesting that higher temperature Si processing steps may have to occur prior to GaAs growth. This technique could, in the long run, realize in single-chip form many of the GaAs/Si hybrid approaches described earlier.

D. Perfect-Shuffle Exchange Network

The realization of a global but fixed interconnect network like the perfect shuffle with optoelectronic techniques could have a large impact on special-purpose, high-performance signal processing systems. The scheme discussed in Section V is just one possible approach. If the processing elements on the VLSI chip are not arranged in a simple linear form, as assumed in that scheme, then the use of LiNbO₃ guided-wave modulators to perform the electrical-to-optical signal transduction, as shown in Fig. 14, could be used. The optical permutation network chip consisting of linear arrays of sources and detectors, as well as the passive network of optical waveguides, can then be physically dissociated from the Si chip. In this way the processing

elements could all be on one chip, or could be distributed among different chips. In any case, the components that need most development are the individual addressable arrays of laser diodes and LiNbO₃ guided-wave modulators that can be directly addressed by a Si chip at high speed.

E. Electrically Addressed Optical Modulators

One of the most important elements of an optical data interconnect system is the external modulator that can operate on guided or free-space optical wavefronts and can be directly addressable by a Si IC. Although several schemes have been proposed and demonstrated, as reviewed in Section IV, special attention is needed to make these modulators compatible with VLSI circuits with respect to size, power consumption, and speed. A matrix-addressed light modulator can also find application in dynamic interconnect schemes involving real-time masks. Depending on the applications, the update rate of these masks could be significantly lower than the clock rates of the VLSI circuits. On the other hand, these masks will be required to have a rather large space-bandwidth product in order to provide high interconnect flexibility. Thus in general future work will be required in developing high-speed, low-power, and small-size modulators.

F. High-Speed ICs

The focus of this paper has been on alleviating the interconnect problem for VLSI MOS circuits which at present operate at clock rates no greater than 30 MHz. A potentially equally important area for optical interconnects may be for smaller scale circuits that operate at very high speeds (100 to 2000 MHz). Si ECL circuits are commercially available with clock rates greater than 100 MHz, and gigahertz circuits in short-gate Si MOS and GaAs MESFET technology have been demonstrated. While the gate counts of such circuits are at least 1 to 2 orders of magnitude smaller than for VLSI chips, their higher speed places severe stress on the isolation and capacitance of the interconnects. In such cases, the inherent high speed and low crosstalk of optical approaches is very appealing. While there are cases where incorporating an optimum number of gates on a chip can minimize the number of high-speed interconnects [57], optics could still play a significant role. Many of the hybridization techniques described earlier could be applied to these high-speed situations. Hybridization of optical chips with fast complex Si circuits is especially attractive due to the recent report of integration of fast CMOS circuits with bipolar circuitry [58]. This achievement could eliminate the need for scaling up the MOS devices (and thus using wafer area and potentially slowing down the devices) in order to drive the optical sources. The most appealing long-term solution for high-speed applications may be the development of all-GaAs systems, since obviously the optical devices could then be integrated, as described above.

ACKNOWLEDGMENT

This paper was the result of an Army Research Office Palantir study. These studies address the physical founda-

tions of approaches to solutions of important technological problems with the aim of stimulating new avenues for progress toward their solution. The participants in the study consisted of the authors of this paper. S. Y. Kung served as chairman. The authors thank Dr. B. Guenther of ARO for his sponsorship and assistance. One of the authors (F. J. L.) wishes to acknowledge partial support by the Department of the Air Force.

REFERENCES

- [1] *Proc. IEEE* (Special Section on Acoustooptic Signal Processing), vol. 69, no. 1, pp. 3–5, 48–118, Jan. 1981.
- [2] N. J. Berg and J. N. Lee, *Acousto-Optic Signal Processing*. New York, NY: Marcel Dekker, 1983.
- [3] L. J. Cutrona, E. N. Leith, L. J. Porcello, and W. E. Vivian, "On the application of coherent optical processing techniques to synthetic aperture radar," *Proc. IEEE*, vol. 54, no. 8, pp. 1026–1032, 1966.
- [4] A. Kozma, E. N. Leith, and N. G. Massey, "Tilted plane optical processor," *Appl. Opt.*, vol. 11, no. 8, pp. 1766–1777, 1972.
- [5] For a review of this area, see H. J. Caulfield, J. A. Neff, and W. T. Rhodes, "Optical computing: The coming revolution in optical processing," *Laser Focus*, vol. 19, no. 11, pp. 100–110, 1983.
- [6] D. Psaltis, D. Casasent, D. Neft, and M. Carlotto, "Accurate numerical computation by optical convolution," *Proc. SPIE*, vol. 232, pp. 160–167, 1980.
- [7] A. Huang, Y. Tsunoda, J. W. Goodman, and S. Ishihara, "Some new methods for performing residue arithmetic operations," *Appl. Opt.*, vol. 18, no. 2, pp. 160–167, 1980.
- [8] C. Y. Yen and S. A. Collins, Jr., "Operation of a numerical optical processor," *Proc. SPIE*, vol. 232, pp. 160–167, 1980.
- [9] A. A. Sawchuk and T. C. Strand, "Fourier optics in nonlinear signal processing," in *Applications of Optical Fourier Transforms*, H. Stark, Ed. New York, NY: Academic Press, 1982, ch. 9.
- [10] C. Warde, A. M. Weiss, and A. D. Fisher, "Optical information processing characteristics of the microchannel spatial light modulator," *Appl. Opt.*, vol. 20, no. 12, pp. 2066–2074, 1981.
- [11] K. C. Saraswat and F. Mohammadi, "Effect of scaling of interconnections on the time delay of VLSI circuits," *IEEE Trans. Electron Devices*, vol. ED-29, no. 4, pp. 645–650, 1982.
- [12] R. W. Keyes, "Communication in computing," *Int. J. Theor. Phys.*, vol. 21, nos. 3/4, pp. 263–273, 1982.
- [13] A. J. Blodgett, Jr., "Microelectronic packaging," *Sci. Amer.*, vol. 249, no. 7, pp. 86–96, 1983.
- [14] C. Mead and L. Conway, *Introduction to VLSI Systems*. Reading, MA: Addison-Wesley, 1980.
- [15] D. J. Kinniment, "VLSI and machine architecture," in *VLSI Architecture*, B. Randell and P. C. Treleaven, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1983, pp. 24–33.
- [16] F. Anceau and R. Ries, "Design strategy for VLSI," in *VLSI Architecture*, B. Randell and P. C. Treleaven, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1983, pp. 128–137.
- [17] R. W. Keyes, "Fundamental limits in digital information processing," *Proc. IEEE*, vol. 69, no. 2, pp. 267–278, 1981.
- [18] P. M. Solomon, "A comparison of semiconductor devices for high-speed logic," *Proc. IEEE*, vol. 70, no. 5, pp. 489–509, 1982.
- [19] G. W. Preston, "The very large scale integrated circuit," *Amer. Scientist*, vol. 71, no. 5, pp. 466–472, 1983.
- [20] R. W. Keyes, "The evolution of digital electronics towards VLSI," *IEEE Trans. Electron Devices*, vol. ED-26, no. 4, pp. 271–278, 1979.
- [21] S. Y. Kung, K. S. Arun, R. J. Gal-Ezer, and D. V. Bhaskar Rao, "Wavefront array processor: language, architecture, and applications," *IEEE Trans. Comput.* (Special Issue on Parallel and Distributed Computers), vol. C-31, no. 11, pp. 1054–1066, 1982.
- [22] P. A. Gilmore, "The massively parallel processor (MPP): A large scale SIMD processor," *Proc. SPIE*, vol. 431, pp. 166–174, 1983.
- [23] R. A. Athale, "Optical matrix algebraic processors," in *Proc. 10th Int. Optical Computing Conf.*, IEEE Cat. No. 83CH1880-4, Apr. 1983, pp. 24–31.
- [24] H. T. Kung, "Why systolic architectures?" *IEEE Computer*, vol. 15, no. 1, pp. 37–46, 1982.
- [25] R. S. Schreiber, "On the systolic arrays of Brent, Luk, and van Loan," *Proc. SPIE*, vol. 431, pp. 72–76, 1983.
- [26] A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975, pp. 284–337.
- [27] H. S. Stone, "Parallel processing with the perfect shuffle," *IEEE Trans. Comput.*, vol. C-20, no. 2, pp. 153–161, 1971.
- [28] K. E. Batchler, "Sorting networks and their applications," in *1968 Spring Joint Comp. Conf., AFIPS Proc.*, vol. 32 (Washington, DC, 1968).
- [29] See, for example, H. Kressel, *Semiconductor Devices for Optical Communications*. Berlin, Germany: Springer-Verlag, 1982.
- [30] D. R. Scifres, R. D. Burnham, and W. Streifer, "Output coupling and distributed feedback utilizing substrate corrugations in double heterostructure GaAs lasers," *Appl. Phys. Lett.*, vol. 27, no. 5, 1975.
- [31] K. Iga, H. Soda, T. Terakado, and S. Shimizu, "Lasing characteristics of improved GaInAsP/InP surface emitting injection lasers," in *Proc. 4th Int. Conf. on Integrated Optics and Optical Fiber Communications* (Tokyo, Japan, June 27–30, 1983), pp. 198–199.
- [32] H. Grothe, G. Muller, and W. Harth, "560 Mb/s transmission experiments using 1.3 μm InGaAsP/InP LED," *Electron. Lett.*, vol. 19, no. 22, pp. 909–911, 1983.
- [33] R. A. Becker, "Broad-band guided-wave electrooptic modulators," to be published in *IEEE J. Quantum Electron.*, vol. QE-20, pp. 723–727, July 1984.
- [34] R. A. Boenning, V. B. Morris, and E. G. Vaerewyck, "PC board test signal extraction," presented at the 29th Int. Instrumentation Symp., Instrument Society of America, Albuquerque, NM, May 1983.
- [35] R. V. Johnson, D. L. Hecht, R. A. Sprague, L. N. Flores, D. L. Steinmetz, and W. D. Turner, "Characteristics of the linear array total internal reflection (TIR) electrooptic spatial light modulator for optical information processing," *Opt. Eng.*, vol. 22, no. 6, pp. 665–674, 1983.
- [36] K. E. Petersen, "Micromechanical light modulator array fabricated on silicon," *Appl. Phys. Lett.*, vol. 31, p. 521, 1977.
- [37] D. R. Pape and L. J. Hornbeck, "Characteristics of the deformable mirror device for optical information processing," *Opt. Eng.*, vol. 22, no. 6, pp. 675–681, 1983.
- [38] R. A. Kingston, B. E. Burke, K. B. Nichols, and F. J. Leonberger, "Spatial light modulation using electroabsorption in a GaAs charge-coupled device," *Appl. Phys. Lett.*, vol. 41, no. 5, pp. 413–415, 1982.
- [39] J. N. Walpole, MIT Lincoln Laboratory.
- [40] M. L. Levy, "An investigation of flaws in complex CMOS devices by a scanning photoexcitation technique," in *Proc. 15th Annu. IEEE Reliability Symp.*, pp. 44–53, 1977.
- [41] R. K. Jain and D. E. Snyder, "Switching characteristics of logic gates addressed by picosecond light pulses," *IEEE J. Quantum Electron.*, vol. QE-19, no. 4, pp. 658–663, 1983.
- [42] B. J. Chang, "Dichromated gelatin as a hologram storage medium," *Proc. SPIE*, vol. 177, pp. 71–81, 1979.
- [43] K. Biedermann, "Silver halide photographic materials," in *Holographic Recording Materials*, H. M. Smith, Ed. Berlin, Germany: Springer-Verlag, 1977, p. 69.
- [44] Y. Fujii, M. Suzuki, and J. Minowa, "A 100 input/output star coupler composed of low-loss slab waveguide," in *Proc. 4th Int. Conf. on Integrated Optics and Optical Fiber Communication*, (Tokyo, Japan, June 27–30, 1983), paper 29C2-4, pp. 342–343.
- [45] See, for example, *Tech. Digest of the Topical Meeting on Guided Wave Optics* (Asilomar, CA), IEEE Cat. No. 82CH 1719-4, 1982.
- [46] S. Dutta, H. E. Jackson, and J. T. Boyd, "Scattering loss reduction in ZnO optical waveguides by laser annealing," *Appl. Phys. Lett.*, vol. 39, no. 3, pp. 206–208, 1981.

- [47] W. E. Ross, D. Psaltis, and R. H. Anderson, "Two-dimensional Magneto-optic spatial light modulator for signal processing," *Proc. SPIE*, vol. 341, p. 191, 1982.
- [48] J. W. Goodman, A.R. Dias, and L. M. Woody, "Fully parallel, high speed incoherent optical method for performing the discrete Fourier transform," *Opt. Lett.*, vol. 2, no. 1, pp. 1-3, 1978.
- [49] M. Kondo, Y. Ohta, M. Fujiwara, and M. Sakaguchi, "Integrated optical switch matrix for single-mode fiber networks," *IEEE J. Quantum Electron.*, vol. QE-18, no. 10, pp. 1759-1765, 1982.
- [50] M. Kim, C. Hong, D. Kasemet, and R. Milano, "GaAlAs/GaAs integrated optoelectronic transmitter using selective MOCVD epitaxy and planar ion implantation," in *Proc. GaAs IC Symp.*, pp. 44-47, Oct. 1983.
- [51] J. Carney, M. Helixal, and R. Kolby, "Gigabit optoelectronic transmitter," in *Proc. GaAs IC Symp.*, pp. 48-51, Oct. 1983.
- [52] For a review, see N. Bar-Chaim, S. Margalit, A. Yariv, and I. Ury, "GaAs integrated optoelectronics," *IEEE Trans. Electron Devices*, vol. ED-29, no. 9, pp. 1372-1381, 1982.
- [53] B. Y. Tsaur, M. W. Geis, J. C. C. Fan, and R. P. Gale, "Hetero-epitaxy of vacuum-evaporated Ge films on single-crystal Si," *Appl. Phys. Lett.*, vol. 38, no. 10, pp. 779-781, 1981.
- [54] B. Y. Tsaur, R. W. McClelland, J. C. C. Fan, R. P. Gale, J. P. Salerno, B. A. Vojak, and C. O. Bozler, "Low dislocation density GaAs epilayers grown on Ge-coated Si substrates by means of lateral epitaxial overgrowth," *Appl. Phys. Lett.*, vol. 41, no. 4, pp. 347-349, 1982.
- [55] D. H. Pommerrenig, D. D. Enders, and T. E. Meinhardt, "Hybrid silicon focal plane development: an update," *Proc. SPIE*, vol. 267, pp. 23-30, 1981.
- [56] A. Chu, MIT Lincoln Lab., private communication.
- [57] B. Gilbert, T. Kinter, S. Hartley, and A. Firstenberg, "Exploitation of GaAs digital integrated circuits in wideband signal processing environments," in *Proc. GaAs IC Symp.*, pp. 58-61, Oct. 1983.
- [58] J. Miyamoto, S. Saitoh, H. Momose, H. Shibata, K. Kanazaki, and S. Kohyama, "A 1.0 μm N-well CMOS/bipolar technology for VLSI circuits," in *Tech. Dig. Int. Electron Device Meet.*, pp. 63-66, Dec. 1983.