Optoelectronic parallel-matching architecture: architecture description, performance estimation, and prototype demonstration

Keiichiro Kagawa, Kouichi Nitta, Yusuke Ogura, Jun Tanida, and Yoshiki Ichioka

We propose an optoelectronic parallel-matching architecture (PMA) that provides powerful processing capabilities in global processing compared with conventional parallel-computing architectures. The PMA is composed of a global processor called a parallel-matching (PM) module and multiple processing elements (PE's). The PM module is implemented by a large-fan-out free-space optical interconnection and a PM smart-pixel array (PM-SPA). In the proposed architecture, by means of the PM module each PE can monitor the other PE's by use of several kinds of global data matching as well as interprocessor communication. Theoretical evaluation of the performance shows that the proposed PMA provides tremendous improvement in global processing. A prototype demonstrator of the PM module is constructed on the basis of state-of-the-art optoelectronic devices and a diffractive optical element. The prototype is assumed for use in a multiple-processor system composed of 4×4 PE's that are completely connected through bit-serial optical communication channels. The PM-SPA is emulated by a complex programmable device and a complementary metal-oxide semiconductor photodetector array. On the prototype demonstrator the fundamental operations of the PM module were verified at 15 MHz.

OCIS codes: 200.4650, 200.2610, 070.4560, 200.3050.

1. Introduction

Parallel distributed processing is an effective method for advancing the performance of computing systems.[1] The performance of parallel-computing systems is dominated by many factors such as network topology, communication bandwidth, task-scheduling methods, and memory architectures.[2] In particular, the bandwidth of the network is critical. However, the performance of parallel-computing systems that are embodied by conventional electronic technologies has been limited because of the bottlenecks associated with electronic planar-interconnection technologies.

For clearing the bottlenecks photonic networks based on three-dimensional free-space optical interconnection (FSOI) and optoelectronic VLSI (OE-VLSI) have become a significantly promising option for parallel computers.[3,4] An OE-VLSI, or what is called a smart-pixel array (SPA), is an optoelectronic device equipped with high-density optical input–output ports on VLSI electronic circuitry, which can provide ultrahigh-speed and dense interconnection, utilizing the large space–bandwidth product and the high-speed propagation of free-space optics. For example, operating speeds greater than 1 GHz[5,6] and integration of more than 1000 pixels on a chip[3] have been reported.

Recently several demonstrators such as an optical backplane,[7] an optoelectronic crossbar network,[8] an optical multimesh hypercube,[9] and the FAST-Net[10] were presented to exploit the potential applicability of the smart-pixel-based FSOI to parallel computers. These demonstrators aimed to show high capabilities of global interconnection with wide communication bandwidths in parallel-computing systems. In most FSOI systems the communication mechanism is implemented by the combination of optical data fan-out, shuffle, and electronic selection.[8,9] To make the best use of the excellent features of large fan-outs and the parallel data shuffle of the FSOI, one should consider the processing algorithm in the system design.

K. Kagawa (kagawa@mls.eng.osaka-u.ac.jp), K. Nitta, Y. Ogura, and J. Tanida are with the Department of Material and Life Science, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan. Y. Ichioka is with the Nara National College of Technology, 22 Yata-cho, Yamatokoriyama, Nara 639-1090, Japan.

Received 30 November 1999; revised manuscript received 1 August 2000.

0003-6935/01/020283-16$15.00/0

From the viewpoint of the system architecture the existing demonstrators do not necessarily utilize optical data fan-out in an effective manner because most of the fanned-out information is discarded.

In this paper we focus on a bandwidth mismatch between the bandwidth of local processing at processing elements (PE's) and the required bandwidth of inter-PE communication in global processing, which is a problem that emerges when a photonic network is introduced into parallel-computing systems. Here we define global processing as the types of operations that require multiple data from multiple PE's. Because communication among the PE's and processing are implemented separately in the conventional parallel-computing architectures, N data from N slave PE's are transferred to the master PE through a network hub to execute global processing. During the procedure the traffic of the communication path between the master PE and the slave PE's is N times as large as the bandwidth of the communication path between the network hub and the PE's. This situation constitutes the bandwidth mismatch, which reduces the throughput of the parallel computing system. This bottleneck caused by the bandwidth mismatch cannot be eliminated by a simple increase in the communication capacity of the network. We can conclude that the conventional parallel-computing architectures are not always suitable for global processing. The existing photonic network demonstrators also inherently have this bottleneck.

We believe that the smart-pixel-based FSOI approach is attractive not only for global data switching but also for global data processing. It has the potential capability to improve the processing performance of parallel-computing systems by use of global processing. The effective method that we propose to eliminate the bottleneck is to integrate the network hub with the master PE by use of the smart-pixel-based FSOI. In this paper we call the hybrid master PE a global processor. Because a huge amount of the data from multiple PE's is processed at once by smart pixels, the input and the processing bandwidths of the global processor will become much larger than the simple master PE. Consequently, the bandwidth of processing at the global processor and the data rate of the fanned-in data become comparable, which means that there is no bandwidth mismatch. This configuration is based on an optoelectronic heterogeneous architecture previously presented by Tanida *et al*.[11]

As an instance of the optoelectronic heterogeneous architecture, we propose an optoelectronic parallel-matching architecture (PMA), which is an effective parallel-computing architecture that is suitable for global processing. A system that is based on the PMA has the ability to execute global processing without degrading the throughput of the whole system. Detection of the PE's that satisfy a given condition and a given summation of absolute differences over the multiple PE's are typical examples of the global processing. The global processor of the PMA is called a parallel-matching (PM) module and consists of a large-fan-out FSOI and a parallel-matching smart-pixel array (PM-SPA). In Section 2 the concept of the PMA and the functions of the PM module are described. In Section 3 an optoelectronic embodiment of the PM module is presented. In Section 4 the performance of the PMA is evaluated theoretically and compared with other optoelectronic networks. In Section 5 the design of the components of the prototype demonstrator is explained. The experimental results of the prototype are presented in Section 6. In Section 7 several issues of the PMA are discussed.

2. Parallel-Matching Architecture

A. Distributed Optimization Problems

We assume a multiple-instruction–multiple-data stream parallel-computing system consisting of N PE's that are embodied as OE-VLSI's. The PE's are connected to each other through a photonic network with a specific network topology. A heuristic optimization algorithm based on the distributed algorithm is a good application of the parallel computing systems. The genetic algorithm[12] is an example of the distributed heuristic optimization algorithms. A heuristic optimization algorithm can be applied to a wide range of problems that do not always have a rigorous method of solution.

A general procedure of the distributed optimization algorithm is composed of distribution of the data, parallel processing (local processing), and integration of the calculated data (global processing). Initially, the contents of the candidates of solutions are set at random to cover the wide volume of solution space. After the candidates of solutions are distributed to the PE's, each PE locally updates the contents of the candidate and calculates its fitness function. To choose the good candidates requires that all the candidates on the multiple PE's be globally compared with each other on the basis of the values of the fitness function. This procedure is repeated until the solutions with the required fitness are obtained. In this procedure parallel processing of the calculation of the candidates of solutions will reduce the total processing time. However, as was mentioned in Section 1, global processing composed of data integration and selection of good candidates will lead to processing bottlenecks in conventional parallel computers without a global processor.

B. Architecture

Figure 1 shows the system compositions for the distributed algorithms by the conventional multiple-instruction–multiple-data stream parallel-computing system and the PMA. The conventional architecture has a hierarchy composed of a master PE and multiple slave PE's, as shown in Fig. 1(a). The roles of the master PE are data distribution, data integration, and global processing. Because the amount of network traffic for the data distribution and integration is N times as large as the processing capacity of the processor, these procedures can become processing bottle-

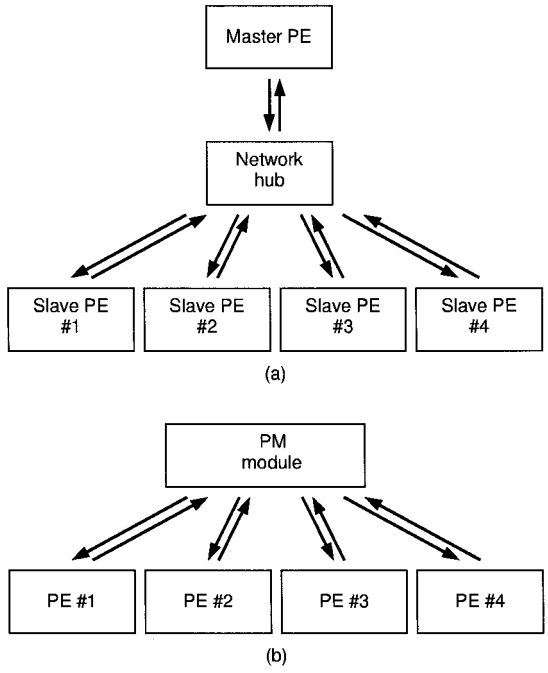

Fig. 1. Configurations of parallel-computing architectures: (a) a conventional master–slave architecture and (b) the PMA.

necks. On the other hand, the PMA has a different composition, as shown in Fig. 1(b). Because global processing is executed inside the PM module, the PE's in the system have the same priority; that is, the system has a flat hierarchy. As a result, there is no bottleneck in the proposed architecture with respect to global processing.

Another important feature of the proposed architecture is its flexibility in selecting its network topology. Because the PMA does not specify its network topology, arbitrary network topologies can be adopted. This feature enables system designers to choose the best network topology to fit their requirements for the parallel-computing system. We believe that the complete connection is promising because the diameter of the network is unity, which means that every PE can communicate directly with arbitrary PE's without relaying the communication packets. Hence we assume complete connection as the network topology of the PMA.

C. Parallel-Matching Operations

We define the datum from each PE as the reference datum and that from the other PE's as the objective datum, as shown in Fig. 2. A set of the reference datum and the objective datum to be compared is called a matching pair. The PM module tests the reference datum and each of the objective data for the following conditions: (1) the reference datum is equivalent to the objective datum, (2) the reference datum is smaller than the objective datum, and (3) the reference datum is larger than the objective datum. The result of the global comparison is expressed by a set of logical values. When the condition is satisfied, the returned value is 1 (true);

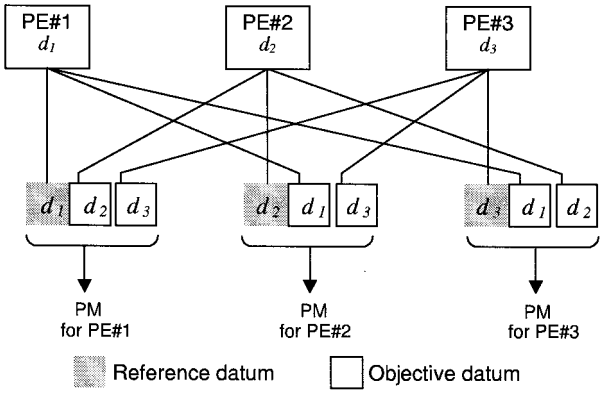

Fig. 2. Reference datum and objective data in PM: d_1, d_2, and d_3 denote the output data of PE 1, PE 2, and PE 3, respectively.

otherwise, it is 0 (false). Operations 1, 2, and 3 are called PM operations and are denoted by pEQU, pMORETHAN, and pLESSTHAN, respectively. The prefix "p" means parallel to distinguish an operation from that for a single data pair. We also define a fourth PM operation: summation of the absolute differences, denoted by pDIFF. This operation provides the summation of the absolute difference between the reference datum and the objective data. Using the pDIFF operation allows each PE to obtain a quantitative measure of the difference. As the primitive operations that constitute the PM operation are performed, a set of matching operations is defined. They execute one-to-one data comparison and return 1-bit logical values, or the absolute difference. They are denoted by EQU, MORETHAN, LESSTHAN, and DIFF.

Figure 3 shows a schematic diagram of PM with five PE's. The numbers in the boxes that represent PE's are the output data from the PE's. In the figure PE-A, PE-B, and PE-C obtain 4-bit binary values that represent the results of PM: pEQU, pMORETHAN, and pLESSTHAN, respectively. PE-D obtains the result of the pDIFF operation. The operating mode of PE-E is different from that of the others. It is the communication mode in which the data from PE-C are sent to PE-E transparently.

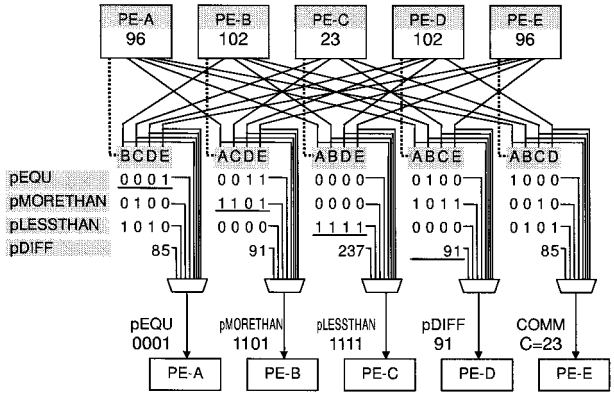

Fig. 3. Fundamental operations of the PM architecture. Mux, multiplexer.

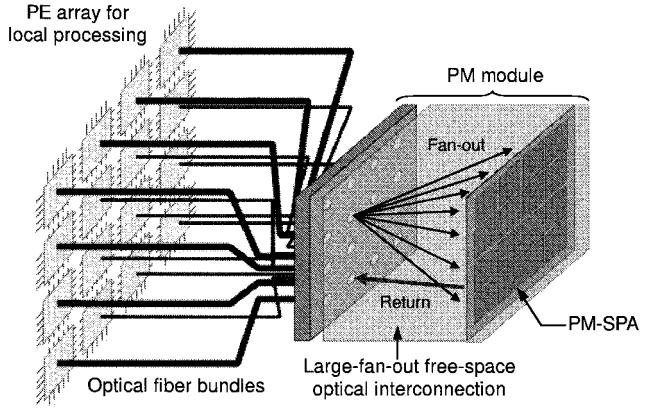

Fig. 4. Target prototype system of the PMA.

The output data from the PE's are fanned out and shuffled. They are concurrently compared by the PM operations in the PM module. Then one of the PM results or the objective datum is selected by the multiplexer on request from the PE's. In general, for an m-bit data format as many as $(m + 1)$ PE's can be compared at the same time. Finally, the selected result is sent back to each PE.

D. Optoelectronic System Composition

Figure 4 shows the system composition of the PMA. For example, the parallel-computing system consists of 4×4 PE's that are connected to the other PE's through the PM module. The PE's are located on a two-dimensional grid, and each PE is connected to the PM module by an optical fiber bundle. It is effective to embody the PM module with a large-fan-out optical interconnection and a SPA called a PM-SPA because the PM module requires a large number of high-speed interconnections. Each PE is also embodied by an OE-VLSI coupled with the optical fiber bundle to put multiple bits in and out simultaneously. As is mentioned in Section 3 below, the photonic network is implemented by optical data fan-out and shuffle. With the optically fanned-out signals the PM operations and the processing for inter-PE communication are executed by the PM-SPA. The resulting data are transmitted from the PM-SPA and returned to the PE's through the optical fiber bundles. In the PM module input and output fiber-bundle arrays are located on the same plane.[13]

Wavelength-division multiplexing is useful for communication between the PE's and the PM module and can enable the replacement of the optical fiber bundle with a single fiber without degrading the communication bandwidth. The introduction of wavelength-division multiplexing will bring the following advantages: The interconnection costs between the PE's and the PM module can be reduced, and easy handling can be achieved because of wiring flexibility. Another advantage is simplification of the optical system of the PM module. Because rotation of the optical fibers does not affect the alignment of the optical system, it is possible to omit most of the alignment mechanisms. Consequently, the volume of the PM module can be reduced significantly.

3. Optoelectronic Parallel-Matching Scheme

The fundamental idea of the proposed optoelectronic implementation of the PM module is that global processing be decomposed into multiple operations on a two-dimensional plane and executed by the same number of smart pixels simultaneously. PM is composed of multiple-element matching operations that can be executed potentially in parallel. A promising technology for embodying the PM module is the smart-pixel-based FSOI because of its huge and dense connectivity. Using the FSOI, we can generate all the matching pairs on an image plane immediately. The PM-SPA then concurrently compares them. Thus a huge number of the operations can be processed at the same time.

The design of the PM-SPA is restricted by various limitations of the VLSI fabrication. Among them, we focus on the device layout of the VLSI. Complicated VLSI wiring consumes the chip area and causes signal latency, so the data processing should be localized to avoid unnecessary wiring complexity. For the OE-VLSI the data shuffle by optical interconnection can be used to satisfy the requirement. After the shuffle, the photodetectors on the OE-VLSI detect the data, and the operations can be executed without complex wiring. On the basis of this idea several optoelectronic systems such as the sorting system[14] and the optical multimesh hypercube[9] have been demonstrated.

In the embodiment of the PM module the matching pair should be located adjacent to each other on the PM-SPA plane. We propose a specific optical interconnection scheme for PM by using two different interconnection patterns. In the proposed scheme the output data from the PE's are fanned out to generate images of the reference and the objective data. For locating the adjacent matching pair for processing on the PM-SPA the fanned-out data are shuffled and overlapped. These procedures can be achieved effectively by optical fan-out elements such as phase-only computer-generated holographic (CGH) filters.[15] Figure 5 shows an example of a schematic of optoelectronic PM in which example values are also shown. As shown in Fig. 5(a), $\sqrt{N} \times \sqrt{N}$ PE's are aligned on a two-dimensional grid. $\mathrm{PE}_{i,j}$ indicates the PE at the position (i, j). We assume a complete connection in which each PE monitors all the other PE's. The number of the monitorable PE's is $N - 1$. Figure 5 shows the case for $N = 4$, and the number of monitorable PE's is 3.

The output datum of $\mathrm{PE}_{i,j}$ is expressed by $d_{i,j}$. The output data from all the PE's are transferred to the PM module and arranged on a grid at the entrance of the PM module, as shown in Fig. 5(b). We define the procedures for generating reference and objective data as reference and objective duplications, respectively. In the complete connection $N - 1$ matching pairs are required for the PM of one PE. Therefore the image generated by reference

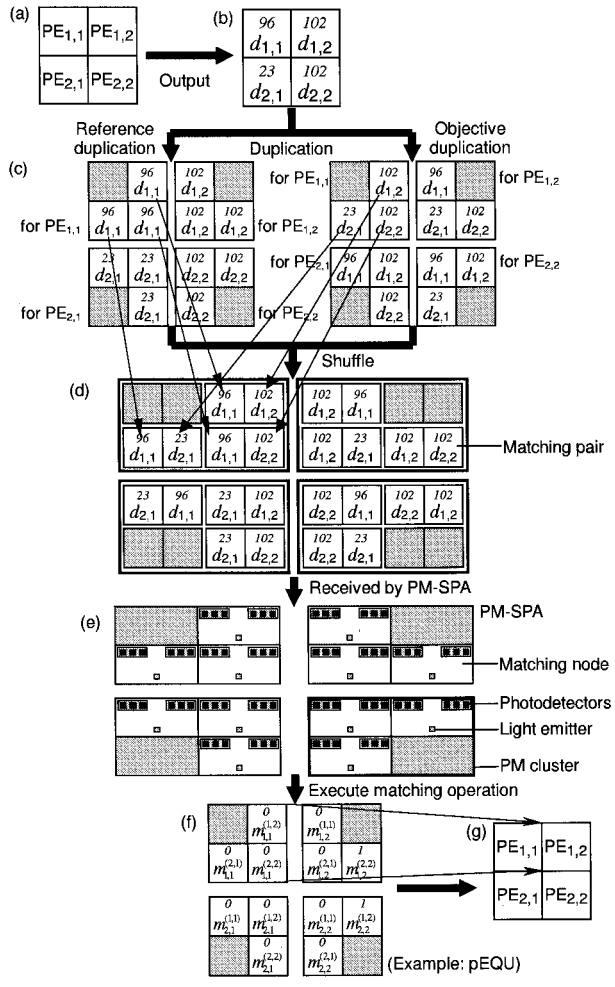

Fig. 5. Schematic diagram of the optoelectronic PM procedure: (a) PE's arranged on a two-dimensional grid, (b) the output data displayed on a light-emitter array, (c) the reference-duplication and the objective-duplication optical patterns, (d) the shuffled optical pattern, (e) the PM-SPA, (f) the optical output pattern of the PM results, and (g) the PE's that receive the optical signals.

duplication is composed of N sets of $N - 1$ copies of the datum of each PE in total. The objective-duplication image consists of N copies of Fig. 5(b). Note that each copy of the objective duplication does not contain the datum that is identical to the reference datum (depicted by the shaded boxes in Fig. 5) because comparison of the datum with itself is meaningless. After two images are shuffled to interleave them, the matching pair is located adjacently, as is shown in Fig. 5(d). For example, the matching pairs for $PE_{1,1}$, $\{d_{1,1}, d_{1,2}\}$, $\{d_{1,1}, d_{2,1}\}$, and $\{d_{1,1}, d_{2,2}\}$, are generated.

Multiple matching nodes are located on the surface of the PM-SPA, as is shown in Fig. 5(e), to detect the matching pairs for the matching operations. The set of matching nodes whose destination PE's are identical is called a PM cluster. Note that the reference data are identical in the same PM cluster and that the PM clusters work independently of each other. The PM-SPA chip is an array of PM clusters and contains at least one PM cluster. The matching nodes execute the matching operations concurrently and put out 1-bit matching results from the light emitters. We express the matching results for the reference $PE_{i,j}$ and the objective $PE_{k,l}$ as $m_{i,j}^{(k,l)}$; the spatial arrangement of the matching results is shown in Fig. 5(f). The PM result for one PE is a set of matching results in the same PM cluster. For example, the PM result for $PE_{1,1}$ is composed of three bits: $m_{1,1}^{(1,2)}$, $m_{1,1}^{(2,1)}$, and $m_{1,1}^{(2,2)}$. The matching results for the same PE are sent back to the PE through the optical communication channel, as shown in Fig. 5(g).

Figure 6(a) shows a block diagram of the matching node. From the matching pair of the reference datum $d_{i,j}$ and the objective datum $d_{k,l}$ four kinds of matching results are calculated by the matching units. Each matching node has multiple photodetectors and a light emitter. In the figure, $d_{i,j}^{(p)}$ denotes the pth bit of $d_{i,j}$ in binary representation. Because the bit lengths of the resultant data are different, two data paths are prepared. The results of the EQU, the MORETHAN, and the LESSTHAN operations are 1 bit long; from among these results $m_{i,j}^{(k,l)}$ is selected by the multiplexer. Then the matching result is put out from the local light emitter. For the PDIFF operation the absolute difference is sent to the accumulator through a local shared bus for the difference summation. As is shown in Fig. 6(b), each matching node is connected to an accumulator. The summation of the absolute differences is returned to the matching nodes through the local shared bus. Each bit of the summation is received by a matching node, and the value is sent back to the PE by the light emitters of the multiple matching nodes in parallel. In the communication the local shared bus is also used to send the selected objective datum to the matching nodes.

4. Performance Evaluation

To compare the performance of the PMA with the other parallel-computing systems that have various network topologies, we constructed an evaluation model and estimated the performance. The estimation is based on the effective communication bandwidth of the network and the processing bandwidth of the processor. Although the evaluation model is simple, it is useful for clarifying the characteristics of the proposed architecture. As examples of the fundamental global processing used in the distributed algorithms, we evaluated global data matching, detection of the maximum (minimum) value, and ranking of the data.

A. Evaluation Model

We evaluate the performance of parallel-computing systems on the basis of the processing bandwidth α [in bits per second (bps)] and the amount of the data for processing β (in bits) for several basic jobs. For simplicity the other overheads such as memory access are neglected in the evaluation. For these quantities the processing time τ (in seconds) is obtained from β/α. Communication time is also ex-

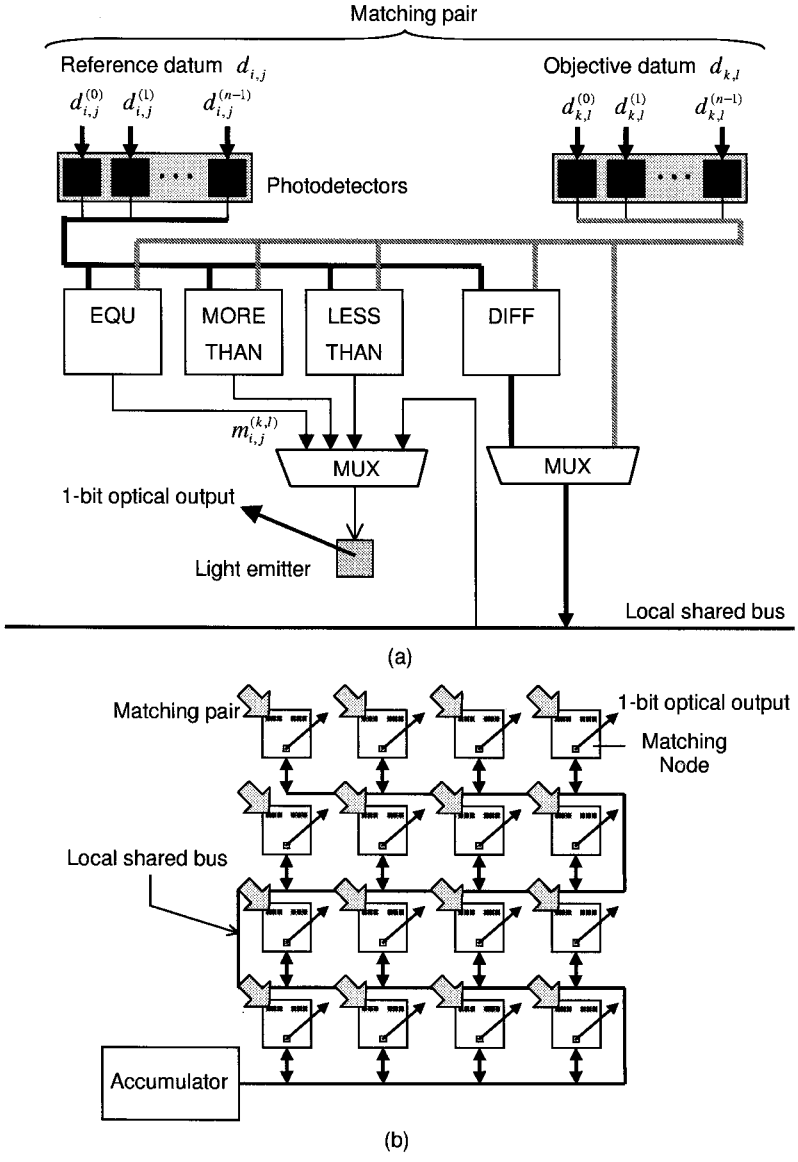

Fig. 6. Structures of (a) a matching node and (b) an array of matching nodes for a PE.

pressed by the same representation if α and β mean the bandwidth of communication and the amount of transferred data, respectively. Therefore we treat both processing and communication in the same manner. The basic jobs considered in the evaluation are classified into two groups: local processing at a PE and inter-PE communication. Inter-PE communication is further classified into single communications, multiple communications, single broadcasts, and multiple broadcasts, as discussed below.

We introduce a cost function for the performance of the system. The processing time τ is used to derive the cost function. The amount of processed data is the product of the bit length of the data w and the required steps s in the algorithm. The cost function is defined as

$$\text{Cost(job; } s) = \frac{ws}{\alpha_{\text{job}}}, \tag{1}$$

where job indicates the job type, w is common to all jobs, and α_{job} is relevant to the first parameter, job. The total cost T is given by the summation of the cost functions for all types of jobs:

$$T = \sum_{\text{job}} \text{Cost(job; } s). \tag{2}$$

We consider three networks: a mesh network, a completely connected network, and the PMA. Although the PMA is applicable to various types of interconnection, in this study, we assume that the network topology of the PMA is complete connection. The networks' bandwidths are expressed by $\alpha_{\text{job}}^{(\text{mesh})}$, $\alpha_{\text{job}}^{(\text{cc})}$, and $\alpha_{\text{job}}^{(\text{pma})}$, respectively. In this notation the superscript and the subscript mean the network topology and the job type, respectively.

All these networks are assumed to be embodied by the optoelectronic technologies. To utilize the resources of the system requires that the processing

Table 1. Elemental Operations of the PE and the Required Number of Steps per Operation

Operation	Number of Steps
Load data to register	1
Store data to register or memory	1
Increase register	1
Decrease register	1
Add to register	1
Count the number of value-1 elements contained in the register	1
Compare the contents of two registers	1
Jump	1
Conditional jump	2

capability of the processor and the communication capabilities of the network be comparable.

We also adopt the following assumptions to simplify the discussion below:

- A parallel-computing system is composed of N PE's.
- Each PE has only one process to execute a given task.
- The data format treated by the PE's is fixed a w bit long ($w \geq N$).
- Each PE has a sufficient number of registers.
- The communication port to and from the network is treated as a register.
- The required number of steps of the operations on the PE are assumed to be one except for the conditional jump of two.
- The overheads for accessing the memory and the network are neglected.

Table 1 lists the PE's operations and their required number of steps.

B. Communication Bandwidth

We evaluate the effective bandwidths of several kinds of fundamental inter-PE communications: single communications ($\alpha_{\text{s.comm}}$), multiple communications ($\alpha_{\text{m.comm}}$), single braodcasts ($\alpha_{\text{s.bcast}}$), and multiple broadcasts ($\alpha_{\text{m.bcast}}$). Figure 7 shows the configurations of these types of communication. In single communication a single PE in the system puts the data out to another PE [Fig. 7(a)]. Data transfer between a slave PE and the master PE is an example of single communication. In multiple communication multiple PE's put data out to another PE simultaneously [Fig. 7(b)]. Note that a PE cannot be assigned as both the source and the destination at the same time. In the single broadcast a single PE in the system broadcasts the data to all the other PE's [Fig. 7(c)]. For example, this kind of communication is used when multiple PE's obtain the same data from one PE. On the other hand, in multiple broadcasts multiple PE's broadcast data to the others simultaneously [Fig. 7(d)]. This type of communication is

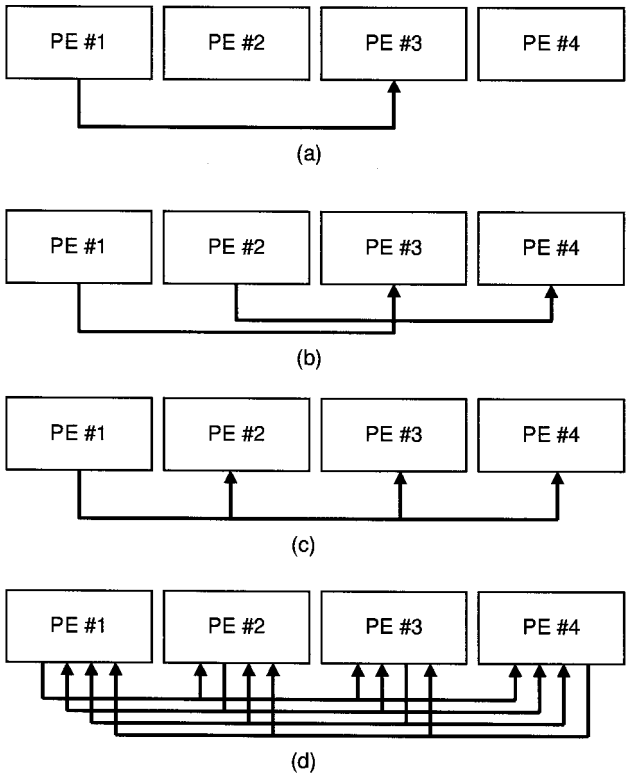

Fig. 7. Communication types: (a) single communication, (b) multiple communication, (c) single broadcast, and (d) multiple broadcast.

required when each PE obtains data from the other PE's.

The clock frequency of the optical data transfer between the PE and the network hub is denoted by r, and the maximum bandwidth of the data path is equal to wr. First, the effective bandwidths of the complete-connection network and the PMA are considered. Because a data path is used by only one PE, $\alpha_{\text{s.comm}}^{(\text{cc})}$, $\alpha_{\text{s.comm}}^{(\text{pma})}$, $\alpha_{\text{m.comm}}^{(\text{cc})}$, $\alpha_{\text{m.comm}}^{(\text{pma})}$, $\alpha_{\text{s.bcast}}^{(\text{cc})}$, and $\alpha_{\text{s.bcast}}^{(\text{pma})}$ are equal to wr. However, when a data path is shared by multiple PE's the effective bandwidth is equal to the maximum bandwidth divided by the number of the PE's in conflict. Therefore $\alpha_{\text{m.bcast}}^{(\text{cc})}$ and $\alpha_{\text{m.bcast}}^{(\text{pma})}$ are equal to $wr/(N-1)$.

The effective communication bandwidths of the mesh network are calculated on the basis of the average communication bandwidth $\alpha_{\text{m.comm}}^{(\text{mesh})}$, which is expressed by

$$\alpha_{\text{m.comm}}^{(\text{mesh})} = \frac{wr}{\sqrt{N}}, \tag{3}$$

as derived in Ref. 16, provided that the packet size is equal to the data-path width and collisions between the packets are neglected. Because the amount of data flow for multiple communication is equal to that of the broadcast from a single PE to the others, $\alpha_{\text{m.comm}}^{(\text{mesh})}$ is identical to $\alpha_{\text{s.bcast}}^{(\text{mesh})}$. Although the actual values of $\alpha_{\text{m.comm}}^{(\text{mesh})}$ are smaller than those obtained with Eq. (3) because of packet collision, we neglect the

Table 2. Throughputs of the Networks

Network Topology	$\alpha_{s.comm}$	$\alpha_{m.comm}$	$\alpha_{s.bcast}$	$\alpha_{m.bcast}$
Mesh	$wr/\sqrt{N}$	$wr/\sqrt{N}$	$wr/\sqrt{N}$	$wr/(N-1)\sqrt{N}$
Complete connection	wr	wr	wr	$wr/(N-1)$
PMA	wr	wr	wr	$wr/(N-1)$

collisions for simplicity. Thus the communication bandwidth for single communication is the same as in Eq. (3). In multiple broadcasts the number of total packets is $N-1$ times as large as that in multiple communications. Therefore $\alpha_{m.bcast}^{(mesh)}$ is expressed by $wr/[(N-1)\sqrt{N}]$. The resultant throughputs of the networks are summarized in Table 2.

C. Estimation of Fundamental Global Processing

We evaluate the performance of several kinds of fundamental global processing. Because the PMA has quite a different architecture from the others, it is considered separately in the following evaluation. We explain the outline of the operations. Fundamental global processing comprises the global data matching, the detection of the maximum (minimum) data, and the ranking of the data, and these operations are evaluated. These are the essential operations in the distributed optimization algorithms. Table 3 summarizes the total costs of the operations.

1. *Global Data Matching*

Non-parallel-matching architectures. When an optoelectronic network with a wide communication bandwidth is available a distributed method is practical for this task, even for the conventional parallel-computing architectures. In the distributed method each PE obtains the data of the other PE's by use of the multiple broadcast to compare the local data at each PE. After the multiple broadcast each PE has complete copies of the data from the other PE's. Each PE locally compares the obtained copies with the local data. Finally, each PE obtains the global-matching result.

Parallel-matching architectures. Because the PMA has the mechanism for global data matching, this task can be executed by the multiple-communication operation to the PM module without local processing at the PE. The process of global data matching is achieved in one frame cycle of the network.

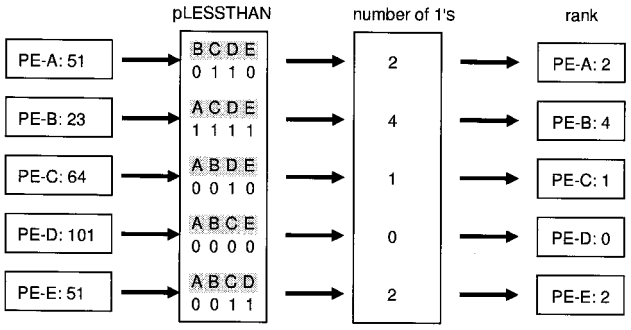

Fig. 8. Schematic diagram of the data-ranking process for the PMA.

2. *Detection of the Maximum (Minimum) Data*

Non-parallel-matching architectures. The maximum (minimum) data can be detected by a sequential search over the data of all the PE's. It is effective to use a master PE in this case. The procedure is as follows: The data from all the PE's are transferred to the master PE. Then the master PE sequentially searches the maximum (minimum) data locally through the received data. Finally, the obtained maximum (minimum) data are returned from the master PE to each PE by single broadcast.

Parallel-matching architecture. This task can be executed concurrently by the PM operations. Each PE can check whether its local data are the maximum (minimum) by the result of the pLESSTHAN (pMORETHAN) operation. For detecting the maximum data the pLESSTHAN operation is performed. If the data of a certain PE are the maximum the matching result of the pLESSTHAN operation does not contain any values of 1. The PE with the maximum data broadcasts its data to the other PE's by single broadcast. The procedure for minimum detection can be executed in the same manner by use of the pMORETHAN operation.

3. *Ranking*

Non-parallel-matching architectures. It is effective to use the master PE to rank the data. The data collected in the master PE from all the slave PE's are sorted to obtain their ranking. Then the rank is returned to each PE by single communication. Fast sorting algorithms that can be used[17] are the quick sort, the merge sort, and the heap sort. It is known that the average number of operations required in these sorting algorithms is approximately linear to $N \log_2 N$.

Table 3. Processing Costs for the Fundamental Matching Operations

Operation	Non-PMA	PMA
Global data matching	Cost(local; $5N-4$) + Cost(m.bcast; 1)	Cost(m.comm; 1)
Maximum (minimum) detection	Cost(local; $6N-4$) + Cost(s.comm; $N-1$) + Cost(s.bcast; 1)	Cost(local; 4) + Cost(m.comm; 1) + Cost(s.bcast; 1)
Ranking	Cost[local; $S(N)$] + Cost(s.comm; $2N-2$)	Cost(local; 1) + Cost(m.comm; 1)

Table 4. Order of the Operation Time Required for Fundamental Global Processing

Operation	Mesh	Complete Connection	PMA
Global data detection	$N^{3/2}$	N	1
Maximum detection	$N^{3/2}$	N	1
Ranking	$N^{3/2}$	$N \log N$	1

Parallel-matching architecture. The procedure for ranking is shown schematically in Fig. 8. The rank of the data can be obtained by the results of the pLESSTHAN or the pMORETHAN operation. For ascent-ordered ranking pMORETHAN is used, whereas for descent-order ranking pLESSTHAN is used. The result of the pLESSTHAN (pMORETHAN) operation of a certain PE indicates the number of PE's that have more (less) data than the PE. The number of value-1 elements included in the PM result shows the rank of the PE. Note that the number representing the rank starts from 0. As shown in Fig. 8, the ranks of PE's with the same values have an identical value.

D. Comparison

We find that the PMA shows excellent processing performance compared with the other architectures that do not have a global processor, as is shown in Tables 2 and 3. Table 4 shows the order estimation of the processing time. For the ranking the sorting time is approximately $\gamma N \log N$, where γ is a proportionality constant that is equal to 2 when the number of data comparisons in sorting is considered. Because the cost functions of the PMA are independent of the scaling factor N, the PMA can execute the tasks in constant time. The counterpart of the PMA is the completely connected network owing to the same network topology. Judging from these architectures, we can conclude that a great advantage is generated by the PMA in global processing.

The communication ratio is defined as the communication time normalized by the total processing time; this ratio can be used for evaluating the load of the network. Table 5 shows approximations of the communication ratios for large N in fundamental global processing. The communication ratio of the PMA is independent of N because the processing time is constant. Communication ratios for the mesh network increase in maximum data detection and ranking as N increases. The results show that communication dominates the total processing time owing to its narrow effective communication bandwidth. In this architecture most of the processing time is wasted on communication. However, the PMA for all tasks and the complete-connection architecture for global data matching and maximum data detection have communication ratios that converge to constant values of less than 1 as N increases. This trend shows that the communication capability is sufficient for the data traffic of the given tasks.

5. Prototype Embodiment

We constructed a prototype of the PM module to demonstrate its fundamental operations. In designing the prototype, we assumed that the multiprocessor system was of the type shown in Fig. 4. The parallel-computing system consists of 4×4 PE's, which are completely connected through the PM module. Because of the limitations of the optoelectronic devices used in the prototype, there are some changes in the specifications from the setup shown in Fig. 4. Each PE is assumed to communicate with the PM module in a bit-serial format, and a vertical-cavity surface-emitting laser (VCSEL) array instead of a fiber array is placed directly at the input plane of the PM module. Objective duplication is not executed optically but electronically in the PM-SPA. The optical system for sending the optical signals of the results back to the PE's is not implemented because the image field of the optical system is not large enough to accommodate both a VCSEL and a photodetector chip for bidirectional communication.

A. Parallel-Matching Scheme in the Prototype

Figure 9 shows a schematic diagram of the optoelectronic complete-connection architecture. Because of the device limitations, the system is simplified from the original shown schematically in Fig. 5. As shown in Fig. 9(a), the optical signals from the 4×4 PE's in the bit-serial format are assumed to be aligned on a two-dimensional grid as an input image to the PM module. Because the whole image of the light signals is required for one PE, 4×4 replica images, shown in Fig. 9(b), were prepared for the 4×4 PE's. Figure 9(c) shows a magnified image of the replica for $PE_{1,4}$. The replica image is detected by the SPA with 4×4 photodetectors and converted to electrical signals. In the PM cluster the reference datum is electronically fanned out to the PM nodes. Then the processing for parallel data matching and the inter-PE communication are executed on the nodes. Because the data are transferred in a bit-serial format in the prototype, the PM result is read out bit by bit from the matching nodes. Therefore there is only one light emitter in a PM cluster. As a result, the output image from whole PM clusters is composed of 4×4 pixels, as shown in Fig. 9(d).

Table 5. Approximated Values of the Communication Ratios of Large *N*

Operation	Mesh	Complete Connection	PMA
Global data detection	$1 - 5/\sqrt{N}$	1/6	1
Maximum detection	$1 - 6/\sqrt{N}$	1/7	1/3
Ranking	$1/(1 + \gamma \log N/2\sqrt{N})$	$1/(1 + \gamma \, log \, N/2)$	1/2

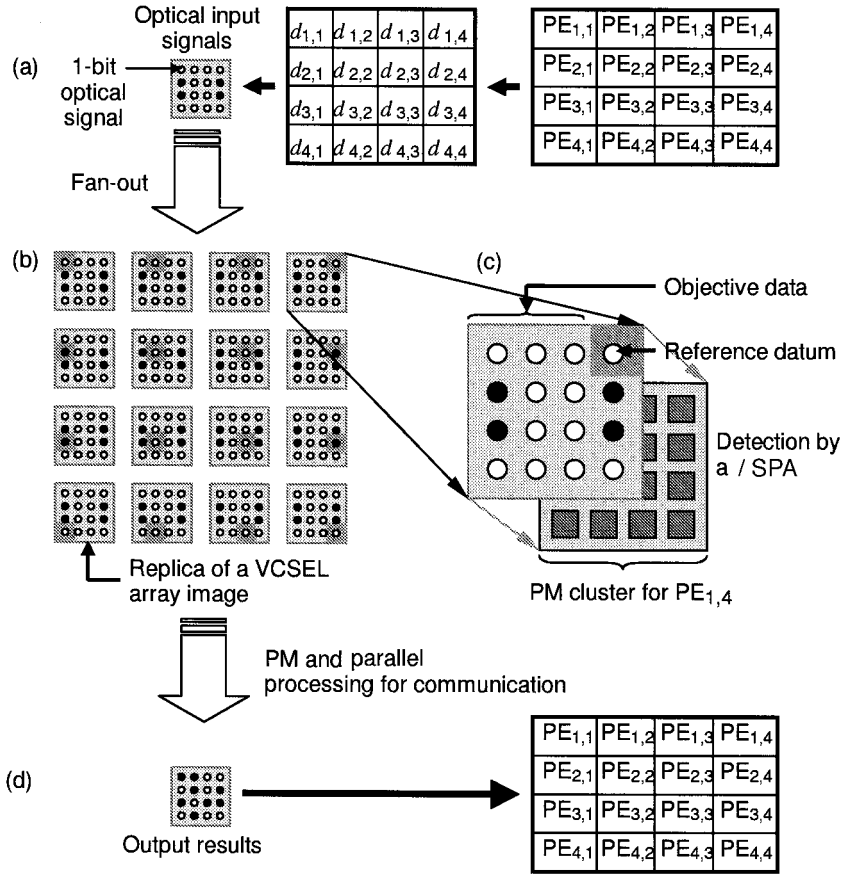

Fig. 9. Schematic diagram of an optoelectronic complete-connection architecture: (a) the output data displayed on the VCSEL array, (b) the replica images of the VCSEL array for the complete-connection network composed of 4 × 4 PE's, (c) a replica of the VCSEL image for $PE_{1,4}$, and (d) the output data.

B. Configuration of Prototype System

Table 6 summarizes the devices used in the prototype system. An 8 × 8 VCSEL array (Micro Optical Devices, Model Gigalase with an emitting wavelength of 850 nm) is used to send the data toward the PM-SPA. The VCSEL array emits the data from the PE's. In the prototype the function of the PM-SPA is emulated by a complex programmable logic device (CPLD) (Cypress, Model FLASH374i) that is coupled with a 4 × 4 complementary metal-oxide semiconductor photodetector (CMOS-PD) array (Model N73CGD) supplied by the U.S.–Japan Optoelectronic Project (U.S.–JOP). Owing to the element number of the photodetector array the effective size of the VCSEL array is limited to 4 × 4 pixels.

For large-fan-out optical interconnection a conventional 4*f* optical correlator is adopted. We constructed a Fourier transform lens system[13,18] whose focal length was 160.0 mm for a wavelength of 850 nm by combining commercially available lenses. For the optical fan-out element that generates the complete-connection pattern shown in Fig. 9(b), we fabricated a phase-only CGH filter with two-level

Table 6. Devices Used in the Prototype

Product	Supplier	Array Size	Pixel Pitch	Pixel Size	Maximum Operating Speed	Other
VCSEL array (Model Gigalase)	Micro Optical Devices	8 × 8	250 μm	ϕ15 μm		Maximum of 3 mW λ = 850 nm
VCSEL driver (Model CLDA2)	JOP	32 input–output channels			150 MHz	
CMOS-PD array	JOP	4 × 4	250 μm	120 μm	15 MHz (at 100 μW and λ = 850 nm)	
CPLD (Model FLASH374i)	Cypress				125 MHz	128 macrocells

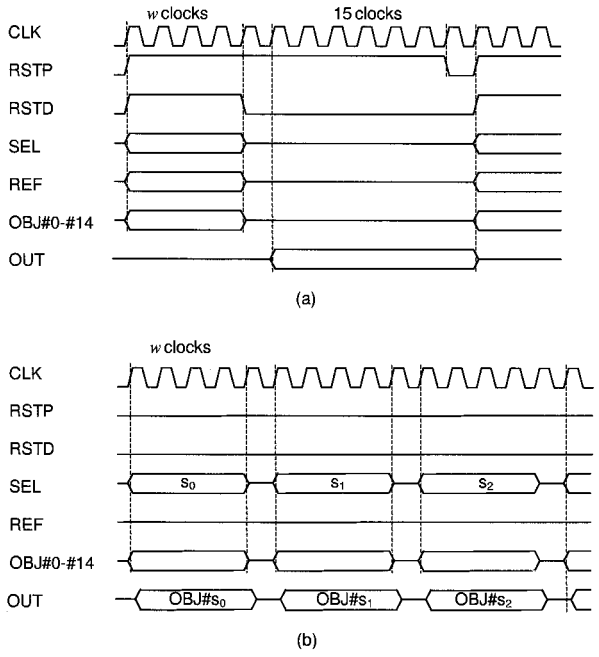

Fig. 10. Timing charts of (a) the PM procedure and (b) the inter-PE communication. s_i denotes the discrimination number of the data-source PE in the communication mode. CLK, clock signal; REF, reference datum; OBJ 0–10, objective data; OUT, output read out from the CPLD; RSTP, timing signal for resetting the PM result; RSTD, timing signal for reading out the PM result; w, word length; SEL, selection signal for determining the type of PM operation in the PM mode; OUT, signal for the PM results; MN, matching node.

phase modulation by using electron-beam (EB) lithography.

C. Emulation of a Parallel-Matching Cluster

The functions of the PM cluster, such as the pEQU, pLESSTHAN, and pMORETHAN operations and the inter-PE communication, are implemented by the CPLD. The CPLD puts the electrical signals detected by the CMOS-PD array in with the control signals and puts the resultant data out. Although the control signals, such as the clock signal and the timing signals for reading out the PM results, should be generated from the reference and the objective data inside the CPLD, in the prototype module they are fed into the system from outside the CPLD to simplify the configuration of the circuits. The PM cluster is composed of 15 matching nodes (MN 0 to MN 14) and one reference node and operates in two functional modes: a PM mode and a communication mode, each of which is selected by the control signal.

Figure 10(a) shows the timing chart of the PM operations. The circuits of the CPLD are synchronized with the clock signal. The output datum of the PM result and the communication datum are read out from the CPLD. Two timing signals determine the timing of resetting and reading out the PM result. The word length w of the target data is variable and is determined by the timing signals. In reading out the PM results, the matching results are output sequentially from matching node 0 to matching node 14. The timing chart in the communication mode is shown in Fig. 10(b). In this case the datum from the source PE that is specified by the selection signals is output directly after a short delay that is as long as the latency of the digital circuits.

D. Design of the Optical System

The magnification and the focal length of the Fourier transform lens system are 1.0 and 160.0 mm, respectively. The first Fourier transform lens is an achromatic-doublet lens (Melles Griot, Model 01LAO149/076) with a diameter of 30.0 mm and a focal length of 160.0 mm. Because the maximum height of the input image is as small as 1.4 mm (which corresponds to a 1° field of view), third-order aberrations are negligible.

On the other hand, the objective height of the output interconnection pattern is as large as 17 mm, which corresponds to a 6° field of view on the second Fourier transform lens. The aberrations of the above achromatic-doublet lens are not corrected for such a large field of view. We designed the Fourier transform lens system by combining commercially available lenses with a heuristic method. In the design, we used CodeV (Optical Research Associates). We sought a good combination of commercial lenses from the database supplied by CodeV. For the final design, we used the combination of an achromatic-doublet lens (Melles Griot, Model 06LAI015/076) with a diameter of 50.0 mm and a focal length of 190.0 mm and a positive-meniscus lens (Melles Griot, Model 01LMP015) with a diameter of 50.0 mm and a focal length of 175.0 mm. The designed Fourier transform lens system provides sufficient optical qualities when used with the CGH filter described below.

E. Design and Fabrication of the Computer-Generated Holographic Filter

We designed the CGH filter for a complete-connection network composed of 4×4 PE's with a bit-serial optical channel. Figure 11(a) shows the mapping of the output plane of the interconnection optics. The output pattern contains 16 copies of the VCSEL image arranged on a grid. Because the equipment used to fabricate the CGH filter does not have an accuracy that is high enough to eliminate zero-order diffraction, the copied images are located so that they do not overlap with the zero-order diffraction image in the design. As shown in Fig. 11(a), each quadrant contains 2×2 copies of the VCSEL image. Each copy corresponds to input-image signals: the reference datum and the object data of the PM cluster. Although only 4×4 pixels of the VCSEL array are effective in the current prototype, the filter is designed for the entire 8×8 pixels of the VCSEL array for future extension of the optical communication bandwidth. The period of each copy of the VCSEL image is 2.5 mm. The margin between adjacent cop-

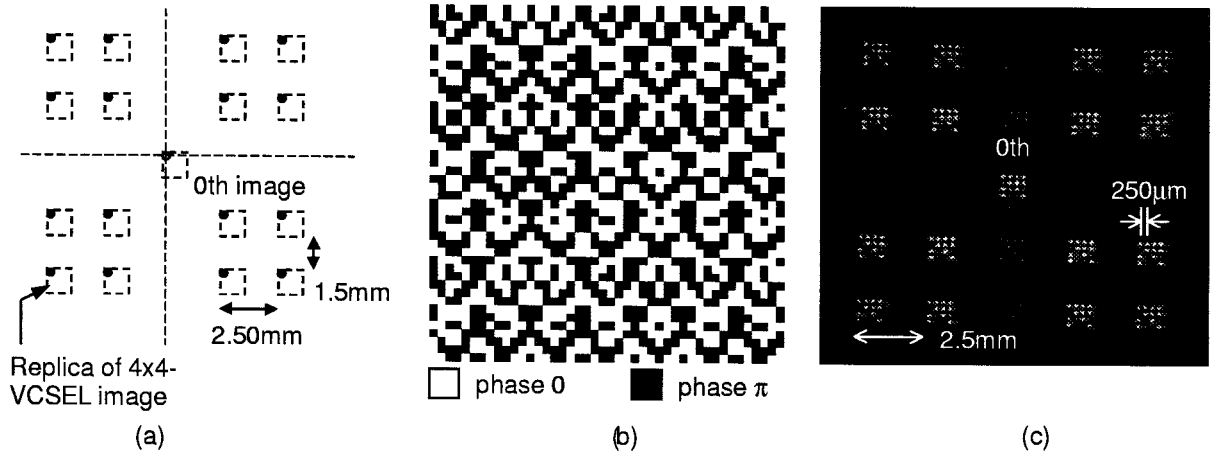

Fig. 11. (a) Designed optical interconnection pattern for a complete-connection network composed of 4×4 PE's. (b) Part of the obtained CGH filter with two-level phase modulation. The pixel size is 8.5 μm, and the filter size is 17.408 μm. (c) Experimental results of the optical interconnection by the CGH filter.

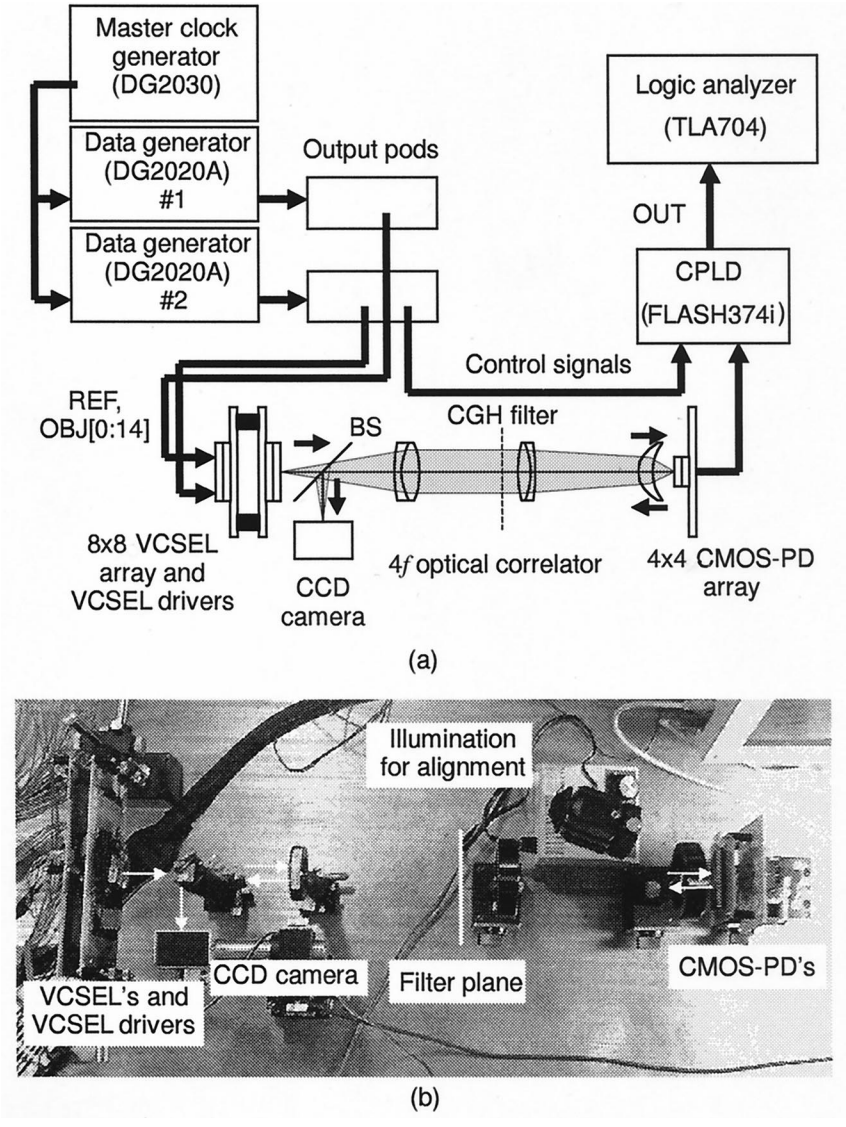

Fig. 12. Schematic diagram of the experimental prototype system: (a) configuration and (b) top view of the optical setup. BS, beam splitter.

ies is 0.5 mm for 8 × 8 VCSEL's or 1.5 mm for 4 × 4 VCSEL's.

We used a design method based on the Gerchberg–Saxton algorithm to calculate the filter pattern.[19] To quantify the phase, we used a stepwise quantization method.[20] In our filter design the pixel size, the number of pixels, and the pattern size were set to 8.5 μm × 8.5 μm, 2048 × 2048 pixels, and 17 mm × 17 mm, respectively. This filter can cover the entrance pupil of the second Fourier transform lens system. The size of the reconstructed plane is 16 mm × 16 mm, which is large enough for the interconnection pattern. Figure 11(b) shows part of the obtained filter pattern.

We fabricated the CGH filter with EB lithography. The positive-type photoresist (Nippon Zeon Co., Ltd., Model ZEP-520-22) was spin coated onto a SiO_2 substrate with a thickness of 5000 Å. The EB writer (JOEL, Model JBX-5000I) was used to expose the filter pattern on the photoresist. After development the relief profile of the photoresist was copied to the glass substrate by plasma dry etching (ULVAC Japan, Ltd., Model NLD-800). The etching gas was a compound of CH_2F_2 and C_4F_8. To obtain a phase modulation of π, we determined the required target etch depth to be 850 nm. The actual etch depth was measured to be 864 nm with a white-light interferometer (Zygo Co., Model New-View5020).

The CGH filter was incorporated into the 4*f* optical correlator. Figure 11(c) shows the reconstructed interconnection pattern of the fabricated CGH filter for 4 × 4 VCSEL's. The results show a good optical quality that is sufficient for detection by the photodetectors. The diffraction efficiency was measured to be 68%. The percentages of the zero-order diffraction and the higher-order light were 6% and 26%, respectively. The zero-order diffraction comes from the etch-depth error.

6. Experiments

To verify the fundamental operations of the prototype system, we executed several experiments. Figure 12 shows the experimental setup. The signal source was two data generators: one (Tektronix, Model DG2020A) with an operating speed of 200 MHz and 36 output channels for which another data generator (Tektronix, Model DG2030) with an operating speed of 400 MHz and four channels supplied the master clock and the start trigger. To monitor the function of the PM-SPA prototype, we used a logic analyzer (Tektronix, Model TLA704) with an operating speed of 500 MHz and 68 input channels.

A. Operation Speeds of the Vertical-Cavity Surface-Emitting Lasers and the Complementary Metal-Oxide Semiconductor Photodetectors

To determine the operating speed of the prototype, we measured the maximum operating speeds of the VCSEL array and the CMOS-PD array. The VCSEL driver (JOP, Model CLDA2) puts the driving current out to the VCSEL's. The swing of the output current was controlled by two voltages: the modulation voltage and the bias voltage. Through the imaging system, the waveform of the optical signal for the square-wave input was measured by an avalanche-photodiode module (Hamamatsu Photonics, Model S5331-01) with a bandwidth of 100 MHz.

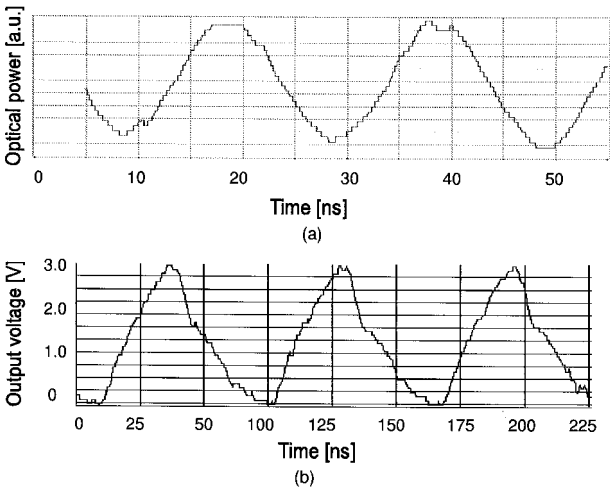

Fig. 13. Observed waveforms of (a) the optical signal emitted by the VCSEL for a rectangular input signal at 50 MHz and (b) the output of the CMOS-PD for the optical signal from the VCSEL at 15 MHz.

The maximum operating speed was 50 MHz for return-to-zero signals. Figure 13(a) shows the optical signal emitted by the VCSEL at 50 MHz. The modulation and the bias voltages were 3.146 and 2.495 V, respectively. The optical power of the VCSEL that was measured at the output plane was approximately 700 μW/pixel. The optical loss of the imaging system that was caused by the finite apertures of the lenses and the reflection at the lens surfaces was 42%.

A beam splitter was inserted into the optical system to monitor the light spots of the VCSEL image on the detector areas of the CMOS-PD array by observation of the CCD image. Because of the optical loss at the beam splitter, the optical power per pixel decreased by one half to 350 μW. For the incident optical power the maximum speed of the CMOS-PD was 15 MHz. Figure 13(b) shows the waveform of the output signal from the CMOS-PD. The photocurrent is amplified and binarized by a series of two comparators in the CMOS-PD. The threshold-signal level can be controlled by two external voltages, which are threshold voltages to the first and the second comparators. These voltages were 1.825 and 1.833 V, respectively. The voltage of the power supply was 3 V. At operational speed's higher than 15 MHz, we could not obtain a common threshold voltage for all the optical channels. We suppose that the problem is caused by nonuniformity of the light intensities of the VCSEL's.

B. Parallel-Matching Operations

We operated the prototype system at 15 MHz without the CGH filter to verify the operation of a single PM

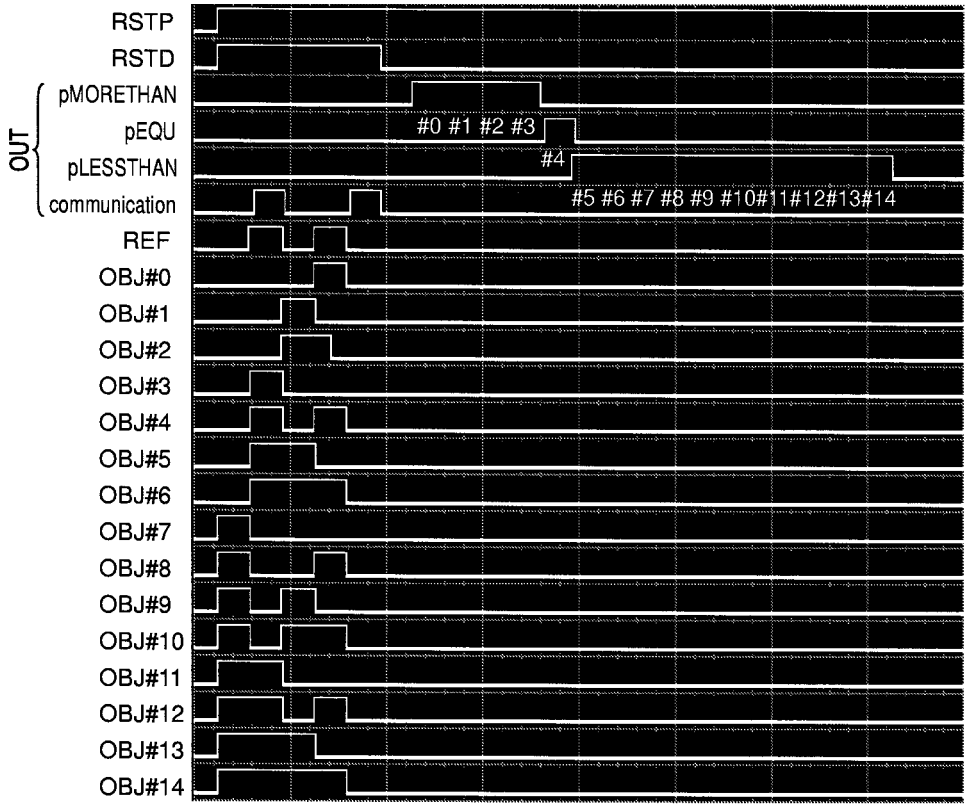

Fig. 14. Experimental results of the PM operations and the inter-PE communication. Data separately measured by the logic analyzer are merged into a single figure.

cluster. Figure 14 summarizes the experimental results of the PM and the inter-PE communications. Because the data transfer is in the bit-serial format, the reference and the objective values and the PM result are expanded on the transverse axis in the figure. The voltages applied to the VCSEL driver and the CMOS-PD array were the same as in the experiments described in Subsection 6.A. The word length w was 4 bits.

To analyze the results of the PM operations easily, we set OBJ 0 to OBJ 14 to 1 (0001 in binary representation) and to 15 (1111), respectively. The reference datum was 5 (0101) in the PM mode. For the pEQU operation, the OUT signal indicated 1 for OBJ 4 and 0 for the others because only OBJ 4 was equal to 5. In this case the PM result was expressed by 000010000000000 (indicated in the row labeled pEQU in Fig. 14) in which the most significant bit corresponds to OBJ 0. The OUT signal for the pMORETHAN operation was 111100000000000 because OBJ 0 to OBJ 3 were less than 5. On the other hand, the OUT signal for the pLESSTHAN operation showed 000001111111111. In the communication mode the reference PE was assumed to communicate with PE 8 so that OBJ 8 was selected by the SEL signal. The OUT signal had a value of 9 (1001), which was equal to OBJ 8, as shown in the row labeled COM in the figure.

We verified that the fundamental operations of the prototype were executed exactly at 15 MHz. The bit rate of communication per PE and the total bit rate of the prototype were 15 Mbps and 240 Mbps, respectively. In this case the frequency of the data switching was 3 MHz. The frequencies of the PM operation for each PE and for the whole system were 0.68 megaoperations/s and 11 megaoperations/s, respectively.

7. Discussion

Scalability is an important factor in parallel-computing systems. In this paper a complete connection was assumed for the network topology of the PMA. However, when we consider the scalability of the parallel-computing system, the complete connection is not always the best network topology. Adding a new PE to an N-PE architecture requires the redesign of the PM module. As was shown in Ref. 9, to solve the unscalability problem, it is effective to embed the unscalable network in a scalable network, e.g., bus, mesh, and ring networks. This technique is also effective for the PMA because the proposed architecture can be combined with arbitrary network topologies. As a result, the PMA can be scalable.

One method of extending the PM module is to increase the number of functions. The extension of the functions is restricted mainly by the number of transistors available for the digital circuits of the PM-SPA. In designing, we must note the following things: The length of the data format of a processor does not increase as fast as does the clock frequency,

and the degree of the PMA's network topology should be as large as the data-format length to monitor the other PE's effectively. The area for digital circuits is restricted by that of the photodetecting and the light-emitting circuits. Because the length of the data format is almost constant for several years, the number of photodetectors and light emitters in a matching node and that of the matching nodes in a PM cluster do not change during that period. When the maximum size of the photodetector is dominated by the diffraction limit of the optical system the area consumed by the photodetectors is almost constant for the same data-format length. Consequently, the number of transistors for the extended functions will increase in proportion to the density of the transistors.

Discussion of the performance evaluation for the general cases remains a future issue. In the performance evaluation, we did not evaluate the case in which the numbers of data are larger than the numbers of PE's. However, if the processing data were divided into several clusters that could be processed by the PMA in a single frame cycle the treatment presented in this paper could be applied. For rigorous evaluation the estimation model must be improved. We should take into account the latencies of the memory access and the network to obtain an accurate speed-up ratio.

The operational speed of the prototype was limited by the CMOS-PD array. The performance can be improved by use of high-speed photodetectors with high sensitivity such as metal–semiconductor–metal photodetectors[7] coupled with transimpedance photoamplifiers. With high-speed photodetectors of 8×8 pixels the bit rate of inter-PE communication and the frequency of the PM operation of the whole prototype are expected to improve to as high as 9.6 Gbps and 2.4 gigaoperations/s, respectively, when operated at 150 MHz.

8. Conclusion

We have proposed an optoelectronic PMA as an effective parallel-computing architecture. This architecture is composed of multiple PE's and a global processor called a PM module, which has a specific mechanism for concurrent data comparison over multiple PE's. The fundamental operations of the PMA are four kinds of PM: pEQU, pMORETHAN, pLESSTHAN, and pDIFF. The PM operations accelerate the execution of distributed algorithms. It has been shown that the smart-pixel-based large-fan-out free-space optical interconnection was effective for the embodiment of the PM module. The performance evaluation has shown that the proposed architecture can reduce the execution time for fundamental global data processing, i.e., global data matching, detection of the maximum (minimum) data, and ranking of the data, compared with other conventional architectures that have photonic networks.

A prototype system of the PM module has been constructed to demonstrate the fundamental global operations on the basis of state-of-the-art optoelectronic devices and a phase-only CGH filter. In the prototype the PM-SPA, which is the core device of the PM module, was emulated by the CPLD and the CMOS-PD array. The prototype module was assumed to be used with 4×4 PE's that are completely connected through bit-serial optical channels. For optical interconnection of the prototype a Fourier transform lens system was designed. For a fan-out element a phase-only CGH filter with two-level phase modulation was designed based on the Gerchberg–Saxton algorithm and was fabricated by EB lithography. We have shown that the prototype performed the fundamental PM operations and the inter-PE communication at 15 MHz. For the whole system the bit rate of the inter-PE communication and the frequency of the PM operation were 240 Mbps and 11 megaoperations/s, respectively.

This research was supported by JOP user funding under the Real World Computing Partnership. The authors acknowledge the activities of the JOP. This study was also supported by Development of Basic *Tera* Optical Information Technologies, Osaka Prefecture Joint-Research Project for Regional Intensive, Japan Science and Technology Corporation. The authors are also grateful to K. Takahara, K. Yamada, H. Toyota, and W. Yu, who were working on the project and helped us fabricate the CGH filter.

References

1. G. S. Almasi and A. Gottlieb, *Highly Parallel Computing* (Benjamin/Cummings, Redwood City, Calif., 1989).
2. K. Hwang and F. A. Briggs, *Computer Architecture and Parallel Processing* (McGraw-Hill, New York, 1985).
3. T. Kurokawa, S. Matso, T. Nakahara, K. Tateno, Y. Ohiso, A. Wakatsuki, and H. Tsuda, "Design approaches for VCSEL's and VCSEL-based smart pixels toward parallel optoelectronic processing systems," Appl. Opt. **37,** 194–204 (1996).
4. J. H. Collet, D. Litaize, J. V. Campenhout, C. Jesshope, M. Desmulliez, H. Thienpont, J. Goodman, and A. Louri, "Architectural approach to the role of optics in monoprocessor and multiprocessor machines," Appl. Opt. **39,** 671–682 (2000).
5. O. Sjölund, D. A. Louderback, E. R. Hegblom, S. Nakagawa, J. Ko, and L. A. Coldren, "Free-space optical interconnect using flip-chip bonded, microlensed arrays of monolithic vertical cavity lasers and resonant photodetectors," in *Digest of the Topical Meeting on Optics in Computing* (Optical Society of America, Washington, D.C., 1999), pp. 215–217.
6. D. A. Louderback, O. Sjölund, E. R. Hegblom, S. Nakagawa, J. Ko, and L. A. Coldren, "Vertical cavity lasers with large bandwidths at low currents for dense free-space optical interconnects," in *Digest of the Topical Meeting on Optics in Computing* (Optical Society of America, Washington, D.C., 1999), pp. 224–226.
7. D. V. Plant, B. Robertson, H. S. Hinton, M. H. Ayliffe, G. C. Boisset, W. Hsiao, D. Kabal, N. H. Kim, Y. S. Liu, M. R. Otazo, D. Pavlasek, A. Z. Shang, J. Simmons, K. Song, D. A. Thompson, and W. M. Robertson, "4×4 vertical-cavity surface-emitting laser (VCSEL) and metal–semiconductor–metal (MSM) optical backplane demonstrator system," Appl. Opt. **35,** 6365–6368 (1996).
8. A. C. Walker, M. P. Y. Desmulliez, M. G. Forbes, S. J. Fancey, G. S. Buller, M. R. Taghizadeh, J. A. B. Dines, C. R. Stanley, G. Pennelli, A. R. Boyd, P. Horan, D. Byrne, J. Hegarty, S. Eitel, H.-P. Gauggel, K.-H. Gulden, A. Gauthier, P. Benabes, J. L.

Gutzwiller, and M. Goetz, "Design and construction of an optoelectronic crossbar switch containing a terabit per second free-space optical interconnect," IEEE J. Sel. Top. Quantum Electron. **5,** 1–13 (1999).
9. A. Louri, S. Furlonge, and C. Neocleous, "Experimental demonstration of the optical multimesh hypercube: scalable interconnection network for multiprocessors and multicomputers," Appl. Opt. **35,** 6909–6919 (1996).
10. M. W. Haney, M. P. Christensen, P. Milojkovic, J. Ekman, P. Chandramani, R. Rozier, F. Kiamilev, Y. Liu, and M. Hibbs-Brenner, "Multichip free-space global optical interconnection demonstration with integrated arrays of vertical-cavity surface-emitting lasers and photodetectors," Appl. Opt. **38,** 6190–6200 (1999).
11. P. Berthomé and A. Ferreira, *Optical Interconnections and Parallel Processing: Trends at the Interface* (Kluwer, London, 1998).
12. L. Chambers, *Practical Handbook of Genetic Algorithms* (CRC Press, Boca Raton, Fla., 1995).
13. D. J. Reiley and J. M. Sasian, "Optical design of a free-space photonic switching system," Appl. Opt. **36,** 4497–4504 (1997).
14. D. T. Neilson, S. M. Prince, D. A. Baillie, and F. A. P. Tooley, "Optical design of a 1024-channel free-space sorting demonstrator," Appl. Opt. **36,** 9243–9252 (1997).
15. D. Prongué, H. P. Herzig, R. Dändliker, and M. T. Gale, "Optimized kinoform structures for highly efficient fan-out elements," Appl. Opt. **31,** 5706–5711 (1992).
16. T. M. Pinkson, M. Raksapatcharawong, and Y. Choi, "WARRP core: optoelectronic implementation of network-router deadlock-handling mechanisms," Appl. Opt. **37,** 276–283 (1998).
17. R. Sedgewick, *Algorithms*, 2nd ed. (Addison-Wesley, New York, 1988).
18. W. J. Smith, *Modern Optical Engineering: The Design of Optical Systems*, 2nd ed. (McGraw-Hill, New York, 1990).
19. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35,** 237–246 (1972).
20. F. Wyrowski, "Diffractive optical elements: iterative calculation of quantized, blazed phase structures," J. Opt. Soc. Am. A **7,** 961–969 (1990).