A Modulator-based Multistage Free-space Optical Interconnection System

A.G.Kirk¹, D.V.Plant¹, T.H.Szymanski², Z.G.Vranesic³, J.A.Trezza⁴, F.A.P.Tooley⁵, D.R.Rolston¹, M.H.Ayliffe¹, F.Lacroix¹, D.Kabal⁶, B.Robertson⁷, E.Bernier⁶, D.F.-Brosseau¹, F.S.J.Michael¹ and E.L.Chuah¹

 Department of Electrical and Computer Engineering, McGill University, Montreal, Canada; 2. Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada; 3. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada; 4. Sanders, Lockheed Martin, Nashua, New Hampshire, U.S.A.; 5. Department of Physics, Heriot-Watt University, Riccarton, Edinburgh, U.K.; 6. Now with Nortel Networks, Ottawa, Canada; 7. Now with Thomas Swan & Co. Ltd., Cambridge University, Cambridge U.K.

ABSTRACT

We describe the design and implementation of a free-space optical interconnect for multi-processor and backplane applications. The system is designed to interconnect 4 nodes in a unidirectional ring, with a total of 256 data channels propagating from node to node. Each node contains an array 512 GaAs electro-absorption modulators and 512 photodetectors, hybridly attached to a silicon integrated circuit. Light is relayed between nodes with a rigid micro-optical system. System results are presented.

Keywords: Free-space optical interconnect, multiprocessor, backplane, micro-optics, quantum-confined Stark effect modulators, optoelectronic-VLSI, diffractive optics.

1. INTRODUCTION

Parallel optical interconnects are capable of providing high bandwidth communication links both within and between high performance electronic systems. In the commercial arena, several companies now supply optical fiber ribbon-based parallel optical data links (PODLs) of 8 to 12 channels, operating at data rates of up to 2.5 Gb/s per channel over distances of 100-1000 m (depending on bit rate). However, there are applications where many more parallel channels are required in an architecture that is more complex than a simple point-to-point link. One such application is the interconnection of processors and memory within multiprocessor computers. The class of processor architectures described as non uniform memory access or NUMA have many important advantages [1]. In such an architecture processors have access to both local and non-local memory. All of the processes are able to access all of the memory. In order to ensure that a processor is not operating on memory that has since been modified by another processor, cache coherency is required. In a NUMA architecture a low latency high bandwidth interconnection network is required in order to broadcast information about the changing memory state to all processors.



Figure 1. Schematic representation of a multihop optical ring interconnection. Each node is a processor or a memory.

Figure 1 above shows one possible interconnection network for a NUMA multiprocessor computer system. This is a multihop ring interconnect in which each node receives high bandwidth data from the previous node and transmits it onwards to the next node. A fraction of the incoming data can be extracted and processed at the node. Similarly data sourced by the node can be transmitted to the next node. Data thus passes around the network in a series of hops. In the context of NUMA multiprocessor, data which passes through at each node must do so with very low latency. This interconnection system can be implemented optically within the hyperplane architecture [2,3]. Each node requires a high-speed, highly parallel point-to-point interconnection to both the next node and the previous node. In contrast, data which is sourced or sunk at each node moves on and off the node at a rate limited by chip-to-board electrical communication rates. This system is thus an example of a 'firehose' architecture [4].

In this paper we describe the design and implementation of a multistage free-space optical interconnection system for use in a NUMA multiprocessor. The development of this system has allowed us to answer a number of important questions about the implementation of highly parallel free-space optical interconnection systems. It also imposed several major challenges. It was decided to implement the system that would interconnect four nodes of a multiprocessor system. However the system itself was designed to scale to at least eight nodes. Each node is designed to inject or extract 32 bits onto or from the ring. In order to support up to 8 nodes the ring was therefore required to transmit 256 optical channels. Dual rail optical is used [5] so that each node has a total of 512 optical inputs and optical outputs. In order for these to be transmitted from a single optoelectronic chip, the cross-section of the optical interconnect must be less than 1 cm². This represents a very high channel density. Another requirement for the system was one of field serviceability. This demand required that the system be designed in such a way that the active components could be removed from the system and reinserted without the need for realignment. The optical interconnect was also designed in order to give the maximum possible tolerance to misalignment.



Figure 2. Schematic representation of the interconnection network, showing the optical and electrical data paths in the system.

2. SYSTEM DESIGN

A schematic diagram of the system is shown in Fig. 2 above. Each node on the network consists of a printed circuit board (PCB) which may contain 1 to 4 processors. Each PCB is interconnected via a flexible printed circuit board to an optoelectronic VLSI circuit which injects and extracts signals into and from an optical ring. The electrical ribbon cable supports 32 input/output lines. To inject data into the ring, the optoelectronic chip converts electrical signals into optical signals. These are then transmitted to the optoelectronic chip at the next node. At this point they can be extracted from the ring or transmitted on to the next node. Signals destined for nodes beyond the nearest neighbour are routed across the optoelectronic chip from a receiver to an appropriate modulator pair with very low latency [2]. They are then transmitted onwards until they reach their final destination at which point they are extracted from the ring. The optical ring transmits 256 data channels, allowing it to support a ring of up to 8 nodes in a non-blocking manner. In the following sections we describe in detail the design and implementation of the electrical, optoelectronic and optical portions of this interconnection system. A photograph of the complete system is shown in Figure 2 below. The various components of the system will be described in the rest of this paper.



Figure 3. Photograph of the system (only one complete chip module is inserted in this picture)

3. ELECTRICAL SYSTEM

3.1 Electrical system overview

The electrical portion of this interconnection system can be divided into three major components. These are: i) the message processor printed circuit boards which source and sink data onto and from the interconnect; ii) the optoelectronic chips which contain the optoelectronic transceiver arrays; iii) the flexible printed circuit boards which connect the message processor boards to the chip whilst also allowing for a mechanical decoupling between the rigid optical system and the electrical system. The optoelectronic chip is attached directly to the flexible printed circuit board using chip-on-board packaging [6].



Figure 4. The layout of the optoelectronic chip [picture will be modified]

3.2 Optoelectronic chip

The VLSI-OE chip is a 9-mm x 9-mm silicon chip, fabricated using an n-substrate (p-well), 5-Volt, 2-metal layer, 1.5 μ m gate-width, CMOS process. The optoelectronic devices are multiple-quantum-well (MQW) P-i-N diodes used in reflection mode. They are patterned on a GaAs substrate and then flip-chip solder-bump bonded to the silicon with substrate removal. There are a total of 1024 MQW diodes and they are patterned as an 8x8 array of 4x4 clusters (see figure 4). The clusters are pitched at 800- μ m in both directions, and the diodes within each cluster are pitched 90- μ m in both directions. The modulator diodes are 50- μ m in diameter and and the detector diodes are 80- μ m x 80- μ m rectangles. The chip is also organized such that columns of modulator clusters alternate with columns of detector clusters. The use of clustering in this system was motivated by the requirements of the optical system [7,8].

The chip is based on the Hyperplane architecture [2] and consists of an array of smart pixels. Each smart pixel requires 4 MQW diodes, for a total of 256 smart pixels. The 256 smart pixels are grouped as 16 logical channels of 16 smart pixels arranged in horizontal rows. The smart pixel consists of a transimpedance amplifier used as an optical receiver, a bit-inversion multiplexer, a D-flip-flop, a by-pass multiplexer, an output concentrator multiplexer, a transmit multiplexer, and a transmitter driver (see figure 5). There are 93 transistors per smart pixel and almost 30,000 transistors on the chip (including bond-pad circuitry). To simplify the testing of the chip, each smart pixel is provide with a mechanism allowing it to operate either synchronously or asynchronously. During asynchronous operation, an external clock is not required. There are three

states of the smart pixel. The first is the "inject-state", which converts electrical input data to optical data on a per channel basis. The second is the "transparent-state", and allows optical input data to pass directly to optical output data. The third state is the "extract-state", which allows optical input data to be converted into electrical output data. Data can be injected into the channels using an input transmit-tree. This is a simple bus structure, 16-bits wide, and distributed into the array. An output concentrator, 16-bits wide, is used to extract data from the array on a per channel basis. Although any channel could be selected to extract data, due to the channel priority, the channel closest to the electrical output port is the one that accesses the port.

There are a total of 232 bond-pads on the chip. Of these, 24 are ground pads, and 20 power pads. There are 16 bond-pads for MQW diode biasing and 2 bond-pads for transimpedance amplifier optical receiver biasing. There are 64-I/O pads for electrical data, and 50 pads used for control and clocking of the channels. A total of 32 bond-pads are used for array test structures and a total of 24 are used for active alignment techniques.

To assist with the alignment of the chip to the optical system, several active and passive structures are designed into the chip and will be described in the section 5.



Figure 5. Schematic diagram of a smart pixel

3.3 Message processor board

In the final implementation of this system the message processor board will be fully compatible with the NUMAchine system [1]. However, in order to evaluate the optical interconnection system we have developed a stand-alone board that allows for more extensive testing of the optoelectronic chip and of the system. This board is a 10-layer PCB with 50 Ohm impedance matched traces. This board is designed to both generate and sink test vectors through the use of an ALTERA FPGA. In order to test system performance, the message processor board is capable routes signals on and off an on-board FPGA via 189 I/O lines while receiving its controlling signals from either an on-board DIP switch array or a PC-based LabVIEW 24-bit user interface. The FPGA can then output six distinct 16-bit test vectors (for which each bit may be activated or masked). The VLSI-optoelectronic chip ready data stream can then be directed to two or more of the 32 available channels.

3.4 Flexible printed circuit board

A flexible printed circuit board is used to connect the message processor board to the optical system and provides the electrical packaging for the optoelectronic chip. It is 18.5 cm long, has 4 layers and is designed with impedance-controlled I/O lines. Time domain reflectometry (TDR) measurements show a line impedance of $53 \pm 3\Omega$. It is connected to the message processor board via a 200-pin mixed technology connector. At the other end of the flexible PCB gold-plated fingers were provided for wire-bonding to an optoelectronic VLSI chip. Decoupling capacitors (0.15 μ F) were also provided in order to minimize switching noise.

4. OPTICAL SYSTEM

4.1 Optical system design

The optical system is a ring, each arm of which is a telecentric 8-f optical relay. A schematic diagram of the optical path which interconnects two nodes is shown in figure 6 below. A clustered optical interconnect design was selected [7] in order to provide the maximum possible optical throw with the greatest tolerance to misalignment. The optical array is divided into 8 x 8 clusters, each of which contains a 4 x 4 array of optical data channels. The system is partitioned into four modules: the chip module, the beam combination module (BCM), the relay module and the optical power supply module (OPS).

Due to the use of electro-absorption modulators, a continuous-wave beam is required to read out the state of the modulators. The optical power supply (OPS) generates a 512 beam right-hand circular polarised cw spot array at a wavelength of 852 nm through the use of two stages of diffractive fan-out [9]. The spots are focused onto the optoelectronic chip via an 8 x 8 array of mini-lenses. These lenses are 4 level diffractive elements, have a focal length of 8.5 mm and a square aperture of 800 μ m. The spots within each cluster are on a 90 x 90 μ m pitch and the cluster to cluster spacing is 800 μ m. The combination of the mini-lens array and the optoelectronic chip (in addition to the electrical and thermal packaging) is described as the chip module.

The beam combination module (BCM) is a polarization-based unit designed to route three arrays of beams: the continuouswave (CW) spot array of beams incoming from the OPS and directed to the modulators, the intensity modulated spot array reflected from the modulators and directed into the relay module and finally the spot array incoming from a previous stage and directed to the receivers. It is comprised of 5 components: 2 quarter-wave plates (QWP), a polarizing beam splitter (PBS), a patterned-mirror grating (PMG) and a corner prism. The patterned mirror grating (PMG) present on the BCM is composed of alternating stripes of diffractive fanout gratings and gold mirrors. The diffractive portion forms part of the optical power supply system. It is used to split each CW beam in the 4x8 array output from the OPS into a 4x4 cluster. The reflective part is used to reflect the spot array incoming from the relay module onto the detectors.

The relay module is a 4-f telecentric relay which contains the same diffractive minilens arrays as in the chip module. It consists of a block of SF56A glass (with a refractive index of 1.76) and a mini-lens array at either end. The over-all length of the relay module is 29.56 mm.



Figure 6. Flattened layout of optical relay between 2 stages.

A three dimensional view of the interconnect hardware is shown in figure 7. Note that the optics (G) are drawn as floating above the baseplate (F) for the sake of clarity but are designed to fit in the grooves (E) and insertion holes (D).

4.2 Optical system assembly

The assembly of the modules within the optical system is described in detail elsewhere [11,12]. A combination of visual alignment features and interferometric alignment [10] is used to precisely align the components within the relay module and the beam combination module. The misalignment tolerant design of the optical system permits the modules to be passively assembled within the baseplate.



Figure 7. Three-dimensional representation of the optical interconnect

5. ELECTRICAL-TO-OPTICAL PACKAGING

The interface between the electronic/optoelectronic portions of this system and the optical interconnect is of critical importance. This function is performed at the chip module. The chip module consists of the optoelectronic VLSI chip, its associated electrical and thermal packaging together with an 8 x 8 minilens array (see Fig. 6 above). The design and performance of this module is described in greater detail elsewhere [6]. The decision to package the minilens array with the chip module arose from a study of the optical alignment tolerances. When the minilens is packaged as part of the chip module the lateral misalignment tolerance between the chip module and the beam combination module is $\pm 26 \,\mu$ m. In contrast, if the minilens array were to be packaged with the beam combination module the lateral misalignment tolerance for the chip module is reduced to $\pm 8 \,\mu$ m. One of the design goals for this system was to ensure that the chip module could be removed and kinematically reinserted into the system. The larger lateral alignment tolerance was therefore an attractive feature of packaging the minilens array with the chip module. The penalty paid for the larger lateral alignment tolerance was a reduced tilt tolerance ($\pm 0.03^{\circ}$) between the minilens and the beam combination module. However, this degree of tilt tolerance can be

achieved by making use of the flat optical surfaces of the minilens array and the flat glass spacer which is attached to the beam combination module as reference surfaces. The required lateral alignment tolerance was achieved through the use of precision machined dowel pins in the chip module and corresponding holes in the optical interconnect baseplate. The chip module was inserted and removed 30 times and the mean deviation of the spots on the detectors was found to be $\pm 12 \mu m$, which is within the required range.

The alignment between the minilens array and the optoelectronic chip is critical. The minilens array must be aligned to a high degree of precision in all 6 degrees of freedom. However it is positioned 8.5 mm from the chip, which makes the use of passive mechanical features problematic. Off-axis Fresnel lenses were used to align the chip to the minilens array. Further details are given in reference [6]. Fig. 8 below shows the completed package module.



Figure 8 Optoelectronic chip package module, showing microlens array, TE cooler and flexible printed circuit board

6. **RESULTS**

6.1 Optical system

Further details of the optical system performance are given in reference 12. Fig. 9(a) below shows the 512 spot array at the receiver plane, after transmission between two stages. A detailed image of one cluster before transmission (at a modulator plane) is shown in Fig 9(b) and the spot cluster after transmission is shown in Fig. 9(c). The stage to stage throughput efficiency was found to be 25%. The dominant source of loss in the optical system was the diffraction efficiency of the 4 level diffractive relay lenses.

Fgure 9 Spot array at the receiver plane after transmission through the optical system (a); detail of spot cluster at modulator plane before transmission (b); detail of spot cluster after transmission (c).

6.2 Optoelectronic chip

Several modes of the optoelectronic chip have been tested. Figure 10 (a) shows the results of receiver mode tests. A binary modulated signal generated by an 850 nm wavelength vertical cavity surface emitting laser was incident on the receiver. The electrical signal generated on the chip was then relayed along the flexible PCB onto the message processor board via the onboard FPGA. The trace shows the output from the board. A data rate of 75 MHz was achieved. This is close to the maximum pad speed permitted by the 1.5 μ m geometry CMOS. Figure 10 (b) shows the results of transmitter mode testing. One of the modulators on the optoelectronic chip was illuminated with a cw beam. Digital signals from the message processor board were relayed to the optoelectronic chip and used to set the state of the appropriate modulator. The operation rate was again limited by the pad speed of the optoelectronic chip. Further results of system operation will be presented at the meeting.

Figure 10. Optoelectronic chip test results: (a) receive mode test; (b) transmission mode test

7. DISCUSSION

In developing this system, we have attempted to demonstrate that free-space optics has the potential to provide high density interconnections at the board to board level. This system has been designed and constructed in order to obtain maximum benefit from the parallelism provided by free-space optics. Although the interconnect is configured as a series of point to point links between boards, the use of the hyperplane architecture allows data to hop over multiple nodes with low latency.

However, there are further improvements that could be made at all levels of the system. The optical components are implemented as a set of modules which can then be integrated together without requiring critical alignment. In the future it would be desirable to implement some of these modules as single components molded in plastic. The relay module is a strong candidate for this process as it consists simply of two diffractive surfaces separated by a transparent block. An example of the integration of free-space optical systems via plastic molding is given in reference [13]. Another example of the way in which the optical system could be improved would be to reduce the size of the optical power supply module. At present each optical power supply module is 20 cm long. A technique for reducing the size of this module by the use of a folded micro optical system is described in reference [14]. However, the optical system would be further simplified if the modulators were replaced with emitting elements such as vertical cavity surface emitting lasers (VCSELs). This would remove the need for the optical power supply altogether. An example of a VCSEL- based system which contains 256 optical channels is given in reference [15]. However, the use of VCSELs would require a modification of the optical design in order to accommodate the high divergence of a VCSEL beam. A final potential modification would be to employ a flexible optical interconnection in place of the current rigid link. Candidate optical technologies include fiber image guides [16-19] and ordered fiber arrays [20,21]. To-date a two-dimensional flexible parallel optical of link containing 512 parallel channels has not yet been demonstrated. However, array sizes of 64 channels have been reported [19]. The use of a flexible optical interconnect would allow the flexible printed circuit board to be eliminated so that the optoelectronic chip could be mounted directly on the message processor board.

Improvements could also be made to the optoelectronic chip. One example of this would be the inclusion of flow control and error correction circuitry. Reference 20 contains a detailed description of optoelectronic chip for parallel interconnect applications which contains such circuitry.

8. CONCLUSION

We have presented a free-space optical interconnection system which is designed to interconnect four boards in a unidirectional ring configuration for optical backplane applications. Each board is connected to an optoelectronic chip which contains an array of 512 modulators and 512 detectors. A dual-rail encoding scheme is used so that each board transmits 256 data channels to the next board in the ring and receives 256 and data channels from the previous board. A rigid micro optical system is used to transmit the optical signals between the optoelectronic chips. This results in a very high density of 1,250 channels/cm². The micro optical components are assembled into modules which are then integrated into the optical system. The system is designed in such a way that critical alignment is only necessary during module assembly. The modules can then be integrated with much looser alignment requirements. The packaging of the optoelectronic chip is a critical part of the system, as it provides the interface between the electrical section of the system and the optical section. In this system improved alignment tolerances are achieved by integrating a microlens array with the optoelectronic chip packaging. The interconnect is partitioned into 16 logical channels each of which contains 32 I/O bits. Control circuitry on each chip allows 32 bits of data to be transmitted optically in parallel from one board to one or more of the other boards on the ring. The operation of individual components of this system has been verified and further details of system operation will be presented at of the meeting.

9. ACKNOWLEDGEMENTS

This work was supported by a grant from the Canadian Institute for Telecommunications Research under the NCE program of Canada, the Nortel/NSERC Industrial Research Chair, NSERC operating grants (OGP0194547), FCAR operating grants (NC-1778). We would like to acknowledge the support of the Canadian Microelectronics Corporation in fabricating the CMOS VLSI chip used in this system.

10. REFERENCES

1. Z.Vranesic, S.Brown, M.Summ, S.Caranci, A.Grbic, R.Grindley,"The NUMAchine Multiprocessor", CSRI Technical Report, University of Toronto

- 2. T.H. Szymanski and H.S. Hinton, "Reconfigurable intelligent optical backplane for parallel computing and communications", Applied Optics, vol. 35, pp. 1253-1268, 1996.
- 3. T.H. Szymanski and H.S. Hinton, "Optoelectronic smart pixel array for a reconfigurable intelligent optical interconnect", U.S. Patent no. 6,016,211, January 18th, 2000.
- A.V. Krishnamoorthy and D.A.B. Miller, "Firehose architectures for free-space optically interconnected VLSI circuits," J. Parallel and Distributed Computing, vol. 41, pp. 109-114, 1997.
- 5. Hinton review paper, Dual rail encoding
- M.H.Ayliffe, D.R.Rolston, E.Lchuah, E.Bernier, F.S.J.Michael, D.Kabal, A.G.Kirk, D.V.Plant, "Packaging of an optoelectronic-VLSI chip supporting a 32 x 32 array of surface-active devices", to be presented at *Optics in Computing* 2000, June 19th-23rd, Quebec, Canada.
- 7. D Rolston, clustering paper
- 8. B. Robertson, "Design of an optical interconnect for photonic backplane applications," Applied Optics, vol.37, no.14, pp. 2974-84, 1998,
- 9. D F.-Brosseau, F Lacroix, M H. Ayliffe, E Bernier, B Robertson, F A. P. Tooley, D V. Plant, and A G. Kirk, 'Design, implementation, and characterization of a kinematically aligned, cascaded spot array generator for a modulator-based free-space optical interconnect' *Applied Optics*, **39** (5), pp 733-745, 2000.
- 10. B Roberston, Interferometric alignment
- F. Lacroix, B. Robertson, M.H. Ayliffe, E. Bernier, F.A.P. Tooley, M. Chateauneuf, D.V. Plant, A.G. Kirk, "Design and implementation of a four-stage clustered free-space optical interconnect", Proceedings of the OSA Topical Meeting on Optics in Computing, vol. 3490, pp. 107-110, 1998.
- 12. E.Bernier et al, OC2000
- 13. D T Neilson, E Schenfeld, 'Free-space optical relay for the interconnection of multimode fibers', *Applied Optics*, **38**, pp 2297, 1999.
- 14. M.Chateauneuf et al, OC2000 Folded OPS
- 15. D.Plant et al, Demo paper OC2000
- T. Maj, A. G. Kirk, D. V. Plant, J. F. Ahadian, C. G. Fonstad, K. L. Lear, K. Tatah, M. S. Robinson, and J. A. Trezza, "Interconnection of a Two-Dimensional Array of Vertical-Cavity Surface-Emitting Lasers to a Receiver Array by Means of a Fiber Image Guide", *Applied Optics* 39(5), pp. 683-689, 2000
- 17. Y. Li, T. Wang, H. Kosaka, S. Kawai, and K Kasahara, "Fiber-image-guide-based bit-parallel optical interconnects," *Appl. Opt.* **35**, 6920-6933 (1996).
- D. M. Chiarulli, S. P. Levitan, P. Derr, R. Hofmann, B. Greiner, and M. Robinson, "Demonstration of a Multichannel Optical Interconnection by use of Imaging Fiber Bundles Butt Coupled to Optoelectronic Circuits", *Applied Optics* 39 (5), pp 698-703, 2000.
- 19. OFC 2000 paper on 8x8 FIG
- J M Sasian, R A Novotny, M G Beckman, S L Walker, M J Wojcik, S J Hinterlong, 'Fabrication of fibre arrays for optical computing and switching applications', OSA International Conference on Optical Computing Technical Digest, pp 229-230, 1994.
- 21. C.V.Cryan, 'Two-dimensional multimode fibre array for optical interconnects', *Electronics Letters*, **34**, pp 586 587, 1998
- 22. T.H.Szymanski, M.Saint-Laurent, V.Tyna, A.Au, B.Supmonchai, "Field-programmable logic devices with optical inputoutput", *Appl. Opt.* **39**, pp 721-732, 2000.