A Distributed Computing Demonstration System Using FSOI Inter−Processor Communication

J. Ekman[1], C. Berger[2], F. Kiamilev[1], X. Wang[1],
H. Spaanenburg[3], P. Marchand[4], S. Esener[2]

[1]University of Delaware, ECE Dept., Newark, DE 19716, USA
[2]University of California San Diego, ECE Dept., La Jolla, CA 92093, USA
[3]Mercury Computer Systems Inc., Chelmsford, MA 01824, USA
[4]Optical Micro Machines, San Diego, CA 92121, USA

Abstract. Presented here is a computational system which uses free−space optical interconnect (FSOI) communication between processing elements to perform distributed calculations. Technologies utilized in the development of this system are integrated two−dimensional Vertical Cavity Surface Emitting Lasers (VCSELs) and MSM−photodetector arrays, custom CMOS ASICs, custom optics, wire−bonded chip−on−board assembly, and FPGA−based control. Emphasis will be placed on the system architecture, processing element features which facilitate the system integration, and the overall goals of this system.

1 Introduction

The area of optical interconnects is continually growing with many advances in optoelectronic devices, integration of CMOS ICs with these devices, and integration of hybrid electrical/optical devices into functional systems. It is clear that the flexibility in terms of scalability, and optical bandwidth which can be achieved by using optical interconnects will lead to changes in system architectures as designers move to to take advantage of this flexibility. As a part of the 3−D OptoElectronic Stacked Processor program[1], a demonstration system is being developed which illustrates the ability to construct distributed computational systems which use optical communication for passing data between processing elements. In this system, the distribution takes the form of linear chains of processors with nearest neighbor communication.

Communication between processors in a multiprocessor system quickly becomes the bottleneck and is therefore an ideal target for the integration of optical communication. One of the goals in developing this system was that of illustrating the use of optical communication in a low cost distributed system as a step toward validation of such architectures.

2 System Topology

This demonstration system consists of two linear chains of five processors each. Three processors in each chain are configured to perform computation and the two remaining (one on each end of the chain) are configured to bring data into and out of each chain. This is accomplished by converting between electrical−domain (digital)

and optical−domain (analog) signals at the ends of each chain (see figure 1). The two chains operate independently, but based on available optoelectronic device arrays, share OptoElectronic (OE) chips for communication. In addition to the ability to lengthen each chain, there is flexibility to scale the number of chains to yield a larger system. The optical chip−to−chip communication is achieved through the use of two dimensional VCSEL and MSM−photodetector arrays provided by Honeywell Technology Center[2] and custom optics designed at UCSD.

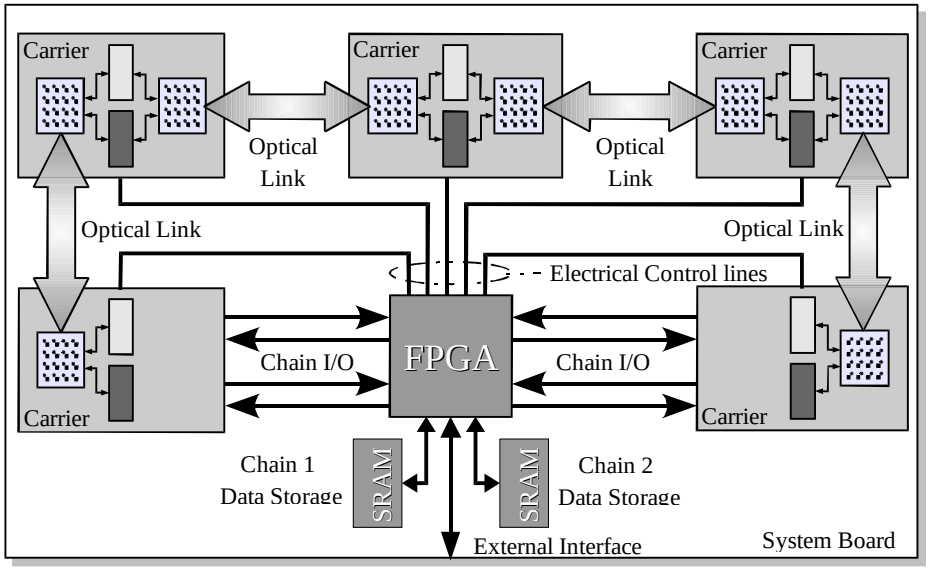

Figure 1. System diagram showig five carrier boards placed on system board. OptoElectronic arrays are shown on left and right sides of carrier boards and processing elements in the center of the carrier boards. The upper *(light)* PEs indicate one chain and the lower *(dark)* PEs indicate the second chain

2.1 Carrier Boards

Each unit in the chain is assembled onto a small "carrier board" where each of these carrier boards contains two processing elements (PEs) and two OE arrays. The OE arrays consist of sixteen VCSELs and sixteen photodetectors in an inter−digitated 4 x 4 array. These parts originally fabricated as a part of the GMU Co−Op program[3]. Each of the chips on the carrier boards are bare die, wire−bonded to contacts on the carrier board. One PE belongs to each of the two chains and the OE arrays are shared among the two chains with dedicated array elements for each chain. These carrier boards are then mounted onto a "system board" which also supports the optics, additional chips to provide control and system interface, and power connectors, etc. For the system described here, there are five carrier boards mounted onto one system board. This is illustrated in figure 1. Another goal of this demo system is to experiment with different opto−mechanics in an effort to

demonstrate the ability to scale−down what has traditionally been a (physically) large part of such systems through the use of "plug−on−top" optical assemblies[4]. The construction of the carrier board modules facilitates this by allowing independent units to be rotated or moved according to a particular optical arrangement.

2.2 System Board

The purpose of the system board is to serve as a substrate for the entire system, supporting the carrier boards and opto−mechanics as well as providing the necessary control to the processing elements, and interfacing with the "outside world" to provide power, data, and system diagnostics. The board itself is a multi−layer printed circuit board (PCB) fabricated commercially. There is electrical and precision mechanical connection of the carrier boards to the system board. The primary components that perform the control and interfacing tasks are a high−end Xilinx Virtex FPGA and commodity SRAM. The Virtex FPGA was chosen for its high pin−count and capacity allowing control of the entire system from one chip and giving great flexibility to re−configure the system. It provides both the data necessary to configure the processing elements initially and control their operation throughout calculations. Additionally, it will provide data to the processor chains, gather results and monitor the results checking for errors. This approach helps reduce risk by allowing for reprogramming of the FPGA and also helps during assembly of such a prototype system. An extension of this system would have built−in controllers with the PEs and allow higher−level programming.

3 Processor Interconnection

The processors in this system are connected in two linear chains with each processor communicating with the one to its left and the one to its right. On the ends of the chain, there is only optical communication in one direction. Data is brought in and taken out of the ends of the chain electrically. The interconnection scheme chosen is meant to facilitate construction of this prototype system and serve as a starting point which can lead to more complex connection schemes which may provide additional benefit to specific applications. The logical connection of the processors in this system and the connection to the FPGA control unit is shown in figure 2. With this connection scheme, all data is brought into the processor chain from the two ends. All data communication within the processor chain is through the FSOI links. This both helps illustrate the viability of optical communication in a multiprocessor system as well as ensure that the links will be heavily utilized. The impact on the system architecture is of course that data must be passed to processors in the center of the chains before they can begin calculations. This is not seen as a serious drawback in this system as it adds only some latency to the beginning of calculations. It should be mentioned here that the application chosen for demonstration on this system is that of a radix−2 butterfly engine as a part of an FFT calculation. With this application, data points are brought into the chain and bounced back and forth between the processors in the chain during calculation and finally output from the ends of the chain. The two chains of this system are utilized to compute real and imaginary points simultaneously.

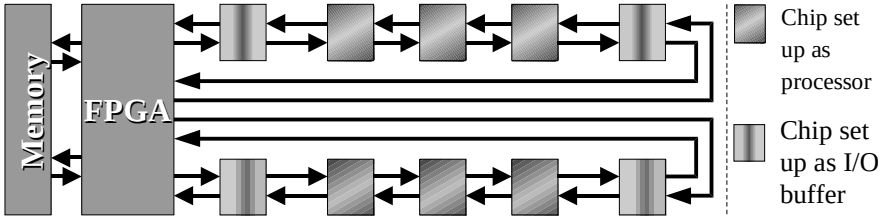

Figure 2. Diagram of logical connection between the multiple processors and the FPGA controller and memory. The upper and lower chains here illustrate the two independant chians in this system

4 Processing Element

The processing element itself is a custom ASIC designed and fabricated for this demo system. It is a 0.5 micron CMOS chip of roughly 10,000 transistors comprising both digital and analog circuitry (shown in figure 3). Some of the design goals for this chip were the ability to interface with the optoelectronic devices to be used in the system, that it provide digital signal processing capability, facilitate system construction and debugging, allow for possible changes to the optical system, and provide the capability to use the chip as an electrical/optical interface at the ends of each chain.

The design of the processing element is divided into the following functional units: Input/Output switching, arithmetic and logic units, optical I/O buffering, and a control interface. Input and output switching provides for the re−organization of data as it is received from, or transmitted over the optical chip−to−chip data links. The arithmetic and logic units provide the calculation capability based on a small instruction set. Translation between the optoelectronic analog domain and the digital domain is accomplished through on−chip receiver and VCSEL driver circuits. The control interface provides for configuration of the input/output switches and the selection of the function performed by the chip.

Input and output switches provided give much of the flexibility achieved in this design. The primary function of the input switches is the correction to input data words that may be necessary due to changes in the optical communication between chips or system I/O. The chip−to−chip communication links are all eight bits wide and the internal datapaths of the PE are six bits wide. The two remaining links out of every eight are devoted to fault tolerance. In the event that a data link is non−operational for any reason, the data being sent over that link can be diverted to one of these two redundant links. In such a case, the input switch would re−assemble the data word before calculation begins. In this manner the calculation is not corrupted or impeded by a loss of a link between chips. This fault tolerance is important in a demo system to ensure that a faulty link does not deteriorate the demonstration, but will also be important in future systems to provide reliability.

The output switches compliment the fault tolerance achieved with the input switches by providing the capability to re−route outgoing data onto a redundant link in the event that a link is known bad. Additionally, the output switches are used to select between outputs of the arithmetic and logic units, the receiver outputs in order

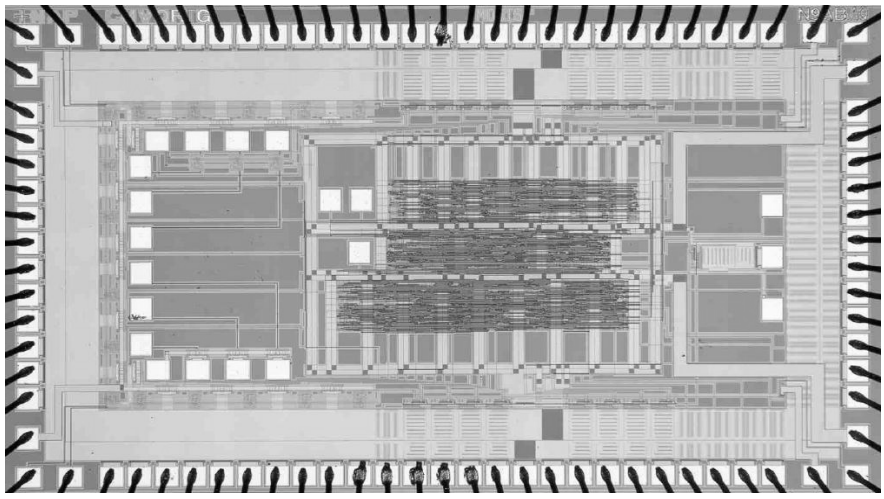

Figure 3. Microphotograph of the CMOS ASIC used as the processing element in this multiprocessor system. Eighty−six wire bonded pads are shown at the chip perimeter. Other unbonded pads are for probe−testing

to completely by−pass the processing functionality, and an auxiliary set of inputs which allow the chip to be used simply as a parallel VCSEL driver. Complete by−pass functionality is included in the PE chip to add flexibility and aid in system construction and debugging as it will allow chips to logically be removed from the chain without changes to the optics and also isolate the optical path from the digital functionality. The dataflow through the PE is shown in the diagram of figure 4.

In addition to the possible loss of an optical data link, changes to the optical system may result in a flipping of the data word during transmission. In order to allow different optical systems to be explored with this system, the ability to account for such flipping is included in the input switches. A final feature of the input switches is the ability to interchange the two inputs before sending them to the Arithmetic and Logic Units.

The arithmetic and logic unit (ALU) is a custom developed component which provides the capability to perform addition, subtraction, and multiplication of signed or unsigned numbers as well as a variety of common logic functions and comparisons for maximum/minimum determination. The unit is a three stage pipeline to increase achievable clock rates which gives the PE its characteristic three−cycle latency on all instructions except complete by−pass. Scan chain registers are used in the ALU and include the capability to generate pseudo−random data to provide testability.

The on−chip analog receiver and VCSEL driver cells included on the CMOS ASIC are previously verified designs from UCSD and UNCC/UDel respectively and were designed to operate with the specific OE elements used in this system. As an additional testability feature, stand−alone copies of these cells have also been placed on the ASIC connected to probe pads.

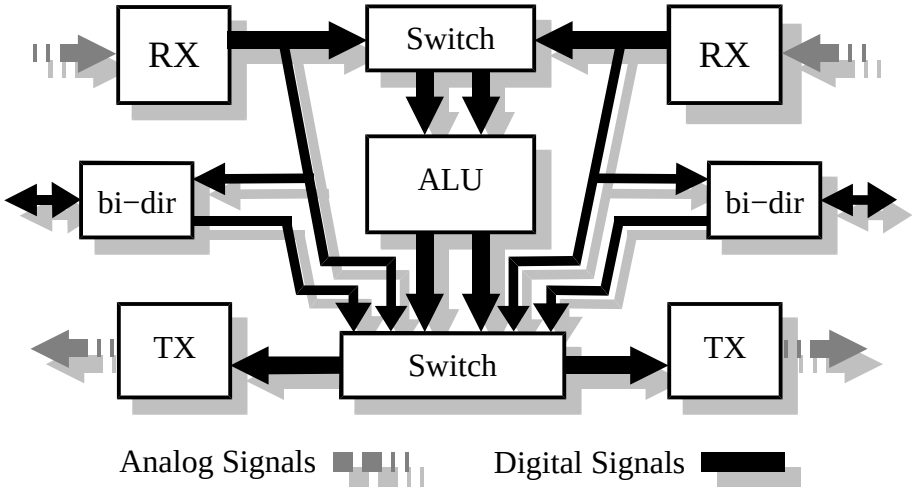

Figure 4. Architecture diagram of the processor element showing dataflow through the chip. *(Thinner)* lines indicate dataflow pattern when by−passing the computational portion of the chip

5 Conclusion

The current state of integration of optical communication with digital CMOS logic affords the ability to build functional systems from which new processing architectures can evolve. We have taken advantage of this to build a prototype multiprocessor demonstration system which utilizes FSOI data communication. This system is currently in the final stages of development and additional results will be presented at the conference.

References

1. 3D−OESP Consortium website: http://soliton.ucsd.edu/3doesp/
2. George Mason University Consortium for Optical and Optoelectronic Technologies in Computing website: http://co−op.gmu.edu/
3. Honeywell Technology Center: http://www.htc.honeywell.com/photonics/
4. C. Berger, J. T. Ekman, P. J. Marchand, F. E. Kiamilev, H. Spaanenburg: Parallel distributed free−space optoelectronic compute engine using flat "plug−on−top" optics package, accepted for presentation at the International Topical Meeting on Optics in Computing in Quebec, Canada, June 2000

Effort sponsored by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory under agreement number F30602−97−2−0122. The US government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon.