

# Systolic Processing Architectures Using Optoelectronic Interconnects

Sek Meng Chai, Abelardo López-Lagunas,  
D. Scott Wills, Nan Marie Jokerst, Martin A. Brooke  
School of Electrical and Computer Engineering  
Microelectronics Research Center  
Georgia Institute of Technology  
Atlanta, Georgia 30332-0250

## Abstract

*Systolic arrays have traditionally provided efficient, high performance execution for computation intensive applications. Despite the extensive research in systolic arrays, system designers must continually incorporate new technological advances to improve node communications, I/O bandwidth, and programmability. This paper presents optoelectronic interconnect as a communication method for systolic arrays in early image processing applications. Optoelectronic interconnects provide potentially high I/O bandwidth required to maintain high utilization rate for systolic arrays. In addition, optoelectronic interconnects provides a two-dimensional focal-plane topology ideal for systolic image processing systems. This paper introduces two new systolic architectures that incorporate integrated optoelectronics to provide an extremely compact, high performance, highly efficient image processing system. Several important early image processing applications developed for these architectures are also described.*

## 1. Introduction

A new emerging class of computational structures, "smart pixels", integrates optoelectronic devices, typically in a plane, with processing electronics. Traditional smart pixel systems are used for image processing where a small array of pixels is mapped to a processing node. Each node, illustrated in Figure 1, contains three components: optical inputs, "smart" electronics, and optical outputs. Optical inputs deliver electronic signals to the processing electronics. The processing electronics consist of either digital or analog functionality for processing images. Results from the processing electronics are then output as optical signals.

The integrated optoelectronic devices offer a potentially lower cost, higher density alternative to traditional wire-based I/O. For image processing applications that require high off-chip bandwidth, these devices provide compact, low latency focal-plane processing that is more difficult to achieve with traditional perimeter wire bonding. Integrated optoelectronics also provide a closer coupling between

optical devices and processors for lower noise, higher sensitivity operations.

Despite the potential of integrated optoelectronics, smart pixel systems generally provide simple processing functions compared to general purpose microprocessors and digital signal processing chips that provide much greater generality at lower performance and efficiency per silicon area. The ideal smart pixel architecture provides not only computational versatility but also an efficient operation. SIMD (Single Instruction Multiple Data) and MIMD (Multiple Instruction Multiple Data) architectures incorporating optoelectronic devices have been reported [4,14,16], but are more versatile than needed for most early vision applications. Systolic computation for early image processing provides excellent fundamental design methodologies for smart pixel systems. Unlike SIMD and MIMD architectures, systolic processing provide compact, efficient, high throughput computation. In addition, systolic design methodologies establish localized communication that allows better scalability for larger image processing applications.

This paper presents a systolic architecture and its compatibility for early image processing in smart pixel systems. A focal plane topology incorporating integrated optoelectronics is introduced as an ideal topology for systolic image processing systems. Focal plane processing provides a three-dimensional computing structure (a two-dimensional computational mesh with data flow in the third dimension) [5]. Two new systolic architectures for early vision image processing, being developed at Georgia Tech, PAMSAC and GT-VISTA, are presented. These systems rely on optoelectronic devices integrated on the focal plane that provide high I/O bandwidth for systolic computation. They use integrated thin film optoelectronic devices interfaced with CMOS analog amplifiers to provide interconnectivity among systolic elements in a three-dimensional processing structure [7,11].

## 2. Early image/vision processing systolic array smart pixel systems

Systolic architectures were introduced by H. T. Kung in 1978 as an architectural approach to exploit the promise of

VLSI technology [13]. The design methodologies were set to alleviate communication penalties by using modular cells without global interconnects, and to decrease design time by reducing complexity. H.T. Kung stressed scalability, high degree of concurrency, and repetitive use of data to efficiently utilize the increasing number of transistors per integrated chip. His landmark paper forms the fundamental basis for early research in systolic architectures. The following describes the three basic design principles of systolic arrays.

- *Modular Cells.* The first design principle to use simple modular cells was set to cope with increased design complexity. For special purpose systems, this is an attractive guideline to reduce the design time by reusing common building blocks. These common building blocks can be optimized for area, power, and speed to create the overall speedup for the entire system.
- *Localized Communication.* Long interconnections consume more power and chip area, and have high latency. The second design principle establishes simple, regular communication to maintain short communication paths. By incorporating the communication flow into the overall design, the outcome may be an efficient, regular implementation without long communication paths.
- *Balanced Computation.* A VLSI computation is feasible only for compute-bound algorithms where there is more computation to be performed than available inputs and outputs. A good systolic design should therefore ensure that for every I/O operation, multiple operations are performed to obtain maximum throughput.

Systolic execution model allows for extremely efficient implementation of systems that solve computational intensive applications. These computations are pipelined in many dimensions in order to obtain high computational throughput. Systolic arrays maximize the processing performed on a datum once it has been obtained from I/O or memory by reusing the datum as it moves through the pipelines in the array. Unlike a regular pipeline structure, however, the input data as well as partial results flow through the array at varying speeds and in multiple directions [10].

Systolic arrays work well with algorithms that require high processing and high throughput, such as early image processing applications. The key components of these image processing algorithms include: large amounts of input data, complex algorithms, and high data rate requirements [15]. These algorithms involve simple operations with the output at each pixel relying on neighborhood pixel values. Examples of early vision algorithms include contrast enhancement, convolution, edge detection, and other simple filtering operations.

Smart pixel systems map image elements to a processing node, and therefore exhibit neighborhood dependency indicative of early vision algorithms.

Consequently, traditional applications for smart pixel systems include early vision algorithms. Hence systolic design methodologies of modular cells, localized communication, and balanced computation are naturally consistent for smart pixel systems. By modifying an image processing application for a smart pixel system as a systolic algorithm, and designing the smart pixel system with systolic design methodologies, a designer can further the smart pixel capabilities with systolic characteristics of high computation rate and high throughput.

In higher level image processing applications, complex operations must be performed on collection of data, usually separated by large distances in terms of memory location or processor nodal location. These operations involve decisions or interpretations based on the relationships between the image data and databases or mathematical models [8].

### 3. Systolic arrays with integrated optoelectronics

Traditional two dimensional systolic arrays rely on perimeter input to feed data into its inner core. The performance of such configuration is limited by the I/O bandwidth, and thus the ultimate performance should be a computation rate that balances the available I/O bandwidth with external devices or memory [13]. Existing research has been focused on the systolic mappings of algorithms that limit the number of I/O rate per computation. In these systems, data items and interim results are stored and used repetitively to alleviate external I/O bandwidth. Figure 2(a) illustrates a two-dimensional mesh of systolic processors with perimeter input.

In a perimeter I/O limited system, the image data is often segmented into small scan-line rows. Each row of cells processes only a portion of the scanned image. Consequently, the processed image typically flows through neighboring rows in order to reach the output boundary cells at the opposite edge of the mesh of processing elements. The algorithm is usually pipelined in rows of cells in order to increase the throughput of the system.

New technologies, such as flip-chip bump bonding and through-wafer optoelectronic devices [7,11], provide additional communication accesses to the inner cells in the 2-D mesh topology. Figure 2(b) shows the focal plane topology with immediate communication access to the inner core [3]. This focal-plane configuration allows an additional dimension of I/O compared to the perimeter I/O. As in the perimeter I/O limited system, this focal-plane topology can be pipelined such that each layer of processing nodes operates on the entire image at an instance. There is no need to segment the image into smaller sub-images for processing. In addition, the systolic computation rate can

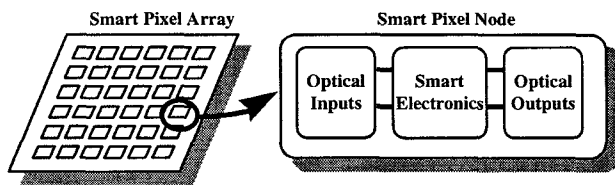


Figure 1. Smart pixel nodal components

rise accordingly to match the already increased I/O rate instead of being limited by the I/O rate in a perimeter I/O limited system.

While this focal-plane topology is common in image processing systems [4,5], our intent is to assert that this topology is the ideal configuration for systolic image processing system. These configurations can provide the short communication path and scalability that are consistent with the systolic design principle. This section will proceed to describe the consistency of systolic design with characteristics of focal plane topology utilizing optoelectronic devices in systolic smart pixel systems.

- **Short Communication Path.** The focal plane topology provides short communication paths among cells between layers of systolic 2D meshes. These cells no longer have to transmit the processed image data along intermediate rows to reach the boundary cells to transmit the output, but rather in a vertical flow towards a new 2D VLSI systolic smart pixel plane.
- **Scalability.** The smart pixel system maintains scalability by incorporating the optical input and output into a cell. The I/O path no longer depends on boundary cells, but instead, individual cells can transmit or receive data. Each cell is modularized since the optical capabilities are granted to every cell. This property of scalability is important for extensibility, since for a large image size, the system only needs to add additional cells to fit a larger problem size.
- **Balance Computation.** The systolic mesh can perform at a higher computation rate to match the increased I/O bandwidth in the inner cells. As a result, the focal plane topology increases the overall system throughput. In addition, programming is simplified since the boundary cells in the systolic mesh no longer have to perform special routines for I/O transmission. The overall processing can be distributed more evenly through the systolic mesh to produce a more balanced computation.

With this focal plane topology for systolic arrays, new programming methodologies are needed for smart pixel algorithms. Algorithm mapping, problem partitioning and cell synchronization for applications such as early image processing must be reviewed in order to utilize desirable characteristics provided by this new topology. For

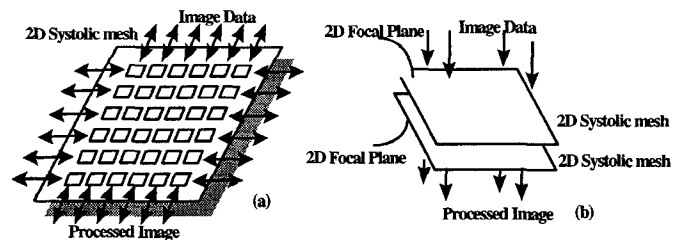


Figure 2. (a) Perimeter I/O via boundary cells and (b) Increase I/O bandwidth with focal plane topology

example, pairs of cells in different layers of the systolic mesh can be coupled to form a processing node. This node, comprised of two systolic cells in two different layers, has more computational capabilities than a single node and may lead to a better algorithmic mapping, partitioning, and synchronization for the application. However, the re-mapping and re-partitioning tasks must be performed for each application in order to achieve better performance.

The 2D systolic mesh can be pipelined in a similar way to the perimeter I/O limited system. In the focal plane topology, an instance of the image is situated in a systolic mesh and is allowed to propagate in a pipeline fashion to the next systolic layer. Instead of processing at a scan row image in the perimeter I/O limited system, the focal plane topology allows pipelining of an entire image.

## 4. Two systolic array smart pixel systems

The following section describes current systolic smart pixel research at Georgia Tech. The first architecture, PAMSAC, is a fixed function systolic array that performs perfect pattern matching. While PAMSAC has modest computational power and limited programming flexibility, it demonstrates the realization of optical interconnection with integrated silicon detectors. The systolic processing core in PAMSAC illustrates the compact, highly efficient computational power of a systolic architecture. The second architecture, GT-VISTA, is multi-layered systolic array with a programmable pixel node. GT-VISTA focuses on early vision processing and provides an optimized, compact datapath to provide programming flexibility while maintaining a degree of efficiency.

### 4.1 The PAMSAC architecture

Figure 3 is the layout and floorplan of PAMSAC [16], a pattern matching chip that incorporates direct optical input of image data via eight on-chip silicon detectors and amplifiers. This  $2252\mu\text{m} \times 2222\mu\text{m}$  chip was implemented through the MOSIS foundry in  $2.0\mu\text{m}$  CMOS technology and operates at 33MHz. Digital logic testing of the systolic core has been completed; the interface to the optoelectronic devices is currently in progress.

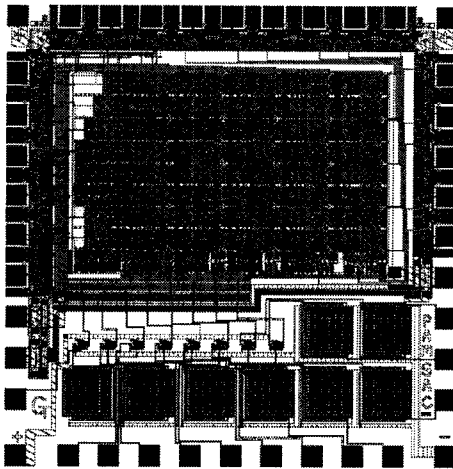


Figure 3. VLSI layout and floorplan of the PAMSAC chip

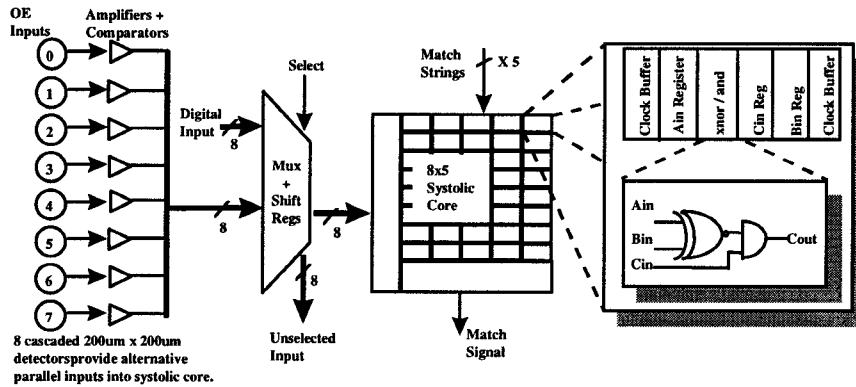


Figure 4. Block diagram of the PAMSAC chip

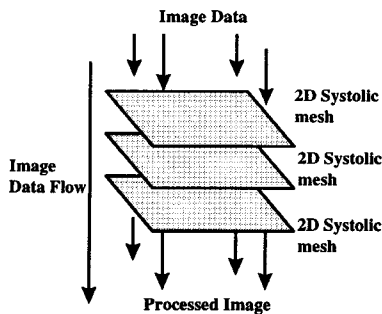


Figure 5: GT-VISTA multi-layer structure

Figure 4 illustrates the block diagram of the PAMSAC chip with a systolic cell logic operation consisting of an XNOR and an AND gate to detect perfect pattern matches. Although the programmability of PAMSAC is fixed, a different digital set of *Match Strings* can be loaded into each cell for comparison to a stream of optical inputs. The systolic core consists of 40 cells to match 40 input bits simultaneously. The system can be scaled to match larger number of optical input bits with additional systolic cells and detectors.

The PAMSAC system demonstrates the realization of optical inputs to provide the high I/O bandwidth requirements. As a systolic system, PAMSAC can maintain a high computational and throughput rate once the system is configured. Each systolic cell is dense, and the overall system size is very compact. PAMSAC utilized localized communication to maintain scalability of the system.

## 4.2 GT-VISTA system overview

The GT-VISTA system consists of multiple layered two-dimension meshes of systolic processors. Each of these layers works collectively to perform low level vision processing. Figure 5 illustrates the proposed 3D topology for the GT-VISTA system. This multi-layered topology is common in image processing, but to the best of our knowledge, GT-VISTA presents the first multi-layered systolic mesh with optoelectronic interconnects. The motivation of the GT-VISTA system came from the desire to harness the capabilities of systolic computation with focal-plane processing.

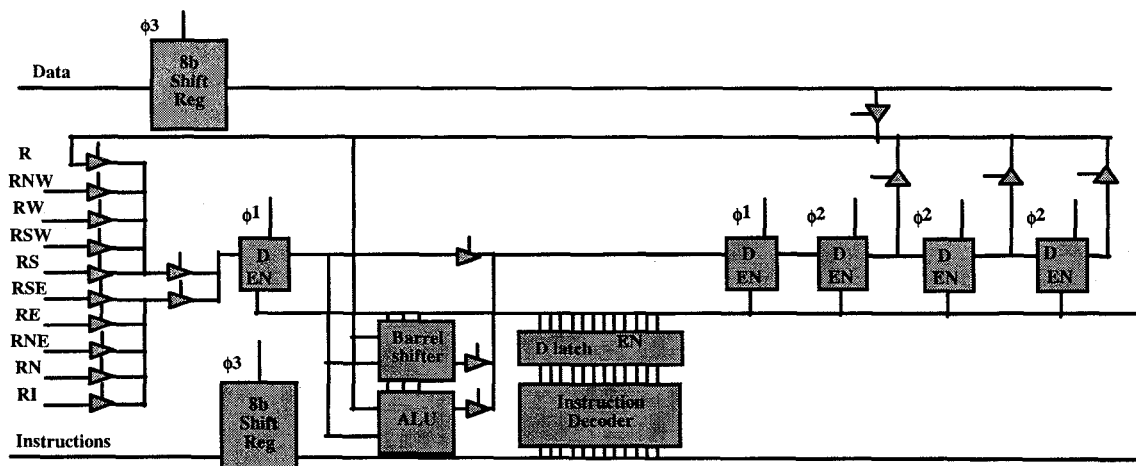


Figure 6: Block diagram of GT-VISTA processor

Communication between systolic layers is achieved with through wafer optoelectronic devices. In the first version of the GT-VISTA system, inter-layer communication is limited to image data flow. Because of the multi-layer topology, high throughput can be achieved by pipelining the image through each 2D systolic mesh. Each pixel is directly mapped onto a systolic element. This mapping is possible because each GT-VISTA systolic element is designed to be small. In addition, the targeted early vision applications typically require per pixel processing. Inter-nodal communication is enabled for all the nearest neighbors: North, East, West, South, North East, North West, South East, and South West. This neighborhood communication is unlike existing NEWS communication because it provides more access than just North, East, West, and South. These neighborhood pixel values require extra clock cycles to load data into the communication registers. Figure 8 illustrates the inter-nodal communication for a systolic node within the mesh.

### 4.3 GT-VISTA microarchitecture

The communication network allows nearest neighbor communication only. Each GT-VISTA node can read the pixel value of its 8 immediate neighbors. The datapath of a node consists of an 8-bit integer ALU, a barrel shifter, and special pipeline registers. This architecture is similar to the single bit MIP morphological system [6], but modified in GT-VISTA with 8 bit data words and a more powerful ALU and barrel shifter. The ALU and shifter will allow enough processing capabilities for early vision processing. Figure 6 illustrates a single node block diagram.

Each GT-VISTA node is mapped to a single thin film detector to form the focal plane processing architecture. Because most detectors generate analog signals from light

intensities, an analog to digital converter (A/D) is required to convert the intensities into digital values [3,7,11]. The prototype GT-VISTA system has light input arriving at a node as a serial sequence of bits that is combined to form an 8 bit data word. Figure 7 illustrates the thin film integration with a GT-VISTA node. Although through-board wiring can be used to maintain the focal plane topology and the 3D structure, the prototype GT-VISTA system is demonstrated with optoelectronic devices as a means of interconnect.

The memory controller is set on the boundary of the 2D systolic mesh. The controller sends program instructions and data only to the boundary of the mesh. This information propagates through each column in the mesh in a systolic fashion. The idea of propagating instructions with moving data sets was introduced in [9], and is utilized here for cell scalability.

### 4.4 GT-VISTA applications

GT-VISTA is designed to perform repetitive neighborhood pixel access operations. Figure 9 illustrates a program segment for a convolution program with a low pass filter mask [2]. The following calculation shows a per node processing capability for a LPF convolution program operating at 100MHz.

Notice that the GT-VISTA instruction set is unlike traditional instruction sets in that many operations such as calculation and register data movement can be performed in a single cycle.

$$\frac{100 \text{ Mop}}{\text{sec}} \times \frac{\text{pixel}}{15\text{ops}} = \frac{6.67\text{Mpixel}}{\text{sec}}$$

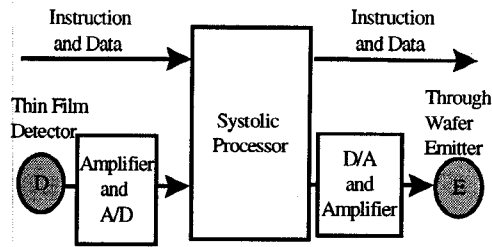


Figure 7: Detector and Emitter pair with systolic processor

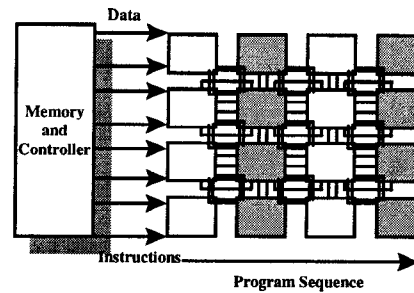


Figure 8: GT-VISTA Neighborhood Communication

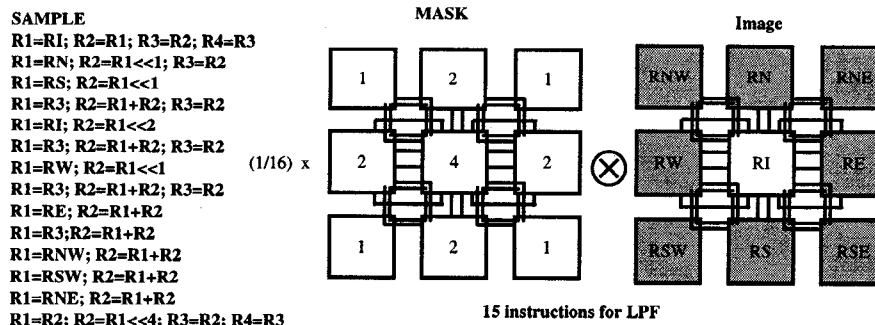


Figure 9: Convolution program with LPF mask

The high frame rate capability of GT-VISTA is derived from the pixel-to-node mapping of a systolic element. For example, if a systolic element is mapped to an image pixel, then an overall frame rate of 6.67Mframe/sec can theoretically be achieved. An array of image pixels can be mapped to a systolic element for lower frame rates. Although off-the-self general purpose processors provide an alternative for the processing core, they do not offer efficient operation and scalability of a GT-VISTA node.

#### 4.5 GT-VISTA VLSI implementation

A completed VLSI design of a GT-VISTA element includes an ALU and barrel-shifter. Figure 10 illustrates the ALU and barrel-shifter units in a 2252 $\mu$ m x 2222 $\mu$ m chip fabricated through the MOSIS foundry in 2.0 $\mu$ m CMOS technology. The functional units operate at a maximum of 21MHz. Table 1 shows the component unit areas (in  $\lambda$  units) and the simulated maximum frequencies of operation. Figure 11 illustrates the VLSI layout for a GT-VISTA systolic processor chip. The functionality of the prototype chip is a limited version of the systolic node shown in Figure 6, but the node is a proof of concept chip, and will be used as a basis of comparison for future systolic designs. The chip will be submitted to the MOSIS foundry (2.0 $\mu$ m CMOS) for fabrication. With optoelectronic devices shown in Figure 7, this single node systolic processor will serve as a building block for larger system.

Table 1. Component unit area and maximum operational frequency

	Height ( $\lambda$ ) x Width ( $\lambda$ )	IRSIM Simulated Freq. (0.8 $\mu$ m Feature Size)
ALU	752 x 396	140 MHz
Barrel Shifter	1472 x 1059	105 MHz
GT-VISTA Node	1841x1797	100 MHz

The ALU design is built upon the Manchester carry chain adder [17]. It has the ability to perform ADD, SUB, AND, OR, and XOR operations. The barrel shifter is capable of both arithmetic and logic shift operations as well as right and left rotations. These operations are sufficient to perform most early vision algorithms efficiently. Both designs were optimized for area and not speed because systolic arrays favor multiple computing elements to obtain higher throughput.

#### 5. Conclusion and future work

In this paper, we described the union of systolic arrays and integrated optoelectronic interconnect for early image processing. Current research in systolic smart pixel architectures, including PAMSAC and GT-VISTA architectures, were introduced. PAMSAC is a special purpose system that demonstrates the realization of optical interconnection with integrated silicon

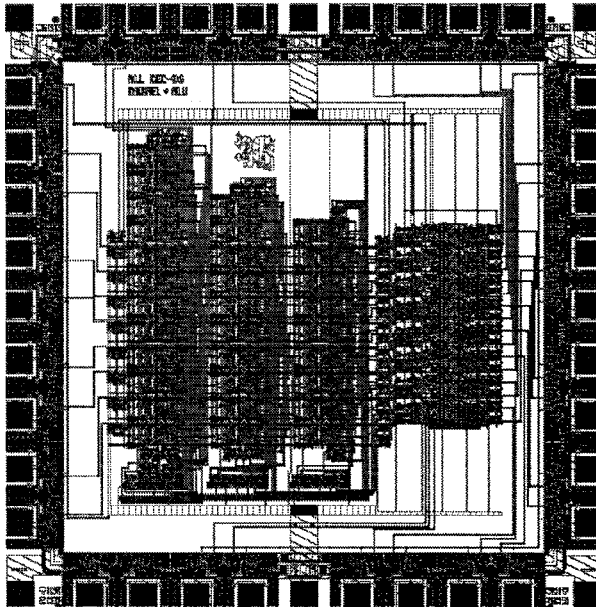


Figure 10. ALU and Barrel Shifter Chip

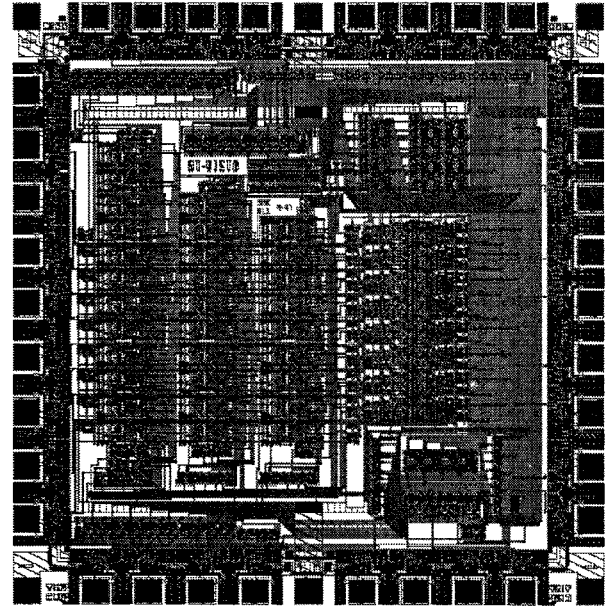


Figure 11. Prototype GT-VISTA VLSI Chip

detectors. PAMSAC advances systolic design methodology of modular cells and localized communications to promote efficient and high throughput operations.

GT-VISTA is a more flexible system for early vision processing, providing modest programmability for high throughput applications. A prototype system is being implemented, beginning with a series of VLSI chips, that have already been fabricated and evaluated, demonstrating both the architecture and optoelectronic interface functions. A complete GT-VISTA node will be submitted for fabrication later this year. Future goals of this project include developing additional programming methodologies for the system in the multi-layer topology and increasing the computational power efficiency and packaging density in a multi-node system.

## 6. References

- [1] J.Basille and S.Castan, "Multilevel Architectures for Image Processing", *Architectures and Algorithms for Digital Image Processing* (1985), v.595, p.46-52
- [2] G.A. Baxes, *Digital Image Processing: Principles and Applications*, John Wiley & Sons Inc., 1994
- [3] B.Buchanan, et. al., "High Density Focal Plane Signal Processing Using 3-D Vertical Interconnects", *Midwest Symposium on Circuits and Systems*, 1994, p.191-194
- [4] H.H.Cat, et. al., "SIMPil: An OE Integrated SIMD Architecture for Focal Plane Processing Applications," *MPPOI*, 1996, p.44-52.
- [5] E.S. Eid and E. Fossum, Real-time focal-plane array image processor, *Proc. of The International. Society for Optical Engineering*, 1989, v.1197, p.2-12
- [6] W.C. Fang, et. al., "VLSI focal-plane array processor for morphological image processing" *Proceedings of the Fifth Annual IEEE International ASIC Conference and Exhibit*, p.423-6, September 1992.
- [7] S.M. Fike, et. al., "8\*8 array of thin-film photodetectors vertically electrically interconnected to silicon circuitry," *IEEE Photonics Tech. Let.*, v.7 n.10, p.1168-70, Oct. 1995
- [8] T.J. Fountain, "The Use of Linear Arrays for Image Processing", *International Conference on Systolic Arrays*, May 1988, p.183-192
- [9] R.Hughey and D.P.Lopresti, "Architecture of a Programmable Systolic Array", *International Conference on Systolic Arrays*, May 1988, p.41-49
- [10] K.T. Johnson, A.R.Hurson, B.Shirazi, "General-purpose systolic arrays," *Computer*, v.26, Nov 1993, p.20-31
- [11] N. M. Jokerst, et. al, "Communication Through Stacked Silicon Circuitry Using Integrated Thin Film InP-based Emitters and Detectors," *IEEE Photonics Tech. Let.*, v.7 n.9, p.1028-1030, Sept. 1995.
- [12] J.Jolion and A.Rosenfeld, *A Pyramid Framework for Early Vision*, Kluwer Academic Publishers, 1993.
- [13] H.T. Kung, "Why systolic architectures?," *Computer*, p.37-46, Jan.1982
- [14] W.S.Lacy, et. al., "A Fine-Grain, High Throughput Architecture Using Through-Wafer Optical Interconnect" *MPPOI*, p.27-36, April 1994
- [15] R.M.Lougheed and C.W.Swonger, "An Analysis of Computer Architectural Factors Contributing to Image Processor Capacity", *Architectures and Algorithms for Digital Image Processing* (1985), v.595, p.3-13
- [16] D. S. Wills, et. al., "Processing Architectures for Smart Pixel Systems", *IEEE Journal of Selected Topics in Quantum Electronics*, v.2 n.1, April 1996, p.24-34
- [17] N.H.E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*, Addison-Wesley Pub. Co., 2<sup>nd</sup> ed., 1993