Silicon VLSI Processing Architectures
Incorporating Integrated Optoelectronic Devices

Huy H. Cao, Myunghee Lee, Brent Buchanan,
D. Scott Wills, Martin Brooke, and Nan Marie Jokerst
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0250

Abstract
Integrated optoelectronic interconnect offers a potentially lower cost, higher density alternative to wire-based technologies for I/O and inter-chip communication. This paper outlines two systems being designed at Georgia Tech which incorporate integrated thin film optoelectronic devices onto high throughput VLSI digital processors. The first system places an array of thin film detectors on top of SIMD processing elements allowing direct area connections between sensors and processors. This allows extremely fast frame processing rates (1-10 thousand frames per second) which are required in high speed and scanned imaging systems. The second system presented incorporates inter-chip IR optoelectronic channels which pass transparently through silicon. These links allow communication between three dimensionally stacked chips supporting high throughput interconnect topologies. This paper demonstrates the potential of optoelectronics integrated VLSI systems for providing extremely dense and lightweight solutions in applications such as image processing.

1. Introduction

Wire-based interconnect technologies have been a key enabler in the widespread use of VLSI circuits over the last 25 years. For most applications, low cost IC packages provide an effective means of interconnecting integrated circuits in a system. However, some applications require higher levels of off-chip I/O bandwidth than can be provided using perimeter bonded ICs. Area bonding techniques (e.g., flip-chip bump bonding) provides greater off-chip I/O, but often complicate the interconnection substrate, requiring more costly fine line technologies. Other applications require interconnection topologies that are not well accommodated by the traditionally planar wire-based interconnects.

This paper presents two processing systems which incorporate optoelectronics. In the first system, an array of GaAs thin film detectors is placed on top of (and electrically bonded to) digital SIMD processors to provide high performance, monolithic focal plane processing system. This processing approach provides greater I/O bandwidth between detectors and processors. It also allows real-time interaction between sensors and processors (e.g., for sensitivity adjustment, etc.) that is not possible with non-integrated systems. While integrated Si detectors have been demonstrated in VLSI processing systems [9], this technique allows Si processing circuitry to be placed underneath the GaAs detector area, providing greater fill factor for better image coverage and more efficient detectors. This integrated approach can lead to extremely compact image processing systems.

In the second system, through-wafer optoelectronic channels provide chip to chip interconnect for an extremely dense, high throughput, three dimensional parallel processing system. Thin film InGaAsP devices, which operate at a wavelength to which silicon is
transparent, are integrated on top of standard foundry Si ICs, so that stacked processor chips communicate vertically. High speed analog interface circuitry on the Si IC provides a high bandwidth link between the devices and the digital processing circuitry. The high bandwidth provided by this optoelectronic interconnect reduces the need for local memory (since more data is provided by I/O), allowing a more productive use of the Si resource for additional processing.

In this paper, the design plans and status of these two systems are presented. The details of several critical system elements are also described, along with an overview of thin-film device integration process.

2. The SIMD Pixel Processor

Current CCD arrays provide low cost, low to medium resolution image capture devices. However, these devices are difficult to integrate with digital VLSI processing, and the charge-coupled devices consume most of the available circuit area. Since detectors and processors are packaged separately, their communication bandwidth is restricted. While sampling at 60 Hz for modest resolutions and pixel levels is possible, high frame rate and scanned systems are not possible. This bandwidth limitation also complicates "active detectors", where a closely coupled processor continually adjusts a detector’s sensitivity, noise rejection, etc.

The SIMD Pixel Processor (SPP) system uses integrated optoelectronics for on-chip conversion and delivery of optical image data to digital processors. This system incorporates an array of GaAs thin-film P-i-N detectors integrated on top of Si VLSI-based SIMD (single instruction stream, multiple data streams) processors. This compact, monolithic system operates at the image focal-plane providing high-performance image processing. The SIMD processors provide greater processing power and programmability than "smart-pixel" systems proposed in the optical computer field. This section describes the SPP processor and the Si-based interface circuits. It also motivates the thin-film integration approach.

2.1 Processor Architecture

The digital component of the SPP consists of an array of SIMD processors. Figure 1 illustrates a block diagram of one SPP node.

![Block diagram of a SIMD Pixel Processor node](image)

Figure 1: Block diagram of a SIMD Pixel Processor node
The node includes a traditional processor datapath plus additional units for interfacing with the thin film detector array. The first implementation of the node includes an 8 bit datapath with an arithmetical, logical, shift unit, and a 16 bit multiply-accumulator (MACC) used in many image processing applications. These functional units access an eight word register file. Each node has 64 words of local memory. (Up to 256 words can be addressed in the instruction set.) SPP nodes communicate through a nearest neighbor NEWS (north, east, west, and south) network using special registers in the datapath.

Each SPP node interfaces to a sub-array of thin film detectors in the focal plane array. In the current SPP implementation, a processor addresses 16 detectors. Future SPP node implementations can address up to 256 detectors. Each processor includes circuitry to convert light intensities at the detectors into digital values for processing. The instruction set architecture provides a SAMPLE instruction to synchronously capture light intensities at each detector of the focal plane array. The SIMD execution model allows the entire image to be sampled synchronously. Figure 2 shows an SPP assembly code fragment that uses the SAMPLE and PLOAD instructions to sample and load detector values to the register file.

```
sample       ;; sample the detectors
pload r0, 0   ;; Loads r0 with detector 0
pload r1, 1   ;; Loads r1 with detector 1
pload r2, 2   ;; Loads r2 with detector 2
pload r3, 3   ;; Loads r3 with detector 3
```

Figure 2: SPP Assembly Code for Sampling Detector Data

Once the detector array data has been digitized, it can be processed on the SIMD nodes in a data parallel fashion. Figure 3 shows assembly code for a two point FFT butterfly.

```
FFT ;; perform FFT butterfly on x[0] and x[1] that
     ;; are stored in local memory
load r7, 0   ;;
load r0, r7   ;; real(x[0])
load r5, 2   ;;
load r1, r6   ;; real(x[2])
add r2, r0, r1 ;; real(x[0]) + real(x[2])
sub r3, r0, r1 ;; real(x[0]) - real(x[2])
store r7, r2   ;; mem[0] = sum
store r6, r3   ;; mem[2] = difference
```

Figure 3: SPP Assembly Code for an FFT Butterfly

An assembler and simulator for this system has been developed and several image processing applications are being implemented (e.g., image compression, filtering, jitter removal). The local memory, register file and logical shift unit, and MACC unit have been submitted to MOSIS as test chips. The complete processor is being implemented on a 1.2 μm tiny chip. An instruction unit is being designed for a host computer (an IBM PC).

### 2.2 Detector Interface Circuitry

The silicon interface circuitry must be small, fast, and sensitive in order to take full advantage of the thin film devices. A block diagram of the interface is shown in Figure 4.
Current through the reverse biased GaAs P-i-N detector is used to control a current controlled oscillator (ICO). A digital counter measures the period of the oscillation to provide a digital representation of the light intensity. In the first SPP implementation, eight bits values will be produced.

Figure 4: Detector Interface Diagram

The core of the oscillator is a loop of three digital inverters, shown in Figure 5, with one of the inverters current limited through current mirrors controlled by the photodetector. The photodetector current thus regulates the switching speed of this inverter stage, which in turn regulates the frequency of the oscillator. Additionally, the current limited inverter is also slowed by the capacitor tied between its input and output. The output of each circuit in the array is also buffered and driven off chip.

The bi-level output signal of this circuit provides high noise immunity by encoding the signal information in pulse width rather than amplitude. High dynamic range is inherent in this data format also, since the limits on signal range can span many orders of magnitude. The lower limit of oscillation frequency is primarily set by the amount of time that the subsequent processing layers are capable of waiting. The upper limit is set by characteristics intrinsic to the fabrication technology. The layout dimensions of this circuit (λ = 1.0) are 58 μm x 48 μm.

A 4 x 4 array of silicon ICOs, integrated with GaAs P-i-N photodetectors has been implemented and tested [1].

2.3 System Integration

The focal plane processing approach to optical interconnect dispenses with the need to electrically convey input matrices to integrated processing circuitry by incorporating photosensitive devices on the same substrate as the processing circuitry. The photodetectors provide I/O data to processors underneath. Optical interconnect technology is ideal for image processing tasks, since it can be used for sampling incident images in real time and in parallel.

The techniques developed to accomplish this primary fall into one of two categories based on the construction of the photodetectors. In the case of photosensitive devices inherent to the circuit's fabrication process [4][9], there are several substantial disadvantages. Primary among them is the processing circuitry's loss of valuable silicon real estate to the photodetectors (the fill factor), which necessarily decreases the complexity of the processing operations as the density of optical channels increases. A complimentary problem is that as the proportion of area sampled by the photodetectors decreases, an increasing amount of the radiant energy available in the optical signal falls on the processing circuits rather than on the photodetectors. Also, fabrication process optimization rarely focuses on the photosensitive devices, leaving the designer few choices relative to their operation.
Three dimensional integration techniques overcome the fill factor problem by removing the photodetectors from the processing plane, placing them on a separate plane above the processing circuitry, and using vertical contacts between the two layers. These techniques are further split into two categories: monolithic and hybrid.

In the monolithic approach [4], GaAs photosensitive devices are grown directly on top of silicon circuits. This requires subjecting the silicon circuitry to degrading conditions, such as high temperature, and may not permit the growth of high quality lattice matched photodetectors.

Hybrid techniques separate the fabrication paths of the two components, allowing for the independent optimization of the fabrication processes of both circuits and photodetectors, after which the photodetectors are integrated directly on top of the silicon circuits.

Using epitaxial lift-off technology (ELO) refined here at Georgia Tech [3], focal plane processors can utilize direct connections between the photodetector and processing circuitry layers. This allows for a high fill factor without detrimentally affecting the area available for signal processing circuitry or inefficient detection of radiant energy in the incident signals, and for the independent optimization of the circuit and photosensitive devices through separate growth processes.

Figure 6 illustrates an Si IC containing nine SIMD processors. This chip can be fabricated by conventional Si VLSI foundries with no special processing. In a separate process, the P-i-N structure of the detectors is grown in GaAs-based material. The thin-film material is then removed from the host substrate through the use of a sacrificial layer. Once removed, the bottom surface of the structure can be accessed for additional processing. The material is then bonded to the top of the Si chip, and the thin film material is pixelated. Finally, electrical contact is made to the top of the detectors. Figure 7 shows the IC after an 8 x 8 array of detectors has been attached. The chip is placed at the focal plane of a projected image.

![Figure 6: Nine SPP nodes on a chip](image)

![Figure 7: Detectors integrated on top of chip](image)

2.4 System Status

Detectors arrays as large as eight by eight have been fabricated and tested. Larger arrays are planned. The detector interface circuit has been fabricated and tested with a four x four detectors array [1]. The major elements of the SPP architecture have been submitted for fabrication and a one node chip is being implemented.

The first prototype SPP system will include a single processing node and a 4 x 4 array of detectors. Future SPP implementations will include up to 64 nodes on a chip, with up to 16 thousand detectors. Both high frame rate, low resolution, and high resolution scanned applications are being explored.
3. Through-Wafer Optoelectronic Interconnection Network

In this second system, optoelectronic interconnect is employed in a general purpose parallel system for chip to chip communication. This section motivates three dimensional optoelectronic interconnection networks and presents one architecture to realize such a network. The circuits used to interface to the optoelectronic devices are also described.

3.1 Approaches to Three Dimensional Interconnects

There are many approaches to implementing a three dimensional interconnection network depending on the network size, system density, and node bandwidth requirements. For low to moderate throughput (0.1 - 2 Gbits/sec off-chip bandwidth), low density systems (10,000-100,000 in²), wire meshes (J-Machine [11]), rings (KSR-1), and hypercubes (NCUBE 2) are appropriate. Portable image processing systems demand high network throughput (5 - 5 Gbits/second off-chip bandwidth) in a high density system (< 100 in²). Few wire-based approaches are being explored for systems of this type (e.g., thermomigrated feedthroughs and microbridge interconnects [6],[10]) and the reliability and manufacturability costs of these technologies are of concern.

Researchers are also exploring optoelectronic interconnect technologies, including both guided-wave and free-space interconnects. Guided-wave approaches are most applicable to planar interconnect applications (e.g., MCMs, wafer-scale integration), and are not well suited to three-dimensional interconnects. Free-space approaches are better suited to three-dimensional connections. Often, microlenses and holograms are used to establish fixed or varied interconnection patterns between processing planes. Yet the alignment sensitivity, coherence demands, size, and complexity of these systems present serious obstacles to their application in high density, low cost interconnects.

The ideal solution would incorporate the low cost, mass-produced manufacturability of traditional Si VLSI. Silicon continues to be the material of choice for digital integrated circuits. However, because of its poor optoelectronic properties, efforts to incorporate integrated optical emitters and detectors have met with limited success. The most effective optoelectronic devices have been demonstrated using GaAs and InP. However, GaAs and InP have been unable to match Si in density or cost for VLSI applications.

3.2 Network Architecture

A new technique for combining GaAs and InP-based optoelectronic devices with Si VLSI has been developed at the Microelectronics Research Center at Georgia Tech [2],[3]. Using epitaxial liftoff, small (100μm x 100μm) GaAs and InP-based light emitting diodes and detector have successfully been separated from their lattice growth substrates. The deposition of these thin film devices onto a silicon host containing circuitry has also been demonstrated. Operating wavelengths for InGaAs and InGaAsP are selected for which Si is transparent, allowing through-wafer transmission. This technique, which has been demonstrated for simple optoelectronic interconnect, offers the promise of low cost and mass producibility. In moderate volume, inexpensive optical devices could be integrated en masse, on Si circuits, using procedures no more complex than existing VLSI processing steps.

This technology can be exploited to form an extremely simple three-dimensional optoelectronic interconnect. Processing elements are designed and fabricated using existing Si VLSI techniques. Bonding pads are provided only for power contacts. (Probe pads are
provided for testing). Holes in the overglass allow GaAs and InP-based emitters and detectors to be attached in a post-processing step. Silicon-based driver circuits connect parallel buses from routers on the chip to the faster optical devices (0.1-1.6 Gbit/second). After the Si chips containing the optical devices are fabricated and tested, they are attached to a larger silicon substrate. This provides physical support and also participates in power distribution and cooling. A fully populated substrate forms one plane of the system. Completed substrates are then stacked to create the third dimension, which are optically connected. Note that the optical interconnections between these layers are not affected by the flow of a transparent gaseous or liquid coolant. To facilitate manufacturing, all chips and substrates are identical.

System I/O is provided by I/O layers which optically inject data directly into the network on the top and bottom surfaces of the stack. These layers can interface to high-bandwidth electrical or optical connections. Alternatively, the input surface can be covered by visible light detectors for direct focal-plane image input. The output surface can incorporate a frame buffer for direct video output. The sides of the stack are used for power distribution and cooling connections. Due to the dense packaging, liquid cooling between layers is anticipated. Optical clock distribution is also incorporated into this system. In-line detectors on each layer detect broadcast clock signals from the bottom of the stack via the through-wafer transmission technique employed by the network. A more complete description of this network, with simulated performance is described elsewhere [14].

3.3 Prototype Interface Circuitry

Interfacing to these relatively fast optoelectronic devices requires specially designed drivers and amplifiers. This section describes analog interface circuits that have been designed for this purpose. It also presents a more advanced design that has been tested at high speeds.

There are several technology alternatives for achieving high speed operation. High speed optical receiver amplifiers for telecommunication applications typically use technologies such as Si-bipolar [12] or a compound semiconductor [13]. However, low-cost, high-complexity silicon CMOS circuitry is more attractive in many cases. A digital CMOS approach gives some advantages over other processes. CMOS is low power so that it can be used at multiple detector sensor array. It also allows putting digital control circuitry on same substrate.

A silicon CMOS detector amplifier is shown in Figure 8. It senses the current from the detector and changes the voltage at the output node accordingly. Current from the photodetector decreases the influence of \( I_{bias1} \) at the output node, allowing \( I_{bias2} \) to pull the voltage low. In the absence of current from the photodetector, \( I_{bias1} \) as mirrored through the p-devices instead pulls the output voltage high. The bias currents are thus constrained to:

\[
(I_{bias1} - I_{det}) < I_{bias2} < I_{bias1}
\]

The detector amplifier also maintains a minimum bias voltage across the photodetector. This is necessary because of the photodiode’s high reverse bias leakage current. It is accomplished by sensing the current through the small (3/2) n-type transistors, which is fed back as a corrective voltage.
The silicon CMOS emitter driver, shown in Figure 9, consists of two analog inverter stages. Current in the first stage is set by \( I_{bias} \) and modulated by the digital signal \( V_{in} \). This current is then mirrored to the second stage and amplified by a factor of ten. The second stage’s output mirror amplifies this current by a factor of ten as well, for a total current gain of 10, which then drives the thin-film light emitting diode directly.

These circuits have been fabricated through MOSIS, and integrated with thin film devices. There are currently being characterized. Unfortunately, these designs are too slow for the network being designed.

A high speed (125 Mbps) transimpedance amplifier for an optical receiver using standard digital CMOS process has been demonstrated [8]. The amplifier was designed, fabricated through MOSIS 1.2 \( \mu \)m digital CMOS process, and tested. This circuit converts the minimum signal current expected from a discrete commercial 1500nm wavelength InGaAsP photodetector to a voltage signal large enough to be converted to digital by a high speed digital comparator circuit.

In order to obtain such an operating speed, design optimization is done using a multi-stage, low-gain per stage design approach. Assuming certain fixed gain and power dissipation limits, an open-loop approach provides better upper 3 dB frequency rather than a feedback circuit. As the number of stages increases the gain required and power available per stage decreases, the bandwidth per stage increases resulting in much higher overall amplifier bandwidth. On the other hand, if too many stages are cascaded, the bandwidth decreases due to the accumulated effect of the parasitic capacitance associated with each amplifier stage. From the analysis as shown in Figure 10, an amplifier with around five stages would be the optimal design for this application.
Figure 10: Frequency response versus number of stages

Figure 11(a) shows a single stage of amplifier. A current mode design approach is employed. Each stage has current gain of 3, which is mainly decided by the gate width ratio of transistors, N1 and N4a, in order to minimize the effect of device mismatch. Since the input capacitance of the stage has significant effect on the overall bandwidth of the amplifier, a buffered input is used to reduce the input capacitance. Without the buffer (transistor N2), the input capacitance would be:

\[ C_{\text{TOTAL}} = 4 \cdot C_{g} + 2 \cdot C_{db} \]

\( C_{g} \) is the gate capacitance and \( C_{db} \) the drain-body capacitance of a transistor. Using the buffer, \( C_{\text{TOTAL}} \) can be reduced to:

\[ C_{\text{TOTAL}} = C_{g} + 2 \cdot C_{db} \]

Since the input resistance of the circuit is \( \frac{1}{g_{m}} \), the bandwidth per stage is:

\[ BW = \frac{R_{o}}{2 \pi \cdot C_{g} + 8C_{db}} \]

Figure 11(b) shows overall circuit diagram of the transimpedance amplifier. A total of five stages are used for the whole amplifier. Therefore, total transimpedance gain is \( 3^5 \times 50 = 12,150 \) \( \Omega \) when a 50 \( \Omega \) load resistor is used. Current gain is achieved only if the input resistance of each stage is significantly lower than the output resistance. To get a high output impedance at each stage, a cascoded output stage (transistor N4) is used. Offset adjustment and bias circuitry are added to the 5-stage amplifier.
Figure 11: (a) Schematic of one stage amplifier. (b) Complete schematic of amplifier.

For layout of the circuit, splitting the power supply rails helps reduce parasitic feedback, which usually causes oscillation. For this amplifier, the power supplies are separated into two halves. The first half serves only those parts of the circuit that amplify small signals, while the second serves the large signal and output parts of the circuit. This prevents the larger output signal swings from generating small feedback signals in the sensitive small signal parts of the circuits. Another aspect of the layout that reduces unwanted coupling into the input signal is the long, thin left-to-right geometry of the layout. Small input signals enter on the far left, while the output signal exits on the far right. This maximizes the separation of the sensitive input stages from the larger signal output stages. Bonding pads were made as small as possible in order to minimize the pad capacitance.

The circuit operates at up to 100 MHz. The uniformity of the output amplitude at several different speeds indicates that the amplifier is operating well within specification for bandwidth. The output signal was developed across a 50Ω load. Approximate measurement of the power dissipation is about 20 mW at speed of 125 Mbps [7].
This design achieves the desired speed, but is still relatively large. New, more compact emitter drivers and detector amplifiers are being designed.

4. Conclusions

This paper has presented two projects which incorporate integrated optoelectronic devices on Si VLSI. While this technique will never replace wire-based interconnects for many applications, it does offer new design alternatives for systems which require high I/O and three dimensional interconnect topologies.

5. References


