National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

VLSI Implementation of Receptive Fields for Peripheral Vision Processes

by

Vivian John Ward
B.A.Sc., Simon Fraser University, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

in the School
of
Engineering Science

Simon Fraser University
October, 1993

ISBN 0-612-01089-9

Approval

NAME: Vivian John Ward

DEGREE: Master of Applied Science (Engineering Science)

TITLE OF THESIS: VLSI Implementation of Receptive Fields for Peripheral Vision Processes

EXAMINING COMMITTEE:

Chairman: Dr. John Jones

Dr. Marek Syrzycki
Senior Supervisor

Dr. Glenn Chapman
Supervisor

Dr. Andrew Rawicz
Examiner

DATE APPROVED: November 8, 1993

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

"VLSI Implementation of Receptive Fields of Peripheral Vision Processes"

Author:

_____
(signature)

Vivian J. Ward
(name)

November 2, 1993
(date)

Abstract

Focal-plane processing is an approach to image processing that incorporates local computation circuitry within an image sensor. This allows the large data bandwidth of visual systems to be dealt with in a parallel fashion, using many simple and efficient processors.

This thesis investigates the feasibility of constructing a hardware-based smart vision sensor that will implement various receptive field processes used for software-based image processors. Simple prototype sensors implementing custom receptive field functions that use very-low-power current-mode analog processing circuitry are examined. These devices use specially-designed multi-mode photodetectors that are capable of reducing over seven decades of illumination intensity to three decades of output current. The sensors use on-chip current-controlled-oscillators to convert the output signal of each receptive field to a quasi-digital format, eliminating the need for on-chip A/D converters, and providing a high degree of noise immunity. The components and sensors described in this thesis were fabricated in a 3 μm digital CMOS process, and evaluated with an optical test bench designed for that purpose. The measured devices were found to operate acceptably.

A structure for microelectronic receptive fields is proposed, and the general approach of focal-plane processing is discussed. This thesis also proposes a framework for constructing an intelligent vision sensor for peripheral vision, and postulates possible technological solutions.

Table of Contents

List of Figures

1. Introduction

1.1 Data Flow in the Visual System- A Concept Model

Vision provides a method by which we interpret our environment. The task performed by our visual system is primarily that of data reduction. From the 100 million or so photoreceptors in the retina to the 1 million nerve fibers in the optic nerve to the understanding that we are observing the face of a friend comprises an extraordinary feat of data reduction. Each stage of the process extracts meaning from the data presented to it, thereby reducing the volume of data passed to the next stage. Without this simplification, higher level processes would face an overwhelming task. It is this process of extracting useful information from the massive amounts of available data that allows us to interact with our environment.

This section describes a model of the visual system based upon data collected from biological systems. This model will represent the data path leading from objects in our environment to our understanding and perception of them.

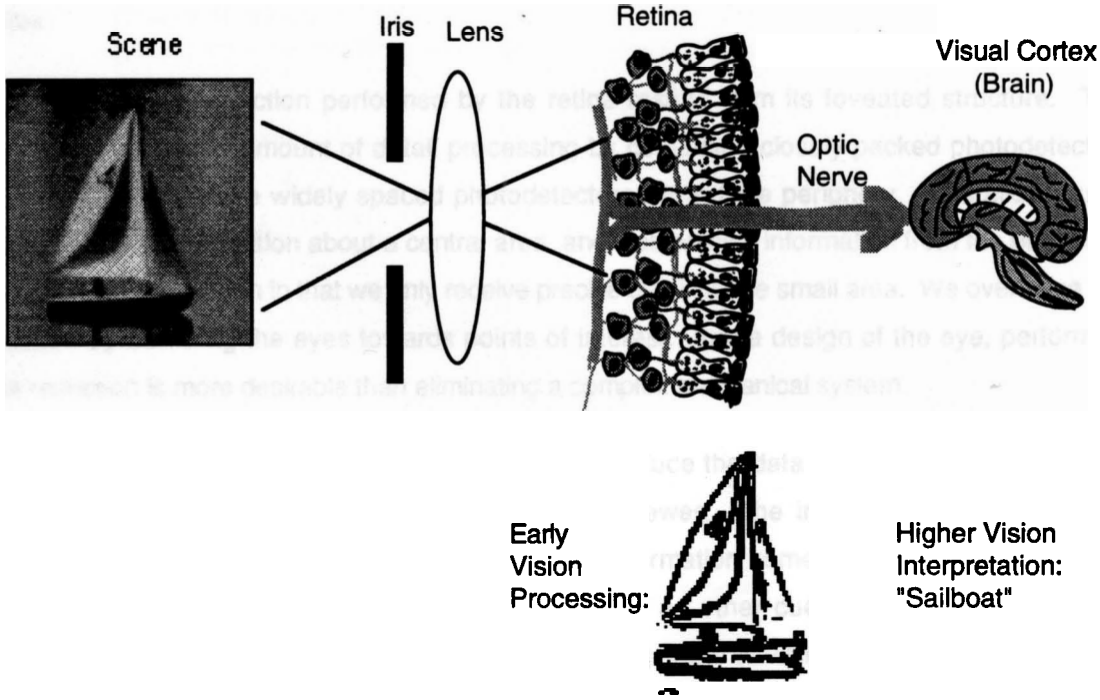

Figure 1-1: Data flow diagram of the visual system

The scene perceived by a viewer is made up from the light reflected or emanating from a set of objects. In the real world, such a scene is generally dynamic, containing either moving objects or changing light intensities. For humans, scenes are never entirely static, as the eye is constantly moving, scanning the interesting features in view. Spatial changes in light intensity and variations

in colour or wavelength also contain information about a scene. Scenes in our environment can also project a broad range of lighting conditions, and much of the specialization of our vision system is adapted towards dealing with night time lighting situations.

The light from the scene passes through the iris and is focused by the lens onto the retina. The iris helps the retina adjust to rapid changes in scene intensity by reacting quickly to restrict the amount of incident light. This mechanism is effective, but can only compensate for approximately one decade of light intensity. A slower reaction that can accommodate larger intensity shifts occurs in the eye's photoreceptors. In addition, the retina has two basic types of photoreceptors, each optimized for a different range of light intensity.

Early vision processes take place in the retina and perform such functions as edge enhancement and motion detection, local gain compensation, as well as feature extraction [19,20]. The data received by the photodetectors consists of the highest bandwidth of the entire vision system. The purpose of early vision processing is to reduce the data bandwidth passed on to the next stage of processing. By extracting relevant features and data from the massive amount of data collected by the rods and cones, the retina greatly simplifies the problem presented to the visual cortex.

Much of the data reduction performed by the retina results from its foveated structure. This structure reduces the amount of detail processing by containing closely packed photodetectors at the center, and more widely spaced photodetectors toward the periphery. This allows us to receive detailed information about a central area, and less precise information from the periphery. This presents a limitation in that we only receive precise detail in one small area. We overcome this limitation by swiveling the eyes towards points of interest. In the design of the eye, performing data reduction is more desirable than eliminating a complex mechanical system.

Higher vision processes occur in the brain to further reduce the data provided by the retina and produce a useful model of the objects around the viewer. The initial steps largely involve assembling depth information from the image. This information comes from the stereo disparity between the image seen by each of our eyes as well as other cues which determine object boundaries. These processes reveal the orientation and depth of visible surfaces in a viewpoint-centered frame of reference [16].

The final processes of vision create a 3D model of the objects in the environment. Rather than dealing with edges and surfaces, the data now represents volumes and objects. Information about the parts of objects not immediately in view is hypothesized. The result is data about objects and their orientations that allows us to interact with them.

In the context of electronic vision systems, the concept of foveated vision can be used to form two categories of vision sensor. The equivalent of the retina's fovea, where photoreceptors are most densely packed to obtain the greatest detail, is central vision. The highest photodetector density in electronic systems is obtained with a CCD array. This technology is well developed, and more than adequate for the task. This CCD array would be coupled to a conventional digital processor that would be sufficient for dealing with its relatively small area. The rest of the retina, or periphery, contains less densely packed photoreceptors, along with more processing and interconnection cells. The electronic equivalent of this region would be a large scale focal plane processor. This circuit would couple photodetectors closely with simple processors, and perform some low level vision operations in a parallel fashion. Together, the two regions would form a foveated smart vision sensor (FSVS).

A sample output of such a vision sensor is shown in figure 1-2. In this diagram, the output of the central region is unprocessed data, as it would be sent to an external processor. In this case, the periphery has been processed with an edge-detection algorithm, although more complex processes could be performed. This thesis will focus on the development of circuits and devices for such a sensor's periphery.

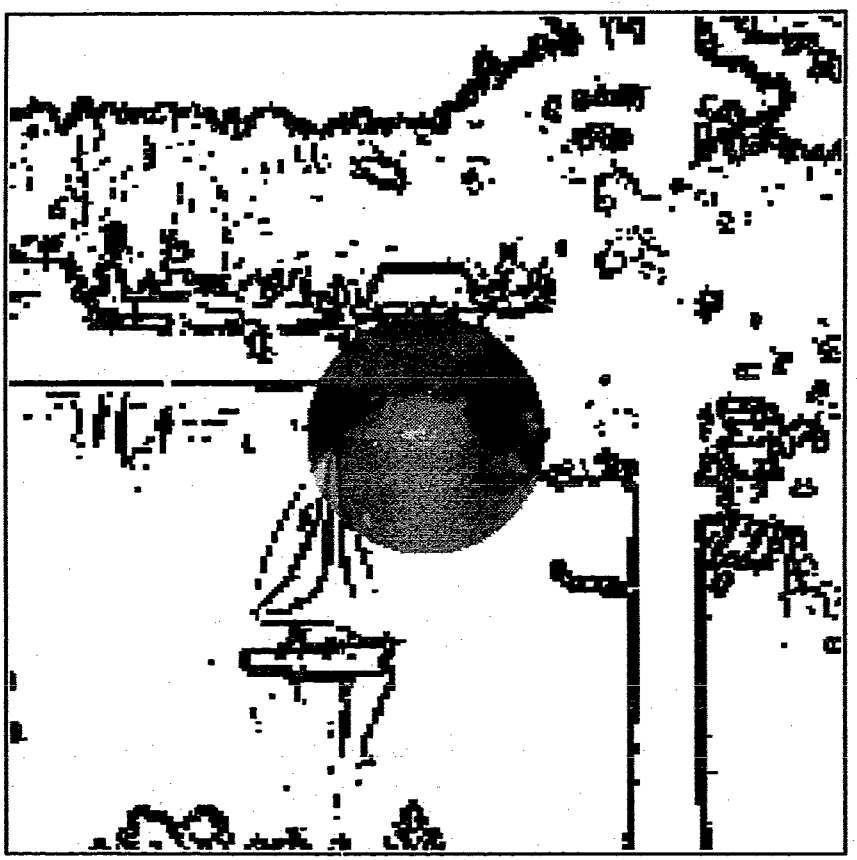

Figure 1-2: Simulated output of a smart vision sensor with central and peripheral regions

VLSI image sensors have evolved to the point where they are ideally suited for providing central vision functions. CCD arrays feature high density and wide dynamic range, and are almost perfect for such applications. For peripheral vision processing, devices are not as well developed. This section will critically review new approaches to Early Vision processing taken with a more intelligent view than conventional software processing. Each one will be examined in the context of its usefulness for peripheral vision functions.

Data reduction and feature extraction are necessary if artificial vision systems are to be used in autonomous applications. For systems to use visual data without human intervention, the information extracted from the data must be represented in a way that allows decisions to be made affecting the environment. These applications require that the data be reduced to representations in which the quantity of data is independent of image complexity [11].

The primary task of a vision system, whether biological or mechanical, is to reduce the data presented to it into a manageable size and form. There are two major approaches to performing this task with computer hardware. The first uses a conventional digital computer to implement processing algorithms on raster images collected from a CCD camera. This approach reduces the task to a software problem, and allows the greatest flexibility and precision. The second group of research involves the fabrication of specialized integrated circuits to perform specific sub-tasks in an analog fashion. Such circuits can feature photosensors tightly coupled to the processing circuitry. As many simple processors can handle the huge data bandwidth in a massively parallel fashion, this approach has the advantage of being much faster, allowing the digital computer to deal only with the reduced bandwidth presented by the specialized circuit.

Previous research done on analog VLSI Early Vision processing [9] has established its merits relative to those of conventional digital processing. The advantages of the analog VLSI approach are primarily its high speed, low power consumption, and the relatively small silicon area consumed. Of these factors, the high processing speed is the most appealing. Coupled with this approach are some disadvantages, which are principally limited accuracy, inflexibility and a lack of storage capability [9].

1.2.1. Image Plane Processors

Carver Mead has designed a silicon retina that uses a resistive network to perform local gain compensation [3]. Mead recognizes that the wiring necessary for computation is the limiting factor in both neural and electronic systems, and he attempts to work around the problem with shared

wiring, such as the resistive network. His designs use a regular hexagonal layout, which is not optimal for peripheral vision tasks, do not perform operations such as edge-detection or feature extraction, and use the subthreshold region of MOS transistors, which accentuates the effects of device mismatches.

Researchers at MIT are pursuing various directions in image-plane processors [9]. Their systems feature parallel computation, and analog circuitry for high throughput and low latency. They choose tasks and algorithms that require low or moderate precision and map naturally to physical processes in silicon. Their approach focuses comprise two subgroups, the clocked and unclocked systems.

The unclocked systems are generally the fastest and use the least power. The MIT group's systems use BJTs for sensors and FETs for processing. These designs are similar to those built by Carver Mead, except that they avoid using the MOSFET's subthreshold region to reduce the effects of device mismatch and attain greater speed. This approach emphasizes data reduction with resistive networks, which can be used to compute the location and orientation of a single object. The primary drawback of this approach is that it can deal only with simple images, in which information about a single object is required.

The clocked systems utilize CCD sensors and emphasize charge-domain signal processing. They have the drawback of requiring 10 to 25 different clock signals, which must be produced off-chip and routed through the sensor. These signals consume power, add to chip area, and increase the chip's complexity. The researchers report a speed advantage of approximately 1.5 orders of magnitude over special-purpose digital systems, although this changes from example to example. Specifically, they have fabricated a CMOS Imager and processor using a 2D resistive grid to find an object's position and orientation at 5000 frames per second with power consumption of 30 mW, and a CMOS/CCD imager and processor performing clocked image smoothing. These results demonstrate the potential benefits of this approach, but require prohibitively large amounts of silicon area for signal routing [9].

Deweerth [11] has fabricated an analog chip that gives analog X and Y voltages corresponding to the brightest point on the sensor's surface. The chip contains 160x160 pixels in 2μm BiCMOS, and uses BJT differential pairs to reduce the transistor mismatch effects common with MOSFETs. This design is not an image sensor, but does use innovative methods of coupling photodetectors and analog circuitry. This chip shows how massive data reduction can be obtained from analog circuitry, although the design is not suited to images containing more than one point of interest.

1.2.2. Innovative Digital Image Processors

Researchers at SRI International [10] have fabricated a chip that constructs a Gaussian or Laplacian pyramid structure from a 512x 480 image at 44 frames per second. Applications of this device include image compression, image enhancement, motion analysis, stereo, and object recognition. This device is a dedicated digital processor that does not contain sensors. Instead, it takes an already digitized image as input. This approach represents an avenue of competition with analog vision processors. Using specialized digital hardware, the performance of analog processors can be approached, but the silicon area required makes this method prohibitively expensive.

This approach might prove useful in a hybrid package in which a wafer-scale photodetector array is coupled to specialized digital processors. The digital processors could be contained on discrete dies, bonded to the wafer so as to accept signals from a local area. One of the primary disadvantages of the digital processor, its intolerance of fabrication defects, would be overcome by using only pre-tested devices in the final package. Such a sensor would require new packaging techniques, but might prove promising in the future.

1.3 A Smart Vision Sensor with Receptive Fields for Peripheral Vision

There are many applications for a smart vision sensor with central and peripheral regions. These include artificial intelligence research, in which an attempt is made to emulate human vision, automation, in which the quick recognition of objects is desirable, and remote sensing, in which the reduced bandwidth produced by the sensor can be transmitted more easily. Rather than concentrating on one area, a more general approach to the concept will be taken.

Figure 1-3 shows a possible structure of a fragment of a complete vision sensor utilizing MicroElectronic Receptive Fields [14]. The receptive fields comprising such a sensor would depend upon the intended application. For example, a vision sensor intended for use exclusively as an edge-detector would exclusively contain edge-detector receptive fields with varying sizes and orientations. A vision sensor intended for motion detection would contain motion detector receptive fields. However, to obtain a vision sensor suitable for general purpose applications, receptive fields should be present to detect many different manners of image primitives, such as lines, circles, and bars.

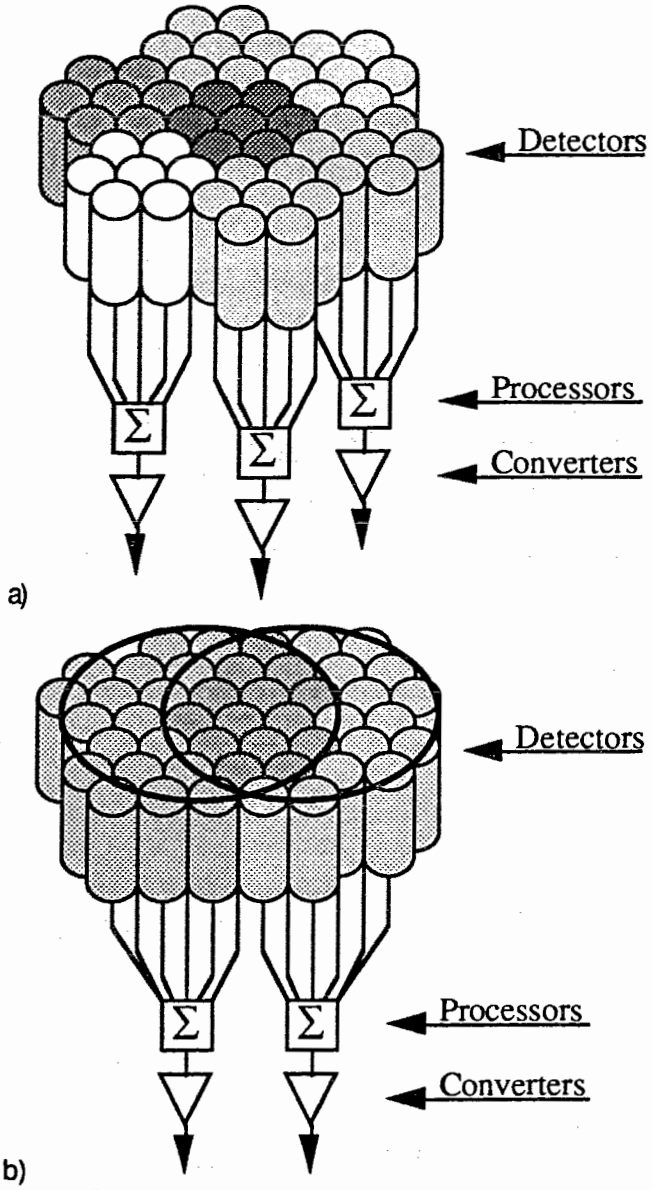

Figure 1-3: A small section of a focal-plane image processor comprised of (a) non-overlapping and (b) overlapping two-dimensional receptive fields

To extend the foveated retina structure to VLSI, the fovea would become a high-density array of photodetectors, likely with external processing. Focal-plane processors would implement the periphery as a lower-density set of receptive fields. This approach would incorporate the strongest points of each approach into a single sensor.

The primary competition to focal plane image processors comes from conventional CCD sensors coupled to off-chip digital processors. The primary advantages of the CCD approach are their high density, and the flexibility to re-program the digital processor for different applications. The disadvantages of the CCD approach are susceptibility to device defects, lower processing speed,

and increased power consumption. In order to retain the advantages of both CCD and MERF sensors, a hybrid design is necessary.

A practical implementation of such a vision sensor would require a hundred thousand or so receptive fields to perform peripheral vision functions. This would not be practical on a conventional VLSI die due to size constraints. To achieve a device of this complexity, it would be necessary to utilize different technologies. Two possible approaches are wafer-scale restructurable systems, and interchangeable silicon modules. The first approach would include redundant circuitry, and use a technique such as laser linking to disconnect defective functional units and connect operative ones [18]. The second approach would be a refinement of the multi-chip module (MCM) concept, and requires some technique for lateral interconnection between die-sized functional units. A drawback of this approach is in the blind spots which would be created along the edges between modules. Most of the lateral data flow would be within each silicon block, but some signals would have to flow between blocks.

1.4 Thesis Structure

The goal of this thesis is propose a hybrid vision sensor incorporating both the high resolution of a CCD array and the processing power of Microelectronic Receptive Fields (MERFs). The MERF concepts will be validated by building and testing circuits containing them. The functional blocks necessary for these circuits will be presented, along with sensors built from these blocks. Rules for the custom design of MERFs will be formulated, and used to help propose a framework for the construction of intelligent vision sensors for peripheral vision processing. The design of an optical test bench necessary for precise sensor measurements will also be described. This equipment will be used to characterize MERF circuits and sensors. From this data, a conclusion will be reached as to the viability of the wafer-scale sensor approach.

Chapter 1 introduced the vision system from a data flow perspective, proposed a model of vision that will guide sensor design, and described previous research that relates to this project. Chapter 2 describes MERFs in more detail, and gives a functional description of the building blocks that would make up the overall sensor. Chapter 3 covers the design and testing of the building blocks. Measurement results are presented for wide-range photodetectors, analog processing circuitry, and analog to digital converters. Chapter 4 presents the design of vision sensors based on these blocks. Chapter 5 describes the design of an optical test bench and the system necessary for vision sensor measurements. In addition, results of prototype small scale vision sensors are presented.

2. Design of Receptive Fields for Vision Signal Processing

This section will describe in more detail a vision sensor based upon the receptive field paradigm, list the tasks that it must perform, and propose the functional units that could be used to perform them.

2.1 The Receptive Field Model

The retina performs many vision processing operations in a massively parallel fashion. Operations such as edge enhancement, motion detection, image averaging and compression are all performed by several layers of cells below the rods and cones. The functionality of our retina is such that we even use it to judge the results of image processing algorithms implemented on machines.

The high performance of the retina is due in part to the interconnections between the photodetectors and surrounding cells, forming networks known as receptive fields (RF). Research on vertebrates has shown that such receptive fields give a signal in response to a particular type of stimulus such as an edge with a certain orientation, a moving object, a bar or a line [19,20]. A mathematical interpretation of these functions can also explain other phenomena, such as Mach bands, which are 'phantom' dark patches appearing between sharp dark edges as a result of local gain compensation mechanisms in the retina. The operations performed by these structures can be simulated mathematically with a computer.

In general, a weighted sum can represent operations performed by biological receptive fields. Each of the neighboring inputs is multiplied by a weighting factor, and the results are summed together, forming the output of the RF:

$$RF = \sum_{i=1}^{k} p_i w_i \qquad (2.1)$$

where the input comes from k photodetectors, $p_1 ... p_k$, each of which has an associated weight, $w_1 ... w_k$. A one- or two-dimensional array specifies the weighting factors and their spatial configuration. Figure 2-1 gives some sample arrays for different receptive field functions [14]. Various smoothing and sharpening operations, as well as detection of image primitives can be performed by selecting appropriate weighting functions.

$$\text{a)}\ \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} \quad \text{b)}\ \begin{bmatrix} & -1/6 & \\ -1/6 & & -1/6\\ & 1 & \\ -1/6 & & -1/6\\ & -1/6 & \end{bmatrix} \quad \text{c)}\ \begin{bmatrix}1 & 0 & -1\\ 2 & 0 & -2\\ 1 & 0 & -1\end{bmatrix}$$

Figure 2-1: Sample receptive field functions for averaging (a), sharpening (b), and vertical edge-detection (c).

These convolutions can be implemented in hardware to retrieve much of the simplicity and speed inherent in the original biological structure. Implementing this process as a software task on a conventional digital computer requires that the convolutions be computed in a sequential fashion. The use of a focal plane processor allows computational parallelism similar to that of the retina.

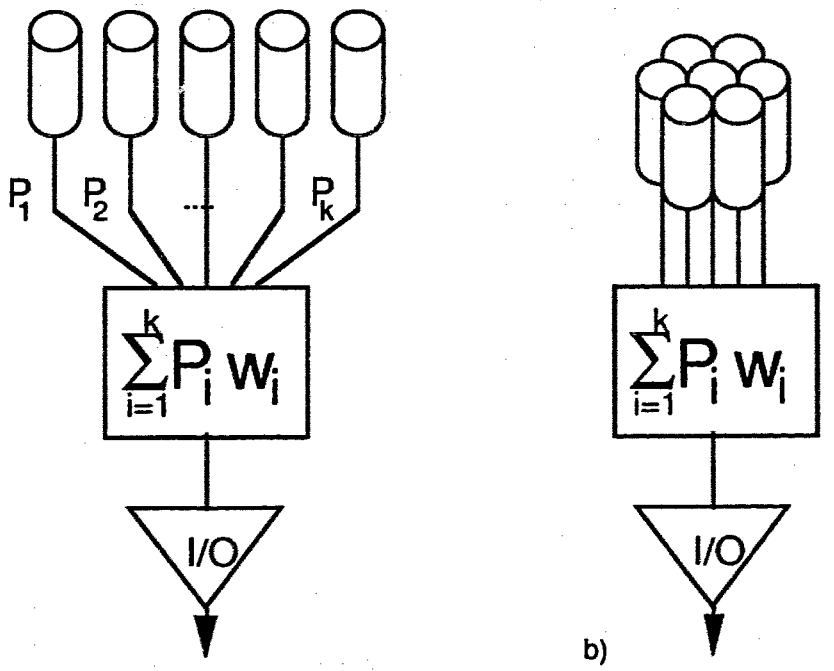

Figure 2-2:. Implementations of (a) a one-dimensional receptive field and (b) a two-dimensional receptive field

2.1.1. Receptive Field Algorithms

Successful candidate algorithms for receptive field implementation in either a biological or electronic system must share three characteristics. First, such algorithms must be physically realizable. This generally means that they must combine a small number of local inputs to reduce the physical interconnections that they contain. The second characteristic is that the algorithm extract information from the data. It must reduce the data bandwidth while retaining information useful for image interpretation. Finally, an algorithm should be resilient. It should not rely on precise spatial orientation or exact input weightings.

One well-known edge-detection algorithm used in digital image processing is Sobel's edge detection [13]. This consists of the convolutions shown in figure 2-3, in which different weightings are used to detect horizontal or vertical edges.

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Horizontal edge-detector

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Vertical edge-detector

Figure 2-3: Sobel's edge-detector for vertically and horizontally aligned edges

This algorithm fulfills the criteria for receptive field implementation. First, it requires that only six local inputs be used, allowing a very simple interconnection strategy. Secondly, this algorithm can reduce the data bandwidth while retaining useful information. Figure 2-5 demonstrates this fact, giving the sum of the vertical and horizontal algorithms as applied to figure 2-4. This procedure has extracted the significant edges of the objects contained in the scene, and they can be represented by a single bit per pixel, instead of the eight bits per pixel required for the original data.

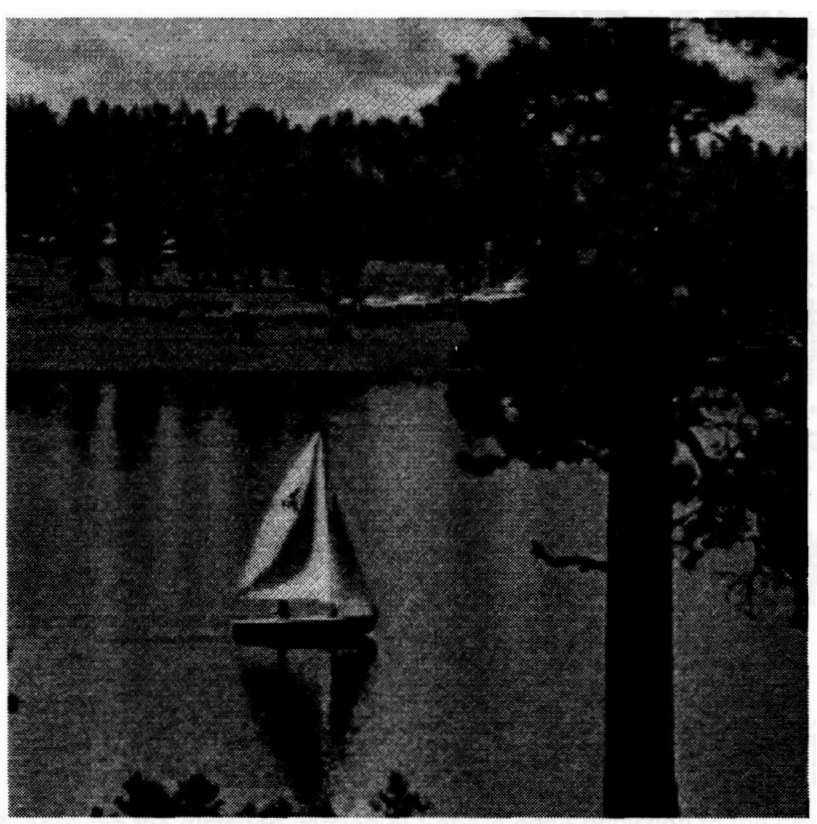

Figure 2-4: The sailboat scene

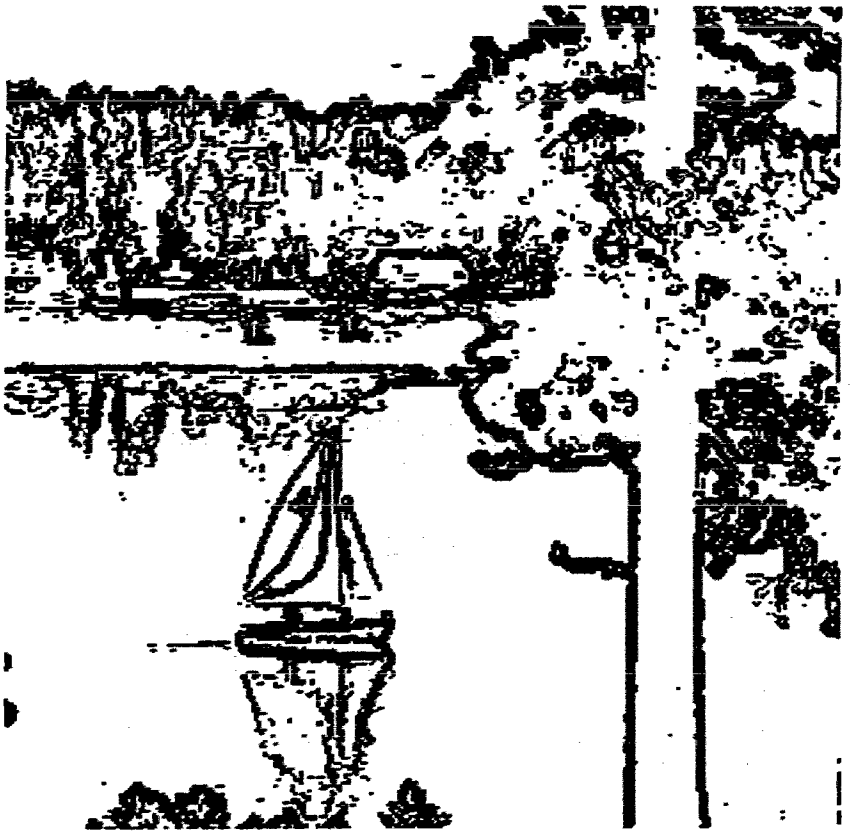

Figure 2-5: Applying Sobel's edge-detection to the sailboat scene

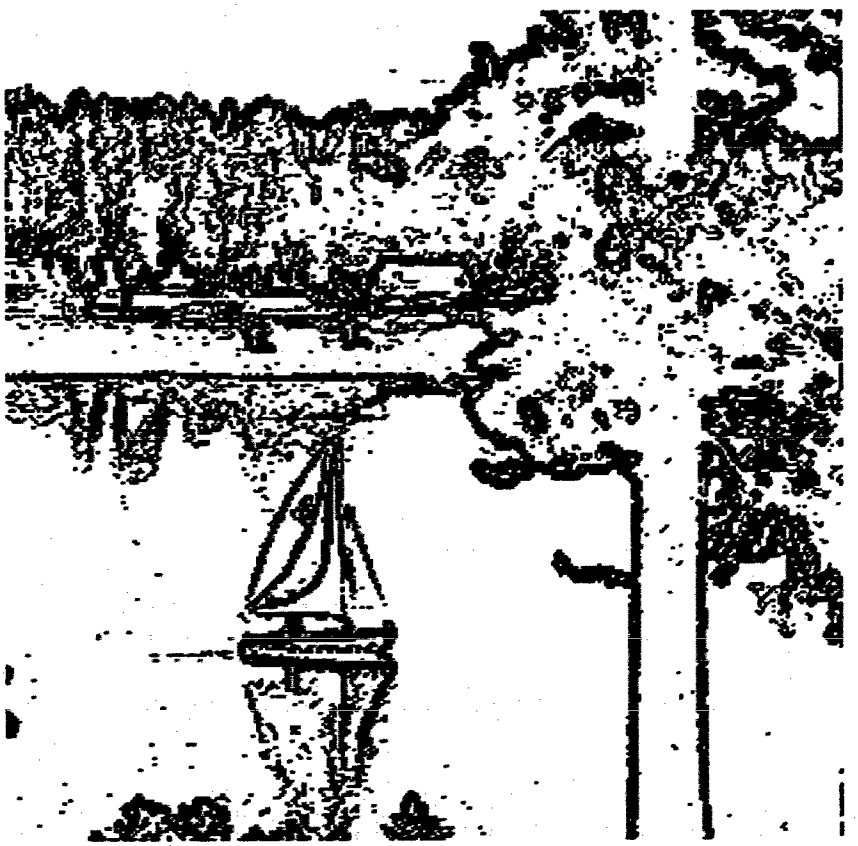

Figure 2-6: Applying the distorted version of Sobel's edge-detection to the sailboat scene

Finally, the application of a distorted version of the same weightings to the original scene demonstrates the resilience of the algorithm. The weightings given in figure 2-7 represent corrupted scaling factors and some non-functional inputs.

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 0 & -3 \\ 0 & 0 & -1 \end{bmatrix}$$

Horizontal edge-detector

$$\begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 0 \\ 0 & -3 & -1 \end{bmatrix}$$

Vertical edge-detector

Figure 2-7: Distorted version of Sobel's edge-detector used to demonstrate resiliency

Visual comparison of figures 2-5 and 2-6 shows that the distorted version of the edge-detection algorithm produces a result almost identical to that of the original algorithm. This demonstrates that Sobel's edge-detector is resilient to distortion.

Sobel's edge-detector is an example of an algorithm that fulfills the requirements for implementation as a receptive field.

2.2 VLSI Components of a Microelectronic Receptive Field

2.2.1. Photodetectors

Photodetectors are optical transducers that convert illumination into an analog electrical signal. The output signal is generally proportional to both the illumination intensity and the device area. The larger the active area of the photodetectors, the greater the output signal for a given illumination, so a larger photodetector can operate in dimmer conditions than a smaller one. Using larger detectors also presents drawbacks. The larger signal requires more power to produce and process, and the increased chip area dedicated to detectors reduces the area available for signal processing. As a human retina will function over a total range of almost 14 decades of light intensity, a photodetector producing an output proportional to illumination intensity over the same range will produce almost 14 decades of analog output signal. The analog signal processor must therefore be able to deal with a large input signal range.

2.2.2. Analog signal processor

The analog signal processor combines signals from neighboring photodetectors to give a single output signal. For the case of a receptive field-type convolution, this consists of performing a weighted sum. This processor must also be able to deal with several decades of input and output signal levels.

The signal from each photodetector can either be sent to a single MERF, or to a few neighboring MERFs. The optimal configuration depends upon the technology in which the sensor is implemented. If a photodetector requires more area than the circuitry necessary to duplicate its signal, it will be desirable to send the signal to the neighboring MERF cells. In the case where photodetectors require less area than the duplication circuitry, it will be more efficient to increase the number of photodetectors, and sent the signal from each to a single MERF. Analog signals can be processed extremely quickly, but such processing is generally less exact than equivalent digital processing.

2.2.3. A/D Converter

The output signal from the analog signal processor is in analog form. Consequently, it must be transformed to digital form before being used in a computer system. If this conversion is coupled closely with analog processor, it will require considerable additional silicon area, as well as digital busses to route signals from the sensors. If A/D conversion is to be performed on another chip, problems such as noise and crosstalk may interfere with the analog signals. In addition, the A/D chip will require many converters to process signals from all of the RF outputs in a timely fashion. An optimal solution requires circuitry which minimizes the drawbacks of each approach.

3. Constructing the Blocks for VLSI Receptive Fields

This section will describe the circuits and hardware necessary to construct each functional unit outlined in chapter 2. These experiments were performed using a 3μm digital CMOS technology. The goals were to find the limits of the MERF approach using a technology that was popular and readily available. The results will show whether this technology is suitable for these purposes or whether they will require other technologies.

3.1 Photodetectors with Built-in Light Adaptation Mechanism

Intelligent vision sensors need photodetectors with the largest possible sensitivity range to react with the large range of light intensity present in our environment. Environmental lighting conditions range across 14 decades of light intensity measured at the earth's surface: from bright June day conditions (10^0 W/cm^2), through full moon light (approximately 10^{-6} W/cm^2) to a darkest night sky illumination (approximately 10^{-14} W/cm^2) [1]. For intelligent vision sensors functioning under the same set of circumstances as human visual systems, their dynamic range must be similar. To meet this need, vertebrates have evolved retinas in which rods and cones have different sensitivity and built-in light adaptation mechanisms. Ideally, an artificial vision system would share these characteristics.

The vertical bipolar transistor and p-n diode are natural byproducts of a standard CMOS process and they share identical structures, with the p-well forming either the base of a vertical npn transistor or part of a diode [3-5]. The design of a photodetector that uses the same physical structure in two different operating modes can exploit this fact. When the structure is operating as a photodiode, either of the two p-n junctions present on the device can provide the output signal. The photocurrent from the substrate to the p-well can form one output signal, and the photocurrent from the n+ diffusion to the p-well can form the other. When the structure is operating as a phototransistor, the substrate forms the collector (which must be biased at +5V), the p-well is left floating to form the base, and the output is taken from the n+ emitter.

The most commonly used photodetectors in CMOS technology are vertical bipolar phototransistors (BJT's) and photodiodes, both featuring a single mode of operation [2-5]. The greatest drawback of the photodiode is that it is relatively insensitive to light, often requiring large detector areas. The vertical phototransistor is a more sensitive device able to detect an optical signal corresponding approximately to a moonlit scene [3], but also requires proportionally more power to operate. When considering very large arrays of photodetectors, the power requirements of the devices can become critical. A single phototransistor or diode with

reasonable dimensions is able to cover 5 - 7 decades of light intensity [2,5], with a similar range in output current. It becomes clear that to improve performance of artificial vision systems new photodetectors with increased dynamic range and minimized cost are necessary. It is also desirable to reduce the range of output current to 2 - 3 decades when using current-mode signal processing techniques.

3.1.1. Structure of the Multi-mode Photodetector

When a MOSFET is placed across the base-emitter junction of a vertical npn phototransistor (Figure 3-1), it can be used to force the device into one of two modes. In the diode mode, the MOSFET permits the substrate to p-well photocurrent to flow directly to the n+ emitter, so no transistor action occurs. The resulting output is simply that due to the reverse-biased substrate to p-well junction. When the MOSFET is turned off, electrons generated by photons in the p-well add to the base current of the vertical transistor, which is amplified by the forward current gain β of the device. Figure 3-1 shows a schematic diagram of these two modes of operation [15].

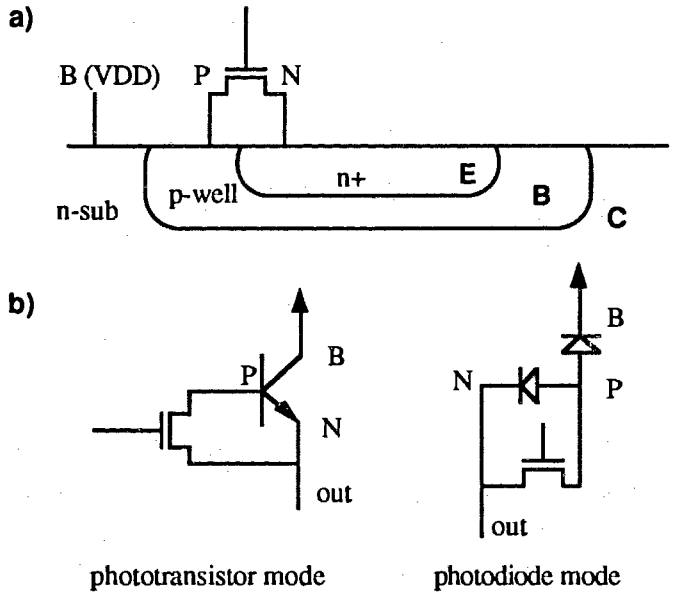

Figure 3-1: Structure and equivalent circuits of a two-mode photodetector

Devices that use a second vertical transistor in the Darlington connection to amplify the output current of the phototransistor can attain further modes of operation. In particular, relatively small BJT devices with moderate β can be cascaded to produce a reasonably large output current for a given illumination level. If this cascade can be selectively connected, allowing the extra gain to be used only at the lowest illumination level, a further increase in the overall dynamic range will result. Hence, the photodetector with three modes of operation will use two MOSFET's to determine the mode of operation of the device.

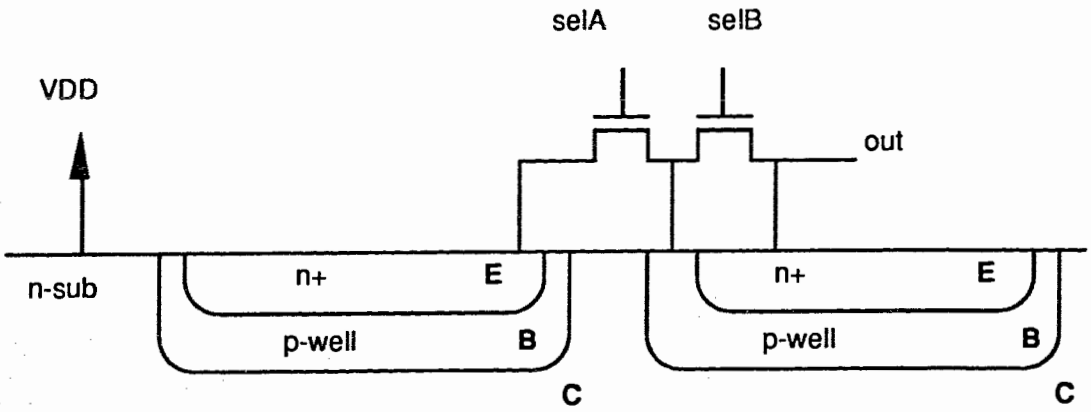

Figure 3-2: Structure of a three-mode photodetector

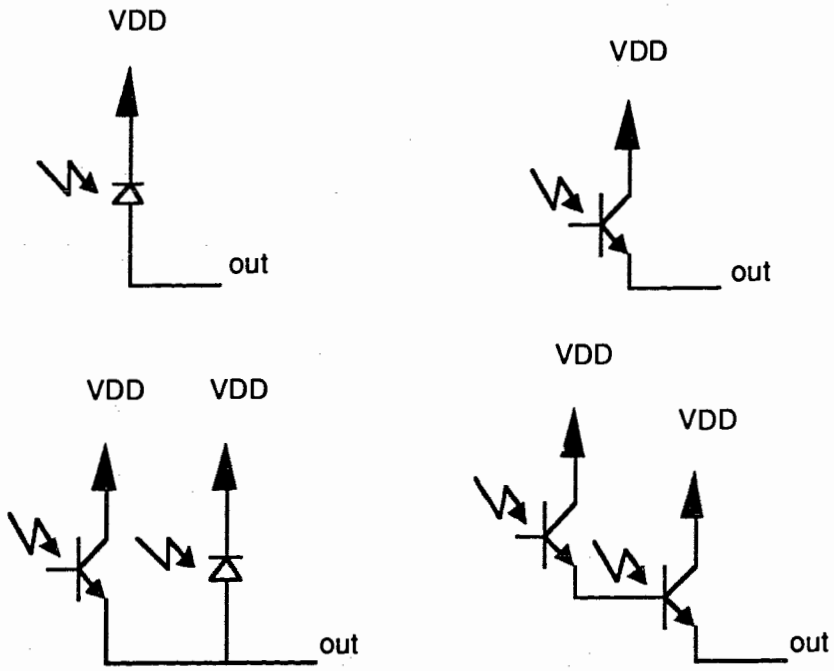

Figure 3-3: Equivalent circuits of a three-mode photodetector

In this design, the two gate inputs can be set in any of four combinations, each resulting in a different mode of operation. In this way, the device can operate either as a photodiode, a vertical npn phototransistor, or the Darlington connection of two phototransistors. A built-in light adaptation mechanism in the photoreceptor could use these three modes of operation within different ranges of light intensities.

In intelligent vision sensors, the photodetector will adapt its mode of operation according to the average light intensity. The phototransistor mode is used only for the low illumination levels, while photodiode mode is appropriate for higher illumination. A switching point between two or more modes can be chosen such that signal processing circuitry following the photoreceptor will receive a similar current range in all modes. This limited current range will simplify vision sensor

design and allow for minimized power consumption. It also makes the use of current-mode signal processing techniques practical. The sensitivity control could either be implemented in a global fashion, or set for local areas, allowing bright or dark areas of a scene to be examined in detail simultaneously [6].

3.1.2. Performance of the Multi-Mode Photodetector

The output of this device operating in phototransistor mode yields a current approximately two decades greater than that resulting from photodiode operation. For the sensor with built-in light adaptation mechanism, this output amplification can be selected through the simple application of a digital input signal. The results prove that two-mode photodetectors with a built-in light adaptation mechanism will benefit from dynamic range increased by approximately two decades.

Results for the three-mode photodetector are given in figure 3-4. These results demonstrate that a second phototransistor stage can be used to increase the output current by a further two decades. The two gain stages allow the output current at low light intensities to be increased by four orders of magnitude. As with the two-mode device, the operation mode is selected through two digital input signals.

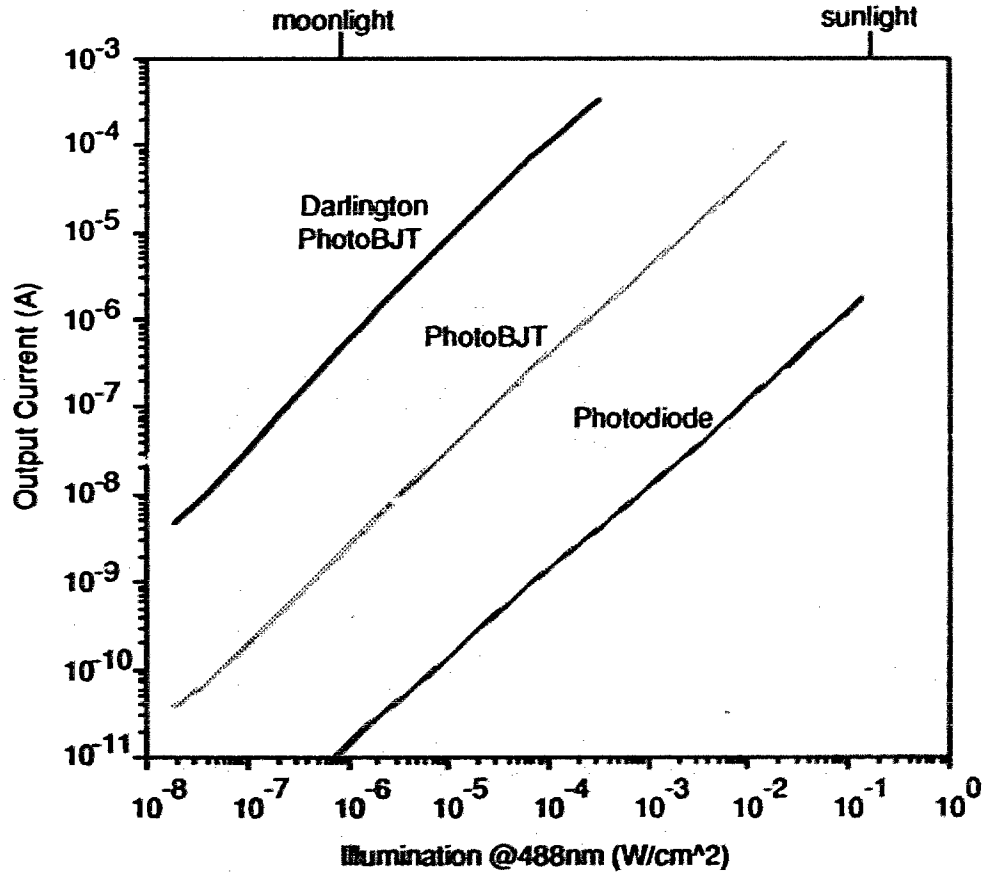

Figure 3-4: Measured response of three-mode photodetector

A multi-mode photodetector can be used to map a large range of illumination intensities onto a relatively small range of output currents, as shown in figure 3-5. The illumination range shown on the graph corresponds roughly to a dark night through to sunlight. The operating mode of the device provides approximately two bits of information, and allows an A/D converter to deal with a reduced analog signal range. The two-mode photodetector measured can reduce more than five decades of light intensity to less than three decades of output current. The addition of a third, Darlington-connected mode increases the allowable illumination range even further, to seven or eight decades of light intensity. It is also worth noting that if the processing circuitry can accept a larger output current range, the illumination range can be increased even further.

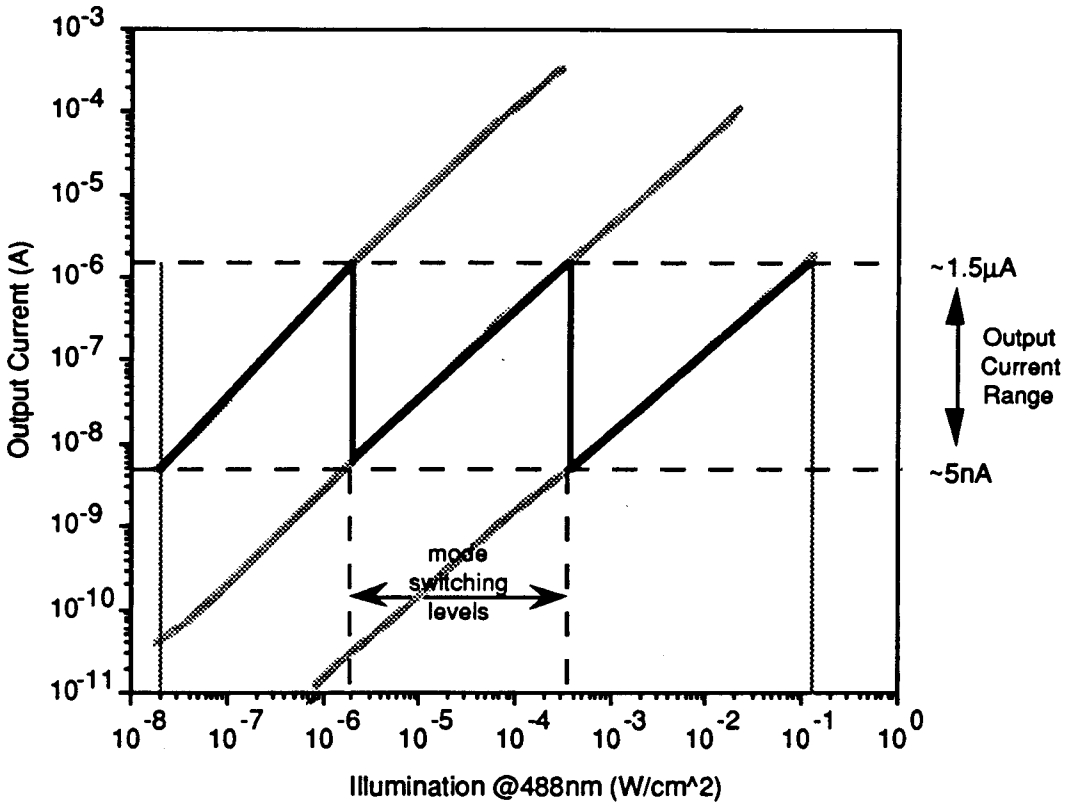

Figure 3-5: Using multiple operation modes to reduce output current range

3.1.3. An Integrated Switching Mechanism Between Operation Modes

An intelligent vision sensor should be usable over a large range of illumination conditions and on scenes containing a large dynamic range. This can be accomplished by multi-mode photodetectors, with an automated transition from one mode to the next. In this way a device that may be operated in more than one range of sensitivity can be converted to one with a built-in light level adaptation mechanism. As the photodetectors discussed here use digital logic levels to

determine their operating mechanism, it is natural that any circuit used to control them will function as a light level controlled switch.

The circuitry required to set the sensitivity control inputs can be relatively straightforward. A block diagram of such a mechanism is shown in figure 3-6. The output of a photodetector operating in phototransistor mode can be compared to high and low thresholds. If the output exceeds a certain value, the photodetectors on the chip can be switched to photodiode operation. Similarly, if the output drops below another threshold, the Darlington operation mode can be enabled. A hysteresis mechanism for each of these levels will be desirable, similar to background adaptation and dark adaptation mechanisms existing in human vision, and can also be implemented in a straightforward manner.

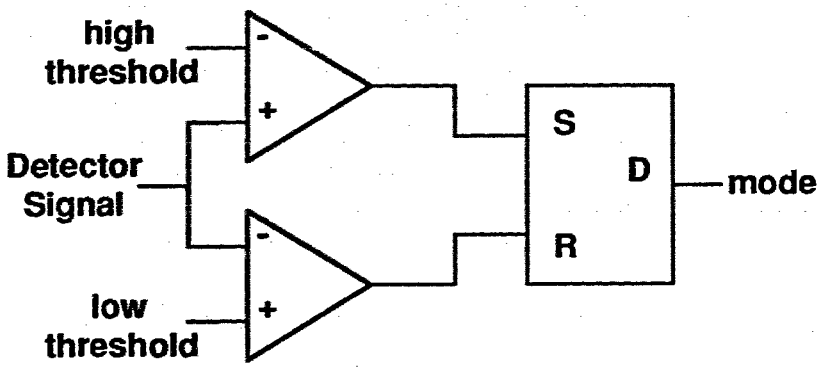

Figure 3-6: A two-mode integrated switching mechanism

The effective cost of the sensitivity control mechanism can be kept low by allowing each circuit to control a group of photodetectors, such as an entire receptive field. This would allow the vision sensor to process scenes containing extremes of light and dark. If local control is not required, a single control circuit could serve a large population of photodetectors in a vision system. The sensitivity control could also be designed with a low-pass filter, allowing the sensor array to respond to a gradually changing environment without over-reacting to a sudden burst of light, such as a flashbulb. Such an event would also cause a sensor array to draw an inordinate level of power if at its most sensitive level, so current limiting circuitry would also be valuable.

Varying levels of sensitivity in a single photodetector mimic the mechanism existing in the vertebrate retina's photodetectors. With the addition of circuitry to select an appropriate sensitivity level, a light sensor could mimic the built-in light adaptation mechanisms present in the retina.

3.2 Analog Signal Processor

Early vision processes can be performed with a conventional CPU, but the time required to perform several operations over an entire array can be formidable. Alternatively, one can take the

fact that most early vision processes involve combining outputs from neighboring detectors, and situate many processors at points evenly distributed across the sensor. This allows the input to be processed in a parallel fashion, greatly reducing the time required, but introduces the further problem that the processors will now consume a greater fraction of the sensor's real estate. Additionally, the interconnections necessary to allow a processor to access 200 or so neighboring cells become very considerable, and a defective processor will result in a relatively large blind spot. Another alternative is to minimize the processors, and give them specific tasks to perform with specific input cells. This is the process taken here. Each processor performs only a specialized function, and consumes a minimum of silicon area.

Most of the MERF processes involve the computation of weighted-sum functions. On silicon, this can be implemented in either the digital or analog domain. The more conventional approach is to use digital processors. This requires that the analog signal from each of the photodetectors be converted to digital form with an A/D converter. The digital data can then be processed with adder circuits and will give a result that is accurate and reasonably fast. The primary drawbacks of the digital approach are in the amount of silicon area required. The A/D converters and adders will require a very large area if distributed among the photodetectors, and create a data bottleneck if they are centralized. In addition, the interconnections required for digital processing require a large amount of space, as each bit needs its own data line.

The analog approach is well suited to this problem. A/D converters are not necessary. The current mode circuitry for weighted sums can be implemented easily using current mirrors, in much less area than a digital adder, although the overall accuracy is smaller. At the cost of a little accuracy, the current mode processor provides higher speed, smaller silicon area, lower power consumption, and far fewer interconnections. The conventional drawbacks of analog signals, such as susceptibility to noise and line resistance are not factors in local, fully-analog computation.

Due to the large amount of silicon area required for digital processors, it is desirable to use analog techniques. It is possible to simulate the results of a digital design, but the implementation would be uneconomical.

3.2.1 Mathematical Operations with Current Signals

The mathematical operations necessary to implement a microelectronic receptive field are scaling, addition and subtraction. When using current mode signals, the processes of addition and subtraction can be performed by simply connecting wires. All currents flowing into a circuit node will sum to the current flowing out of it. The subtraction operation is implemented by inverting the sign of the signal to be subtracted, and then proceeding as with addition. Signal scaling can also

be implemented using current mirrors. Figure 3-7 shows circuits using current mirrors to implement mathematical functions. The deviations of the measured output of these circuits from the expected value are shown in figure 3-8.

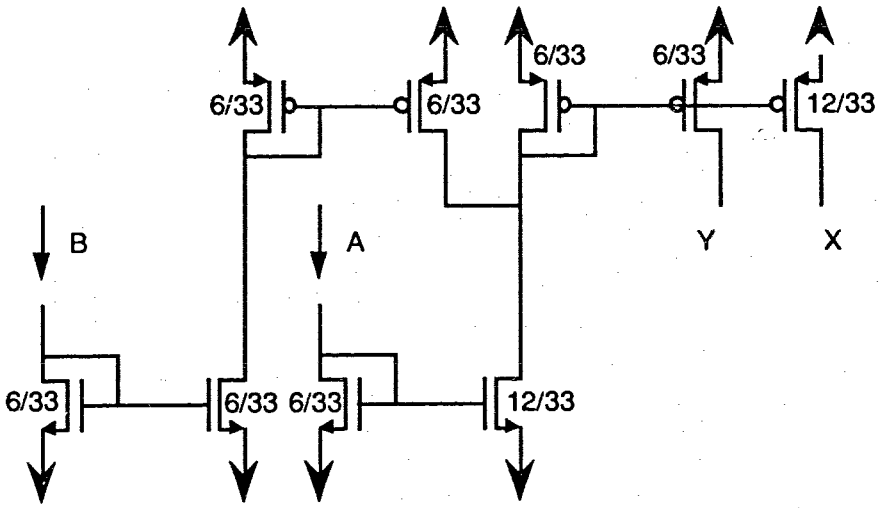

Figure 3-7: Current mirror circuit implementing X = 2A-B and Y = 2(2A-B)

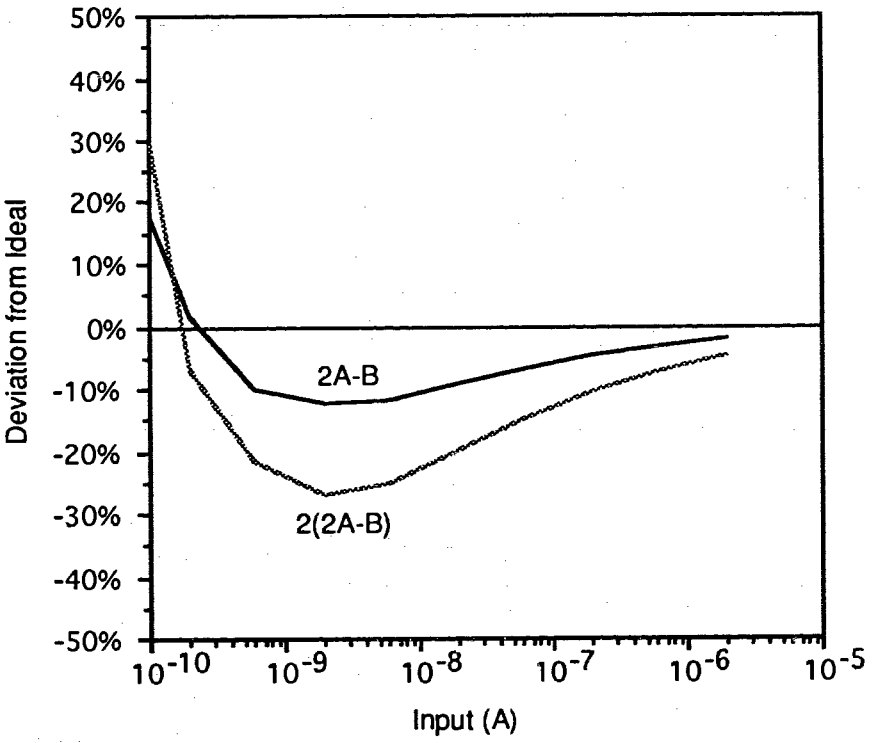

Figure 3-8: Deviation of output from the ideal value in the circuits of figure 3-8

The deviations from the ideal value vary from less than 10% to more than 30%, depending on the size of the currents being examined. This circuit will operate with an error less than 30% over five decades of input current from 10^{-10}A to 10^{-5}A. For other operations, the deviation is related to the complexity of the mathematical operation performed, and is sufficiently small for the

implementation of robust receptive field algorithms, such as Sobel's edge-detector described in chapter 2.

3.2.2 Current Mirror Design

In many situations it is desirable to duplicate a current mode signal to transmit it to multiple destinations. This can be done with a simple circuit known as a current mirror. A CMOS current mirror consists of two or more transistors connected so that they all have the same gate to source voltage (Vgs). A reference current is applied to the drain of the input transistor, which is also connected to the common gates. In this way the reference current determines the Vgs necessary to allow it to flow, and each of the output transistors receives the same gate voltage, as in figure 3-9.

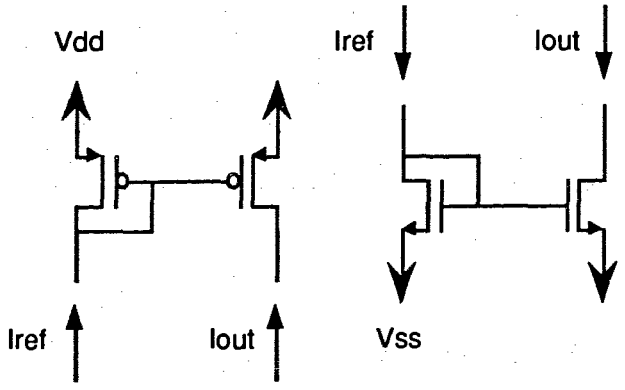

Figure 3-9: Current mirrors using (a) p-channel and (b) n-channel devices

As transistor mismatches are accentuated in the subthreshold region of operation, it is desirable that current mirrors be kept above this region. By using multimode photodetectors, the range of photocurrent values can be reduced from seven decades to three. Consequently, by using long-channel transistors, the need for subthreshold operation can be avoided entirely.

The equation governing above threshold operation of an enhancement MOSFET is given in equation 3.1. The equation governing subthreshold operation of an enhancement MOSFET is given in equation 3.2.

$$I_D = \left(\frac{W}{L}\right)\frac{\mu_n C_{ox}}{2}\left(v_{gs} - V_T\right)^2\left(1 + \frac{v_{DS}}{V_A}\right) \tag{3.1}$$

$$I_D = \left(\frac{W}{L}\right)Ke^{-\kappa v_G/V_T}\left(e^{v_S/V_T} - e^{v_D/V_T}\right) \tag{3.2}$$

From these equations, current scaling can be done by altering the aspect ratios (W/L) of the output and reference transistors. This is not effective in the subthreshold region. Instead, it was

found to be necessary to use parallel transistors. This is due to edge effects of the transistors, whereby two channels next to one another react differently than if they were touching. This effect can be seen in figure 3-10, where the output ratio of two n-channel x2 scaling current mirrors is shown for circuits consisting of two parallel output transistors and an output transistor with an aspect ratio twice that of the reference transistor. The output of the two parallel transistors provides good matching for currents as small as 100 pA, while the transistor with a doubled aspect ratio provides an output which is closer to 2.5x for most levels.

Had the circuits in figure 3-7 been implemented with parallel output transistors, rather than ones with double the aspect ratio, the results in figure 3-8 would be much closer to ideal.

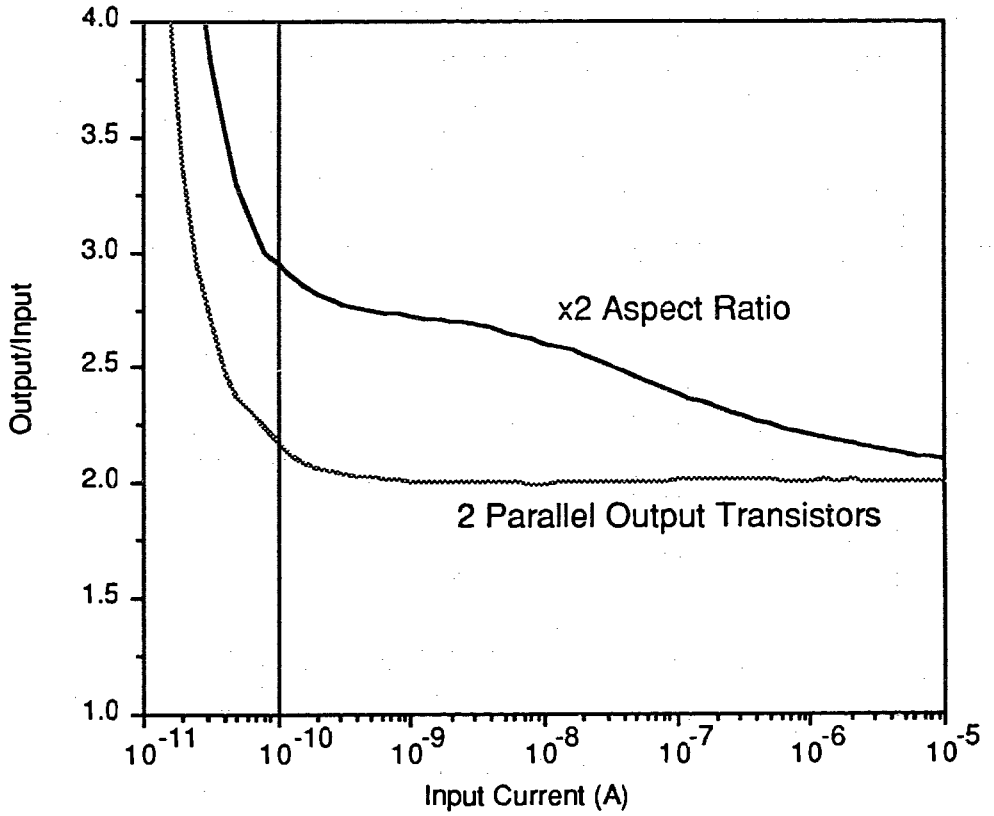

Figure 3-10 Comparison of n-channel x2 scaling current mirrors, one using two parallel output transistors and one using a doubled aspect ratio

3.3 A/D Converter

Although current mode signals allow focal plane processing to be simply and efficiently implemented, the analog domain is not necessarily the optimal form for sensor output. Analog signals are vulnerable to noise and sensitive to line resistance or current leakage. These factors make it difficult to preserve the large dynamic range of the output while transmitting it over the

relatively large distance to an adjacent chip. In addition, analog signal transmission requires the use of one or many fast off-chip A/D converters.

One solution to this problem is to perform a full A/D conversion on the image sensor chip. This approach also has drawbacks. If A/D converters are distributed one per cell, the silicon area required for converters and bus connections becomes overwhelming, and the power requirements will start to become a limiting factor. If the A/D converters are distributed one per row, the silicon area required is greatly reduced, but the A/D conversion will still provide a data bottleneck. For example, in a moderately sized 100x100 cell sensor, each A/D would have to service 100 cells. As each cell may be providing an output current of 1 nA or less to a bus with a capacitance on the order of 1 pF, the accumulated settling time between conversions could be as great as 0.1 seconds. The additional capacitance of the output transistor and the A/D converter input will increase this latency period further. Overall, it is unlikely that a scan rate greater than 10 Hz could be attained.

Alternatively, a partial Analog to Digital conversion could be performed on chip. This could be done if each cell converted its signal to the frequency domain. The resulting signal would be quasidigital (digital in amplitude, analog in time), and would combine the flexibility of an analog signal with the noise immunity of a digital signal. Such signals require only one output line for transmission and can be easily converted to a fully digital signal by an off-chip circuit. If desired, they can also be converted back to a fully analog signal (with a PLL) after having been cleanly transmitted from the sensor chip.

The advantages of a frequency mode output are that only one output line is required for a signal, and the conversion can be done with a small amount of silicon area. The drawbacks of this approach are that each frequency-domain signal must either be timed or counted to form a fully digital signal, and lower frequency signals will require more time to interpret. For example, if the output frequencies from a cell range over 6 decades between 10 Hz and 10 MHz, the time required to count one cycle will range from 0.1 μs to 0.1 s. If the circuitry performing this timing must be multiplexed between many signals, the lower frequencies will require large latency times. Depending upon the intended application, this may present a problem. The solution to this problem is either to have the sensor transmit data to another fully parallel system, so that each input must examine only a small number of outputs, or to design a circuit which produces a smaller range of output frequencies.

3.3.1. Current to Frequency Converters

Analog-domain signals can be translated to frequency-domain signals through the use of relatively simple voltage-controlled ring-oscillator circuits. A pseudo-DTL CMOS voltage controlled oscillator [14] is shown in figure 3-11. The implementation of this circuit requires little silicon area, and a standard cell layout consuming 150 x 96 μm^2 is given in figure 3-12 The primary drawback of this circuit design is its high susceptibility to variations in transistor threshold voltages. The operating frequency of the device is directly related to the current flowing through the 7/25 transistors, which is in turn sensitive to their threshold voltage.

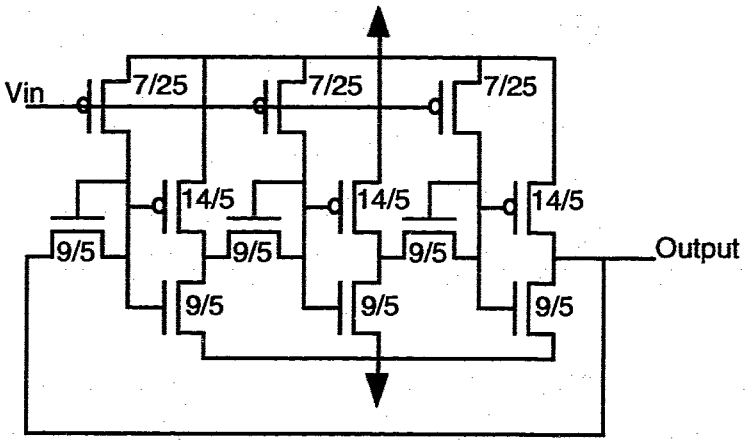

Figure 3-11: Circuit diagram of voltage-controlled oscillator

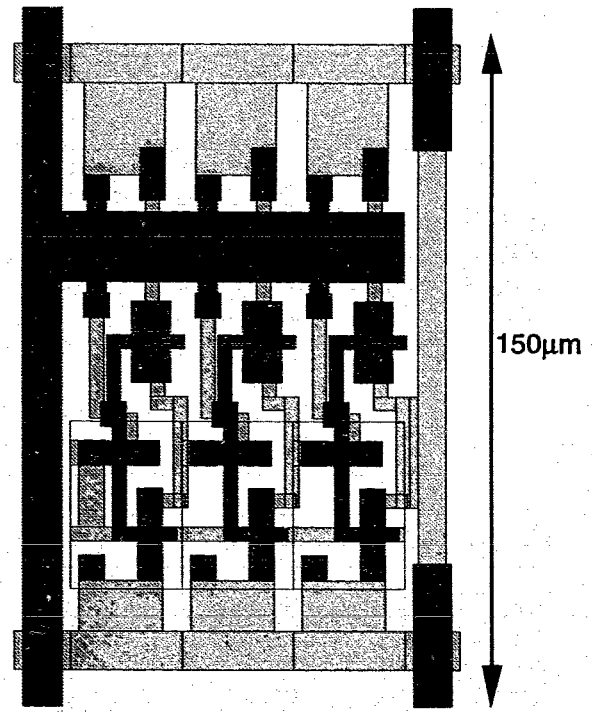

Figure 3-12: Layout of voltage-controlled oscillator

3.3.2. Performance of Current to Frequency Converters

The VCO can be used as a current controlled oscillator through the addition of a long channel MOSFET at the input of the VCO, as in figure 3-13. This device converts a logarithmic change in input current into a linear change in voltage. As the VCO converts a linear voltage change into a logarithmic change in frequency, the combined circuit translates a linear change in current to a linear change in frequency over at least six decades. Characterization results for this circuit are given in figure 3-14.

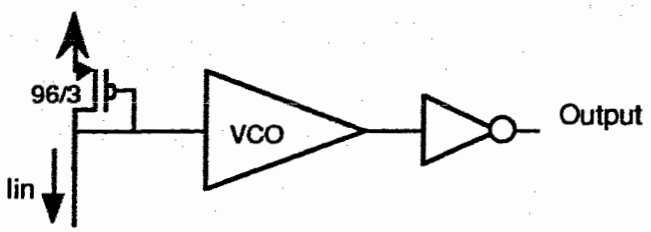

Figure 3-13: VCO connected as current controlled oscillator

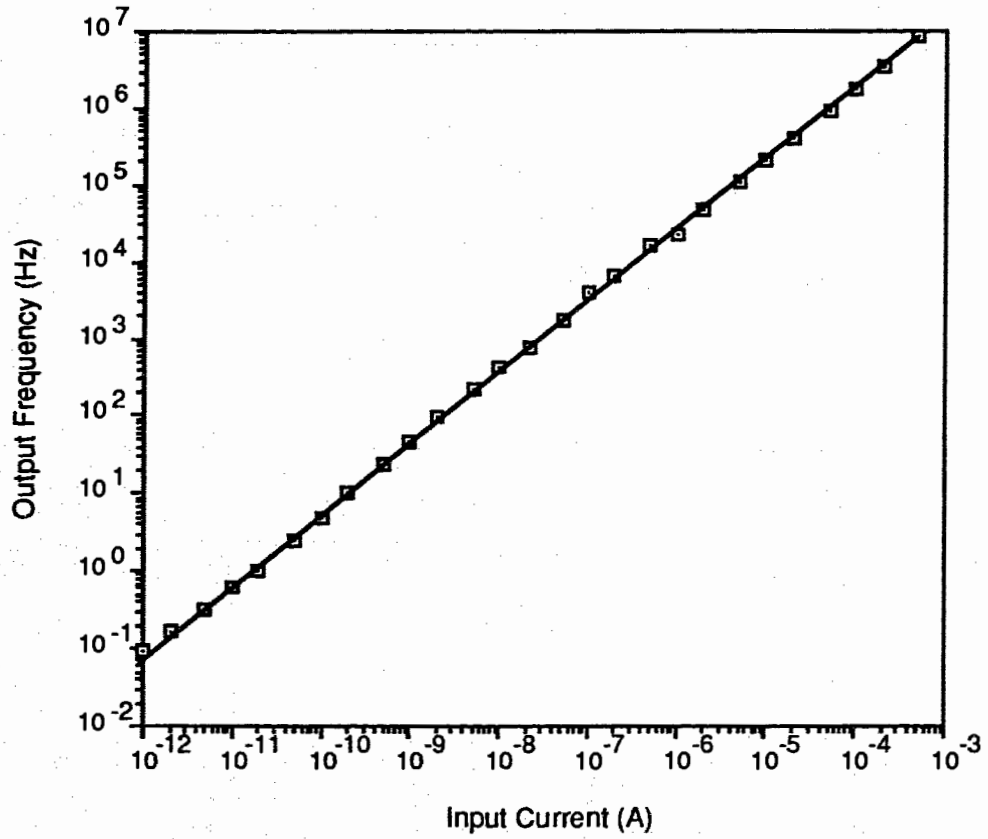

Figure 3-14: Output of VCO connected as current controlled oscillator

4. Prototype Vision Sensors using Receptive Fields

In order to experimentally test local vision processing with receptive fields, a few prototype vision sensors have been designed and fabricated. This chapter describes their structure, while their performance will be assessed from measurement results in chapter 5.

4.1 Design Philosophy and Approach

The design philosophy for this research allows concepts to be tested at different levels. The first level tested the component blocks described in Chapter 3 as individual units. The second level tests prototype small-scale sensors made from these components, built for concept verification purposes. The next levels, not covered in this thesis, would transform the prototype sensors into devices with enough MERF cells to perform real world applications, such as peripheral vision functions in a smart vision sensor.

As the vision sensors presented in this thesis are designed to verify MERF concepts, high device density is not one of the goals. Instead, a modular structure based upon the standard cell approach is used. This allows sensor components to be treated as interchangeable blocks, but silicon area is not used efficiently. The combination of the standard cell approach and the relatively large scale 3µm CMOS process used restrict the number of sensor cells per chip to 8x8 or so. This severely limits the complexity of scenes which may be imaged, so most tests are done using simple image gradients, in which one side of the chip is exposed to more light than the other.

Over the course of this research, nine chips, each with an area of 4.5 x 4.5 mm^2 have been designed, fabricated, and tested. They contained various combinations of building blocks designed as standard and custom cells, as well as prototypes of small-scale MERF-based sensors. These circuits targeted several experiments intended to demonstrate the feasibility of the MERF approach, MERF resistance to defects and device variations, and the operating range of the sensors.

Due to the long turn around time in IC fabrication, it was necessary to fabricate complete sensors, as well as sufficient quantities of MERF components to allow the assembly of complete MERFs. Without these additional components, it would be impossible to determine the cause of a non-functional sensor, or retrieve useful data from it. This modular approach allows complete MERFs to be built up from a selection of functional components. As the MERF components generally

performed well, it was possible to measure useful data from the more complex sensors, as well as for more basic ones.

4.2 A Multi Sensitivity Image Sensor with Frequency Mode Output

This circuit was designed to test concepts of the receptive field based sensor without performing any image processing. The analog processor in this circuit is effectively removed, resulting in an image sensor in which each photodetector controls one output oscillator. In addition, two-mode photodetectors are used to boost the effective dynamic range of the sensor.

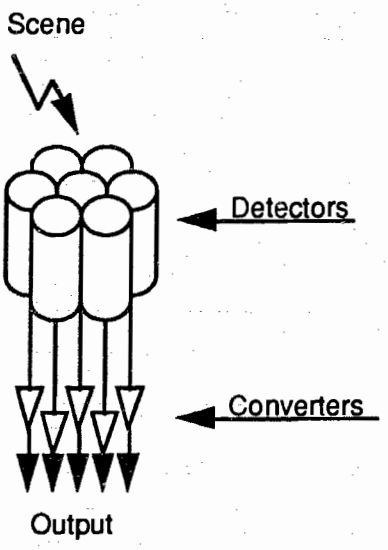

Figure 4-1: Schematic diagram of a section of the multi sensitivity image sensor

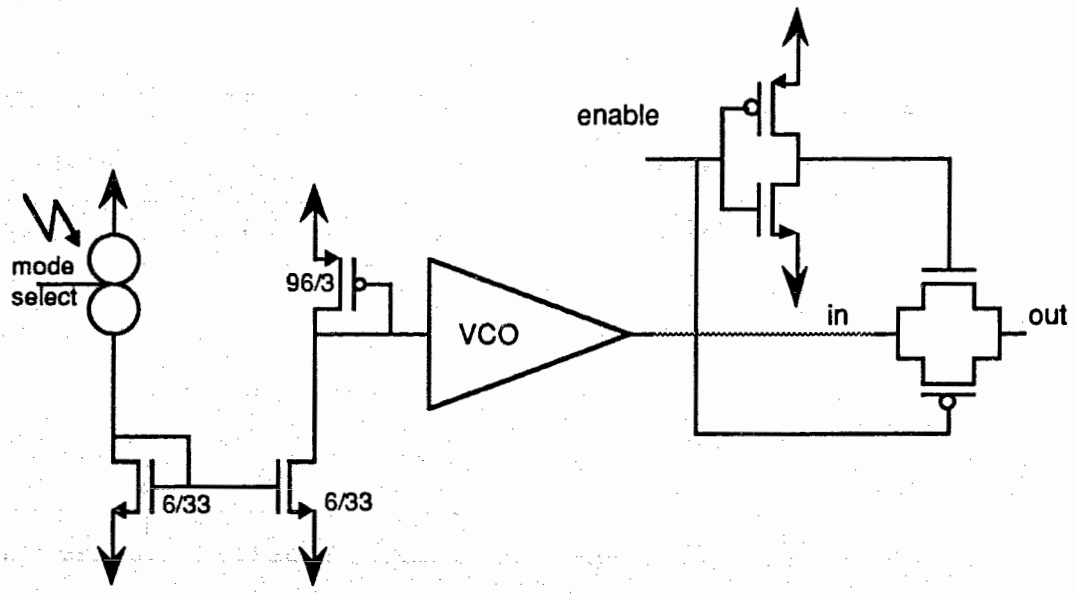

Figure 4-2: Circuit diagram of a multi sensitivity image sensor cell

The circuit diagram of a single cell of this sensor is given in figure 4-2. The current-mode signal from the photodetector is mirrored through the two n-channel transistors, and passed to the VCO,

which is connected to a wide-channel to form a current controlled oscillator. The output of the oscillator passes through a transmission gate to a common output bus. In this design, only one cell can be read at a time. This is not an optimal design for a smart vision sensor, as it forms a data bottleneck at the output bus, but it allows much simpler external processing circuitry to be used in a testing situation

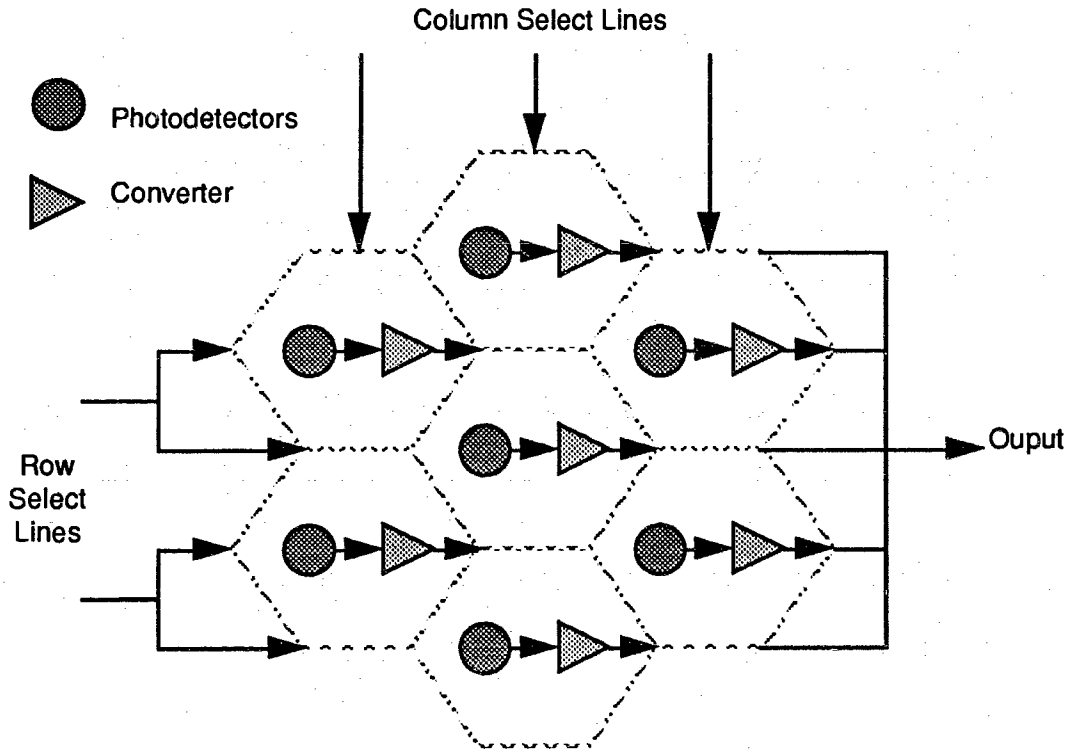

Figure 4-3: Block diagram of a section of the multi sensitivity image sensor

The block diagram of a section of this sensor is given in figure 4-3. The cells are laid out in a hexagonal format, in which each one has six neighbours. Data is not shared between adjacent cells, as it would be in a MERF-based sensor. Row and column select lines determine which cell should place its signal on the output bus.

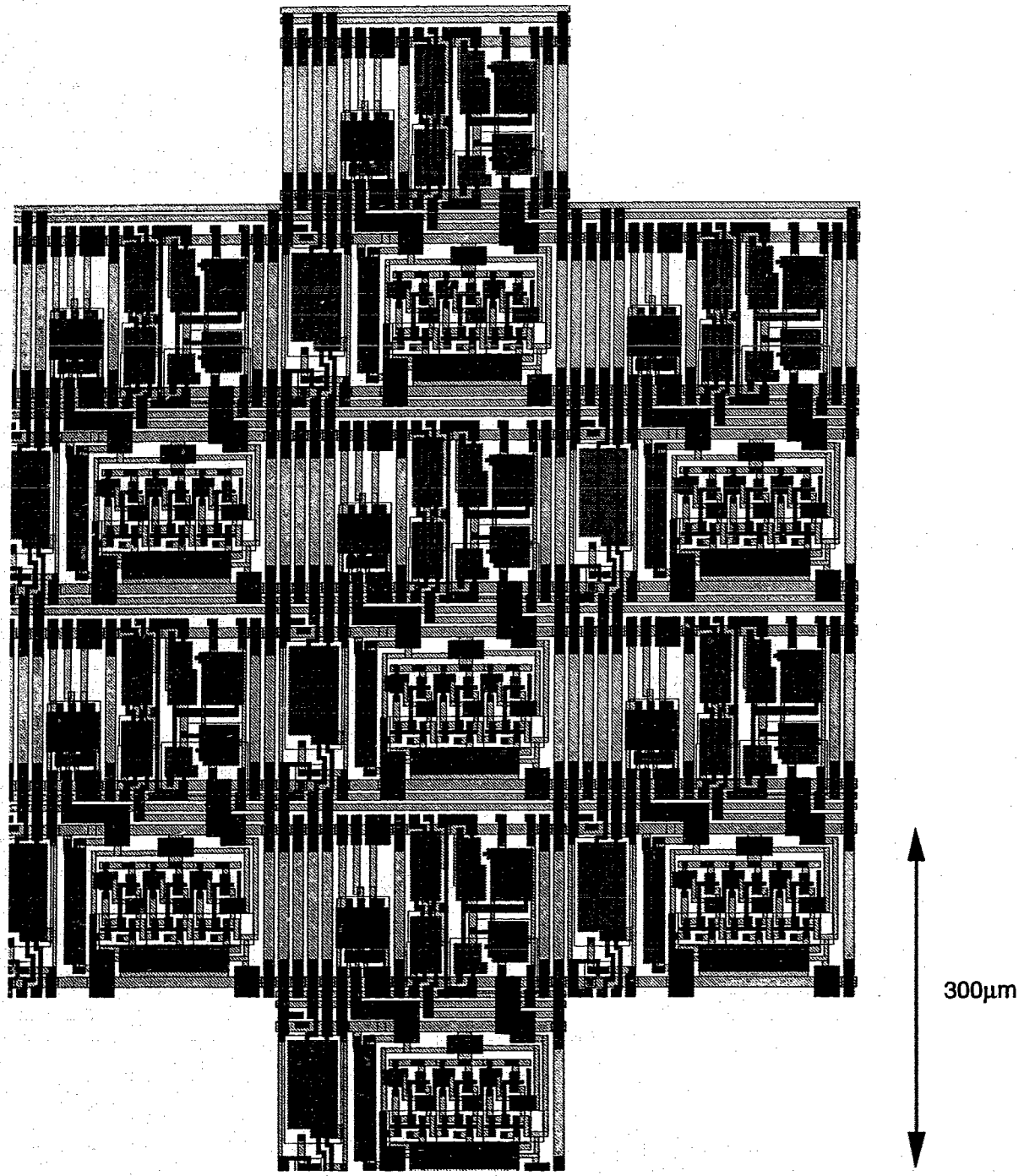

Figure 4-4: Layout diagram of a section of the multi sensitivity image sensor

The layout diagram of a section of the image sensor is given in figure 4-4. The cells are arranged in a staggered fashion to attain the hexagonal layout pattern. Control signals are transmitted vertically and horizontally to each cell, and the output bus runs horizontally.

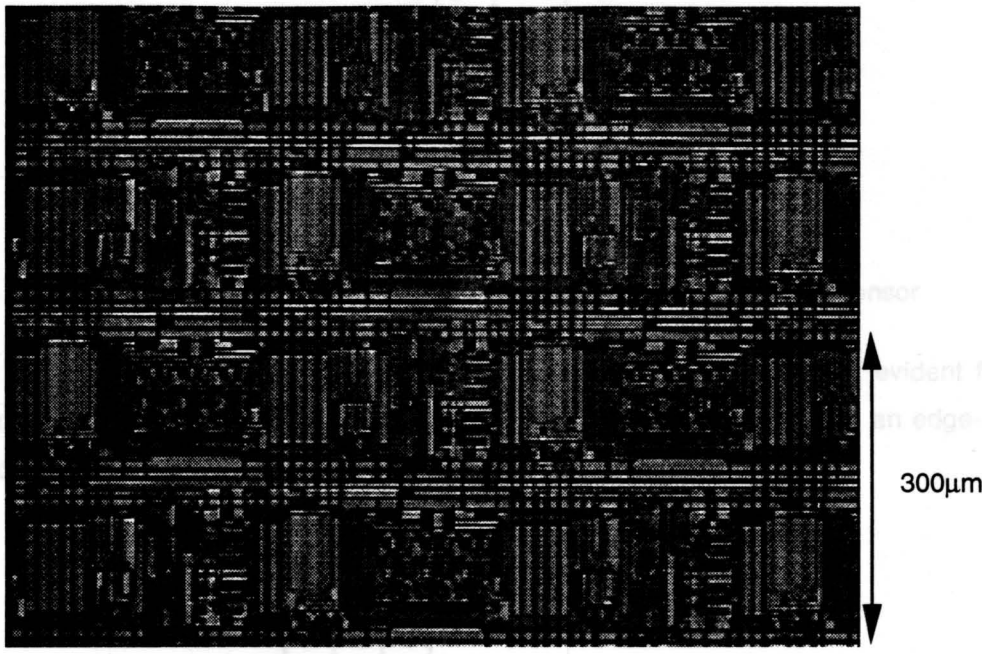

Figure 4-5: Photomicrograph of a section of the multi sensitivity image sensor

A photomicrograph of a section of the image sensor is given in figure 4-5. All of the sensor area except for the photodetectors themselves is covered with a metalization layer. This blocks light from the other active devices in the circuit, preventing photo-induced currents from influencing the output value.

4.3 A Smart Vision Sensor Providing Edge-Detection

This circuit was designed to test concepts of the receptive field based sensor by performing a simple image processing algorithm. The analog processor in this circuit is wired to perform Sobel's edge-detection algorithm. This consists of two convolutions, one for vertical edges and one for horizontal edges. These functions are shown in figure 4-6, each combines the signals from six photodetectors. A digital signal determines which function will be performed by the sensor and passed to the output oscillator. The frequency-mode output signals are combined to an output bus to make measurement more convenient. This sensor uses three-mode photodetectors to allow a greater range of light intensity to be examined.

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Horizontal edge-detector

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Vertical edge-detector

Figure 4-6: Sobel's edge-detector as implemented in the vision sensor

Figure 4-7 is a schematic diagram of an edge-detector cell. Although not evident from the diagram, each photodetector passes its output to all eight neighboring cells; an edge-detector cell is centered over each photodetector in the sensor.

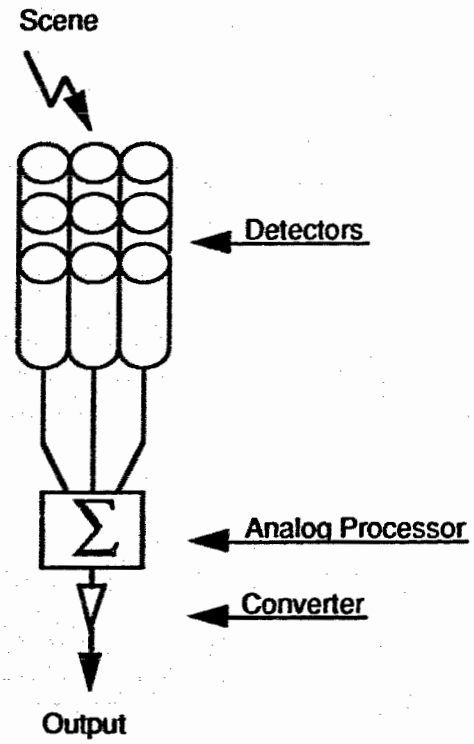

Figure 4-7: Schematic diagram of a section of the edge-detector image sensor

This overlap is more evident in the block diagram of figure 4-8, which shows a section of the sensor.

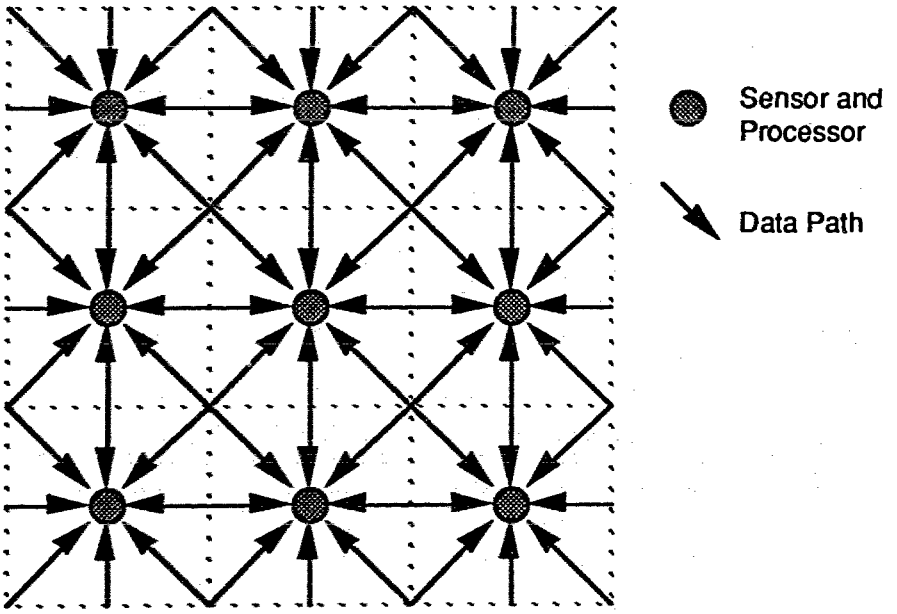

Figure 4-8: Block diagram of a section of the edge-detector image sensor

The circuit diagram of a single cell of this sensor is given in figure 4-9. In 4-9 (a) the current-mode signal from the photodetector is duplicated with a set of n- and p-channel transistors, creating current signals which are transmitted to neighboring cells. The signals from the surrounding cells are combined in 4-9 (b) to form the output signal. The function select input selects which of the two edge-detection convolutions will form the output. In addition, a bias signal is applied to another n-channel device which serves to keep a unipolar current signal for the oscillator. Output select circuitry in 4-9 (c) places the buffered frequency-mode signal onto the output bus.

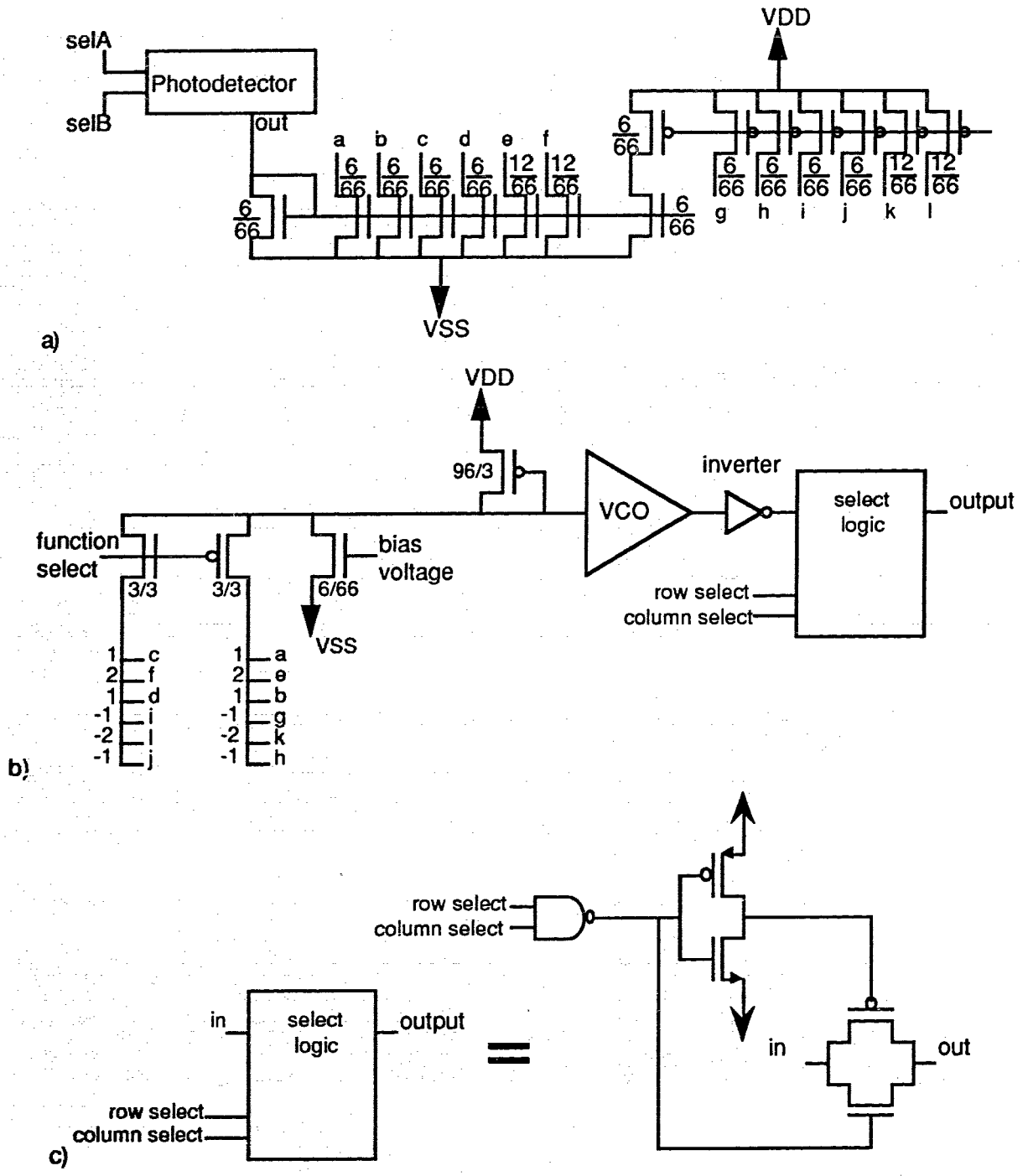

Figure 4-9: Circuit diagram of a edge-detector sensor cell, showing (a) duplication of photodetector output for neighboring cells, (b) signals from neighboring cells being combined to form the output signal, and (c) output select circuitry.

The layout diagram of a section of the image sensor is given in figure 4-10. The cells are arranged in a regular x-y grid which reflects the convolution being performed. Control signals are transmitted vertically and horizontally to each cell, and the output bus runs horizontally.

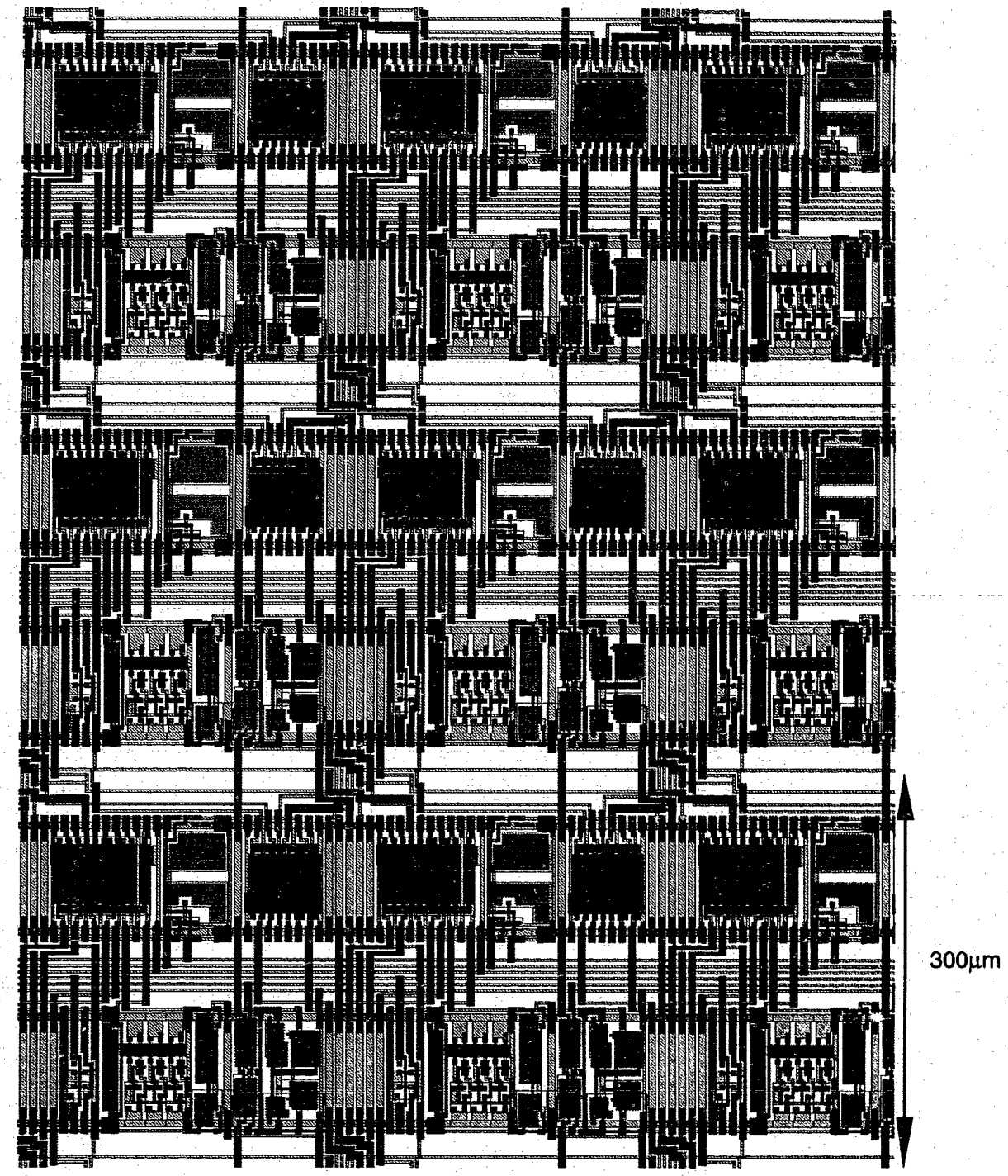

Figure 4-10: Layout diagram of a section of the multi sensitivity image sensor

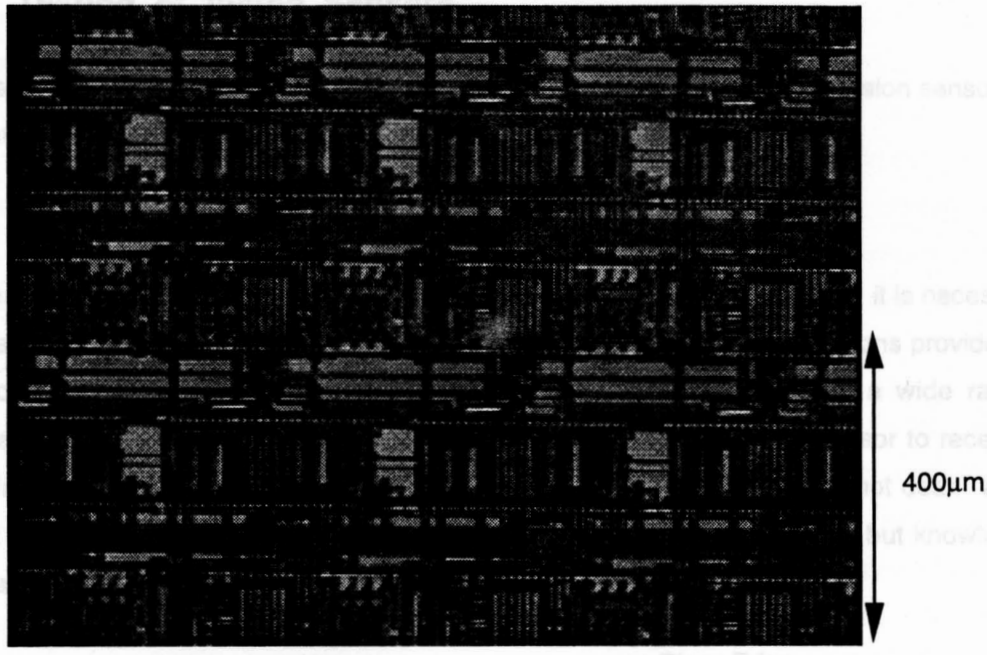

Figure 4-11: Photomicrograph of a section of the edge-detector image sensor.

A photomicrograph of a section of the image sensor is given in figure 4-11. All of the sensor area except for the photodetectors themselves is covered with a metalization layer. This blocks light from the other active devices in the circuit, preventing photo-induced currents from influencing the output value.

5. Testing of Vision Sensors

This section will describe the equipment necessary to measure the prototype vision sensors, and present their performance.

5.1 Scene Generation Optics

In order to characterize either individual photodetectors or entire vision sensors, it is necessary to provide a scene to the sensor. This scene should be able to mimic the conditions provided by a real-world scene. Desirable features in such a scene generator include a wide range of illumination intensities, a uniform beam, allowing all detectors in a vision sensor to receive the same input, as well as known spectral characteristics. As this research has not dealt with the effects of colour, control of the wavelengths present in a scene is not critical, but knowledge of the beam wavelength is important.

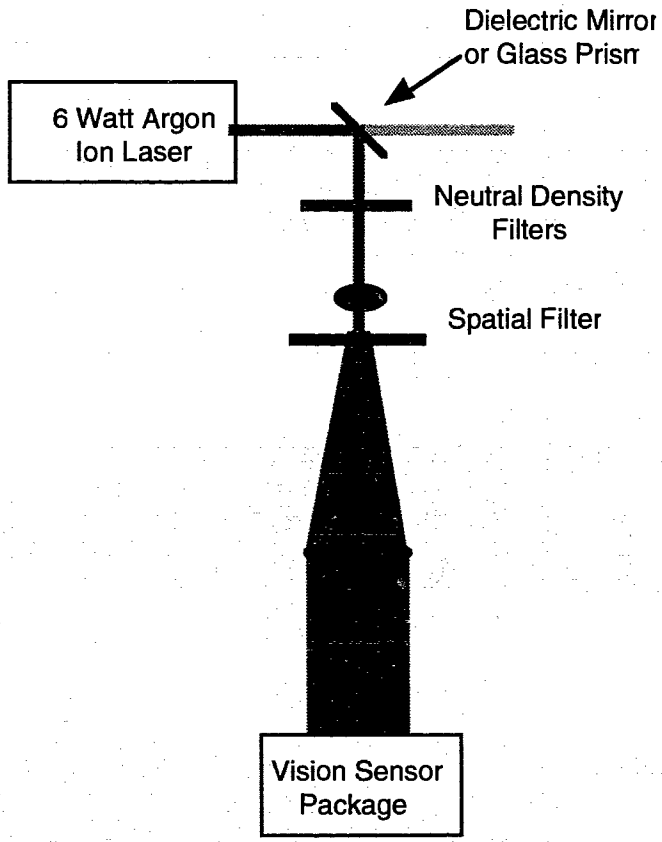

Figure 5-1: Schematic diagram of optical test bench

In order to meet these requirements, an optical test bed was set up using a 6 Watt Argon Ion Laser (λ = 488 nm / 514 nm) as the illumination source. The laser power itself is controlled over two decades, and is internally stabilized to keep power fluctuations to less than 1%. A glass prism is

substituted for the dielectric mirror (figure 5-1) to reduce the range of laser power by approximately two orders of magnitude. Further filters allow the useful range to be expanded to over seven decades. A beam expander is used to enlarge the beam from a two millimeters in diameter to an area approximately five centimeters in diameter. In theory, a laser beam has a Gaussian intensity profile, and this will be preserved as the beam is expanded. In order to obtain a uniform intensity over a chip-sized area (~5 mm x 5 mm), the beam must be expanded sufficiently for the peak of the gaussian to be flat over this area. It is also of note that the power incident on the chip decreases with the square of the expanded beam area, so over-expanding the beam is not a desirable solution.

The expanded beam was characterized by sliding a small-area photodetector across it, and recording the intensity at each point. This procedure revealed that while the beam was roughly Gaussian, it did not present a smooth curve. Instead, interference fringes were present in the beam, and they resulted in spatial "noise" when the beam profile was measured. In order to remove these effects, it was necessary to pass the beam through a spatial filter. A spatial filter consists of a pinhole through which the beam is focused. The process of focusing the beam to a point effectively performs a Fourier transform in the spatial domain, and the pinhole is of a size sufficient to pass the low spatial-frequency beam energy, and block the higher spatial-frequency noise components. Spatial filtering commonly results in an energy loss of 5 - 30%, depending upon the initial beam quality, the precision to which the beam is focused, and the thickness of the pinhole material.

The profile of the expanded and filtered beam is shown in figure 5-2. During measurements, the beam power was measured by placing a Newport 840 optical power meter in place of the sensor being measured.

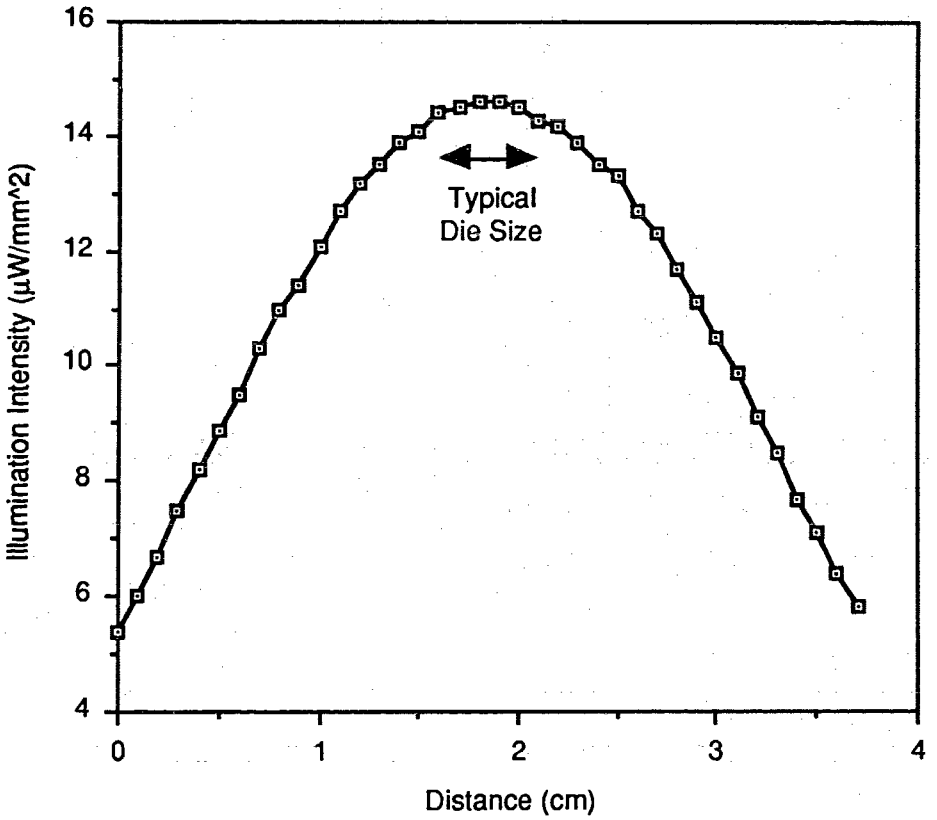

Figure 5-2: Expanded and spatially filtered laser beam profile

It can be seen from the beam profile that there is a region with a diameter of about 0.5 cm over which the illumination intensity varies by only a few percent. This is the area in which chips being measured are placed.

5.2 Sensor Measurement Techniques

In order to simplify the process of collecting data from image sensor chips an external data collection system was set up. This system is based upon a Motorola 68HC11 microcontroller (6811), which forms a link between the image sensor chip and a personal computer (PC). The 6811 sets up the control lines for the image sensor, and examines the pulses on the output bus to calculate the selected oscillator's frequency. This data is then passed to the PC through a serial link. The PC stores the incoming data, and allows the user to program the 6811 board.

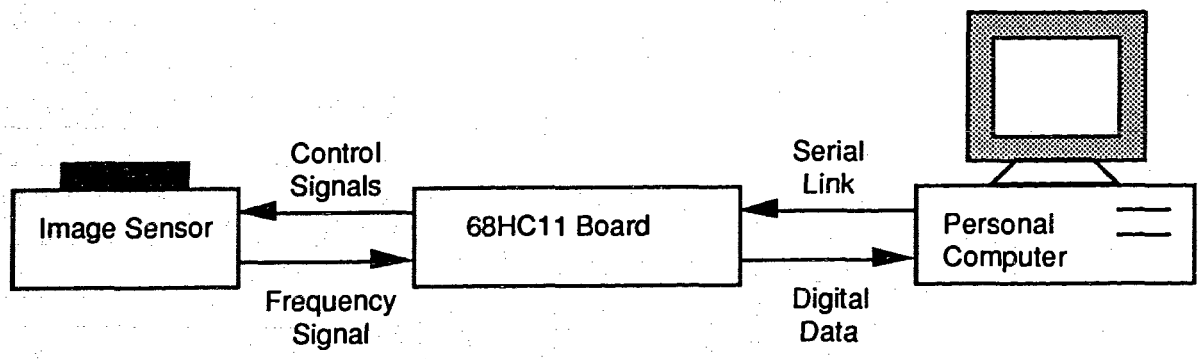

Figure 5-3: Data flow to and from image sensor chip

The complete system allows data to be read easily and efficiently from the image sensor chip, and is well suited to a testing environment. However, two data bottlenecks are present in this system. In a situation where scan rates of 30 Hz or greater were required, it would be necessary to bypass the data bottlenecks. The first is between the 6811 and the sensor chip, since each output cell must share a common output bus. To eliminate this it would be necessary to examine more output signals simultaneously. This could be done with multiple output lines, or ultimately, by integrating digital processors with the sensor. The second bottleneck is the serial line between the 6811 and the PC. This isn't a serious problem, since a practical system based on such a sensor would use a digital processor with a higher speed link between the initial digital processors and the controlling processor.

5.3 Vision Sensor Designs and Their Capabilities

The various vision sensors and associated components described in this thesis were fabricated using a standard 3μm CMOS process available through the Canadian Microelectronics Corporation. It provides a double metal layer polysilicon gate technology with a minimum feature size of 3μm. The standard cell approach consists of designing each circuit component to fit within a certain area, using the same connection conventions for each component [8]. In this way, designing a circuit consists largely of selecting the component blocks and wiring them together. Standard cells also allow components to be replaced and updated without modifying the entire circuit. The main drawback of this approach is that it severely restricts circuit density.

Another limitation of these sensors is due to the device variation of the VCO. This is described in more detail in section 3.3.1. This variation could be reduced either through the use of an oscillator design less susceptible to threshold voltage variation, or by multiplexing MERF outputs to a reduced number of oscillators, although this would reduce the output bandwidth.

This is an hexagonal array of 60 photodetectors, connected to oscillators through simplified RF processors with unity transfer functions. This sensor is described fully in section 4.2. The operating characteristics of a single sensor cell are given in figure 5-4. This graph shows the output of a single cell in each of its two sensitivity modes, over 5 decades of light intensity.

The power consumption of this chip varies with illumination intensity from a low of approximately 20mW, to a peak consumption of 40 mW.

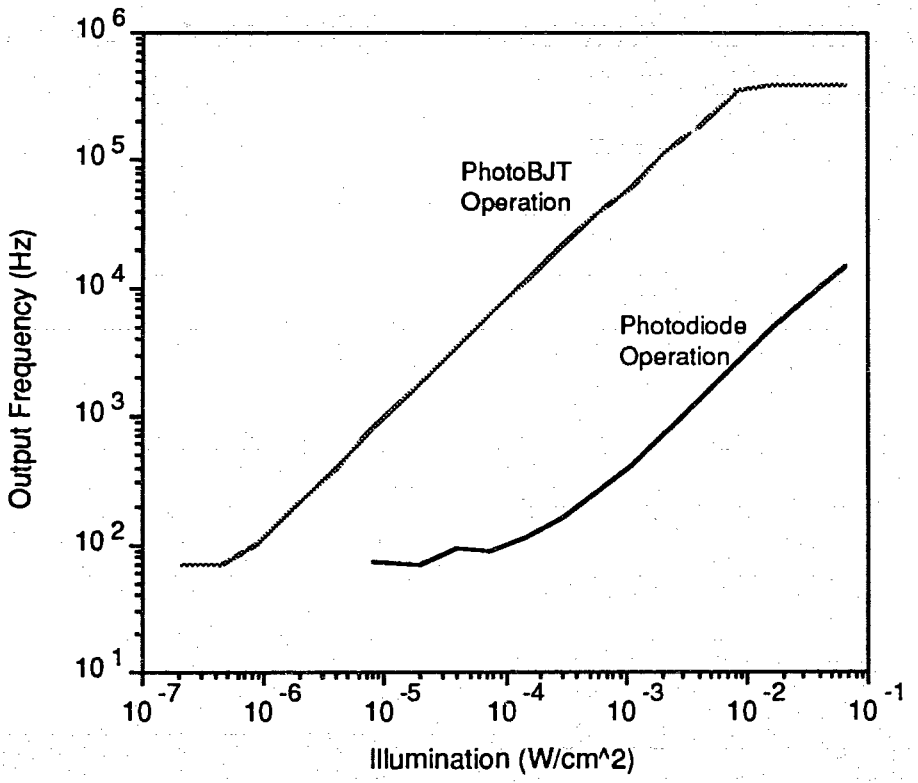

Figure 5-4: Measured output of a single cell of the multi-sensitivity image sensor in photodiode and phototransistor operation modes.

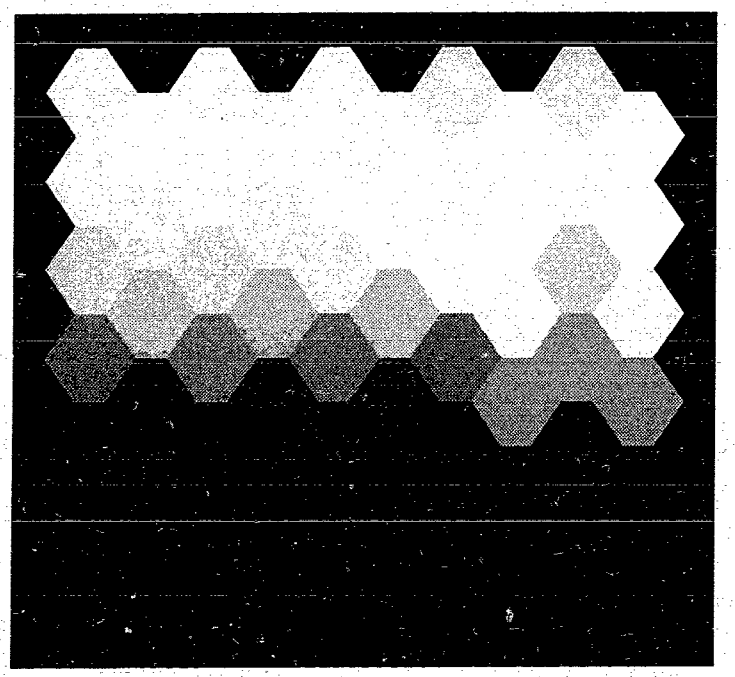

Figure 5-5: Output of the multi-sensitivity image sensor with an illumination gradient across it.

The output of the entire sensor array with an illumination gradient placed across it is given in figure 5-5. The output frequencies have been replaced with grayscale values such that the lowest frequency output is displayed as black, and the highest frequency output appears white.

These results show that the concepts used in this circuit, the multi-mode photodetector, and the current-controlled oscillator operate effectively in an image sensor.

5.3.2. A single edge-detector MERF

This structure performs the Sobel edge-detection method. The function implemented is $\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$ by the circuit given in figure 4-6. The center to center photodetector spacing is 80μm.

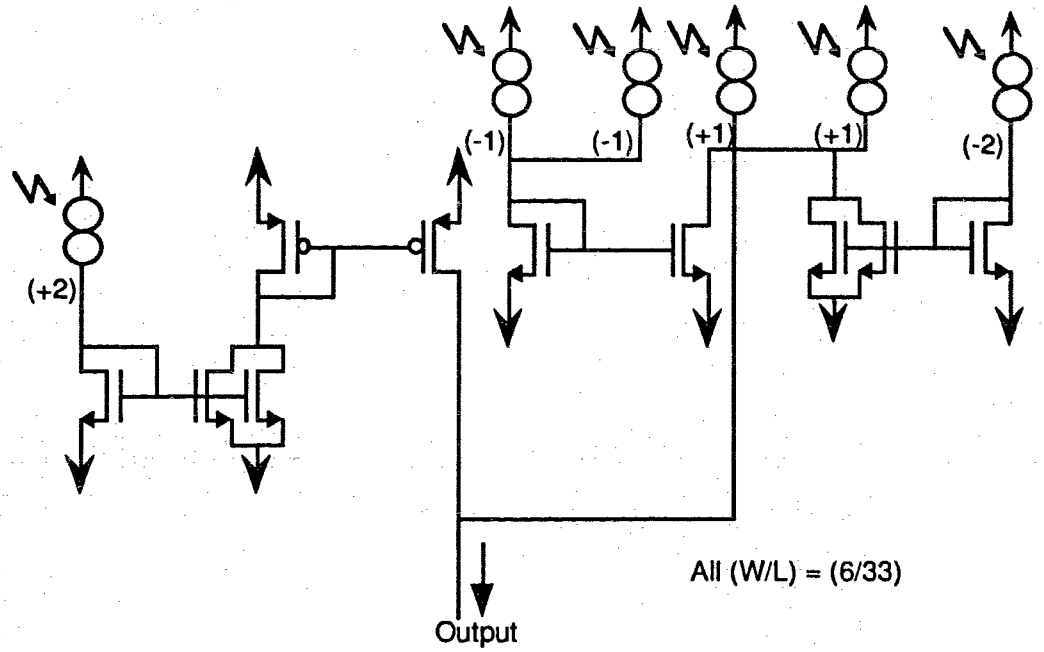

Figure 5-6: Circuit diagram of an edge-detector MERF.

This circuit produces a bipolar output current proportional to the value of the convolution performed on photodetector currents. It is an example of a function which could be useful in a smart vision sensor. A simulated output of Sobel's algorithm applied to an image is given in figure 2-5.

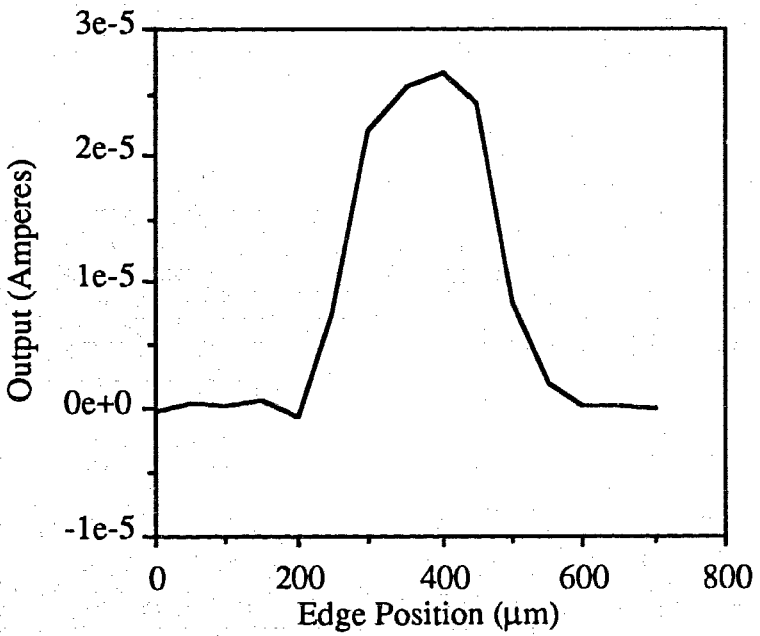

Figure 5-7: Output of 3x3 RF edge-detector as an illumination gradient is moved past it

The output of this MERF versus edge position is given in figure 4-7. In this case, an illumination gradient, or edge, has been moved past the detector, giving its output for each possible edge

location. It can be seen that near zero output is produced when the edge is not incident on the MERF, and a reasonable large output value occurs when the edge lies across the MERF.

This circuit demonstrates that the analog processor unit of a MERF can be effectively implemented in the current domain.

5.3.3. A smart vision sensor providing Edge-Detection

As described in section 4.4, this sensor was designed to perform Sobel's edge-detection method. However, due to a layout error, two signals of opposite sign were switched, and the actual convolutions performed are those shown in figure 5-8.

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Horizontal edge-detector

$$\begin{bmatrix} -1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Vertical edge-detector

Figure 5-8: The convolution algorithms performed by the edge-detector sensor

For the one-dimensional case, these convolutions reduce to $[4 \quad 0 \quad -2]$ and $[2 \quad 0 \quad -4]$. As all of the experiments described here dealt with the one-dimensional case, these values are adequate for establishing expected sensor performance.

The first experiment was performed by setting the sensor's bias input such that a uniform background produced a 60 kHz output signal. The output of a single cell was then examined as fragment of glass patterned with photoresist was moved past it in 20 μm steps. The photoresist transmitted about 1% of the incident light, and was patterned such that a 120 μm wide bar was projected across the chip. Figure 5-9 gives the results, in which positive and negative output is given by variation from the bias value of 60 kHz. Measurements which were repeated differed by less than 2% from the original values. The data shows a clear peak as the light bar crosses the edge-detector, as well as a somewhat regular pattern of secondary peaks. These secondary peaks are caused in part by diffraction of coherent light through the 120 μm aperture, and in part by secondary reflections within the glass screen. Subsequent experiments used non-reflective materials for patterning, which eliminated any internal reflection effects. In addition, the patterning

materials used larger apertures and were positioned much closer to the sensor chip, both of which reduce diffraction effects .

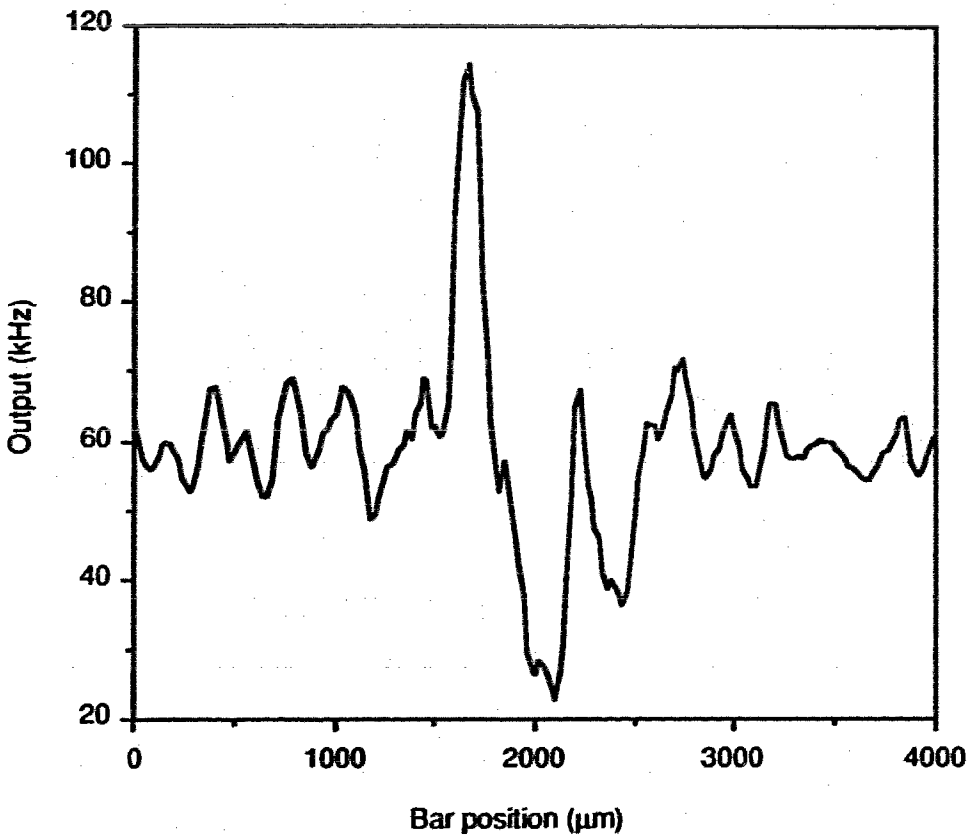

Figure 5-9: Output of an edge-detector cell as a light bar is moved across it

In the next experiment the sensor was biased at 120 kHz to allow a greater output swing. A light bar approximately 2 mm wide was passed across the sensor in 25 µm steps. Figure 5-10 gives the measurement results, along with expected values predicted through numerical simulation. This graph demonstrates reasonably good agreement between the measured and expected results, and the interference fringes observed in the previous experiment have been eliminated.

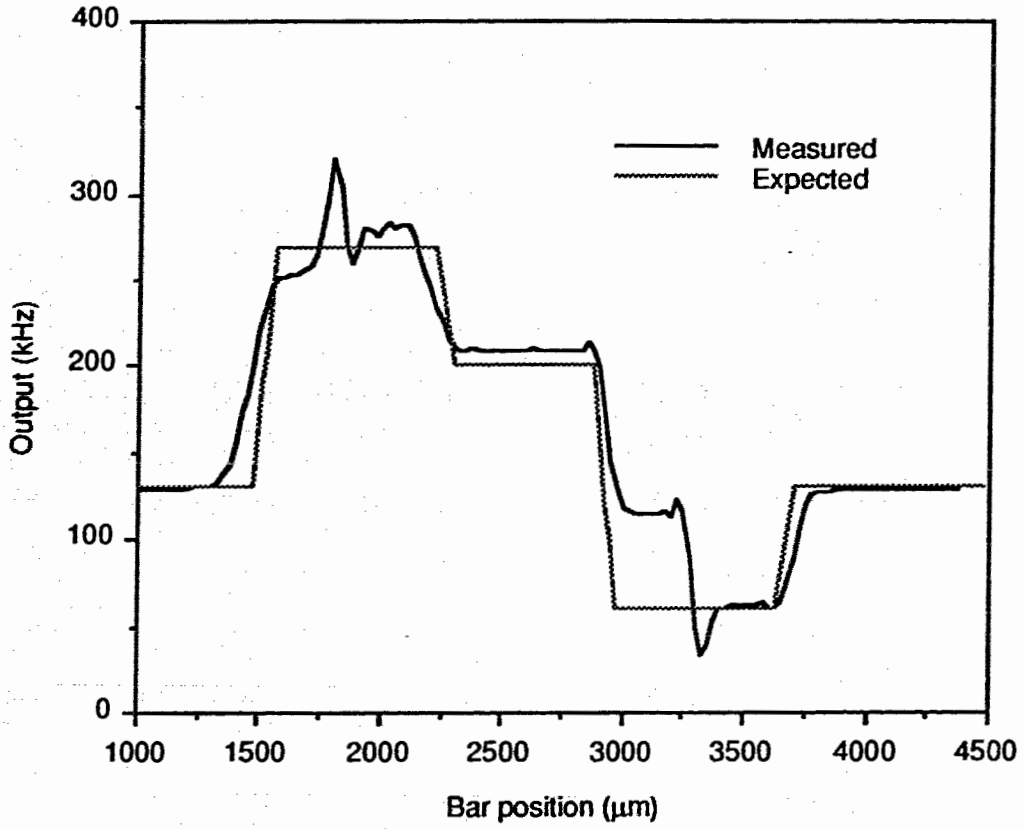

Figure 5-10: Measured and expected output as a light bar is moved across an edge-detector cell

The final experiment used the same setup as the previous one, except that data from four adjacent cells was collected. Figure 5-11 shows the results of these measurements. All four of the cells produce similarly shaped output waveforms, each separated by the distance between the cells (384μm). The variation in signal magnitude of the four cells is due to differences in device characteristics, of which variations in VCO characteristics will dominate.

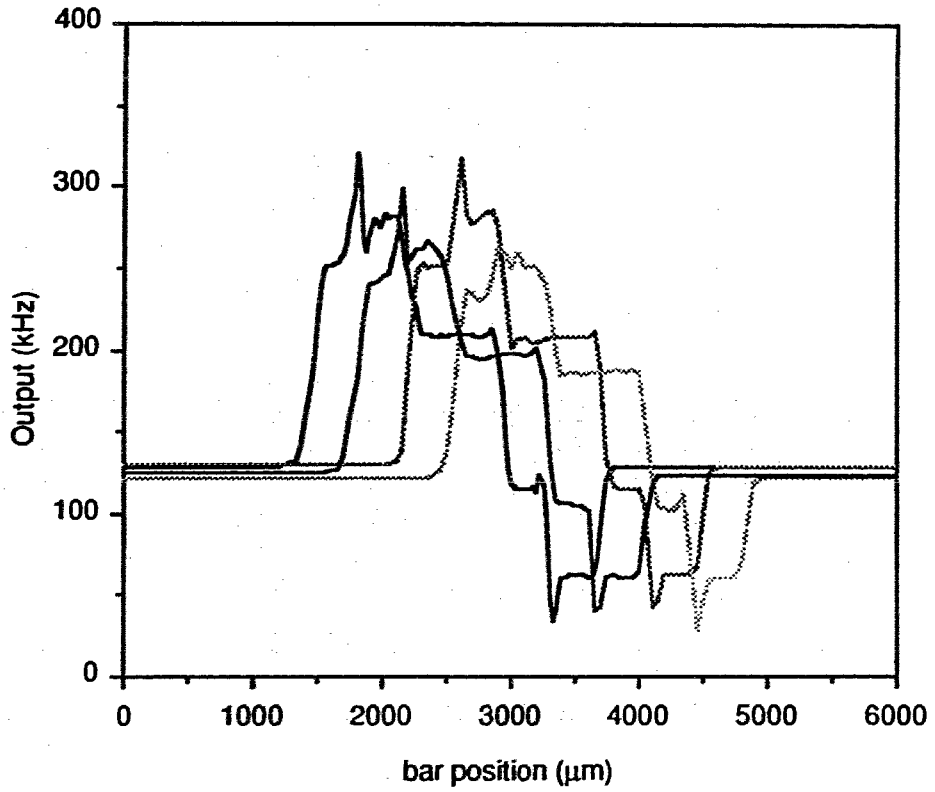

Figure 5-11: Output of four adjacent edge-detector cells as a light bar is moved across them

These are the operations necessary at a pixel level to perform processing over an entire image. This experiment therefore forms a micro-scale version of the simulations given in chapter 2. The same processes, performed over a much larger area would produce a result similar to the edge-detector output shown in figure 2-5.

6. Conclusions

This thesis described the VLSI implementation of receptive fields for focal-plane image processor applications. A microelectronic receptive field (MERF) consists of a set of photodetector inputs, an analog processor to combine those inputs, and an A/D converter to convert the output signal to a useful form. Circuit blocks performing the above functions were designed and characterized, and prototype vision sensors built from these blocks were also tested. Optical experiments were conducted on a custom designed optical test bench necessary for precise sensor measurements. This equipment was used to characterize MERF circuits and sensors over seven decades of light intensity.

The photodetector circuits used in this research were two- or three-mode CMOS devices which can operate as either photodiodes, phototransistors, or Darlington-connected phototransistors featuring different light sensitivity in each mode [15]. This allows a much greater range of light intensity to be processed by the MERF. Three-mode photodetectors are able to compress seven decades of light intensity into three decades of output current, which greatly reduces the task required of the analog processors. This is a new concept in image processors which allows CMOS current mode techniques to be used over a wider range of light intensities.

The analog processor blocks use current-mode circuit techniques to perform weighted-sum operations on photodetector signals. The use of long-channel transistors allows mathematical operations to be performed over four decades of input current while keeping the output error to less than 30%. This accuracy is sufficient if appropriate receptive field functions are chosen for implementation. In particular, MERF functions for vertical and horizontal edge-detection were implemented and performed well.

The A/D conversion technique found to be most appropriate for MERFs was a current to frequency conversion. This is performed by connecting a long p-channel transistor to a pseudo-DTL voltage-controlled oscillator [14]. This circuit allows a current-mode signal to be converted to a frequency-mode signal over at least six decades. The frequency-mode signal provides immunity to noise and crosstalk not offered by analog signals, while requiring a small converter area and only one output line, which are advantages over more conventional A/D converters. After transmission from the sensor, the frequency-mode signal can be easily converted to a fully digital representation by counting output pulses or timing a period, but this will require special hardware on the external computer.

If appropriate algorithms are chosen for MERF implementation, MERF-based sensors show considerable resistance to large variations in device parameters. For example, the sensor designed to perform an edge-detection operation in which incorrect weightings were assigned to its analog processors was still found to operate effectively. In future, vision sensors containing more cells could be measured more easily, by mounting the sensor in a camera body and examining real scenes.

Three micron digital CMOS was found to be sufficient to prove MERF concepts, but it is not the best suited to smart vision sensor design. To achieve both large numbers of photodetector cells and high densities it would be necessary to use a technology with a smaller feature size, and designs optimized for high density. This technology was used because it was a readily available and well characterized.

The MERF circuits and sensors tested in this thesis validate the concepts used, and show that there is potential for further development of vision sensors based on the receptive field paradigm. Due to their high processing speed, low power, and resistance to defects, large area sensors containing thousands of MERFs may become an alternative for smart vision sensors performing peripheral vision functions.

7. Bibliography

[1] A. Fein, E. Szuts "Photoreceptors: their role in vision", New York Press Syndicate of the University of Cambridge, 1982, p.8.

[2] S. Chamberlain and J. Lee, "A Novel Wide Dynamic Range Silicon Photodetector and Linear Imaging Array", IEEE Journal of Solid-State Circuits, vol. 19, no. 1, pp. 41-48, February 1984.

[3] C. Mead "Analog VLSI and Neural Systems", Addison-Wesley Publ. Co., 1989, Ch. 15.

[4] A. Chandna, G. Jullien, W.Miller "Opto-programmable Neural Networks: An Initial Study", Proc. Canadian Conf. on Very Large Scale Integration, CCVLSI'89, Vancouver, Canada, Oct. 22-24, 1989, pp.41-48.

[5] M. Grigoleit, M. Syrzycki "Smart Vision Sensors in Digital CMOS Technology", Proc of the IAESTED Int. Symp. Computers, Electronics, Communication and Control, Calgary, Canada, April 8-10, 1991, pp.10-14.

[6] R. Ginosar, Y. Zeevi "Adaptive Sensitivity/ Intelligent Scan Imaging Sensor Chips", Proc. of Visual Communications and Image Processing '88, SPIE vol. 1001, pp. 462-468.

[7] M. Trembley D. Poussart "M.A.R. : An Early Vision System with Integrated Optics and Processing", Proc. Canadian Conference on Very Large Scale Integration, CCVLSI'89, Vancouver, Canada, Oct.22-24, 1989, pp.33-40.

[8] G. Chapman, M. Syrzycki, M. Parameswaran, P. Juneja "Development of a Standard Transducer Cell Library for ASIC's", Proc. Fourth International Forum on ASIC and Transducer Technology, ASICTT'91, Leuven, Belgium, May 13-15, 1991, pp.179-189.

[9] Analog VLSI Systems for Image Acquisition and Fast Early Vision Processing Wyatt et al, MIT, 1992

[10] A VLSI Pyramid Chip for Multiresolution Image Analysis, Van der Wal and Burt, SRI International, 1992

[11] Analog VLSI Circuits for Stimulus Localization and Centroid Computation, Deweerth, Cal Inst. Tech, 1992

[12] Computing Motion Using Analog VLSI Vision Chips: An Experimental Comparison Among Different Approaches Horiuchi et al, Cal Inst. Tech, 1992

[13] W. K. Pratt, "Digital Image Processing", 2nd Ed., John Wiley and Sons, Inc, 1991.

[14] M. Grigoleit "Integrated Smart Image Sensors in Digital CMOS Technology", M.A.Sc Thesis, School of Engineering Science, Simon Fraser University, 1991

[15] V. Ward, M. Syrzycki, G. Chapman "CMOS photodetector with built-in light adaptation mechanism", Microelectronics Journal, vol. 24, no.5, pp. 547-553, August 1993.

[16] D. Marr, "Vision". W. H. Freeman and Company, 1982.

[17] V. Ward, M. Syrzycki "VLSI Implementation of Receptive Fields for Smart Vision Sensors", Proc of CCECE'93, Vancouver, Canada, Sept 14-17, 1993, pp1184-1187.

[18] G. Chapman, M. Parameswaran, M. Syrzycki "A Wafer Scale Dynamic Thermal Scene Generator", Proc of International Conference on Wafer Scale Integration, San Francisco, USA, Jan 22-24, 1992, pp300-309.

[19] D. Hubel, T. Wiesel, "Receptive fields and functional architecture of monkey striate cortex". Journal of Physiology, vol 195, pp.215-243, 1968.

[20] J. Jones, L. Palmer, "The two-dimensional spatial structure of simple receptive fields in cat striate-cortex cells", Journal of Neurophysiology, vol 58, pp.1187-1211, 1987.