



The University of
Nottingham

**CMOS OPTICAL CENTROID PROCESSOR FOR AN
INTEGRATED SHACK-HARTMANN WAVEFRONT
SENSOR**

Boon Hean Pui, B.Eng (Hons)

*Thesis submitted to the University of Nottingham for
the degree of Doctor of Philosophy*

September 2004

ABSTRACT

A Shack Hartmann wavefront sensor is used to detect the distortion of light in an optical wavefront. It does this by sampling the wavefront with an array of lenslets and measuring the displacement of focused spots from reference positions. These displacements are linearly related to the local wavefront tilts from which the entire wavefront can be reconstructed. In most Shack Hartmann wavefront sensors, a CCD is used to sample the entire wavefront, typically at a rate of 25 to 60 Hz, and a whole frame of light spots is read out before their positions are processed. This results in a data bottleneck. In this design, parallel processing is achieved by incorporating local centroid processing for each focused spot, thereby requiring only reduced bandwidth data to be transferred off-chip at a high rate. To incorporate centroid processing at the sensor level requires high levels of circuit integration not possible with a CCD technology. Instead a standard $0.7\mu\text{m}$ CMOS technology was used but photodetector structures for this technology are not well characterised. As such characterisation of several common photodiode structures was carried out which showed good responsivity of the order of 0.3 A/W. Prior to fabrication on-chip, a hardware emulation system using a reprogrammable FPGA was built which implemented the centroiding algorithm successfully. Subsequently, the design was implemented as a single-chip CMOS solution. The fabricated optical centroid processor successfully computed and transmitted the centroids at a rate of more than 2.4 kHz, which when integrated as an array of tilt sensors will allow a data rate that is independent of the number of tilt sensors employed. Besides removing the data bottleneck present in current systems, the design also offers advantages in terms of power consumption, system size and cost. The design was also shown to be extremely scalable to a complete low cost real time adaptive optics system.

ACKNOWLEDGEMENTS

If I have seen further it is by standing on the shoulders of giants.

-Isaac Newton

I would like to thank Dr. Barrie Hayes Gill for his support, guidance, and not least, patience throughout my research work. I would also like to express my gratitude to Professor Mike Somekh and Dr. Chung Wah See for their help and guidance. To Matt Clark, all your help and brilliant insights are greatly appreciated. To the many research staff, colleagues and technicians who have in one way or another been involved in my work, I thank you.

The research work has been supported by the University of Nottingham, University of Nottingham International Office, University of Nottingham in Malaysia and the Engineering and Physical Sciences Research Council (EPSRC), UK and I would like to thank them for making this possible.

Finally, my sincere thanks go to my friends and family for making this journey bearable, and my deepest gratitude and love to my parents and brother for their belief in me.

TABLE OF CONTENTS

ABSTRACT.....	I
1 INTRODUCTION.....	1
1.1 ADAPTIVE OPTICS	1
1.2 APPLICATIONS.....	3
1.2.1 ASTRONOMY	3
1.2.2 OPHTHALMOLOGY.....	6
1.2.3 BEAM QUALITY CONTROL	7
1.2.4 MICROSCOPY.....	9
1.3 WAVEFRONT SENSING.....	10
1.3.1 SHACK-HARTMANN WAVEFRONT SENSOR.....	10
1.3.2 OTHER WAVEFRONT SENSORS	13
1.4 CENTROID DETECTION	15
1.4.1 LATERAL EFFECT PHOTODIODES (LEP)	16
1.4.2 MULTI-ELEMENT PSD	16
1.4.3 MULTI-ELEMENT PSD PERFORMANCE	18
1.4.4 CENTROID PROCESSING.....	22
1.5 PHOTODETECTION	34
1.5.1 OPTICAL ABSORPTION.....	34
1.5.2 QUANTUM EFFICIENCY AND RESPONSITIVITY	36
1.5.3 NOISE AND PHOTODIODE EQUIVALENT CIRCUIT	39
1.5.4 PHOTODIODE MEASUREMENTS.....	42
1.5.5 TECHNOLOGY AND MATERIALS	44
1.6 PIXEL ARCHITECTURES IN CMOS	51
1.6.1 PASSIVE PIXEL SENSORS (PPS).....	52
1.6.2 ACTIVE PIXEL SENSORS (APS)	53
1.6.3 NOISE REMOVAL AND EXTENDING DYNAMIC RANGE.....	57
1.7 CHAPTER SUMMARY	60
2 CHARACTERISATION OF CMOS PHOTODIODES.....	63

4.3.3	LAYOUT AND TEST BOARD	154
4.4	ASIC CAD ENVIRONMENT AND DESIGN FLOW.....	157
4.4.1	DESIGN AND LAYOUT ISSUES	159
4.5	RESULTS OF CENTROID ASIC	164
4.5.1	POSITION RESOLUTION.....	166
4.5.2	SPEED.....	169
4.5.3	DYNAMIC RANGE.....	171
4.5.4	SCALABILITY.....	173
4.6	CHAPTER SUMMARY	173
5	WAVEFRONT RECONSTRUCTION.....	174
5.1	INTRODUCTION.....	174
5.2	WAVEFRONT DESCRIPTION	174
5.3	DEFORMABLE MIRRORS	175
5.4	WAVEFRONT RECONSTRUCTION	177
5.4.1	MODAL APPROACH	178
5.4.2	ZONAL APPROACH	179
5.4.3	RECONSTRUCTION PROCEDURE.....	180
5.5	COMPLETE AO SYSTEM	183
5.5.1	SPEED.....	186
5.5.2	AREA AND COST.....	191
5.6	CHAPTER SUMMARY	192
6	CONCLUSIONS	193
6.1	DISCUSSION.....	193
6.1.1	DESIGN SPECIFICATIONS	193
6.1.2	CHARACTERISATION OF CMOS PHOTODIODE STRUCTURES	194
6.1.3	DESIGN PROTOTYPING	195
6.1.4	COMPLETE AO SYSTEM.....	196
6.2	FURTHER WORK.....	197
6.3	CONCLUSIONS	198

APPENDICES

CHAPTER 1

INTRODUCTION

1.1 ADAPTIVE OPTICS

Since Galileo pointed his telescope to the heavens some 400 years ago, man has been trying to see further and further into the stars and in greater detail. The fundamental limit of resolving the images is known as the diffraction limit and is governed by the diameter of the lens used. However it was observed that as larger telescope lenses were used, the astronomical images did not get any sharper when the lenses exceeded about 20cm in diameter [Angel 2000]. There was something distorting the images in a seemingly random manner. This was the air around us. Variations in temperature in the atmosphere cause random fluctuations in wind velocity and hence, changes in the refractive index [Tyson 1998]. This leads to distortion in the images obtained. Fortunately there was something we could do about it and it is called adaptive optics.

Adaptive optics (AO), which has been heavily developed over the last 30 years, allows automatic compensation of atmospheric systems. It deals with the control of light in a real-time closed-loop fashion and is made up of three fundamental components, the wavefront sensor, the control computer and a corrector element such as a deformable mirror. The wavefront sensor acts like the eyes detecting light from the object of interest, such as an astronomical object or a satellite, and transducing the intensity information of the wavefront into phase information of the aberration in the wavefront. The control computer then calculates the necessary changes required to correct this aberration, and passes this on to the corrector or the deformable mirror where these changes are made. Figure 1.1 shows the components of a typical adaptive optics system as used in a telescope [O'Byrne 1996]. Often a tilt-tip mirror is used to rapidly remove beam wander in the incoming beam of light while the deformable mirror performs the higher order corrections.

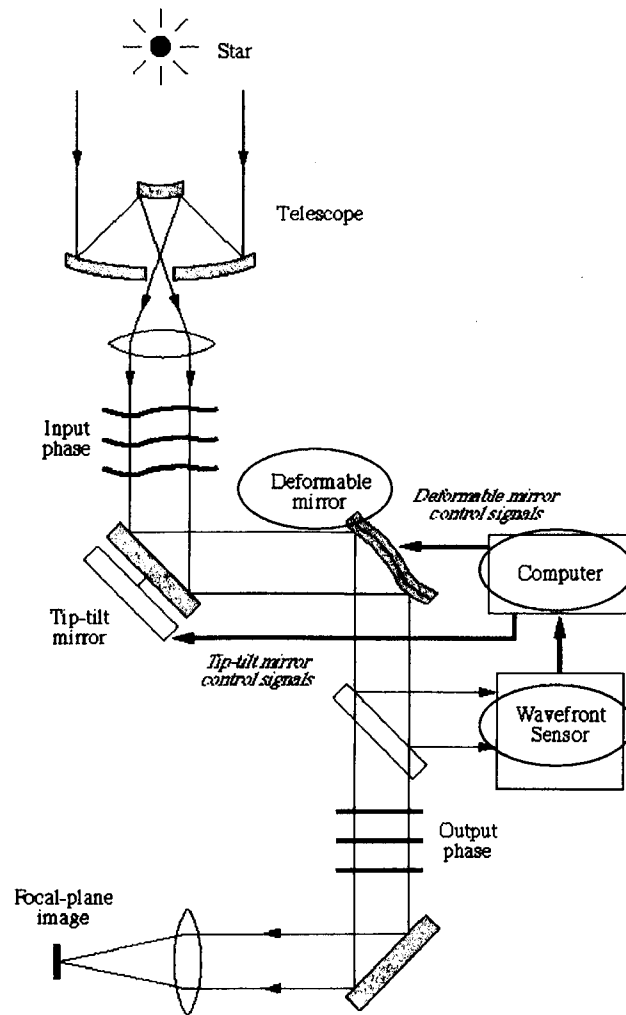


Figure 1.1 A typical adaptive optics system with its fundamental components highlighted [O'Byrne 1996]

The most widely used wavefront sensor in adaptive optics is the Shack-Hartmann [Platt 2001] and currently with most of these systems a CCD is used to sample the wavefront and a frame grabber is used to acquire and digitise the image before it is transferred to a PC for reconstruction of the wavefront. The bandwidth of these AO systems is often limited to some tens of Hz [Nirmaier 2003]. Integration of these systems with processing at the detector level will reduce the bandwidth of the data to be transferred off-chip thus allowing fast real-time wavefront detection and correction and is the topic of this research. Furthermore, integration of the wavefront sensor with wavefront reconstruction will reduce the size and cost of the system even further; realising the concept of a System-on-a-Chip (SoC).

Adaptive optics has traditionally been known for its role in compensating wavefront distortions for astronomical applications. The main reason for this is the cost of the key elements of an adaptive optics system - deformable mirrors, wavefront sensors and control systems requiring high-speed computers. AO systems with a reasonable bandwidth (greater than a few Hz) were extremely expensive, with a component cost of $>£10^5$ [Munro 1999]. Applications of adaptive optics however are not limited to astronomy or defence initiatives and a number of potential applications are surfacing which will benefit from some form of cheap, fast, adaptive optics systems. These range from laser communications, to medical imaging of the retina, to industrial inspection to the development of more efficient lasers as well as underwater imaging devices and better microscopes. Basically adaptive optics can be used wherever light passes through a distorting medium. Section 1.2 will cover some of the application areas where an adaptive optics system can be applied. In Section 1.3 the concept of wavefront sensing is described paying particular attention to the mechanics of a Shack-Hartmann wavefront sensor and how integration will remove the bottleneck in traditional CCD systems. The process of detecting a centroid which is a fundamental component of a Shack-Hartmann wavefront sensor is covered under Section 1.4. Section 1.5 and 1.6 will then review the theory behind photodetection and the possible implementation structures for this. Section 1.7 summarises the chapter while Section 1.8 will detail the layout of the rest of the chapters.

1.2 APPLICATIONS

In addition to system integration, the development of new low-cost technologies such as Micro-Opto-Electro-Mechanical Systems (MOEMS), liquid crystal wavefront correctors and micromachined deformable mirrors [Anderson 1999, Hatcher 2001, Vdovin 1997] will further open up new areas of applications. Some of the key applications for an adaptive optics system are discussed in the following subsections.

1.2.1 ASTRONOMY

The field of astronomy gave birth to the technique of adaptive optics and is widely used in correcting the imaging capabilities of ground-based telescopes. The image

quality of all ground-based telescopes suffers from atmospheric turbulence, which is the fundamental reason for placing the Hubble Space Telescope in space and the fact that ground-based telescopes are built high in the mountaintops with clear air regions. The spatial resolution of uncompensated telescopes can be more than 10 times better on mountains than at sea level [Tyson 2000].

The structure and statistics of turbulence as well as its corresponding effects can be described by a model by Kolmogorov [Tyson 1998]. The effect of this turbulence is to cause high spatial frequency beam spreading, low spatial frequency beam wander, and intensity variations which limits the ability of telescopes to resolve fine details. The level of turbulence at a particular site can be described by a parameter introduced by Fried called the Fried coherence length, r_0 , [Fried 1965] and is the maximum diameter of the aperture that can be used for collection of the wavefront before atmospheric distortion seriously limits its performance. This parameter defines the limit of the achievable resolution without compensation, as shown in Figure 1.2 by the sketch of the typical point spread function of a star being imaged by an astronomical telescope. The Fried coherence length is $\sim 2\text{cm}$ under poor seeing conditions to $\sim 20\text{cm}$ under good seeing conditions [Mansell 2000]. Figure 1.3 shows the uncompensated and compensated image of a binary star as taken at the Starfire Optical Range [Air Force Research Laboratory Directed Energy Directorate 1997]. With compensation, the image halo or beam spread, as in Figure 1.2, has been corrected for and the two distinct stars of the binary star k-Peg can be discerned.

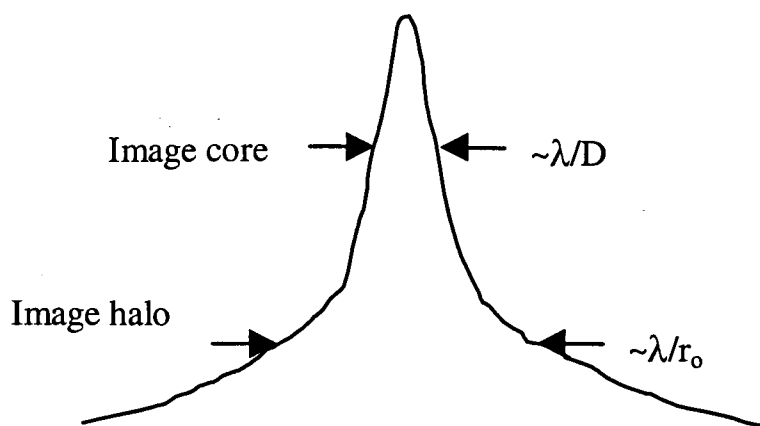


Figure 1.2 Beam spread due to atmospheric turbulence limits the resolution for an aperture of diameter D

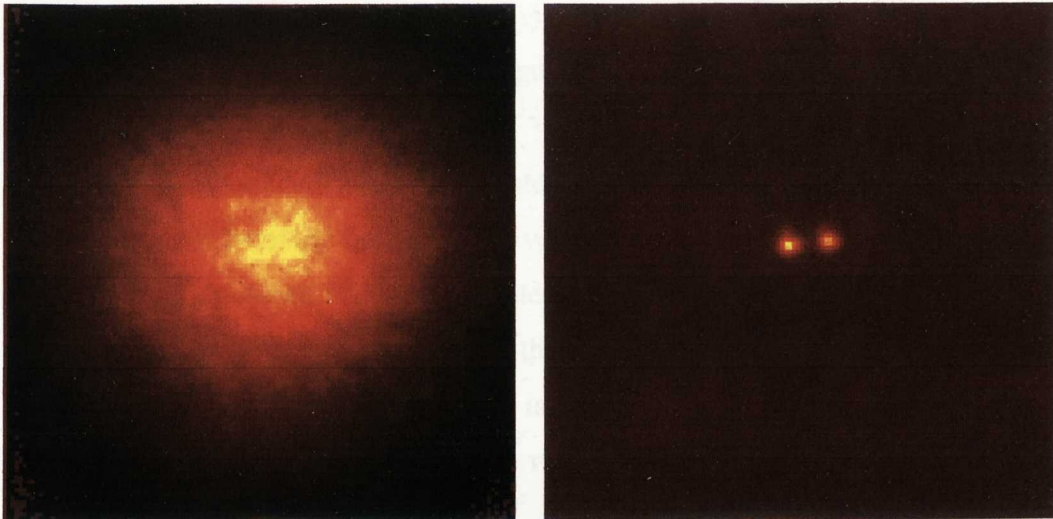


Figure 1.3 Uncompensated (left) and compensated (right) images of the binary star k-Peg as taken by the Starfire Optical Range [Air Force Research Laboratory Directed Energy Directorate 1997]

For adaptive optics to work, the aberrations that are caused by the turbulence have to be measured faster than they can change. This is given by the Greenwood frequency f_G , which is strongly dependent on the velocity of the wind, and can range from tens to hundreds of hertz under fair viewing conditions [Tyson 2000]. Another important factor to consider in the design of atmospheric adaptive optics systems is the isoplanatic angle θ_0 , which determines the maximum angle that we can look away at our object point and still measure the correct wavefront [Tyson 2000]. Because the isoplanatic patch for the atmosphere is so small, only a tiny fraction of the sky will be near suitably bright stars that can serve as reference beacons. A way of overcoming this is to produce artificial guide stars using powerful lasers to illuminate the sky. Two types of artificial guide stars exist. One using Rayleigh scattering of ultraviolet or visible light illuminates the sky at a height of 5 to 15 kilometres in the atmosphere. The other uses resonant scattering of light from a layer of sodium atoms that sits in the upper mesosphere at about 90 to 100 kilometres in altitude. The second scheme has the advantage of putting the reference beacon higher, thus sampling a larger portion of the path of light from a celestial object in space to a telescope on Earth [Olivier 1999]. The disadvantage is that it is more expensive and requires laser at a specific wavelength of 589nm for excitation of sodium atoms. An emerging technique called Multi-Conjugate Adaptive Optics (MCAO) which uses several guide stars and

wavefront sensors allows the field of view to be extended and could overcome the disadvantage of having to use artificial guide stars [Berkefeld 2001].

Besides atmospheric imaging, underwater imaging and fluid mechanics [Neal 1993] will also benefit from the field of adaptive optics. And just as how the advancement of lasers, imaging devices and optical materials has pushed the frontiers of the field of adaptive optics for astronomy, the theories and techniques developed for the correction of atmospheric turbulence is directly applicable to that of other non-astronomical applications enabling their rapid development.

1.2.2 OPHTHALMOLOGY

Imperfections in the cornea and the eye leads to refractive errors which causes image blurring. This gives rise to long and short sightedness which needs correction with glasses or contact lenses. It is now possible to perform these corrections through eye surgery. Laser-Assisted In-Situ Keratomileusis, or LASIK as it is commonly known, is the procedure of reshaping the cornea with a laser beam to correct for these errors. Typically LASIK corrects for low-order aberrations and in the course of reshaping the cornea to correct these, refractive surgeries can inadvertently increase higher-order aberrations. A wavefront sensor can be used to measure these higher-order aberrations and to allow doctors to have a more detailed and quantitative view of the topography of the cornea before it is operated upon. The first commercial ophthalmic Shack-Hartmann aberrometer, the Complete Ophthalmic Analysis System (COAS), manufactured by WaveFront Sciences, Inc. became available in early 2000 and incorporates a CCD-based Shack-Hartmann wavefront sensor [Salmon]. The human eye is a non-static optical system and the corrections need to be done at a bandwidth of at least several hundred Hz [Nirmaier 2003]. Real-time wavefront correction in the human eye will also allow a better diagnosis of eye diseases like the common glaucoma and will allow the development of the next generation of customised wavefront-guided contact lenses [Thibos 2003].

1.2.3 BEAM QUALITY CONTROL

The beam quality and output power of lasers can be degraded by optical aberrations within the laser resonator [Kudryashov 2002]. Adaptive optics allow the correction of these aberrations using either intracavity or extracavity control of the beam. Intracavity control involves using an adaptive mirror as one of the end mirrors of the laser resonator as shown in Figure 1.4.

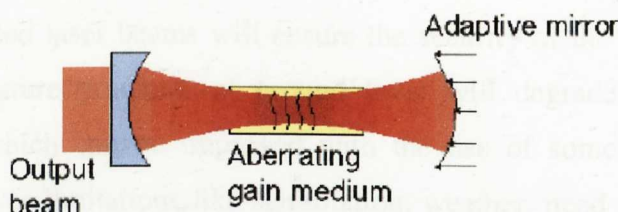


Figure 1.4 Intracavity laser beam correction [Applied Optics Group; Imperial College]

Intracavity control is able to influence the geometry of the output modes and stabilise the output energy. Also the output parameters of the beam can be changed without the need to reconstruct the entire cavity or altering the power supply block which is costly and time consuming. Intracavity beam control will also aid in the generation of beams with a super-gaussian distribution [Cherezova 1997], which has lower side lobe intensities than a typical Gaussian beam and consequently, a reduction in higher spatial frequencies and a higher intensity profile. This is very attractive for industrial applications.

For lower orders of aberration, extracavity control is easier to implement. Extracavity correction involves performing correction outside the cavity of the resonator. Extracavity control will allow beams to be accurately focused on a sample as well as maintaining beam quality over long distances. For instance, extracavity control will also be used on the Laser Interferometer Gravitational-Wave Observatory (LIGO) system for the detection of gravitational waves [Mansell 1999]. Gravitational waves are produced by events such as collapses, explosions or collisions of celestial objects and its observation will allow a better view of the universe and its beginnings. They are less attenuated than electromagnetic waves like radio waves but the predicted

magnitudes of such waves are extremely small. As such very sensitive means of detection are necessary to detect these waves and typically laser interferometry with large kilometre sized arms is used. It is necessary to maintain the beam quality and its coherence over the length of the arms making adaptive optics necessary.

Another field that has received a lot of attention lately is that of free space optical communications which will allow high-speed transmission of large bandwidths of data in the order of gigabits and without the need for cables [Weyrauch 2002]. The use of highly collimated laser beams will ensure the security of the communication. Air flow and temperature gradients at ground level will degrade the quality of the communication which can be improved with the use of some form of wavefront correction. However limitations like scintillation, weather, need for line-of-sight and sun-blindness needs to be addressed. In free-space optoelectronic interconnects, a key challenge is maintaining precise alignment of the opto-mechanical system, which requires high tolerances of optical components and opto-mechanics. Correcting any misalignment dynamically using adaptive optics will help reduce the specifications and tolerance requirements of the opto-mechanical system and improve the cost/performance trade-off [Gourlay 2000].

In laser fusion, pulse shaping and precision focus of the high-energy lasers involved will ensure the quality of the laser pulse as it goes through the amplification process and will allow safe testing of nuclear devices as well as aid fusion energy research [Metrologic Instruments Inc.]. Industrial applications of laser beam control include laser welding and cutting [Haferkamp 1993]. For pulse piercing technology using deformable mirrors, the piercing time can be reduced and for laser cutting technology the thickness of high-quality cutting can be increased. Adaptive optics was used to laser cut thicknesses up to 16 mm in mild steel without decrease of the cut surface with a thickness increase by maintaining focus of the laser beam [Geiger 1996]. Commercially, adaptive optics can also be applied to optical data storage such as in CD drives.

1.2.4 MICROSCOPY

In microscopy, an adaptive optical system can aid in the sensing and correction of aberrations due to imperfections and misalignment in components and the mismatch of refractive indices between the media and the sample to be observed [Booth 2002a, Booth 2002b]. For instance, in a confocal microscope a pinhole is used to block out light from the specimen that are not within the focal plane. This allows strong rejection of multiple scattered light and gives significant improvements in resolution over conventional microscopes [Diaspro 2001]. Its principle is illustrated in Figure 1.5. By scanning the specimen a full 3D image of the specimen can be built up. However, even small amounts of spherical aberration are enough to produce considerable degradation of the imaging performance in the depth direction. Also, confocal microscopes are often operated in reflection because aberrations caused by the refractive index structures within the specimen make imaging in transmission difficult. This results in a loss of phase information only available in transmission. The use of an adaptive optical system would overcome this and allow the compensation of the aberrations introduced by the specimen as well as any misalignment of optical components in the microscope [O'Byrne 1999, Sheppard 1991].

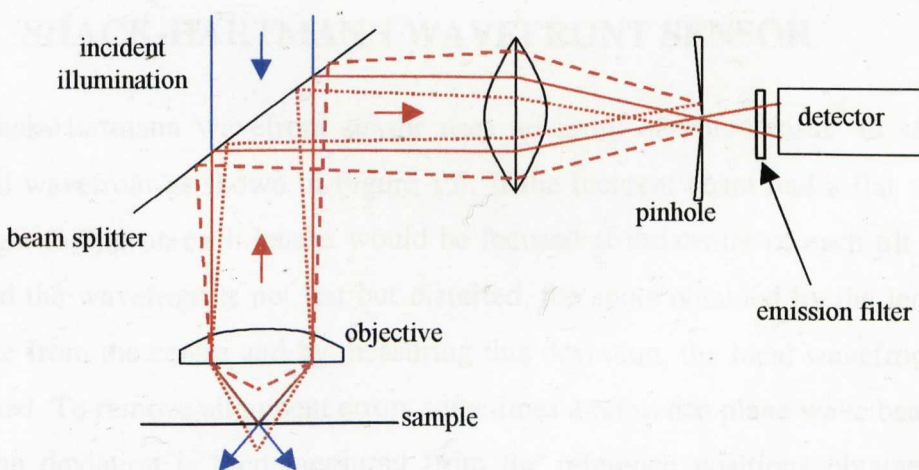


Figure 1.5 Principle of the confocal microscope

In multiphoton fluorescence microscopy, a point source is scanned through the sample volume and the resulting fluorescence is imaged. The localised excitation provides high spatial resolution, efficient background rejection, reduced photobleaching and

increased penetration depth in specimens compared to conventional microscopes. It allows the elimination of the confocal aperture and hence does not limit the number of photons detected. However specimen induced aberration again reduces the achievable resolution as well as increases the necessary laser power to achieve imaging. Aberration correction using feedback will allow the imaging depth to be extended and increase the efficiency of the system [Marsh 2003].

1.3 WAVEFRONT SENSING

As mentioned previously, an integral part of an adaptive optics system is the wavefront sensor which quantitatively measures the amount of aberration present in the wavefront. Wavefront sensing can be either modal or zonal [Tyson 1998]. In modal sensing the wavefront is expressed in terms of coefficients of the modes of a polynomial expansion each representing one of the known aberrations (e.g. tip, tilt, defocus, astigmatism, coma etc.), whose magnitudes are measured separately. Current modal sensors can only sense low-order aberrations. In zonal sensing the wavefront is divided into a number of zones, and the slope or the curvature of the local wavefront is measured in each zone. The Shack-Hartmann wavefront sensor is one such sensor.

1.3.1 SHACK-HARTMANN WAVEFRONT SENSOR

A Shack-Hartmann wavefront sensor uses an array of microlenses¹ to sample the optical wavefront as shown in Figure 1.6. If the incident beam had a flat wavefront, the light falling on each lenslet would be focused at the centre of each tilt sensor. If instead the wavefront is not flat but distorted, the spots obtained by the lenslets will deviate from the centre and by measuring this deviation, the local wavefront tilts are obtained. To remove alignment errors sometimes a reference plane wave beam is used and the deviation is then measured from the reference positions obtained [Tyson 1998].

¹ The Shack-Hartmann wavefront sensor is an improvement over the basic Hartmann test which uses an array of hard apertures instead of the lenslet array. The Shack-Hartmann samples the entire wavefront and has the advantage of better photon efficiency. The disadvantage is in the cost of the microlenses and the difficulty in the optical alignment.

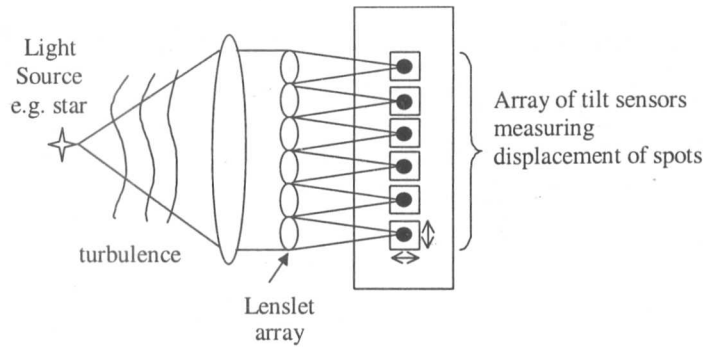
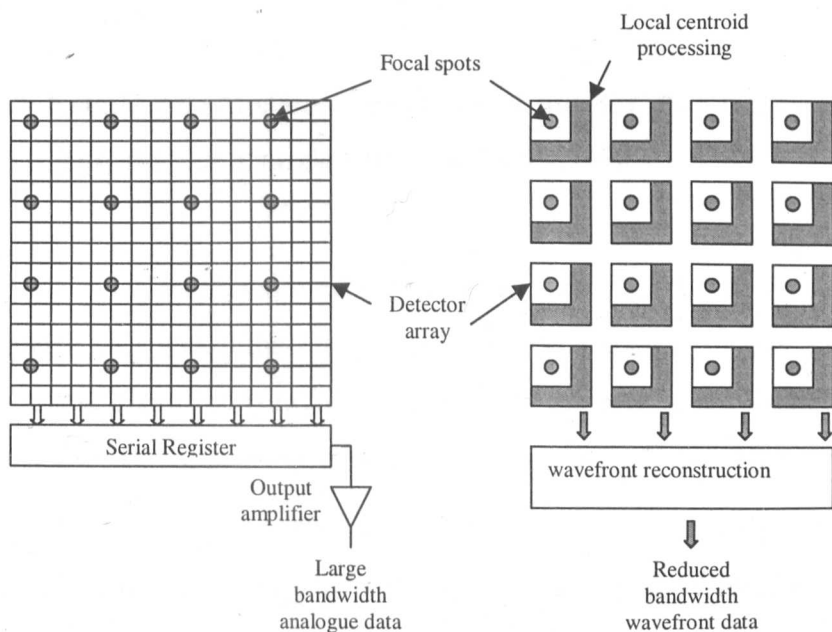


Figure 1.6 Shack-Hartmann wavefront sensor

Traditional CCD systems for Shack-Hartmann wavefront sensing use the CCD to sample the entire wavefront and entire array of spots need to be read out before they are processed leading to a data bottleneck. This bottleneck is illustrated in Figure 1.7 in comparison with our proposed system, where each local wavefront tilt is measured by a local tilt sensor with its own detector array and local centroid processing. The parallel readout and processing of the raw data into reduced bandwidth centroid data will allow faster frame rates to be achieved. In addition, the array of tilt sensors can be linked to a matrix processor to reconstruct the estimate of the complete wavefront. Once calculated, the reduced bandwidth wavefront data can then be transferred off-chip. Hence, as a result of parallel processing, the data rate is independent of the number of tilt sensors employed.



(a) Traditional CCD systems

(b) Wavefront sensing with local centroid processing

Figure 1.7 Integration of on-chip centroid processing to remove data bottleneck

Assuming that at each tiny local portion of the wavefront the only aberration is the tilt, the local wavefront tilt can be linearly related to the displacement of the centroid position from its centre or reference position, as illustrated in Figure 1.8 and given by:

$$\text{Tilt} = \frac{dW}{dx} = \frac{\Delta x}{f} \quad (1.1)$$

where x is the displacement of the centroid and f is the distance of the subaperture from the focal or measurement plane and $f \gg$ maximum dW over the entire subaperture. From these local wavefront tilts, the entire wavefront can be reconstructed and this will be covered further in Chapter 5.

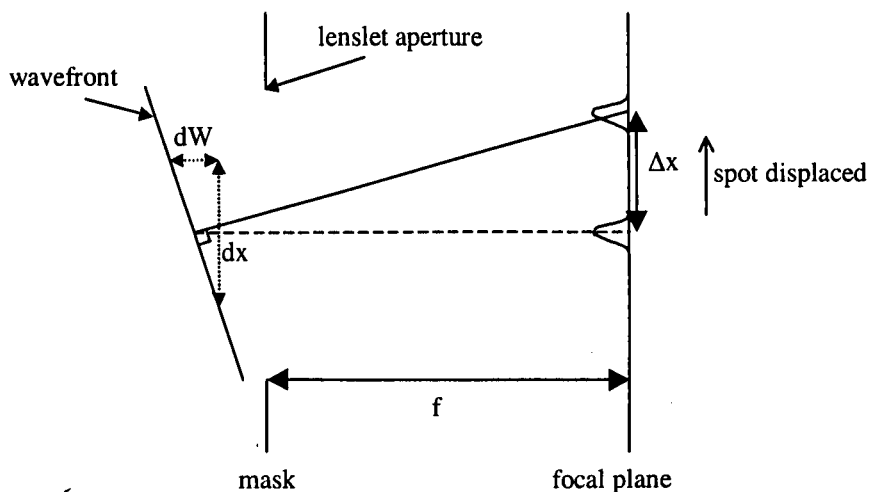


Figure 1.8 Relationship between local wavefront tilt and displacement of the centroid (for a single lenslet of Figure 1.6)

The size of the subaperture required for correct measurement of the wavefront is given by the distance over which the subaperture can pass a coherent beam, i.e. over which the optical phase distortion is highly correlated. In the case of atmospheric optics, this is given by Fried's coherence length, r_0 , which has a dependence of $\lambda^{6/5}$ with wavelength, λ , and as such astronomical adaptive optics is usually performed in the infrared. Another factor to consider is the number of degrees of freedom required, that is, the number of actuators in the wavefront corrector, and this is closely related to the number of subapertures required. There should be roughly one actuator corresponding

to each patch of sky equal in size to Fried's coherence length [Mansell 2000], so the number of subapertures required, N , will be:

$$N \sim (D/r_0)^2 \quad (1.2)$$

where D is the size of the entire pupil or wavefront. Hence the longer the wavelength the lower the complexity.

The Shack-Hartmann wavefront sensor is simple to construct, robust with no moving parts, compact and is by far the most common and established wavefront sensor. It offers high accuracy, reproducibility and a wide dynamic range [de Lima Monteiro 2002]. The work done in this thesis focuses on the use of a Shack-Hartmann wavefront sensor because of the high level of integration possible but it is by no means the only option open to designers of adaptive optic systems. The following section will briefly describe the other wavefront sensing techniques available and why these are less suitable for the purpose of this work.

1.3.2 OTHER WAVEFRONT SENSORS

The choice of wavefront sensor is very much dependent on the application. Several other common wavefront sensing techniques include interferometers, phase diversity, curvature wavefront sensors and the relatively new pyramid wavefront sensors. Interferometric methods include the lateral shear interferometer which measures the wavefront slope or the first derivative of the phase and the point diffraction interferometer which measures the phase of the wavefront directly [Tyson 1998]. The lateral shear interferometer works by splitting the beam and introducing a lateral shear on one arm and measuring the difference or interference between these two beams. The point diffraction interferometer also generates its own reference but does this by capturing a small part of the beam and expanding this as a plane wave reference. In general, interferometric methods of wavefront sensing require monochromatic, highly coherent sources to work making them unsuitable for certain applications such as astronomical imaging. They are also vibration sensitive, expensive and wavefront extraction is complicated so real-time analysis is difficult. Unlike the Shack-Hartmann, they suffer from phase ambiguity of phases exceeding 2π and they cannot be used for pulsed sources. However, the point diffraction interferometer for example,

performs better than the Shack-Hartmann wavefront sensor in strong scintillation where phase discontinuities make the use of linear reconstruction difficult.

Another technique called phase diversity retrieves the phase from the analysis of two simultaneous images, one in-focus and the other defocused [Jefferies 2002]. This method has the advantage of not having any particular requirement on the optical beam and can be used with greatly extended sources. But the algorithm is non-linear and hence slow so it is often used as a post-processing technique for measuring aberrations and deblurring images.

The curvature wavefront sensor works by measuring the irradiances at two planes at the same distance but on opposite sides of the focal point [Rodier 1998b]. By solving the irradiance transport equation that relates the irradiances on the two planes, the curvature of the wavefront can be obtained. They have the advantage of being cheaper and more sensitive than the Shack-Hartmann. However, the equation is non-linear and its solution is not trivial [de Lima Monteiro 2002], and they are difficult to implement for systems that require large number of degrees of freedom such as in highly segmented telescopes [Jefferies 2002] and are only suited for low order systems. On highly segmented mirrors they could still be used for the tip/tilt alignment or the alignment of the primary mirror segment. In confocal microscopy, curvature sensing does not work well due to strong diffraction effects.

Pyramid wavefront sensors work by focusing the wavefront onto the central vertex of a glass pyramid which splits the beam into its four parts with the four edges acting like four Foucault knife edge tests and the images contain information of the aberration present in the wavefront. Pyramid wavefront sensors offer higher sensitivity than Shack Hartmann wavefront sensors and also allow variable gain which makes them useful in wide field adaptive optics. However the fabrication of the pyramids is no simple matter. The quality of the edges between the faces of the pyramids and the size of the roof at the apex of the pyramid are critical [Canadian VLOT Working Group 2003]. Manufacturing of single pyramid structures using the classical figuring and polishing technique is a time consuming process and the production of a large number of identical pyramids is still being developed.

A new development, the hybrid curvature and gradient sensor enables one to obtain information on the local curvature as well as the local wavefront tilts or gradients while maintaining the simplicity of the Shack-Hartmann wavefront sensor [Paterson 2000]. The sensor uses quad cells placed at the foci of an array of astigmatic lenslets and the curvature signal is obtained from the difference of the pair of diagonal elements of the quad cell. Experimental results of this design have yet to be published.

Several factors make the Shack-Hartmann wavefront sensor the choice for an integrated wavefront sensor not least of which is that it requires only simple processing in finding the spot positions which can easily be integrated at the sensor level to reduce the amount of data to be sent off chip. Lower resolution imagers can be used in finding the centroid position, instead of obtaining complicated fringe data in interferometric methods for example. The linear relationship between the spot displacement and the local wavefront tilt also means a simple linear reconstruction technique can be used. This translates to fast real-time correction of wavefront aberrations. Integration could also lead to a reduction in size and costs in many applications.

1.4 CENTROID DETECTION

The fundamental process performed in a Shack-Hartmann wavefront sensor is the detection of the optical centroids. Optical position-sensitive detectors (PSDs) detect the centroid position of a light spot projected on their surface and can be divided into two broad categories namely lateral-effect PSDs and multi-element PSDs [Sharman 2002]. Besides adaptive optics, optical position sensing has numerous commercial, industrial and laboratory applications. In the manufacturing process position-sensitive devices are used to characterize lasers, align optical systems, and calibrate and analyze machinery. PSDs are also used as triangulating sensors in various domestic appliances for switching the appliances on and off by detecting the presence of a body. They are also used in the feeding of paper in fax machines and printers and in the reading of disc tracks in CD players.

1.4.1 LATERAL EFFECT PHOTODIODES (LEP)

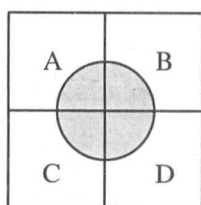
LEPs, as shown in Figure 1.9 (c), consist of a single resistive sheet formed by a p-n junction. The photogenerated charge carriers in the silicon move towards the appropriate electrode where the photocurrent at each electrode is inversely proportional to the distance between that electrode and the centroid of the incident light beam. Lateral effect PSDs are usually operated under reverse bias. Different geometries and positioning of the electrodes in lateral effect PSDs will give rise to tradeoffs in terms of linearity, sensitivity and resolution [Wang 1989].

A lateral effect PSD requires large uniform sheet resistance for linear operation, which is not readily available in a standard CMOS process making integration with circuitry difficult [de Lima Monteiro 2002] and hence unsuitable for the aims of this work. However, the performance of the LEP shall be compared with other PSD structures in Section 1.4.3.4.

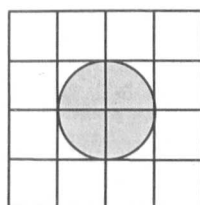
1.4.2 MULTI-ELEMENT PSD

Multi-element PSDs consists of separate active areas. The simplest two-dimensional multi-element structure would be the quad cell, shown in Figure 1.9 (a). Larger structures are termed multi-pixel arrays. Like LEPs, quad cells have simple readout schemes. The position of the incident spot is determined by the comparison of the signals from the four quadrants as illustrated in Figure 1.9(a) and described below:

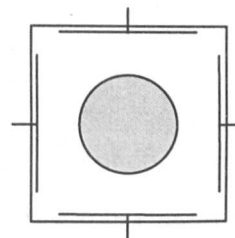
$$\begin{aligned} x &= [(B+D) - (A+C)] / [A+B+C+D] \\ y &= [(A+B) - (C+D)] / [A+B+C+D] \end{aligned} \quad (1.3)$$



(a) Quad cell



(b) Multi-pixel array



(c) Lateral Effect Photodiodes (LEP)

Figure 1.9 Different position sensitive detector (PSD) structures

For multi-pixel arrays, the position of the spot can be found either by simply finding the maximum signal in the array, and this is termed binary position sensing [Makynen 1998], or by finding the normalized first order moment of the signals of all the pixels in the array [Horn 1986] and this is given by:

$$C(x) = \frac{\sum r_{xn} I_n}{\sum I_n}; \quad C(y) = \frac{\sum r_{yn} I_n}{\sum I_n}; \quad (1.4)$$

where r_{xn} is the displacement in the x-direction of pixel n

r_{yn} is the displacement in the y-direction of pixel n

I_n is the light (photocurrent) level of pixel n

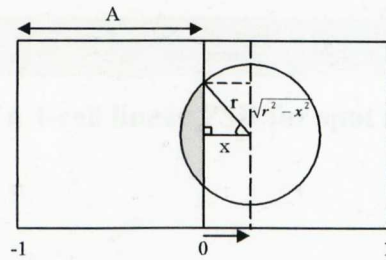
This essentially finds the weighted average of the different elements. Finding the weighted average offers the advantage of subpixel accuracy at the expense of more complicated processing. Higher order moments can also be found. The second order moment for example can be used to give the axis of least inertia or orientation of the imaged object [Standley 1991]. In the field of computer vision, the centroid and higher order moments are often used for character and object recognition [Cash 1987, Dudani 1977, Low 1998] as well as image compression [Karadimitiou 1998].

Other methods for computing a centroid from multi-pixel arrays also exist, such as the median-sum method used by the students of Johns Hopkins University [Dickinson 2003] for tracking objects, which was motivated by the Robocup competition where robots are built to play soccer. In this method, the row and column currents are summed and the median of these currents represent the centroid. This technique has the advantage of not requiring complex mathematical processing but is only accurate when a large number of pixels are used. Also this technique does not provide subpixel accuracy.

Another technique for determining the centroid of an object is by fitting a suitably defined PSF to a series of images [Fosu 2004]; a Gaussian function for stellar images for example. This method can only be used when the image is spread over more than four pixels but is said to give better accuracy than the moment analysis method. However, it is computationally intensive and complex making integration and real-time operation difficult.

1.4.3 MULTI-ELEMENT PSD PERFORMANCE

Pixelated position sensitive devices are typically evaluated in terms of linearity, positional sensitivity and positional range. These are affected by the detector size², the cell density i.e. the number of cells for a given detector size, the gap between the cells and the intensity profile of the spot. Consider a uniform circular beam incident on a bi-cell, which is basically a 2-cell device which measures position in 1 dimension. The results of sweeping the beam of varying sizes across the cells are simulated and shown in Figure 1.10. This case is then extended to a 4-cell linear array and the results are shown in Figure 1.11. Note that for the simulations, truncation of the beam in the vertical direction is ignored. That is the height of the cells are infinitely long and the problem is limited to one dimension. These results shall be discussed in terms of spot size, cell density and beam intensity profile.



$$\text{Shaded area} = r^2 \cos\left(\frac{x}{r}\right) - x\sqrt{r^2 - x^2}$$

where r is the radius of the beam and x is the lateral displacement

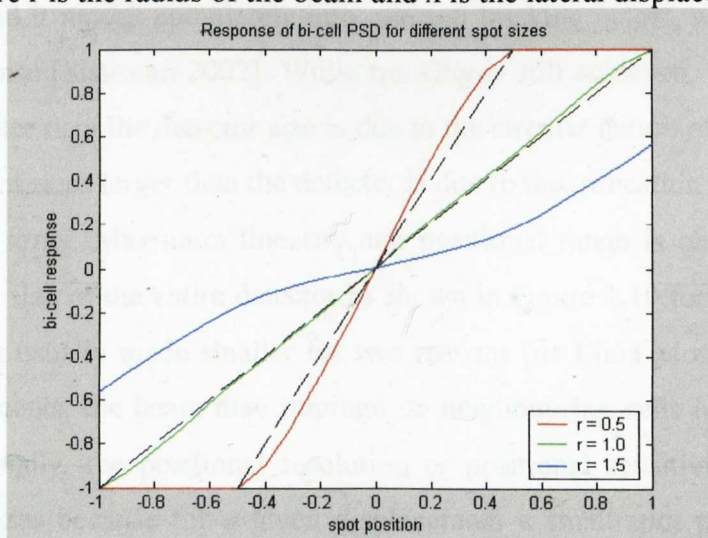


Figure 1.10 Response of a bi-cell PSD for spot sizes of different radius, r

² Detector size is the size of the entire array whereas the cell size is the size of a single element or pixel in the array.

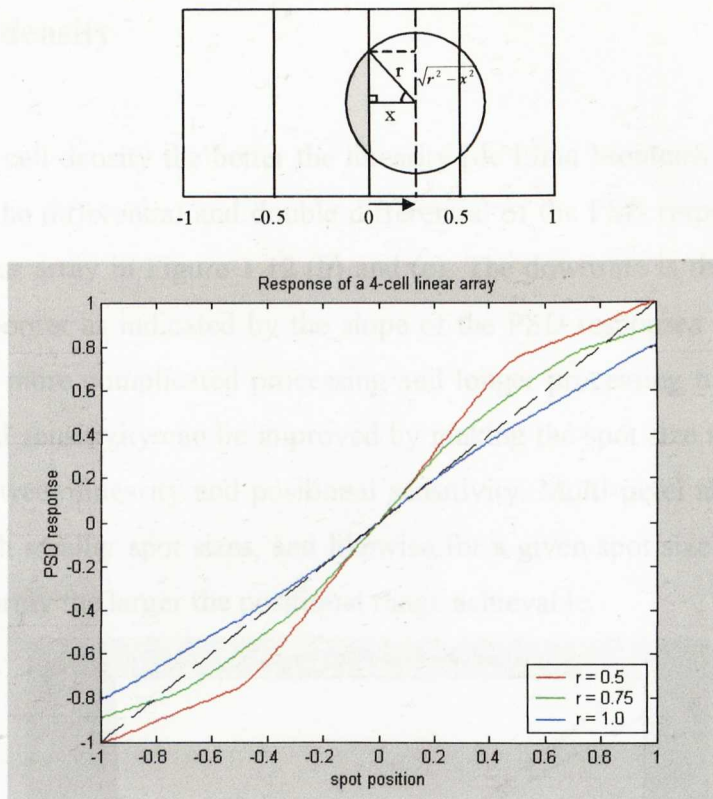


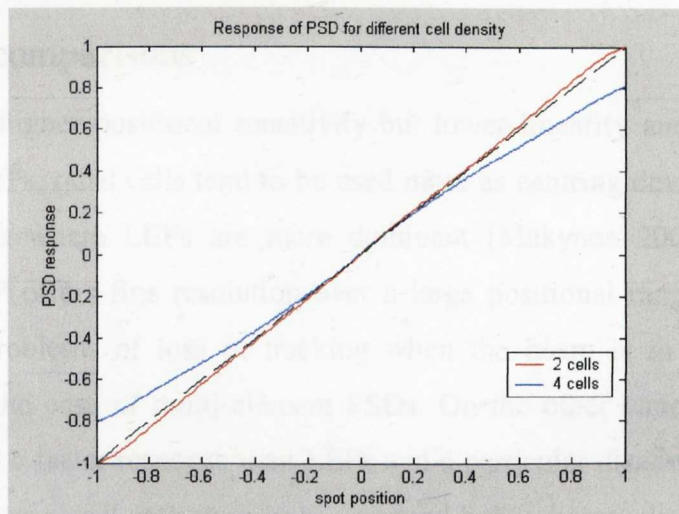
Figure 1.11 Response of a 4-cell linear PSD for spot sizes of different radius, r

1.4.3.1 Spot size

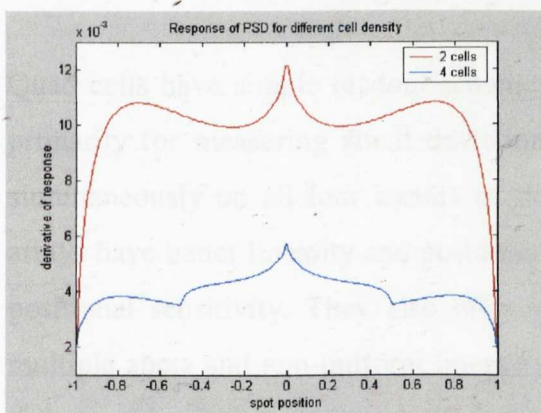
From Figures 1.10 and 1.11, we can see that when a spot is smaller than the size of a cell or pixel and it moves completely into one cell tracking is lost, which results in a step-like response [Sharman 2002]. While tracking is still achieved, non-linearity for spot sizes smaller than the detector size is due to the circular nature of the beam. Non-linearity for spot sizes larger than the detector is due to the truncation of the beam as it moves off the array. Maximum linearity and positional range is obtained when the spot size is the size of the entire detector as shown in Figure 1.10 for $r = 1$. However, the spot size is usually made smaller for two reasons [de Lima Monteiro 2002]. For large displacements, the beam may impinge on neighbouring cells leading to optical crosstalk. Secondly, the positional resolution or positional sensitivity is higher for smaller spot sizes because for a given displacement a small spot produces a much bigger differential signal.

1.4.3.2 Cell density

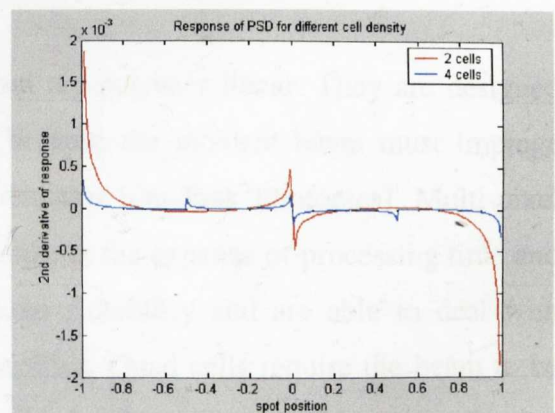
The larger the cell density the better the linearity [de Lima Monteiro 2002]. This can be seen from the differential and double differential of the PSD response of a bi-cell and 4-cell linear array in Figure 1.12 (b) and (c). The downside is that the positional sensitivity is poorer as indicated by the slope of the PSD responses. Also larger cell density means more complicated processing and longer processing time. As we have seen, positional sensitivity can be improved by making the spot size smaller. There is a trade-off between linearity and positional sensitivity. Multi-pixel arrays are able to deal better with smaller spot sizes, and likewise for a given spot size of a few pixels, the larger the array the larger the positional range achievable.



(a) Bi-cell and a 4-cell linear array PSD response



(b) Differential of bi-cell and 4-cell PSD response (Fig. (a))



(c) Double differential of bi-cell and 4-cell PSD response (Fig. (a))

Figure 1.12 Comparison of a bi-cell and a 4-cell linear array PSD response

1.4.3.3 Intensity profile

The effect of the beam shape and intensity profile also needs to be considered. The response of a quad cell is only linear over the whole range for a rectangular or square beam. With a circular beam, linearity is only achieved over the central region of the quad cell. The situation is even worse for laser beams which have a Gaussian profile [de Lima Monteiro 2002]. For a Gaussian beam, maximum linearity is not obtained with a beam the size of the quad cell but of that smaller due to the infinite extent of a Gaussian beam. With a Gaussian beam incident on a multi-pixel array, typically a spot size of about 1 to 2 pixels would then be suitable for maximum linearity, sensitivity and positional range.

1.4.3.4 PSD comparisons

Because of its higher positional sensitivity but lower linearity and positional range compared to LEPs, quad cells tend to be used more as centring devices than as linear position sensors where LEPs are more dominant [Mäkynen 2000]. As a custom device, the LEP offers fine resolution over a large positional range as there are no gaps and no problems of loss of tracking when the beam is in a single detector segment as in the case of multi-element PSDs. On the other hand, quad cells have lower noise and a faster response than LEPs and a particular disadvantage of LEPs is that it does not cope well with stray or background light whereas discrete detectors are able to remove this somewhat by applying a threshold.

Quad cells have simple readout schemes but are not very linear. They are designed primarily for measuring small deviations because the incident beam must impinge simultaneously on all four sectors of the detector [On-Trak Photonics]. Multi-pixel arrays have better linearity and positional range at the expense of processing time and positional sensitivity. They also offer greater flexibility and are able to deal with multiple spots and non-uniform intensity profiles. Quad cells require the beam to be defocused in order to achieve sufficient linearity making it susceptible to illumination fluctuations [Mäkynen 2000], that is, smaller spot sizes deal better with scintillations due to atmospheric turbulence. The relative performance of the different PSD structures can be summarised as in Figure 1.13.

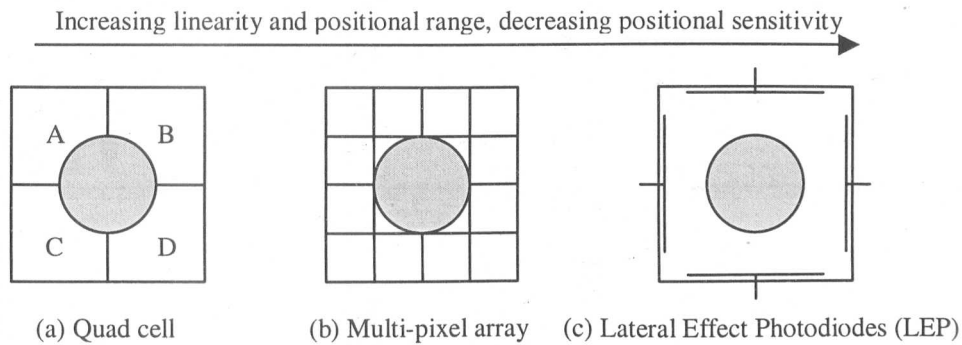


Figure 1.13 Performance of the different PSD structures

With any multi-element detector, the issue of crosstalk arises and requires mentioning. There are two possible sources of crosstalk; crosstalk from other elements or cells and crosstalk from the substrate. Crosstalk from outside the array due to diffused carriers actually improves the linearity by increasing the signal at the edges and gives the appearance of larger pixel size at the edges. However, crosstalk from within the array serves to average the centroid value towards the centre leading to a reduction in positional sensitivity.

1.4.4 CENTROID PROCESSING

In the previous section it was shown how lateral effect photodiodes (LEP), quad cells and multi-pixel arrays are used for the purpose of centroid detection. In this section the processing techniques in computing the centroid from these architectures are presented. Table 1.1 shows a summary of the work done by other groups capable of obtaining optical centroids using standard CMOS or BiCMOS processes. Most LEP systems have the processing performed off-chip because the LEP itself is not usually fabricated on a standard CMOS process due to the high non-linearity obtained. Turner [Turner 1994] demonstrated a LEP in standard CMOS with photocurrents measured externally at a maximum bandwidth of 2.4 kHz. The reported resolution was approximately $0.25\mu\text{m}$ but with non-linearity at the edges reaching 40%. Centroid processing for quad cell and multi-pixel array architectures, on the other hand, can be readily integrated on-chip.

Author, Year	Location	Architecture	Process	Array size	Pixel size	Bandwidth	Positional resolution	Application
Johansson, 2003	Integrated Vision Products (IVP), Sweden	Centre of gravity calculation on 3-10 bit digital values	0.35µm CMOS	1536x512 pixels	9.5µm (fill factor: 60%)	N/A (Bit-serial processor in each column. Multiplication takes $3b^2$ cycles of 33MHz clock)	-	General purpose machine vision sensor
Nirmaier, 2003	Kirchoff Institute of Physics	Winner Take All (WTA) circuitry	0.35µm CMOS	8x8 array of 21x21 pixels	17µm	300 Hz (30pW detectable power)	17µm (1.25 pixel uncertainty, alternate pixels for x and y centroid)	Ophthalmology
Akita, 2002	Future-University-Hakodate, Japan	Thresholding (point masking algorithm to detect multiple light), pulse width modulation (lower switching noise but slower)	0.6µm CMOS	23x23 pixels	120µm (fill factor: 5.7%)	20 kHz	120µm (1 pixel)	Robot vision

Ambundo, 2002	Mixed-Signal-Wireless (MSW), Texas Instruments	Quad cell (on-chip reconstruction using resistive grid to solve 2 nd derivative of phase from adjacent quad cell centroids)	AMI 0.5µm n-well CMOS for translinear amplifier (rest of chip not yet fabricated)	-	-	-	-	Adaptive optical systems
de Lima Monteiro, 2002	Electronic Instrumentation Laboratory, Delft University of Technology	Quad cells (off chip centroid computation by PC)	1.6µm CMOS process	8x8 quad cells	600µm x 600µm	200 Hz closed-loop correction, 260 Hz wavefront reconstruction, 3.125 kHz sensor readout	1.4µm for a 7µW spot	Optical profile testing
Oike, 2002	University of Tokyo	Thresholding (with logarithmic sensor correlation circuit background suppression)	0.5µm CMOS	64x64 pixels	40µm (fill factor: 18.05%)	2 kHz	40µm (1 pixel)	3D measurement systems
Droste, 2001	Kirchoff Institute of Physics	WTA	0.6µm n-well CMOS	16x16 array of 19x19 pixels	17.6µm	500 Hz (1nW detectable power)	17.6µm (1.85 pixel uncertainty, alternate x/y pixels)	Ophthalmology

Pain, 2000	NASA's Propulsion Laboratory, California	Jet	Capacitive array (centroid computation for up to 9x9 pixels)	HP 0.5µm CMOS	128x128 pixels	12µm	20 kHz (9x9 pixels) to 50 kHz (3x3 pixels)	0.6µm (0.05 pixel, 850mV saturation)	Object tracking, navigation, deep-space optical communications
Furth, 1998	The School of Electrical & Computer Engineering, New Mexico State University	Klipsch of	Quad cell (current-to-voltage conversion and quad differencing circuits integrated on-chip)	1.2µm CMOS	2x2 (single quad cell)	160µm x 160µm	-	-	Adaptive optical systems
Makynen, 1998	University of Oulu, Finland	of	Thresholding, off-chip centroid calculation of binary image map	1.2µm n-well CMOS	16x16 pixels	50µm (30% fill factor)	-	4.3µm (uniform 280µm spot, 100Hz)	Position sensing applications with simultaneous multiple spot detection
Turner, 1994	Optoelectronic Computing Systems Center, University of Colorado	of	WTA for coarse resolution, LEP for fine resolution	2µm n-well CMOS	2x1 LEPs	400µm x 180µm	2.4 kHz for a single 2D LEP (photocurrents measured externally)	0.25µm (up to 40% non-linearity at edges)	Pattern recognition in optical correlator systems

Deweerth, 1992	Physics Computation Laboratory, California	Resistive array (Differential transistor pair aggregates image data to two output currents. Resistive array encodes position)	2 μ m BiCMOS process	160x160 pixels	-	-	-	Development of biologically inspired vision systems
Forcheimer, 1992	Integrated Vision Products (IVP), Sweden	Centre of gravity calculation on 8 bit data	-	256x256 pixels	32 μ m (fill factor: 48%)	250 Hz	-	General purpose machine vision sensor
Standley, 1991	Massachusetts Institute Technology	Uniform resistive line for centroid, quadratic resistive line for orientation (resistive grid aggregates 2D image to 1D problem)	2 μ m CMOS process	29x29 pixels	190 μ m x 190 μ m	5 kHz	$\pm 0.3\%$ of a 25 x 25 square	Robot vision
Gonnason, 1990	Alberta Microelectronic Centre, Canada	Resistive array	3 μ m p-well CMOS process	10x10 pixels	84 μ m x 90 μ m	3.33 kHz	1 μ m	Laser and process control instrumentation, tool alignment

Table 1.1 Other research efforts capable of optical centroid detection

1.4.4.1 Quad-Cell Centroid Processing

Processing using quad cells are relatively simple requiring only a minimum number of signals; two from each axis. De Lima Monteiro [de Lima Monteiro 2002] demonstrated an approach for an integrated Shack-Hartmann wavefront sensor using an array of 8 x 8 quad cells in a 1.6 μm CMOS process. The sensor can be read out at a rate of 3.125 kHz but the current-to-voltage conversion and serial conversion of the analogue voltages into digital format was performed off-chip and the centroid computation was carried out on a PC. The resulting operating frequency of 260Hz was limited by the data acquisition card. Another quad cell centroiding approach, this time in analogue using a 1.2 μm CMOS process, by Furth [Furth 1998], integrates the current-to-voltage conversion on-chip using passive and active loads as well as differencing circuits which computes the difference between the photocurrents in the x and y-direction. The differencing circuits consist of double-differential transconductance amplifiers. Experimental results were not reported. However, recently, Ambundo and Furth [Ambundo 2002] have incorporated the wavefront reconstruction on-chip by finding the second derivative of the phase by taking the difference between the centroid currents of neighbouring quad cells and injecting this result into a resistive grid which solves this second derivative to obtain the phase. Normalization allows the centroid computation to be independent of light intensity and was achieved using a modified current amplifier to divide the sum of the four photocurrents in the quad cell. Currently the system has only been simulated and yet to be fabricated and no performance results were shown.

Charge-coupled devices (CCD) are multi-pixel arrays but when used in a Shack-Hartmann wavefront sensing system CCDs are typically³ used as an array of quad cells with guard row and column pixels between them [Thompson 2002]. Due to the serial readout nature of CCDs, the entire wavefront has to be sampled and a whole frame of light spots read out before they can be processed. This results in a data

³ There are applications where the use of more pixels per subaperture than a quad cell is needed such as in varying seeing conditions. A multi-pixel array can easily be adapted for such circumstances at the expense of reduced signal-to-noise ratio and increased computational load.

bottleneck. Processing of individual light spot positions at the sensor level would alleviate this problem as only reduced bandwidth data need to be transmitted off-chip. However, circuit integration on CCDs remains difficult (see Section 1.5.5.1).

1.4.4.2 Multi-Pixel Array Centroid Processing

Processing using quad cells offer limited displacement range and require careful alignment of the null point of the system [Tyson 1998] as large offsets from the null point will reduce the dynamic range of the system and lead to significant non-linearity [Dillon 1999]. Using a multi-pixel array will allow the system to cope better with varying aberrations and seeing conditions. Efforts into incorporating centroid computation for multi-pixel arrays at the sensor level can be categorised into two basic approaches, analogue and digital and several different sub-approaches, as illustrated in Figure 1.14.

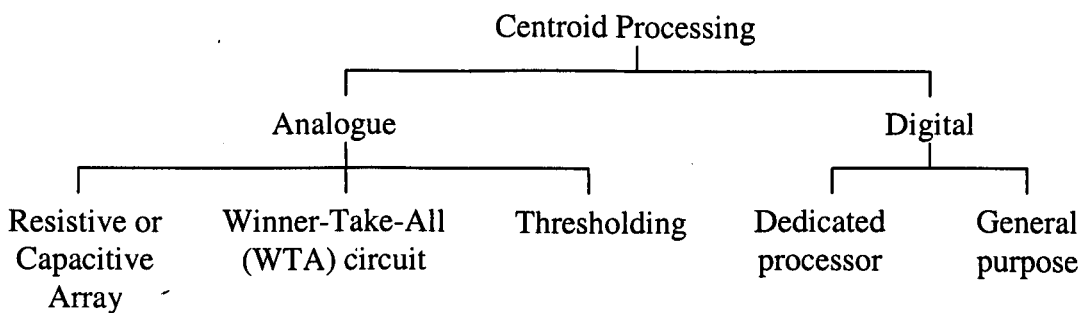


Figure 1.14 Different approaches for optical centroid processing using multi-pixel arrays

1.4.4.2.1 Analogue Centroid Processing for Multi-Pixel Arrays

Most multi-pixel array approaches are performed in analogue, using either an analogue current division method capable of subpixel accuracy, or discrete binary position sensing techniques⁴. With the analogue current dividing method, photocurrents are divided on a uniform resistive array [Gonnason 1990, Standley 1991] or a linearly varying capacitive array [Pain 2000]. Both effectively compute the first order moment of the array photocurrents. With the uniform resistive array, the photocurrent of a pixel is divided on the line and the difference in output currents of the ends of the resistive line is directly related to the position of the incident light on the array. With a quadratic resistor line, the second order moment can be obtained and used to determine the orientation of the object [Standley 1991]. In addition to the basic resistive line, Deweerth [Deweerth 1992] used a current mirror and differential transistor pairs to establish feedback allowing the system to continuously respond to changes in spot position. However, non-idealities and mismatch in these additional circuitry caused offsets in the system. With the linear capacitive array, the pixel voltages are sampled onto separate sampling capacitors, the sizes of which are proportional to the integer row and column addresses, hence giving the inner products of the centroid computation of equation (1.4).

Binary position sensing effectively uses a form of thresholding technique to reject all photocurrent levels below a certain threshold level or below the largest signal level in the array or a collection of pixels. Many variations are possible but two commonly used circuits are the winner-take-all (WTA) circuit [Droste 2002, Nirmaier 2003] or some form of on-pixel comparator [Burns 2003, Makynen 1998]. Figure 1.15 shows the basic form of the WTA circuit and its I_D vs. V_{DS} characteristic. A WTA circuit consists of an array of competing cells with each cell consisting of two MOSFETs M_S and M_F . M_S senses the input current I_i while M_F , if activated, draws the output current I_o .

⁴ Digital centroid computation in this thesis refers to the computation of the centroid from several bits of digitized pixel values and not from a binary image map such as in the case of binary position sensing.

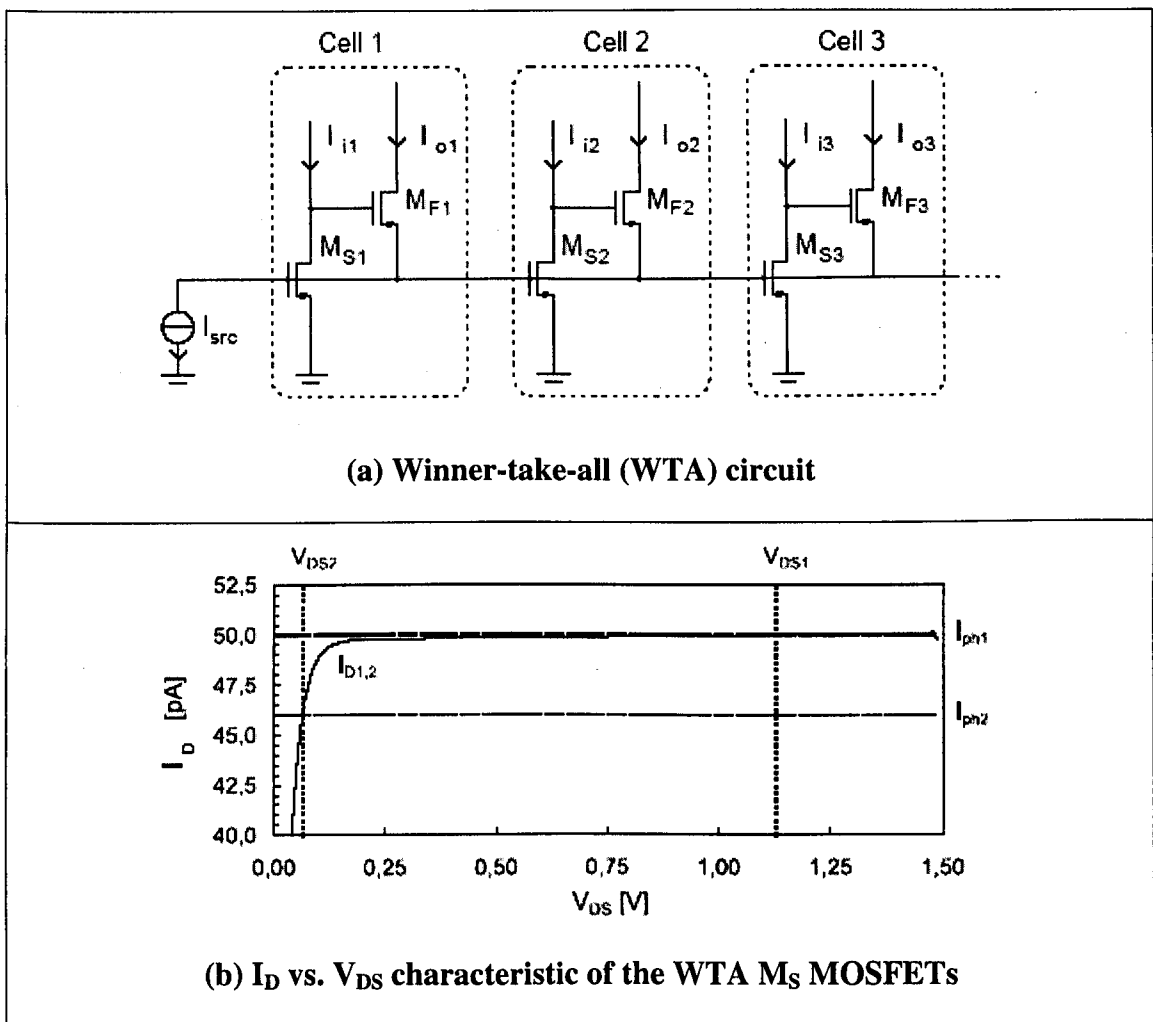


Figure 1.15 Basic topology and operation of a WTA circuit [Droste 2002]

Because all the M_S are identical and are gate-connected, they have the same I_D vs. V_{DS} characteristic. The one with the highest input current will generate the highest drain potential and hence the highest V_{GS} of all the M_F , therefore sinking most of the current source I_{src} and shutting off all other M_F . The computation is continuous in time and the winning output encodes the logarithm of its associated input since all the M_F are operating in subthreshold [Lazzaro 1988]. Saturation of the pixel is determined by the saturation of the WTA M_S MOSFETs and positional accuracy is limited to that of a single pixel. With this circuit, a very slow response time (several hundred ms) is obtained due to the large photodiode capacitance seen at the drain of M_S . The capacitance seen was reduced by using a regulated cascode configuration. Response time can be improved further by setting the drain of M_S to a defined value at startup and by introducing positive feedback into the WTA. But enabling feedback reduces accuracy of position detection due to mismatches. Nirmaier et al. [Nirmaier 2003]

introduced an interdigitated topology to the WTA concept by splitting the single WTA circuit into several groups. This has the advantage of increased robustness against defective outputs, reduced sensitivity to mismatch and faster response.

For analogue centroid computation utilizing destructive readout such as in the current division method, or for those utilizing the WTA algorithm, two discrete photodiodes are needed per pixel. One for the x-centroid and one for the y-centroid. This results in lower fill factor, sensitivity and a non-linear spatial response. De Lima Monteiro [de Lima Monteiro 2002] proposed the use of a spiral structure to reduce the non-linearity. With these architectures, the pixels in each row and column are tied together and the photocurrents along each row and column are summed so only two sets of current division or WTA circuits are needed per array, one for each axis, as illustrated in Figure 1.16. In a CCD this would be equivalent to binning all pixels in the row or column [Dillon 1999].

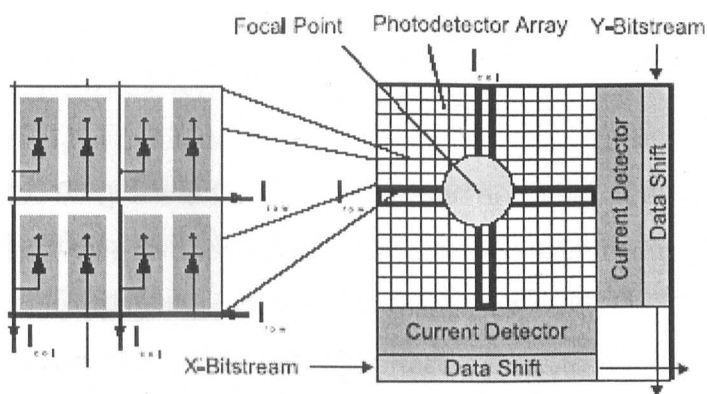


Figure 1.16 Use of two photodiodes per pixel and the summation of photocurrents along each row and column with analogue centroid computation [Droste 2002]

Standley [Standley 1991] used a uniform grid of resistors to aggregate the photocurrents in both the x and y-dimensions, hence eliminating the need for two photodiodes per pixel. However, this suffers from non-linearity due to the tolerance of on-chip resistors as well as increased power consumption and thermal noise. It also required the use of two resistive lines per axis instead of just one. This technique has limited usage in position sensing because the advantage of increased fill factor and sensitivity from removing the need of a second structure is lost by the need to

integrate a resistor at each pixel. However, its use in neural network structures for vision chips is common as interconnectivity between neighbouring pixels is desired.

Makynen [Makynen 1998] used global threshold current comparison per pixel to generate a binary image map and off-chip moment calculation of the binary map to obtain sub-pixel accuracy. Unlike the WTA circuit approach, it is able to deal with multiple beam spots and it does not require two structures per pixel. However, it does not deal well with non-uniform intensity profiles due to its binary representation and the extensive circuitry per pixel leads to low fill factor and sensitivity. With position sensing using on-pixel comparators, it is possible to use a ramp function of the threshold value, to obtain a more accurate centroid estimate as well as deal with non-uniform intensity profiles by obtaining several binary image maps at different threshold levels [Burns 2003]. However, this requires post-processing and several readouts of the array.

1.4.4.2.2 Digital Centroid Processing for Multi-Pixel Arrays

Analogue centroid computation offers the advantage of high speed and high functional density but suffers from lack of flexibility and imprecision due to mismatches and tight tolerances of components. De Lima Monteiro [de Lima Monteiro 2002] found that there was significant spatial variation in on-chip polysilicon-array resistance which leads not only to the shifting of the zero response but also to the slope of the response curve, as per Figure 1.10. Well structures offer higher sheet resistance but has greater spatial variation and poorer temperature and voltage coefficients. Also, as CMOS technology scales, the advantage of speed and functional density of analogue over digital diminishes.

Digital centroid computation involves the analogue-to-digital conversion of the pixel values into several bits of data and computing a weighted average of the photogenerated signals. A generic 256 x 256 pixel array system with an on-chip image processor has been designed which performs several common image processing algorithms including centroiding at 250 frames/s [Forcheimer 1993, Forcheimer 1992]. Recently an even more advanced and larger array sized programmable image

sensor and processor has been developed by the same group [Johansson 2002]. However, in an adaptive optics system such as the Shack Hartmann wavefront sensor, a large number of tilt sensors are required but the pixel count of each tilt sensor can be minimal. Nonetheless, the work presented by the group is encouraging because it shows that it is possible to integrate complex digital circuits alongside a CMOS image sensor and still achieve low noise.

Another generic structure for image processing is the cellular neural network (CNN) architecture where each cell (pixel) senses a point of the input image and interacts with neighbouring cells to perform parallel-processing tasks on the input image [Roska 1993]. All cells operate in parallel and in continuous time so that high operation speeds are obtained [Dominguez-Castro 1997]. However, due to the locality of the connections, global image processing tasks such as centroid detection require longer processing times, and generic structures in general are not optimised for any particular task⁵.

The approach taken in this work is to integrate dedicated local digital centroid processing at each subaperture to measure the local wavefront tilt. By performing the centroid computation of the subapertures in parallel, the processing speed is maximised and the amount of data to be sent off-chip is reduced. In addition to an increase in speed, a single-chip system will have an advantage of reduced system size, costs and power consumption over multi-chip systems. This work represents the only dedicated digital centroid processor designed and fabricated to date⁶.

⁵ However, post-processing of a Shack-Hartmann subaperture image using artificial neural networks is capable of providing a more accurate estimate of the centroid location than with conventional linear estimators (1st moment calculation) [Montera 1996].

⁶ There are digital chips that compute the first, second and higher order moments, e.g. [Hatamian 1986], but these do not have on-chip photodetection and are not dedicated centroid processors.

1.5 PHOTODETECTION

When determining the centroid in a given subaperture, the relative light intensities incident on each pixel in the array needs to be measured accurately. So an understanding of the mechanisms involved in the photogeneration of carriers is needed and this section will examine this.

1.5.1 OPTICAL ABSORPTION

When a photon is incident on a piece of semiconductor, there is a possibility that the photon will be absorbed if its energy is greater than the bandgap energy of the semiconductor. When a photon is absorbed, a bound electron in the valence band is excited to the conduction band where it is free to move randomly or under the influence of an electric field. The excited electron leaves behind a vacancy, or hole, in the valence band, which is also mobile. Hence an electron-hole (e-h) pair is generated. The electron-hole pair will then either recombine, diffuse or get separated by an electric field. Silicon is an indirect bandgap material so a phonon is required in the optical absorption process reducing transition probability and making the process strongly temperature dependent. For crystalline silicon, the bandgap energy, E_g is 1.12 eV making the cut-off wavelength above which no photons can be absorbed to be $\lambda_c \sim 1.11 \mu\text{m}$ ⁷. The optical absorption process can be quantified as follows. The carrier generation rate $g(x)$ at a depth of x in the silicon must equal the rate of change of the photon flux $\phi(x)$ with x and at the same time proportional to $\phi(x)$ [Bar-Lev 1984] as given by:

$$g(x) = -\frac{d\phi}{dx} = \alpha(\lambda)\phi(x) \quad (1.5)$$

where the proportionality constant $\alpha(\lambda)$ (cm^{-1}) is called the absorption coefficient and is dependent on the material and the wavelength, λ . The solution of this shows an exponential decay of photon flux with penetration depth as follows:

⁷ Energy of a photon, $E = hf = \frac{hc}{\lambda q} eV = \frac{1.24}{\lambda(\mu\text{m})} eV$; Cut-off wavelength, $\lambda_c = \frac{1.24}{E_g (eV)} \mu\text{m}$

$$\phi(x) = T\phi_0 \exp(-\alpha x) \quad (1.6)$$

where T is the transmission coefficient⁸ and ϕ_0 is the photon flux at the surface ($x=0$). This then gives a carrier generation rate of:

$$g(x) = -\frac{d\phi}{dx} = T\alpha\phi_0 \exp(-\alpha x) \quad (1.7)$$

The absorption coefficients of several common semiconductor materials and compound semiconductors are shown in Figure 1.17. For wavelengths exceeding λ_c , α becomes negligible and the material becomes transparent to those wavelengths. For shorter wavelengths, α becomes very large which means photons of shorter wavelengths get absorbed closer to the surface. The slow increase of α with photon energy in silicon is due to the fact that Si is an indirect bandgap semiconductor.

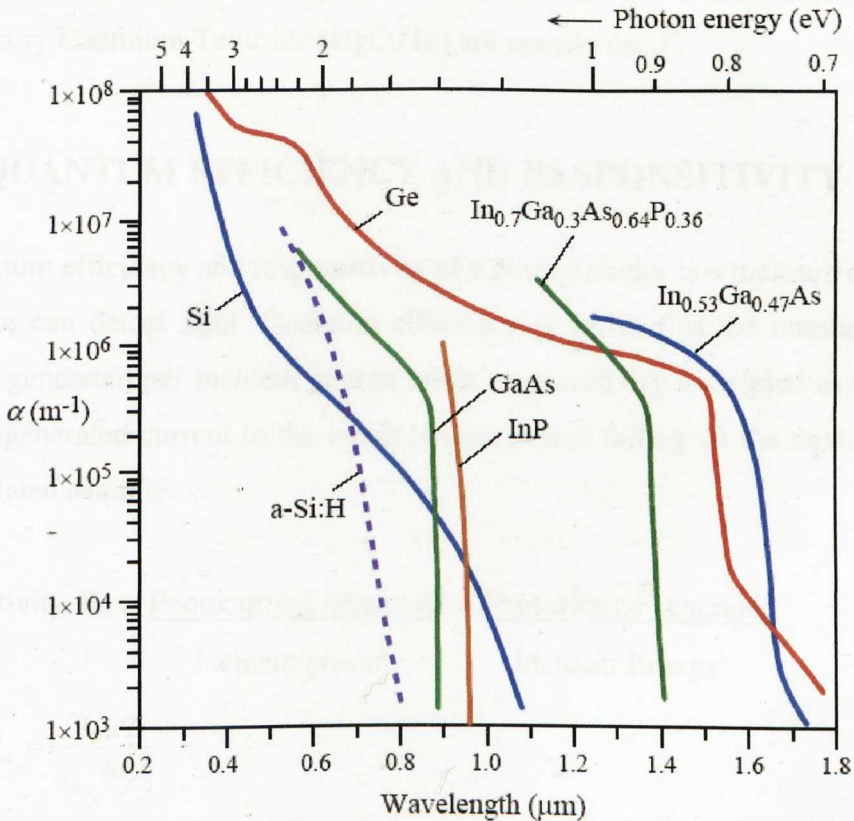


Figure 1.17 Absorption coefficient, α , for various semiconductor materials at 300K [Kasap 2001]

⁸ The transmission coefficient or transmittance is the ratio of the amount of transmitted light to the amount of incident light i.e. the fraction of incident photons on the surface that is not reflected. With antireflection coatings, $T=1-R \approx 1$, where R is the reflectance.

Figure 1.17 also shows that different semiconductor materials can be used to detect incident radiation over different wavelength regions with silicon having a characteristic wavelength range of about 250 nm to 1100 nm. Visible wavelengths range from 400nm (blue) to 750nm (red). Typically, blue light penetrates to a depth of about 0.2 μ m while red light penetrates more than 10 μ m. This difference in penetration depths can be utilized for the design of colour sensors by stacking charge collection layers at different depths, as pursued by Foveon Inc. in their commercially available Foveon X3 direct image sensors [Hubel].

The choice of silicon in this work is due to the high level of circuit integration required and available with the Complementary Metal Oxide Semiconductor (CMOS) silicon process technology. In the near infrared and infrared, compound semiconductors like Indium Gallium Arsenide (InGaAs), Indium Antimonide (InSb) and Mercury Cadmium Telluride (HgCdTe) are usually used⁹.

1.5.2 QUANTUM EFFICIENCY AND RESPONSIVITY

The quantum efficiency and responsivity of a photodetector is a measure of how well the device can detect light. Quantum efficiency is defined as the number of signal electrons generated per incident photon while responsivity is defined as the ratio of the photogenerated current to the incident light power falling on the device and they can be related as follows:

$$\text{Responsivity, } R_\lambda = \frac{\text{Photocurrent generated}}{\text{Incident power}} = \frac{\text{Photocharge generated}}{\text{Incident Energy}}$$

$$\text{Hence, } R_\lambda = \frac{n_{gen}q}{n_{inc} \frac{hc}{\lambda}} \quad (1.8)$$

$$\text{Therefore, the quantum efficiency } \eta = \frac{n_{gen}}{n_{inc}} = R_\lambda \frac{hc}{\lambda q} = 1.24 \times 10^{-6} \frac{R_\lambda}{\lambda} \quad (1.9)$$

⁹ These types of detectors are called quantum detectors. Thermal detectors like bolometers and thermopiles are also used for far infrared detection.

where n_{gen} is the number of electron-hole pairs generated and n_{inc} is the number of incident photons, λ is the incident wavelength (m), h is Planck's constant = 6.626068×10^{-34} m²kg/s, c is the speed of light = 3×10^8 m/s, and q is the electron charge = 1.6×10^{-19} Coulombs.

There are various types of photodetector structures that can be implemented in silicon such as p-n junction photodiodes, Schottky photodiodes, p-i-n photodiodes, avalanche photodiodes (APD), metal-oxide semiconductor (MOS) capacitors and phototransistors [Bar-Lev 1984, Sze 1981].

1.5.2.1 P-N Junction Photodiode

The p-n junction photodiode is by far the most common structure because of its low cost, visible wavelength range and its easy availability in standard silicon processes. In a junction photodiode, a p-n junction is used as the photodetection region as the depletion region provides an electric field to efficiently separate and collect the electron-hole pairs generated and to prevent recombination. However, electron-hole pairs generated outside the depletion region can also diffuse to the depletion region and be collected but less efficiently.

The quantum efficiency of a photodiode structure can be derived by solving for the drift current inside the depletion region and the diffusion current outside the depletion region¹⁰. The quantum efficiency for a vertical p-n photodiode with a very narrow p-region and n-type bulk substrate can be shown to be [Sze 1981]:

$$\eta = 1 - \frac{\exp(-\alpha W)}{1 + \alpha L_p} \quad (1.10)$$

where η is the quantum efficiency, α is the absorption coefficient, W is the depletion width of the junction and L_p is the diffusion length of the minority holes in the n-substrate. Hence the quantum efficiency of a photodiode can be increased by

¹⁰ The drift current is obtained from integrating the carrier generation rate of equation (1.7) across the depletion region. The diffusion current is found by solving the diffusion equation for the minority carrier concentration using boundary conditions.

increasing the depletion width, which is dependent on the doping levels and the reverse bias voltage applied.

The speed of a photodiode is limited by three factors: diffusion of carriers, drift time in the depletion region, and capacitance of the depletion region [UDT Sensors Inc. 1982]. Carriers generated outside the depletion region must diffuse to the junction resulting in considerable time delay. The wider the depletion region, the more light is absorbed and the larger the spectral bandwidth. However, the depletion region must not be too wide or transit-time effects will limit the frequency response. It also should not be too thin or excessive photodiode capacitance C will result in a large RC time constant.

1.5.2.2 Other photodetector structures

The Schottky photodiode is formed by the interface of a doped semiconductor with a metal layer and is capable of high speeds of the order of GHz but suffers from lower quantum efficiency and higher dark current. A p-i-n photodiode has a thick or lightly doped intrinsic (i) layer between the p and n-regions that serves to provide the device with a large depletion region and a low junction capacitance. This results in faster response times and higher quantum efficiency. However, the intrinsic layer which is usually tailored to be fully depleted is not a standard feature in the CMOS fabrication process. An avalanche photodiode (APD) achieves internal gain by operating under high reverse bias in the avalanche region where multiplication of charge carriers occurs through impact ionization. APDs have large dark current and integration of electronic circuitry with an APD is not straightforward due to the high reverse voltage requirement [de Lima Monteiro 2002]. A MOS capacitor detects light by storing photogenerated charges in a potential well that is formed when a voltage is applied to its gate. It is capable of high sensitivity and is the basis of the charge-coupled device (CCD) which will be discussed later on in Section 1.5.5.1. Phototransistors provide internal gain but only carriers generated in the base-collector space-charge region is amplified and phototransistors are slower and less linear than photodiodes and have a large dark current.

1.5.3 NOISE AND PHOTODIODE EQUIVALENT CIRCUIT

For modelling of a junction photodiode, an equivalent circuit is needed and one that is typically used is shown in Figure 1.18. The different noise sources have been collectively represented by the current source I_N and the photocurrent is modelled as the current source I_{ph} .

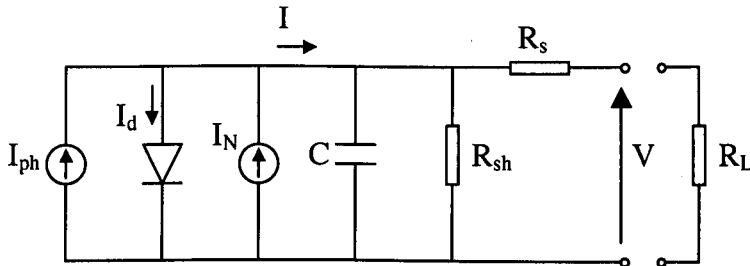


Figure 1.18 Photodiode equivalent circuit

I_d represents the diode current and is given by:

$$I_d = I_o \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] \quad (1.11)$$

where $k=1.38 \times 10^{-23}$ J/K is the Boltzmann constant, T is the absolute temperature in Kelvin, $q=1.6 \times 10^{-19}$ C is the electron charge, V is the voltage across the photodiode and I_o is the process dependent diode saturation current. In the reverse bias the diode current converges to $-I_o$, which is equivalent to the dark current of the photodiode.

The resultant output current is given by the sum of the individual currents:

$$I = I_{ph} - I_d + I_N \quad (1.12)$$

The capacitance, C , of the photodiode is the junction capacitance of the depletion region formed and the shunt resistance R_{sh} represents the resistance of this depletion layer and is usually very large of the order of $10\text{M}\Omega$ to $10\text{G}\Omega$ [de Lima Monteiro 2002]. The series resistance, R_s , which is the resistance of the undepleted region between the edge of the depletion layer and the metal contact, has a value ranging from several Ohms to several hundred Ohms. There are two main sources of noise in a photodiode, shot noise and thermal noise [UDT Sensors Inc. 1982, de Lima Monteiro 2002, Hornsey 1999c]. In addition, there is also $1/f$ noise, reset noise and spatial noise.

1.5.3.1 Shot Noise

Shot noise, I_s , is due to the statistical fluctuation of both the photocurrent I_{ph} and the dark current I_d , and is expressed by:

$$I_s = \sqrt{2q(I_{ph} + I_d)B} \quad (1.13)$$

where q is the electron charge and B is the noise measurement bandwidth.

1.5.3.2 Thermal Noise

The thermal noise or Johnson noise of the photodiode, V_t , is due to the random motion of carriers in resistive electric materials and it increases with temperature. In a photodiode the thermal noise associated with the load resistance R_L is given by¹¹:

$$V_t = \sqrt{4kTR_L B} \quad (1.14)$$

where k is the Boltzmann constant, T is the absolute temperature in Kelvin and B is the noise measurement bandwidth.

1.5.3.3 Reset Noise

Capacitors are usually thought of as noise-free devices. In the case of sampling systems, however, they exhibit a theoretical noise because the capacitor is periodically reset (see Section 1.6.3). In most image sensor pixel architectures, signal detection will involve the reset of the photodiode capacitive node. This operation gives rise to reset noise and is due to the thermal noise of the resistance of the switch used to reset the photodiode.

The noise equivalent bandwidth, B , of a circuit is defined as the voltage-gain-squared of the circuit as follows [Hornsey 1999c]:

$$B = \frac{1}{A(0)^2} \int_0^{\infty} |A(f)|^2 df \quad (1.15)$$

where A is the voltage gain of the circuit and f is frequency. For an RC circuit like that of a photodiode being reset through a switch, the noise equivalent bandwidth can be shown to have the value of $1/4RC$ and substituting this into the expression for thermal noise voltage we get the ever popular 'kTC' noise figure of:

¹¹ Assuming the load resistance R_L is significantly smaller than the R_{sh} , which is a reasonable assumption in most cases.

$$V_i = \sqrt{4kTR_{eq}B} = \sqrt{\frac{4kTR_{eq}}{4R_{eq}C}} = \sqrt{\frac{kT}{C}} \quad (1.16)$$

where C is the capacitance of the photodiode (Figure 1.18) and R_{eq} is the equivalent resistance of the circuit.

1.5.3.4 1/f Noise

Another noise source which exists but is given only brief mention here is the 1/f noise or flicker noise. The causes of this noise are not well understood and it has been proposed that it comes from carrier fluctuations at the surface interface traps or by mobility fluctuations. It derives its name from the fact that its magnitude is inversely proportional to its frequency and structures with a larger area are less prone to its effects. Also, it is more significant in lateral shallow devices (e.g. MOS transistors) and less important in bare photodiodes.

1.5.3.5 Spatial Noise

The sources of noise talked about so far are forms of temporal noise. When an array of photodiodes is used, spatial noise¹² needs to be considered as well. This consists of fixed pattern noise (FPN) which is the pixel-to-pixel variations in the absence of illumination and photoresponse non-uniformity (PRNU) which is a function of the incident light level. The main causes of FPN are variations in photodetector geometry, dark current¹³ and threshold voltages, V_T , while the non-uniformity in the photoresponse of CMOS photodiodes is caused mainly by light interference in the passivation layers as well as threshold variations [Makynen 1998]. Typical non-uniformity of a CMOS photodetector responsivity is <5%. With integrated on-pixel circuitry, threshold variations dominate the spatial non-uniformity¹⁴. Good matching in general requires close spacing and non-minimum size which is prohibitive with on-pixel circuitry. Devices operating in the subthreshold have higher threshold voltage or current variations such as in the logarithmic active pixel sensor (see Section 1.6.2.3).

¹² Also known as pattern noise

¹³ Variations in photodetector geometry and dark current are smaller for larger sized devices.

¹⁴ Threshold variation and circuit mismatches have a larger effect on spatial noise than do the other photodiode parameters. For instance, variation of photodiode well capacity across the array does not matter if only half of the total well capacity is used for the desired application.

However, there are means to remove these spatial noise sources with the use of additional circuitry at the column or chip level which will be discussed in Section 1.6.3.

1.5.4 PHOTODIODE MEASUREMENTS

In the testing of photodiodes it is important to understand the different measurement techniques possible when measuring photocurrents directly without integration of charge. The best place to start would be with the general I-V characteristic of a photodiode as shown in Figure 1.19. A photodiode can be operated in either quadrant 3 or quadrant 4 of the diode I-V response. In quadrant 4, one can either measure the open-circuit voltage $V_{o/c}$ or the short-circuit current, $I_{s/c}$.

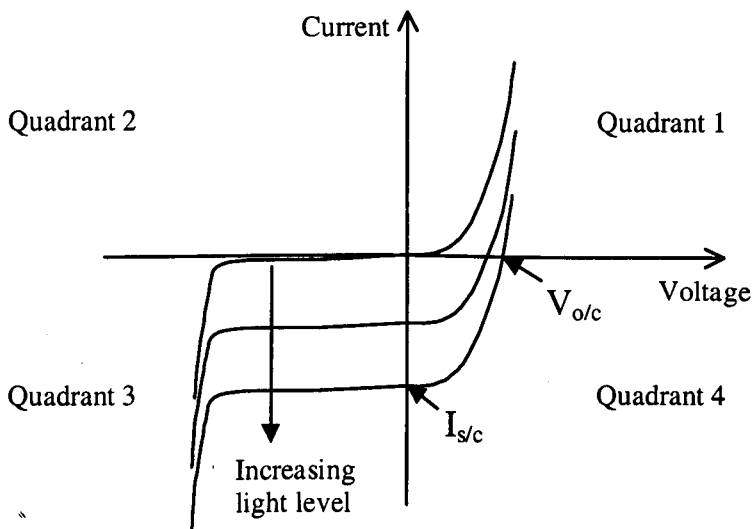


Figure 1.19 I-V characteristics of a photodiode

When measuring the open-circuit voltage ($I=0$ in Figure 1.18 and Figure 1.19), the load resistance is very large, for example that of a high input impedance multimeter. Ignoring noise, and from equations (1.11) and (1.12), the photogenerated open-circuit voltage obtained is:

$$V_{o/c} = \frac{kT}{q} \ln \left(\frac{I_{ph}}{I_o} + 1 \right) \quad (1.17)$$

In effect, what is happening is that the generated photocurrent cancels out the forward bias diode current for very small forward bias voltages. The problem with obtaining the photocurrent this way is that the measurements now depend on temperature as well as I_0 , which in turn depends on process parameters like doping concentration and minority carrier lifetimes. The open circuit voltage, V_{oc} , is also highly non-linear.

In order to get a linear response with respect to photocurrent, it is more suitable to measure the short circuit current, $I_{s/c}$, i.e. measuring the change in photocurrent along the y-axis of Figure 1.19. In order to do so, a very low load resistance is required. An op-amp is typically used to achieve this low load resistance by keeping the voltage across the diode fixed, as shown in Figure 1.20, using the virtual earth principle. In the short-circuit mode, $I_{s/c} = I_{ph}$ ($I_d=0$).

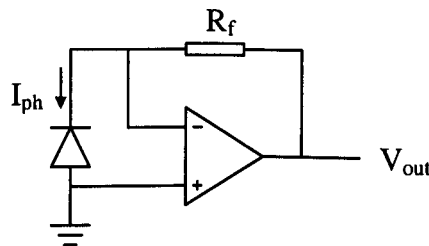


Figure 1.20 Photocurrent measurements using an operational amplifier

The photodiode can also be operated under reverse bias in quadrant 3 with a linear response. In this region, the diode current, I_d is approximately equal to the leakage current, I_0 . An op-amp can again be used to obtain a low load resistance line. The advantage of operating a photodiode under reverse bias is its high speed of response as well as larger generated photocurrent. Both of these are due to the increasing depletion width with reverse bias voltage. However, the disadvantage is that the leakage current is also increased and hence the noise.

When charge integration of a photodiode is to be measured, usually a capacitor performs the charge integration and this could be the photodiode capacitance itself, and a buffer or amplifier is used to readout the signal. The amplifier could be a sophisticated off-the-shelf component which has the advantage of low noise or a simple source follower buffer which lends itself to on-chip integration with the photodiode.

1.5.5 TECHNOLOGY AND MATERIALS

Even in silicon several different technologies and fabrication processes are available to the designer. These include the charge-coupled device (CCD), BiCMOS and CMOS technologies as well modifications to the standard CCD and CMOS process and even a combined CCD/CMOS process. A newer development, borne out of the move towards smaller feature sizes and silicon-on-insulator (SOI) technology, is the Thin Film on ASIC (TFA) technology [Wong 1996]. In the following sections, these technologies and their applicability to the work will be discussed.

1.5.5.1 Charge-Coupled Device (CCD)

Invented in the late 1960s by researchers at Bell Labs, the charge-coupled device (CCD) was initially intended for use as a memory circuit. But its potential in imaging soon became clear and it has since become the industry standard in image sensor technology. The basis of a CCD is the accumulation, storage and transfer of charges using closely spaced metal-oxide-semiconductor (MOS) capacitors. A MOS capacitor is simply a semiconductor substrate with an overlying thin oxide layer and a top metal contact, also known as the gate. When the structure has a p-type substrate, an n-type MOS capacitor is formed. To operate the CCD these MOS capacitors are pulsed with a positive gate voltage and driven into deep depletion (empty potential well). This is a non-equilibrium phase and the structure is able to collect any available minority carriers (electrons). The empty potential well can either be filled up by thermally generated electrons or photo-generated electrons. Fortunately thermal generation of electrons is relatively slow. It takes several seconds at room temperature to collect enough thermally generated electrons for inversion of the MOS capacitor to occur. During this time the potential well is available to collect photo-generated electrons. For low light level applications long integration times may be necessary and cooling is used to reduce the thermally generated dark current.

Once the charge has been stored, the next step is for the charge to be transferred to the output amplifier to be read off-chip. The transfer mechanisms in CCDs are well documented [Theuwissen 1995]. Figure 1.21 illustrates the charge transfer mechanism for a three-phase CCD system. A typical analogy used to describe the transfer of charge in a CCD is that of transferring water using buckets. By varying the voltages

applied to the gate electrodes in a properly timed sequence, the stored charges are shuttled across the array to the output register and finally to the output amplifier. There are various transport systems possible for a CCD, from the classical four-phase system all the way to a single phase system. They have relative tradeoffs between fill factor, charge handling capacity, fabrication complexity and clocking requirements. But by far the most common is the four-phase system for transfer in the array and the two-phase system in the output register.

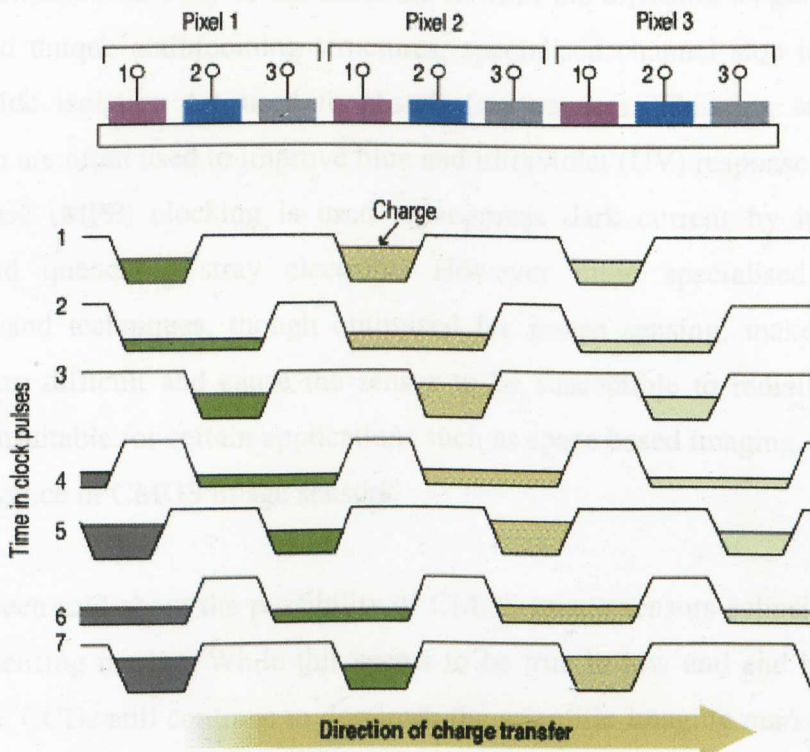


Figure 1.21 Three-phase charge transport mechanism in a CCD (Source: Eastman Kodak)

Besides the various transport mechanisms, there are several architectures possible in a CCD imager, the main architectures being the full frame, frame transfer and interline transfer CCDs. Full frame CCDs represents the basic architecture where the image is directly transferred to the readout register and has the problem of image smear as the sensor is still exposed to illumination as the image is transferred out, necessitating the use of a shutter and making them unsuitable for video applications. The other architectures aim to correct this by having fast intermediate transfer to on-chip storage area before the image is serially readout.

Since its inception, CCDs has had its fabrication process specially tailored towards imaging. CCD fabrication is complex with typically 15 - 25 masks [Hornsey 1999b]. To name but a few unique features; closely spaced or overlapping gates and large clocking voltages (10-20V) are necessary to produce high charge transfer efficiencies, large operating voltages means the gate oxide thickness has to be large (80nm), compared to 10nm in CMOS, and a buried channel structure reduces surface traps and improves charge transfer efficiency. Crosstalk is reduced by controlling the doping concentration and resistivity of the substrate to limit the diffusion length of minority carriers, and unique antiblooming structures, specialised channel stop implants and stepped oxide isolation are used to absorb free carriers. Thinning and backside illumination are often used to improve blue and ultraviolet (UV) response while Multi Pinned Phase (MPP) clocking is used to suppress dark current by inverting the channel and quenching stray electrons. However these specialised fabrication procedures and techniques, though optimised for image sensing, make integration with circuitry difficult and cause the sensor to be susceptible to radiation damage, making it unsuitable for certain applications such as space based imaging. This has led to the resurgence of CMOS image sensors.

Much has been said about the possibility of CMOS image sensors eclipsing CCDs in the image sensing market. While this seems to be true in low end and high volume applications, CCDs still continue to dominate the scientific imaging market. For sure, developments into improving the performance of CCDs are still ongoing with several innovations being introduced. Roper Scientific's deep depletion CCDs use a high resistivity silicon substrate to reduce diffusion of charge carriers and improve quantum efficiency in the near-infrared (NIR). Kodak's Microelectronics Technology Division developed a gate structure based on indium tin oxide (ITO) which is more transparent than polysilicon hence giving better sensitivity in the blue/green region. Fujifilm's 3rd Generation Super CCD System uses octagonal-shaped photodiodes in an interwoven layout to achieve higher sensitivity and equal resolution in both horizontal/vertical direction and diagonal direction. The orthogonal transfer CCD (OTCCD), developed by Tonry, Burke and Schechter [Tonry 1997], permits parallel clocking in both the horizontal and vertical direction by replacing the channel stop between columns of pixels by an additional gate and was used to remove image motion caused by atmospheric turbulence at rates of up to 100Hz. The low light level

CCD (LLLCCD) from E2V is able to achieve sub-electron readout noise levels even at MHz pixel rates using on-chip charge multiplication and is currently being incorporated into the NAOMI wavefront sensor at the Isaac Newton Group of Telescopes (ING). Sony has introduced its HAD (Hole Accumulation Diode), Super HAD and EXview HAD CCD technology where an additional accumulation layer has been included to drain off thermally generated currents. The newer Super HAD and EXview HAD technology also incorporates two layers of on-chip microlenses for better light collection. Kodak integrated clock drivers on-chip with its interline KAI2020 CCD chip. Research is also being done into making CCDs more radiation tolerant. All these mean the predicted demise of CCDs is far from certain. However, for the purpose of this work, CCDs do not offer the level of integration needed to allow parallel processing of subaperture centroids. The disadvantages of this process will be highlighted further in Section 1.5.5.3 when the CMOS process technology is discussed.

1.5.5.2 BiCMOS

The BiCMOS process was introduced to combine the performance, high packing density and low power dissipation of the CMOS process with the high current drive, high switching speed and low mismatch of the bipolar device [Gray 1992]. However, the use of BiCMOS processes for imaging has been limited [Biber 2000, Chou 1991, Guidash 1995, Kuo 1991, Tanaka 1989, Wohl 2003] due to its complexity and cost with no obvious advantage in possible photosensing structures. The process is not yet mature and, unfortunately, many of the improvements in CMOS fabrication techniques do not directly transfer to BiCMOS fabrication. Also the large area required for each bipolar transistor makes them unattractive in large vision chips [Moini 1999]. The bipolar image sensor did achieve some commercial success with the base-stored image sensor (BASIS) [Tanaka 1989] which was used in Canon's EOS line of autofocus sensors but has since been dropped in favour of CMOS sensors. The imager achieves amplification using a vertical bipolar transistor structure with the optically generated holes being integrated on the base.

1.5.5.3 CMOS

Complementary Metal-Oxide Semiconductor (CMOS) technology is the dominant technology in integrated circuit (IC) fabrication and is continuing to mature. CMOS

image sensors, on the other hand, are relatively immature having been sidelined for the better image quality of CCD sensors. However, these devices are making a comeback and an in-depth historical account of the birth and development of CMOS image sensors is given by Fossum [Fossum 1997].

Unlike CCDs, standard CMOS processes are not tailored for imaging purposes. For example, in a standard CMOS process, a shallow epi-layer substrate (see Figure 2.1) is used to mitigate latch-up reducing the response in the red, while heavily doped junctions which enable denser, shorter gate-length devices reduces the response in the blue/green region. Furthermore, CMOS imagers suffer from high temporal noise and 1/f noise because signals are transferred to the outside world via multiple transistor stages. However, CMOS imagers offer higher levels of integration and compared to multi-chip systems, a reduction in system size and power consumption [Janesick 2002]. CMOS imagers are more suited for high volume, space-constrained applications where imaging quality is less important such as in security cameras, PC peripherals, toys, fax machines, and some automotive applications [Litwiller 2001]. A summary of the relative advantages and disadvantages of the CMOS and CCD processes are given in Table 1.2.

CMOS Advantages	CCD Advantages
Capable of on-chip circuit integration Low power consumption Random access to pixel regions of interest (ROI)	High light sensitivity and low noise Low dark current High uniformity
CMOS Disadvantages	CCD Disadvantages
Higher noise levels Larger dark current Lower fill factor	Circuit integration difficult High power consumption Require large multiple supply voltages and complex timing signals Pixel defects could render entire row/column unusable

Table 1.2 Comparison of advantages and disadvantages of CMOS and CCD image sensors

The cost advantage of CMOS over CCDs is not very well understood. CMOS imagers would be much cheaper if they could be produced on the same high-volume wafer processing lines as mainstream logic or memory chips. However, typically for improved performance, CMOS imagers would require additional modifications to the basic process such as optical packaging, on-chip colour filter arrays and on-chip microlenses. So at the chip or sensor level costs are similar but at the system level CMOS imagers are generally cheaper due to the additional related circuitry required for CCD operation [Litwiller 2001].

An issue which faces CMOS imagers is that of decreasing feature sizes. Smaller feature sizes mean higher packing densities, improved fill factors, lower power consumptions, faster speeds¹⁵ and reduced crosstalk¹⁶. However, reduced voltage swing due to downscaling reduces dynamic range and smaller junction depths means reduced volume for photocharge collection and increase in surface effects [Wong 1996] as well as shifting of quantum efficiency curves to shorter wavelengths. Short channel effects lead to off leakage currents and tunneling currents which contribute to the dark current of pixels. Furthermore, for these processes, opaque silicide layers (WSi_2 , $TiSi_2$, $CoSi_2$) are used to reduce contact and sheet resistances of source/drain regions and gates. Hence, as technology scales beyond $0.5\mu m$, modifications to the fabrication process are needed to enable good quality imaging [Lule 2000, Wong 1996] such as the removal of the silicide layer.

The use of CMOS imagers currently proves difficult for low-light level applications such as astronomy due to the high-level of noise in CMOS compared to CCDs. However, CMOS imaging is a relatively new development and noise reduction techniques by means of specialized circuitry are being heavily researched [Bursky 1999, Lai 2002, Meynants 2001, Pain 2003, Rullmann 2003]. Watabe et al. [Watabe 2003] mentioned the overlaying of a high-gain avalanche rushing amorphous photoconductor (HARP) film on top of a CMOS image sensor to produce an ultrahigh

¹⁵ This is due to lower capacitance which leads to better conversion efficiency (q/C) between electron charge and output voltage.

¹⁶ This is due to higher doping levels which lead to reduced diffusion lengths.

sensitivity CMOS image sensor. So it may be that in the not too distant future CMOS image sensors will achieve the level of sensitivity now only seen in CCDs.

1.5.5.4 CCD/CMOS

Modifications of the basic CCD and CMOS process in order to allow more flexible readout in CCDs or improved imaging quality in CMOS include the charge injection device (CID), static induction transistor (SIT), charge modulation device (CMD), pinned photodiodes and many more. The CID uses MOS capacitors like in CCDs but allow X-Y addressing and non-destructive readout [Theuwissen 1995]. The SIT achieves current amplification by placing a light sensitive MOS capacitor on top of a bipolar transistor. A CMD sensor, developed by Olympus, consists of a MOSFET structure where photogenerated charges collected under the gate of the device modulates the current flowing through the transistor [Hornsey 1999a]. Amplification is achieved and the device is compact requiring only two transistors per pixel but suffers from large dark current and fixed pattern noise. The pinned photodiode¹⁷ developed by JPL/Kodak offers high quantum efficiency, low dark current and low noise readout [Fossum 1997]. However none of these sensors are fully compatible on a standard CMOS process and additional fabrication steps are required.

Several efforts have been made to combine CCD and CMOS processes to make use of their relative advantages, in particular the better imaging quality of CCD sensors with the high level of integration of the CMOS process. However this is not without its difficulties [Hornsey 1999b, Moini 1999]. CCD/CMOS processes do not provide an optimised CCD structure. In fact, neither process is fully optimised in a combined device and the approach represents more of a compromise than an improvement. Also, the high clocking pulses needed for CCD operation induces noise into any circuitry that is integrated. Being highly capacitive devices, CCD structures will cause adjacent CMOS circuits to dissipate too much power. Furthermore, combining CMOS and CCD processes to obtain the best of both worlds would require almost all the stages

¹⁷ Pinned photodiode has a p^+np^+ structure where the voltage applied to the n-layer fully depletes the n-layer and the voltage is pinned. Photogenerated majority carriers are then stored in this depletion region decreasing the pinned voltage. This is different than a p-i-n photodiode which utilizes an intrinsic layer between a p-layer and n-layer (typically $p^+n^-n^+$) and photogenerated minority carriers are swept across the depletion region and collected by electrodes connected to the p and n-layers.

from both processes, which means an excessive number of fabrication masks and the resulting process tends to be more expensive than either the standard CMOS process or CCD manufacturing. High volume production is highly unlikely.

One of the approaches taken by NASA is their concept of "Hybrid imaging technology" (HIT). Instead of uniting CCD and CMOS devices at the device-fabrication-process level, the devices are fabricated separately and then joined mechanically and electrically (hybridized) by standard bump bonding techniques where indium bumps are deposited on matching bump-bond pads formed on the CCD imager and CMOS chips.

1.5.5.5 Thin Film on ASIC (TFA)

TFA (Thin Film on ASIC) image sensors consists of a hydrogenated amorphous silicon (a-Si:H) photodiode with the a-Si:H layers directly deposited on the CMOS chip to give fill factors approaching 100% [Wong 1996]. Furthermore, detectors and electronic circuitry can be developed independently, with the potential of obtaining very low dark currents due to the higher energy gap of a-Si:H (1.75eV). However this represents a relatively immature technology and is not widely available, hence costs are still high. But for downscaled processes this technology holds great promise.

1.6 PIXEL ARCHITECTURES IN CMOS

Of the technologies discussed, CMOS offers the highest possibility of on-chip processing at a reasonable cost and performance. This section describes the possible pixel architectures in a CMOS process. Readout for CMOS photodetector structures can either be made in the direct readout mode or in the charge integration mode. The advantage of charge integration readout is that it offers higher signal sensitivity [Fossum 1997] and allows the dynamic range to be controlled by changing integration times and it has low sensitivity to device mismatch because the integration time depends on the input capacitance, which has less mismatch than other parameters of the circuit [Moini 1999]. Also it has a linear transfer characteristic and integration acts as a low-pass filter which removes the high frequency components of the noise.

1.6.1 PASSIVE PIXEL SENSORS (PPS)

The passive pixel sensor (PPS) first introduced by Weckler in 1967 [Weckler 1967] represents the early form of the CMOS imager and is responsible for much of its initial criticism due to its poor noise performance. Passive pixel sensors have one transistor per pixel for addressing purposes as shown in Figure 1.22. Operating the passive pixel sensor in a direct or continuous mode usually involves the use of a transimpedance amplifier, with the feedback resistance providing the current-to-voltage conversion. However, this technique does not lend itself to on-chip integration due to the difficulty in incorporating the large feedback resistance required. A more common approach is to operate the sensor in a charge integration mode using a charge amplifier with feedback capacitance at the column or chip level [Hornsey 1999b] as in Figure 1.22.

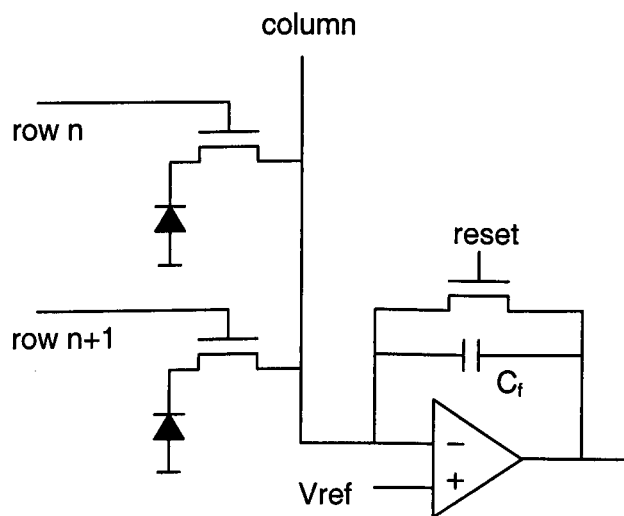


Figure 1.22 Passive pixel sensors with column-level charge amplifier readout circuitry

The photocharge integrated on the photodiode capacitance is transferred to the feedback capacitance of the charge amplifier and output as a voltage. Gain is provided by the ratio of the photodiode capacitance to the feedback resistance. With passive pixel sensors, parasitic capacitances of the data line is a major concern as it limits the speed at which the pixel can be read out, increases readout noise as well as reduces the charge seen at the output. As such, passive pixel devices does not scale well to larger array sizes and is not usually the architecture of choice except where fill factor is a limitation or current readout is desired. Integration of a charge amplifier at the column

level has the advantage of reduced bus capacitance but the disadvantage of mismatches between the amplifiers and limited space, and hence performance, available for each amplifier.

1.6.2 ACTIVE PIXEL SENSORS (APS)

Active pixel sensors incorporate an active amplifier or buffer at each pixel, typically a source follower, to overcome the large bus capacitance of the passive pixel sensors. The initial problem with active pixel sensors was the poor fill factor caused by the incorporation of the on-pixel amplifier but decreasing feature sizes means more and more functionality can now be built into a single pixel. Pixels as small as 4 microns have been fabricated [Endo 2003]. There are three major types of active pixel sensors, namely the photodiode APS, photogate APS and logarithmic APS.

1.6.2.1 Photodiode APS

The structure of a photodiode APS is shown in Figure 1.23. Light incident on the photodiode generates charge carriers which are collected on the photodiode capacitance. After the integration time has elapsed the voltage on the capacitor is read out and is linearly related to the charge collected and hence the incident illumination. After readout, the reset line is pulsed high to reset the photodiode to the supply voltage. The integration may then be repeated.

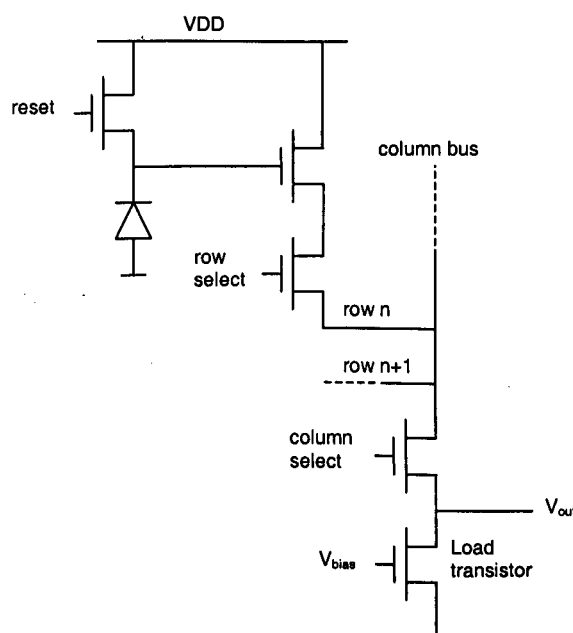


Figure 1.23 A photodiode APS with array row/column selection

Due to the added circuitry and their threshold drops, the dynamic range of an active pixel sensor is normally limited by the voltage swing of the circuit rather than the full well capacity of the photodiode. Methods to extend this output swing include using a complementary PMOS readout structure in addition to the regular NMOS source follower readout structure [Xu 2002]. This however causes a reduction in fill factor. Numerous modifications to the basic photodiode APS have been carried out in order to improve its functionality or its performance, as detailed in [Fossum 1997].

1.6.2.2 Photogate APS

A photogate APS is based on a CCD device where photogenerated charge is collected in a potential well when a voltage is applied to the photogate (PG). The structure of the photogate APS is shown in Figure 1.24. After integration, the floating diffusion is reset, and its reset voltage is stored. A transfer gate is then pulsed to transfer the stored photogenerated charge to the floating diffusion and this voltage is then read. Readout of the reset and signal voltages are performed through a source follower buffer and a row select transistor like in the photodiode APS of Figure 1.23. The difference in the reset and signal voltages is the output of the sensor. This approach is called correlated double sampling (see Section 1.63) and it suppresses reset noise, $1/f$ noise and FPN. However the photogate APS has a lower fill factor, higher mismatch¹⁸ and lower quantum efficiency, particularly in the blue, than the photodiode APS due to the additional circuitry and the overlying polysilicon gate. However, it has better noise suppression and charge conversion efficiency¹⁹ making it suitable for low-light level applications.

¹⁸ This is due to the surface states at the Si-SiO₂ interface contributing to the recombination of stored carriers

¹⁹ This is because it has a separate smaller output node (floating diffusion) which means a smaller capacitance (Charge conversion efficiency = q/C in V/e^-)

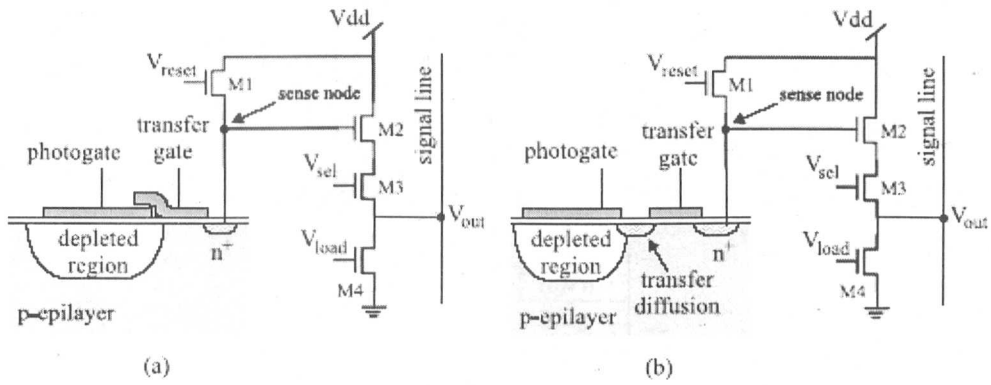


Figure 1.24 Photogate APS with (a) overlapping transfer gate and (b) with n+ transfer diffusion [de Lima Monteiro 2002]

Ideally the transfer gate should overlap the photogate to ensure effective charge transfer. This would require a double poly process. However, the need for an additional gate can be avoided by utilizing an intermediate 'bridging' diffusion, as shown in Figure 1.24 (b). This has little effect on the performance of the pixel except for the possible introduction of image lag [Mendis 1997].

1.6.2.3 Logarithmic APS

The logarithmic pixel is a modification of the linear photodiode active pixel sensor where the gate of the reset transistor is connected to the supply voltage giving continuous readout of the photocurrent and is depicted in Figure 1.25. The small photocurrent causes the reset transistor to operate in the weak inversion or subthreshold region where the MOS current flow is dependent upon the exponential of V_{DS} . The voltage at the photodiode node therefore varies logarithmically with the photocurrent, giving the pixel a very large dynamic range and can be expressed by the following equation [Hornsey 1999b]:

$$V_s = VDD - \frac{kT}{q} \ln \left(\frac{i_{photo}}{i_o} \right) \quad (1.18)$$

where k is the Boltzmann constant, T is the absolute temperature in Kelvin, q is the electron charge, VDD is the supply voltage, i_{photo} is the generated photocurrent and i_o is a process dependent parameter.

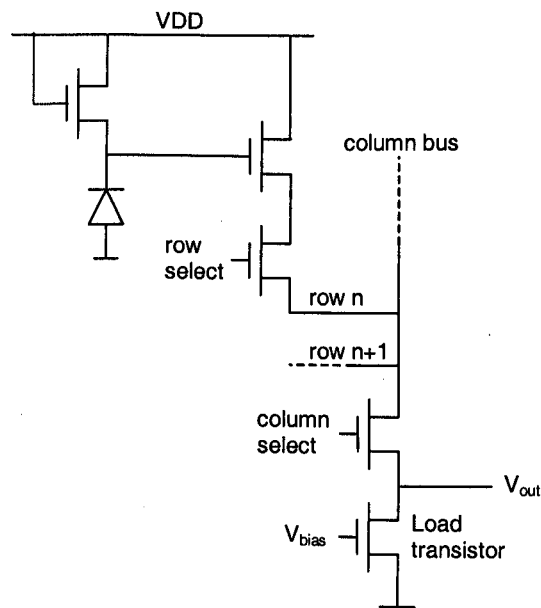


Figure 1.25 Logarithmic APS

Logarithmic pixels can measure illumination over 5 orders of magnitude, an order of magnitude more than ordinary APS [Hornsey 1999b]. In addition, logarithmic pixels do not require a reset line and have simpler timing and operation as well as larger fill factor. Since logarithmic pixels operate in continuous time, they are randomly accessible both in time and in space. This also means they are able to operate at a higher sampling rate. On the downside, because of the subthreshold operation of the MOSFET and its dependence on temperature and process parameters such as threshold voltage and oxide thickness, logarithmic sensors suffer from large pixel offset non-uniformity or FPN. So though its dynamic range is larger, typically logarithmic pixels have lower SNR. This FPN cannot be removed by correlated double sampling because of its continuous time operation. This offset, however, can be removed by storing the offset in memory and subtracting when the pixel is read. It can be performed by software but for the highest possible speed, a parallel hardware correction method is used. Dierickx et. al. [Dierickx 1996] used an external PROM and a dedicated co-processor while Ricquier et. al. [Ricquier 1995] performed the non-uniformity correction on-chip.

Another disadvantage of the logarithmic APS is its speed under low illumination levels because of the small photocurrent available for charging/discharging of the sensing node [Hornsey 1999b]. Delbrück, however, used feedback to improve the

speed response. An adaptive element was also used in order to give compression for slowly varying signals and higher gain for larger frequencies making it useful for biological vision systems and motion detection [Moini 1999]. In fact, logarithmic sensors are the preferred sensors for modelling biologically inspired vision systems as it mimics its large dynamic range response.

An inverted logarithmic APS structure, where the positions of the photodiode and the load (transistor in subthreshold) is reversed, was used to reduce pattern noise and improve output voltage swing by reducing signal compression [Hong 2001]. The electrical sensitivity of the conventional structure can be improved by increasing the number of subthreshold diode connected MOS transistors (MOS diodes) in the pixel at the expense of reduced fill factor and speed of response. With an inverted structure the effect is less pronounced (no increase in sensitivity) but instead the subthreshold region of operation is extended over a wider region offering an even larger dynamic range. Again at the expense of reduced fill factor.

1.6.3 NOISE REMOVAL AND EXTENDING DYNAMIC RANGE

Noise has been the weak point of CMOS imagers compared to the highly sensitive CCDs. However, there are various means to achieve noise removal in CMOS sensors. New techniques are constantly being developed but two well established methods are the Correlated Double Sampling (CDS) and Delta-Difference Sampling (DDS) techniques for removing the 'kTC' reset noise²⁰ and FPN. Figure 1.26 shows the typical circuit for performing CDS and DDS.

²⁰ It is known as kTC noise because the number of noise electrons generated $n = \frac{\sqrt{kTC}}{q}$, though the

noise voltage at the output is given by $V_n = \sqrt{\frac{kT}{C}}$ (from $Q=CV$). Since signal electrons increases proportionately with area but reset noise electrons increase as a square root of area (capacitance), SNR improves with a larger photodetection area.

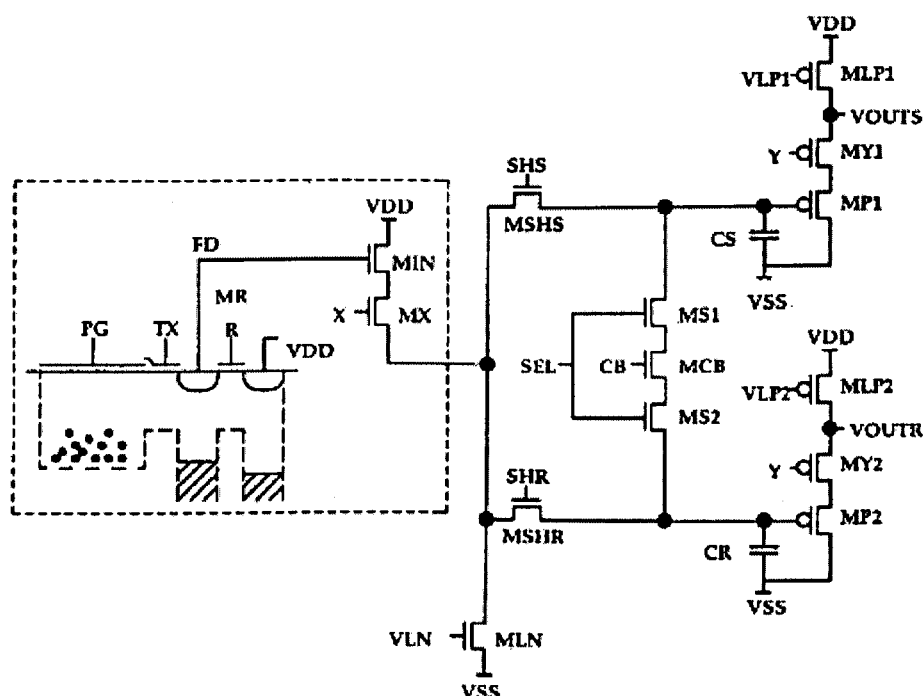


Figure 1.26 Correlated Double Sampling (CDS) and Delta-Difference Sampling (DDS) applied to a photogate APS [Mendis 1997]

Correlated double sampling is usually performed at the column level and works by differentially reading out the reset and signal levels. However, due to the threshold voltage variations between the two readout circuits, column-wise FPN is generated. Delta-difference sampling removes this by shorting the two sample and hold capacitors (by pulsing CB and SEL in Figure 1.26) and taking another differential reading. This reading is proportional to the threshold voltage difference between the two circuits and subtracting this from the initial reading gives the final offset free output.

Reset noise of an APS is the thermal noise (see equation (1.16)) associated with the finite resistance of the reset switch. This noise is transferred to the capacitor when the reset switch opens. In the case of a photogate APS, the 'kTC' noise freezes when the reset transistor switches off because the effective noise bandwidth, $B = 1/4RC$ (equation (1.15)), drops significantly ($R_{off} \gg R_{on}$). Figure 1.27 (a) illustrates this. However, in a photodiode APS the charge is integrated on the output node such that when the reset signal goes low the photodiode immediately starts discharging the stored charge. Removal of reset noise would require sampling right at the instant reset

is switched off and the photodiode discharges, which is difficult to do. However, it is still possible to use double sampling (not correlated) to remove $1/f$ noise and fixed pattern noise from the photodiode pixel.

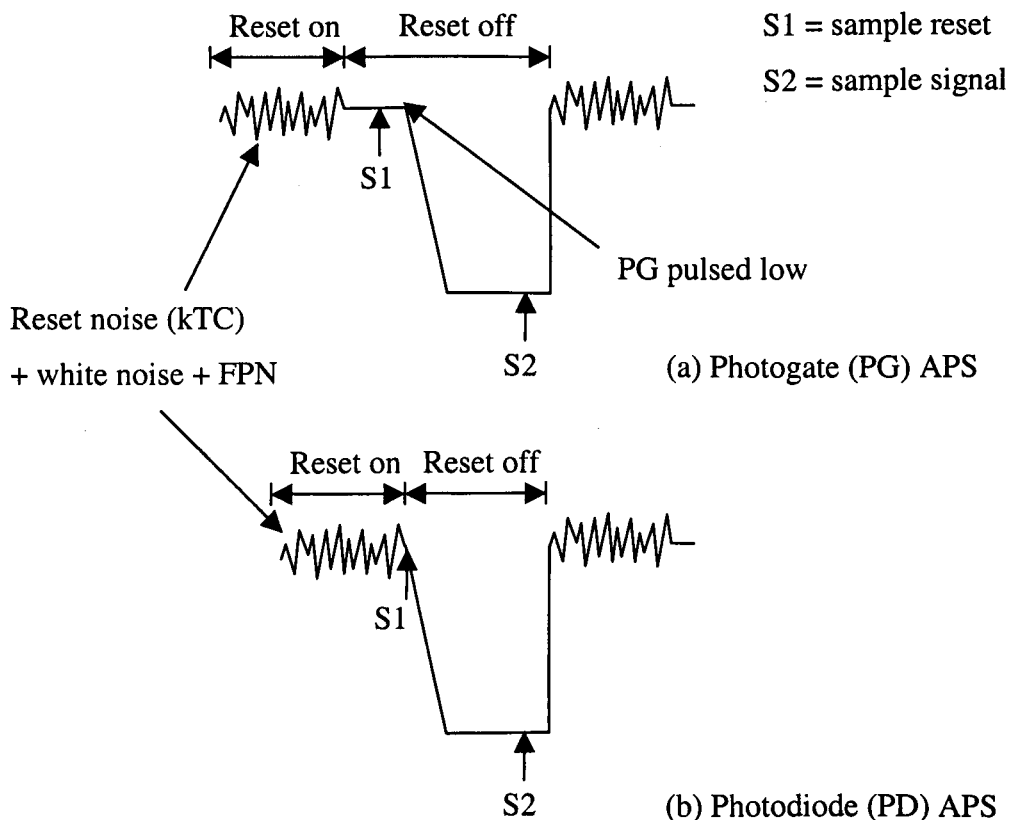


Figure 1.27 Double sampling to remove noise in (a) a photogate APS and (b) a photodiode APS

Besides noise, another important characteristic of image sensors is the dynamic range. There are several means to extend the dynamic range of active pixel sensors [Yadid-Pecht 1999]. These include the logarithmic pixel discussed previously, multi mode sensors, clipping a sensor's response, having a variable integration time [Yasuda 2003], and conversion of the sensor output to a pulse frequency [Yang 1994]. Multi mode sensors allow the photodetector structure used to be operated under different modes. One such example makes use of the fact that it is simple to switch between the linear and logarithmic mode of the active pixel sensor by proper biasing of the reset/subthreshold transistor. This has been commercially marketed under the label LINLOG technology by Photonfocus AG and it uses a linear response at low illumination levels and logarithmic compression at high intensities. Clipping sensors

have anti-blooming structures that bleed off excess charge as it builds up. Control of integration time to extend dynamic range works on the fact that increasing integration time allows more charge to be stored in the pixel and this can be done either globally or locally. The advantage of controlling the integration time locally is that if the scene being captured consists of different illumination levels, the dynamic range at the brighter part of the scene is extended while the resolution at the darker regions is maintained. Most dynamic range enhancement efforts, specifically those requiring on-pixel circuitry, suffer from reduced fill factor, sensitivity and spatial resolution as well as increased mismatch.

It is clear that the ability to integrate circuitry on-chip with CMOS imagers has opened the doors to a wide range of applications and possibilities. Its flexibility has meant enhanced functionality of devices. From adaptive photocircuits and foveated pixels for robotic vision [Moini 1999], to unique readout and pixel reset structures [Yadid-Pecht 2003], to pixel-level ADCs for high frame rates [Kleinfelder 2001], to on-chip or in-pixel analogue memory [Simoni 1995] for motion detection, extended dynamic range and electronic shuttering; the possibilities seem endless for CMOS imaging.

1.7 CHAPTER SUMMARY

This chapter has emphasized the need for adaptive optics (AO) highlighting several key application areas where low cost real-time AO systems would be useful such as astronomy, ophthalmology, intra and extra-cavity laser correction, free space optical communications and microscopy. A fundamental part of any AO system is the wavefront sensor and with current Shack-Hartmann wavefront sensors, conventional imagers are used with limited frame rates ranging from 25 to 60 Hz. Using a dedicated CCD increases the frame rate but at the expense of increased cost, and the need for an image-processing step and special hardware still remains [de Lima Monteiro 2002]. In this thesis, a solution to the data bottleneck is proposed by integrating local centroid processing at the detector level.

There are several possible structures for implementing a position sensitive device (PSD) such as the lateral effect photodiode (LEP), the quad cell and the multi-pixel array. A lateral effect PSD requires large uniform sheet resistance for linear operation, which is not readily available in a standard CMOS process making integration with circuitry difficult. Quad cells have simple readout schemes but are not very linear. Multi pixel arrays have better linearity and positional range, which translates to larger tilt measurement capability. They also offer greater flexibility and are able to deal with multiple spots and non-uniform intensity profiles. The drawback is the increased computational load but for moderate array sizes this is reasonable and this was the architecture chosen for our system. A 5 x 5 pixel array was selected as a tradeoff between linearity and circuit complexity.

Several technological options were highlighted and the standard CMOS process was chosen as the technology of choice as it allows high levels of circuit integration needed to implement the local centroid processing. There have been various efforts to implement centroid detection on a CMOS process for numerous applications. In general, analogue multi-pixel array approaches suffer from low fill factor and sensitivity, requiring either separate x and y pixels or on-pixel circuitry such as a comparator or resistors. In addition, binary position sensing techniques using Winner Take All (WTA) circuitry or an on-pixel comparator do not offer subpixel accuracy and cannot cope with multiple spots or non-uniform spots. A dedicated digital centroid processor has yet to be demonstrated to date, though several generic image processors exist, and this research explores this approach. A dedicated digital centroid processor offers high accuracy and greater flexibility. Also the processor can be made programmable and additional image processing tasks can easily implemented if necessary.

The fundamentals of the photodetection mechanism were described along with issues of response, noise and operation. The junction photodiode structure was chosen as the basis of the imaging component as it is readily available in a standard CMOS process and offers good quantum efficiency as well as high linearity and dynamic range. In terms of pixel architectures, the CMOS active pixel sensor (APS) was selected as it offers high fill factor and low mismatch compared to other APS types. Ideally, the

pixel size has to be sufficiently large in order to achieve a large fill factor and sufficient tilt dynamic range. A large fill factor also means less mismatch.

In summary, in the proposed design each tilt sensor will consist of: i) a 5 x 5 photodiode active pixel sensor array in a standard CMOS process ii) a dedicated on-chip digital centroid processor to remove the data bottleneck. A discussion of the data bottleneck in current CCD systems and how our system addresses this is given in Appendix A1.1. The following chapters of this thesis will cover the design, fabrication and implementation of the proposed system.

Chapter 2 will discuss the results from the characterisation of fabricated full custom photodiodes in a standard CMOS process. Their suitability and performance are assessed. Chapter 3 then describes the use of a hardware emulation system to validate the functionality of the design prior to committing the design to silicon. The emulation system consists of a photodiode array as the front-end for light detection and a Field Programmable Gate Array (FPGA) as the digital backend that performs the centroid computation. The system was tested using both a commercial photodiode array and a fabricated full custom CMOS photodiode array. Chapter 4 then details the integration of a full custom CMOS photodiode array with on-chip digital centroid processing. Chapter 5 discusses the reconstruction of an optical wavefront from an array of centroid data and finally Chapter 6 will offer some concluding remarks and some discussions on possible further developments and improvements.

CHAPTER 2

CHARACTERISATION OF CMOS PHOTODIODES

2.1 INTRODUCTION

In the design of any complete system, particularly in VLSI, the individual parts of the system needs to be evaluated and characterised before the complete system is fabricated. One of the fundamental building blocks of any optoelectronic system is that of the photodetector. As mentioned in the previous chapter, these photodetectors are formed in a CMOS process by the generation of a p-n junction and are typically the “well-substrate” or “diffusion-substrate” or the “diffusion-well” photodiode types. This chapter covers the characterisation of these discrete photodiodes and the selection of the optimum device prior to the addition of any circuitry or processing.

2.2 FABRICATION OF TEST STRUCTURES

The process used for the test structures and also for the fabrication of the centroid processor is the Alcatel Microelectronics (Mietec)²¹ 0.7 μ m self-aligned twin-well, single-poly, double-metal layer CMOS process with LOCOS isolation [Europractice IC Service]. This process is accessed via IMEC in Belgium through the Europractice IC Service. The Europractice Multi-Project Wafer (MPW) service enables the prototyping to be carried out at a reduced cost. The main electrical and physical parameters of this process such as the resistivity, threshold voltage and transistor transconductance are highlighted in Appendix A2.1. However, to give a clearer view

²¹ Now known as AMI Semiconductor (AMIS) after AMIS acquired Alcatel Microelectronics' mixed-signal business activities from STMicroelectronics.

of the characterisation results the junction depths of the process have been illustrated in a typical CMOS cross section shown in Figure 2.1.

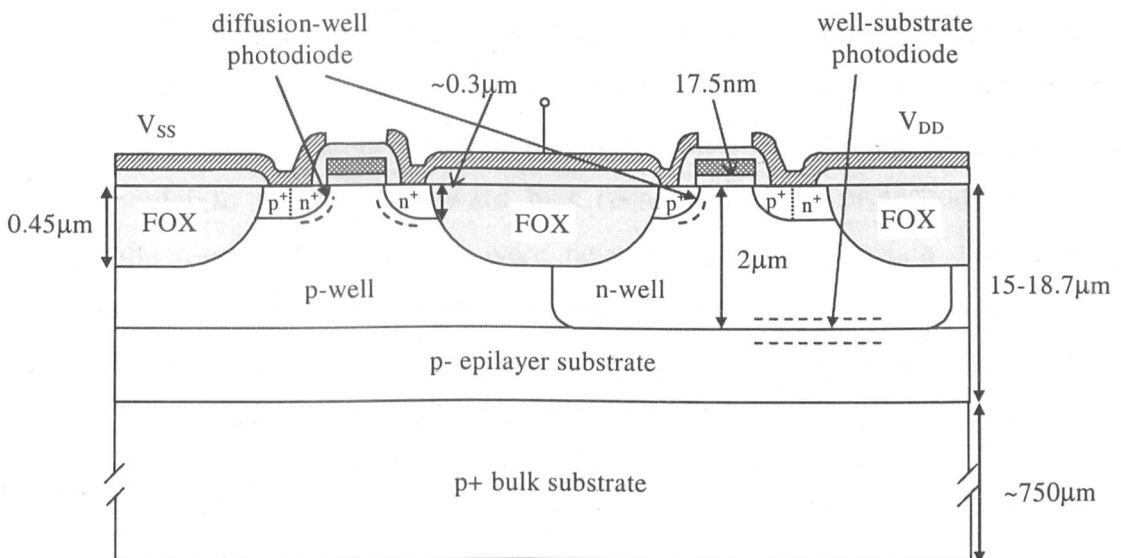


Figure 2.1 Junction depths of the Mietec 0.7 μ m CMOS process²²

2.2.1 FIRST CHARACTERISATION CHIP (PDFINAL)

Figure 2.2 shows the layout of the first test chip PDFinal. This chip contained the following: 1. well-substrate photodiodes 2. diffusion-substrate photodiode 3. combined well-substrate and diffusion-well photodiode 4. lateral effect photodiode (LEP) 5. active pixel sensor 6. 5-by-5 array of combined well-substrate and diffusion-well photodiodes. The various junction photodiodes were included in order to determine their relative response and characteristics as well as their individual variation with area and periphery. In addition, a combined device was designed and included in order to capture a longer range of wavelengths than either the well-substrate (deep) photodiode or the diffusion-substrate (shallow) photodiode, and will be discussed further in Section 2.2.3. The LEP is commonly used for position sensing as a custom device and was included for characterisation in a CMOS process but was not used in this work as the multi-pixel array approach was chosen for our application. Finally, the 5-by-5 photodiode array was included for use in the hardware emulation system of the centroid processor which is described in Chapter 3.

²² The term p-substrate will be used frequently in this thesis and this will refer to the p-epilayer substrate and not the bulk substrate.

Initial characterisation of this chip showed several issues. Firstly light being absorbed in the substrate and diffusing to the photodiode active region gave rise to crosstalk and a larger signal than expected. Secondly the pads used were those available in the library and contained a diode protection structure (see Figure 2.3), which if biased incorrectly may interfere with the characterisation of the raw devices. Also when operating the photodiode in reverse bias it was necessary to power up the protection circuit in order to avoid any forward bias current from the protection structure affecting the results. However, we were nevertheless able to obtain satisfactory responsivity values from the photodiodes and the array on this chip allowed us to proceed with the development of the centroiding system, as will be discussed in the next chapter. The chip size was $2513.8\mu\text{m} \times 2412.2\mu\text{m}$ (Area = 6.0638mm^2) and was packaged in a 44-pin ceramic J-led chip carrier (JLCC 44).

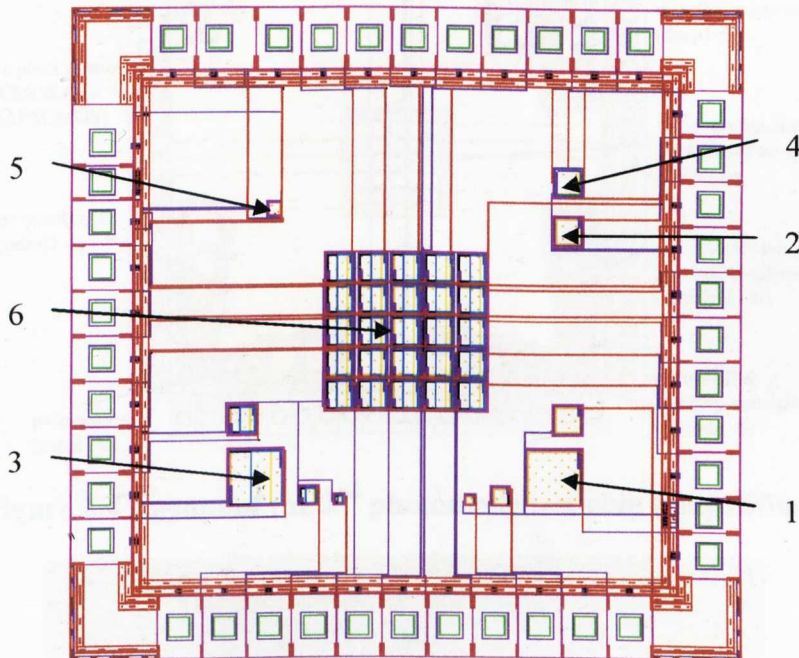


Figure 2.2 Layout of 1st photodiode test chip (PDfinal)

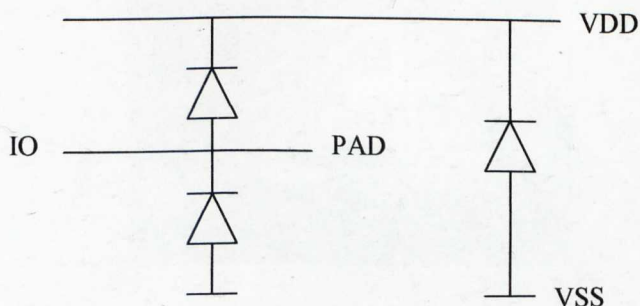


Figure 2.3 Diode protection structure present in pads used in PDfinal

2.2.2 SECOND CHARACTERISATION CHIP (CHIP1BFINAL)

Figure 2.4 shows the layout of the 2nd characterisation chip (chip1bfinal). In this chip several changes were made. Firstly, a metal light shield surrounding each structure was incorporated. However the process required that holes be included in the metal every 25 μm to relieve mechanical stress, which meant total blockage was not possible. Secondly, more structures were incorporated and the number of different sized devices for each structure was increased in order to better determine any area and perimeter scaling effects. Finally it was necessary to design a pad without any additional circuitry on it to allow accurate characterisation of the photodiode test structures.

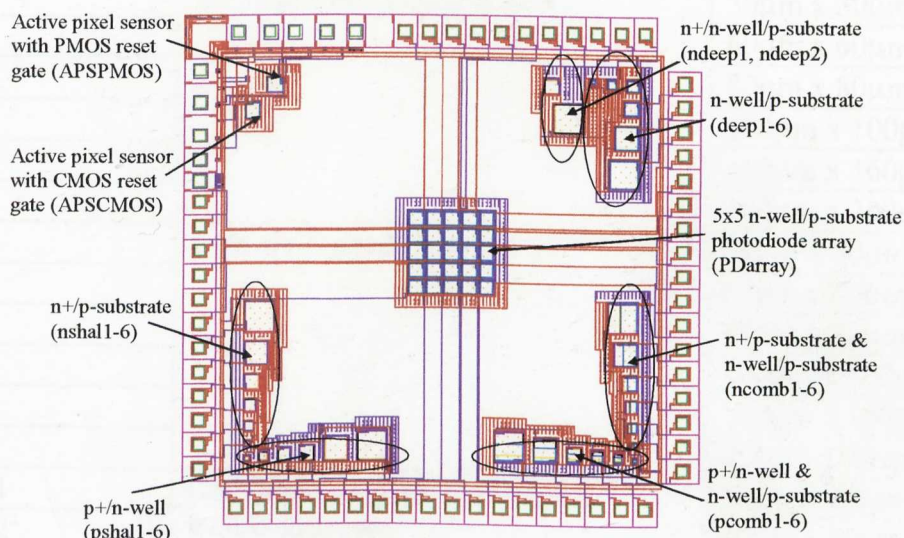


Figure 2.4 Layout of the 2nd photodiode test chip (chip1bfinal)

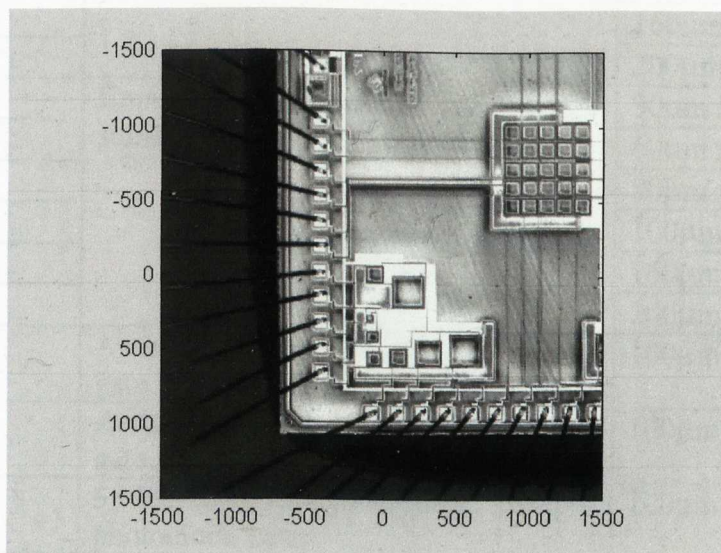


Figure 2.5 Optical image of well-substrate photodiodes on test chip (transposed)

The size of this chip was $3640\mu\text{m} \times 3543\mu\text{m}$, which is an area of approximately 12.9mm^2 and was packaged in a 68-pin ceramic J-leaded chip carrier (JLCC 68). The full list of devices present on this chip are summarised in Table 2.1 and will be referred to by its assigned short name from henceforth - and is also used in Figure 2.4.

Short name	Photodiode type	Size of photodiode
deep1	n-well/p-substrate with n+ removed (deep)	$30\mu\text{m} \times 30\mu\text{m}$
deep2		$60\mu\text{m} \times 60\mu\text{m}$
deep3		$80\mu\text{m} \times 80\mu\text{m}$
deep4		$100\mu\text{m} \times 100\mu\text{m}$
deep5		$160\mu\text{m} \times 160\mu\text{m}$
deep6		$200\mu\text{m} \times 200\mu\text{m}$
ndeep1	n-well/p-substrate with n+ across (deep with n+)	$100\mu\text{m} \times 100\mu\text{m}$
ndeep2		$200\mu\text{m} \times 200\mu\text{m}$
nshal1	n+/p-substrate (shallow n+)	$30\mu\text{m} \times 30\mu\text{m}$
nshal2		$60\mu\text{m} \times 60\mu\text{m}$
nshal3		$80\mu\text{m} \times 80\mu\text{m}$
nshal4		$100\mu\text{m} \times 100\mu\text{m}$
nshal5		$160\mu\text{m} \times 160\mu\text{m}$
nshal6		$200\mu\text{m} \times 200\mu\text{m}$
pshal1	p+/n-well (shallow p+)	$30\mu\text{m} \times 30\mu\text{m}$
pshal2		$60\mu\text{m} \times 60\mu\text{m}$
pshal3		$80\mu\text{m} \times 80\mu\text{m}$
pshal4		$100\mu\text{m} \times 100\mu\text{m}$
pshal5		$160\mu\text{m} \times 160\mu\text{m}$
pshal6		$200\mu\text{m} \times 200\mu\text{m}$
ncomb1	Combined n+/p-substrate and n-well/p-substrate	$30\mu\text{m} \times 30\mu\text{m}$
ncomb2		$60\mu\text{m} \times 60\mu\text{m}$
ncomb3		$80\mu\text{m} \times 80\mu\text{m}$
ncomb4		$100\mu\text{m} \times 100\mu\text{m}$
ncomb5		$160\mu\text{m} \times 160\mu\text{m}$
ncomb6		$200\mu\text{m} \times 200\mu\text{m}$
pcomb1	Combined p+/n-well and n-well/p-substrate	$30\mu\text{m} \times 30\mu\text{m}$
pcomb2		$60\mu\text{m} \times 60\mu\text{m}$
pcomb3		$80\mu\text{m} \times 80\mu\text{m}$
pcomb4		$100\mu\text{m} \times 100\mu\text{m}$
pcomb5		$160\mu\text{m} \times 160\mu\text{m}$
pcomb6		$200\mu\text{m} \times 200\mu\text{m}$
APSPMOS	Active pixel sensor with PMOS reset gate (n-well/p-substrate photodiode)	$100\mu\text{m} \times 100\mu\text{m}$
APSCMOS	Active pixel sensor with CMOS reset gate (n-well/p-substrate photodiode)	$100\mu\text{m} \times 100\mu\text{m}$
PDarray	5 by 5 n-well/p-substrate (deep4) photodiode array	$100\mu\text{m} \times 100\mu\text{m}$

Table 2.1 Devices present in characterisation chip 'chip1bfinal'

Figure 2.6 shows the layout and cross section of the deep photodiode with n^+ removed from the active region except for the cathode contact region (i.e. deep1-6 and those used in PDarray). In addition, the simplified cross sections of the other photodiode types present on the chip are shown in Figure 2.7.

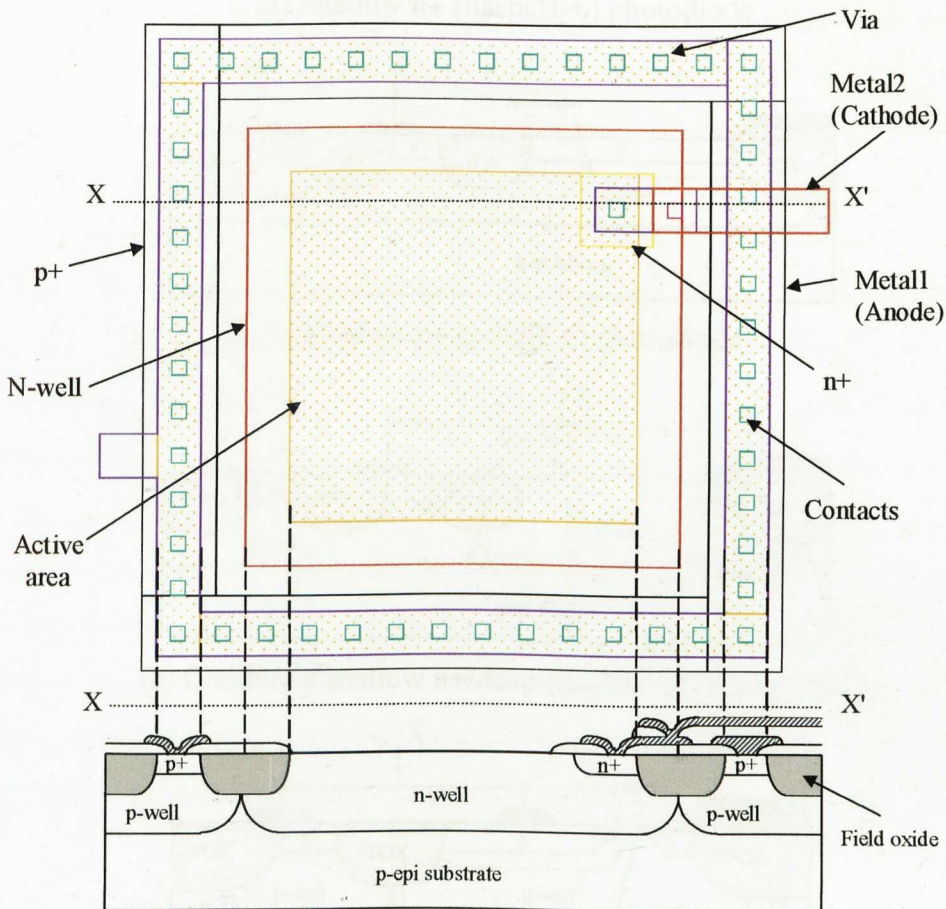
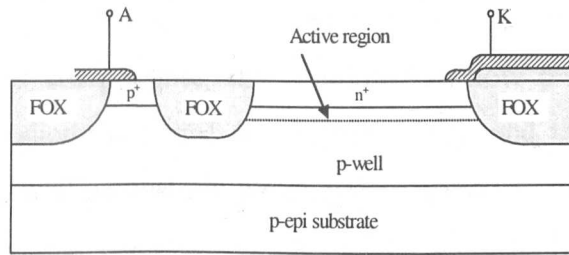
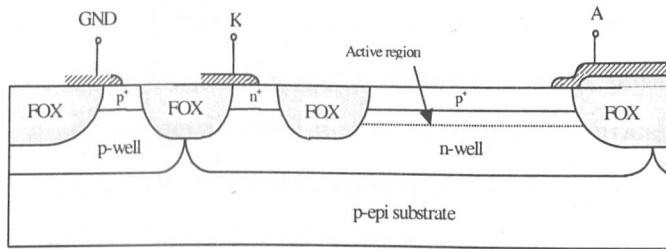


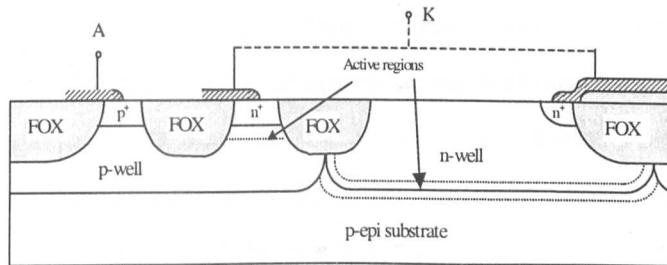
Figure 2.6 Layout and cross section of fabricated well-substrate (deep1-6 and PDarray) photodiode



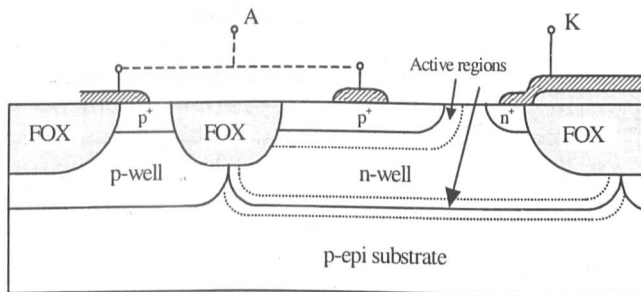
(a) Shallow n+ (nashal1-6) photodiode



(b) Shallow p+ (pshal1-6) photodiode



(c) Combined shallow n+/deep (ncomb1-6) photodiode



(d) Combined shallow p+/deep (pcomb1-6) photodiode

Figure 2.7 Other photodiodes present on the characterisation chip

2.2.3 COMBINED PHOTODIODES

In Section 1.4.1 we observed that the absorption depth of a photon depends upon the wavelength. As a result the two different junction depths (at $0.3\mu\text{m}$ and $2\mu\text{m}$) in theory lend themselves to being sensitive to different wavelengths. Hence the combined devices were designed and fabricated in order to extend the spectral

response of the typical photodiode over a wider range of wavelengths. Two types of combined photodiodes are possible, the combined shallow p⁺/n-well and deep n-well/p-substrate photodiode (pcomb1-6 – Figure 2.7(d)) and the combined shallow n⁺/p-well and deep n-well/p-substrate photodiode (ncomb1-6 – Figure 2.7(c)). In the first chip, the former was included. In the second chip, both were included. The ncomb devices are in effect two discrete photodiodes laterally adjacent to each other and will not provide any additional advantages when focused light is used such as in the intended application. In fact, in the intended application of finding a centroid, the use of this device would be detrimental due to its non-symmetrical spatial response. As for the pcomb devices, when tested in room light, these devices were found to be very leaky with increasing reverse bias voltage. This observed effect is shown in Figure 2.8. This is believed to be due to the depletion region being formed at the surface of the photodiode leading to a large leakage current as a result of the large electrical field caused by increased mechanical stress and increased number of surface traps present [Bogaerts 2000, Pain 2001]. As the reverse bias voltage increases the leakage current increases as a result of the larger depletion width. As a consequence of these observations, the majority of the characterisation work presented henceforth will be focused on the deep (Figure 2.6) and shallow devices (Figure 2.7(a) and 2.7(b)). However, results on these combined devices will be presented where deemed relevant to highlight its uniqueness or simply for completeness.

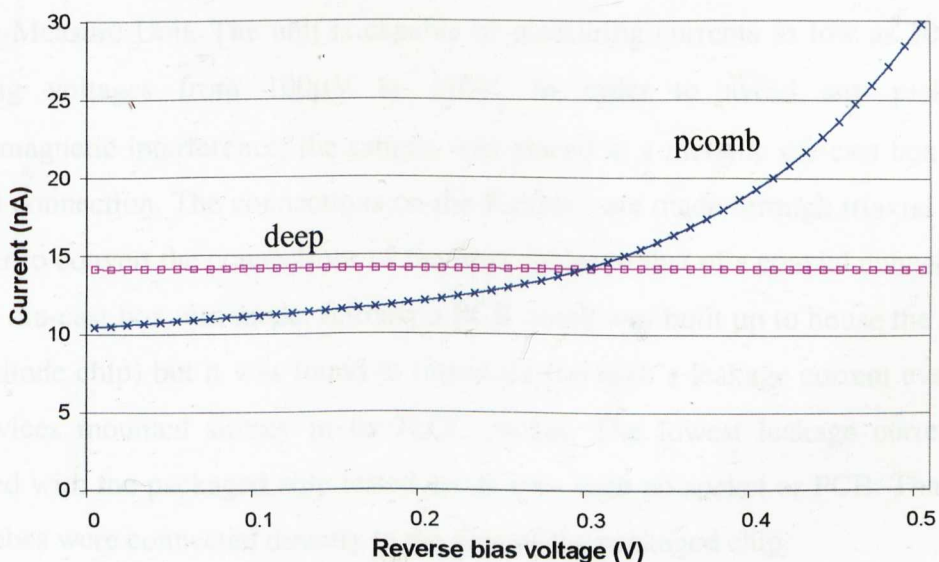


Figure 2.8 I-V response in room light showing increased leakage of combined shallow p⁺/deep (pcomb1-6) device compared to the deep (deep1-6) device for reverse bias operation (PD_{final})

2.3 DARK RESPONSE OF PHOTODIODES

The response of a photodiode can be evaluated under dark or illuminated conditions. Its response in the dark is assessed by its current-voltage (I-V) characteristics and its capacitance-voltage (C-V) characteristics.

2.3.1 DARK I-V MEASUREMENTS

The dark current of a photodiode determines the smallest detectable photocurrent and hence the dynamic range achievable. Dark current also gives rise to shot noise. Hence it is necessary, particularly for low light level applications, to quantify the amount of dark current present in the system. Hence, I-V measurements of the devices under no illumination i.e. in the dark, were carried out. Note that the direction of the current and voltage on the I-V plots to be shown is such that a positive current and a positive voltage represents the photodiode operating in reverse bias i.e. in quadrant 3 of a typical I-V plot of a photodiode (the plot is therefore transposed).

2.3.1.1 Experimental setup for dark I-V measurements

Figure 2.9 shows the setup. The dark current was measured using a Keithley 236 Source-Measure Unit. The unit is capable of measuring currents as low as 10fA and sourcing voltages from 100 μ V to 110V. In order to avoid any pickup of electromagnetic interference, the sample was placed in a metallic die-cast box with a coaxial connection. The connections on the Keithley are made through triaxial cables. In order to convert the connections of the triax cables to that of a coaxial connection, a second die-cast box was made. Initially a PCB board was built up to house the sample (photodiode chip) but it was found to introduce too high a leakage current even with the devices mounted simply in its JLCC socket. The lowest leakage current was obtained with the packaged chip tested on its own with no socket or PCB. That is the test probes were connected directly to the pins of the packaged chip.

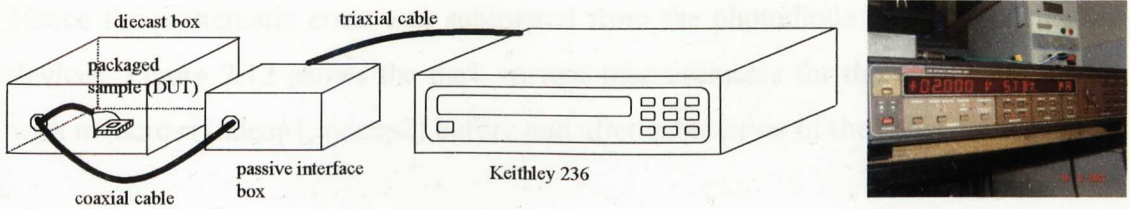


Figure 2.9 Experimental setup (left) and Keithley 236 Source-Measure Unit (right)

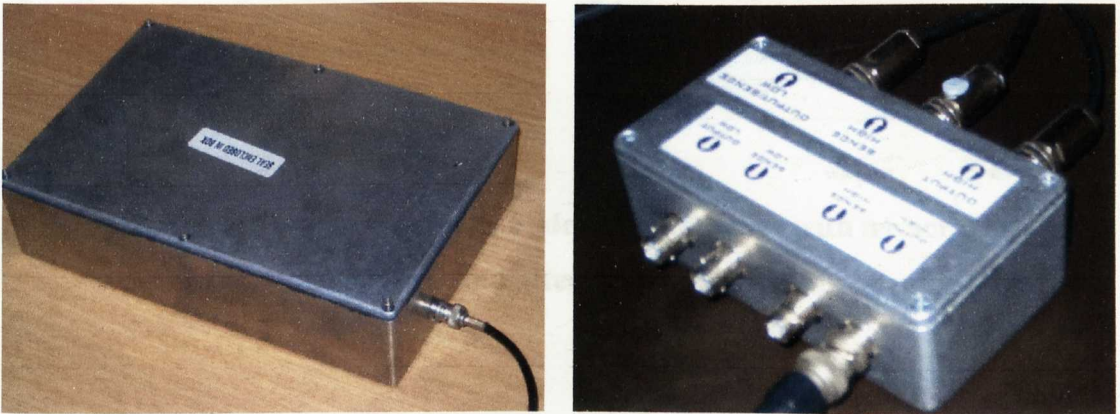


Figure 2.10 Die-cast box to hold sample (left) and die-cast box for triax to coaxial connection (right)

The measured dark current of the fabricated photodiodes is typically less than 1pA. With the added DC leakage from the cabling, packaging and housing, the actual leakage current can be much larger. As a result the system DC leakage was measured with no sample (photodiode) attached i.e. just the cables and die cast box. These results are shown in Figure 2.11. Here we can see the DC leakage of the system is of the order of 833G Ω with a 1pA offset.

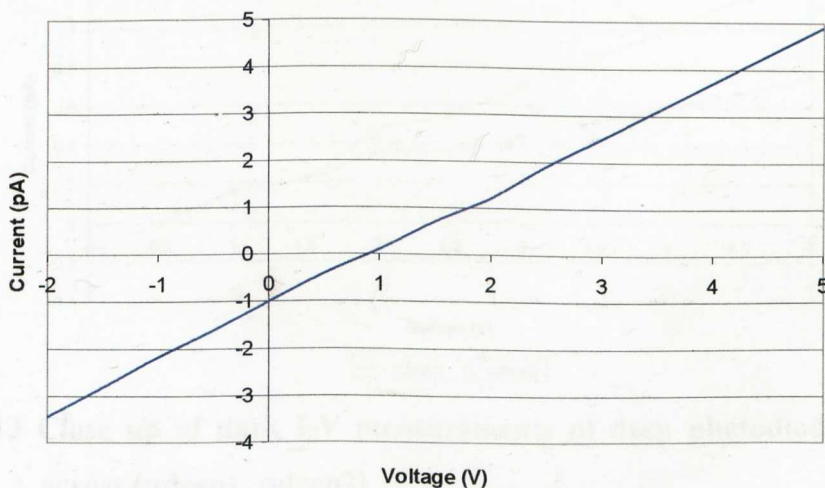


Figure 2.11 I-V measurement of the cables showing systematic error in the setup

Hence this systematic error was subtracted from the photodiode readings of the test devices. Figure 2.12 shows the dark current measurements for the deep photodiodes with n+ across (ndeep1, ndeep2) before and after subtraction of the cable offset.

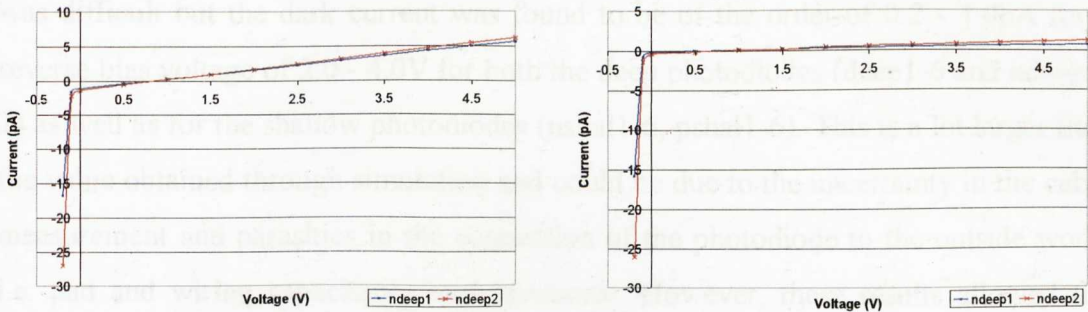


Figure 2.12 Dark I-V measurements of deep photodiodes with n+ across (ndeep1, ndeep2) before (left) and after (right) subtraction of cable offset

2.3.1.2 Results and discussion of dark I-V measurements

From a closer look at the dark current measurement of the deep photodiodes with n+ across in Figure 2.13, it can be seen that the plots do not pass through the origin indicating that a systematic error in the reading still exists. We can however see that at 2V the dark current of ndeep1 and ndeep2 is estimated to be 0.35pA and 0.5pA respectively.

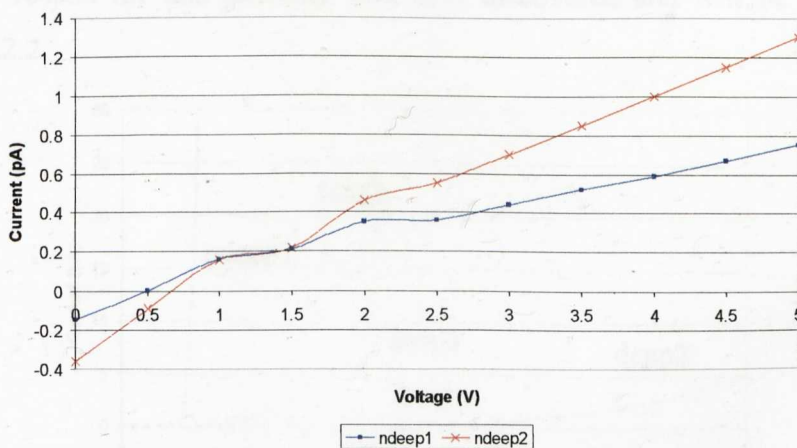


Figure 2.13 Close up of dark I-V measurements of deep photodiodes with n+ across (ndeep1, ndeep2)

It was also found that the measurements of the dark current were affected by the position of the cable in the die-cast box due to possible triboelectric effects at such low currents. Hence it was necessary to measure the offset introduced by the cable for every measurement of the sample with the cable in roughly the same position. This was difficult but the dark current was found to be of the order of 0.2 - 1.0pA for a reverse bias voltage of 2.0 - 4.0V for both the deep photodiodes (deep1-6 and ndeep1-2) as well as for the shallow photodiodes (nshal1-6, pshal1-6). This is a lot larger than the value obtained through simulation and could be due to the uncertainty in the cable measurement and parasitics in the connection of the photodiode to the outside world i.e. pad and wiring capacitance and resistance. However, these results allowed the typical measurement accuracy of the characterisation system to be determined and to obtain a figure for the dark current limits for deciding the next stage of the design.

An interesting observation was made in the forward bias currents of the deep devices with n⁺ removed (deep1-6). The forward bias current in these devices does not rise exponentially as in a typical forward biased diode but was significantly smaller. Initially, because of its somewhat linear response, this was thought to be due to a large load resistance in series with the diode introduced somewhere in the design or in the setup. However when modelled for this, it showed that this was not the case as a large resistance would make the response linear at an early stage of the bias. Forward I-V plots of deep2 and deep3 are shown in Figure 2.14 and are shown in comparison to ndeep1 and a simulation plot of a 50k Ω resistor in series with a deep3 photodiode model. The reason for this anomaly was later discovered and will be explained in Section 2.3.2.2.

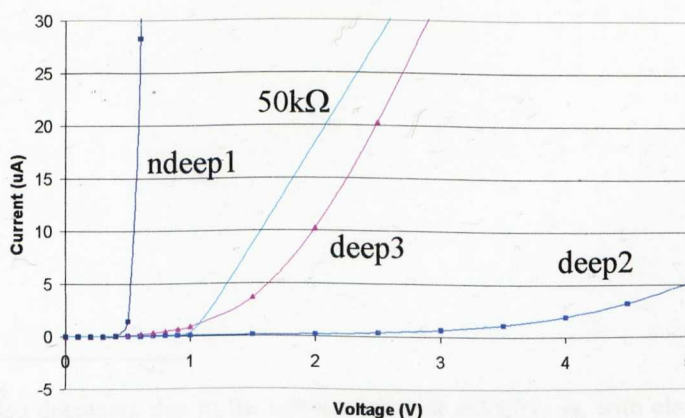


Figure 2.14 Forward bias currents of deep2 and deep3 showing abnormal behaviour compared to ndeep1

2.3.2 DARK C-V MEASUREMENTS

The response time of a photodiode is dependent on the drift time of charge carriers across its depletion region, the charge collection time of carriers outside of the depletion region diffusing to it, and the RC time constant of the photodiode and the circuit [Centronic Ltd. 1998, UDT Sensors Inc., Zimmermann 2000]. This response time is highly dependent on the applied bias voltage. By increasing the applied reverse bias, the depletion region of the diode increases thereby reducing the diffusion time²³ of the photodiode. The RC time constant also decreases because the capacitance of the photodiode, which arises from the junction capacitance of the depletion region, is inversely proportional to the width of the depletion region [Sze 1981]. Depending on the circuitry connected to the photodiode the RC time constant could very well dominate the response time of the system. As such it was necessary to characterise the photodiodes in terms of their C-V characteristics. This would also allow one to determine a suitable operating voltage for the photodiode depending on the application.

The C-V characteristics of the various junction diodes in the Mietec 0.7 μm CMOS process were simulated using PSpice and the models provided. Figure 2.15 shows the C-V plots for the p+/n-well (pshal1-6), n+/p-well (nshal1-6) and the n-well/p-substrate (ndeep1, ndeep2) junction diodes. The values shown are for an area of 10000 μm^2 (100 μm x 100 μm).

²³ The drift time also decreases due to the increase of drift velocity, v_d , with electric field, E , applied ($v_d = \mu E$ where μ is the mobility of carriers). However, once saturation is reached, the drift velocity does not increase further and drift time, t_d , increases with depletion width, w ($t_d = w/v_d$).

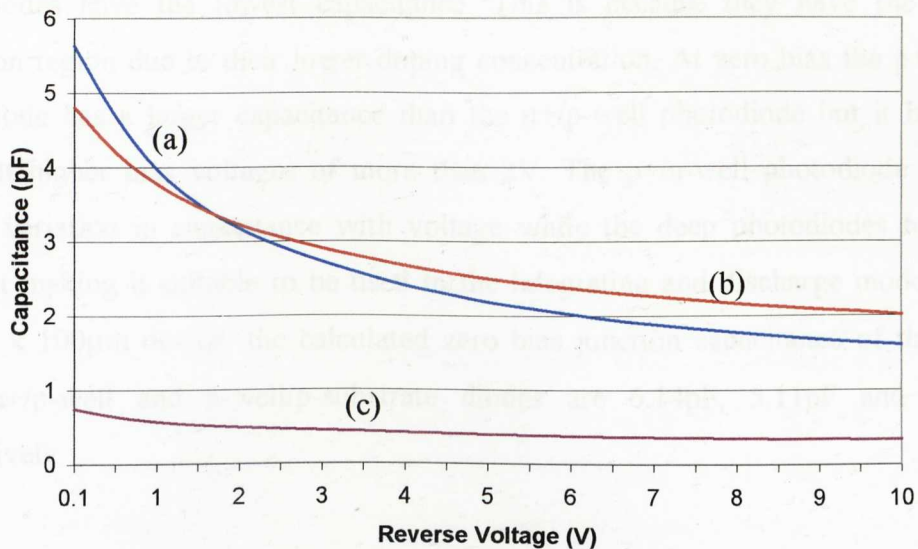


Figure 2.15 Simulated C-V plots for the (a) p+/n-well (pshall1-6), (b) n+/p-well (nshall1-6) and (c) n-well/p-substrate (ndeep1-2) junction diodes (all devices are of area $100\mu\text{m} \times 100\mu\text{m}$)

The software does not simulate periphery capacitance but it took into account scaling factors and grading coefficients provided in the models. Table 2.2 summarises the parameters from the models and the results of the simulation.

Photodiode type	Process datasheet (0V)		Simulations (PSpice)	
	C_j ($\text{pF}/\mu\text{m}^2$)	$C_{j\text{sw}}$ ($\text{pF}/\mu\text{m}$)	C_j at 0V ($\text{pF}/\mu\text{m}^2$)	C_j at 2V ($\text{pF}/\mu\text{m}^2$)
p+/n-well (pshall1-6)	6.0×10^{-4}	3.6×10^{-4}	5.6×10^{-4}	3.2×10^{-4}
n+/p-well (nshall1-6)	5.0×10^{-4}	2.8×10^{-4}	4.8×10^{-4}	3.2×10^{-4}
n-well/p-substrate (ndeep1, ndeep2)	7.89×10^{-5}	7.33×10^{-4}	7.4×10^{-5}	4.9×10^{-5}

Table 2.2 Junction capacitance of diodes based on model parameters and simulations

Based on these values, the periphery capacitance of the shallow photodiodes (nshall1-6, pshall1-6) only starts to dominate the total capacitance value for areas $< 1\mu\text{m} \times 1\mu\text{m}$. But with the deep photodiode (ndeep1, ndeep2) the periphery capacitance remains the dominant capacitance for areas up to $10\mu\text{m} \times 10\mu\text{m}$. The deep

photodiodes have the lowest capacitance. This is because they have the largest depletion region due to their lower doping concentration. At zero bias the p+/n-well photodiode has a larger capacitance than the n+/p-well photodiode but it becomes lower at higher bias voltages of more than 2V. The p+/n-well photodiode has the largest variation in capacitance with voltage while the deep photodiodes have the smallest making it suitable to be used in the integrating and discharge mode. For a $100\mu\text{m} \times 100\mu\text{m}$ device, the calculated zero bias junction capacitance of the p+/n-well, n+/p-well and n-well/p-substrate diodes are 6.14pF, 5.11pF and 1.08pF respectively.

It should be noted that the process parameters can vary from run to run and with external conditions such as temperature. This makes the simulations an estimate at best. Mietec provide a set of models to account for the variation in process parameters such as threshold voltage, gate oxide thickness and gate lengths drawn. The models provided are TYP for nominal process conditions, FAST for fast devices (to estimate worst-case power dissipation) and SLOW for slow devices (to estimate worst-case delay). Simulations were mainly performed and shown for TYP but illustrated in Figure 2.16 is the effect of process variations on the C-V characteristics of an n+/p-well junction diode.

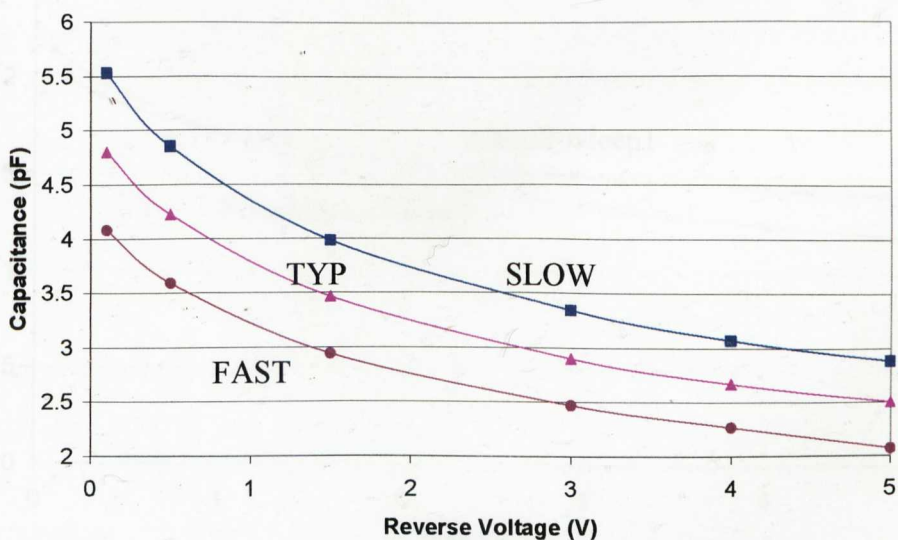


Figure 2.16 Simulations of the C-V characteristics of a $100\mu\text{m} \times 100\mu\text{m}$ n+/p-well diode

2.3.2.1 Experimental setup for C-V measurements

In order to verify the values given by the simulations, the capacitance was measured using a Boonton Electronics Capacitance Meter (Model 72B) capable of measuring capacitance down to a resolution of 0.01pF. In order to eliminate the measurement of any parasitic capacitance (cable, package and bondpads), a differential measurement was obtained. The capacitance meter allows a direct difference measurement to be made at its two terminals. So by connecting one photodiode of a particular size to one terminal and another photodiode of a different size to the other terminal, a reading corresponding to the capacitance of the difference in size of the two photodiodes is obtained. This of course assumes the stray cable capacitance in both connections are similar. A reverse bias voltage bias was applied to bias the photodiodes via the back of the capacitance meter.

2.3.2.2 Results and discussions for C-V measurements

Figure 2.17 shows the differential capacitance measurement between ndeep2 and ndeep1. The result is hence equivalent to the effective capacitance of a $30000\mu\text{m}^2$ (or 0.03mm^2) deep photodiode with n+ across. The C-V response obtained has the characteristic inverse shape of a C-V curve we would expect from a junction diode and it agrees satisfactorily with simulations.

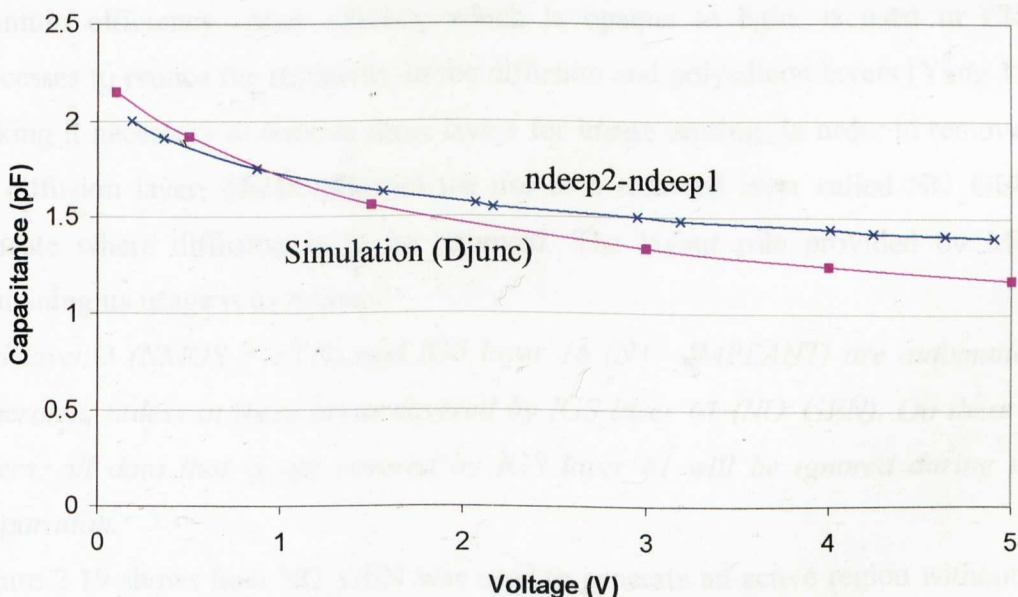


Figure 2.17 Differential C-V measurements of deep photodiodes with n+ across (effective area of $30000\mu\text{m}^2 \approx 170\mu\text{m} \times 170\mu\text{m}$) compared with the simulation model

However, when the deep photodiodes without n+ (deep1-6) were tested, an unusual C-V response was obtained just as we saw with the I-V plots for these devices. The response observed was essentially and uncharacteristically flat, as shown in Figure 2.18. In general the capacitance for these devices were lower than that obtained with the deep photodiodes with n+ across (ndeepl, ndeepl2).

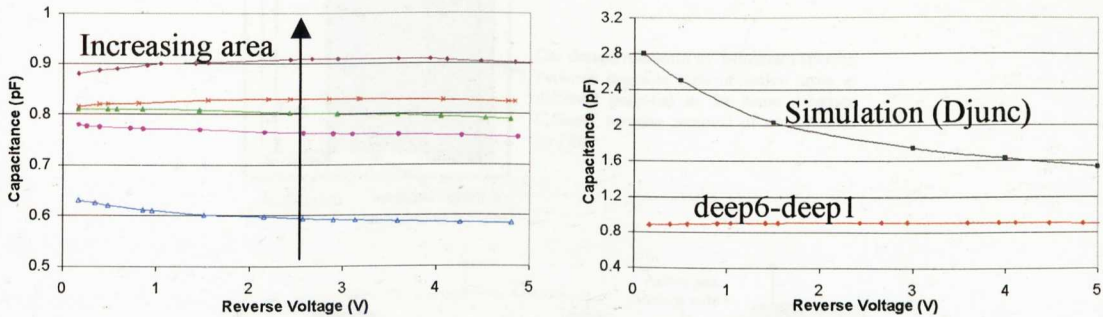


Figure 2.18 Differential C-V measurements (left) of deep photodiodes with n+ removed (deep1-6) and compared to simulations (right) for an effective area of $39100\mu\text{m}^2$ ($\approx 200\mu\text{m} \times 200\mu\text{m}$)

The reasons for requesting the removal of the n+ layer was to maximize the light entering the substrate without being strongly absorbed at the surface and to reduce recombination of photogenerated carriers in this region and hence improve the overall quantum efficiency. Also, silicide, which is opaque to light, is used in CMOS processes to reduce the resistivity in the diffusion and polysilicon layers [Yang 1996] making it necessary to remove these layers for image sensing. In order to remove the n+ diffusion layer, Mietec allowed the use of a reserved layer called NO_GEN to indicate where diffusion is to be removed. The layout rule provided by Mietec containing its usage is as follows:

IGS layer 3 (NMOS_FIELD) and IGS layer 16 (N+ _IMPLANT) are automatically generated, unless in these areas covered by IGS layer 61 (NO_GEN). On these IGS layers, all data that is not covered by IGS layer 61 will be ignored during mask preparation.

Figure 2.19 shows how NO_GEN was used to generate an active region without any diffusion layer. Note the use of NO_GEN as shown gave a design rule warning because this was a reserved layer that wasn't recognised by the design rule checker

used. However when checked via IMEC's Dracula design rule checker the layer was recognised and the diffusion layer was removed at that junction.

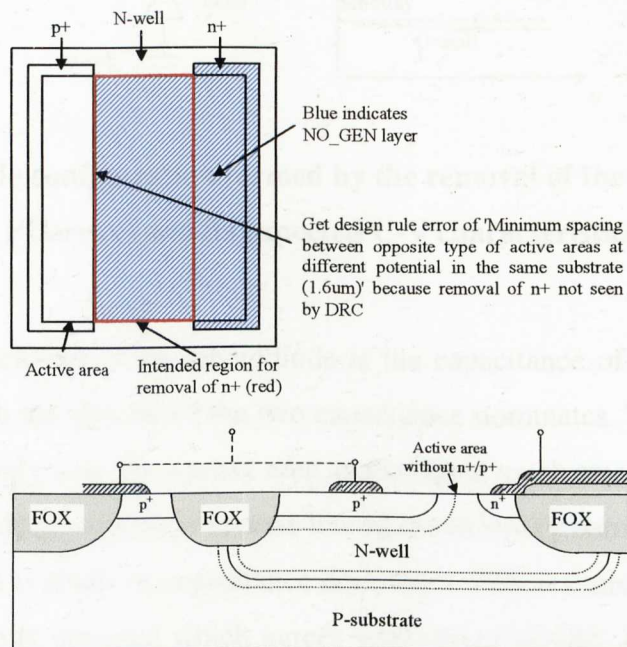


Figure 2.19 Use of NO_GEN to remove diffusion layers in an active area

It was later established that the Mietec 0.7 μm CMOS process used was a polycide process and not a silicide²⁴ process. But the abnormal results obtained with the deep devices with n+ removed necessitated closer inspection of the device. It seems that the use of NO_GEN layer over the explicitly drawn n+ region had removed the n+ diffusion layer here as well, despite the description of the layout rule. As a consequence a Schottky barrier diode was formed between the n-well and the cathode (K) contacts. Hence two diodes in series were formed as illustrated in Figure 2.20, which explains the C-V characteristics as well as the forward bias current obtained.

²⁴ Silicidization is the process of depositing metal (typically titanium or cobalt) on to the silicon in order to lower the resistance of the polysilicon interconnect or the source-drain contact. In a polycide process only the polysilicon is silicided. In a silicide process both polysilicon gate and source-drain regions are silicided. If this silicide process is a self-aligned process, it is usually termed silicide.

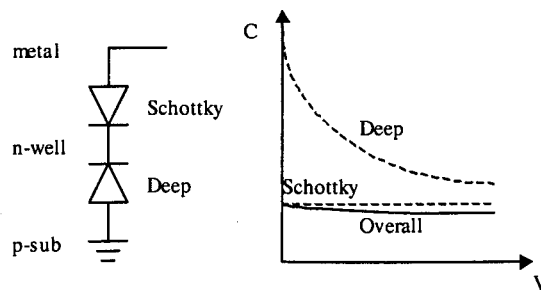


Figure 2.20 Diode configuration formed by the removal of the n+ layer (deep1-6 and PDarray) and its expected C-V characteristics

The overall capacitance of the photodiode is the capacitance of these two diodes in series and as such the smaller of the two capacitance dominates. The Schottky barrier diode is formed only over the contact area and is hence much smaller than the junction diode. When the deep junction is reverse biased the Schottky barrier is forward biased. However, since it is small its capacitance dominates. Hence a smaller and more linear capacitance value is obtained which agrees with that observed. As the deep junction capacitance drops with increasing bias it will come into play in determining the overall capacitance. The effect of the Schottky diode on the photoresponse in the reverse bias operation of the photodiode is not observable which is reasonable because under these conditions the Schottky diode is forward biased. This will be shown later. Paradoxically, the lower capacitance obtained with these deep devices is a useful by-product for high-speed applications and increased charge conversion efficiency. Also the linear C-V curve obtained will give rise to a linear discharge curve when the photodiode is used in an integrating mode, such as in an integrating active pixel sensor.

It should be pointed out that the diode capacitances measured do not scale linearly with the effective area and is not expected to because a differential measurement will have a lower periphery component than a direct measurement. Consider a difference measurement between a $200\mu\text{m} \times 200\mu\text{m}$ and a $100\mu\text{m} \times 100\mu\text{m}$ diode, the effective area measured will be $30000\mu\text{m}^2$ and the effective periphery measured will be the difference in periphery, which is $400\mu\text{m}$. However, for an area of $30000\mu\text{m}^2$ the periphery expected would be close to $700\mu\text{m}$. A more accurate estimate of the capacitance is obtained by considering the measurement for the largest difference in

area, for example the measurement for deep6 – deep1. Figures 2.21, 2.22 and 2.23 show the measured C-V characteristics of the shallow n⁺, the shallow p⁺ and both the combined photodiodes respectively.

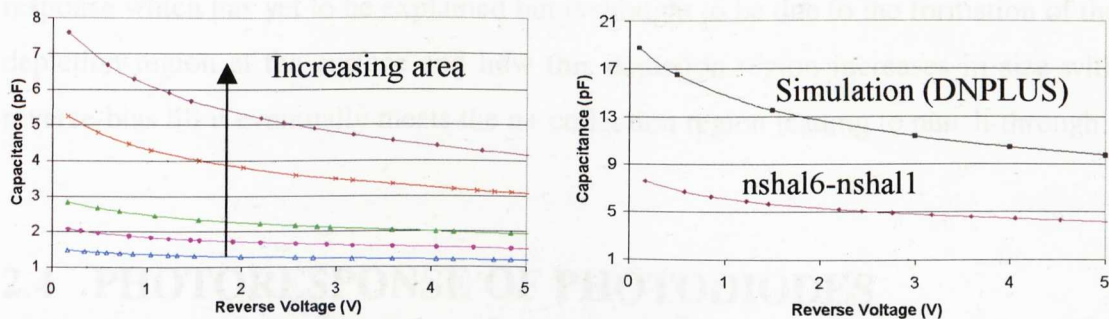


Figure 2.21 Differential C-V measurements of shallow n⁺ photodiodes (nshal1-6) (left) and compared to simulations for an effective area of 39100 μm² (right)

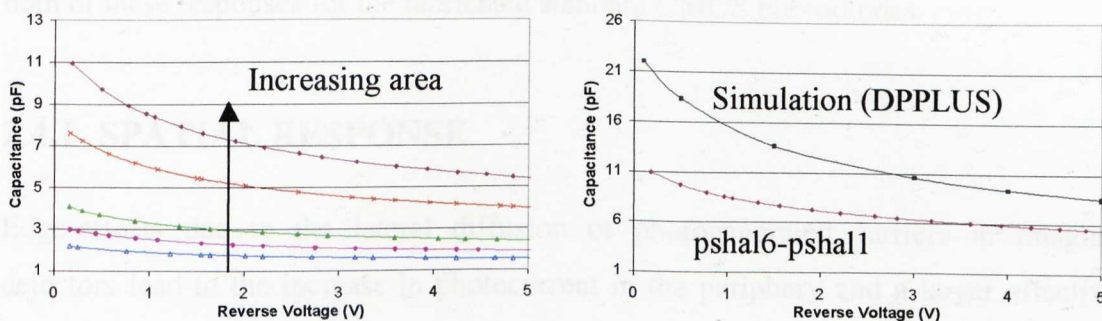


Figure 2.22 Differential C-V measurements of shallow p⁺ photodiodes (pshal1-6) (left) and compared to simulations for an effective area of 39100 μm² (right)

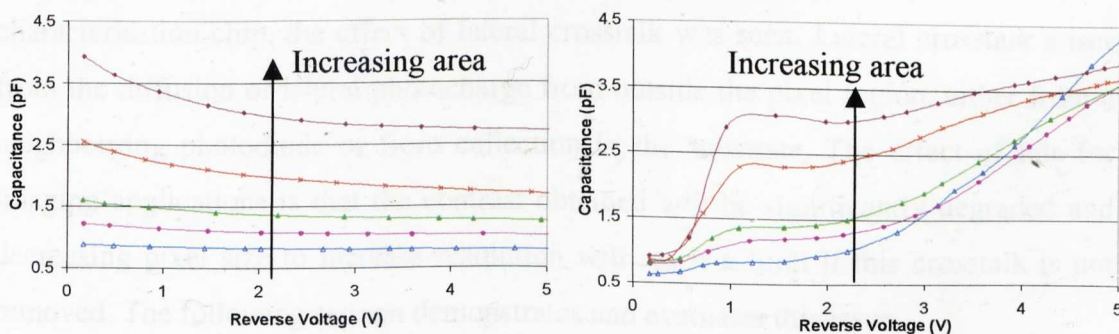


Figure 2.23 Differential C-V measurements of combined shallow n⁺/deep photodiodes (ncomb1-6) (left) and combined shallow p⁺/deep photodiodes (pcomb1-6) (right)

The p+/n-well (pshall1-6) photodiode has a larger capacitance than the n+/p-well (nshall1-6) photodiode and both have a larger capacitance per unit area than the deep photodiodes, which agrees with what simulations suggest. The measured capacitance of the combined devices are also shown. The pcomb devices exhibit a strange response which has yet to be explained but is thought to be due to the formation of the depletion region at the surface and how this depletion region increases in size with reverse-bias till it eventually meets the n+ collection region leading to punch-through.

2.4 PHOTORESPONSE OF PHOTODIODES

The photoresponse of a photodiode can be evaluated in terms of its spatial and spectral sensitivity. The following sections detail experimental work carried out in determining both of these responses for the fabricated standard CMOS photodiodes.

2.4.1 SPATIAL RESPONSE

Edge-effects due to the lateral diffusion of photogenerated carriers in imaging detectors lead to the increase in photocurrent in the periphery and a larger effective charge collection area than the actual geometry of the photodiode [Holloway 1983]. This effect is expected to be more pronounced in small photodiodes which has a larger perimeter-to-area ratio. A series of photodiodes of varying sizes were included in the second characterisation chip (chip1bfinal) in order to evaluate this. Also with the first characterisation chip, the effect of lateral crosstalk was seen. Lateral crosstalk arises from the diffusion of lateral photocharge from outside the pixel region, either from a neighbouring photodiode or from collection in the substrate. The effect of this for imaging applications is that the contrast obtained will be significantly degraded and decreasing pixel size to increase resolution will reach a limit if this crosstalk is not removed. The following section demonstrates and evaluates this issue.

2.4.1.1 Experimental setup of spatial response test

In order to determine the spatial photoresponse of the photodiodes, a laser beam (667nm) was focused to a spot of approximately $5\mu\text{m}$ and scanned across the area of the photodiode. This was done by placing the sample on a scanning stage and adjusting the height and position of the stage such that the laser is focused on the sample. The stage is controlled by a PC to move in 2 dimensions to cover the scanning area with the focused laser remaining fixed. The scanning stage is capable of moving in step sizes as small as $1\mu\text{m}$ but mainly a step size of $5\mu\text{m}$ was used, as too small a step would lead to excessively long scan times. Also it would be unnecessary to make the step size too small when the spot size is limited to $5\mu\text{m}$ anyway. The PRO8000 laser diode controller from Profile, Germany, was used to control the laser output power over a range of 2 decades (0.4mW - 10mW). It also maintains the temperature of the laser at a specified level for stability and a room temperature of 25°C was chosen. Figure 2.24 shows the setup of the scanning system. The power of the laser diode was set at 0.4mW with no neutral density filter (NDF) in the optical path but after going through the optics the power incident on the chip was approximately $82\mu\text{W}$. The reflected beam was imaged on a reference photodiode to obtain an image of the scan and to determine if the setup was in focus. As with the dark current measurements, the Keithley Source-Measure Unit was used to apply a bias voltage and take the current measurement. The scan was performed with a reverse bias voltage of 2V applied to the test photodiode. However, unlike the dark current measurements, the Keithley was controlled through the IEEE 488.2 GPIB (General Purpose Interface Bus) serial interface [Keithley Instruments Inc. 2001] to allow automatic collection of data²⁵. However, it was necessary to wait for a period of at least 3s after setting the bias conditions before taking a reading from the Keithley as the bus remains busy for this period. Consequently a time between readings of 5s was used throughout. A test board allowed each photodiode on the test chip to be tested in

²⁵ When controlling the Keithley through the GPIB, the autoranging feature of the Keithley would fail at low measured currents and an arbitrary value of $+0.001\text{mA}$ is obtained. When that occurred, it was necessary to change the measurement range. The easiest way to do this was to check the reading obtained and if the reading was out of range, the program would switch to an appropriate measurement range.

turn by connecting the appropriate jumper. The schematic and PCB of the test board for the scanning experiment is included in Appendix A2.2.

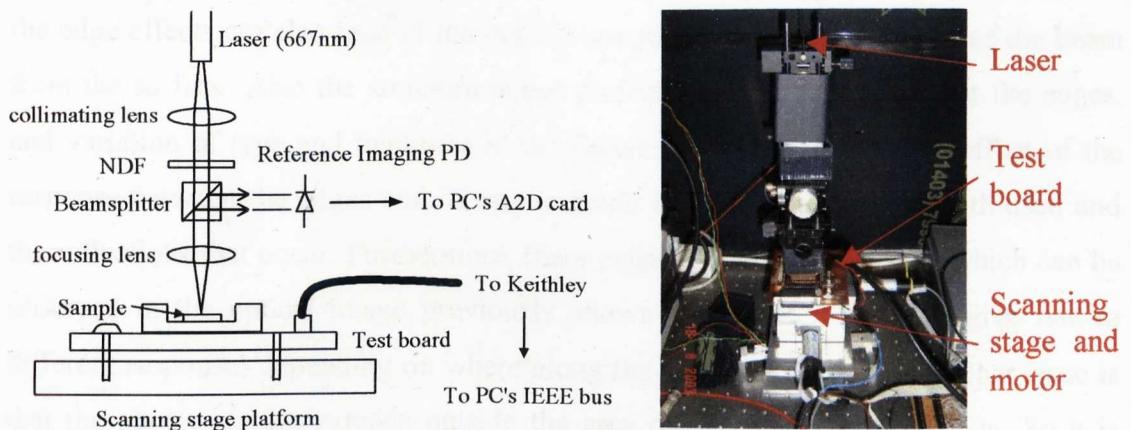


Figure 2.24 Setup of scanning system for the characterisation of fabricated photodiodes for spatial response measurements

2.4.1.2 Results and discussion of spatial sensitivity measurements

Figure 2.25 shows the optical image obtained from a scan of the deep4 photodiode and Figure 2.26 shows the spatial results in both x and y direction obtained from this scan. The increase of photocurrent at the edges is due to the side-wall of the photodiode providing a larger volume depletion region (see Figure 2.1) and hence collection region. Also the large number of defects at the edges, particularly at the surface and at the field-oxide/well-junction interface, could contribute to its presence. During the oxidation process in chip fabrication, stresses are generated that slightly lift the protective nitride at its edges, creating a tapered oxide called a bird's beak. The LOCOS or bird's beak region is the transition between the field oxide and the thin oxide that covers the n^+ implant and is under elevated mechanical stress. The presence of this can lead to a larger leakage current. In a recent paper by Hornsey and Renshaw [Lee 2003 (Part II)], it was observed that the edge-effect in CMOS photodiodes is significantly affected by surface recombination and mobility degradation along the Si-SiO₂ interface.

It can be seen that edge effects are more significant in the x-direction than the y-direction. It is not yet clear why this is so. However it is felt that shadowing effects in the optics made the observed edge effects more pronounced than they actually are as the edge effects are also seen in the optical image which is the reflection of the beam from the surface. Also the structure is not perfectly planar, particularly at the edges, and variation of type and thickness in the layers will mean the relative effect of the response between the edges and its centre could depend on the wavelength used and the reflections that occur. Furthermore, there exists a grain in the wafer, which can be observed in the optical image previously shown (Figure 2.5) and will give rise to different responses depending on where along the grain the spot lies. Another issue is that the photoresponse extends outside the area of the exposed photodiode. So it is possible that at the edges, diffraction effects and multiple reflections in the passivation layers are occurring, and not discounting the possibility that the light spot is diffused significantly more than expected by the imaging optics. Stray and scattered light was also an issue in the experiment.

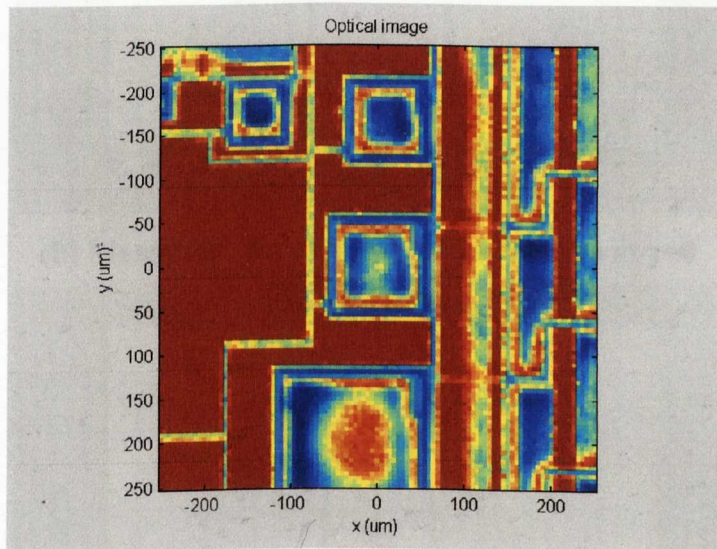
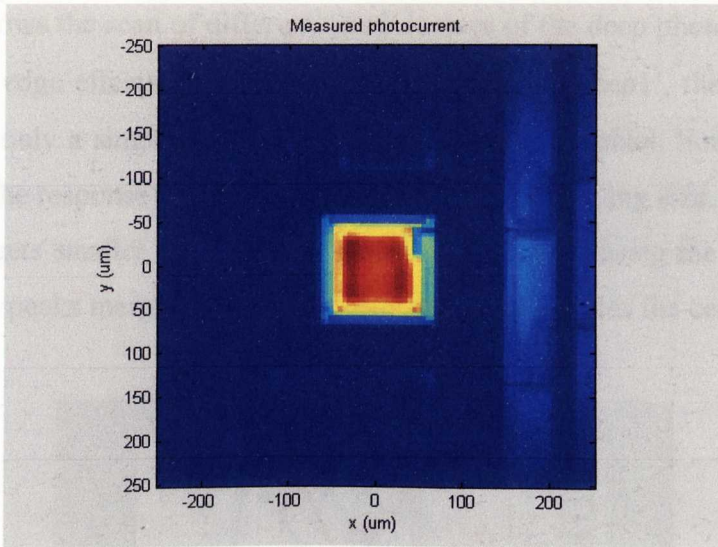


Figure 2.25 Optical image from reflected beam



(a) Measured photocurrent image map

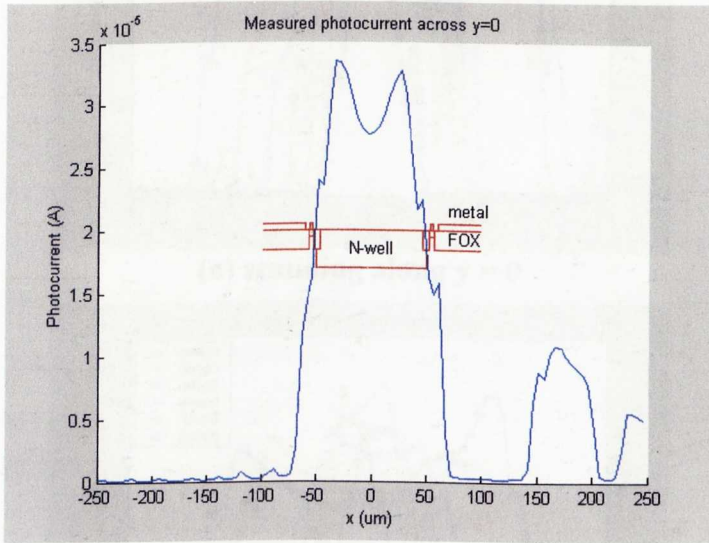
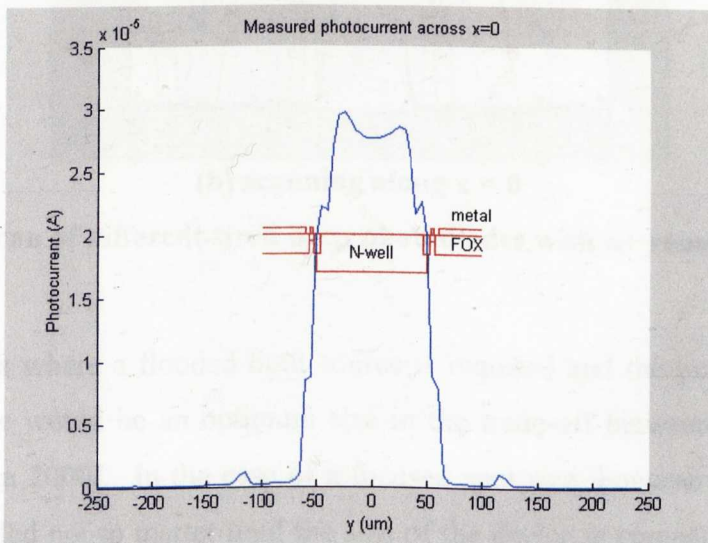
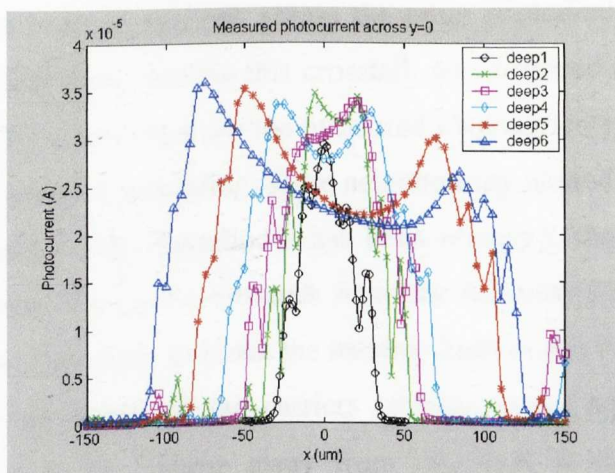
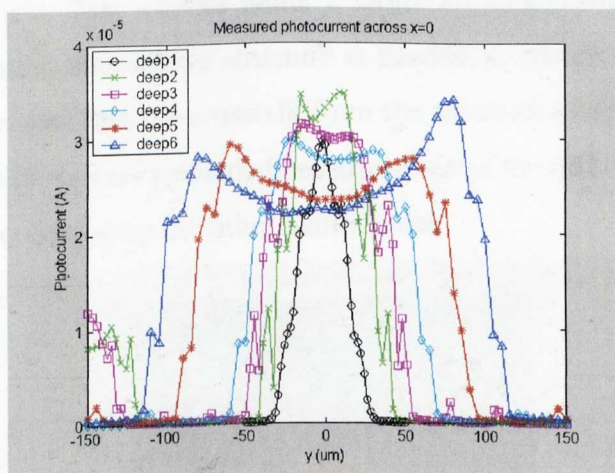
(b) Measured photocurrent in x direction at $y=0$ (c) Measured photocurrent in y direction at $x=0$ Figure 2.26 Scan of $100\mu\text{m} \times 100\mu\text{m}$ deep photodiode with n+ removed (deep4)

Figure 2.27 shows the scan of different sized devices of the deep photodiodes with n^+ removed. The edge effects can again be seen, except for 'deep1', the $30\mu\text{m} \times 30\mu\text{m}$ device, where only a single peak exists at the centre of the pixel. For the other sizes (i.e. deep2-6) the response at the centre decreases with increasing size. It seems that as the pixel size gets smaller the peaks get closer together increasing the response at the centre until the peaks merge and further decrease in size reduces the central response.



(a) scanning along $y = 0$



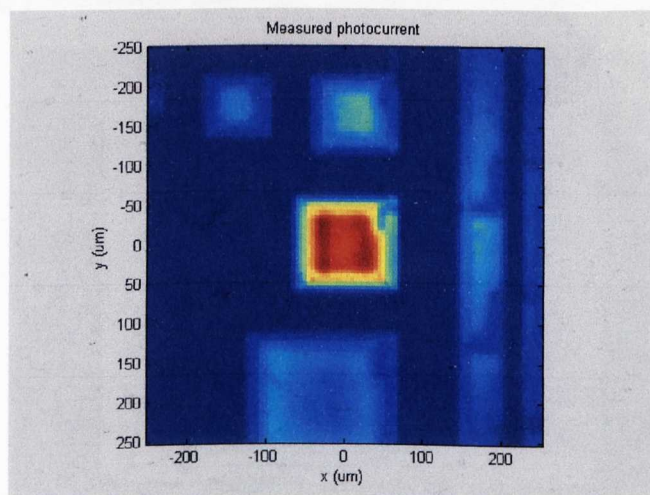
(b) scanning along $x = 0$

Figure 2.27 Scan of different sized deep photodiodes with n^+ removed (deep1-6)

For applications where a flooded light source is required and the pixel size dictated resolution, there would be an optimum size in the trade-off between sensitivity and resolution [Chen 2000]. In the case of a focused spot size, however, the size of the device is expected not to matter until the size of the device is comparable to the spot size. However, because of the edge effects and non-uniform response, it may be prudent to make the detector size somewhat larger than the spot size.

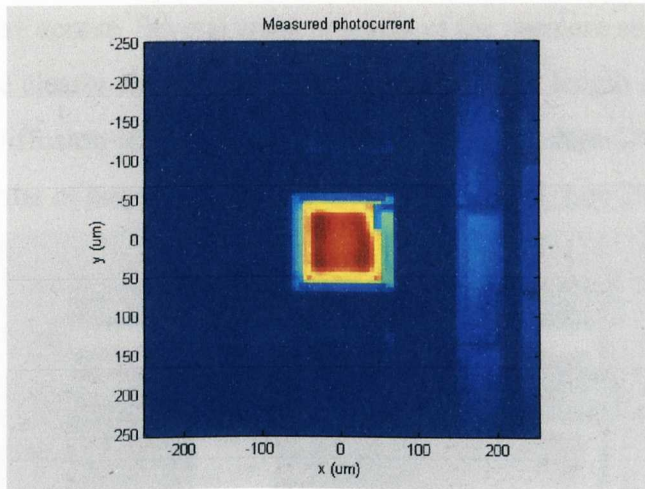
Crosstalk

In scanning a laser beam across the chip, two sources of crosstalk could be seen: diffusion of photogenerated carriers from neighbouring photodiodes and diffusion of carriers from the exposed substrate. The crosstalk from neighbouring photodiodes could be removed by grounding the neighbouring photodiodes. This is illustrated in Figure 2.28 (a) and (b). Figure 2.28 (a) shows the photocurrent detected by the device in the centre as the beam is scanned across the other photodiodes which were left floating. By grounding these devices this crosstalk was removed as shown in Figure 2.28 (b). Figure 2.28 (c) and (d) show the measured photocurrents along $y=0$ and $x=0$ of the scan before and after grounding of the neighbouring photodiodes. Although the crosstalk from neighbouring photodiodes has been removed, the crosstalk from the substrate still remains. To remove crosstalk from the substrate a metal light shield is placed around each photodiode to block the incident light in this region. However due to the large diffusion length of the carriers relative to the scale of the devices, substrate current as far as $300\mu\text{m}$ away from the pixel is still detected by the photodiode under test. This implies either a larger area light shield is required or a guard ring or parasitic photodiode structure is needed to absorb the leakage current. However, it can be seen that the crosstalk from the substrate is also reduced when the neighbouring photodiodes are grounded because some of the diffused substrate current is now drawn and collected by the other photodiodes.

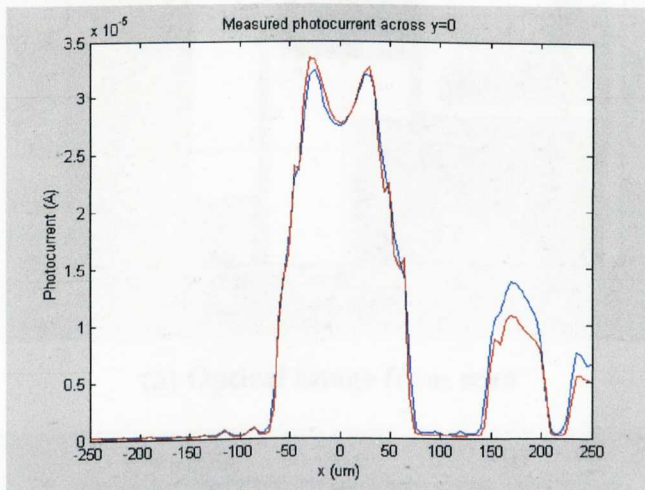


(a) Neighbouring photodiodes floating

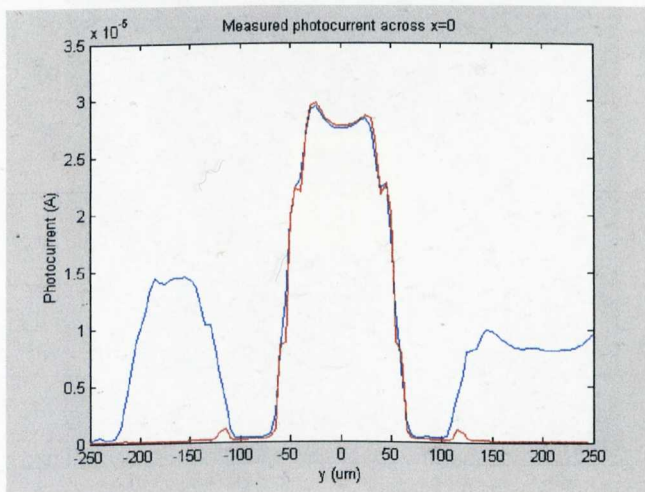
Figure 2.28 Scan of 'deep' photodiode with crosstalk present and crosstalk removed



(b) Neighbouring photodiodes grounded



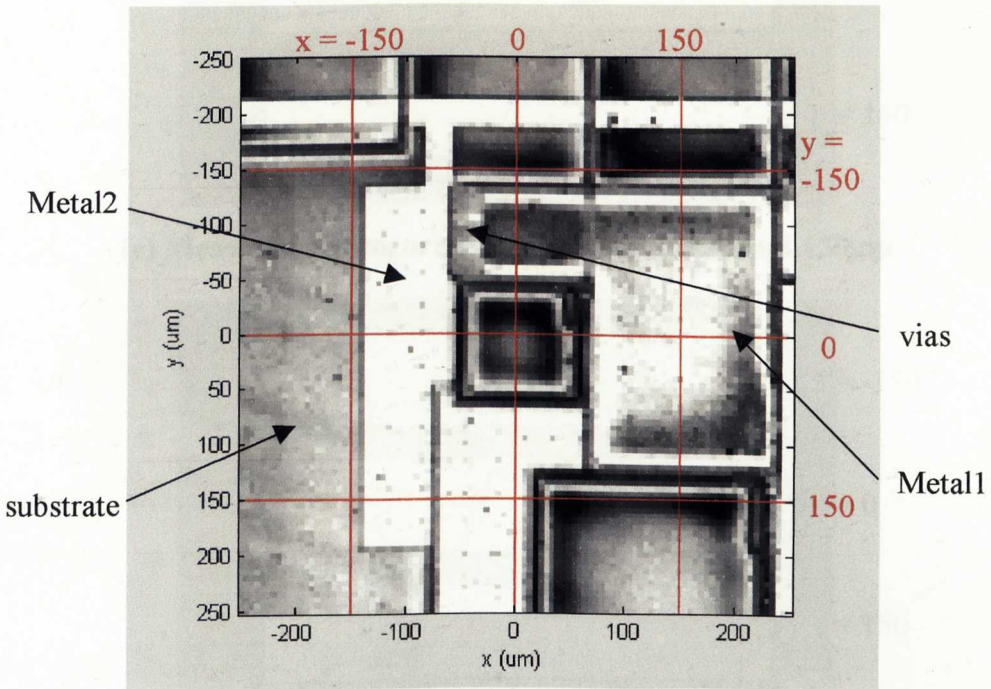
(c) Scan along $y = 0$ before and after grounding of neighbouring photodiodes



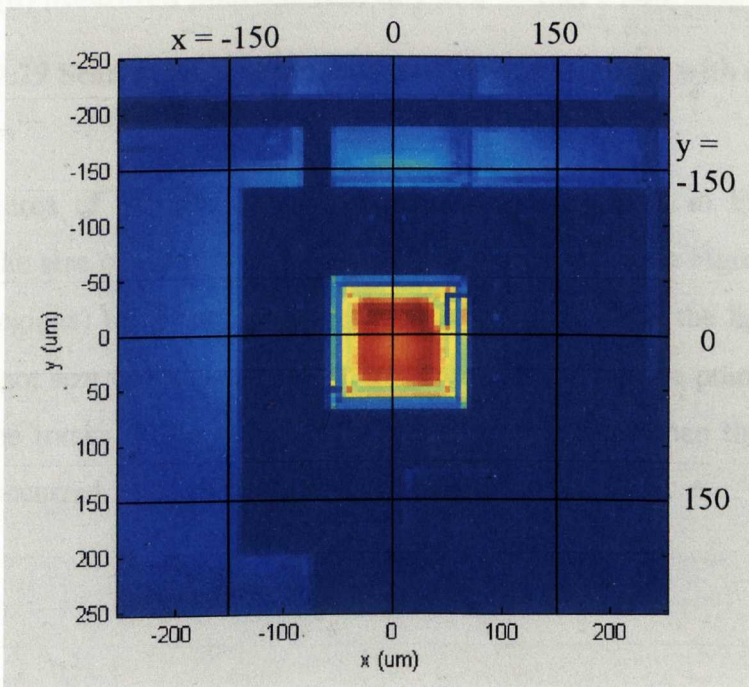
(d) Scan along $x = 0$ before and after grounding of neighbouring photodiodes

Figure 2.28 Scan of 'deep4' photodiode with crosstalk present and crosstalk removed

Figure 2.29 shows the response obtained from the scan of the $100\mu\text{m} \times 100\mu\text{m}$ deep photodiode with n^+ across. Several cross sections of the scan are shown. The crosstalk from the substrate clearly shows the extent of the diffusion length of the carriers. The minority carrier diffusion lengths of epitaxial silicon in modern CMOS processes are typically in the order of hundreds of micrometers [El Gamal, Lee 2003 (Part I)].

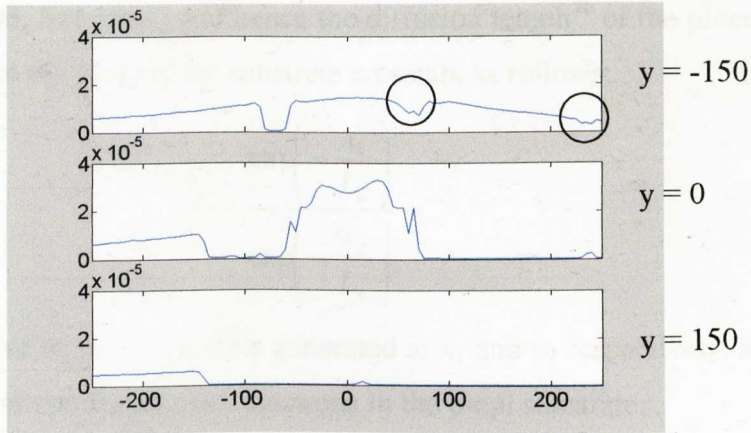


(a) Optical image from scan

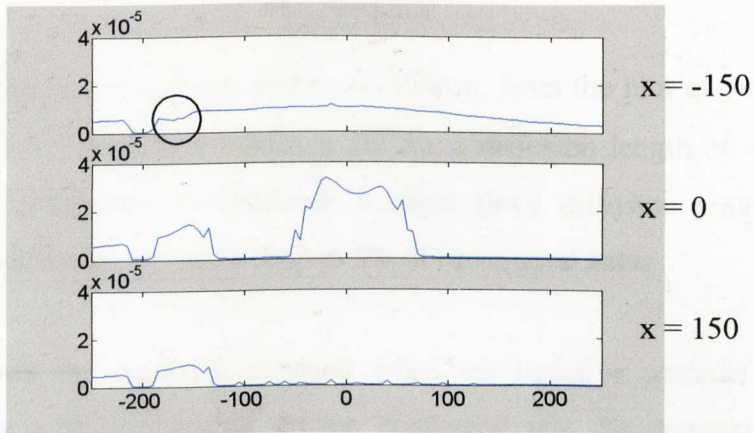


(b) Measured photocurrent image map

Figure 2.29 Scan of the $100\mu\text{m} \times 100\mu\text{m}$ deep photodiode with n^+ across



(c) Measured photocurrent in x at $y = -150, 0$ and $150\mu\text{m}$



(d) Measured photocurrent in y at $x = -150, 0$ and $150\mu\text{m}$

Figure 2.29 Scan of the $100\mu\text{m} \times 100\mu\text{m}$ deep photodiode with n+ across

The contact area of the photodiode can clearly be discerned in Figure 2.29 (b) establishing the size of the spot to be less than $6\mu\text{m}$. Note that in Figure 2.29 (c) and (d) (circled regions) the presence of metal tracks did not block the light completely because the spot size was larger than the track size ($2\mu\text{m}$) at these points. This causes the size of the tracks in the optical image to appear broader than they are. Spatial filtering has occurred.

The diffusion of minority carriers follows an exponential decay with length [Shcherback 2003, Sze 1981] and hence the diffusion length²⁶ of the process used can be estimated from the plots of the substrate crosstalk as follows:

$$\frac{I_1}{I_2} = \frac{\exp\left(-\frac{x_1}{L_n}\right)}{\exp\left(-\frac{x_2}{L_n}\right)} \quad (2.1)$$

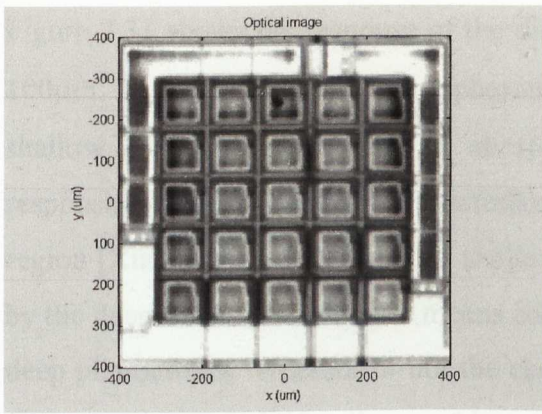
where I_1 and I_2 are the photocurrents generated at x_1 and x_2 respectively, and L_n is the diffusion length of minority carrier electrons in the p-epi substrate.

Therefore, the diffusion length, $L_n = \frac{x_2 - x_1}{\ln\left(\frac{I_1}{I_2}\right)}$ (2.2)

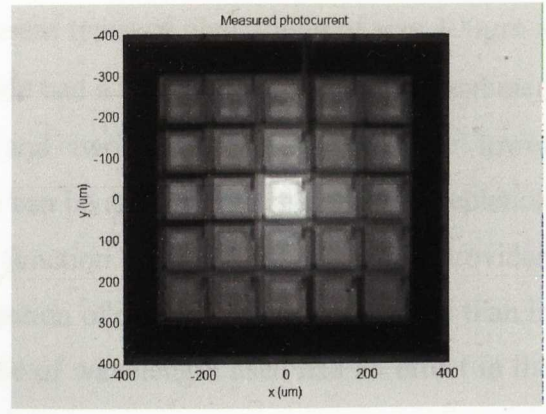
Choosing two points, $x_2 = -200\mu\text{m}$ and $x_1 = -150\mu\text{m}$, from the plot of $y=0\mu\text{m}$ of the scan ($I_2 = 1.0111 \times 10^{-5}\text{A}$, $I_1 = 8.0628 \times 10^{-6}\text{A}$), a diffusion length of $\sim 220\mu\text{m}$ for electrons in the p-substrate is obtained. It takes three diffusion lengths for the concentration of diffused carriers to drop to 5% of its original value.

Figure 2.30 shows the crosstalk obtained when the beam is scanned across the photodiode array with the central device connected and the remaining devices floating. The lateral crosstalk is significant with adjacent pixels reaching more than 50% of the central pixel value. Furthermore, the response of the pixel under test is lower than that obtained in the isolated pixel case. This is possibly because part of the photoresponse is due to the diffusion of carriers outside the depletion region and this is now being collected by the p-n junctions (depletion regions) of neighbouring photodiodes. The diffusion process, though contributing to the photocurrent, acts as a spatial filter performing spatial averaging of the image. It is also interesting to note, from the 3D image obtained (Figure 2.30 (e)), that the edge-effects are most prominent at the corners of a photodiode pixel where the electric field stresses are higher [Shcherback 2002].

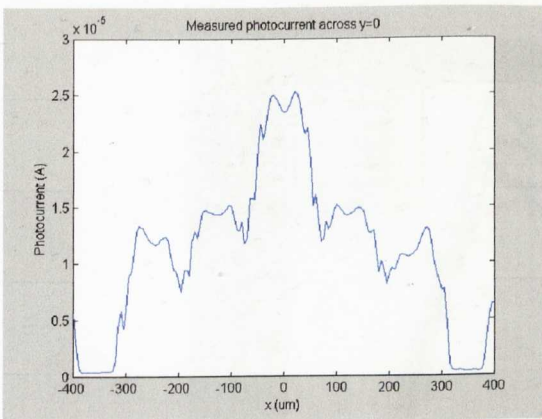
²⁶ The distance over which concentration of injected free charge carriers injected into semiconductor falls to $1/e$ (37%) of its original value.



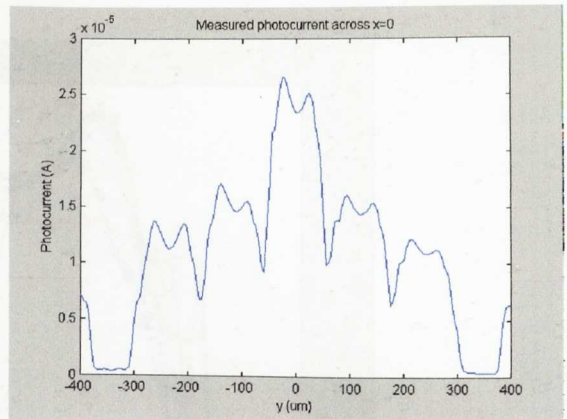
(a) Optical image



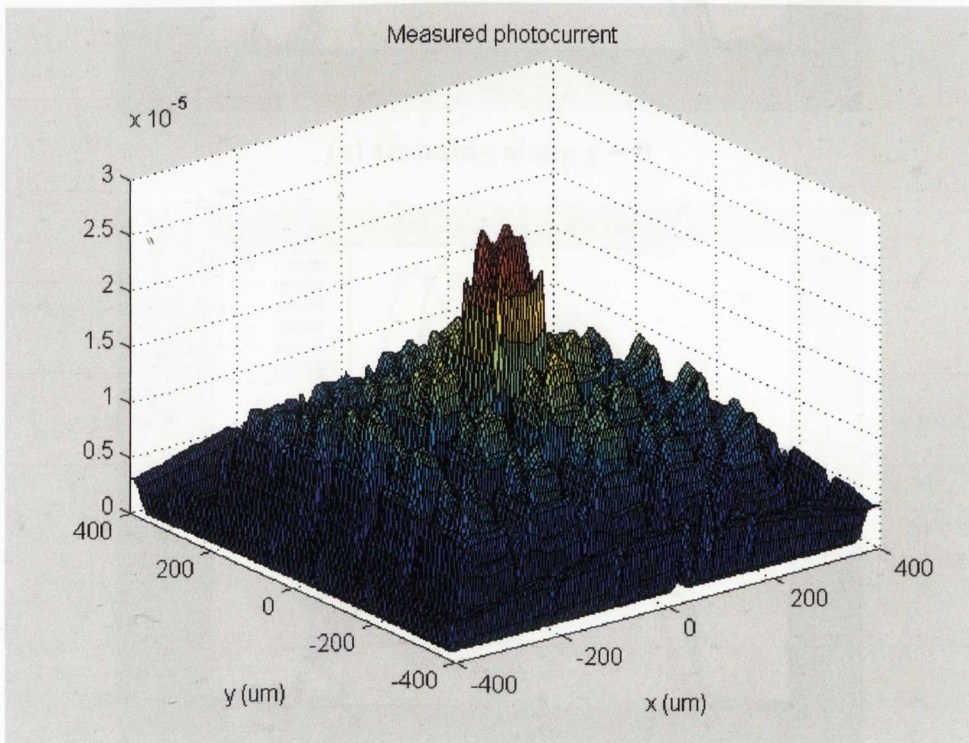
(b) Image of photocurrent measured



(c) Measured photocurrent along x



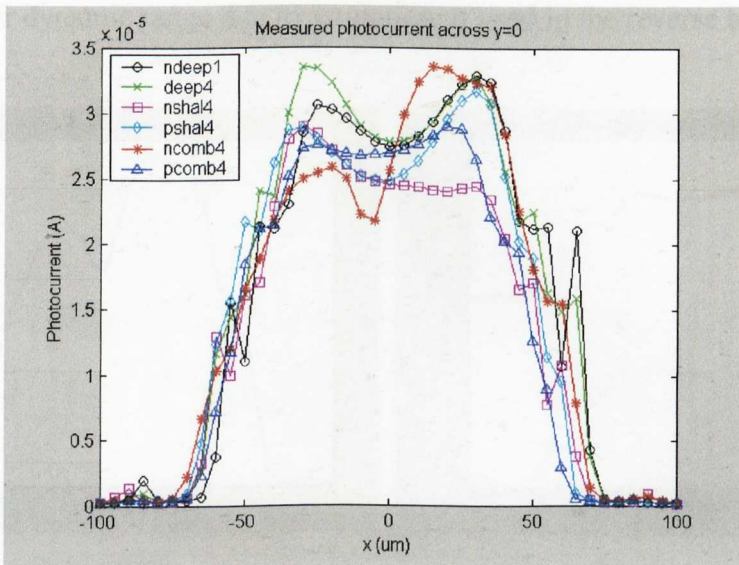
(d) Measured photocurrent along y



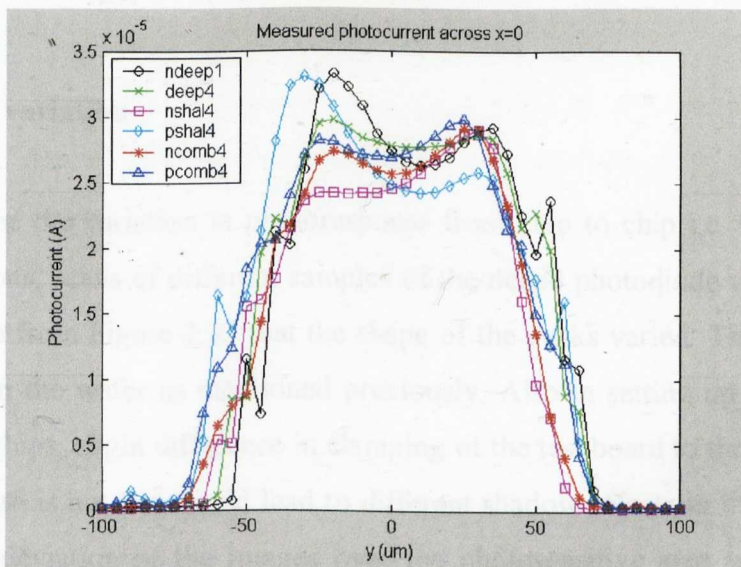
(e) 3D image of crosstalk obtained

Figure 2.30 Crosstalk between neighbouring pixels of the photodiode array (PDarray)

Figure 2.31 shows the response of the different types of photodiode of size $100\mu\text{m} \times 100\mu\text{m}$. In general the shallow n^+ photodiode had a lower response and the combined shallow n^+ /deep device has an abrupt and two distinct responses. The lower responsivity of the shallow n^+ photodiode can be attributed to its narrower depletion region [Xiangliang 2002]. Also its shallow junction depths and the isolation provided by the deep field oxide trenches means collection of diffusion carriers is poorer than in deep photodiodes. Whether or not the choice of wavelength used had an effect in the response obtained will be discussed later.



(a) scanning along $y = 0$



(b) scanning along $x = 0$

Figure 2.31 Photoresponse of the different $100\mu\text{m} \times 100\mu\text{m}$ sized photodiodes

Figure 2.32 shows the measured photoresponse in the combined devices. In the case of the combined shallow n+/deep (ncomb) device, the response is due to the fact that this photodiode consists of two distinct photodiodes next to each other (see photodiode cross-section in Figure 2.7 (c)) with the deep device having a higher responsivity than the shallow n+ photodiode as mentioned. It is also interesting to observe the breakdown effect in the combined shallow p+/deep (pcomb) devices as mentioned previously in Section 2.2.3. With the pcomb device, a large background current is obtained with reverse bias voltage but with no amplification of the photocurrent. This device will suffer from poor signal-to-noise ratio due to large shot noise and poor dynamic range due to saturation, if used in the reverse bias mode.

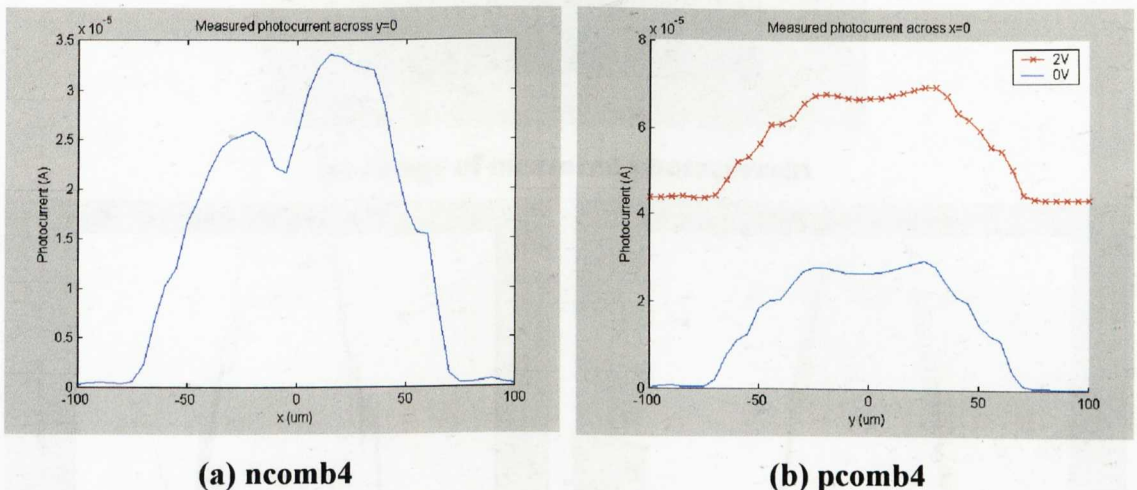
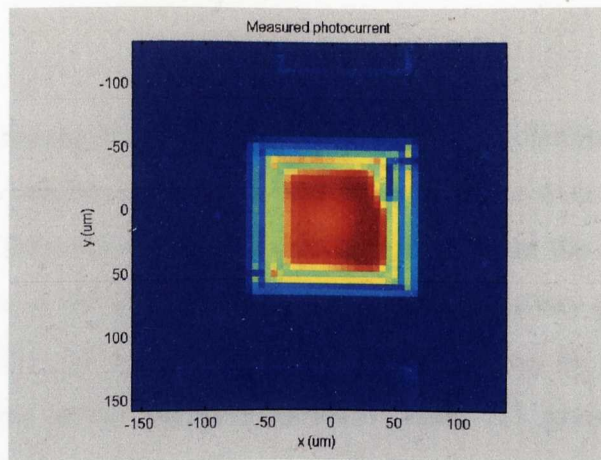


Figure 2.32 Photoresponse of the $100\mu\text{m} \times 100\mu\text{m}$ combined photodiodes

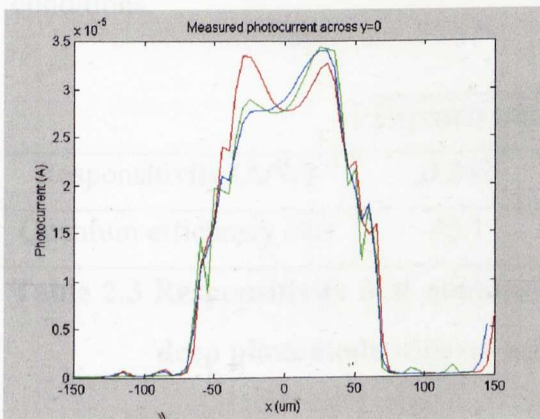
Chip to chip variation

In order to see the variation in photoresponse from chip to chip i.e. with process or wafer variations, scans of different samples of the deep4 photodiode were performed. It can be seen from Figure 2.33 that the shape of the peaks varied. This could be due to the grain in the wafer as mentioned previously. Also in setting up the experiment for different chips, slight difference in clamping of the test board to the scanning stage i.e. if the board is not flat, could lead to different shadow effects in the scan. Overall the standard deviation of the images over the photosensitive area is still less than 1.7%. From the image of the measured photocurrent of Figure 2.33 (a), the absorption of light through the metal holes can also be seen. Also with the neighbouring

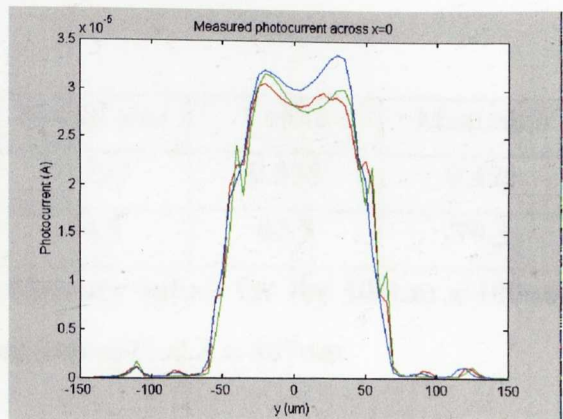
photodiodes grounded, contribution from the edges of these devices is still visible though significantly reduced as seen in Figure 2.33 (b) and (c). Perfect removal of crosstalk just by grounding is not possible. Diffusion follows a statistical process and a very small proportion still diffuses to the test photodiode.



(a) Image of measured photocurrent



(b) Measured photocurrent along $y=0$



(c) Measured photocurrent along $x=0$

Figure 2.33 Scan of the deep $100\mu\text{m} \times 100\mu\text{m}$ photodiode with n+ removed (deep4) on various chips

Responsivity

From the spatial sensitivity experimental setup, responsivity values can be obtained. However, there are several means to determine this value due to the spatial nature of the response. The responsivity can be obtained by taking a mean over the area of the photodiode. But it is difficult to determine exactly the size of the photodiode as the light entering the substrate outside the defined photodiode area can also be picked up.

Thus far the photodiode size has been defined by the size of the n-well as this corresponds to the location of the p-n junction or depletion region. But the photodiode exposed area is slightly larger than this because of the necessary substrate contacts around the periphery of the device – see Figure 2.6. For our design, the exposed region of each photodiode is about 20 μm larger than the stated n-well width in both directions.

Four different conditions are defined for the possible calculation of the responsivity. The responsivity can be calculated based on an average over a defined area. Here two will be used: the total area exposed to light i.e. before the boundary of the light shield and the area of the drawn n-well. Or the responsivity can be obtained from specific points on the photoresponse scan. Intuitively, either the centre of the pixel or the maximum value across the scan is used. Table 2.3 gives the responsivities obtained for a 100 μm x 100 μm deep photodiode with n+ across for these different conditions.

	Exposed area	N-well area	Centre	Maximum
Responsivity (A/W)	0.242	0.293	0.335	0.426
Quantum efficiency (%)	45.1	54.5	62.3	79.3

Table 2.3 Responsivity and quantum efficiency values for the 100 μm x 100 μm deep photodiode with n+ across (ndeep1) at $\lambda = 667\text{nm}$

It is felt that the average value obtained using the n-well area gives a fair and good estimate of the responsivity and will be used from now on. In the case of photodiodes without an n-well, the equivalent active area drawn defines the area. Quantum efficiency values are also shown and are obtained from the measured responsivity values using equation (1.9).

Table 2.4 shows the responsivities obtained for the photodiodes tested. The average responsivity of the deep devices is 0.298 A/W. The responsivity of the shallow n+ device is slightly lower as it has a smaller depletion region to collect the photogenerated charges. The shallow p+ device has a comparable responsivity to the deep devices because of the presence of the parasitic n-well/p-substrate junction. A

more detailed analysis of the responsivity will be given later when the spectral response of the devices is observed.

Photodiode	Responsivity (A/W) at 667nm
ndeep1 (100 μ m x 100 μ m)	0.293
ndeep2 (200 μ m x 200 μ m)	0.307
deep1 (30 μ m x 30 μ m)	0.259
deep2 (60 μ m x 60 μ m)	0.287
deep3 (80 μ m x 80 μ m)	0.281
deep4 (100 μ m x 100 μ m)	0.307
deep5 (160 μ m x 160 μ m)	0.306
deep6 (200 μ m x 200 μ m)	0.320
nshal4 (100 μ m x 100 μ m)	0.258
pshal4 (100 μ m x 100 μ m)	0.303
ncomb4 (100 μ m x 100 μ m)	0.282
pcomb4 (100 μ m x 100 μ m)	0.273

Table 2.4 Responsivities of photodiodes tested at 667nm (based on n-well area)

2.4.2 I-V CHARACTERISTICS

By obtaining the I-V characteristics of a photodiode (see Section 1.5.4) under varied illumination levels, its linearity and suitable operating range in terms of light intensity and bias voltage can be determined.

2.4.2.1 Experimental setup

The same scanning system employed in the measurement of the spatial photoresponse (see Section 2.4.1.1) was used to obtain the I-V characteristics of the photodiodes in light. However, the incident power on the sample is now adjusted by varying the output power of the laser diode and by placing various NDFs in the optical path. NDFs with optical densities, D of 3.2, 0.8 and 0.4 where the NDF transmittance, $T =$

10^{-D} , was used to give an incident power range of 5 decades (60nW to 2.7mW). The focused laser beam is imaged onto the centre of the pixel under test and then an I-V sweep is performed using the Keithley with a step size of 0.1V up to a reverse bias voltage of 5V and a forward bias of 1V (5V for the deep and pcomb devices with Schottky diodes). The results obtained from the characterisation will be shown in the following section.

2.4.2.2 Results and discussion of I-V characterisation in light

Figure 2.34 shows the I-V characteristics obtained for various illumination levels (82 μ W to 2.7mW) of the deep 100 μ m x 100 μ m photodiode with n+ removed (ndeepl). As expected, the larger the incident power, the larger the photocurrent. The response appears linear and this will be investigated further. What is interesting to note is that there is some response to light in the forward bias because the Schottky diode (as a consequence of no n+ under the contacts) can act as a photodiode as well. However, its response is weak, partly due to the size of the device and partly due to the fact that it is completely covered in metal with no interdigitated structure required for proper photodiode operation of Schottky photodiodes. The main mechanism for light detection here is probably due to the diffusion of carriers from outside the contact area.

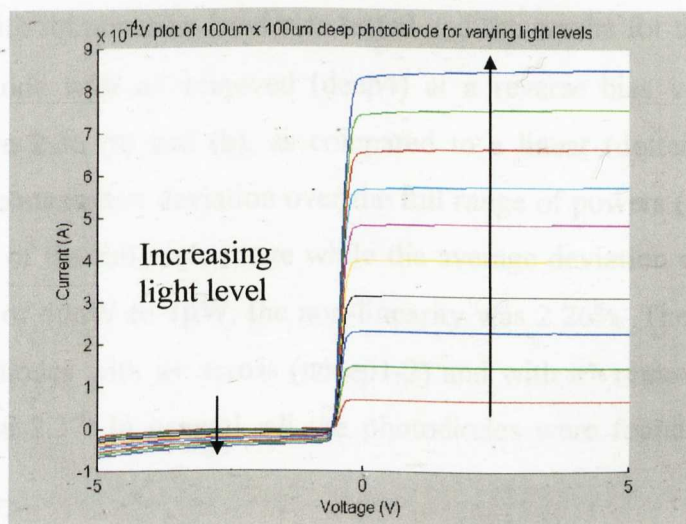


Figure 2.34 I-V characteristics for deep 100 μ m x 100 μ m photodiode with n+ removed (ndeepl) for increasing light level

Figure 2.35 shows the comparison between the I-V characteristics of the deep $100\mu\text{m} \times 100\mu\text{m}$ photodiode with n^+ across (ndeep1) and that without (deep4) for incident light powers of $41.6\mu\text{W}$ and $494\mu\text{W}$. The deep photodiode with n^+ across is slightly more responsive (8.6%). It is also observed that the deep photodiode with n^+ removed has significantly less photoresponse in 'quadrant 4' of the I-V plot and has a fixed open-circuit voltage in the presence of illumination. This is due to the reverse bias action of the Schottky diode in this region. This limits the operating range of this device as a photodiode.

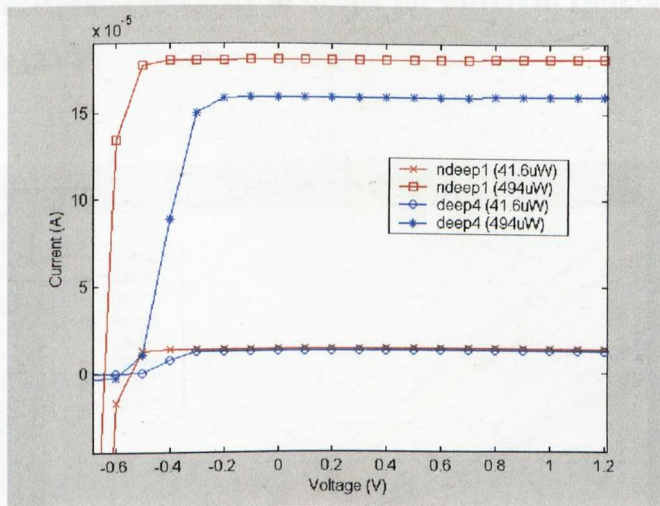
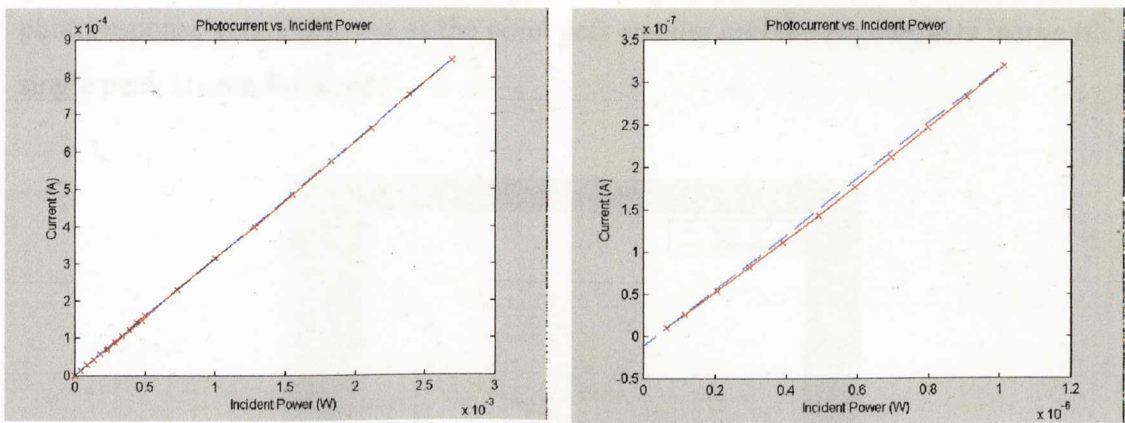


Figure 2.35 I-V characteristics of the deep $100\mu\text{m} \times 100\mu\text{m}$ photodiode with n^+ across (ndeep1) and without (deep4)

The linearity with illumination level was tested and the results for the deep $100\mu\text{m} \times 100\mu\text{m}$ photodiode with n^+ removed (deep4) at a reverse bias voltage of 2V are shown in Figure 2.36 (a) and (b), as compared to a linear (dotted) line. The non-linearity²⁷ or the maximum deviation over the full range of powers (60nW to 2.7mW) tested is 0.73% of the full scale range while the average deviation was about 0.13%. Over the range of 60nW to $1\mu\text{W}$, the non-linearity was 2.26%. The linearity of both the deep photodiodes with n^+ across (ndeep1-2) and with n^+ removed (deep1-6) are shown in Figure 2.37. In general, all the photodiodes were found to be of similar linearity.

²⁷ Non-linearity is defined as the maximum deviation of the transmitter output from the reference line (terminal or best-fit straight line) and is reported as a percentage of the unit's full-scale range.



(a) Incident power of 60nW to 2.7mW

(b) Incident power of 60nW to 1 μ W

Figure 2.36 Photoresponse linearity of deep4 for different ranges of incident light levels (2V)

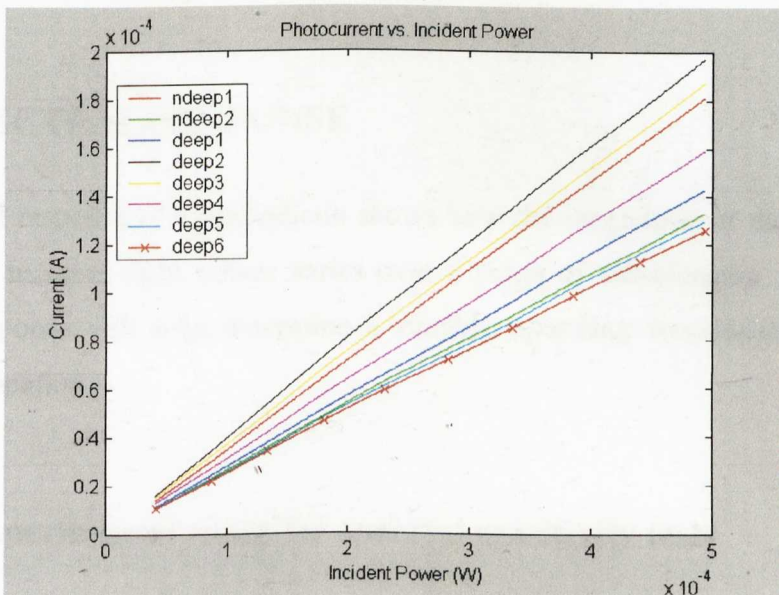


Figure 2.37 Photoresponse linearity of the deep photodiodes with n+ across (ndeep1-2) and without (deep1-6) for incident power of 40 μ W to 460 μ W (2V)

When the photocurrent measured is plotted against area as shown in Figure 2.38, a non-linear response was obtained. Also the response seems to be larger for smaller sized devices with the exception of the 30 μ m x 30 μ m device. This test was repeated on a separate chip with similar results. In actual fact, this is the same response that was seen with the spatial photoresponse of the deep devices in Figure 2.27 because it is the photoresponse at the centre of the pixel that is being measured. The dip in the

photoresponse gets shallower as the pixel gets smaller and the edge effects merge till a single peak is seen for 'deep1'.

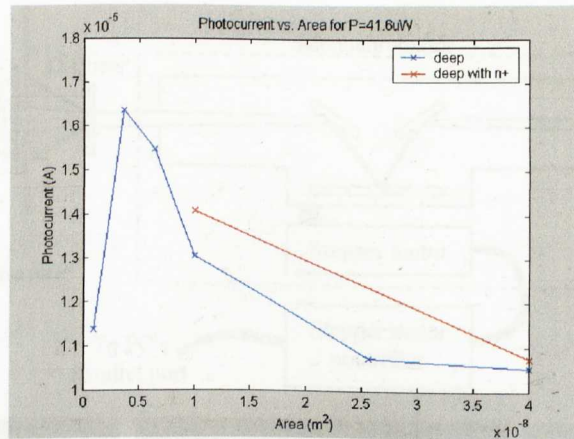


Figure 2.38 Variation of photocurrent with area for deep devices (2V)

2.4.3 SPECTRAL RESPONSE

The spectral response of a photodiode shows how the magnitude of the photocurrent for a given incident light power varies over a range of wavelengths. Obtaining the spectral response will help determine a suitable operating wavelength to use in a chosen application.

2.4.3.1 Experimental setup for spectral sensitivity tests

This section describes the setup and testing of the experimental apparatus for obtaining the spectral response of the full custom photodiodes, as shown in Figure 2.39. The first part of the setup involves providing a monochromatic or single wavelength output over a wide range of wavelengths. The H20 IR Jobin Yvon monochromator [Jobin Yvon (Horiba) New Jersey, USA] with a grating of 600 lines/mm and an output wavelength range of 400nm to 1100nm was used. The monochromator takes in white light from a 70W tungsten-halogen lamp through its entrance slit. Mirrors inside the monochromator direct the light to a diffraction grating, which divides the white light into its spectrum. Another set of mirrors direct the light to the exit slit where the spectrum is narrowed down to a near-

monochromatic light. The wavelengths exiting the monochromator are selected by rotating the grating which is controlled by the dial on the monochromator.

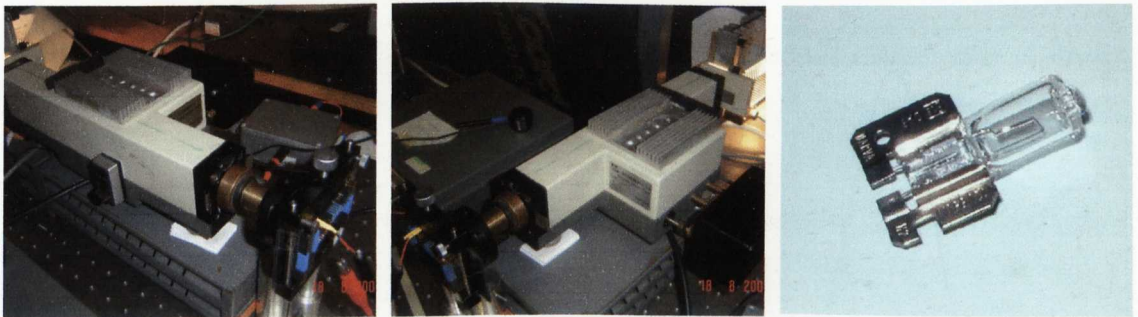
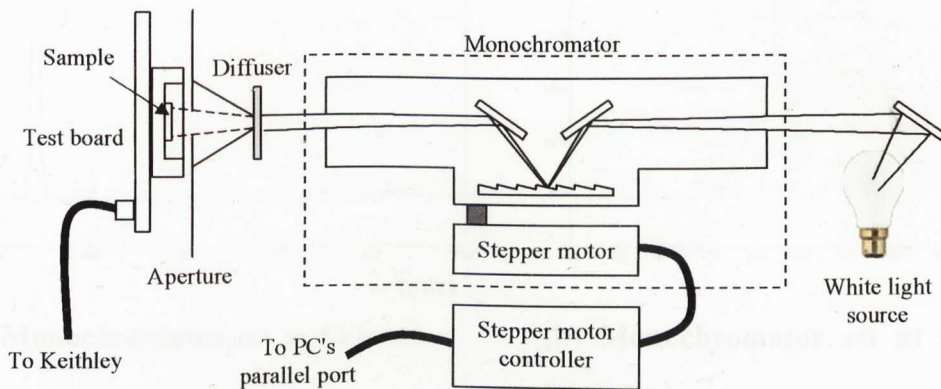
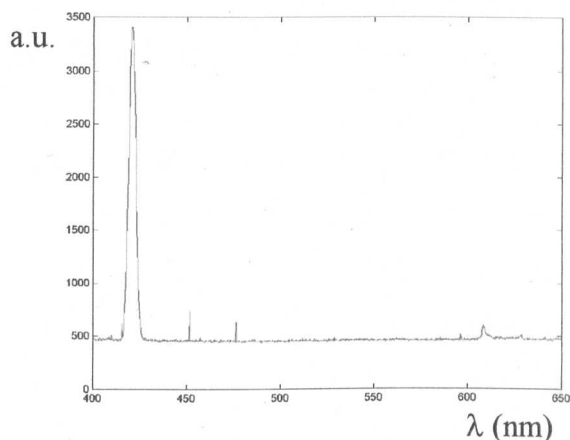
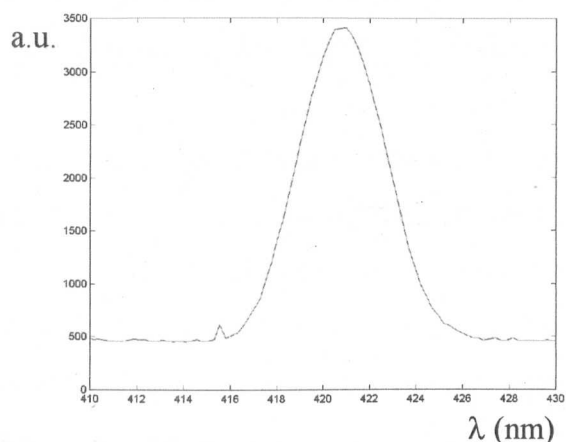
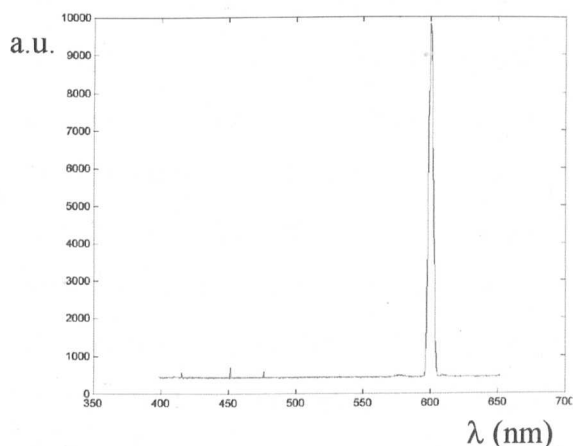


Figure 2.39 Setup to determine spectral response of test photodiodes

The resolution of the monochromator is specified as 0.5nm for a wavelength of 500nm and a diffraction grating of 1200 lines/mm. In order to observe and confirm the resolution of the monochromator for the diffraction grating used, its output was observed through a spectrometer. The grating used in the spectrometer allowed a range of 400nm to 700nm to be observed. Figure 2.40 shows the resolution observed for wavelengths of 420nm and 600nm. The resolution, specified as the full width at half maximum (FWHM), is approximately 5nm at both wavelengths. In addition, the output beam of the monochromator was diverging and non-uniform. The non-uniformity was partly due to the image of the grating appearing on the output beam and the position of this changes with wavelength. Hence a diffuser was used in order to produce a uniform beam of light over the area of the sample. The disadvantage of this is less light gets through.



(a) Monochromator set at 420nm

(b) Monochromator set at 420nm
(close up)

(c) Monochromator set at 600nm

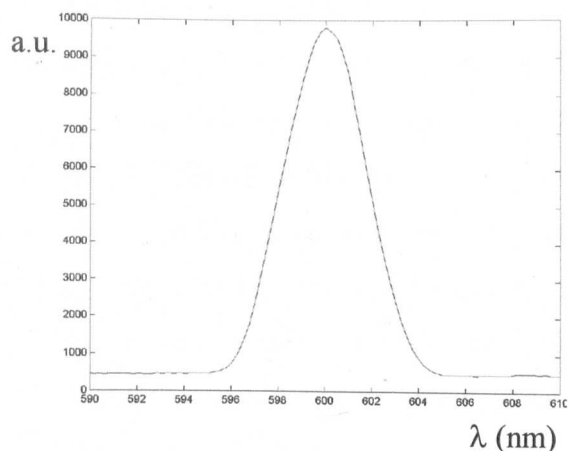
(d) Monochromator set at 600nm
(close up)

Figure 2.40 Spectral output of the monochromator viewed through a spectrometer

A stepper motor was used to automatically rotate the grating and step through the wavelengths. The UCN5804B BiMOS II Unipolar Stepper Motor Driver is used to convert CMOS/TTL logic inputs into a stepper motor drive format to drive the four-phase unipolar stepper motor attached to the monochromator grating turret [Chen 2002]. The format used was the two-phase drive format which has better torque performance and less susceptible to motor resonance. The driver accepts two signals from the PC's parallel port. One is to control the rotation sequence of the outputs and hence the direction of the motor i.e. whether the wavelength is increased or decreased and the second is to advance the sequence position of the outputs by one position with

every high-to-low transition. Six step pulses were needed to advance the wavelength by 1nm. Step sizes of 5nm were used in the measurement of the spectral response.

The second part of the setup is the measurement of the sample's photocurrent and the incident light power. The Keithley was used to obtain the photocurrent measurements. In theory, fluctuations in the source (wavelength and intensity) can be compensated for by splitting the light and simultaneously measuring (and cross calibrating) the photocurrent generated on the test photodiode and on a calibrated reference photodiode. However, the Keithley only allows for measurements on one channel and so this cannot be carried out. The power of the light incident on the sample was measured with a Newport Optical Power Meter (Model 835) that used an 818-SL detector type with an active area of 1cm^2 . Readings were double-checked with a second power meter, namely the Coherent LabMaster Ultima Power Meter which had a detector aperture of 7.9mm, a spectral range of 400-1064nm and a resolution of 1nW . Measuring the power of a non-uniform beam as illustrated in Figure 2.41 gives rise to an incorrectly higher responsivity value because the power measured is averaged across the beam but the power incident on the detector, which is smaller than the aperture, is higher. An aperture was used so that a more uniform area of illumination is obtained and this also allowed a more accurate determination of the illuminated area when measuring the incident light power with the power meter.

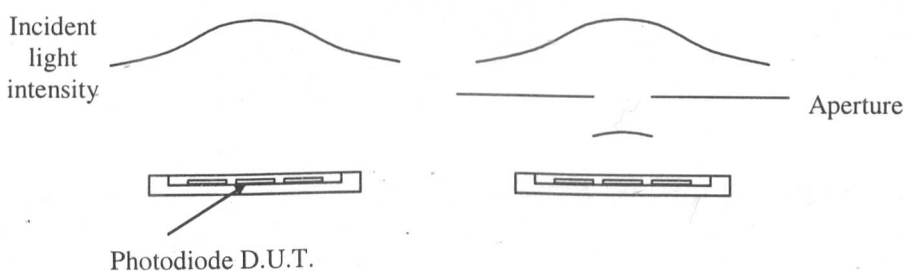


Figure 2.41 The use of an aperture to obtain more uniform power measurements

Initial tests were made with the Temic BPW34 photodiode in order to use its datasheet values for comparison. However, this showed a significantly larger response than that specified in its datasheet. This turned out to be due to the source not being accurately imaged on the entrance slit. This resulted in a more non-uniform beam at the output. However the response obtained after correcting this was still high as shown in Figure

2.42 (a). Also shown are measurements taken at different times (correctly imaged on the slit) which shows that the temporal variations of power or wavelength could not account for this discrepancy. A calibrated photodiode from Hamamatsu (S6058 4-quadrant Si PIN Photodiode) was then used for further tests but this still showed similar results of the response being higher than expected. This is shown in Figure 2.42 (b). It was thought that the uniformity of the beam was still affecting the reading and a smaller aperture and a second diffuser was used. This improved the reading as shown. However, some non-uniformity probably still existed. As such it was decided to scale the spectral response curves obtained with the value of the responsivity obtained with the spatial response experiment as the light source was focused in that experiment and the incident power could be determined more accurately. The scaled result is also shown in Figure 2.42 (b). Now the curve is slightly lower but this is marginal and can be explained by scattered light, which gets measured by the power meter, and losses in the photocurrent measurement.

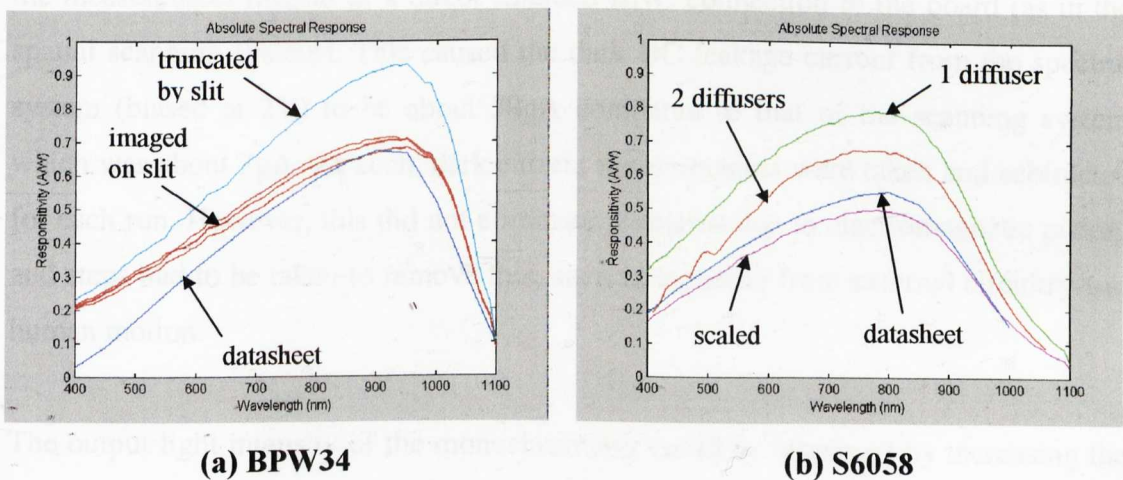


Figure 2.42 Comparing measured and documented spectral response of reference photodiodes

Table 2.6 shows the responsivity values obtained from the scanning system compared to that quoted on the datasheet for the reference photodiodes. The incident power in this test was measured at $84\mu\text{W}$. It was concluded that the best and most accurate means of determining the spectral response of any test photodiodes was to scale the response obtained with the responsivity value obtained from the previous spatial scanning experiment. This removed the need for the second diffuser, which

meant more light could be directed at the sample and the signal-to-noise ratio of the system is improved.

	Reference photodiode	
	BPW34	S6058
Measured photocurrent	35.4 μA	37.5 μA
Measured responsivity	0.421 A/W	0.446 A/W
Responsivity from datasheet	0.425 A/W	0.488 A/W
Percentage error	0.94 %	8.6 %

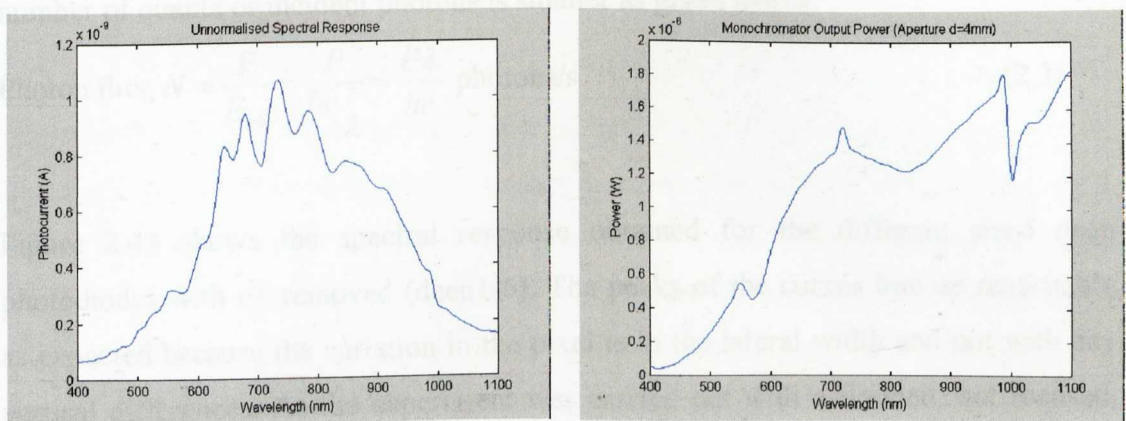
Table 2.5 Comparing measured responsivity values of the reference photodiodes using the scanning system with their quoted datasheet values at 667nm

Due to size restrictions in the setup, unshielded flying leads had to be used to make the measurement instead of a direct shielded BNC connection to the board (as in the spatial scanning system). This caused the dark DC leakage current from the spectral system (biased at 2V) to be about 50pA compared to that of the scanning system which was about 3pA. As such, dark current measurements were taken and subtracted for each run. However, this did not eliminate problems due to electromagnetic pickup and steps had to be taken to remove this, such as isolating from external circuitry and human motion.

The output light intensity of the monochromator could be increased by increasing the entrance and exit slit widths of the monochromator but this has the effect of reducing the resolution of the wavelength selection. Using a higher power (wattage) light bulb may not necessarily increase the input light intensity because a higher power light bulb may just have a larger filament area whose image is truncated by the input slit anyway.

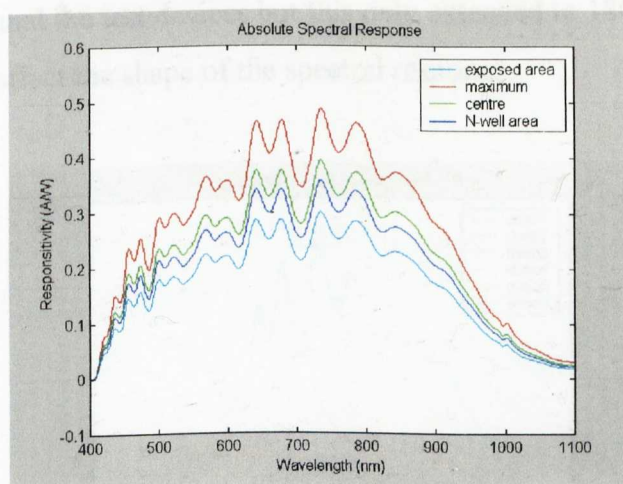
2.4.3.2 Results and discussion of spectral sensitivity tests

The spectral response of the deep $100\mu\text{m} \times 100\mu\text{m}$ photodiode with n+ removed (deep4) from 400nm to 1100nm is shown in Figure 2.43 (c). The response curves when scaled to the different responsivity values are shown. Figures 2.43 (a) and (b) show the measured photocurrent and incident light power respectively. The readings go below 100pA for wavelengths of 450nm or less for the $200\mu\text{m} \times 200\mu\text{m}$ devices and 470nm or less for the $100\mu\text{m} \times 100\mu\text{m}$ devices making it susceptible to the noise level at this range. However this corresponds to where the response curves quickly drop off anyway. An internal spline function in Matlab was used to interpolate the data and get a smoother, more detailed waveform.



(a) Measured photocurrent

(b) Measured light power



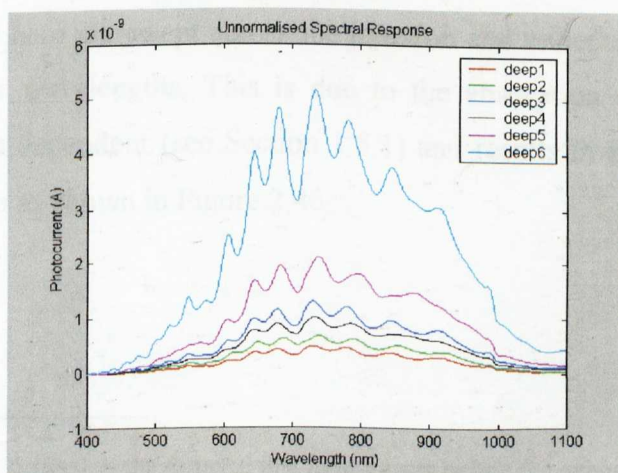
(c) Spectral response curve obtained

Figure 2.43 Spectral response of the deep $100\mu\text{m} \times 100\mu\text{m}$ photodiode with n+ removed (deep4)

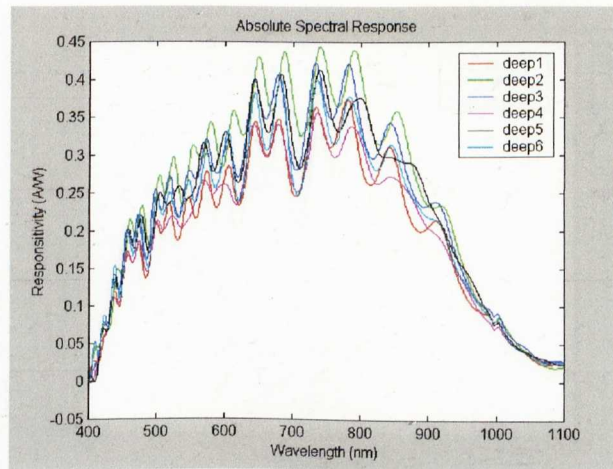
The spectral response curve of the device is typical of a photodiode fabricated in a standard CMOS process [Lee 2003 (Part I), Stoppa 2002]. The peaks and troughs in the response seen are due to the interference of the reflections within the passivation layers covering the active area of the photodiode. The response drops off at wavelengths longer than 1100nm as it approaches the cut-off wavelength of silicon (see Section 1.5.1). Photons with energies smaller than the bandgap energy of 1.12eV at room temperature will not be absorbed at all. At the other end, the reason for the drop-off of responsivity at lower wavelengths is twofold. Firstly light at that wavelength gets absorbed closer to the surface and the photogenerated carriers do not diffuse to the depletion region but are lost due to surface recombination at the Si-SiO₂ interface. Secondly, for a given amount of power P incident on the detector, the shorter the wavelength, λ , the more energy, E_{ph} , the photons have and hence the number of quanta or incident photons is smaller as given below:

$$\text{Photon flux, } N = \frac{P}{E_{ph}} = \frac{P}{hc/\lambda} = \frac{P\lambda}{hc} \text{ photons/s} \quad (2.3)$$

Figure 2.44 shows the spectral response obtained for the different sized deep photodiodes with n+ removed (deep1-6). The peaks of the curves line up reasonably as expected because the variation in the pixel is in the lateral width and not with any vertical differences. As the experiment was carried out with a flooded, not focused, light source, it was susceptible to crosstalk from the substrate – an optical blocking layer was used around the test devices but this only extended to 180 μ m. However, this is not expected to affect the shape of the spectral response.



(a) Measured photocurrent

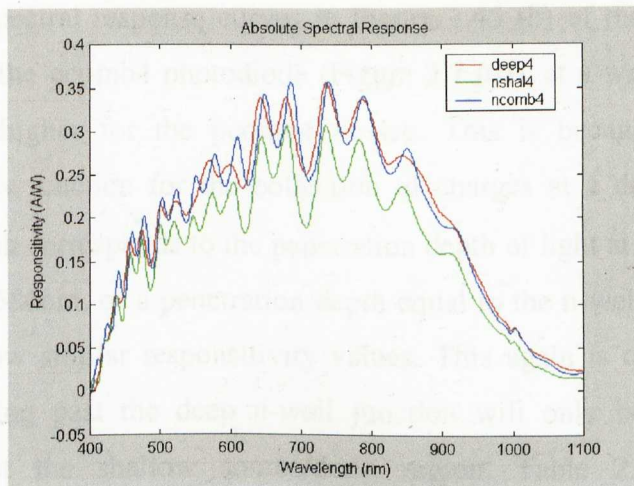


(b) Spectral response

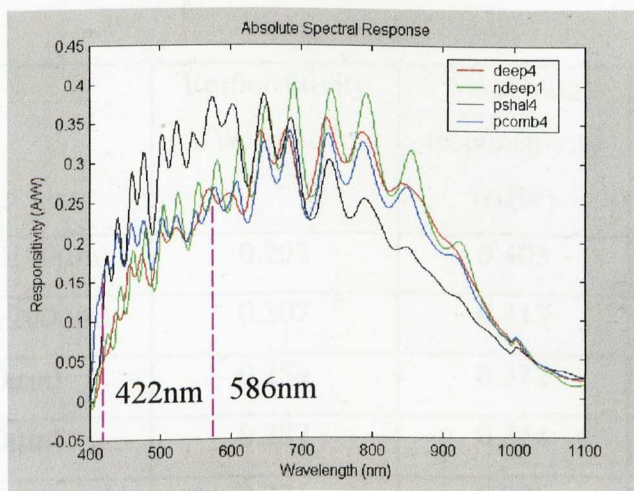
Figure 2.44 Measured spectral response of the deep photodiodes with n+ removed (deep1-6)

Figure 2.45 shows the measured spectral response of the various $100\mu\text{m} \times 100\mu\text{m}$ photodiodes. It shows that the 'deep' and 'ncomb' devices show similar responses with the 'nshal' device showing a lower response (maximum of 0.3081A/W at 735nm). The lower response of the 'nshal' device is believed to be due to the higher doping concentration of the p-well compared to the p-substrate leading to a potential barrier for the collection of diffusion electrons in the substrate [Dierickx 1997]. Also the 'pshal' device shows better response at shorter wavelengths and the 'pcomb' device showing an overall wider response than the others. Overall the 'ndeep1' device gave the best response at longer wavelengths while the 'pshal' device was better at shorter wavelengths. This response is consistent with the fact that light of longer wavelength penetrates deeper into the substrate where the junction of the deep photodiode lies so charges generated here are swept across the junction and collected, while the reverse is true for shorter wavelengths. This is due to the absorption coefficient which is highly wavelength dependent (see Section 1.5.1) and results in a penetration depth²⁸ of light into silicon as shown in Figure 2.46.

²⁸ Penetration depth is defined as the distance that light travels before the intensity falls to 37% ($1/e$) of its original value at the surface.



(a) deep4, nshal4, ncomb4



(b) deep4, ndeep1, pshal4, pcomb4

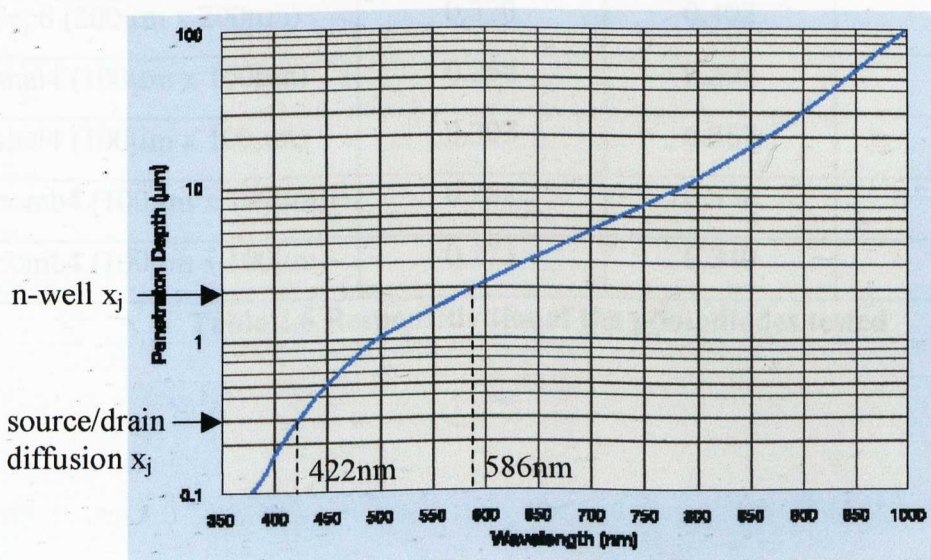
Figure 2.45 Measured spectral response of the $100\mu\text{m} \times 100\mu\text{m}$ photodiodes

Figure 2.46 Penetration depth of light into the silicon substrate at various wavelengths [UDT Sensors Inc.]

Comparing the spectral response curves in Figure 2.45 (b) of the deep4 photodiode (Figure 2.6) and the pcomb4 photodiode (Figure 2.7 (d)), at a wavelength of 422nm the response is higher for the pcomb4 device. This is because pcomb4 has an additional shallow junction for the collection of charges at a depth of 0.3 μm (see Figure 2.46) which corresponds to the penetration depth of light at this wavelength. At a wavelength of 586nm, or a penetration depth equal to the n-well junction depth, x_j , of 2 μm , they show similar responsivity values. This again is consistent since any photons penetrating past the deep n-well junction will only be collected by this junction and not the shallow source/drain region. Table 2.6 summarises the responsivities obtained with the test photodiodes.

Photodiode	Responsivity (A/W) at 667nm	Maximum responsivity (A/W)	Wavelength of max. responsivity (nm)
ndeep1 (100 μm x 100 μm)	0.293	0.403	690
ndeep2 (200 μm x 200 μm)	0.307	0.413	683
deep1 (30 μm x 30 μm)	0.259	0.312	779
deep2 (60 μm x 60 μm)	0.287	0.354	740
deep3 (80 μm x 80 μm)	0.281	0.326	733
deep4 (100 μm x 100 μm)	0.307	0.357	735
deep5 (160 μm x 160 μm)	0.306	0.377	739
deep6 (200 μm x 200 μm)	0.320	0.402	681
nshal4 (100 μm x 100 μm)	0.258	0.308	735
pshal4 (100 μm x 100 μm)	0.303	0.387	647
ncomb4 (100 μm x 100 μm)	0.282	0.358	686
pcomb4 (100 μm x 100 μm)	0.273	0.340	685

Table 2.6 Responsivities of the photodiodes tested

2.5 CHAPTER SUMMARY

Unlike CCDs which have a specially tailored process and structures such as buried channels and surface state pinning to achieve very low dark current levels, photodetectors in a standard CMOS process make use of the parasitic junctions that exists. The work done in this chapter was carried out in order to evaluate the design and use of these junction photodiodes from a standard CMOS process. Several factors need to be considered such as the dark current, the capacitance and its variation with bias, the responsivity of the device and its spatial and chromatic variation. The dark current determines the minimum sensitivity of the device and has two main sources [Shcherback 2002]: dark current from the diffusion of carriers across the depletion region which depends on the doping concentration, bandgap, temperature, bias voltage and active area, and stress induced or defect generated leakage current which depends on the active area shape and bias voltage. For the same fill factor, the smoother the shape the lower the leakage current. This was clearly seen from the observed spatial response of the devices where edge effects showed increase leakage current. The dark current for the devices tested was of the order of 1pA or less for a reverse bias voltage of 2 - 4V.

For applications where speed of response is important, the junction capacitance of the devices needs to be small for a fast response time. The capacitance of the deep device was shown to be smaller than the shallow devices with the presence of the inadvertent Schottky barrier diode in reverse lowering the capacitance further. In determining the spatial response of a device, the issue of crosstalk between pixels and from the substrate was highlighted. Due to the large diffusion lengths in silicon, either a metal shield or a guard ring is required to prevent degradation of the contrast of the image obtained. Also, due to the presence of edge effects the response of the photodiodes does not scale linearly with area but is affected by the peripheral response. So in cases where there is a trade-off between sensitivity and resolution, this must be taken into account. The junction photodiodes were also shown to be very linear with light power, with saturation level not yet reached for an incident light power of 2.7mW. The linearity range of a photodiode can be extended slightly by applying a reverse bias voltage [UDT Sensors Inc.].

The deep or well-substrate photodiode showed better responsivity than the shallow devices due to its wide depletion region caused by the relatively low carrier concentration in the n-well. Since it is deep it is also able to collect the minority carriers photogenerated deep in the substrate provided that they are generated within a diffusion length of the depletion region. In terms of spectral response, the deep photodiode has better spectral response at longer wavelengths while the shallow performed better at shorter wavelengths. This is due to the absorption coefficient and penetration depth of light into silicon, where light of longer wavelength penetrates deeper into the substrate. The deep photodiode is sensitive to substrate noise and crosstalk from the neighbouring photodiodes due to its large and deep collection region while the shallow n+/p-substrate photodiode has good substrate noise immunity due to the presence of the deep field oxide (FOX) implants. Also the presence of the diffusion implant at the surface helps reduce the collection of dark current generated at the surface states of the Si-SiO₂ interface. The shallow p+/n-well photodiode is the least sensitive to substrate noise and crosstalk with neighbouring pixels because each junction is isolated within its own n-well [de Lima Monteiro 2002]. However the presence of the n-well also means that arrays using these photodiodes are less dense with the n+/p-substrate photodiode offering the best packing density. Noise was not characterised as the noise components in a bare photodiode without any additional circuitry are small and the large shunt resistance of the photodiode gives rise to a very small noise bandwidth (see Section 1.5.3). In most applications, where there is sufficient light budget, the imaging system tends to be photon shot noise limited [Hornsey 1999c]. In addition, the connection of the photodiode to readout circuitry will induce another form of noise known as read noise, which limits the noise at low-light levels.

In summary, this chapter has demonstrated that the photodiodes present in a standard CMOS process offer great potential as an optical detector. This work provides an essential foundation to the rest of this thesis. In the next chapter, the design of a hardware emulation system of the optical centroid processor will be presented which makes use of a full custom array of photodiodes as the front end.

CHAPTER 3

HARDWARE EMULATION SYSTEM

3.1 INTRODUCTION

The fabrication of an ASIC (Application Specific Integrated Circuit), especially one that contains analogue components, carries the possibility that the design may fall outside specifications and hence more than one fabrication iteration may be required before a satisfactory operating circuit can be realised. This carries a heavy cost penalty. Hence, a more conservative approach of a hardware emulation system prior to ASIC fabrication has been adopted in order to reduce the number of iterations needed. The hardware emulation system consists of a photodiode array as the optical front end and a reconfigurable digital device (called a Field Programmable Gate Array or FPGA) for the digital centroid processing. Once the emulation hardware confirms the satisfactory performance of a design in its intended application, it can then be converted into a mask programmed CMOS integrated circuit. Due to the re-programmable nature of the FPGA the hardware emulation environment can also be used to evaluate many other optical processing algorithms prior to ASIC fabrication.

3.2 SYSTEM OVERVIEW

The hardware emulation system is shown in Figure 3.1 and consists of two printed circuit boards: a main motherboard and a smaller daughter board. The motherboard contains a single channel 16-bit analogue-to-digital converter (ADC), a Field Programmable Gate Array (FPGA), an RS232 transceiver for a PC serial interface, LED displays for debugging purposes and miscellaneous switches for initiating various test routines under user control. The second, smaller, daughter board contains an optical front end with a 5 by 5 photodetector array for optical light detection, multiplexers for pixel access and current-to-voltage conversion prior to digitisation. The individual parts of the system will be described in the following sections.

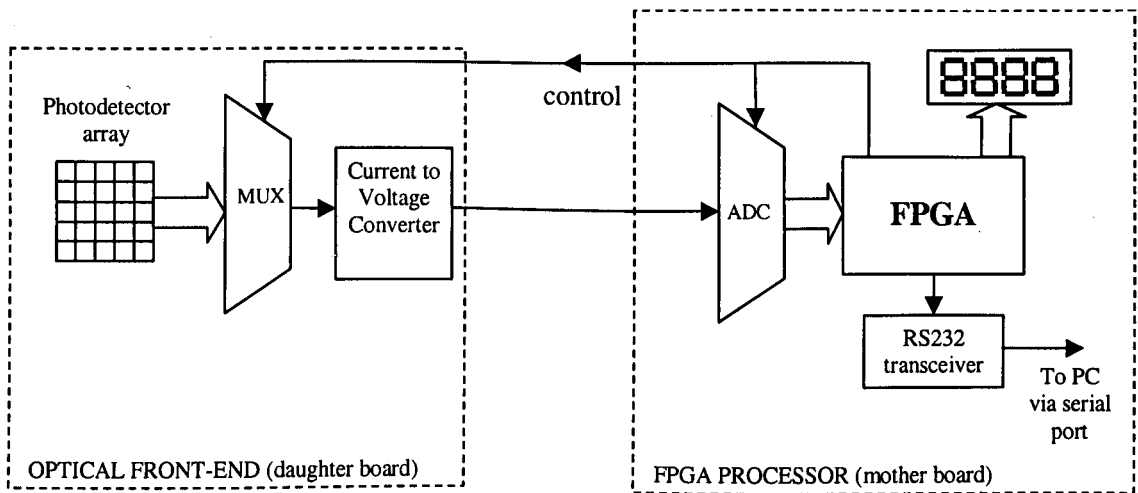


Figure 3.1 Block diagram of centroid emulation hardware

3.2.1 OPTICAL FRONT END

Initially a commercial photodetector array from Centronics (part number MD25-5T) [Centronic Ltd. 1998] was used in the front end. This device is a 5 by 5 photodiode array with a pixel size of 2.7mm x 2.7mm, a wavelength range of 340nm to 1100nm and a quoted responsivity of 0.18A/W at 436nm. This allowed the design and testing of the hardware emulation system to be carried out in order to locate any possible problems. Once confirmed the fabrication of a full custom photodetector array was carried out and once the fabricated array (5 by 5 photodiode array in PDfinal)²⁹ was tested, it was incorporated into the emulation system in place of the commercial photodetector array. The size of the array chosen is a tradeoff between linearity and positional range with complexity and the desired centroid processing time as discussed in Section 1.4.3.2. Each pixel in the full custom array has a size of 100 μ m x 100 μ m and though the exact size of the photodiodes is not crucial, there is a compromise between ease of focusing, efficient use of silicon area and light budget when deciding upon the pixel size.

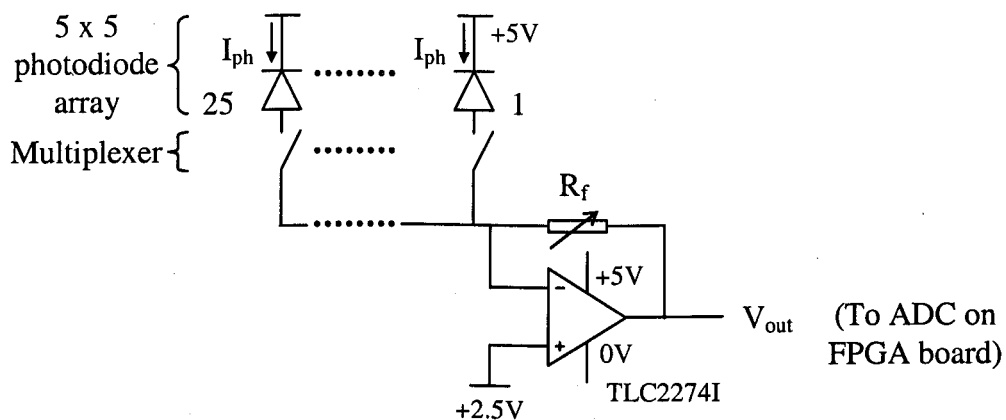
Both the commercial and the full custom array have a passive pixel architecture with no active circuitry at each photodiode. Current to voltage conversion is achieved using an op-amp in the transimpedance mode i.e. with a feedback resistance. A 5M Ω

²⁹ Note that the photodiode type used in the full custom array is the combined device which is leaky in the reverse bias. However, in the emulation system the photodiodes were biased at 0V where the operation of the combined devices is acceptable.

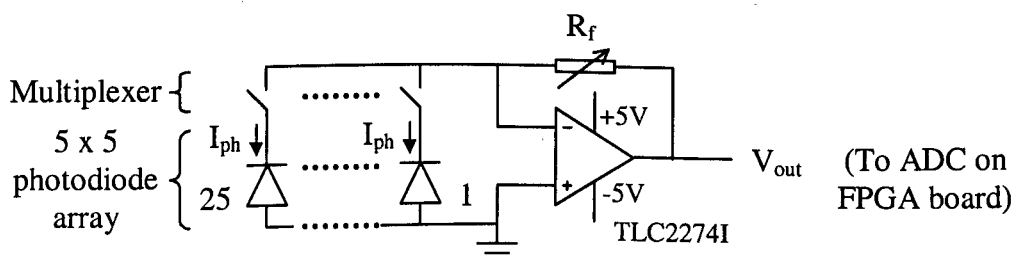
variable resistor was used to allow the design to cope with different photocurrent levels and hence light intensities. A single transimpedance amplifier is used and multiplexers are used to select each photodiode output in turn to be converted. Serial multiplexers (Maxim MAX349) [Maxim Integrated Products Inc. 1998] were used to reduce the number of control signals needed. The schematics and PCB layouts for the daughter board with the commercial photodiode array and for the daughter board with the full custom photodiode array are shown in Appendix A3.1 and A3.2 respectively.

The choice of op-amps is crucial when using a large feedback resistance to detect a small photocurrent. An op-amp with low input bias current is necessary. The input bias current should be significantly smaller than the photocurrent that is to be converted because the large feedback resistance will convert this input bias current into a dc offset voltage at the output of the op-amp for every pixel. This background offset will significantly affect the centroid algorithm by shifting the centroid position towards the centre. Initially the Texas Instruments TLE2024Y op-amp [Texas Instruments Inc. 1997] was used in the front end (Appendix A3.1 and 3.2) which had an input bias current of 50nA but this was replaced with the pin-compatible TLC2274I [Texas Instruments Inc. 2000] which had an input bias current of only 1pA. In addition, the TLC2274I has a low noise voltage of $9\text{nV}/\sqrt{\text{Hz}}$ and rail-to-rail output voltage, hence providing a larger dynamic range.

The Centronics photodiode array has a common cathode configuration and hence the photodiodes must be wired as shown in Figure 3.2(a). If the non-inverting input of the transimpedance amplifier is biased at 0V, the output of the transimpedance amplifier will be a negative voltage. However, the input voltage range for the ADC on the FPGA board was hardwired for 0 to 5V operation (see Section 3.2.2.2). So either a 2nd op-amp configured as an analogue inverter is used, but this would also require generating a -5V supply for the daughter board, or the non-inverting input of the transimpedance amplifier is biased at 2.5V. The latter was chosen though this meant a decrease in dynamic range by half. But for testing purposes this was adequate. The MAX873 voltage reference generator [Maxim Integrated Products Inc. 1992] was used to generate the $2.5\text{V} \pm 1.5\text{mV}$ reference.



(a) Common cathode configuration of Centronics photodiode array



(b) Common anode configuration of the full custom photodiode array

Figure 3.2 Connection of photodiode array pixels to the current-to-voltage converter on the daughter boards

In the case of the full custom array, the photodiodes have a common anode configuration (Figure 3.2(b)) and the output of the transimpedance amplifier goes from 0 to 5V. However, due to the on-resistance of the switches (60Ω for MAX349 and 10Ω for MAX4514 [Maxim Integrated Products Inc. 1996c]), the input voltage at the switches go negative when photocurrent is drawn. So this meant the switches had to be able to cope with a negative input voltage range. So a MAX660 voltage inverter [Maxim Integrated Products Inc. 1996b] was used to generate a -5V supply while the MAX349s were configured for $\pm 5V$ operation and the MAX4514 was replaced with the dual supply DG418DY [Maxim Integrated Products Inc. 1996a].

3.2.2 FPGA PROCESSOR

The FPGA processor board consists of an ADC for digitising the analogue signal voltages, an FPGA to perform the centroid processing and an RS232 transceiver for transmitting the computed centroids to a PC. Peripheral circuitry includes LEDs for debugging purposes, switches for control and power supply protection. An onboard 25MHz crystal oscillator provides a clock input for the FPGA. The schematic and PCB layout for the FPGA motherboard is given in Appendix 3.3 and the following sections will discuss the construction of the different components of the board.

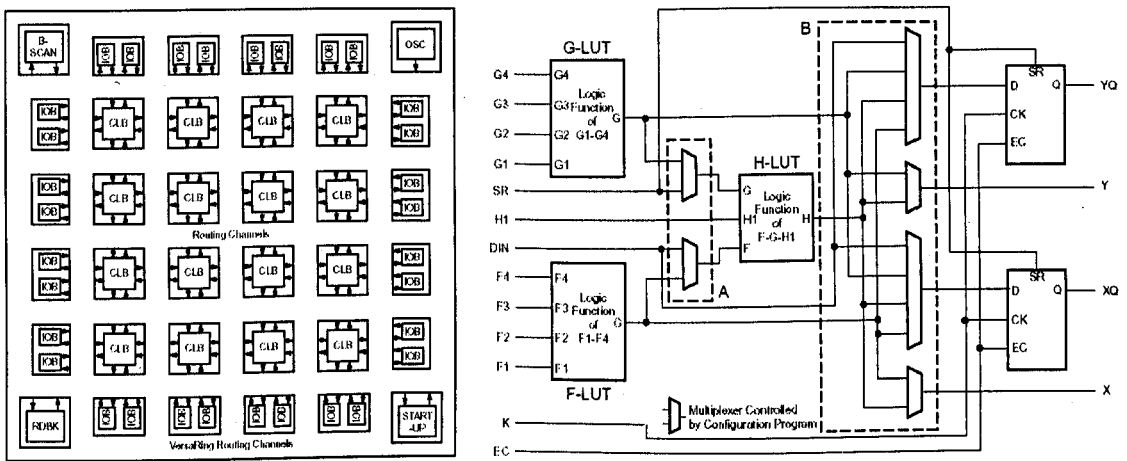
3.2.2.1 FPGA

The processor selected to perform the centroid processing in the hardware emulation system is the Xilinx Spartan XCS40-3PQ208C FPGA [Xilinx Inc. 1999] with 40,000 system gates³⁰ or 784 Configurable Logic Blocks (CLBs)³¹. CLBs are used to implement most of the logic in the FPGA and are organised as a two dimensional array interconnected by routing channels and surrounded by a perimeter of programmable Input/Output Blocks (IOBs). Figure 3.3(a) shows the basic block diagram of a Spartan FPGA. Each CLB consists of primitive hardware elements such as look-up tables (LUT) and positive-edge triggered flip flops as shown in Figure 3.3(b). Each IOB controls one package pin and can be configured for input, output, or bidirectional signals³².

³⁰ This is the quoted maximum value but the typical gate range can vary from 13,000 – 40,000 logic and RAM gates depending on how much of the resources can be utilised in a design. It is more common to quote the number of CLBs used.

³¹ The Spartan devices with speed grade -3 have a specified minimum clock high time and clock low time of 4.0ns. So theoretically these devices can be run up to a speed of 125MHz.

³² Note that if an I/O is unused after configuration, it is configured as an input with a pull-up resistor activated.



(a) Block diagram of Spartan FPGA (b) Simplified Spartan CLB Logic Diagram

Figure 3.3 Architecture of a Xilinx Spartan FPGA

The FPGA is programmed by loading configuration data into its internal static memory cells. The values stored in these memory cells determine the logic functions and interconnections implemented in the FPGA. The board is designed to allow configuration from a Xilinx XCS17S40-PD8C [Xilinx Inc. 1999] serial PROM (Master Serial mode) or from an external device such as a PC via an XChecker cable (Slave Serial mode) as shown in Figure 3.4. In the Master Serial mode, the FPGA's internal oscillator generates a Configuration Clock (CCLK) for driving the serial-configuration PROM (SPROM) while in the Slave Serial mode, CCLK is driven by an external signal. Clearing of the configuration memory is done using the PROGRAM pin which is controlled by a pushbutton on the board. *INIT and DONE provide status outputs during configuration of the FPGA. Connecting the *RESET of the SPROM to the *INIT output of the Spartan device ensures that the SPROM address counter is reset before the start of any configuration.

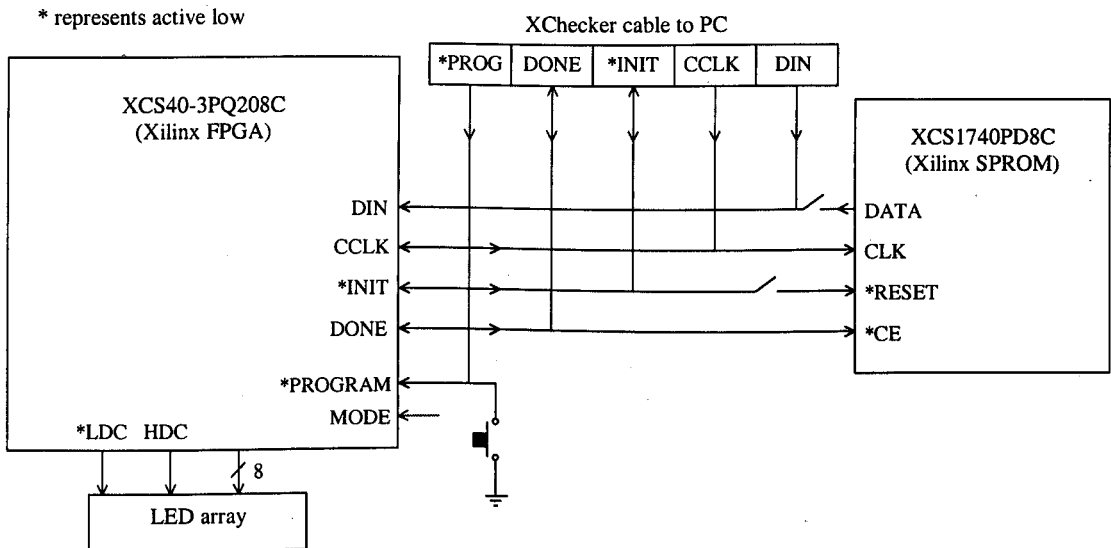


Figure 3.4 Configuration of the Spartan FPGA via a PC or a SPROM

3.2.2.2 Analogue-to-Digital Converter (ADC)

The ADC used for the hardware emulation system is the Burr-Brown ADS7807UB 16-bit sampling successive approximation ADC [Burr Brown Corporation 1994]. Figure 3.5 shows how the ADC is controlled and connected to the FPGA and also how the processed data from the ADC is to be displayed or transmitted to a PC. The ADC was hardwired for an input voltage range of 0 to 5V. The ADS7807UB can acquire and convert 16-bits in $25\mu\text{s}$ (40kHz) while consuming only 35mW (max) with a maximum integral non-linearity error of $\pm 1.5\text{LSB}$ and no missing codes. It has 8 parallel output lines and a BYTE signal that has to be controlled to read the high byte and low byte in turn. Conversion is initiated by controlling a convert signal, R/C, with the 25MHz clock used to run the FPGA, as shown in Figure 3.6.

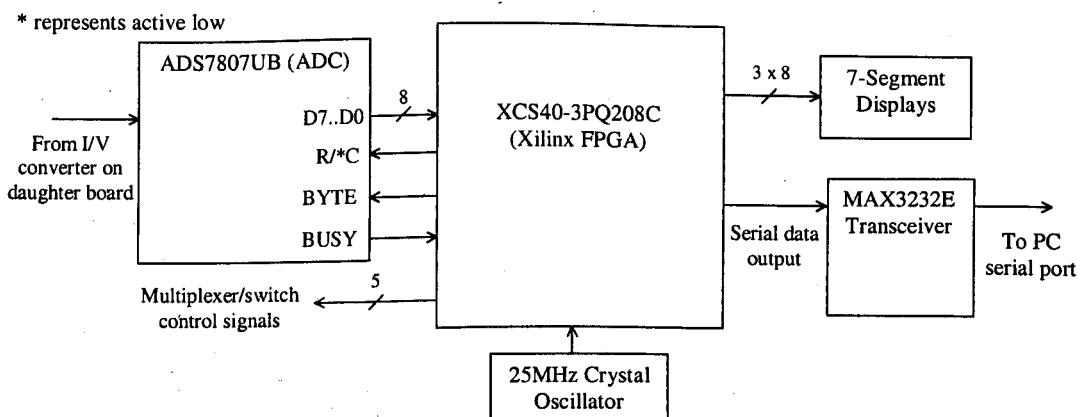


Figure 3.5 Connection of ADC to the FPGA motherboard

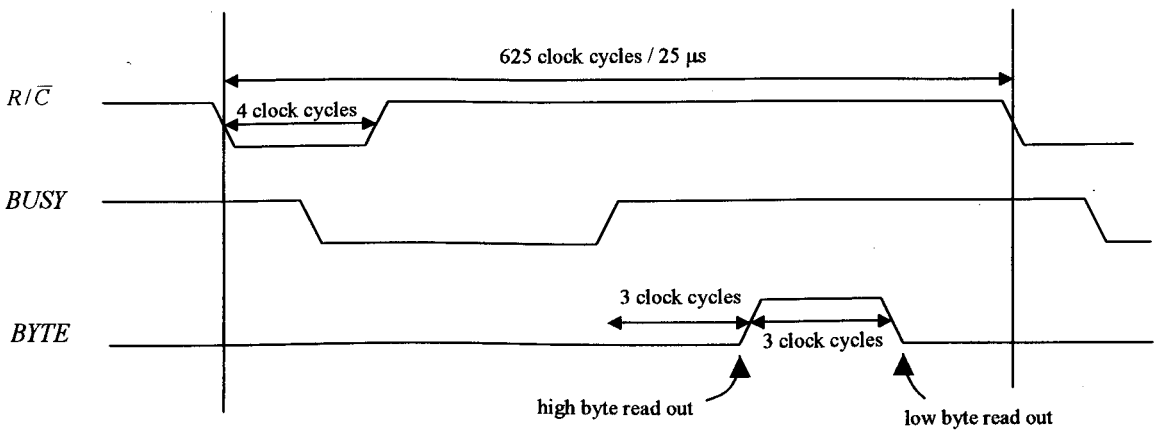


Figure 3.6 Control of ADC by FPGA using a 25MHz clock

The ADC used had a low input impedance of only 20k Ω . But the output impedance of the transimpedance amplifier was significantly smaller (130 Ω) so there was relatively no volt drop of the input voltage due to this and no buffering was required. An anti-aliasing filter was incorporated into the front end of the ADC. The switches of the multiplexers on the daughter board are updated 10 clock cycles before the ADC convert signal is sent.

3.2.2.3 RS232 Transceivers

The RS232 (or EIA232) standard was introduced to ensure reliable serial communication between devices. In the RS232 standard³³, voltages of -3V to -25V with respect to signal ground (pin 7 on DB25 connectors or pin 5 on DB9 connectors) are considered logic '1' while voltages of +3V to +25V are considered logic '0'. An RS232 transceiver is a level converter IC which converts CMOS level voltages to RS232 level voltages and vice versa, and for this purpose, a MAX3232E transceiver [Maxim Integrated Products Inc. 2000] was employed which has two receivers and two drivers guaranteed to run at data rates of 250kbps while maintaining RS-232 output levels³⁴.

³³ In RS232 the start bit is logic '0' and stop bit is logic '1' and the least significant bit is always the first bit sent.

³⁴ The output voltage swing of the transmitters is ± 5.4 V (typ).

For serial communication with a PC, a null modem connection is made. For synchronising, the receiver on the PC scans the incoming data for valid start and stop bit pairs. The receiver uses a 16x clock for detecting the incoming start bit, so the occurrence of the start bit will be located within the $\pm 1/2$ 16x clock cycle or $\pm 1/32$ bit or $\pm 3.125\%$. The design of the transmitter for generating and sending the output data will be discussed in Section 3.3 when the design of the digital centroid processor is presented.

3.2.2.4 Peripheral Circuitry

LED bargraphs and 7 segment displays are used to display results and for troubleshooting³⁵. A logic low level on an FPGA output connected to the LEDs draws current through the LEDs turning them on³⁶. In addition, an 8-way rocker DIL (Dual In-Line) switch and 3 tactile pushbutton switches were included to allow the user to easily control several input pins of the FPGA. Some of these input switches are used as mode or control inputs during the configuration of the FPGA.

All the components on the board operate on a 5V supply. Power supply protection is incorporated to protect these devices from voltage surges and incorrect powering of the board. The power supply protection circuit is shown in Figure 3.7 and includes a fuse, a zener diode, a varistor (voltage-dependent resistor) and a PMOS power MOSFET. When there is a power surge or a large voltage spike, the zener diode goes into breakdown and the varistor's resistance rapidly decreases creating a shunt path for the over-voltage. The PMOS SI9430SDY is used to protect the board from incorrect connection of the power supply³⁷. If the supply is connected correctly and the gate of the PMOS is connected to 0V and the source is connected to the +5V input, then $V_{gs} < 0$ and the PMOS is on. Else if it is connected in reverse say, $V_{gs} > 0$ and the PMOS turns off cutting the supply to the board.

³⁵ To display the output data in decimal on the 7 segment display, binary to BCD (binary coded decimal) conversion is performed using the FPGA prior to output. Only three 7 segment displays were available, requiring data larger than 12 bits to be truncated. If the data is to be displayed in units other than binary, say volts, the effect of this truncation has to be taken into account when scaling.

³⁶ Resistors are used to limit the current drawn from the LEDs.

³⁷ $R_{ds(on)} = 0.1\Omega$

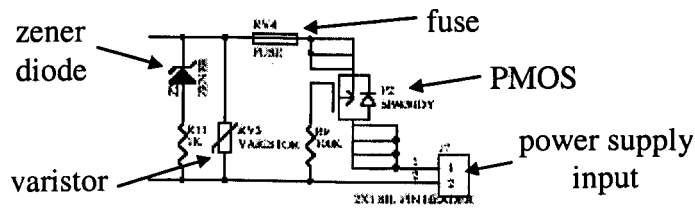


Figure 3.7 Power supply protection circuitry

Under quiescent conditions, the FPGA board drew 105.6mA from a 5V regulated supply. With all the LEDs on, it drew 189.6mA and with the Centronics array daughter board connected, 213.5mA was drawn. The board also allows users to use an unregulated power supply or a non-compliant supply voltage such as a 9V battery. It does this by having a second power supply input with the same power supply protection circuitry but connected to a MAX667 voltage regulator [Maxim Integrated Products Inc. 1994] prior to connection to the board's power lines. The power supply protection in the voltage regulator path has the same structure as that in the unregulated supply path but with different ratings. For example, the zener diodes have a zener voltage of 5.1V and 16V respectively in the regulated and unregulated supply paths. The MAX667 accepts a +3.5V to +16.5V input and has a maximum dropout voltage of 350mV and maximum supply current of 250mA, sufficient for the design needs.

3.2.2.5 Layout and Testing of FPGA Board

The PCBs were designed in Protel and the PCB motherboard was sent away for fabrication while the daughter boards were built in-house. When laying out the board, several basic rules were adhered to for reducing EMI (Electromagnetic Interference) and crosstalk:

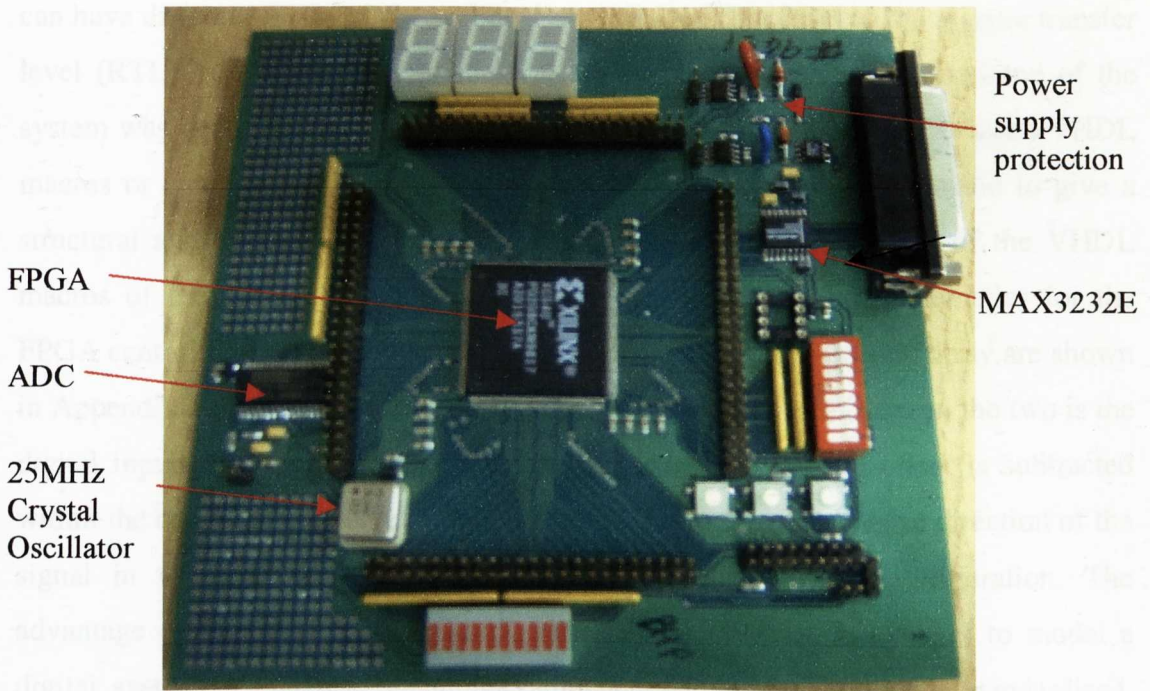
- Use a large ground plane.
- Make power supply tracks large.
- As far as possible keep signals away from power lines.
- Avoid creating a loop when routing the power line³⁸. Use a star configuration.

³⁸ Routing traces in a loop around the board can increase the board's susceptibility to external fields as well as increase the generation of them.

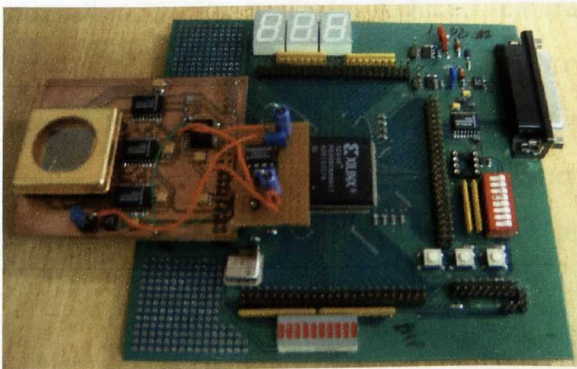
- Two decoupling capacitors (0.1uF ceramic and 10uF tantalum) are placed as close as practically possible to each power and ground pin of all the IC components³⁹.
- Use of surface mount technology (SMT) components is preferred over through-hole mounting due to its shorter lead length and hence lower inductance.
- Use a regulated power supply. If not, use the onboard voltage regulator.

Figures 3.8(a), (b) and (c) respectively show photos of the FPGA board, the FPGA board with the Centronics array board attached and the FPGA board with the full custom array daughter board connected. A basic test of the FPGA was to generate a counter and observe the outputs on the LEDs. The serial link was tested by generating and sending data from the FPGA to a PC [Goodwin 1992] via the onboard transceivers. For the testing of the ADC, various analogue test signals such as a triangle wave and a sine wave input were acquired and converted and the results transmitted via the RS232 port to the PC. The analogue front end was tested by applying fixed voltages via resistors to generate an input current at the bare photodiode array sockets and observing the multiplexer switching and I-to-V conversion outputs. Once the system was fully tested, it was inserted into an optical bench setup for obtaining centroids.

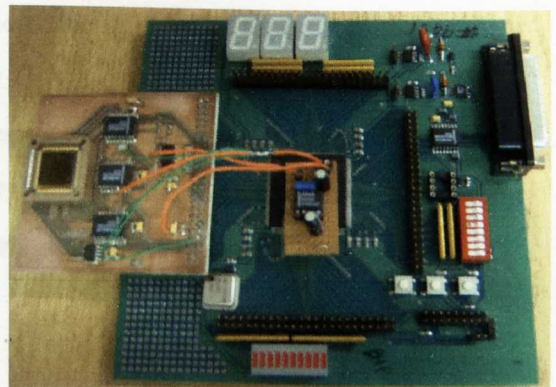
³⁹ A real capacitor includes both an inductor and resistor in the form of leads, traces, and even ground planes in series with it. This means that, in a circuit, a capacitor acts as a low-impedance element only over a limited range of frequencies. To extend this frequency range, many references propose adding a second capacitor to bypass frequencies outside the limited range of the single capacitor.



(a) FPGA processor board



(b) FPGA board with "commercial photodiode array daughter board connected



(c) FPGA board with full custom photodiode array daughter board connected

Figure 3.8 Photographs of the boards designed and built for the centroid hardware emulation system

3.3 DIGITAL CENTROID PROCESSOR DESIGN

This section discusses the design of the digital centroid processor on the FPGA. Digital systems can be specified in three domains [Yalamanchili 2001]. Under the functional domain, the system is described in terms of its operation or behaviour. Under the structural domain, the system is specified by the interconnection and hierarchy of its components and finally in the physical or geometrical domain, it is specified by the physical layout of the components. At the same time, a digital system

can have different levels of abstraction from the algorithm level to the register transfer level (RTL)⁴⁰ to the boolean logic level. In this design, the digital backend of the system was described in VHDL where the individual behavioural RTL-based VHDL macros or components are placed and connected together on a schematic to give a structural and graphical description of the system. A block diagram of the VHDL macros of the centroiding system is shown in Figure 3.9. The schematics for the FPGA centroiding system of the Centronics array and the full custom array are shown in Appendix A3.4 and A3.5 respectively⁴¹. The only difference between the two is the digital inputs in the Centronics array case are inverted and an offset is subtracted within the centroid processor block in order to account for the inverse direction of the signal in the Centronics array due to its common cathode configuration. The advantage of using VHDL (VHSIC⁴² Hardware Description Language) to model a digital system is that it is technology independent hence allowing a standardised, portable model of electronic systems. Technology independence will allow technology migration to, for example, reduced feature lengths in ASICs (deep sub-micron) or from, say, a Field Programmable Gate Array (FPGA) to ASIC where an FPGA has been used to prove the functionality of a design. In addition, a VHDL model of a digital system can be described both structurally and behaviourally and at different levels of abstraction, providing a means of managing large, complex designs.

⁴⁰ At the register transfer level (RTL) a digital system is represented by a set of registers and a set of transfer functions describing the flow of data between the registers.

⁴¹ As the inputs of the transimpedance amplifier for the Centronics array are biased at 2.5V instead of 0V, its output voltage goes from 2.5V to 0V (10000000 to 00000000) for an effective signal level of 00000000 to 10000000 so inversion (shown in schematic) and subtraction of an offset of 01111111 from the digitised data bits is performed before centroid processing is carried out.

⁴² VHSIC: Very High Speed Integrated Circuits

* represents active low

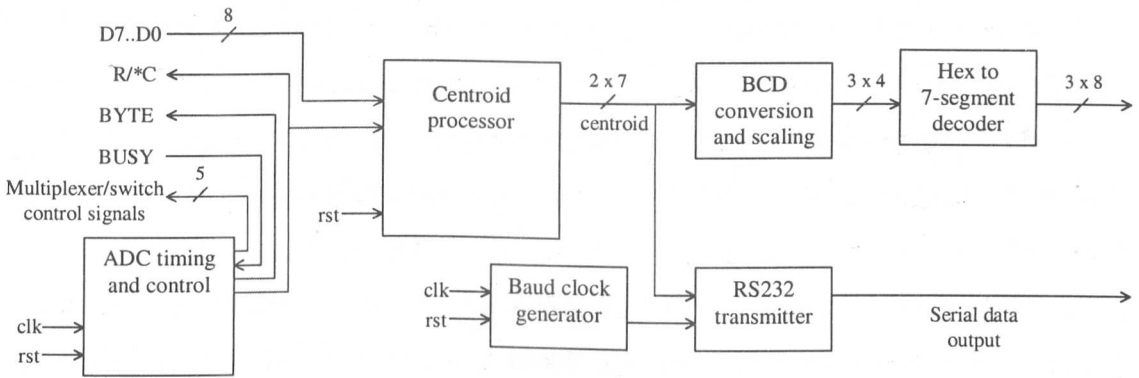


Figure 3.9 Block diagram of the centroid processor implemented in the FPGA

To obtain a centroid from incident light levels of a photodetector array, the 1st order moment of the light levels has to be calculated as described in Section 1.4.2 and given by equation (1.4). A simplified example of a centroid calculation is shown in Figure 3.10. In this example⁴³, a 4 x 4 photodiode array (shaded) with arbitrary light intensity given by the decimal numbers in the top right hand corner of each pixel produces centroid positions of $C(x)=2.53$ and $C(y)=2.68$.

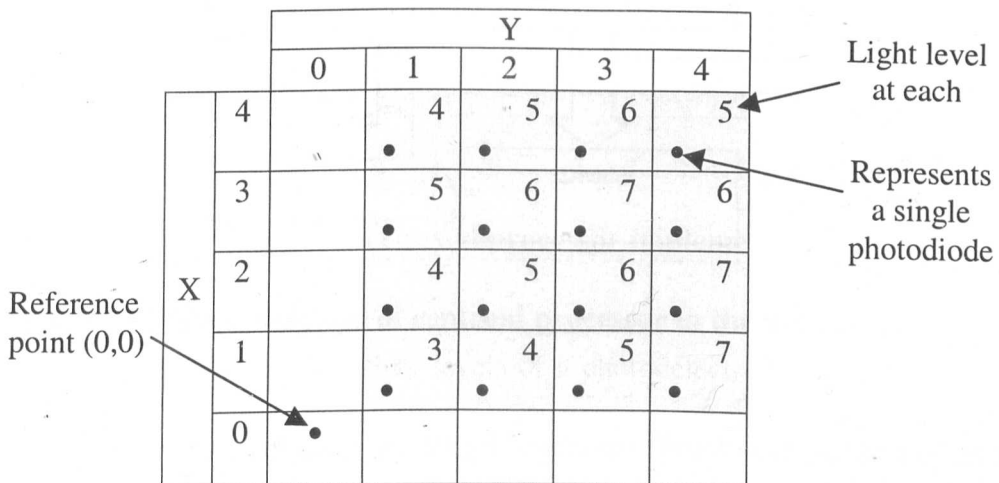


Figure 3.10 Example centroid calculation for a 4 by 4 photodiode array giving centroid positions of $C(x)=2.53$ and $C(y)=2.68$

⁴³ Note that the reference point is chosen outside of the array. If the reference point was chosen as the centre of the array with positive and negative coordinate ranges, this reference point will not carry any weighting in the centroid calculation, leading to poor noise characteristics when the spot is close to the centre.

If the light levels are now represented digitally then these centroid moments can be implemented using the block diagram shown in Figure 3.11 for the x-coordinate and another duplicate block (not shown) for the y-coordinate. Photocurrent data is clocked in sequentially from each photo-detector and multiplied by a counter (Mod $N^{1/2}$, where N is the number of pixel elements in the array) that holds the position of the detector relative to the reference point in the x-direction. The output of this multiplier is continually accumulated via an adder block and the result is divided by the total photocurrent acquired via a separate and parallel running accumulator. The resultant division represents the x-centroid coordinate. A second centroid processing block calculates, in parallel, the y-centroid coordinate.

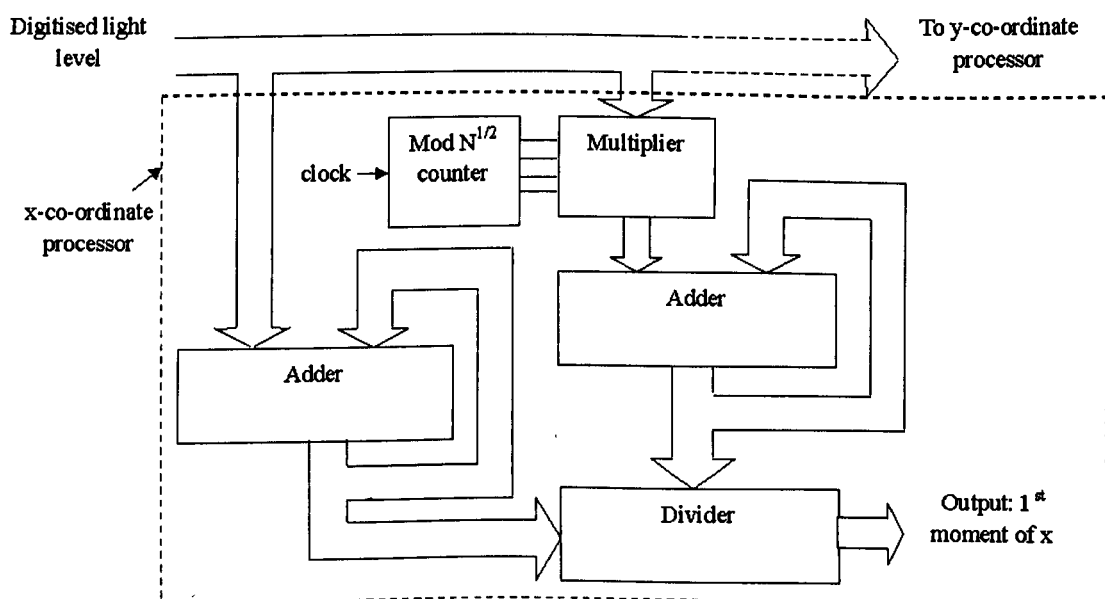


Figure 3.11 Block diagram of centroid processor in the x-direction

For binary addition and multiplication, VHDL operators (functions) within the IEEE numeric_std package are used. The addition effectively synthesises carry look-ahead adders⁴⁴ while the binary multiplication process is effectively a shift and add procedure [Chang 1999]. Binary division however was not supported and division is

⁴⁴ Adders can be implemented using a ripple structure which is small but slow or carry look-ahead adders which is faster but larger.

implemented using shift and conditional subtract operations of long division⁴⁵ [Dewey 1997]. For a 5 x 5 array with a digitised 8-bit input light level, 15 bits are required for the numerator ($255 \times 15 \times 5 + 1$ levels) and 13 bits for the denominator ($255 \times 25 + 1$ levels). This results in a non-floating point quotient output of 3 bits and corresponds to the coordinate range of 1 to 5 of the array, or 001 to 101 in binary. To increase the number of quotient bits and hence the precision of the division process, additional shift and subtract cycles are performed. This represents an increase in the number of cycles of operation with minimal increase in hardware as the dividend and divisor size remains the same (as long as 8-bit representation of light level is sufficient). A 7-bit representation of the centroid coordinates was chosen with 3 non-floating point bits and 4 floating point bits giving a positional resolution of 0.0625 of a pixel.

A centroid is obtained after $N+5$ conversion cycles or pixel cycles from the start of the frame where N , the number of photodiodes is 25. For a 40 kHz ($25\mu\text{s}$) conversion rate, a centroid is obtained after 0.75ms from the start of the frame. The 5 remaining cycles are required to allow the division process to complete. However, a new frame is started after $N+1$ or 26 cycles by making use of the latency during the division process, so centroids are updated every 26 cycles or at a rate 1.54 kHz. The additional cycle in this case is for the latching and reset of the dividend and divisor result prior to division and the start of the next frame⁴⁶.

The calculated centroid positions are then converted into the serial RS232 format with one start bit (logic '0'), 8 data bits, no parity bits and 1 stop bit (logic '1'). The MSB of each byte sent is used to indicate whether it is x or y-data while the remaining 7-bits are for the actual centroid data. A standard RS232 baud clock of 19,200 bits/s is generated to transmit the centroid coordinates, which limits the frame rate to 960 Hz

⁴⁵ Like in long division, the divisor needs to be aligned to the dividend before subtraction can be carried out. This is done by buffering or padding the divisor with additional zeros.

⁴⁶ A conversion cycle was used for convenience sake and a shorter cycle could be utilised by controlling the final latching of the dividend and divisor on a faster clock.

$(19,200 \text{ bits/s} \div 20 \text{ bits})^{47}$. When a baud rate of 38,400 bits/s is used, the full frame rate of the centroid processing is utilised. The serial centroid data is then sent off the FPGA chip to the MAX3232E for RS232 level conversion. In addition to the computation and transmission of the centroid, the digital processor had to control the ADC for the digitisation and acquisition of the photocurrents as well as the serial multiplexers for selecting the individual pixels in turn.

3.4 FPGA CAD ENVIRONMENT AND DESIGN FLOW

The CAD environment used for the development and programming of the FPGA system is the Xilinx Foundation Series 2.1i software [Xilinx Inc.] which fully supports the use of the Spartan device. The design flow for an FPGA design environment is shown in Figure 3.12. VHDL programs are analyzed to check for syntax errors and compiled to a form executable by a VHDL simulator. The analyzed design is synthesized to a library of components, typically gates, latches or flip-flops. Hierarchical designs are synthesized in a bottom up fashion, that is lower level components are synthesized before higher level components. Once the design is synthesized we have a gate-level netlist. This gate-level netlist can now be simulated⁴⁸. Functional simulation is possible but accurate timing simulation is not possible at this point because the actual timing characteristics are determined by the physical placement of this design within the FPGA chip.

⁴⁷ The 25MHz clock is divided by 1302 to obtain a baud rate clock of 19201.23 bits/s. This represents an error in the bit rate of 0.0064% which is not significant and in addition, RS232 receivers are designed to synchronise the transmission at the start of each new byte sent by clocking in the start bit at 16x the baud rate clock.

⁴⁸ Xilinx simulation script files (.cmd) are used to ease the input of test vectors as well as allow simulations to be repeated or modified quickly.

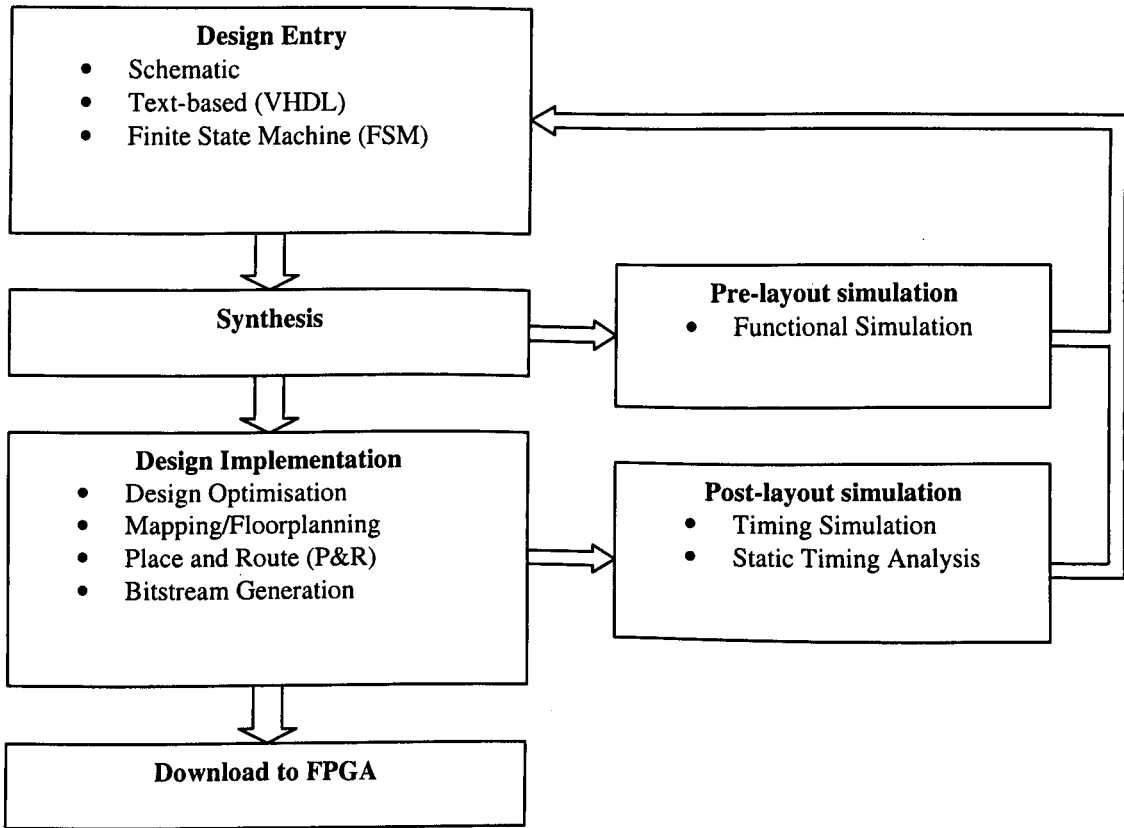


Figure 3.12 FPGA design flow

Once the gate-level netlist is obtained the next step is to map this design onto the FPGA. Mapping a design onto an FPGA involves translating the gate-level netlist produced by the synthesis compiler into a netlist of FPGA primitive hardware components. Locking of the input and output ports on the design schematic to specific physical pins on the FPGA chip is done using the LOC property on the individual ports. The LOC property is a property provided for within the Xilinx Foundation Series software for assigning pin numbers to each input and output pin. This option was preferred over the use of the Constraint Editor to lock the pinouts as it did not always register or store the values entered.

In the Place and Route stage of the design, these primitive hardware components are assigned to actual physical primitives on the FPGA chip and the interconnections between these components are made. The Spartan device, and FPGAs in general, have different types of routing channels from single-length lines between each CLB to long or global lines which run the entire length of the array. These global routing networks can be used to route and distribute critical nets such as clock signals and high fanout signals throughout the device with minimal skew. This is done by placing global

buffers from the library on the design schematic. Besides layout constraints, timing constraints can also be placed in the User Constraints File (UCF) for controlling and optimising the placement and routing. In addition, Xilinx Foundation Series 2.1i allows several options to be selected during design implementation such as optimisation for area or speed, number of place and route passes to make, configuration of input/output pins (TTL or CMOS) and so on.

After place and route, the design can be simulated with propagation delays of the routed signals incorporated. Two types of post-layout simulation are possible in the Xilinx Foundation Series design environment, namely logic timing simulation and static timing analysis. In the timing simulation, user defined test vectors are dynamically propagated through the circuit and the resulting output waveforms are observed. The time required to perform the simulation limits the number of input vectors and circuit operating modes, and the length of circuit operation that can be simulated. Static timing analysis on the other hand does not have a simulation cycle and therefore do not schedule events. Instead of evaluating logic functions, static tools sum up and compare delays through paths, relative to pre-defined clocks. Static timing analysis will determine the critical paths in the design and verify that the design meets the timing constraints set. Static timing analysis is faster and provides a wider coverage but is less comprehensive and may generate false paths. Note that the Timing Analyzer (i.e. static timing analysis) in Xilinx Foundation 2.1 does not detect setup and hold violations but these violations are highlighted during logic simulation. Xilinx FPGA chips come with different speed grades and the static timing analyser can provide a quick analysis of the effect of different speed grades on the same design.

Once the design has been properly verified, the generated configuration bits during implementation can be downloaded onto the FPGA via a Xilinx XChecker cable. Due to the reprogrammability of the device, the design can be verified in-circuit using real data and any errors can easily be corrected and the device reprogrammed until the desired performance is achieved.

3.5 RESULTS OF HARDWARE EMULATION SYSTEM

For testing the hardware emulation system, a $20\mu\text{m}$ diameter laser beam (a double YAG laser at 532nm with approximate output power of 0.86mW) was scanned across the array at a speed of $2000\mu\text{m}/\text{sec}$. Figure 3.13 shows the experimental setup used for testing the hardware emulation system. Centroid values were computed by the FPGA and serially transmitted in real time to a PC at a rate $38,400$ bits/s. Initially the Centronics photodiode array was incorporated in order to test the VHDL centroid algorithm. Then the full custom photodiode array was included to evaluate the performance of a full custom array for centroid detection. The results of these tests are shown in the following sections.

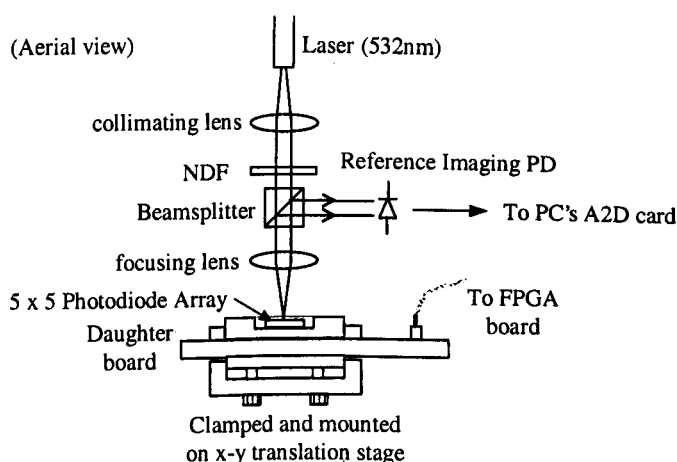


Figure 3.13 Experimental setup of scanning system

3.5.1 COMMERCIAL PHOTODIODE ARRAY (CENTRONICS)

Figure 3.14 shows a grey scale map of the centroid values successfully recorded at each position on the array – each photodiode was of size $2.7\text{mm} \times 2.7\text{mm}$. The dark regions correspond to larger centroid coordinates whilst lighter regions correspond to small centroid coordinates. As expected, as we scan in the x-direction, the x-centroid values increases while the y-centroid values remain constant and vice versa. Since the laser beam size ($20\mu\text{m}$) is less than the size of one pixel then a stepped appearance can be seen as the beam moves across the array passing from one discrete detector to another. The guard rings of the Centronics array was left floating (Appendix A3.1)

and this gave rise to crosstalk but the purpose of the test was to check the functionality of the centroid processing algorithm which shows the desired response.

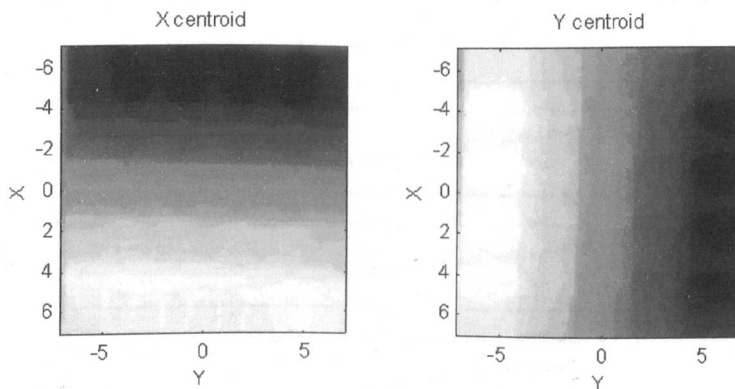


Figure 3.14 Image map of x and y centroids for the Centronics array

3.5.2 FULL CUSTOM PHOTODIODE ARRAY

Centroid values were again calculated in real time by the FPGA with the 532nm laser scanned across the custom made CMOS array – the pixel size is $100\mu\text{m} \times 100\mu\text{m}$. Figure 3.15 shows the y-coordinate centroid values plotted as a function of pixel position for different beam diameter sizes. The array goes from $-250\mu\text{m}$ to $250\mu\text{m}$. Near the edges we can see non-linearity effects as the beam falls off the edge of the array. This effect is more pronounced for larger beam sizes because these will fall off the edge first. For very small beam sizes, we obtain discrete steps in the waveform as we would expect as the beam passes from 1 pixel to another. The steeper rise in centroid value occurs when the beam lands in-between two pixels. As the beam size increases the response becomes more linear in the centre of the array.

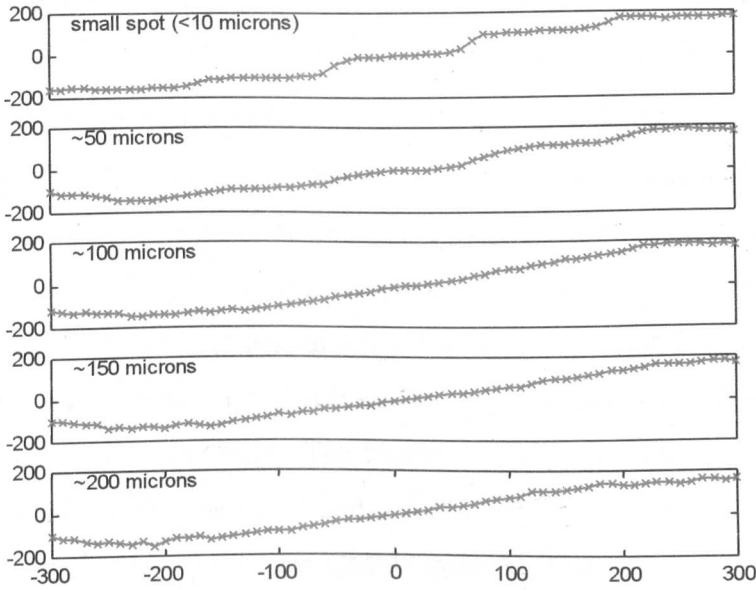


Figure 3.15 Measured position vs. actual position for different beam sizes

Figure 3.16 shows the grey scale centroid maps for both the x and y centroid coordinates obtained in real time by the CMOS array via the FPGA. Again the darker regions correspond to large centroid coordinates and vice-versa for light regions. Again we can see the stepped appearance with small beam size and a smoother appearance for larger beam size. From these results it can be seen that the hardware emulation system has proven the functionality of the digital centroid processor on the FPGA in computing the required centroids.

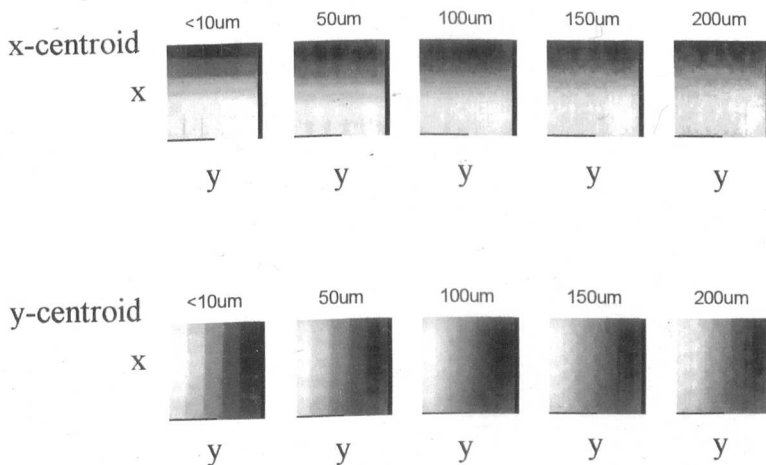


Figure 3.16 Image map of x and y centroids for different beam sizes

3.6 CHAPTER SUMMARY

A hardware emulation system allows the designer to test out various processing algorithms prior to IC fabrication. Although conservative, this approach aims to reduce the number of design iterations needed to produce a working design, and hence can lead to a reduction in cost and time-to-market. The hardware emulation system for the optical centroid detector consists of a 5 x 5 photodiode array, I-to-V conversion of the photocurrent, ADC of the signal voltage and the centroid calculation of the digital data with a reprogrammable FPGA. The hardware emulation system was tested with both a commercial photodiode array and a full custom standard CMOS photodiode array fabricated in the Mietec 0.7 μ m CMOS process. Current-to-voltage conversion was achieved using a transimpedance amplifier with a feedback resistance.

The centroid processor successfully computed the centroids at a rate of 1.54kHz which was limited by the maximum conversion frequency of the ADC of 40kHz. But even at this speed, an array of these centroid detection systems operating in parallel will enable fast low cost adaptive optical systems to be built. The centroid data was then transmitted off-chip to a PC using a RS232 transmitter. Having proven the functionality of the digital centroid processor and the use of a full custom photodiode array, the next step in the design was to integrate the full custom array with the digital centroid processor onto a single CMOS IC chip.

CHAPTER 4

DESIGN, FABRICATION AND TEST OF CENTROID ASIC

4.1 INTRODUCTION

After the performance of a full custom photodiode array and the centroid processing algorithm was verified by the hardware emulation system, the next stage is to integrate the full custom array and the processing onto a single piece of silicon. A block diagram of the overall system is shown in Figure 4.1. The top level schematic of the system is shown in Appendix A4.1 which shows the centroid chip divided into an analogue front end and a digital backend. The analogue front end consists of an active pixel sensor array and analogue-to-digital conversion circuitry. The digital backend consists of the centroid processor and a serial link for transmitting and receiving data off-chip. The individual components of the system are discussed in greater detail in the following sections.

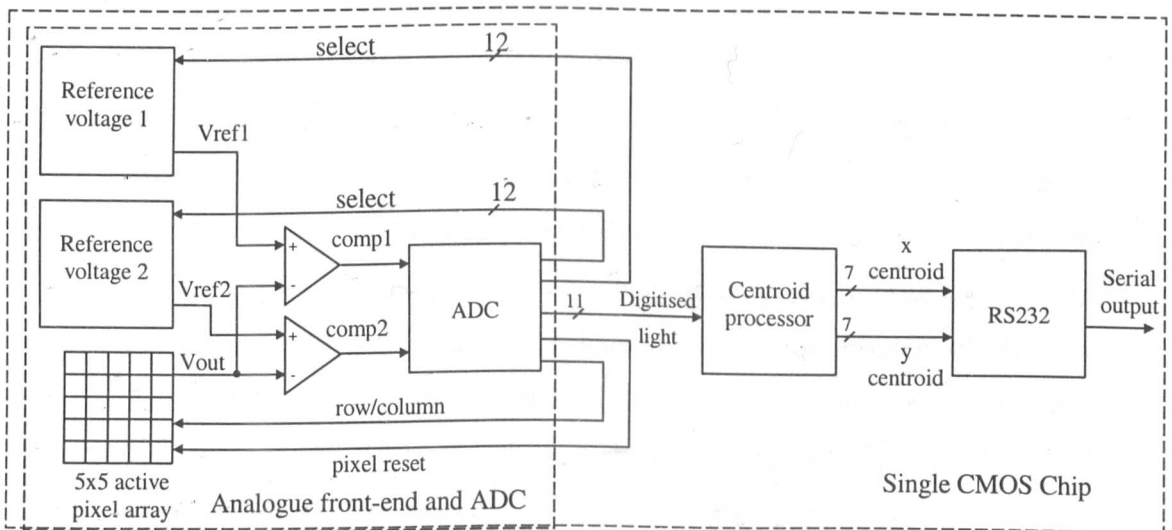


Figure 4.1 Block diagram of the single CMOS chip optical centroid processor

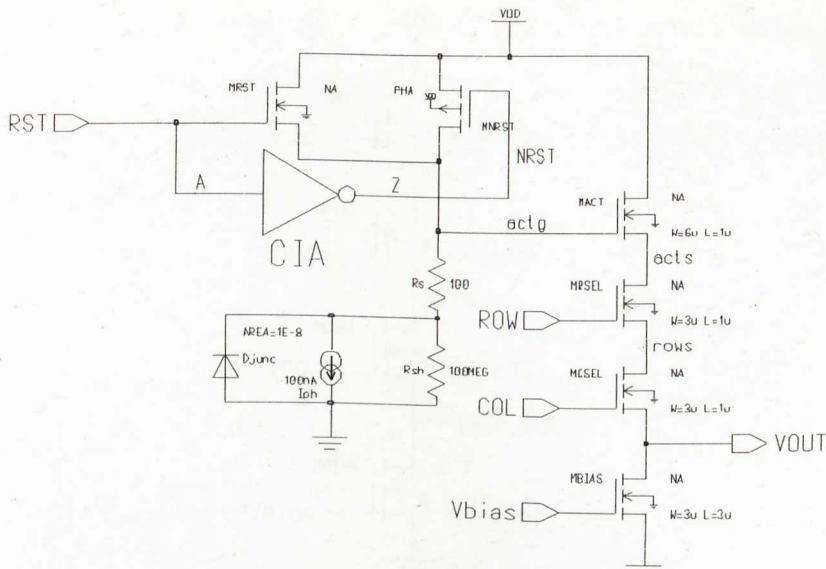
4.2 ANALOGUE FRONT END

The analogue front end in the hardware emulation system consists of a passive pixel array and a transimpedance amplifier with a large feedback resistance. However, it is difficult to integrate a large resistor on silicon and the use of an external resistor would lead to parasitics and noise. Instead, an integrating photodiode APS array was used which incorporates a buffer at each pixel to convert the photocurrent to a discharging voltage output (see Section 1.6.2.1). The optimisation of this architecture is presented in Section 4.2.1 while the digitisation of the output signal is discussed in Section 4.2.2.

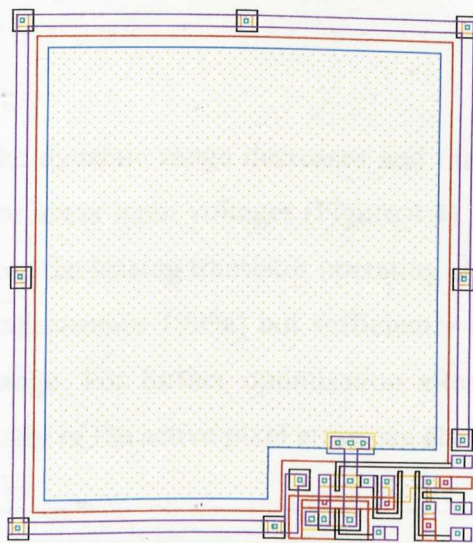
4.2.1 PHOTODIODE AND PIXEL ARCHITECTURE

For the photodiode array, the deep n-well to p-substrate photodiode is used. The deep photodiode has better responsivity and lower junction capacitance than the shallow devices due to its wide depletion region caused by the relatively low carrier concentration in the n-well. Also, it does not suffer from a large leakage current in the reverse bias as the combined shallow p+/n-well and deep n-well/p-substrate photodiode in the original array does. Furthermore, its higher responsivity at longer wavelengths make them suitable for adaptive optical systems where longer wavelength operation means less stringent requirements.

Each pixel of the APS array, shown in Figure 4.2 (a), has a size of $100\mu\text{m} \times 100\mu\text{m}$ and consists of the deep photodiode (Djunc), a complementary NMOS/PMOS reset gate (MRST, MNRST), a source follower (MACT, MBIAS) and two select transistors (MRSEL, MCSEL). All pixels are reset globally and the inverter output and the bias transistor (MBIAS) are shared with all pixels. Having a CMOS transmission gate allows the pixel to take advantage of a wider dynamic range by pulling the pixel up to the 5V supply voltage (VDD) during reset. This eliminates the problem obtained with using only an NMOS reset transistor whereby the reset level varies with light intensity [Tian 2001]. The layout of a single pixel is shown in Figure 4.2(b). Circuitry other than the photodetector is light shielded using one of the two available metal layers but this is not shown in Figure 4.2(b).



(a) Active pixel circuit with global reset and global bias transistor



(b) Layout of a single pixel

Figure 4.2 Active pixel circuit and its layout

The backend of the active pixel sensor, shown in Figure 4.3, acts as a current-to-voltage converter and buffer for the photodiode node. It consists of the source follower active transistor (MACT), a row select transistor (MRSEL), a column select transistor (MCSEL) and a bias transistor (MBIAS) which is shared by all pixels. To optimize the design of the backend, simulations were carried out to find the optimum W/L of the transistors and biasing voltage. For the simulations, the gates of the row-column access transistors were held at VDD i.e. 5V. To verify and confirm the simulation results, first order analysis of the circuit was carried out. The results of the simulation and the circuit analysis are presented in the following subsections.

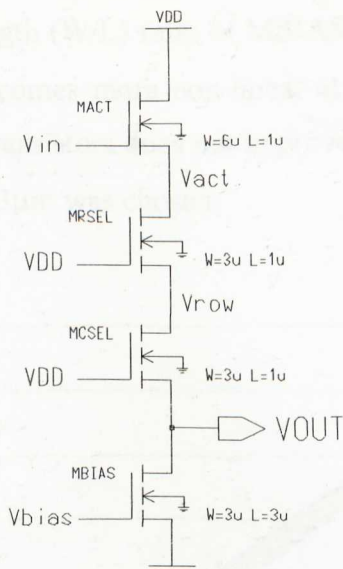


Figure 4.3 Backend of the active pixel sensor

Optimisation of MBIAS:

As V_{bias} is increased the dynamic range decreases and the response becomes more non-linear particularly for lower input voltages (Figure 4.4). An optimum bias voltage of 1V was selected to keep the biasing transistor operating above the threshold voltage of 0.76V [Alcatel Microelectronics 1999a] but sufficiently small so as to maintain a wide linear operating range. For further optimisation after fabrication, the applied voltage of the bias transistor of the active pixel array can also be applied externally.

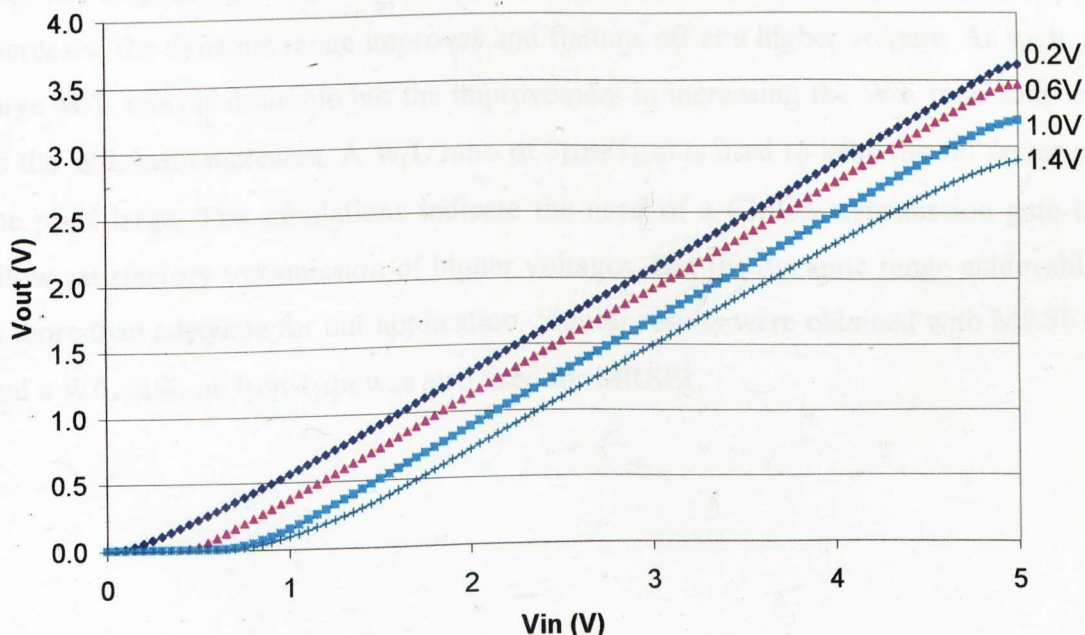


Figure 4.4 V_{OUT} against V_{in} for different values of V_{bias} ($W/L = 3\mu m/3\mu m$)

Similarly, as the width-to-length (W/L) ratio of MBIAS increases, the dynamic range reduces and the response becomes more non-linear at lower input voltages (Figure 4.5). Also using larger size transistors does not improve the linearity of the response. As such, a W/L ratio of $3\mu\text{m}/3\mu\text{m}$ was chosen.

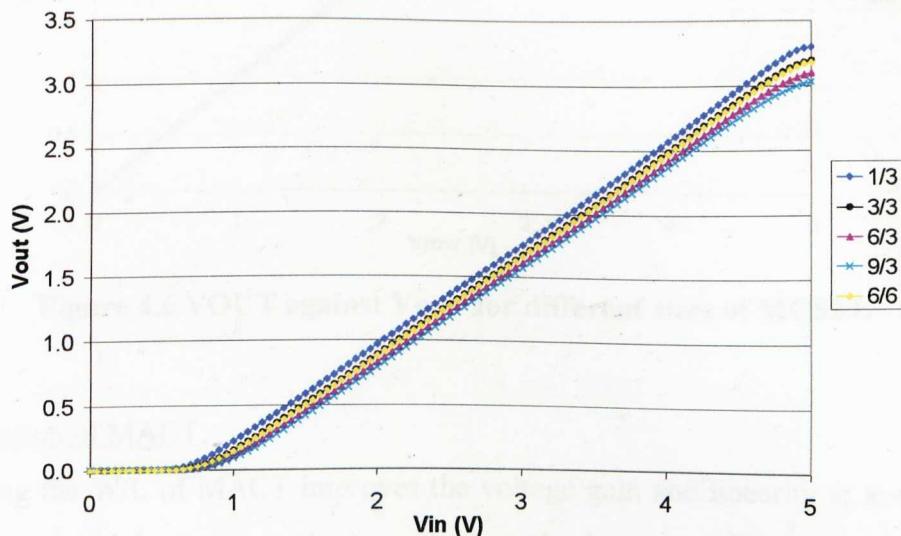


Figure 4.5 VOUT against Vin for different sizes of the MBIAS transistor

Optimisation of MCSEL and MRSEL:

The effect of sweeping the voltage at V_{row} on the output V_{OUT} is shown in Figure 4.6. The response is fairly linear at voltages below 3V. As the W/L ratio of MCSEL increases, the dynamic range improves and flattens off at a higher voltage. As such, a large W/L ratio is desirable but the improvement in increasing the W/L ratio reduces as the W/L ratio increases. A W/L ratio of $3\mu\text{m}/1\mu\text{m}$ is used to keep the fill factor of the pixel large. The simulations indicate the need of a CMOS transmission gate to allow satisfactory transmission of higher voltages. But the dynamic range achievable is more than adequate for our application. Similar results were obtained with MRSEL and a W/L ratio of $3\mu\text{m}/1\mu\text{m}$ was also used for MRSEL.

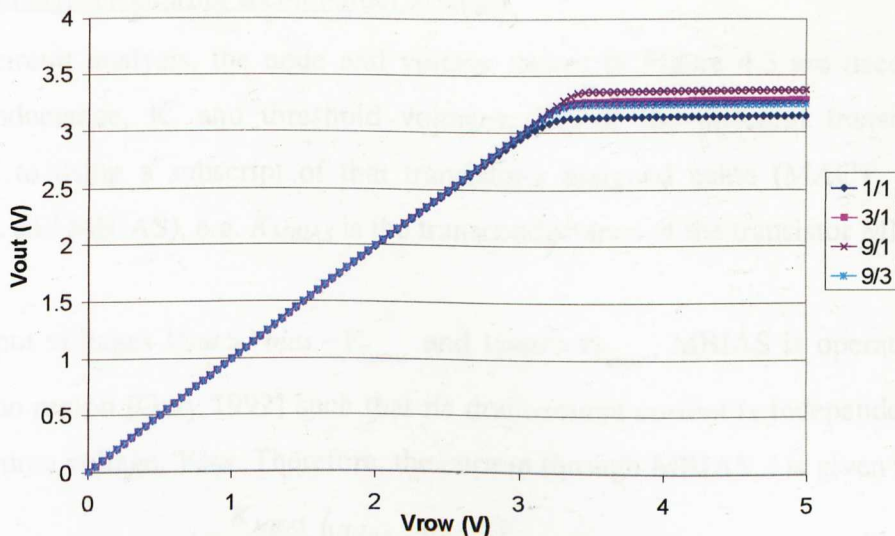


Figure 4.6 VOUT against Vrow for different sizes of MACT

Optimisation of MACT:

Increasing the W/L of MACT improves the voltage gain and linearity at lower input voltages but at higher voltages the linearity degrades because of the poor transmission of high voltages by the NMOS row-column select transistors (Figure 4.7). The improvement in increasing the W/L ratio reduces as the ratio increases. Using a larger size transistor with the same W/L ratio does not give better results. A W/L ratio of $6\mu\text{m}/1\mu\text{m}$ was chosen for MACT. The selected W/L ratios of the transistors are shown in Figure 4.3.

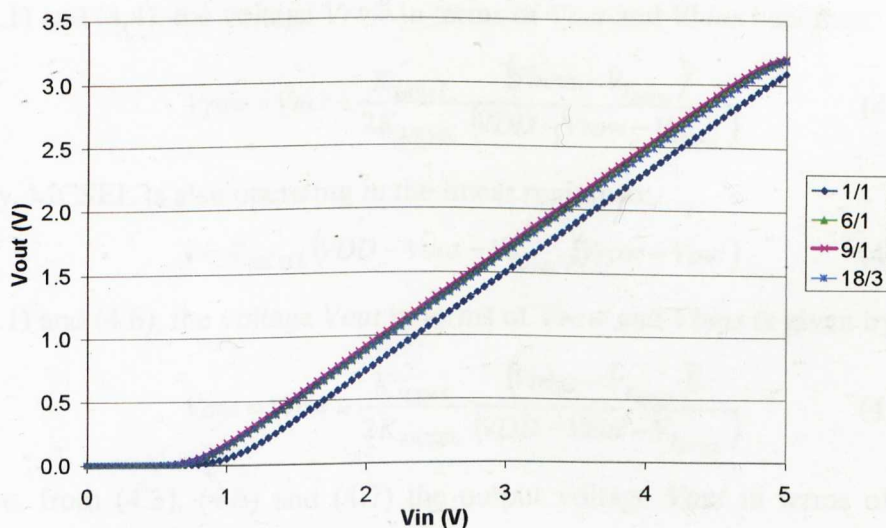


Figure 4.7 VOUT against Vin for different sizes of MACT

Circuit analysis (ignoring second order effects):

In the circuit analysis, the node and voltage names in Figure 4.3 are used and the transconductance, K , and threshold voltages, V_T , of the different transistors are referred to using a subscript of that transistor's assigned name (MACT, MRSEL, MCSEL and MBIAS), e.g. K_{MBIAS} is the transconductance of the transistor MBIAS.

For output voltages $V_{out} > V_{bias} - V_{T_{MBIAS}}$ and $V_{bias} > V_{T_{MBIAS}}$, MBIAS is operated in the saturation region [Gray 1992] such that its drain-source current is independent of the drain-source voltage, V_{out} . Therefore, the current through MBIAS, i is given by:

$$i = \frac{K_{MBIAS}}{2} (V_{bias} - V_{T_{MBIAS}})^2 \quad (4.1)$$

MACT is also in saturation and because the voltage at V_{out} is buffered, the current through MACT is equal to the current through MBIAS, MRSEL and MCSEL. Therefore:

$$i = \frac{K_{MACT}}{2} (V_{in} - V_{act} - V_{T_{MACT}})^2 \quad (4.2)$$

From (4.1) and (4.2), the voltage at V_{act} in terms of V_{in} and V_{bias} is obtained as follows:

$$V_{act} = V_{in} - V_{T_{MACT}} - \sqrt{\frac{K_{MBIAS}}{K_{MACT}}} (V_{bias} - V_{T_{MBIAS}}) \quad (4.3)$$

MRSEL is operating in the linear region. Hence:

$$i \approx K_{MRSEL} (V_{DD} - V_{row} - V_{T_{MRSEL}}) (V_{act} - V_{row}) \quad (4.4)$$

From (4.1) and (4.4), the voltage V_{row} in terms of V_{act} and V_{bias} becomes:

$$V_{row} = V_{act} - \frac{K_{MBIAS}}{2K_{MRSEL}} \frac{(V_{bias} - V_{T_{MBIAS}})^2}{(V_{DD} - V_{row} - V_{T_{MRSEL}})} \quad (4.5)$$

Similarly, MCSEL is also operating in the linear region so:

$$i \approx K_{MCSEL} (V_{DD} - V_{out} - V_{T_{MCSEL}}) (V_{row} - V_{out}) \quad (4.6)$$

From (4.1) and (4.6), the voltage V_{out} in terms of V_{row} and V_{bias} is given by:

$$V_{out} = V_{row} - \frac{K_{MBIAS}}{2K_{MCSEL}} \frac{(V_{bias} - V_{T_{MBIAS}})^2}{(V_{DD} - V_{out} - V_{T_{MCSEL}})} \quad (4.7)$$

Therefore, from (4.3), (4.5) and (4.7) the output voltage V_{out} in terms of V_{in} and V_{bias} is now obtained as follows:

$$V_{out} = V_{in} - V_{T_{MACT}} - \sqrt{\frac{K_{M_{BIAS}}}{K_{MACT}} (V_{bias} - V_{T_{M_{BIAS}}})} - \frac{K_{M_{BIAS}}}{2K_{MRSEL}} \frac{(V_{bias} - V_{T_{M_{BIAS}}})^2}{(V_{DD} - V_{row} - V_{T_{MRSEL}})} - \frac{K_{M_{BIAS}}}{2K_{MCSEL}} \frac{(V_{bias} - V_{T_{M_{BIAS}}})^2}{(V_{DD} - V_{out} - V_{T_{MCSEL}})} \quad (4.8)$$

This shows that as V_{bias} or the W/L ratio of MBIAS is increased, the dynamic range is reduced as indicated by Figure 4.4. It also confirms that increasing the W/L ratios of MACT, MRSEL, MCSEL improves the voltage gain and linearity of the transfer function. Furthermore, the last term shows that as the output voltage drops, the non-linearity increases. This agrees with the simulation results presented earlier. The non-linearity for V_{in} values close to the supply voltage is due to the use of NMOS select transistors which do not pass high voltages very well.

4.2.2 ANALOGUE-TO-DIGITAL CONVERSION

The pixel is operated in the charge integration mode. Each pixel is globally reset to 5V for $8\mu s$ (with a 32MHz clock) after which the pixel photodiode is allowed to discharge through its own photocurrent as shown in Figure 4.8. The discharge rate is proportional to the photocurrent of that pixel, which in turn is proportional to its incident light level. The discharge curve is approximately linear for voltages above 1V. This is because the photodiode capacitance varies inversely with the square root of the diode voltage. Also, as the diode and the pixel output voltage drops, the bias transistor starts to operate in the linear region and is no longer independent of the output voltage.

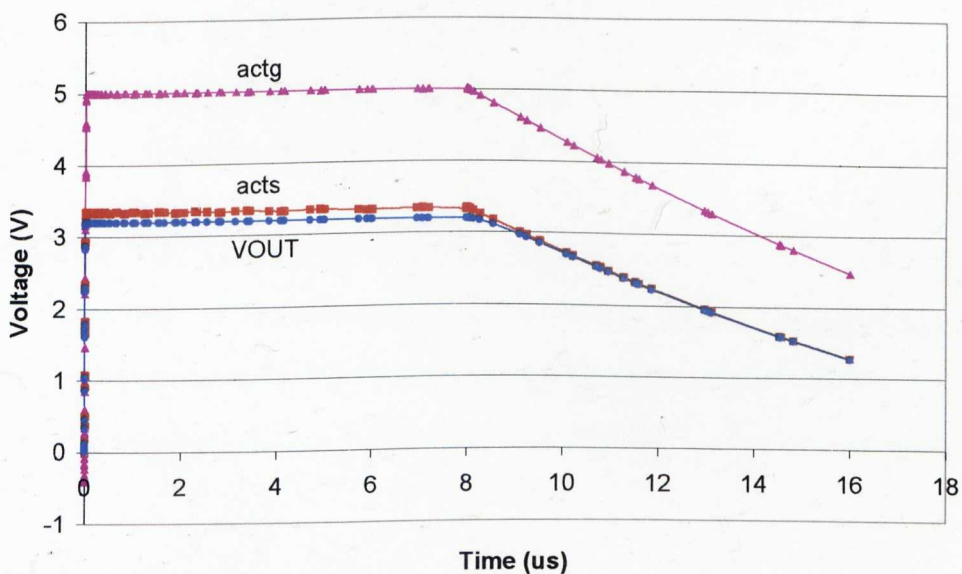


Figure 4.8 Discharge curve of the active pixel circuit used

When digitising the pixel light level from a discharge curve, either the final output voltage is digitised using standard ADC techniques such as successive approximation, dual slope or flash techniques or the time taken for the discharge to occur is measured. The second method was preferred due to the simple circuitry required and by integrating the ADC into the discharge curve of the pixel using a counter technique the digital output is immediately available after the discharge period. The discharge time is measured by starting an 8-bit counter when it passes through an initial voltage level and stopping it when it passes a second lower voltage level. These voltage levels are set by 2 sets of reference voltage generators.

Each set of reference voltage generators can generate a voltage between 1V and 3.75V with a step size of 0.25V and each connects to a comparator input. Six levels of one of these reference generators are shown in Figure 4.9. The reference voltage generator consists of a set of voltage dividers implemented using active resistors (transistors in saturation) and transmission gates for selecting the desired voltage. These transmission gates are used to select the switching points of the comparators during analogue-to-digital conversion. This style of reference generation was used because of its efficient use of space [Allen 1987]. Increasing the number of devices in an active resistor can reduce the total required area by reducing the voltage across the transistors and changing the ratio required for the desired output⁴⁹. For testability, external reference voltages can also be applied in place of the internal ones.

⁴⁹ If the transistors are identical in size the volt drop across the transistors are equal.

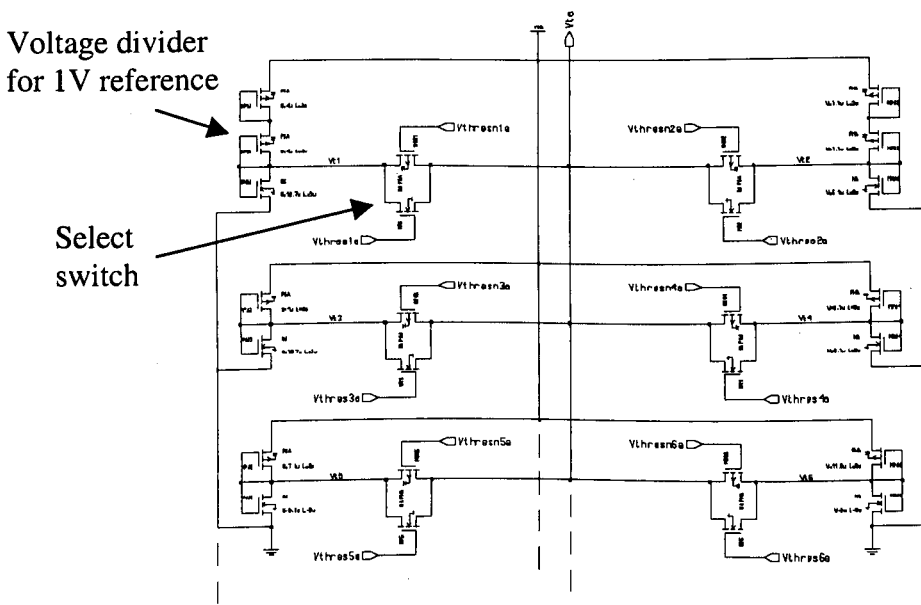


Figure 4.9 Reference voltage generator

When reset is fired, one reference voltage (V_{ref1}) will be set at 3.75V while the second reference voltage (V_{ref2}) will be set at 3.5V. Then the reference voltages are decreased until both reference voltages are below the pixel-reset level. This is determined by when the comparators switch over. In order to cope with a wide range of light levels, three modes of operation have been designed. In the first mode, the counter is started when the reset is removed and stopped when the discharge curve passes the 1st reference voltage. In the second mode, the counter is started when the discharge curve passes the 1st reference voltage and stopped when it passes the 2nd reference voltage. This has the advantage that if the reset level varies from pixel to pixel, the reading will be independent of this offset. In the third mode, a 2-cycle approach is used. In the 1st cycle, a reading is obtained as in mode 2. In the 2nd cycle the value of V_{ref2} is adjusted such that a larger dynamic range is obtained thereby increasing the resolution for higher light levels.

The simulated comparator delay is $0.38\mu\text{s}$ and the pixel reset period is sufficiently long for the setting of the threshold levels. The default discharge time of $8\mu\text{s}$ for the system was chosen for detecting photocurrents of the order of 10nA to $10\mu\text{A}$ in Mode 1 and 2. The minimum and maximum detectable current, I_{min} and I_{max} respectively, is given by:

$$I_{min} = \frac{C\Delta V}{\Delta T_{max}} = \frac{0.5\text{pF}(0.25\text{V})}{8\mu\text{s}} = 15.6\text{nA} \quad (4.9)$$

$$I_{\max} = \frac{C\Delta V}{\Delta T_{\min}} = \frac{0.5 \text{ pF}(0.25\text{V})}{8\mu\text{s}/256} = 4\mu\text{A} \quad (4.10)$$

where C is the capacitance of the photodiode at 2V (taken as the average voltage on the discharge curve) and is obtained from the dark C-V measurements (see Section 2.3.2), ΔV is the volt drop over which the time is measured, and ΔT_{\max} and ΔT_{\min} are the maximum and minimum measurable time step respectively. Mode 3 extends this dynamic range by 8 times. Increasing ΔV to increase the minimum (and maximum) detectable current is in effect a form of thresholding and can be used to remove background signals. With this ADC technique, the sensitivity at low light intensity is limited by the spacing of the reference voltage levels while the sensitivity at high light intensity is limited by the speed of the counter.

Two of the main sources of noise in an active pixel sensor are shot noise and reset noise. Shot noise due to integration is given by [Droste 2002]:

$$\begin{aligned} \overline{V}_n &= \sqrt{\frac{q}{C^2} \int_0^T (I_{ph} + i_{dc}) dt} \\ &= \sqrt{\frac{q\Delta V}{C}} \end{aligned} \quad (4.11)$$

where $q = 1.6 \times 10^{-19} \text{C}$ is the electron charge, C is the photodiode capacitance, I_{ph} is the pixel photocurrent, i_{dc} is the pixel dark current, T is the integration period and ΔV is the signal volt drop over the integration period. For a signal volt drop of 0.25V, the shot noise voltage is:

$$\overline{V}_n = \sqrt{\frac{1.6 \times 10^{-19}}{0.5 \text{ pF}} \times 0.25} = 0.283 \text{ mV} \quad (4.12)$$

The reset voltage (equation (1.16)) on the other hand works out to be:

$$V_{\text{reset}} = \sqrt{\frac{kT}{C}} = \sqrt{\frac{1.38 \times 10^{-23} \times 300}{0.5 \times 10^{-12}}} = 91 \mu\text{V} \quad (4.13)$$

For a signal volt drop of 0.25V, this represents a total SNR of more than 58.9dB.

A programmable discharge clock is used such that for lower light levels, a slower clock is used to measure the slower discharge curve over a longer period, thus allowing optimum resolution to be maintained for different intensity levels. This is yet another advantage of using CMOS processing which allows on-chip programmability

to be incorporated. Four programmable discharge clock frequencies are possible, which are internally selected by two mode registers. The states of which can be read via the on-chip RS232 receiver.

4.2.3 APPLICABILITY OF DESIGN

Given that the sensor was designed to operate with photocurrents of 10nA to 10μA in Mode 1 and 2, the typical incident light levels for several applications are examined in order to determine which applications are feasible. In astronomy, the number of photons reaching the Earth's surface in a given area in unit time is given by the astronomical brightness, B_{astro} and this is defined for a visible passband by:

$$B_{astro} = (4 \times 10^6) 10^{-m_v / 2.5} \text{ photons / cm}^2 - \text{sec} \quad (4.14)$$

where m_v is the visual magnitude of the observed star and a visual magnitude of 14 is roughly the brightness of a sunlit geosynchronous satellite [Tyson 1998]. For $m_v = 14$, $B_{astro} = 10 \text{ photons/cm}^2\text{s}$. For an 8m telescope, the photon flux will be $5 \times 10^6 \text{ photons/s}$ and assuming the Fried's coherence length, r_o , is 15cm i.e. the size of a single subaperture, the photon flux per subaperture will be 2500 photons/s or about 0.8fW of incident power. For a responsivity of approximately 0.3A/W, the photocurrent generated will be about 0.2fA! In order to detect this level of intensity, the integration time needs to increase by 10^8 times i.e. 3s.

According to [Nirmaier 2003], specifications for ophthalmic applications state that a safely applicable laser power results in 200pW per spot or about 60pA of photocurrent, which is about 250 times below the measurement limit. With faster clocks, shorter integration times, moving to smaller lower capacitance pixels and better readout techniques, it is possible with future designs to improve the sensitivity to this level and enable its use in ophthalmic applications.

In free-space optical (FSO) communications, the requirement is for the system to be eye safe as per IEC 60825 Class 1 or Class 1M specifications (up to 2mW/cm^2). In addition, FSO systems operate at longer wavelengths, either 780-850nm and near the 1550nm band in order to be completely eye safe. For example, at approximately 1550nm, the regulatory agencies allow approximately 100 times higher power for

"eye safe" lasers. This is because at this wavelength, the aqueous fluid of the eye absorbs much more of the energy of the beam, preventing it from travelling to the retina and inflicting damage.

With further work it is foreseen that the design will be applicable to the fields of free space optical communications, microscopy and ophthalmology, but it is unlikely to be used in astronomy, where currently CMOS imagers are not as sensitive as CCDs, and the cooling and long integration times required negates the benefits of using the low-cost highly-integrated CMOS option.

4.3 DIGITAL BACKEND

The digital backend consists of the centroid processor which was previously verified by the hardware emulation system, counters and control for the ADC, the required clock dividers, serial transmitters for sending output data off-chip and serial receivers for receiving control signals. The block diagram of the digital backend is shown in Figure 4.10.

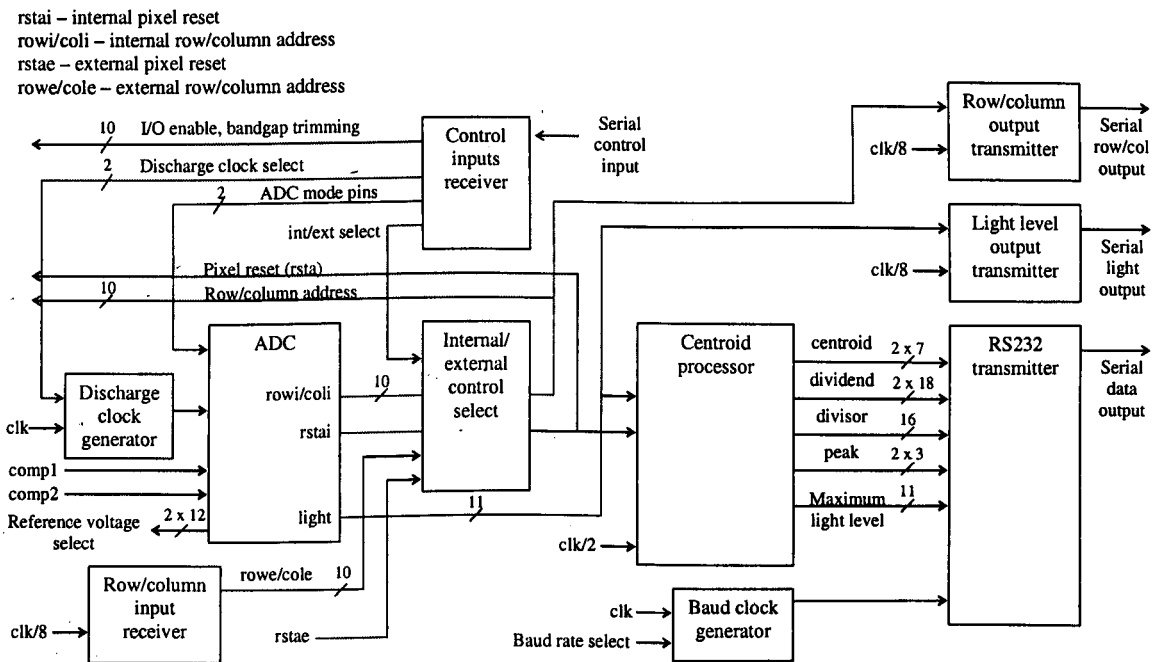


Figure 4.10 Block diagram of digital backend of the ASIC

4.3.1 ASIC CENTROID PROCESSOR

The centroid processor computes the first order moment of the light intensity (see Section 1.4.2) and was successfully demonstrated in the hardware emulation system. The x and y-centroids are calculated in parallel with separate processors. In addition to finding the centroid, the position of the pixel with the highest intensity is also found. In the ASIC processor, the number of bits is extended to 11 bits due to the extended dynamic range. This extends the number of bits required for the dividend and divisor to 18 bits and 16 bits respectively. The only change required to the centroid processor in the hardware emulation system is larger storage while the operations performed remain the same.

The centroid multiplications and summations are carried out as the pixels are read while the division process of the centroid calculation is performed once every frame, after all the pixel values have been read and the final dividend and divisor values obtained. After the division process is completed, the centroid values are latched out. A frame lasts 26 pixel periods. A pixel period consists of a reset period that lasts 256 clock cycles (8 μ s) and a discharge period that lasts up to 256 clock cycles depending on the light level. For the full 32MHz clock, this gives a frame rate of between 2.4kHz (32MHz / (512*26)) and 4.8kHz (32MHz / (256*26)) depending on the incident light level of every pixel. The use of 26 pixel periods per frame was out of convenience and it is possible to reduce this to 25 pixel periods by using a faster clock to latch out the data and clear the registers, thereby allowing a new frame to start immediately after 25 pixel periods.

In the hardware emulation system, the division process was performed asynchronously due to the style of coding and required about 125 μ s (5 conversion cycles of the ADC) to settle to its result⁵⁰. The VHDL division process was modified in the centroid ASIC for synchronous operation, leading not only to an improvement in speed but also a reduction in gate count despite the increase in number of bits. The division process now needed just 15 clock cycles of a 16MHz clock to complete. The

⁵⁰ Although the division process in the hardware emulation system is slower, the update rate of the centroid values was also 26 ADC conversion cycles or pixel access cycles.

clock frequency used was half the external 32MHz clock frequency as the critical path delay was found to be 31.37ns in the division process (Appendix A4.3). At 16MHz, 15 clock cycles takes just 0.9375 μ s to complete. With a pixel period of 8 μ s to 16 μ s, there is plenty of latency to be utilised allowing the division processor to be shared among several centroid processors and this will be discussed further in Section 4.3.3. For testability, the reset of the pixels and the row-column addressing can also be controlled externally.

4.3.2 DATA TRANSMISSION

Serial transmission was preferred over parallel outputs in order to minimise the number of pinouts required. Centroid data and intermediate values of centroid processing (such as the x-dividend, y-dividend, divisor, individual pixel light level and peak pixel position) are transmitted in RS232 format⁵¹ with, as before, one start bit (logic '0'), 8 data bits, no parity bits and 1 stop bit (logic '1') at a selectable baud rate of 115200, 76800, 57600, 38400, 19200, 9600, 4800 or 2400 bits/s selected by three external control pins. The default startup rate is 115200 bits/s and this is the only RS232 baud rate capable of transmitting the centroid data in real time. The minimum possible data rate needed to transmit the centroid data in real time = $(32 \text{ MHz} / (256 \times 26)) \times 20 \text{ bits} = 96154 \text{ bits/s}$.

The row/column address and digitised light levels are also transmitted off-chip so external processing of the centroids can be performed. Unlike the centroid values, these need to be transmitted at the pixel rate. So the light level data was transmitted in a modified serial format with one start bit (logic '0'), 11 data bits (10 for the row/column address), no parity bits and 1 stop bit (logic '1') and the minimum possible data rate required in this case is $(32 \text{ MHz} / 256) \times 13 \text{ bits} = 1.625 \text{ Mbits/s}$. Hence, a baud clock of 4 MHz was used to transmit the row/column address and

⁵¹ A single transmitter is used to transmit the centroid and intermediate values by controlling three external pins to select the data type to send. Some of the intermediate values have more than two bytes (or just one byte for the pixel position of maximum intensity) to transmit and as before, the MSBs are coded to distinguish between the packets transmitted. clkd10, clkd20 and clkd30 signals are used to ensure that a complete set of bytes are sent before newly available or updated data are transmitted.

digitised light levels. This is a non-standard format intended for use with an FPGA or DSP to perform the external centroid processing.

Two serial receivers were implemented. One for obtaining control signals for selecting the ADC mode of operation, the discharge clock rate, the accessibility of certain input/output pins, and the enabling of external processor control. The second receiver is used to receive external row/column address inputs, and like the row/column address transmitter, receives data at 4 Mbits/s to enable real-time operation with an FPGA say. The control signal receiver, on the other hand, does not need to operate very fast and is designed to detect a serial RS232 format input of one start bit, 8 data bits and 1 stop bit at a baud rate of 19200 bits/s. Two bytes of input data (or 16 bits) provide the required control signals. The receivers synchronise the baud rate clock at every start bit using the 32 MHz clock and starts looking for the next start bit at the centre of a stop bit⁵².

4.3.3 LAYOUT AND TEST BOARD

The ASIC centroid processor represents a single tilt sensor in an integrated Shack Hartmann wavefront sensor and the core-limited layout of this tilt sensor is shown Figure 4.11. The chip contained 7200 logic gates and has a size of $4500\mu\text{m} \times 4000\mu\text{m}$. The core area is $3800\mu\text{m} \times 3400\mu\text{m}$ and the photodiode array took up $530\mu\text{m} \times 600\mu\text{m}$. The digital circuitry which makes a significant portion of the chip will scale favourably as technology scales and as we move towards a triple metal process with improved routing capabilities. The main centroid processor block, for example, takes up an area of $1800\mu\text{m} \times 1900\mu\text{m}$ or 3.42mm^2 and the division circuitry takes up about 44% of this area or 1.5mm^2 . However, since the division is performed only once every frame and requires only 15 cycles of the 16MHz clock or $0.9375\mu\text{s}$ to complete, there is significant latency in the use of the divider. Assuming that it is necessary to obtain all the centroid/tilt outputs within one pixel cycle ($256 \times 1/32\text{MHz} = 8\mu\text{s}$) in order that the tilts correctly represent the same wavefront, about 8 centroid processors could

⁵² The received data is latched out at this point as well. Certain control signals were not received serially but as parallel inputs, such as the baud rate clock select, the transmit data type, the transmit trigger signal and the global reset signal.

time-share one divider without any loss in data rate. However, such a large number of centroid processors to one divider would lead to significant routing and crosstalk issues. Also circuitry like the clock dividers, receiver and control logic can be shared alongside the divider while the transmitters are only needed for the final output signals. So overall there is a 50% fill factor in the circuitry that can be shared. Integrating 4 tilt sensors for every divider such as illustrated in Figure 4.12 will give a fill factor of about 20% which is a reasonable solution.

For testing of the centroid processor, the fabricated ASIC is incorporated into a test board with power supply protection, input switches, serial port connection and a 32MHz crystal oscillator, as shown in Figure 4.13. The pushbuttons control reset signals and to ensure default values are used on startup, an active-high power-on reset circuit with an RC time constant of 2.7ms is used as shown in Figure 4.14. The schematic and PCB layout for this test board is shown in Appendix A4.2.

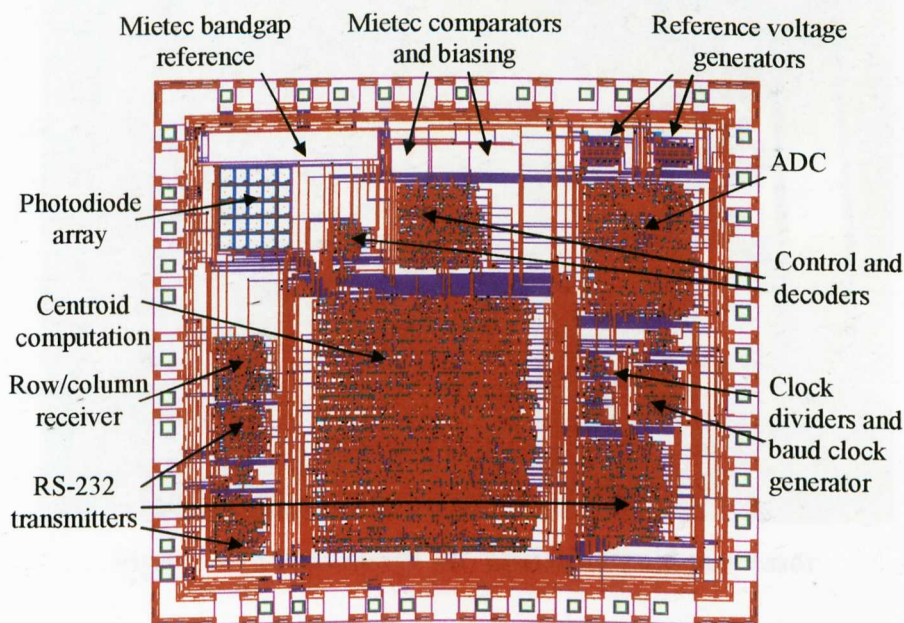


Figure 4.11 Layout of ASIC optical centroid processor (a single tilt sensor)

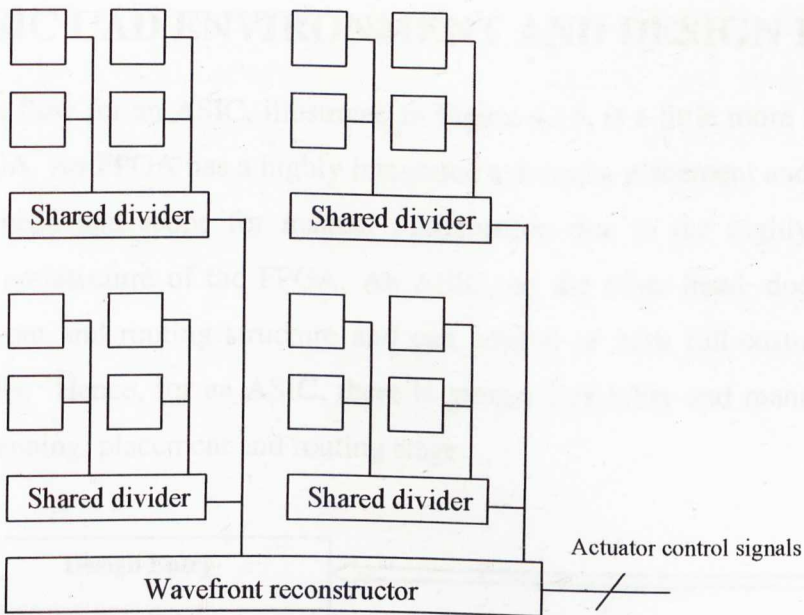


Figure 4.12 Scalability of design to a complete wavefront sensing and reconstruction system

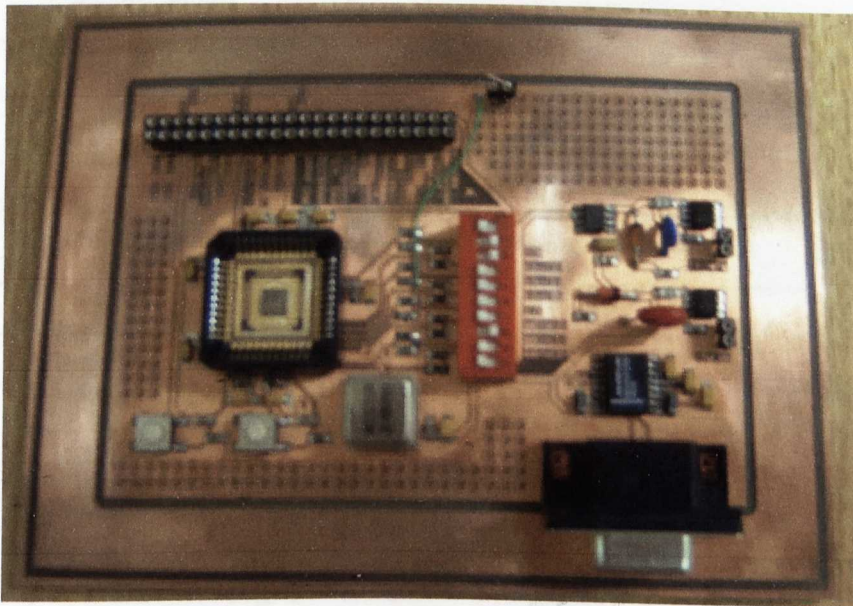


Figure 4.13 Test board for ASIC centroid processor

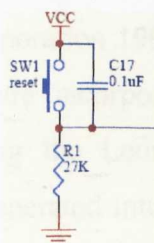


Figure 4.14 Power-on reset used in centroid ASIC test board

4.4 ASIC CAD ENVIRONMENT AND DESIGN FLOW

The design flow for an ASIC, illustrated in Figure 4.15, is a little more involved than for an FPGA. An FPGA has a highly integrated synthesis, placement and routing flow with less need and scope for manual optimisation due to the highly regular and structured architecture of the FPGA. An ASIC, on the other hand, does not have a regular layout and routing structure and can consist of both full-custom and semi-custom cells. Hence, for an ASIC, there is greater flexibility and manual control in the floorplanning, placement and routing stage.

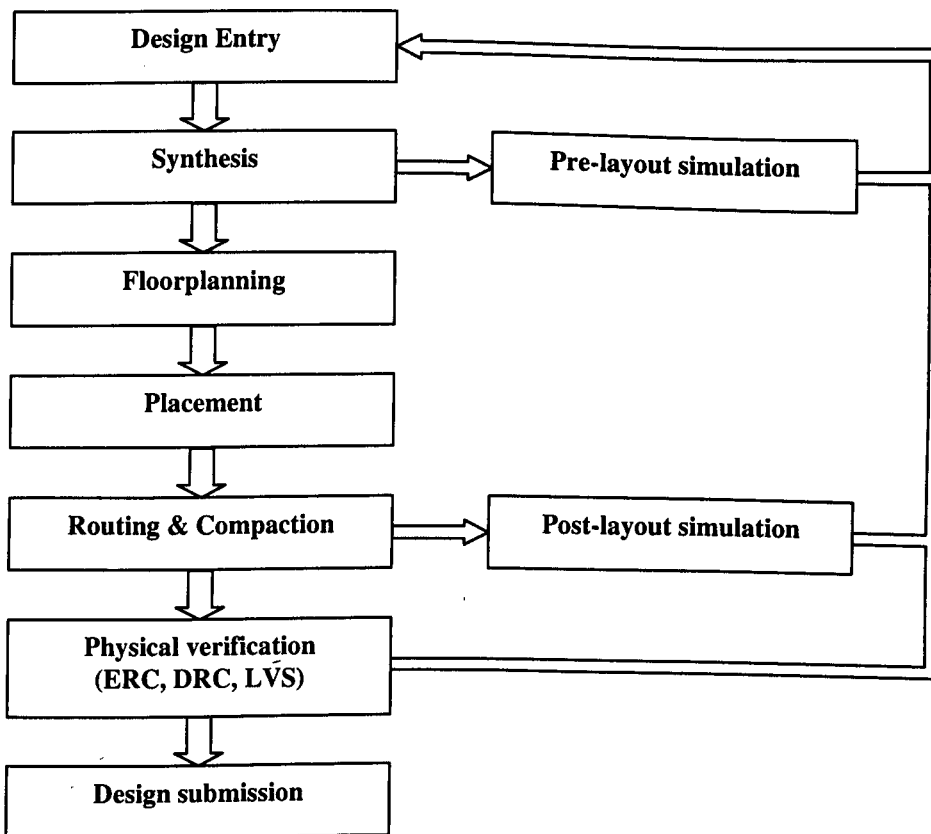


Figure 4.15 ASIC design flow

The CAD environment used for the design of the ASIC was the Mentor Graphics C4 suite of tools [Mentor Graphics Corporation 1998]. Design entry is made via Design Architect v8.6_4. VHDL macros are incorporated into the design schematic by synthesizing the VHDL code using the Leonardo Spectrum synthesis tool and converting the EDIF netlist (.edf) generated into Mentor's proprietary netlist format, Electronic Design Data Model (EDDM), prior to schematic and symbol generation. The simulation tools used were Mentor's Accusim v8.6_3, QuickSim II v8.6_4 and

QuickPath v8.5_1 for analogue simulation, digital simulation and static timing analysis respectively. Mixed analogue and digital simulation was not available so the analogue and digital parts of the design had to be simulated separately. Layout was carried out using Mentor's IC Station v8.7_3 family of tools, which include ICgraph for full custom editing, ICplan for floorplanning⁵³, ICblock and ICroute for automatic placement and routing of standard cells and blocks, and ICcompact for layout compaction. The target technology for the design was the same as in the first two chips (Section 2.2.1 and 2.2.2), that is the Alcatel Microelectronics (Mietec) 0.7 μ m CMOS process.

Due to the added nature of full custom editing, the physical verification stage is an important part of any ASIC design and can be divided into three tasks: Electrical Rule Checks (ERC), Design Rule Checks (DRC) and Layout Versus Schematic (LVS) checks. The ERC checks for simple circuit violations such as short circuits, open circuits and correct power and ground connections. A DRC is to ensure the design meets the layout rules set by the foundry such as minimum spacing and minimum lengths of any given mask [Alcatel Microelectronics 1999b], while the LVS checks if the layout structure matches with the original schematic design. In IC Station the verification toolset is called ICverify and consists of the ICtrace, ICrules and ICextract tools for ERC, DRC and LVS checks respectively. ICextract is also the extraction tool used for extracting parasitic resistance and capacitance from the layout for backannotation⁵⁴ into the design schematic. Post layout simulations were not carried out as the tools for backannotation were not properly setup but LVS and DRC were performed. LVS had to be performed separately on each individual block before the top level check, that is, a hierarchical check had to be done. Finally when fully verified, the design was exported in GDSII format and submitted to the chosen foundry, IMEC in Belgium, via the Europractice IC Service. There further ERC and

⁵³ Floorplanning is the process of placing groups of circuits on a die, and analyzing the effect of that placement in terms of design performance and routability. Floorplanning also helps to monitor the actual size of a design.

⁵⁴ Backannotation is the process of extracting timing information from the layout back into the design schematic for post-layout simulation.

DRC checks were carried out using their Cadence Dracula set of tools before fabrication can commence.

4.4.1 DESIGN AND LAYOUT ISSUES

The design is a mixed semi-custom and full-custom design with full custom cells (photodiode array, voltage reference generators), Mietec analogue cells (comparators, biasing) and digital standard cells (VHDL macros). With such a mixed design, several design and layout issues need to be addressed [Baker 1997, Johns 1997], such as the combining, partitioning and routing of the design in the Mentor Environment, clock tree buffering of digital circuitry and power consumption. These issues are discussed in the following subsections.

4.4.1.1 Design Entry and Full Custom Editing

In the Mentor environment, specific properties are used on the components at different levels of hierarchy such as the PWR_NET and GND_NET properties to specify which global power supply nets to use, SUB_COMP and CELL_COMP to specify the lowest level of hierarchy, MODEL, INSTPAR and INST properties for analogue simulation, PLACE properties to specify placement on the layout, etc. When designing full custom cells, these properties had to be incorporated and the layout instance pin names must match the schematic symbol pin names for correct LVS. In addition, equivalent circuit models of the full custom cells are included in subcircuit schematics for simulation purposes. At the top level schematic (Appendix A4.1), the full custom components, Mietec components and VHDL macros are connected and external input/output and power supply pads are attached. Each VHDL macro corresponds to a single layout block where the standard cells within each individual block are autoplaced and autorouted.

4.4.1.2 Design Partitioning

The design was partitioned into blocks, which aids the separation of the analogue and digital sections of the design and avoids coupling of high frequency digital switching onto sensitive analogue lines. Clocks and transmitters were placed further away from the analogue portion of the circuit but generated clocks were kept close enough to the ADC and centroid processing blocks to prevent excessive clock skews and delays. Keeping the ADC and centroid computation blocks together also helps to minimise

wire lengths and delays. Once the individual blocks were placed, routing was carried out.

4.4.1.3 Power Supply Routing and Protection

Power supply nets and critical nets were routed manually before autorouting of all nets was performed. By using a 'keep pre-routes' option, the router uses the widths of the manual routing but moves it as it sees fit. When the autorouter has finished, further manual changes are made where necessary. Mietec specifies the maximum current for the metal layers, the contacts and the vias under different conditions and this works out to be approximately $1\text{mA}/\mu\text{m}$, $0.30\text{mA}/\text{contact}$ and $0.35\text{mA}/\text{via}$ respectively [Alcatel Microelectronics 1999]. To ensure low resistance and inductance, a conservative approach was taken and power supply nets were widened to at least $25\mu\text{m}$, expanding as power supply buses join, and clock lines were widened to $10\mu\text{m}$. In the design, four sets of power supplies (VDD1-4, VSS1-4) are used for the digital I/O, one pair for the digital core cells (VDD, VSS) and three pairs (VDDA1-3, VSSA1-3) for the analogue cells, namely the photodiode array, the comparators (with biasing) and the voltage reference generators⁵⁵ respectively. The Mietec P_SUPPROT power supply protection structure was placed between each power and ground pair for better immunity against electrostatic discharge (ESD) [Alcatel Microelectronics 1995a].

4.4.1.4 Mietec Analogue Cells

The comparators used were the analog CFCMP1 cell provided for in Mietec's MTC22500 Analog Library [Alcatel Microelectronics 1995b]. In order to use this cell the cell had to be biased according to the biasing strategy stated by Mietec. This includes a bandgap voltage reference (1.20V), a master bias generator for providing a 'sinking' current source and a slave bias cell to convert the reference current to bias voltages for the analogue cells. The slave bias cells and the analogue cells that they bias are placed close together and in the same row of cells, since voltage drops in the supply lines can cause errors in the bias current.

⁵⁵ VDDA3-VSSA3 was also used to power the analogue I/Os.

4.4.1.5 Clock Tree Planning

Clock tree planning is necessary as it is impossible to use an ideal clock to drive all the latches due to issues of routability, circuit drive strength and clock latency and skew. In the FPGA design environment, global buffers and routing channels are used to distribute critical nets such as input clock signals on the device with minimum skew. In the ASIC environment, no clock-tree synthesis tool was available so clock tree synthesis had to be done manually by calculating the load of every input and output signal of each block and inserting clock drivers or buffers (inverters) into the design in a tree-like manner.

Several buffers are available within the Mietec standard cell library [Alcatel Microelectronics 1998] as shown in Table 4.1. CBTSA, CBTSB and CBTSC are positive enabled tristate buffers with low drive, 2X drive and 3X drive respectively, while CIA, CIB and CIC are inverters with low drive, 2X drive and 3X drive respectively. CIB was chosen as a compromise between high output drive, low propagation delay, minimum area and low power consumption. A maximum load of 20SL was allowed on a net before buffers were inserted to keep the maximum load of each arm of the clock tree to 20SL.

Cell name	Area (μm^2)	Power ($\mu\text{W}/\text{MHz}$)	Propagation delay for 32SL (ns)		Input capacitance (SL)*	Output drive (SL)*
			t PLH	t PHL		
CBTSA	432	2.1360	3.69	2.74	2.0	19
CBTSB	649	3.7925	1.85	1.52	4.0	42
CBTSC	1189	9.6915	0.97	0.83	9.7	108
CIA	216	0.8878	2.29	1.65	2.0	31
CIB	324	1.2632	1.12	0.96	3.3	63
CIC	757	4.5378	0.16	0.16	11.6	172

* SL is defined in the Mietec library as a standard load (SL) of 0.029pF

Table 4.1 Buffer types in the Mietec MTC23000 standard cell library

Only part of the clock tree buffering circuit can be seen in the top level schematic (Appendix 4.1) with additional clock tree buffering extending into the lower level schematics of certain blocks. Only input and output signals of the blocks were

analysed by hand. Internal signals within each block were not dealt with manually but static timing analysis highlights nets with long propagation delays and heavy loads.

4.4.1.6 Power Consumption

Another issue in circuit design is the power consumption of the circuit. A power analysis tool was not available and manual calculations of the chip's power consumption were carried out. Based on simulations and the specified supply currents of the analogue cells, the majority of the power consumption is expected to come from the digital circuits which also makes up a bigger proportion of the circuitry. The power dissipation of a cell in the Mietec MTC23000 CMOS Standard Cell library is given by:

Total Power Dissipation per cell =

$$[\text{POW (value from datasheet)} + (\text{VDD}^2 * \text{Cext})] * \text{FREQ}$$

where Cext = load capacitance in pF for each cell

POW = power in $\mu\text{W}/\text{MHz}$

FREQ = switching frequency in MHz

The first term represents power dissipation due to the internal circuitry of the cell while the second term is the power consumption due to the charging and discharging of the load capacitance of the cell. Except for heavily loaded lines, the 2nd term can be ignored and for the purpose of these power consumption calculations was not taken into account. Calculation of the total power consumption of the digital circuitry is shown in Appendix A4.3 where the power dissipation of the components in each block are summed and multiplied by their operating frequency. A conservative approach was used where the highest frequency which any part of the block runs at was taken as the operating frequency of the whole block. For example, though the calculation of the dividend and divisor runs at the pixel access rate of 2.4kHz to 4.8kHz, the switching frequency of the block was taken as the frequency at which the division process occurs i.e. at $32\text{MHz}/2 = 16\text{MHz}$. Furthermore, not every part of the circuit is constantly being operated and the transmitters, the receiver and the control logic only operate when activated. A summary of the results as well as the number of gates and the critical path delay of each block is shown in Table 4.2.

Block name	POW(μ W/MHz)	Frequency, f (MHz)	Power Consumption (mW) at frequency, f (MHz)	No. of Gates	Critical path delay (ns)
clkdiv1	72.9231	32	2.3335392	47	6.46
CTR5by5a	5131.3544	16	82.1016704	3641	31.37
CTRa2d	1554.306	32	49.737792	1028	17.45
ctrlrcvr	1029.2836	32	32.9370752	625	13.2
div2	12.632	32	0.404224	7	2.37
div8	58.0278	32	1.8568896	35	5.02
div10	60.6702	0.1152	0.006989207	37	4.86
div20	79.6368	0.1152	0.009174159	48	5.51
div30	74.3362	0.1152	0.00856353	46	5.69
divbaud	307.2977	32	9.8335264	191	13.76
opctrlsyn	150.0007	32	4.8000224	92	2.05
RCrcvr	554.6075	32	17.74744	335	7.56
txall	869.5475	0.1152	0.100171872	590	11.02
txlightout	399.6852	4	1.5987408	256	8.84
txrowcol	381.9001	4	1.5276004	239	8.5
Total			205.0034192	7217	

Table 4.2 Power consumption, gate count and critical path delay of digital blocks

These figures are purely estimates at best but showed that the chip could cope with power supply demands. Certainly, compared to the hefty power consumption of a PC (typically 100W) there is significant power savings to be made by going to a single chip solution. When the chip had been fabricated and tested on a circuit board, the supply current drawn from a 5V supply was 30mA (rms) when computing and transmitting centroid data. Hence, the power drawn was less than estimated, which is reasonable as a conservative approach of over-estimating the frequency of operation was taken.

4.5 RESULTS OF CENTROID ASIC

A $3\mu\text{m}$ diameter beam from a 633nm HeNe laser was scanned across the array at a speed of $2000\mu\text{m}/\text{sec}$. Centroid values were computed by the processor and serially transmitted in real time to a PC running Linux with a baud rate of 115,200 bits/s. Figure 4.16 shows the experimental setup used to scan the beam across the ASIC and obtain the centroids. A neutral density filter (NDF) was used to bring the power incident on the ASIC down to about $1\mu\text{W}$.

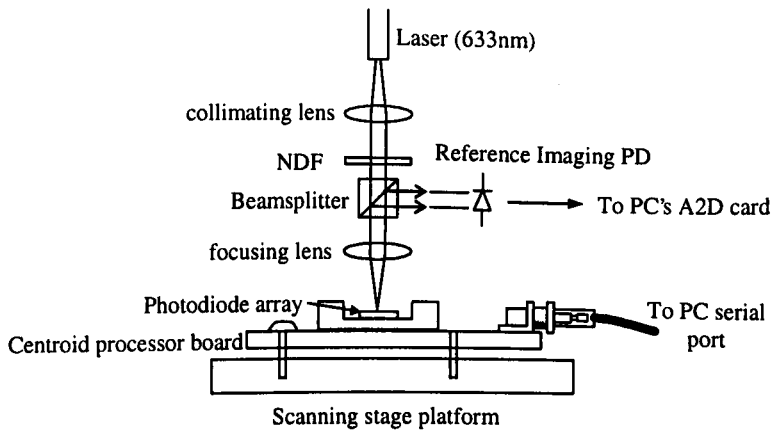


Figure 4.16 Experimental scanning setup for testing the centroid ASIC

Figures 4.17 shows the optical image obtained from the reference photodiode as the beam is scanned across the array. Figures 4.18 and 4.19 show grey scale maps of the x and y-centroid values successfully recorded at each position on the array, with the device operating under mode 1 of the digitisation procedure. The dark regions correspond to smaller centroid coordinates whilst lighter regions correspond to larger centroid coordinates. As expected, as we scan in the x-direction, the x-centroid values increases while the y-centroid values remain constant and vice versa. Since the laser beam size is less than the size of one pixel, a stepped appearance can be seen as the beam moves across the array passing from one discrete detector to another. Figures 4.19 and 4.20 show the averaged x and y-centroid values plotted as a function of pixel position. Also shown in Figures 4.19 and 4.20 are the error bars for the measurement of the position across the array.

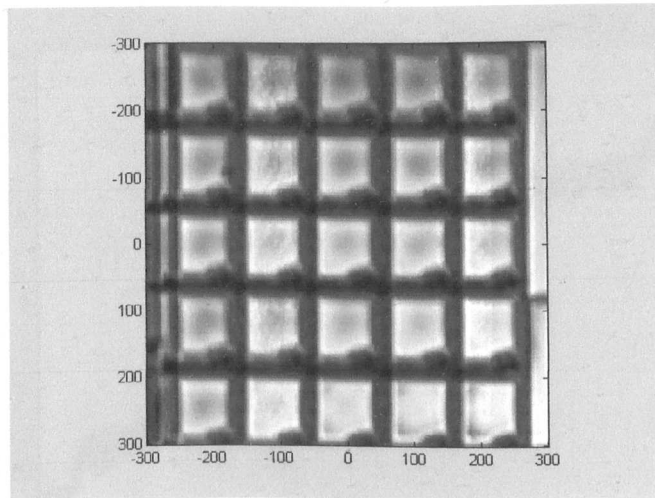


Figure 4.17 Optical image of scan

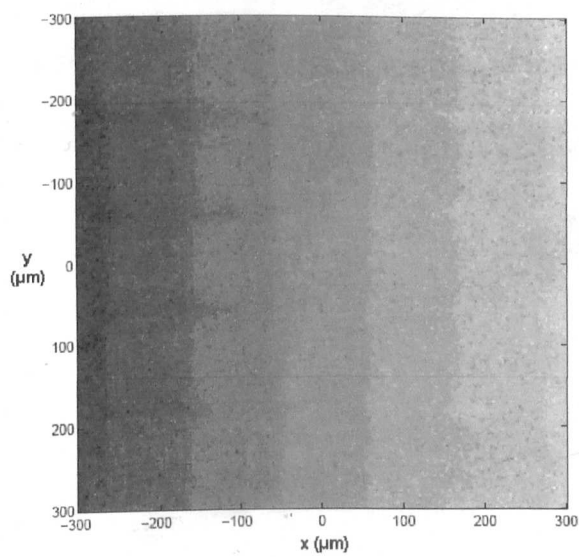


Figure 4.18 Image map of x-centroids

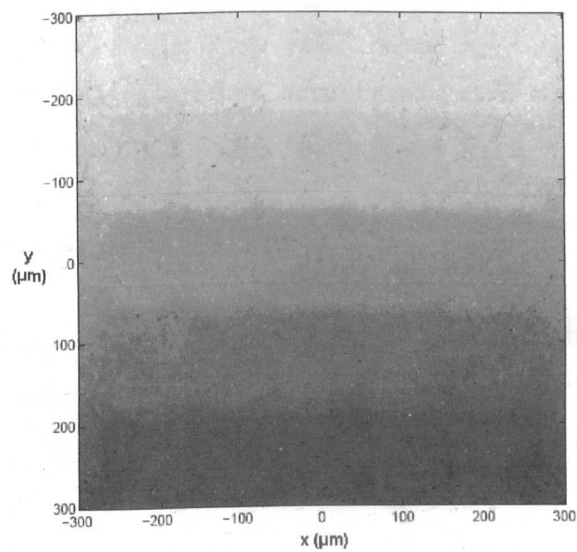


Figure 4.19 Image map of y-centroids

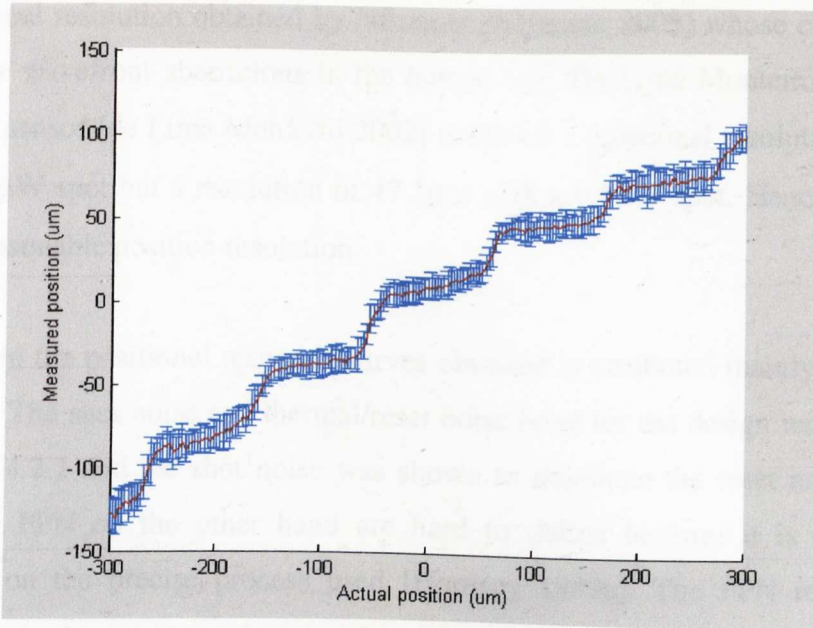


Figure 4.20 Measured vs. actual position of x-centroids

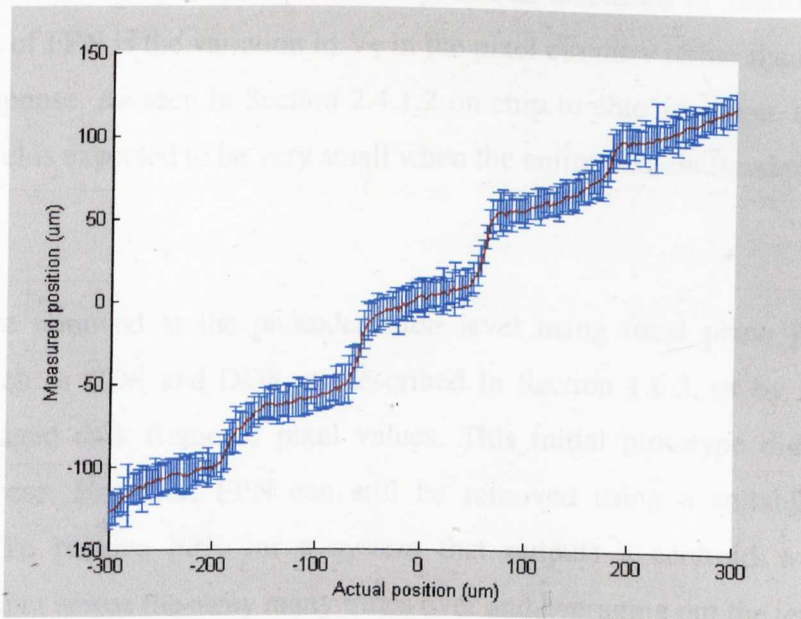


Figure 4.21 Measured vs. actual position of y-centroids

4.5.1 POSITION RESOLUTION

The positional resolution is obtained by finding the average of the maximum deviation (error bar) in the position response across the array in a single scan and this is found to be $19.5\mu\text{m}$ in the y (2.1LSB) and $14.9\mu\text{m}$ in the x (1.8LSB)⁵⁶. This is comparable to

⁵⁶ The difference in resolution in the x and y is accentuated by the fact that the array size is $530\mu\text{m}$ in the x and $600\mu\text{m}$ in the y.

the positional resolution obtained by Nirmaier [Nirmaier 2003] whose chip was used to measure wavefront aberrations in the human eye. De Lima Monteiro's integrated wavefront sensor [de Lima Monteiro 2002] achieved a positional resolution of $1.4\mu\text{m}$ with a $7.0\mu\text{W}$ spot but a resolution of $47.1\mu\text{m}$ with a $0.2\mu\text{W}$ spot. Hence, the design showed reasonable position resolution.

The noise in the positional response curves obtained is attributed mainly to FPN and shot noise. The shot noise and thermal/reset noise level for the design were discussed in Section 4.2.2 and the shot noise was shown to dominate the reset noise. Typical figures for FPN on the other hand are hard to define because it is significantly dependent on the precise process used [Hornsey 1999c]. The FPN reported with CMOS image sensors in a submicron process was approximately 2-3% of saturation for raw data without FPN removal circuitry, and as discussed in Section 1.5.3.5, the main cause of FPN is the variation in V_T in the pixel circuitry rather than the variation in photoresponse. As seen in Section 2.4.1.2 on chip to chip variation, the FPN from pixel to pixel is expected to be very small when the entire pixel is flooded i.e. for large spot sizes.

FPN can be removed at the photodetection level using focal plane FPN removal circuitry such as CDS and DDS, as described in Section 1.6.3, or by subtracting a stored averaged dark frame of pixel values. This initial prototype did not include either of these. However, FPN can still be removed using a suitable calibration technique. To remove FPN for a system that outputs a centroid would require scanning a spot across the array many times over and averaging out the temporal noise in the 2D images of centroids acquired to obtain a single 2D image map consisting of the positional response, the dark FPN component and the PRNU component. Assuming the PRNU component is negligible and this image map is applicable to all other intensity levels, curve fitting can be used to fit ideal or average curves through the positional response curves obtained, as illustrated in Figure 4.22, and the difference between the curves can be stored in memory and subsequently subtracted from future centroid readings. However, this is only accurate for a particular spot size for which the design is tailored for. Note that for larger spot sizes the effect of the FPN is expected to decrease.

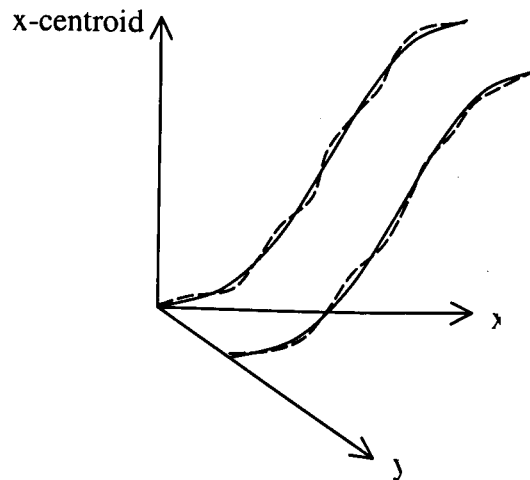


Figure 4.22 Curve fitting of calibration curves to obtain FPN values from time-averaged positional response curves

In addition, it is possible to increase the resolution by reducing the size of the pixels but this, however, reduces the maximum detectable tilt and hence, the maximum measurable aberration magnitude. Also, the resolution was inherently limited by the number of bits in the centroid representation (7 bits) and to increase this requires very little overhead. That is, just an additional shift-and-subtract cycle in the divider is required for each additional bit in the result, without the need for additional storage for the dividend and divisor.

Under conditions of low signal-to-noise ratio (SNR) it is possible to improve the accuracy of the centroiding by removing the background signal through thresholding. With a digital centroid processor this is easily achieved by subtracting a programmable, even adaptive, offset to the digitised input or by setting all pixel values below a certain threshold to zero before the centre of gravity is found. It is also possible to apply windowing around the pixel of maximum intensity to improve the SNR further.

4.5.2 SPEED

The design achieved a frame rate of more than 2.4kHz which when scaled to an array of centroid processors or tilt sensors in parallel will achieve a frame rate which is independent of the number of tilt sensors employed, allowing fast real time adaptive optical systems to be built.

The speed of the design is limited by the frame readout and digitisation technique and not the centroid computation. In the initial design, 26 pixel periods were required per frame but this can be reduced to 25 pixel periods by using a faster clock to latch out the data and clear the registers, thereby allowing a new frame to start immediately after 25 pixel periods.

Although the system is able to remove the data bottleneck in present systems, it is possible for the design to go even faster as currently with the ADC technique used, only 1 pixel is digitised per discharge cycle and 25 separate discharge cycles were required before a frame was readout, as illustrated in Figure 4.23 (a). By measuring the time to discharge to a particular voltage level, different pixels with different discharge rates (due to different incident light intensities) complete the ADC conversion at different times making it difficult to sequentially measure every pixel during one cycle. It would be possible to reduce the current system to a single integration period by comparing every pixel during each count as shown in Figure 4.23 (b), but this would require a very fast clock (800MHz for the default 8 μ s range) and an equally fast comparator. This may be feasible for long integration times but is not considered a suitable alternative. Instead a design which requires only one discharge cycle but a separate conventional ADC [Hoeschele 1994] that does not incorporate the discharge curve into the digitisation process is proposed. This is illustrated in Figure 4.23 (c) and Figure 4.24. By starting the integration period of the pixels at different times and using a fixed integration period, pixel values can be readout and digitised sequentially. Variable integration time is achieved by controlling the position at which integration is started and stopped. The integration time is given by the number of pixel access times between these points. With this technique, the frame delay is reduced to a maximum of 25 pixel access times for a 5 x 5 array, which is significantly faster than the current design.

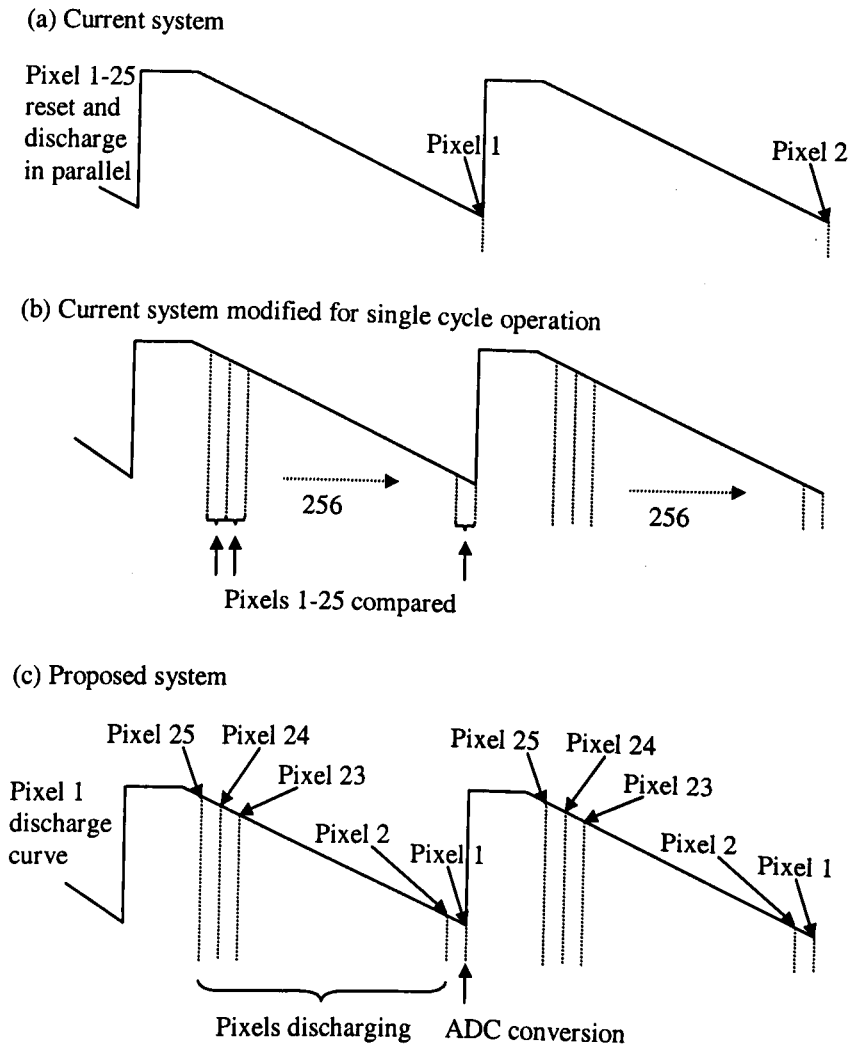


Figure 4.23 Pixel discharge and access with (a) the current system, (b) current system modified for single discharge cycle operation and (c) a proposed sequential digitisation structure

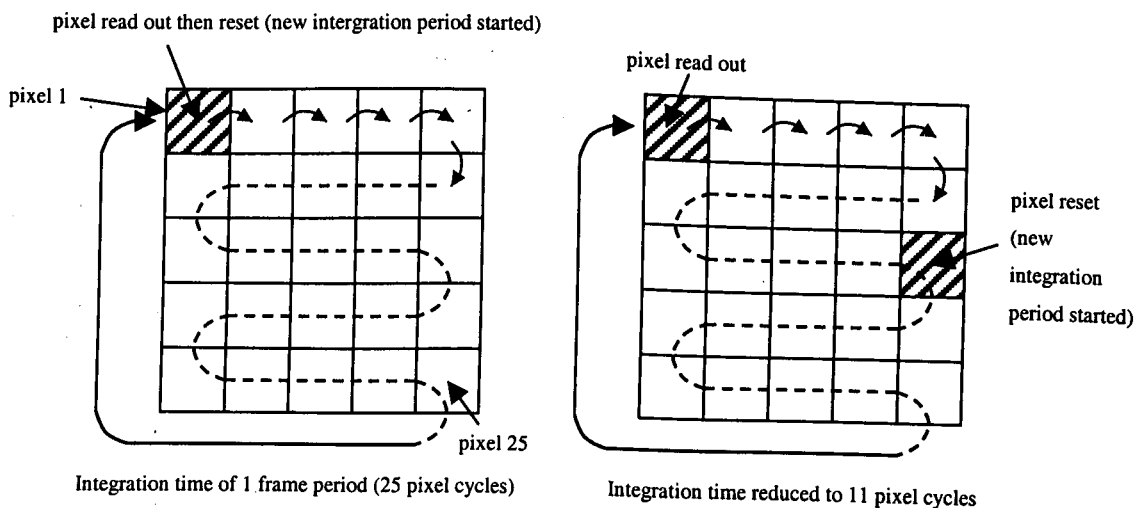


Figure 4.24 Alternative pixel access and digitisation structure

4.5.3 DYNAMIC RANGE

The spatial dynamic range of the centroid outputs as shown in Figures 4.20 and 4.21 was limited leading to limited positional sensitivity⁵⁷. A limited spatial dynamic range translates to a limited tilt measurement range. There are several reasons for the reduced spatial dynamic range. Stray light in the test system will lead to a large background signal on all pixels, shifting the centroid values towards the centre. Secondly, as a global reset was used, all the pixels discharge simultaneously whether or not they are read and during this discharge the photodiode node is floating and the pixel current can diffuse to neighbouring pixels. This crosstalk will lead to a larger background reading in all the signals, once again shifting the centroid output towards the centre. Also, as simulation results may vary from actual values, the voltage drop in mode 1 may be significantly smaller than that designed for, leading to a limited dynamic range in this mode. In order to have a minimum of 0.25V volt drop in mode 1, the 2nd reference voltage level should be used instead of the 1st reference voltage level. Finally, as a consequence of the digitising technique used where the discharge time is measured for a given light level, a 1/x compression of the input photocurrent is achieved leading to a high light intensity dynamic range [Forchheimer 1994] but also a smaller signal to background ratio, as expressed by:

$$\Delta T = T_{\max} - \frac{C\Delta V}{I_{ph}} \quad (4.14)$$

where ΔT is the measured discharge time, T_{\max} is the maximum discharge time for a given volt drop ΔV , photodiode capacitance C and photocurrent I_{ph} . For the case of the current system $C = 0.5\text{pF}$ at 2V and $\Delta V = 0.25\text{V}$, a digital output count, x , is obtained as follows (and shown in Figure 4.25):

$$x = 255 - \frac{4\mu\text{A}}{I_{ph}} \quad (4.15)$$

⁵⁷ Poor positional sensitivity does not imply poor positional resolution and the accurate positional resolution of spots could still be obtained by careful calibration of the response curves.

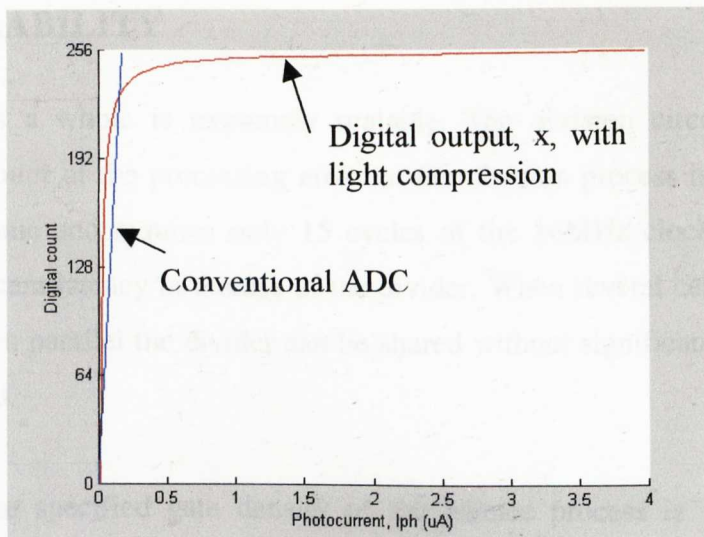


Figure 4.25 Conversion of photocurrent to digital output by measuring discharge time showing large dynamic range compared to conventional ADC techniques of measuring voltage drop

The issue of limited spatial dynamic range can be addressed in various manners. In order to remove any background signals, a thresholding technique can be used, or alternatively, an initial frame of dark (or background) readings is stored and subtracted from subsequent readings of each pixel. To resolve the problem of pixel crosstalk, the pixels need to be reset individually [Yadid-Pecht 1997] such that when one pixel is discharging the other pixels remain under reset. Also, a suitably biased guard ring structure can be incorporated between pixels to mop up any crosstalk current. As for the compression effect of the current digitisation technique, this has little benefit for determining centroids but would be a useful feature in imaging where it is desired to capture both the bright and dark regions of an image. Hence, an alternate digitisation structure like that proposed in Section 4.5.2 and Figure 4.23 is preferred. Note that the proposed pixel access and digitisation structure does not allow crosstalk to be removed by having individual pixel reset. Instead a guard ring structure must be used with every pixel.

4.5.4 SCALABILITY

The system as a whole is extremely scalable. The division circuitry takes up a significant amount of the processing area but the division process is performed only once every frame and requires only 15 cycles of the 16MHz clock to complete so there is significant latency in the use of the divider. When several centroid processors are integrated in parallel the divider can be shared without significant increase in size or loss of speed.

In addition, the specified gate density of the Mietec process is 1250 gates/mm². Migration to smaller feature sizes will mean greater packing density. The austriamicrosystems (AMS) 0.35µm CMOS process, for example, has a gate density of 18k gates/mm². This is a 14 times reduction in size of the digital circuitry, making the integration of a large number of centroid processors for a complete wavefront sensing system feasible.

4.6 CHAPTER SUMMARY

A real time VLSI optical centroid processor was successfully designed and fabricated for integration into a proposed Shack-Hartmann wavefront sensor. The chip consists of an optimised 5 x 5 active pixel array and analogue-to-digital conversion circuitry integrated with the centroid processor previously demonstrated using a hardware emulation system. Centroid values can be obtained at a rate of 2.4 - 4.8 kHz with a position resolution of less than 20µm or 0.2 of a pixel, allowing real time performance of the adaptive optical system. By replacing the use of a CCD, a frame grabber and a PC with a dedicated on-chip centroid processor, significant savings in power, size and cost can also be achieved.

CHAPTER 5

WAVEFRONT RECONSTRUCTION

5.1 INTRODUCTION

Once an array of optical centroid processors has obtained the wavefront slopes, the next step in the process of an adaptive optics system is the reconstruction of the aberrated wavefront. The main aim of wavefront reconstruction is to generate the required actuator signals to deform a flexible mirror to compensate for the distortions in the wavefront. Hence, in order to understand the process of wavefront reconstruction, one needs to understand how wavefronts are described and how deformable mirrors are used to perform the correction before delving into the reconstruction techniques available. The following sections discuss the concepts of wavefront reconstruction and how this process can be incorporated into the design, which will enable the design of a complete, compact, fast and low-cost adaptive optical system.

5.2 WAVEFRONT DESCRIPTION

A wavefront can be described using a zonal approach or a modal approach [Tyson 1998]. In a zonal approach, the wavefront is expressed in terms of the phase over a small spatial area or zone and by combining all the zones within the aperture, a complete wavefront is described. If the number of zones approaches infinity, the wavefront is exactly represented. In the modal approach, the wavefront is expressed in terms of a weighted sum of spatial modes such as tip/tilt, defocus, etc. where each mode is defined over the entire aperture. For wavefronts with low spatial frequencies, the entire wavefront can be adequately represented by a few low-order modes whereas if high spatial frequencies are present a large number of terms are needed and a zonal approach may be preferable [Geary 1995]. This weighted sum of modes is expressed

as a suitable polynomial expansion and one such expansion is the sum of Zernike polynomials, Z_k of order k :

$$\phi(\rho, \theta) = \sum_k A_k Z_k(\rho, \theta) \quad (5.1)$$

where ρ , θ are polar coordinates and the coefficients A_i is a time varying parameter which is typically smaller for higher orders. Z_1 and Z_2 correspond to the tilt of the wavefront in the x and y -directions, Z_3 to defocus, Z_4 and Z_5 to astigmatism and so on [Noll 1976]. Zernike polynomials are a popular choice because the polynomials are defined over a unit circle similar to the circular aperture of a telescope making it straightforward to express such wavefronts in terms of Zernike polynomials. They can also take into account the effect of the annulus present in telescopes. The orthogonality of the polynomials over a unit circle is also useful for incorporating higher order terms that are independent of the lower order terms and Zernike polynomials also allow easy calculation of the wavefront variance or error.

5.3 DEFORMABLE MIRRORS

Deformable mirrors are used to produce the phase conjugate of the aberrated wavefront in order to produce a plane wave. There are different types of deformable mirrors used in adaptive optics and these include segmented mirrors, continuous faceplate mirrors and bimorph mirrors [Tyson 2000].

Segmented mirrors can be manufactured to tight tolerances and each segment acts independently so the control computer is simplified. However, they do not provide a smooth surface transition and the gap between segments can have an adverse effect on the optical beam because its regular pattern acts somewhat like a diffraction grating by imparting diffractive modes into the beam. In addition, segmented mirrors need more actuators than continuous faceplate mirrors. The continuous faceplate deformable mirror eliminates the gaps and the optical problems associated with segmented mirrors at the expense of more complicated control. The shape of the continuous deformable mirrors is described by its influence function which describes the influence of one actuator on the surrounding surface.

A bimorph mirror consists of two thin layers of material bonded together. The layers can be oppositely polarized piezoelectric wafers or a piezoelectric wafer bonded with an optical surface made from glass or silicon. An array of electrodes is deposited between the two wafers and when a voltage is applied to an electrode, one wafer expands relative to the other producing a curvature proportional to the voltage applied. For a given number of electrodes bimorph mirrors achieve the highest degree of turbulence compensation. Compared to other deformable mirror technologies such as membrane mirrors, bimorph mirror fabrication uses lower cost components and involves fewer and much simpler processes. Bimorph mirrors produce a curvature (which follows a Poisson equation) making it suitable for use with curvature wavefront sensors [Roddier 1998a] without the need of complex reconstruction circuitry but less suitable with other wavefront sensors which requires the Poisson equation to be solved. Also, the geometry of the actuators in bimorph mirrors is radial-circular which conveniently matches the circular telescope apertures with a central annulus. However, the number of modes or actuators remains limited.

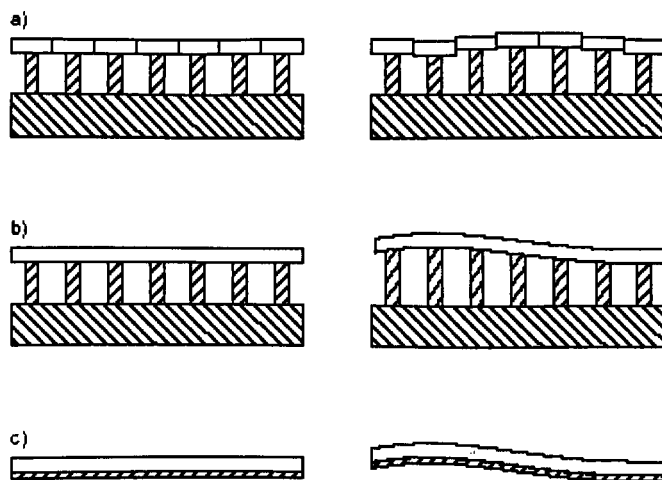


Figure 5.1 Comparison between (a) segmented mirrors, (b) continuous faceplate mirrors and (c) bimorph mirrors [Doelman 2000]

Micromachined deformable mirrors are a new class of deformable mirrors fabricated in silicon Micro-Electro-Mechanical Systems (MEMS) technology where small mirror elements are deflected by electrostatic forces. They offer potential for low cost and large number of actuators [Hatcher 2001, Mansell 2000]. But currently insufficient stroke and the small size of the elements remain a limitation.

Liquid crystal (LC) spatial light modulators (SLM) are another way to control the phase of light [Dayton 1997]. They operate based on the fact that an applied voltage will change the alignment of the long thin LC molecules and hence change its index of refraction. They can have a large number of elements but the phase shifts introduced by liquid crystals remain too small and wavelength-dependent.

5.4 WAVEFRONT RECONSTRUCTION

As described in Section 1.3.1, the Shack Hartmann wavefront sensor obtains local wavefront tilts from focal spot position displacements. The reconstruction of a wavefront from a set of local wavefront tilts involves solving a system of linear equations [Tyson 2000], which expressed in matrix algebra has the form:

$$\mathbf{s} = \mathbf{B} \mathbf{a} \quad (5.2)$$

where \mathbf{s} is a vector of the local wavefront tilts, \mathbf{a} is a vector of the required actuator commands (if modal reconstruction is used modes are obtained instead of phases and have to be converted into actuator commands) and \mathbf{B} is called the reconstruction matrix that contains information on how the tilts are related to the actuator signals.

The system is usually overdetermined with the system having more equations than unknowns such that \mathbf{s} has a higher dimension than \mathbf{a} . A least-squares approximation can be used to solve for vector \mathbf{a} and this is equivalent to calculating:

$$\mathbf{a} = [\mathbf{B}^T \mathbf{B}]^{-1} \mathbf{B}^T \mathbf{s} \quad (5.3)$$

where \mathbf{B}^T is the transpose and $[\mathbf{B}^T \mathbf{B}]^{-1} \mathbf{B}^T$ is the *pseudo-inverse* of the reconstruction matrix \mathbf{B} . The equation is valid on the condition that $\mathbf{B}^T \mathbf{B}$ is invertible (not singular). If this condition is not met, a method called singular value decomposition (SVD) is used. Otherwise, direct inversion methods like Gaussian elimination can be used [de Lima Monteiro 2002]. The pseudo-inverse matrix $[\mathbf{B}^T \mathbf{B}]^{-1} \mathbf{B}^T$ only needs to be calculated once for a given configuration (sensor-actuator geometry and reconstruction method). After which, the system only needs to compute the centroids, the associated tilts and evaluate a matrix multiplication for the actuator commands. There are two types of reconstruction methods that can be used, namely the *zonal* or *modal* reconstruction. The choice of which depends very much on the choice of deformable mirror and the choice of wavefront sensor.

5.4.1 MODAL APPROACH

In a modal reconstructor, the coefficients of a polynomial function for describing the wavefront, such as the Zernike polynomials, are obtained. From equation (5.1) the local tilts, S_{ix} and S_{iy} , can be related to the local derivatives of the phase and hence the local derivatives of the Zernike polynomials of subaperture i , as follows:

$$S_{ix} = \left. \frac{d\phi}{dx} \right|_i = \sum_k A_k \left. \frac{dZ_k}{dx} \right|_i \quad (5.4)$$

$$S_{iy} = \left. \frac{d\phi}{dy} \right|_i = \sum_k A_k \left. \frac{dZ_k}{dy} \right|_i \quad (5.5)$$

In matrix form of N subapertures and M modes, this can be written as:

$$\begin{bmatrix} S_{1x} \\ S_{1y} \\ S_{2x} \\ S_{2y} \\ \vdots \\ S_{Nx} \\ S_{Ny} \end{bmatrix} = \begin{bmatrix} \left. \frac{dZ_1}{dx} \right|_1 & \left. \frac{dZ_2}{dx} \right|_1 & \left. \frac{dZ_3}{dx} \right|_1 & \dots & \left. \frac{dZ_M}{dx} \right|_1 \\ \left. \frac{dZ_1}{dy} \right|_1 & \left. \frac{dZ_2}{dy} \right|_1 & \left. \frac{dZ_3}{dy} \right|_1 & \dots & \left. \frac{dZ_M}{dy} \right|_1 \\ \left. \frac{dZ_1}{dx} \right|_2 & \left. \frac{dZ_2}{dx} \right|_2 & \left. \frac{dZ_3}{dx} \right|_2 & \dots & \left. \frac{dZ_M}{dx} \right|_2 \\ \left. \frac{dZ_1}{dy} \right|_2 & \left. \frac{dZ_2}{dy} \right|_2 & \left. \frac{dZ_3}{dy} \right|_2 & \dots & \left. \frac{dZ_M}{dy} \right|_2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \left. \frac{dZ_1}{dx} \right|_N & \left. \frac{dZ_2}{dx} \right|_N & \left. \frac{dZ_3}{dx} \right|_N & \dots & \left. \frac{dZ_M}{dx} \right|_N \\ \left. \frac{dZ_1}{dy} \right|_N & \left. \frac{dZ_2}{dy} \right|_N & \left. \frac{dZ_3}{dy} \right|_N & \dots & \left. \frac{dZ_M}{dy} \right|_N \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_M \end{bmatrix} \quad (5.6)$$

By solving for vector A_k using equation (5.3), the coefficients of the Zernike polynomial are obtained. At the same time, the influence function $\varphi_l(x, y)$ of each actuator in a deformable mirror can also be expressed as a function of the Zernike polynomial as follows [Zhu 1999]:

$$\varphi_l(x, y) = \sum_{k=1}^M b_{kl} Z_k(x, y) \quad (5.7)$$

where b_{kl} is the coefficient corresponding to the k th Zernike polynomial due to the control signal of the l th channel of the mirror. Assuming that the total deflection of the mirror is a linear superposition of the deflections from all the control channels, the mirror surface deflection $\Delta\phi(x, y)$ can be expressed as:

$$\begin{aligned}
\Delta\phi(x, y) &= \sum_{l=1}^P c_l \phi_l(x, y) \\
&= \sum_{l=1}^P c_l \sum_{k=1}^M b_{kl} Z_k(x, y) \\
&= \sum_{k=1}^M \left(\sum_{l=1}^P c_l b_{kl} \right) Z_k(x, y) \quad (5.8)
\end{aligned}$$

where c_l is the control signal of the l th channel of the deformable mirror. Therefore, the Zernike coefficients obtained from solving equation (5.6) can be related to the control signals c_l as follows:

$$A_k = \sum_{l=1}^P c_l b_{kl} \quad (5.9)$$

where b_{kl} is experimentally determined and the equation is solved using equation (5.3) once again to obtain the required control signals c_l to perform the corrections.

5.4.2 ZONAL APPROACH

In a zonal reconstructor, the phase at regular grid points across the aperture is evaluated and several sensor-actuator geometries such as the Hudgin geometry and the Fried geometry shown in Figure 5.2 for a 3 x 3 actuator system. Here, a represents the actuator positions and S_i represents the slopes of subaperture i .

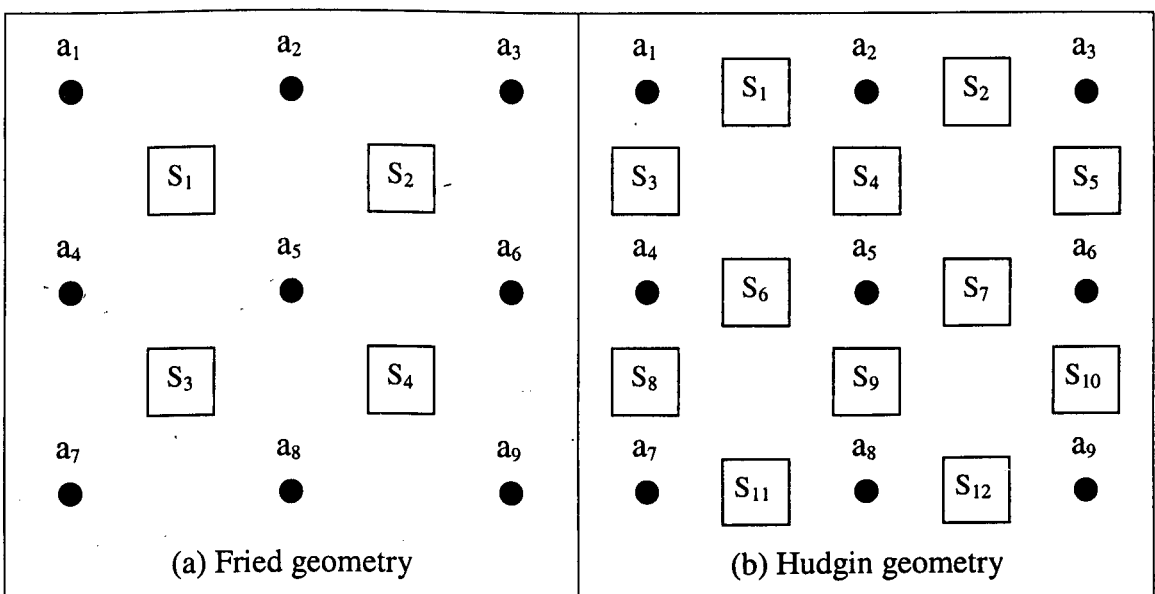


Figure 5.2 Wavefront sensor-actuator geometries

With the Hudgin geometry only one centroid per subaperture is used and $S_1, S_2, S_6, S_7, S_{11}$ and S_{12} represent slopes in the x-direction while S_3, S_4, S_5, S_8, S_9 and S_{10} represent slopes in the y-direction. For $N \times N$ actuators, a Hudgin geometry requires $2N(N-1)$ subapertures ($N(N-1)$ x-centroids and $N(N-1)$ y-centroids) while the Fried geometry requires $(N-1)^2$ subapertures so the Hudgin geometry requires more subapertures but less processing per subaperture.

From these configurations, the equations (equation (5.1)) that relate the wavefront sensor signals to the actuator commands can be developed. Besides the geometry and alignment of the sensor subapertures and the actuators, the type of mirror used for reconstruction determines how the tilt values are related to the required actuator signals and hence the reconstruction matrix \mathbf{B} [Tyson 2000]. In the case of a segmented mirror, the slope of a particular subaperture only depends on the influence of the neighbouring actuator signals and the reconstruction matrix \mathbf{B} is sparse. In the case of the continuous faceplate deformable mirror, the remaining elements in \mathbf{B} are not zero but dependent on the influence function of the mirror⁵⁸.

5.4.3 RECONSTRUCTION PROCEDURE

A Shack Hartmann wavefront sensor measures local wavefront slopes, which provide a zonal description of the aberrated wavefront and lends itself to zonal reconstruction procedures [Geary 1995]. To illustrate the architecture required for wavefront reconstruction from a set of Shack Hartmann wavefront tilts, the process of generating the reconstruction matrix and finding its pseudo-inverse for a chosen architecture is carried out, and the implementation of this structure is considered.

For this purpose, a segmented mirror system with 3×3 actuators and a Hudgin geometry (requiring 12 tilt sensors) as shown in Figure 5.2(b) is selected. In the matrix form this is expressed as:

⁵⁸ For example, B_{11} of the reconstruction matrix represents the influence of actuator 1 on the 1st slope in the x-direction, B_{12} represents the influence of actuator 2 on the 1st slope in the x-direction and so on.

$$s = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \\ S_{10} \\ S_{11} \\ S_{12} \\ 0 \end{bmatrix}, B = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \end{bmatrix} \quad (5.10)$$

The row of 1's at the bottom of the reconstruction matrix is used to force the average surface of the wavefront to a specific shape or value and to keep the reconstruction matrix from being singular⁵⁹. The pseudo-inverse matrix of **B** is found as follows:

$$[B]^{-1} = [B^T B]^{-1} B^T =$$

$$\begin{bmatrix} 0.4444 & 0.1806 & 0.4444 & 0.1528 & 0.0694 & 0.1528 & 0.0972 & 0.1806 & 0.0972 & 0.0556 & 0.0694 & 0.0556 & 0.1111 \\ -0.2639 & 0.2639 & 0.1528 & 0.3611 & 0.1528 & -0.0556 & 0.0556 & 0.0972 & 0.1389 & 0.0972 & -0.0139 & 0.0139 & 0.1111 \\ -0.1806 & -0.4444 & 0.0694 & 0.1528 & 0.4444 & -0.0972 & -0.1528 & 0.0556 & 0.0972 & 0.1806 & -0.0556 & -0.0694 & 0.1111 \\ 0.1528 & 0.0972 & -0.2639 & -0.0556 & -0.0139 & 0.3611 & 0.1389 & 0.2639 & 0.0556 & 0.0139 & 0.1528 & 0.0972 & 0.1111 \\ -0.0556 & 0.0556 & -0.0556 & -0.2222 & -0.0556 & -0.2222 & 0.2222 & 0.0556 & 0.2222 & 0.0556 & -0.0556 & 0.0556 & 0.1111 \\ -0.0972 & -0.1528 & -0.0139 & -0.0556 & -0.2639 & -0.1389 & -0.3611 & 0.0139 & 0.0556 & 0.2639 & -0.0972 & -0.1528 & 0.1111 \\ 0.0694 & 0.0556 & -0.1806 & -0.0972 & -0.0556 & 0.1528 & 0.0972 & -0.4444 & -0.1528 & -0.0694 & 0.4444 & 0.1806 & 0.1111 \\ -0.0139 & 0.0139 & -0.0972 & -0.1389 & -0.0972 & -0.0556 & 0.0556 & -0.1528 & -0.3611 & -0.1528 & -0.2639 & 0.2639 & 0.1111 \\ 0.0556 & -0.0694 & -0.0556 & -0.0972 & -0.1806 & -0.0972 & -0.1528 & -0.0694 & -0.1528 & -0.4444 & -0.1806 & -0.4444 & 0.1111 \end{bmatrix}$$

Consequently, the solution of the actuator commands from equation (5.3) is reduced to a straightforward matrix multiplication. In the case of this example, 117 floating point operations (and a further 108 additions) are required. For on-chip implementation, the values of the matrix is stored in memory and the choice of the number of bits to represent the values depends on the accuracy of the centroid

⁵⁹ The piston component of the mirror can take on any value and still match the wavefront shape. Hence the mirror has to be constrained to an average surface height.

calculation as well as the control specifications of the deformable mirror used. In order that the error in the reconstruction matrix does not propagate through the reconstruction algorithm, and the reconstruction error is mainly due to the error in the position response, the maximum fractional uncertainty in $[\mathbf{B}]^{-1}$ must be sufficiently less than the minimum fractional uncertainty in the wavefront slopes s . At this current stage of development, the centroid processor achieved a positional resolution of 2.1LSB in the y and 1.8LSB in the x . Hence the minimum fractional uncertainty in the centroid measurement is given by:

$$\left| \frac{\delta s}{s_{\max}} \right| = \frac{1.8}{80} = 0.0225 \quad (5.11)$$

where s_{\max} is the maximum position output from the centroid processor and this corresponds to a pixel position of 5 (1010000 or 80). $[\mathbf{B}]^{-1}$ has a minimum element value of 0.0139 and a full scale range of 0.8888. Hence the maximum allowable uncertainty in $[\mathbf{B}]^{-1}$ is given by:

$$\delta B^{-1}_{\max} = \left| B^{-1}_{\min} \right| \left| \frac{\delta s}{s_{\max}} \right| = 0.0139 \times \frac{1.8}{80} = 3.1275 \times 10^{-4} \quad (5.12)$$

As such the minimum number of bits required for $[\mathbf{B}]^{-1}$ is given by:

$$N = \log_2 \frac{0.8888}{3.1275 \times 10^{-4}} \approx 12 \text{ bits} \quad (5.13)$$

However, it turns out that for this configuration and this set of values of $[\mathbf{B}]^{-1}$, the round off error is very small and the same accuracy is obtained if 10 bits are used. Also, although the result of the matrix multiplication will consist of 7 (slopes) + 12 (reconstruction matrix) = 19 bits, a typical deformable mirror usually requires 8 bits or less of control input and the result is usually truncated.

If signed 10-bit encoding is used, for example, the signed integer representation of the inverse matrix values⁶⁰ become:

512	208	512	176	80	176	112	208	112	64	80	64	128
-304	304	176	416	176	-64	64	112	160	112	-16	16	128
-208	-512	80	176	512	-112	-176	64	112	208	-64	-80	128
176	112	-304	-64	-16	416	160	304	64	16	176	112	128
-64	64	-64	-256	-64	-256	256	64	256	64	-64	64	128
-112	-176	-16	-64	-304	-160	-416	16	64	304	-112	-176	128
80	64	-208	-112	-64	176	112	-512	-176	-80	512	208	128
-16	16	-112	-160	-112	-64	64	-176	-416	-176	-304	304	128
-64	-80	-64	-112	-208	-112	-176	-80	-176	-512	-208	-512	128

This can then be converted into signed binary or two's complement for hardware implementation. In general the number of bits of memory required to implement wavefront reconstruction on-chip is given by:

No. of actuators x (No. of subapertures + 1 piston term) x No. of bits required

The same concepts can be applied for the Fried geometry but in this case the relationship between the tilts and the actuator signals is given by (see Figure 5.2(a)):

$$S_{1x} = a_2 + a_5 - a_1 + a_4$$

$$S_{1y} = a_1 + a_2 - a_4 + a_5$$

... (5.14)

5.5 COMPLETE AO SYSTEM

Traditional systems require data to be transmitted from the optical sensor (i.e. a CCD) to a host computer by means of an analogue video line, an analogue-to-digital converter (ADC) and a frame memory and hence are invariably slow and costly. Figure 5.3 shows the structure of our proposed AO system and that of a traditional system. By partitioning the design into its function and incorporating processing at the

⁶⁰ The values have a minimum of -0.4444, maximum of 0.4444 and full scale of 0.8888. For 10-bit encoding (or 1024 levels), an accuracy limit of $0.8888/1024 = 0.0009$ is obtained.

sensor level, the data bottleneck present in traditional system can be alleviated. The final integrated wavefront sensor (iWFS) will consist of an array of tilt sensors with local centroid processing, and wavefront reconstruction circuitry implemented either on-chip or on a dedicated processor such as an FPGA, as illustrated in Figure 5.4. There is a further reduction in data bandwidth of 2 after reconstruction of the wavefront from the wavefront slopes. On-chip implementation will allow higher speeds and a more compact design, while an FPGA implementation has greater flexibility in allowing the system to cope with different mirror and optical configurations.

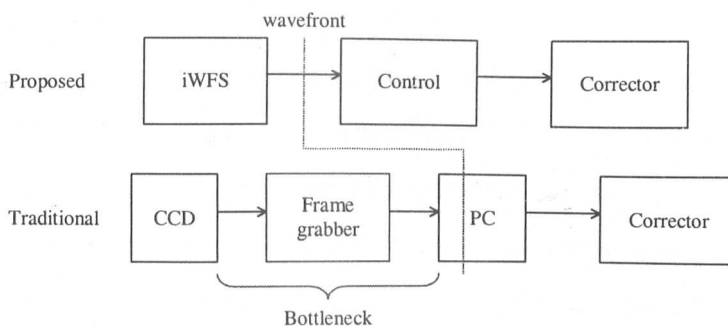


Figure 5.3 Partitioning the AO system by function instead of hardware reduces the data bottleneck

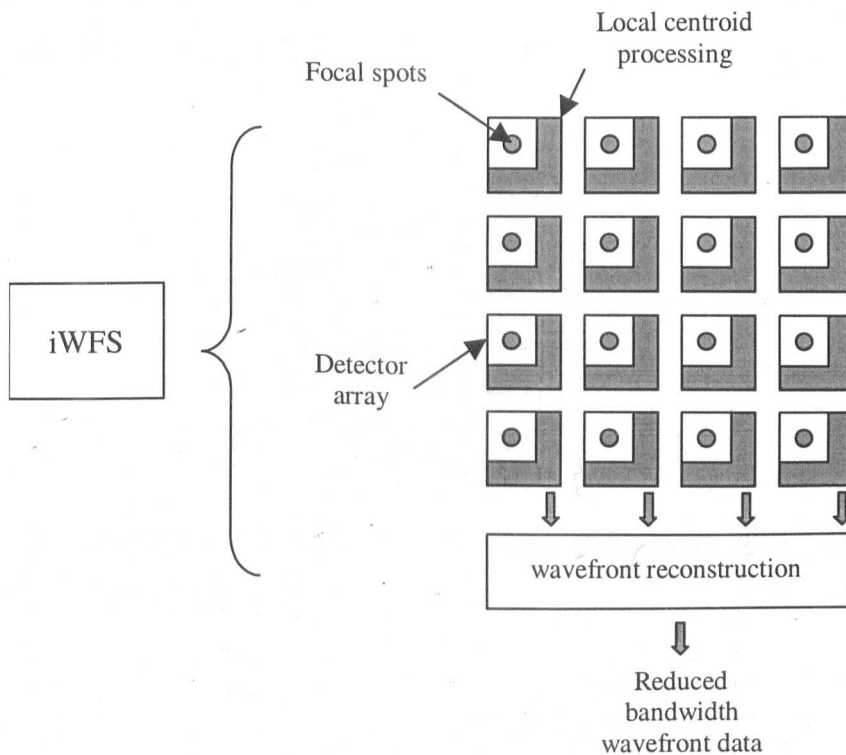


Figure 5.4 Proposed integrated wavefront sensor

There was insufficient time to build a complete working AO system. However, in order to highlight the potential of this system, a proposed architecture incorporating the fabricated tilt sensor and a continuous deformable mirror will be discussed in terms of its speed, its area and its cost. Scalability of this design will also be considered. The deformable mirror selected is the 37-channel 15mm micromachined membrane deformable mirror from OKO Technologies with a settling time of 1ms and is a device which is widely used in other adaptive optical systems [Dayton 2002, de Lima Monteiro 2002, Paterson 2000, Rhoadarmer 1999]. The cost of the mirror is EUR4850 or £3300 including control electronics [Flexible Optical BV]. Each channel is driven by an 8-bit input signal.

The proposed geometry to be used for alignment of the subapertures with the mirror is shown in Figure 5.5, which requires 37 subapertures for the 37 actuators, and has been used by Rhoadarmer et. al. [Rhoadarmer 1999] for testing the wavefront sensor hardware and software for the new Multiple Mirror Telescope adaptive optics system. The DM actuators and WFS subapertures have been projected onto the entrance pupil. The large circle represents the entrance pupil diameter. The hexagons are the DM actuators and the squares are the WFS subapertures. The small circles mark the centers of the actuators. Dayton et. al. [Dayton 2002] used a slightly different geometry with 32 actuators as shown in Figure 5.6.

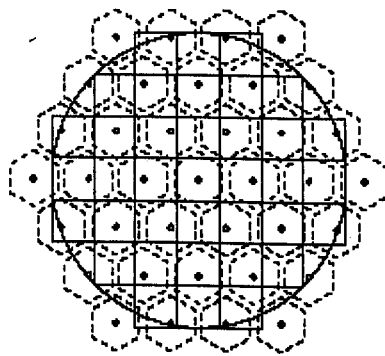


Figure 5.5 Proposed subaperture-actuator geometry for AO system [Rhoadarmer 1999]

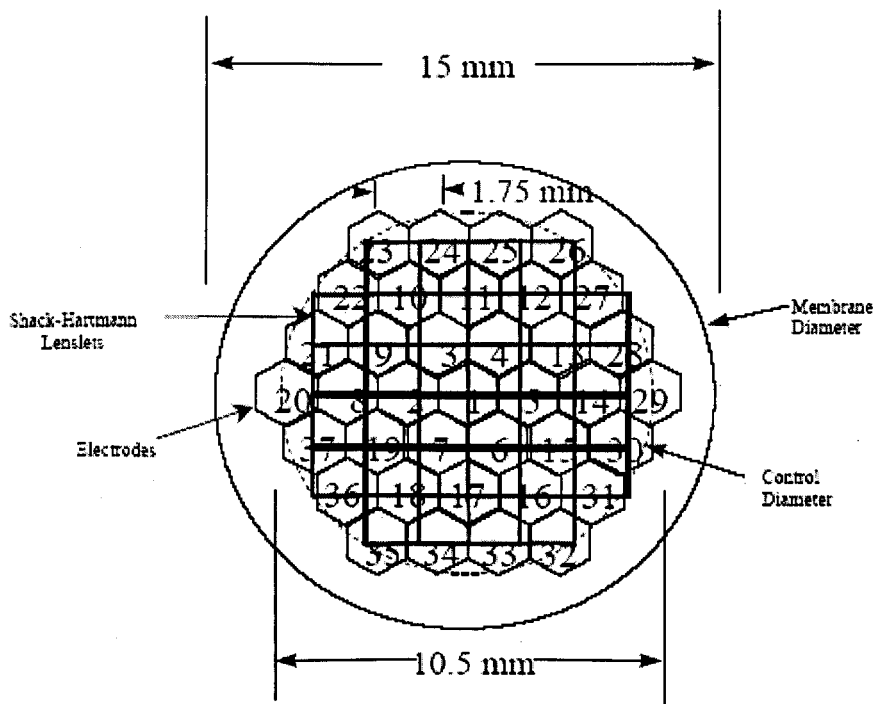


Figure 5.6 Alternative sensor-actuator geometry used by [Dayton 2002]

5.5.1 SPEED

A zonal reconstruction procedure is assumed and the closed loop response time of the system is estimated as follows:

Closed loop response time = Frame time of tilt sensors + Readout time of tilt sensors + Reconstruction time + Mirror settling time

5.5.1.1 Frame time of tilt sensors

All the tilt sensors in the prototype design operate in parallel at a minimum frame rate of 2.4kHz, that is, a frame time of 0.416ms.

5.5.1.2 Readout time of tilt sensors

In the current design, each tilt sensor produces two (x and y) 7-bit centroid values, although the number of bits used to represent the centroids may be increased in subsequent designs to achieve better positional resolution. Parallel 7-bit readout is possible but a serial readout using the 32MHz clock will be assumed here giving a readout time of:

$$T_{\text{readout}} = (14 \text{ bits/subaperture} \times 37 \text{ subapertures}) / 32\text{MHz} = 16.2\mu\text{s} \quad (5.15)$$

5.5.1.3 Reconstruction time

The reconstruction procedure involves the matrix multiplication of the 7-bit centroid values with the inverse reconstruction matrix as described in Section 5.4. Hence, the reconstruction time, T_{recon} , required is given by:

$$T_{\text{recon}} = N_{\text{mult}} \times (T_{\text{mult}} + T_{\text{mem}}) \quad (5.16)$$

where N_{mult} is the number of multiplication procedures required, T_{mult} is the time required for a single multiplication process and T_{mem} is the memory access time for accessing the stored inverse reconstruction matrix values, assuming a single multiplier operating serially on the entire matrix. If instead more than one multiplier unit is used, then the reconstruction time required is given by:

$$T_{\text{recon}} = (N_{\text{mult}} / N_{\text{unit}}) \times (T_{\text{mult}} + T_{\text{mem}}) \quad (5.17)$$

where N_{unit} is the number of multiplier units operating in parallel.

i) Number of multiplication procedures, N_{mult}

The number of multiplication procedures, N_{mult} , required is given by:

$$N_{\text{mult}} = \text{number of actuators} \times (\text{number of slope terms} \times 1 \text{ piston term}) \quad (5.18)$$

Typically, there are two slopes obtained per subaperture (For the Hudgin geometry only one slope per subaperture is required) and hence for the configuration proposed, $N_{\text{mult}} = 37 \times (2 \times 37 + 1) = 2775$ procedures.

ii) Multiplication delay time, T_{mult}

If 12-bits are used for the inverse reconstruction matrix values, then the result of the multiplication will be 19-bits long. For the multiplication, a shift and add method can be used and for the addition, a carry-look-ahead adder is preferred over a ripple adder for its parallelism and hence, shorter delay. A 16-bit 2-level carry-look-ahead (CLA) adder⁶¹ requires just 9 gate delays to complete [Parhami 2000] and this can easily be extended to a 19-bit value by incorporating another 4-bit CLA adder in the 1st level. This does not entail significant delay overhead because although the fanout and the

⁶¹ This consists of four 4-bit CLA adders in the first level and a 2nd stage carry-look-ahead generator in the second level.

individual gate delay increases slightly, the number of levels or the number of gate delays remains the same. An alternative to the carry-look-ahead adder is the carry-save adder which has the advantage of reduced number of gates at the expense of longer delay times. Also the pipelining of carry-save adders is a simple matter and is suitable where the result of a carry-save addition is immediately re-used in another addition e.g. in multiplication. In the case of a 12-bit by 7-bit multiplication, a carry-save adder approach will require 2×19 clock cycles while the carry-look-ahead adder technique requires 2×7 clock cycles but the cycle can be made faster for the case of the pipelined carry-save adders. Other complicated optimised multipliers and adders [Parhami 2000] are also possible. However, for the purpose of this investigation only the carry-look-ahead adder will be considered. With a typical gate delay of $<1\text{ns}$, the addition process can be performed within one cycle of the 32MHz clock ($1/32\text{MHz} = 31.25\text{ns}$) and the shift and add multiplication requires 7×2 clock cycles with a 7-bit quotient (centroid) and 2 cycles per bit for shift and add. Hence the multiplication delay time, T_{mult} , becomes:

$$T_{\text{mult}} = 14 / 32\text{MHz} = 0.4375\mu\text{s} \quad (5.19)$$

iii) Memory access time, T_{mem}

The inverse reconstruction matrix is fixed for a given configuration and these values can be stored in memory for fast reconstruction computation. Foundries often provide service for the generation of on-chip random access memory (RAM) and in the AMS $0.35\mu\text{m}$ CMOS (C35) process, for example, a 2775 word \times 12-bit single port RAM will take up an area of 1.86mm^2 with an access time of 5.75ns [Austriamicrosystems – Memory Compiler].

iv) Reconstruction time, T_{recon}

Therefore, with a single serial multiplier the reconstruction time, T_{recon} , required for this design is:

$$T_{\text{recon}} = 2775 \times (0.4375\mu\text{s} + 5.75\text{ns}) = 1.23\text{ms} \quad (5.20)$$

It is possible to move to a faster clock to perform the arithmetic calculations as well as use several parallel multiplier units to reduce the delay time. De Lima Monteiro [de Lima Monteiro 2002] performed modal reconstruction on PC using a 750MHz Pentium III Processor for 64 quad cells (128 slopes), 9 Zernike modes and 37 mirror control signals, and this took a time of 134 μ s, so it is possible to go much faster. Also, according to de Lima Monteiro, the control and feedback algorithms run in a 750MHz Pentium III PC, under Linux, are not of concern compared to the other elements of the system.

5.5.1.4 Mirror settling time

The mirror settling time was 1ms [de Lima Monteiro 2002]. The mirror has to be stable during WFS integration so the mirror actuation and settling cannot be pipelined with the wavefront sensing and reconstruction.

5.5.1.5 Closed loop response time

Therefore, the closed loop response time = 0.416ms + 16.2 μ s + 1.23ms + 1ms = 2.66ms. Hence, the closed loop bandwidth = 376Hz. Rhoadarmer et. al. [Rhoadarmer 1999] used a 80 x 80 array 4-port, split frame transfer CCD with 1kHz frame rate with zonal reconstruction and achieved a closed loop bandwidth of 5Hz, while Dayton et. al. [Dayton 2002] achieved a closed loop bandwidth of 80Hz. De Lima Monteiro [de Lima Monteiro 2002] achieved an operational frequency (sensor readout and wavefront reconstruction) of 370Hz and a closed loop bandwidth (sensor readout, wavefront reconstruction, mirror actuation and settling time) of 260Hz with 44 quad cells and modal reconstruction of 9 Zernike modes. Paterson et. al. [Paterson 2000] used a 128 x 128 CCD with a maximum frame rate of 800Hz and the closed loop bandwidth achieved was 50Hz. Hence the system compares favourably to other similar systems.

Assuming the number of subapertures is equal to the number of actuators and there are two slopes per subaperture, the delay of the system as the number of degrees of freedom (actuators) increases can be shown. Figure 5.7 shows the delay of the system when a 32MHz clock with a single multiplier is used while Figure 5.8 shows the delay

when a 200MHz clock and 10 multiplier units are used. It can be seen that moving to faster clock speeds and using parallel multiplier units will remove the delay bottleneck from the reconstruction procedure to the mirror settling time. Also, by moving to faster ADC techniques for the tilt sensors, it is possible to remove this as a key delay as well.

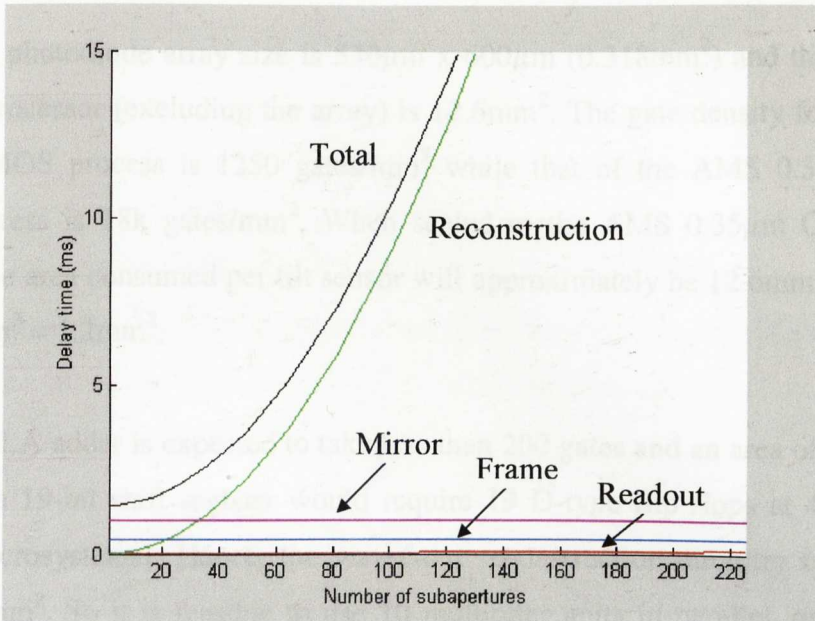


Figure 5.7 Delay times for the AO system when a 32MHz clock and 1 multiplier is used

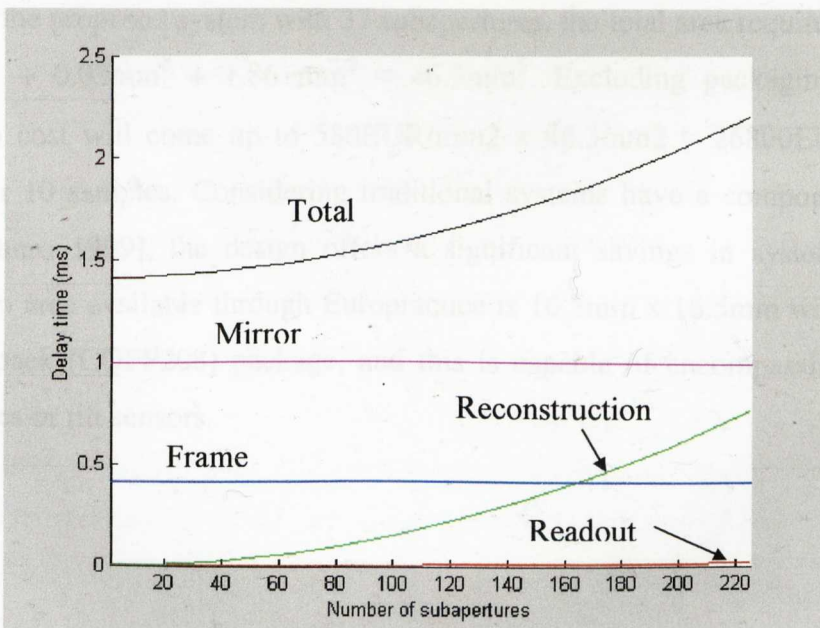


Figure 5.8 Delay times for the AO system when a 200MHz clock and 10 multipliers are used

5.5.2 AREA AND COST

The integrated wavefront sensor consists of 3 main components; the fabricated tilt sensors, the wavefront reconstruction circuitry and memory storage for the reconstruction matrix.

The 5 x 5 photodiode array size is $530\mu\text{m} \times 600\mu\text{m}$ (0.318mm^2) and the area of the centroid processor (excluding the array) is 12.6mm^2 . The gate density for the Mietec $0.7\mu\text{m}$ CMOS process is $1250 \text{ gates}/\text{mm}^2$ while that of the AMS $0.35\mu\text{m}$ CMOS (C35) process is $18\text{k gates}/\text{mm}^2$. When scaled to the AMS $0.35\mu\text{m}$ CMOS (C35) process, the area consumed per tilt sensor will approximately be $12.6\text{mm}^2 \times 1250/18\text{k} + 0.318\text{mm}^2 = 1.2\text{mm}^2$.

A 19-bit CLA adder is expected to take less than 200 gates and an area of $100\mu\text{m}^2$ per gate, and a 19-bit shift register would require 19 D-type flip flops at $400\mu\text{m}^2$ each [Austriamicrosystems]. Hence, the wavefront reconstruction circuitry takes up less than 0.03mm^2 . So it is feasible to use 10 multiplier units in parallel, or even more. Ultimately, the bottleneck of the system will lie with the settling time of the mirror except for the case where a very large number of actuators are needed.

Hence for the proposed system with 37 subapertures, the total area required will be $37 \times 1.2\text{mm}^2 + 0.03\text{mm}^2 + 1.86 \text{ mm}^2 \approx 46.3\text{mm}^2$. Excluding packaging costs, the fabrication cost will come up to $580\text{EUR}/\text{mm}^2 \times 46.3\text{mm}^2 \approx 26800\text{EUR}$ or about $\pounds 18000$ for 10 samples. Considering traditional systems have a component costs of $>\pounds 10^5$ [Munro 1999], the design offers a significant savings in system cost. The largest chip area available through Europractice is $16.5\text{mm} \times 16.5\text{mm}$ with a ceramic quad flat pack (CQFP208) package, and this is capable of encompassing over 220 subapertures or tilt sensors.

5.6 CHAPTER SUMMARY

In this chapter, the concepts for wavefront reconstruction were introduced. A wavefront can be described using a zonal approach where the wavefront is divided up into subapertures (zones) or using a modal approach where the wavefront is treated as a sum of basis functions (modes) with Zernike polynomials being a popular choice.

Wavefront reconstruction usually involves the solution of a linear system of equations in matrix form ($s = \mathbf{B} \mathbf{a}$) where for an overdetermined system, a linear least-square approximation can be used to solve for the actuator signals \mathbf{a} by finding the pseudo-inverse of the reconstruction matrix \mathbf{B} and multiplying this by the measured wavefront tilts s ($\mathbf{a} = [\mathbf{B}^T \mathbf{B}]^{-1} \mathbf{B}^T s$). For a given configuration, the pseudoinverse matrix $[\mathbf{B}^T \mathbf{B}]^{-1} \mathbf{B}^T$ need only be calculated once reducing the wavefront reconstruction computation to a single matrix multiplication.

Two types of reconstruction techniques from wavefront slopes are possible and these are the modal and zonal techniques. With the modal technique, coefficients of the polynomial function for describing the wavefront (Zernike) are obtained and these need to be converted into actuator commands for driving the deformable mirror. With the zonal technique, the phases of the wavefront at regular discrete points on the aperture are obtained and these translate directly into actuator commands. In the zonal approach, the sensor-actuator geometry and choice of deformable mirror directly affects the generation of the reconstruction matrix \mathbf{B} and an example for the reconstruction of a 3 x 3 actuator system with a Hudgin geometry was shown. Once the pseudo-inverse matrix of \mathbf{B} is obtained, it can be converted into binary values and stored in on-chip RAM for wavefront reconstruction allowing a complete, cheap, fast, low cost adaptive optics system to be built.

The structure for our proposed AO system was then presented and it was shown that the parallel processing achieved with the system allowed a closed loop bandwidth of more than 370Hz and at a fraction of the cost of traditional AO systems. The design is able to remove the bottleneck from the readout and processing of the wavefront to its fundamental limit of the mirror settling time.

CHAPTER 6

CONCLUSIONS

6.1 DISCUSSION

The research covered in this thesis addresses the need for a fast, low cost integrated wavefront sensor for use in an adaptive optical system. An adaptive optical (AO) system corrects for wavefront distortion in the imaging medium, such as the atmosphere, by having a closed loop detection and correction scheme. A Shack Hartmann wavefront sensor uses an array of small lenslets to sample the optical wavefront and by detecting the deviation of the focused spots from reference positions, the local wavefront tilts are obtained. Currently with most of these systems, a single CCD is used to sample the entire wavefront before it can be processed, resulting in a data bottleneck. This research addresses this issue by integrating local centroid processing for each local wavefront tilt which will allow parallel processing of the wavefront. In addition, removing the need for a CCD-frame grabber-PC architecture will lead to a reduction in system size, cost and power consumption. Adaptive optics has traditionally been known for its use in astronomical and military applications mainly because of the high cost of the components in the system. With a low-cost real-time adaptive optical system, many new application areas such as ophthalmology, intra and extra-cavity laser correction, free space optical communications and microscopy, will become feasible. The design stages of the system are summarised below.

6.1.1 DESIGN SPECIFICATIONS

There are several possible structures for implementing a position sensitive device (PSD) such as the lateral effect photodiode (LEP), the quad cell and the multi-pixel array. A lateral effect PSD requires large uniform sheet resistance for linear operation,

which is not readily available in a standard CMOS process making integration with circuitry difficult. Quad cells have simple readout schemes but are not very linear. Multi pixel arrays have better linearity, sub-pixel accuracy and positional range. They also offer greater flexibility and are able to deal with multiple spots and non-uniform intensity profiles. The drawback is the increased computational load but for moderate array sizes this is reasonable and this was the architecture chosen for our system. A 5 x 5 pixel array was selected as a tradeoff between linearity and circuit complexity.

In terms of centroid processing, there have been various efforts to implement centroid detection on a CMOS process for numerous applications. In general, analogue multi-pixel array approaches suffer from low fill factor and poor linearity due to poor tolerance of components such as polysilicon resistors and capacitors. Binary position sensing techniques using Winner Take All (WTA) circuitry or an on-pixel comparator does not offer subpixel accuracy and cannot cope with multiple spots or non-uniform spots. This research explores the approach of a dedicated digital centroid processor which offers high accuracy and greater flexibility and programmability for various image processing tasks. In terms of pixel architectures, the CMOS active pixel sensor (APS) was selected as it offers high fill factor and low mismatch compared to other APS types.

6.1.2 CHARACTERISATION OF CMOS PHOTODIODE STRUCTURES

An important design requirement for this work is the integration of circuitry at the sensor level and this is difficult to achieve with CCDs. As such, a standard CMOS process was used. However, CMOS processes have been optimised towards microelectronic circuitry rather than imaging. Hence, characterisation of photodetector structures in a standard CMOS process was necessary. The CMOS process selected for the work was the Mietec 0.7 μ m CMOS process accessed via the Europractice IC Service. The CMOS photodiode structures were characterised for dark current, capacitance, spatial response, responsivity and spectral response. The dark current for the devices tested was of the order of 1pA or less for a reverse bias voltage of 2 - 4V. The capacitance of the deep device (0.5pF for a 100 μ m x 100 μ m photodiode at

2V reverse bias) was shown to be smaller than the shallow devices (3.2pF for a 100 μm x 100 μm photodiode at 2V reverse bias) making them more suitable for high speed applications. The presence of an inadvertent Schottky barrier diode lowered the capacitance further. The photodiode structures were also shown to be highly linear with incident light intensity and to have saturation levels higher than 2.7mW of light power. The results of the characterisation work showed that without the need for any process modifications photodiodes in standard CMOS showed good responsivity of the order of 0.3A/W. In terms of spectral response, the deep photodiode has better spectral response at longer wavelengths while the shallow performed better at shorter wavelengths. This is due to the absorption coefficient and penetration depth of light into silicon, where light of longer wavelength penetrates deeper into the substrate. The deep well-substrate photodiode was chosen for integration of the ASIC because of its low capacitance, low leakage in reverse bias and high responsivity particularly at longer wavelengths.

6.1.3 DESIGN PROTOTYPING

To achieve the goal of fabricating a single IC optical centroid processor, a design philosophy of functional validation via a hardware emulation system prior to chip fabrication was employed. This reduces the risk and the number of iterations and fabrication runs needed to produce a working centroid processor. The hardware emulation system consists of a 5 x 5 photodiode array, a transimpedance amplifier for current-to-voltage conversion, an analogue-to-digital converter (ADC) and a reprogrammable FPGA processor for calculating the centroid. The hardware emulation system was tested with a commercial photodiode array and a full custom standard CMOS photodiode array fabricated in the Mietec 0.7 μm CMOS process. The centroid processor successfully computed the centroids at a rate of 1.54kHz, which was limited by the maximum conversion frequency of the ADC of 40kHz. Having proven the functionality of the digital centroid processor and the use of a full custom photodiode array, the next stage in the design was to integrate the full custom array with the digital centroid processor onto a single CMOS IC chip.

For the ASIC centroid processor, an active pixel sensor array was used for buffering of the pixel and current-to-voltage conversion. The pixel architecture was optimised according to simulation results and circuit analysis. Digitisation of the pixel output was done using a counter and two comparators to measure the discharge time of the pixel. The dynamic range of the pixel output could be extended using a two cycle adaptive technique. The digitised pixel values were then computed by the digital centroid processor which was previously verified by the hardware emulation system. The ASIC allowed different modes of operation and various control signals for increased testability and observability. Being a mixed full-custom and semi-custom design, several layout and design issues had to be considered such as design partitioning, power supply management, physical design verification and clock tree planning. The fabricated optical centroid processor successfully obtained and transmitted the centroids at a rate of 2.4 - 4.8 kHz allowing real-time operation in many applications.

6.1.4 COMPLETE AO SYSTEM

It was shown that wavefront reconstruction for a Shack Hartmann wavefront sensor can be reduced to a simple matrix multiplication with on-chip memory storage so integration of wavefront sensing and wavefront reconstruction can easily be achieved, leading to cheap and fast adaptive optical systems. The structure for a proposed AO system was presented to illustrate the scalability of the design and the advantage drawn from processing the centroids in parallel. The system was capable of achieving a closed-loop bandwidth of more than 370Hz and at a fraction of the cost of traditional AO systems. The design is able to remove the bottleneck from the readout and processing of the wavefront to its fundamental limit of the mirror settling time.

6.2 FURTHER WORK

The recommended further work for this design shall be summarised below:

- 1) To improve the positional resolution of the processor, the number of bits in the centroid representation can be increased. This requires very little overhead. That is, just an additional shift-and-subtract cycle in the divider is required for each additional bit in the result, without the need for additional storage for the dividend and divisor.

- 2) The limited spatial dynamic range of the design was attributed to stray light, crosstalk between pixels and the choice of ADC technique which compresses higher intensity light levels hence reducing the signal to background ratio. The problem of stray light can be overcome by improving the optical setup. To resolve the problem of pixel crosstalk, individual pixel reset can be implemented, such that when one pixel is discharging the other pixels remain under reset. Also, a suitably biased guard ring structure can be incorporated between pixels to mop up any crosstalk current. FPN noise removal circuitry or pixel offset subtraction should be incorporated in future designs.

- 3) To overcome the speed limitation of the current ADC technique, an alternate digitisation structure can be implemented where a conventional ADC is used to digitise the final discharge voltage, which would also allow pixel values to be readout and digitised sequentially without the need of separate discharge cycles per pixel.

- 4) In the initial design, 26 pixel periods were required per frame but this can be reduced to 25 pixel periods by using the fast clock instead of the pixel reset signal to latch out the data and clear the registers, thereby allowing a new frame to start immediately after 25 pixel periods.

- 5) Under conditions of low signal-to-noise ratio (SNR) it is possible to improve the accuracy of the centroiding by removing the background signal through thresholding. With a digital centroid processor this is easily achieved by subtracting a programmable, even adaptive, offset from the digitised input or by setting all pixel values below a certain threshold to zero before the centre of gravity is found. It is also possible to apply windowing to improve the SNR further.

- 6) Testing of mode 2 and mode 3 operation of the ADC should be carried out, which are expected to give better noise rejection and increased dynamic range capability respectively.
- 7) The current design should be tested with different laser beam sizes in order to characterise and quantify the linearity of the device as a centroid detector.
- 8) Faster, more robust off-chip readout techniques can be considered in place of the RS232 link which has limited data rates, such as the Universal Serial Bus (USB) which has a data rate of up to 480Mbps.
- 9) Finally, an array of tilt sensors can be integrated along with wavefront reconstruction to form a complete low-cost real-time adaptive optical system.

6.3 CONCLUSIONS

The main conclusions from this work are highlighted below:

- 1) CMOS photodiode structures offer satisfactory responsivity (about 0.3 A/W) for the intended application while allowing high levels of circuit integration not possible with the CCD process. This has allowed the use of parallel processing to remove the data bottleneck in traditional CCD systems.
- 2) A hardware emulation system was used to confirm the performance of the design prior to ASIC fabrication, hence reducing the risk and the number of iterations needed to produce a working centroid processor. The hardware emulation system successfully computed centroids at a rate of 1.54kHz which was limited by the speed of the ADC used [Pui 2002]. Due to the re-programmable nature of the FPGA the hardware emulation environment can also be used for prototyping many other optical processing algorithms.

3) This work represents the only dedicated digital centroid processor fabricated to date and it was integrated with an on-chip CMOS photodiode array and the system successfully processed and transmitted centroids at a rate of 2.4 – 4.8 kHz [Pui 2004], removing the data bottleneck present in traditional CCD systems and allowing real-time operation in many applications.

4) The centroid processor has the potential to be scaled to a complete cheap and fast AO system. The division process of the centroid processor can make use of latency in the design to be shared among several centroid processors. In addition, moving to smaller feature sizes and improved routing capability will lead to a significant reduction in the size of the digital centroid processor, which is an advantage not offered by analogue approaches due to the tolerance of its components. When integrated with an array of tilt sensors operating in parallel, the frame rate of the design is not limited by the number of tilt sensors employed. In fact, the speed advantage over traditional systems increases with the number of tilt sensors required.

REFERENCES

- [Air Force Research Laboratory Directed Energy Directorate 1997] Air Force Research Laboratory Directed Energy Directorate. (September, 1997). "Starfire Optical Range - Imagery (Binary Star - k-Peg)". Available at <http://www.de.afrl.af.mil/SOR/binary.htm>
- [Alcatel Microelectronics 1995a] Alcatel Microelectronics, "Design Guidelines for Supply Protection Configuration," *Document GP13055, Revision 1*, 1995.
- [Alcatel Microelectronics 1995b] Alcatel Microelectronics, "MTC22500 Analog Library Manual," *Revision 1.0*, 1995.
- [Alcatel Microelectronics 1998] Alcatel Microelectronics, "Standard Cell Design Data Book (0.7 μ m CMOS) - MTC23000 Compact Core Cells," *Revision 1.0*, 1998.
- [Alcatel Microelectronics 1999a] Alcatel Microelectronics, "Electrical parameters CMOS 0.7 μ m - C07MA and C07MD," *Document DS13291, Revision 8*, 1999.
- [Alcatel Microelectronics 1999b] Alcatel Microelectronics, "Layout rules CMOS 0.7 μ m," *Document DS13290, Revision 14*, 1999.
- [Allen 1987] Allen, P. E., Holberg, D. R., "CMOS Analog Circuit Design," Oxford University Press, 1987.
- [Ambundo 2002] Ambundo, A. J., Furth, P. M., "Fully Integrated Current-Mode Subaperture Centroid Circuits and Phase Reconstructor," *10th NASA Symposium on VLSI Design*, March 2002.
- [Anderson 1999] Anderson, M. H., "ADAPTIVE OPTICS: Liquid crystals lower the cost of adaptive optics," *Laser Focus World*, December, 1999.
- [Angel 2000] Angel, R., Fugate, B., "Adaptive Optics," *Science*, Vol. 288, pp. 455-456, 2000.
- [Applied Optics Group; Imperial College] Applied Optics Group; Imperial College. "Low-Cost Adaptive Optics". Available at <http://op.ph.ic.ac.uk/lcao/lowcost.html>
- [Austriamicrosystems 2004] Austriamicrosystems. (May 2004). "C35 Core Cells". 2.0, Available at http://asic.austriamicrosystems.com/databooks/c35/databook_c35_33/index.html
- [Baker 1997] Baker, R. J., Li, H. W., Boyce, D. E., "CMOS Circuit Design, Layout, and Simulation," Wiley-IEEE Press, 1997.

-
- [Bar-Lev 1984] Bar-Lev, A., "Semiconductors and Electronic Devices," 2nd ed: Prentice Hall, 1984.
- [Berkefeld 2001] Berkefeld, T., Glindemann, A., Hippler, S., "Multi-Conjugate Adaptive Optics with Two Deformable Mirrors – Requirements and Performance," *Experimental Astronomy*, Vol. 11, No. 1, pp. 1-21, February 2001.
- [Biber 2000] Biber, A., Seitz, P., Jäckel, H., "Avalanche Photodiode Image Sensor in Standard BiCMOS Technology," *IEEE Transaction on Electron Devices*, Vol. 47, pp. 2241-2243, 2000.
- [Bogaerts 2000] Bogaerts, J., Dierickx, B., Van Hoof, C., "Radiation Induced Dark Current Increase in CMOS Active Pixel Sensors," *Proceedings of SPIE*, Vol. 4134, October 2000.
- [Booth 2002a] Booth, M. J., Neil, M. A. A., Juskaitis, R., Wilson, T., "Adaptive aberration correction in a confocal microscope," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99, No. 9, pp. 5788-5792, 30 April 2002.
- [Booth 2002b] Booth, M. J., Neil, M. A. A., Wilson, T., "New modal wave-front sensor: application to adaptive confocal fluorescence microscopy and two-photon excitation fluorescence microscopy," *Journal of the Optical Society of America A*, Vol. 19, No. 10, pp. 2112-2120, October 2002.
- [Burns 2003] Burns, R., Hornsey, R., "CMOS Image Sensor With Cumulative Cross Section Readout," *IEEE Workshop on CCDs and Advanced Image Sensors*, May 15-17, 2003.
- [Burr Brown Corporation 1994] Burr Brown Corporation, "ADS7807 Low-Power 16-Bit Sampling CMOS Analog-to-Digital Converter," *Product Datasheet*, 1994.
- [Bursky 1999] Bursky, D., "CMOS Megapixel Image Sensors Deliver Nearly Noise-Free Pictures.," *Electronic Design*, Nov 22, 1999.
- [Canadian VLOT Working Group 2003] Canadian VLOT Working Group, "Very Large Optical Telescope (VLOT) Book," Vol. 11 (Adaptive Optics), November 2003.
- [Cash 1987] Cash, G. L., Hatamian, M., "Optical character recognition by the method of moments," *Computer Vision, Graphics, and Image Processing*, Vol. 39, No. 3, pp. 291-310, 1987.
- [Centronic Ltd. 1998] Centronic Ltd., "High Performance Silicon Photodetectors," *Catalogue, European Edition 2*, 1998.
- [Chang 1999] Chang, K. C., "Digital Systems Design with VHDL and Synthesis: An Integrated Approach," Wiley-IEEE Computer Society, 1999.

- [Chen 2002] Chen Foo Sen, "IOS (integrated optical sensors) test bed - spectral sensitivity test bed," *Thesis*, University of Nottingham, 2002.
- [Chen 2000] Chen, T., Catrysse, P., El Gamal, A., Wandell, B., "How Small Should Pixel Size Be?," Electronic Imaging 2000 conference, San Jose, CA, January 2000.
- [Cherezova 1997] Cherezova, T. Y., Chesnokov, S. S., Kaptsov, L. N., Kudryashov, A., "Doughnut-like laser beam output formation by intracavity flexible controlled mirror," *Optics Express*, Vol. 3, No. 5, pp. 180-189, 1997.
- [Chou 1991] Chou, T. L., Wong, E. J., Lee, W. C., Kuo, J. B., "A BiCMOS image sensor with a chopper-stabilized edge detector and a correlated-double-sampling readout circuit for neural network VLSI operating at 77 K," Custom Integrated Circuits Conference, San Diego, USA, 1991.
- [Dayton 2002] Dayton, D., Gonglewski, J., Restaino, S., Martin, J., Phillips, J., Hartman M., Browne, S., Kervin, P., Snodgrass, J., Heimann, N., Shilko, M., Pohle, R., Carrion, B., Smith, C., Thiel, D., "Demonstration of new technology MEMS and liquid crystal adaptive optics on bright astronomical objects and satellites," *Optics Express*, Vol. 10, No. 25, pp. 1508-1519, 2002.
- [Dayton 1997] Dayton, D., Sandven, S., Gonglewski, J., Browne, S., Rogers, S., McDermott, S., "Adaptive optics using a liquid crystal phase modulator in conjunction with a Shack-Hartmann wave-front sensor and zonal control algorithm," *Optics Express*, Vol. 1, No. 11, pp. 338-346, 1997.
- [de Lima Monteiro 2002] de Lima Monteiro, D., "CMOS-Based Integrated Wavefront Sensor," Delft University Press, 2002.
- [Deweerth 1992] Deweerth, S. P., "Analog VLSI Circuits for Stimulus Localization and Centroid Computation," *International Journal of Computer Vision*, Vol. 8, No. 2, pp. 191-202, 1992.
- [Dewey 1997] Dewey, A. M., "Analysis and design of digital systems with VHDL," PWS publishing company, 1997.
- [Diaspro 2001] Diaspro, A., "Confocal and Two-Photon Microscopy : Foundations, Applications and Advances," Wiley-Liss, 2001.
- [Dickinson 2003] Dickinson, J., Goldin, J., Lund, R., Myhr, S., "Centroid and Volume Tracking Imager," *Design Projects, 520.490: Analog and Digital VLSI Systems and Architecture*, Johns Hopkins University, Fall 2003.
- [Dierickx 1997] Dierickx, B., Meynants, G., Scheffer, D., "Near 100% fill factor CMOS active pixels," IEEE CCD & AIS workshop, Brugge, Belgium, 5-7 June, 1997.

- [Dierickx 1996] Dierickx, B., Scheffer, D., Meynants, G., Ogiers, W., Vlummens, J., "Random addressable active pixel image sensors," *Proceedings of SPIE*, Vol. 2950, pp. 2-7, 1996.
- [Dillon 1999] Dillon, N., "1-D Centroiding Shack Hartmann Wavefront Sensor for Adaptive Optics," *Gemini Acquisition and Guidance (A&G) Reports*, 1999.
- [Doelman 2000] Doelman, N. J., Mulder, E. H., "Control strategies for Adaptive Optical systems," *Adaptronic Congress 2000*, Potsdam, 2000.
- [Dominguez-Castro 1997] Dominguez-Castro, R., Espejo, S., Rodríguez-Vázquez, A., Carmona, R. A., Földesy, P., Zarándy, A., Szolgay, P., Szirányi, T., Roska, T., "A 0.8-um CMOS Two-Dimensional Programmable Mixed-Signal Focal-Plane Array Processor with On-Chip Binary Imaging and Instructions Storage," *IEEE Journal of Solid-State Circuits*, Vol. 32, pp. 1013-1025, July 1997.
- [Droste 2002] Droste, D., Bille, J., "An ASIC for Hartmann-Shack wavefront detection," *IEEE Journal of Solid State Circuits*, Vol. 37, No. 2, pp. 173-182, 2002.
- [Dudani 1977] Dudani, S. A., Breeding, K. J., McGhee, R. B., "Aircraft identification by moment invariants," *IEEE Transactions on Computers*, Vol. 26, pp. 39-45, 1977.
- [El Gamal] El Gamal, A., "Generic 0.5 m N-well CMOS Process Information," *Handout #3, EE392b: Introduction to Image Sensors and Digital Cameras, Vision, Imaging Science and Technology Activities (VISTA), Stanford University*.
- [Endo 2003] Endo, Y., Nitta, Y., Kubo, H., Murao, T., Shimomura K., Kimura, M., Watanabe, K., Yamamoto, S., Komori, S., "4-micron pixel CMOS image sensor with low-image lag and high-temperature operability," *Proceedings of SPIE*, Vol. 5017, pp. 196-204, 2003.
- [Europractice IC Service] Europractice IC Service, "AMI Semiconductor 0.7 μ m CMOS Technology Description," http://www.europractice.imec.be/europractice/on-line-docs/prototyping/ti/ti_mtc0u7.html.
- [Flexible Optical BV 2003] Flexible Optical BV, "Pricelist OKO Technologies as of 10 November, 2003," 2003.
- [Forcheimer 1993] Forcheimer, R., Chen, K., Svensson, C., Odmark, A., "Single Chip Image Sensors With a Digital Processor Array," *Journal of VLSI Signal Processing*, Vol. 5, pp. 121-131, 1993.
- [Forcheimer 1992] Forcheimer, R., Ingelhag, P., Jansson, C., "MAPP2200 - A second generation smart optical sensor," *Proceedings of SPIE*, Vol. 1659, pp. 2-11, February 1992.

-
- [Forchheimer 1994] Forchheimer, R., Åström, A., "Near-Sensor Image Processing: A New Paradigm," *IEEE Transactions on Image Processing*, Vol. 3, No. 6, pp. 736-746, 1994.
- [Fossum 1997] Fossum, E. R., "CMOS Image Sensors: Electronic Camera-On-A-Chip," *IEEE Transactions on Electron Devices*, Vol. 44, No. 10, pp. 1689-1698, October 1997.
- [Fosu 2004] Fosu, C., Hein, G. W., Eissfeller, B., "Determination of Centroid of CCD Star Images," Proceedings of the XXth ISPRS Congress, Istanbul, Turkey, 12-23 July 2004.
- [Fried 1965] Fried D. L., "The Effect of Wavefront Distortion on the Performance of an Ideal Optical Heterodyne Receiver and an Ideal Camera," Conf. on Atmospheric Limitations to Optical Propagation, 1965.
- [Furth 1998] Furth, P. M., Clark, N., "Analog VLSI subaperture centroid circuits," *7th NASA Symposium on VLSI Design*, October 1998.
- [Geary 1995] Geary, J. M., "Introduction to Wavefront Sensors," in *Tutorial Texts in Optical Engineering*, vol. TT18: Society of Photo-Optical Instrumentation Engineers (SPIE), 1995.
- [Geiger 1996] Geiger, M., Schuberth, S., Hutfless, J., "CO₂ laser beam sawing of thick sheet metal with adaptive optics," *Welding in the World*, pp. 5-11, 1996.
- [Gonnason 1990] Gonnason, W. R., Haslett, J. W., Trofimenkoff, F. N., "A Low Cost High Resolution Optical Position Sensor," *IEEE Transactions on Instrumentation and Measurement*, Vol. 39, pp. 658-663, August 1990.
- [Goodwin 1992] Goodwin, M., "Serial Communications in C and C++," Hungry Minds, Inc., 1992.
- [Gourlay 2000] Gourlay, J., Yang, T., Ishikawa, M., Walker, A. C., "Low-order Adaptive Optics for Free-Space Optoelectronic Interconnects," *Applied Optics*, Vol. 39, No. 5, pp. 714-720, February 2000.
- [Gray 1992] Gray, P. R., Meyer, R. G., "Analysis and design of analog integrated circuits," 3rd ed: John Wiley & Sons, Inc., 1992.
- [Guidash 1995] Guidash, R. M., Lee, P. P., Andrus, J. M., Ciccarelli, A. S., Erhardt, H. J., Fischer, J. R., Meisenzahl, E. J., Philbrick, R. H., Kenney, T. J., "Modular high-performance 2-um CCD-BiCMOS process technology for application-specific image sensors and image sensor systems on a chip," *Proceedings of SPIE, Charge-Coupled Devices and Solid State Optical Sensors V*, Vol. 2415, pp. 256-264, 1995.

-
- [Haferkamp 1993] Haferkamp, H., Schmidt, H., Seebaum, D., Homburg, A., "Beam delivery using adaptive optics for material processing applications with high power CO lasers," *Proceedings of SPIE's International Symposium on Optical Tools for Manufacturing and Advanced Automation*, pp. 14-22, September 1993.
- [Hatamian 1986] Hatamian, M., "A real time two-dimensional moment generating algorithm and its single chip implementation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-34, No. 3, June 1986.
- [Hatcher 2001] Hatcher, M., "Deformable mirrors flex low-cost potential," *Opto & Laser Europe*, May 2001.
- [Hoeschele 1994] Hoeschele, D. F., "Analog-to-Digital and Digital-to-Analog Conversion Techniques," 2nd ed: John Wiley & Sons, Inc., 1994.
- [Holloway 1983] Holloway, H., Brailsford, A. D., "Peripheral photoresponse of a p-n junction," *Journal of Applied Physics*, Vol. 54, No. 8, pp. 4641-4656, August 1983.
- [Hong 2001] Hong, C., Hornsey, R. I., "Inverted Logarithmic Active Pixel with Current Readout," *IEEE Workshop on CCDs and Advanced Image Sensors*, June 2001.
- [Horn 1986] Horn, B. K. P., "Robot Vision." Cambridge, Massachusetts: MIT Press, 1986.
- [Hornsey 1999a] Hornsey, R. I., "Appendix C - Other sensor types," *Short Course Notes, Two-day short course presented at the Waterloo Institute for Computer Research*, May 1999.
- [Hornsey 1999b] Hornsey, R. I., "Fabrication Technology and Pixel Design," *Short Course Notes, Two-day short course presented at the Waterloo Institute for Computer Research*, May 1999.
- [Hornsey 1999c] Hornsey, R. I., "Noise in Image Sensors," *Short Course Notes, Two-day short course presented at the Waterloo Institute for Computer Research*, May 1999.
- [Hubel] Hubel, P. M., Liu, J., Guttosch, R. J., "Spatial Frequency Response of Color Image Sensors: Bayer Color Filters and Foveon X3," *Foveon Technical Papers*.
- [Janesick 2002] Janesick, J., "Dueling Detectors: CCD or CMOS?," *SPIE OE Magazine*, February 2002.
- [Jefferies 2002] Jefferies, S. M., Lloyd-Hart, M., Hege, E. K., "Sensing wave-front amplitude and phase with phase diversity," *Applied Optics*, Vol. 41, pp. 2095-2102, 2002.

-
- [Jobin Yvon (Horiba) New Jersey, USA] Jobin Yvon (Horiba), "H20 Series Monochromators datasheet," New Jersey, USA.
<http://www.jobinyvon.com/jy/mono/images/h20.pdf>
- [Johansson 2002] Johansson, R., Lindgren, L., Melander, J., Moeller, B., "A Multi-Resolution 100 GOPS 4 Gpixels/s Programmable CMOS Image Sensor for Machine Vision," *2003 IEEE Workshop on CCDs & Advanced Image Sensors*, May 15-17 2002.
- [Johns 1997] Johns, D. A., Martin, K., "Analog Integrated Circuit Design," John Wiley and Sons, 1997.
- [Karadimitiou 1998] Karadimitiou, K., Tyler, J. M., "The centroid method for compressing sets of similar images," *Pattern Recognition Letters*, Vol. 19, pp. 585-593, 1998.
- [Kasap 2001] Kasap, S. O., "Optoelectronics and Photonics: Principles and Practices," Prentice-Hall Inc., 2001.
- [Keithley Instruments Inc. 2001] Keithley Instruments Inc., "Model 236/237/238 Source Measure Units Operator's Manual," *Revision E*, March 2001.
- [Kleinfelder 2001] Kleinfelder, S., Lim, S., Liu, X., Gamal, A. E., "A 10,000 Frames/s CMOS Digital Pixel Sensor," *IEEE Journal of Solid State Circuits*, Vol. 36, No. 12, December 2001.
- [Kudryashov 2002] Kudryashov, A. V., Panchenko, V. Y., Zavalova, V. Y., "Shack-Hartmann wavefront sensor for beam quality measurements," *Proceedings of SPIE*, Vol. 4900, pp. 331-338, 2002.
- [Kuo 1991] Kuo, J. B., Wong, E. J., Chou, T. L., "A BiCMOS image sensor circuit for pattern recognition neural network," *IEEE Transactions on Circuits and Systems*, Vol. 38, No. 12, pp. 1554-1556, 1991.
- [Lai 2002] Lai, L.-W., King, Y.-C., "A Novel Logarithmic Response CMOS Image Sensor With High Output Voltage Swing and In-pixel Fixed Pattern Noise Reduction," *IEEE Asia-Pacific Conference on ASIC*, 2002.
- [Lazaro 1988] Lazaro, J., Ryckebusch, S., Mahowald, M. A., Mead, C. A., "Winner-take-all networks of O(N) complexity," *Advances in Neural Information Processing Systems*, pp. 703-711, 1988.
- [Lee 2003 (Part I)] Lee, J. S., Hornsey, R., Renshaw, D., "Analysis of CMOS Photodiodes - Part I: Quantum Efficiency," *IEEE Transactions on Electron Device*, Vol. 50, No. 5, pp. 1233-1238, May 2003 (Part I).
- [Lee 2003 (Part II)] Lee, J. S., Hornsey, R., Renshaw, D., "Analysis of CMOS Photodiodes - Part II: Lateral Photoresponse," *IEEE Transactions on Electron Device*, Vol. 50, No. 5, pp. 1239-1245, May 2003 (Part II).
-

-
- [Litwiller 2001] Litwiller, D., "CCD vs. CMOS: Facts and Fiction," *Photonics Spectra*, January 2001.
- [Low 1998] Low, S. H., Maxemchuk, N. F., Lapone, A. M., "Document Identification for Copyright Protection using Centroid Detection," *IEEE Transactions on Communication*, Vol. 46, No. 3, pp. 372-383, March 1998.
- [Lule 2000] Lule, T., Benthien, S., Keller, H., Mutze, F., Rieve, P., Seibel, K., Sommer, M., Bohm, M., "Sensitivity of CMOS Based Imagers and Scaling Perspectives," *IEEE Transactions on Electron Devices*, Vol. 47, No. 11, pp. 2110-2122, November 2000.
- [Mäkynen 2000] Mäkynen, A., "Position-sensitive devices and sensor systems for optical tracking and displacement sensing applications," *PhD Thesis*, University of Oulu: Department of Electrical Engineering, Oulu, Finland, 2000.
- [Makynen 1998] Makynen, A., Rahkonen, T., Kostamovaara J., "A binary photodetector array for position sensing," *Sensors and Actuators A*, Vol. 65, pp. 45-53, 1998.
- [Mansell 1999] Mansell, J. D., Byer, R. L., "Adaptive Optics For LIGO," *LSC Meeting*, Spring 1999.
- [Mansell 2000] Mansell, J. D., Catrysse, P. B., Gustafson, E. K., Byer, R. L., "Silicon Deformable Mirrors and CMOS Wavefront Sensors," *Proceedings of SPIE*, Vol. 4124, 2000.
- [Marsh 2003] Marsh, P. N., Burns, D., Girkin, J. M., "Practical implementation of adaptive optics in multiphoton microscopy," *Optics Express*, Vol. 11, No. 10, 19 May 2003.
- [Maxim Integrated Products Inc. 1992] Maxim Integrated Products Inc., "MAX873, MAX875, MAX876 Low-Power, Low-Drift, +2.5V/+5V/+10V Precision Voltage References," *Maxim Integrated Products Datasheet, Rev. 1*, 1992.
- [Maxim Integrated Products Inc. 1994] Maxim Integrated Products Inc., "MAX667 +5V/Programmable Low-Dropout Voltage Regulator," *Maxim Integrated Products Datasheet, Rev. 3*, 1994.
- [Maxim Integrated Products Inc. 1996a] Maxim Integrated Products Inc., "DG417, DG418, DG419 Improved, SPST/SPDT Analog Switches," *Maxim Integrated Products Datasheet, Rev. 2*, 1996.
- [Maxim Integrated Products Inc. 1996b] Maxim Integrated Products Inc., "MAX660 CMOS Monolithic Voltage Converter," *Maxim Integrated Products Datasheet, Rev. 2*, 1996.

- [Maxim Integrated Products Inc. 1996c] Maxim Integrated Products Inc., "MAX4514, MAX4515 Low-Voltage, Low-On-Resistance, SPST, CMOS Analog Switches," *Maxim Integrated Products Datasheet*, 1996.
- [Maxim Integrated Products Inc. 1998] Maxim Integrated Products Inc., "MAX349, MAX350 Serially Controlled, Low-Voltage, 8-Channel/Dual 4-Channel Multiplexers," *Maxim Integrated Products Datasheet, Rev. 1*, 1998.
- [Maxim Integrated Products Inc. 2000] Maxim Integrated Products Inc., "MAX3222E, MAX3232E, MAX3237E, MAX3241E $\pm 15\text{kV}$ ESD-Protected, Down to 10nA, 3.0V to 5.5V, Up to 1Mbps, True RS-232 Transceivers," *Maxim Integrated Products Datasheet, Rev. 3a*, 2000.
- [Mendis 1997] Mendis, S. K., Kemeny, S. E., Gee, R. C., Pain, B., Staller, C. O., Kim, Q., Fossum E. R., "CMOS Active Pixel Image Sensors for Highly Integrated Imaging Systems," *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 2, pp. 187-197, February 1997.
- [Mentor Graphics Corporation 1998] Mentor Graphics Corporation, *Mentor Graphics C4 documentation and user manuals*, 1998.
- [Metrologic Instruments Inc. 2002] Metrologic Instruments Inc., "AOA Awarded Orders from LLNL for National Ignition Facility," *Press Release*, January 24, 2002.
- [Meynants 2001] Meynants, G., Dierickx, B., Uwaerts, D., Bogaerts, J., "Fixed pattern noise suppression by a differential readout chain for a radiation-tolerant image sensor," *IEEE Workshop on CCDs and AISs*, 2001.
- [Moini 1999] Moini, A., "Vision Chips," Kluwer Academic Publishers, October 1999.
- [Montera 1996] Montera, D. A., Welsh, B. M., Roggemann, M. C., "Use of artificial neural networks for Hartmann sensor lenslet centroid estimation," *Applied Optics*, Vol. 35, pp. 5747-5757, 1996.
- [Munro 1999] Munro, I., Paterson C., Dainty, J. C., "Low-cost adaptive optics breadboard system," ICO XVIII, San Francisco, 2-6 August 1999.
- [Neal 1993] Neal, D. R., O'Hern, T. J., Torczynski, J. R., Warren, M. E., Shul, R., "Wavefront sensors for optical diagnostics in fluid mechanics: applicaiton to heated flow, turbulence and droplet evaporation," *Proceedings of SPIE*, Vol. 2005, 1993.
- [Nirmaier 2003] Nirmaier, T., Pudasaini, G., Bille, J., "CMOS-based Hartmann-Shack Sensor for Real Time Adaptive Optics," *13th IEEE-NPSS Real Time Conference*, 2003.
- [Noll 1976] Noll, R. J., "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.*, Vol. 66, pp. 207-211, 1976.

-
- [O'Bryne 1999] O'Bryne, J. W., Fekete, P. W., Arnison, M. R., Zhao, H., Serrano, M., Philp, D., Sudiarta, W., Cogswell, C. J., "Adaptive Optics in Confocal Microscopy," *Adaptive Optics in Industry and Medicine*, Durham, UK, July 1999.
- [O'Byrne 1996] O'Byrne, J., "Sharper Eyes on the Sky," *Sky & Space magazine*, pp. 20-24, December 1996.
- [Olivier 1999] Olivier, S., "A New View of the Universe," *Science & Technology Review*, July/August 1999.
- [On-Trak Photonics] On-Trak Photonics, "PSD vs CCD," *Application Notes* <http://www.on-trak.com/appnote2.html>.
- [Pain 2003] Pain, B., "CMOS Devices," *Workshop on Innovative Designs for the Next Large Aperture UV/Optical Telescope (NHST)*, April 10-11, 2003.
- [Pain 2001] Pain, B., Hancock, B., Cunningham, T., "Radiation-Tolerant CMOS Image Sensors," *NASA Tech Brief NPO30185*, December 2001.
- [Pain 2000] Pain, B., Sun, C., Yang, G., "CMOS APS With Integrated Centroid-Computation Circuits," *NASA Tech Briefs*, Vol. 24, No. 9, September 2000.
- [Parhami 2000] Parhami, B., "Computer Arithmetic: Algorithms and Hardware Designs." New York: Oxford University Press, 2000.
- [Paterson 2000a] Paterson, C., Dainty, J. C., "Hybrid curvature and gradient wavefront sensor," *Optics Letters*, Vol. 25, No. 23, pp. 1687-1689, December 2000.
- [Paterson 2000b] Paterson, C., Munro, I., Dainty, J. C., "A low cost adaptive optics system using a membrane mirror," *Optics Express*, Vol. 6, No. 9, 2000.
- [Platt 2001] Platt, B. C., Shack, R., "History and Principles of Shack-Hartmann Wavefront Sensing," *Journal of Refractive Surgery*, Vol. 17, September/October 2001.
- [Pui 2004] Pui, B. H., Hayes-Gill, B. R., Clark, M., Somekh, M., See, C. W., Morgan, S., Ng, A., "Integration of a Photodiode Array and Centroid Processing on a Single CMOS Chip for a Real-Time Shack Hartmann Wavefront Sensor," *accepted for publication in IEEE Sensors Journal*, Vol. 4, No. 8, December 2004.
- [Pui 2002] Pui, B. H., Hayes-Gill, B. R., Clark, M., Somekh, M., See, C., Morgan, S., Ng, A., "The design and characterisation of an optical VLSI processor for real time centroid detection," *Analog Integrated Circuits and Signal Processing*, Vol. 32, No. 1, pp. 67-75, July 2002.

-
- [Rhoadarmer 1999] Rhoadarmer, T. A., McGuire, P. C., Hughes, J. M., Lloyd-Hart, M., Angel, J. R. P., Schaller, S., Kenworthy, M. A., "Laboratory Adaptive Optics System for Testing the Wavefront Sensor of the New MMT," *Proceedings of SPIE*, Vol. 3762, No. 161-173, 1999.
- [Ricquier 1995] Ricquier, N., Dierickx, B., "Active pixel CMOS image sensor with on-chip non-uniformity correction," *IEEE Workshop on CCDs and Advanced Image Sensors*, April 20-21, 1995.
- [Roddiier 1998] Roddiier, F., "Curvature Sensing: a new concept in adaptive optics," *Applied Optics*, Vol. 27, pp. 1223-1225, 1998.
- [Roska 1993] Roska, T., Chua, L. O., "The CNN universal machine: An analogic array computer," *IEEE Transactions on Circuits and Systems - Part II*, Vol. 40, No. 3, pp. 163-173, 1993.
- [Rullmann 2003] Rullmann, M., Schlußler, J.-U., Schüffny, R., "On-chip digital noise reduction for integrated CMOS Cameras," *Proceedings of the SPIE*, Vol. 5150, pp. 1620-1629, 2003.
- [Salmon] Salmon, T. O. "A Primer on Using Wavefront Analysis for Refractive Surgery and Other Ophthalmic Applications". Available at <http://www.opt.pacificu.edu/ce/catalog/10260-RS/WavefrontSalmon.html>
- [Sharman 2002] Sharman, P., "Position sensing with photodiodes," *Laser Focus World*, February 2002.
- [Shcherback 2002] Shcherback, I., Belenky, A., Yadid-Pecht, O., "Active Area Shape Influence on the Dark Current of CMOS Imagers," *Proceedings of SPIE*, Vol. 4669, pp. 117-124, April 2002.
- [Shcherback 2003] Shcherback, I., Yadid-Pecht, O., "CMOS APS Crosstalk Characterization Via a Unique Submicron Scanning System," *IEEE Transactions on Electron Devices*, Vol. 50, No. 9, pp. 1994-1997, September 2003.
- [Sheppard 1991] Sheppard, C. J. R., Gu, M., "Aberration compensation in confocal microscopy," *Applied Optics*, Vol. 30, pp. 3563-3568, 1991.
- [Simoni 1995] Simoni, A., Torelli, G., Maloberti, F., Sartori, A., Plevridis, S. E., Birbas A. N., "A single-chip optical sensor with analog memory for motion detection," *IEEE Journal of Solid State Circuits*, Vol. 30, No. 7, pp. 800-806, July 1995.
- [Standley 1991] Standley, D. L., "An Object Position and Orientation IC with Embedded Imager," *IEEE Journal of Solid-State Circuits*, Vol. 26, pp. 1853-1859, December 1991.

- [Stoppa 2002] Stoppa, D., Simoni, A., Gonzo, L., Gottardi, M., Dalla Betta, G.-F., "Novel CMOS image sensor with a 132-dB dynamic range," *IEEE Journal of Solid-State Circuits*, Vol. 37, No. 12, pp. 1846-1852, 2002.
- [Sze 1981] Sze, S. M., "Physics of Semiconductor Devices," 2nd ed: John Wiley & Sons, Inc, 1981.
- [Tanaka 1989] Tanaka, N., Hashimoto, S., Shinohara, M., Sugawa, S., Morishita, M., Matsumoto, S., Nakamura, Y., Ohmi, T., "A 310k pixel bipolar imager (BASIS)," 36th IEEE International Solid-State Circuits Conference, New York, USA, 1989.
- [Texas Instruments Inc. 1997] Texas Instruments Inc., "TLE202x, TLE202xA, TLE202xB, TLE202xY Excalibur High-Speed Low-Power Precision Operational Amplifiers," *Product Datasheet*, 1997.
- [Texas Instruments Inc. 2000] Texas Instruments Inc., "TLC227x, TLC227xA Advanced LinCMOS Rail-to-Rail Operational Amplifiers," *Product Datasheet*, 2000.
- [Theuwissen 1995] Theuwissen, A. J. P., "Solid-State Imaging with Charge-Coupled Devices," Kluwer Academic Publishers, March 1995.
- [Thibos 2003] Thibos, L. N., "Wavefront-guided contact lens design: Principles, techniques and limitations," *Optometry Today*, 24th January 2003.
- [Thompson 2002] Thompson, L. A., Teare, S. W., "Rayleigh Laser Guide Star Systems: Application to UnISIS," *Publications of the Astronomical Society of the Pacific*, Vol. 114, pp. 1029-1042, 2002.
- [Tian 2001] Tian, H., Fowler, B., El Gamal, A., "Analysis of Temporal Noise in CMOS Photodiode Active Pixel Sensor," *IEEE Journal of Solid State Circuits*, Vol. 36, No. 1, pp. 92-101, 2001.
- [Tonry 1997] Tonry, J., Burke, Barry E., Schechter, Paul L., "The Orthogonal Transfer CCD," *Publications of the Astronomical Society of the Pacific*, Vol. 109, pp. 1154-1164, 1997.
- [Turner 1994] Turner, R. M., Johnson, K. M., "CMOS Photodetectors for Correlation Peak Location," *IEEE Photonics Technology Letters*, Vol. 6, pp. 552-554, April 1994.
- [Tyson 1998] Tyson, R. K., "Principles of Adaptive Optics," 2nd ed: Academic Press, 1998.
- [Tyson 2000] Tyson, R. K., "Introduction to Adaptive Optics," in *Tutorial Texts in Optical Engineering*, vol. TT41: SPIE, 2000.
- [UDT Sensors Inc.] UDT Sensors Inc., "Standard Photodetector Catalog," *Hawthorne, California, USA*.

-
- [UDT Sensors Inc. 1982] UDT Sensors Inc., "Photodiode Characteristics and Applications," *2003 UDT Sensors Catalog*, April 1982.
- [Vdovin 1997] Vdovin, G., Sarro, P. M., Middelhoek, S., "Technology and applications of micromachined adaptive mirrors," *Optical Engineering*, Vol. 36, pp. 1382-1390, 1997.
- [Wang 1989] Wang, W., Busch-Vishniac, I.J., "The linearity and sensitivity of lateral effect position sensitive devices - an improved geometry," *IEEE Transactions on Electron Devices*, Vol. 36, No. 11, pp. 2475-2480, November 1989.
- [Watabe 2003] Watabe, T., Goto, M., Ohtake, H., Maruyama, H., Abe, M., Tanioka, K., Egami, N., "New signal readout method for ultrahigh-sensitivity CMOS image sensor," *IEEE Transactions on Electron Devices*, Vol. 50, No. 1, pp. 63-69, 2003.
- [Weckler 1967] Weckler, G. P., "Operation of p-n Junction Photodetectors in a Photon Flux Integrating Mode," *IEEE Journal of Solid State Circuits*, Vol. 2, No. 3, pp. 65-73, 1967.
- [Weyrauch 2002] Weyrauch, T., Vorontsov, M. A., Gowens, J., Bifano, T. G., "Fiber coupling with adaptive optics for free-space optical communication," *Proceedings of SPIE*, Vol. 4489, pp. 177-184, 2002.
- [Wohl 2003] Wohl, G., Parry, C., Kasper, E., Jutzi, M., Berroth, M., "SiGe pin-photodetectors integrated on silicon substrates for optical fiber links," *IEEE International Solid-State Circuits Conference 2003 (ISSCC 2003)*, Vol. 1, pp. 374-375, 2003.
- [Wong 1996] Wong, H.-S., "Technology and Device Scaling Considerations for CMOS Imagers," *IEEE Transactions on Electron Devices*, Vol. 43, No. 12, pp. 2131-2142, December 1996.
- [Xiangliang 2002] Xiangliang, J., Jie C., Yulin, Q., "Sensitivity and Photodetector Considerations for CMOS imager Sensor," Second Joint Symposium on Opto- & Microelectronic Devices and Circuits, Stuttgart, Germany, March 10-16, 2002.
- [Xilinx Inc.] Xilinx Inc., "Software manuals and documentation for Foundation Series 2.1i," http://support.xilinx.com/support/sw_manuals/2_1i/download/index.htm.
- [Xilinx Inc. 1999] Xilinx Inc., "The Programmable Logic Data Book," 1999.
- [Xu 2002] Xu, C., Ki, W.-H., Chan, M., "A low-voltage CMOS complementary active pixel sensor (CAPS) fabricated using a 0.25 μm CMOS technology," *IEEE Electron Device Letters*, Vol. 23, No. 7, pp. 398-400, 2002.
- [Yadid-Pecht 1999] Yadid-Pecht, O., "Wide-dynamic-range sensors," *Optical Engineering*, Vol. 38, No. 10, pp. 1650-1660, October 1999.
-

-
- [Yadid-Pecht 2003] Yadid-Pecht, O., Belenky, A., "In-Pixel Autoexposure CMOS APS," *IEEE Journal of Solid-State Circuits*, Vol. 38, No. 8, pp. 1-4, August 2003.
- [Yadid-Pecht 1997] Yadid-Pecht, O., Pain, B., Staller, C., Clark, C., Fossum, E., "CMOS active pixel sensor star tracker with regional electronic shutter," *IEEE Journal of Solid State Circuits*, Vol. 32, No. 2, pp. 285-288, 1997.
- [Yalamanchili 2001] Yalamanchili, S., "Introductory VHDL: From Simulation to Synthesis," Prentice Hall, 2001.
- [Yang 1996] Yang, D., Min, H., Fowler, B., El Gamal, A., Beiley, M., Cham, K., "Test Structures for Characterization and Comparative Analysis of CMOS Image Sensors," Proceedings of Advanced Focal Plane Array European Conference, Berlin, Germany, October 1996.
- [Yang 1994] Yang W., "A wide-dynamic-range, low power photosensor array," *IEEE ISSCC*, Vol. 37, 1994.
- [Yasuda 2003] Yasuda, T., Hamamoto, T., Aizawa, K., "Adaptive integration time image sensor with real time reconstruction function," *IEEE Transaction on Electron Devices*, Vol. 50, No. 1, pp. 111-120, 2003.
- [Zhu 1999] Zhu, L., Sun, P.-C., Bartsch, D.-U., Freeman, W. R., Fainman, Y., "Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation," *Applied Optics*, Vol. 38, No. 1, pp. 168-176, 1999.
- [Zimmermann 2000] Zimmermann, H., "Integrated Silicon Optoelectronics," Springer-Verlag, May 2000.

Appendix A1.1: Removal of data bottleneck in traditional wavefront sensors

In this section, the data bottleneck in current CCD AO systems is quantified and compared with our proposed system. Figure A1.1 and A1.2 shows the structure of our system and that of a single sensor, typically a CCD, system respectively. N_{light} is the number of bits used to represent the light level while $N_{centroid}$ is the number of bits that make up the centroid values and typically, $N_{light} > N_{centroid}$. For each centroid processing block two centroid values are obtained (x and y). Figure A1.3 shows the timing diagram associated with both. The off-chip centroid computation time is ignored, as it is possible to make use of the latency in the frame acquisition and readout, just as the parallel on-chip computation does, and to use several parallel processors/CPU/DSP units off-chip.

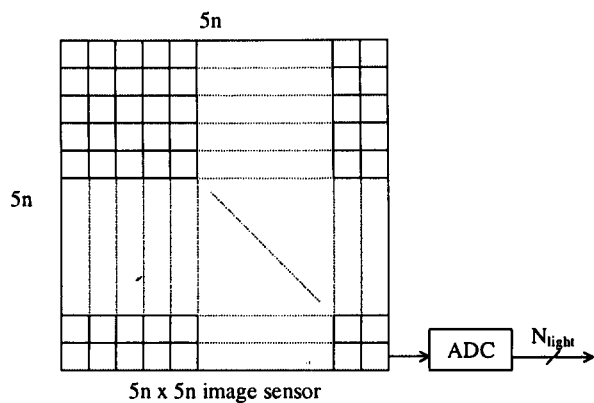


Figure A1.1 Traditional systems with conventional CCD readout architecture

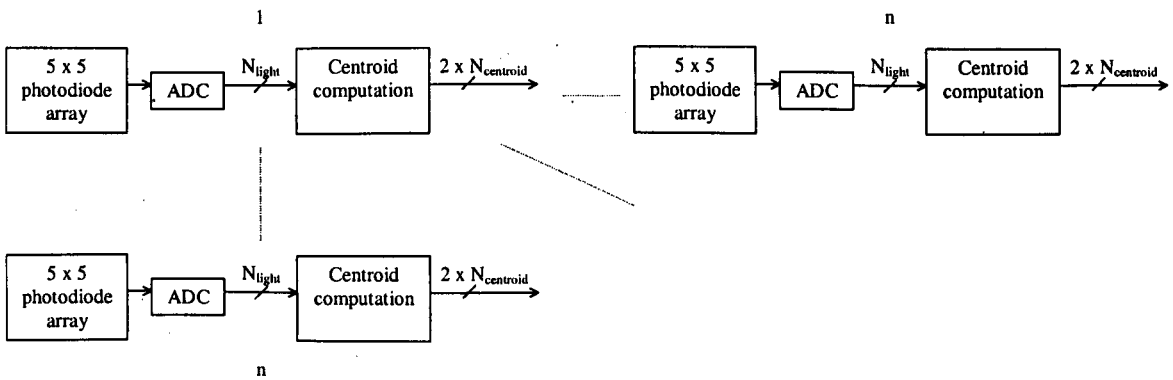


Figure A1.2 Our proposed system with parallel centroid processing

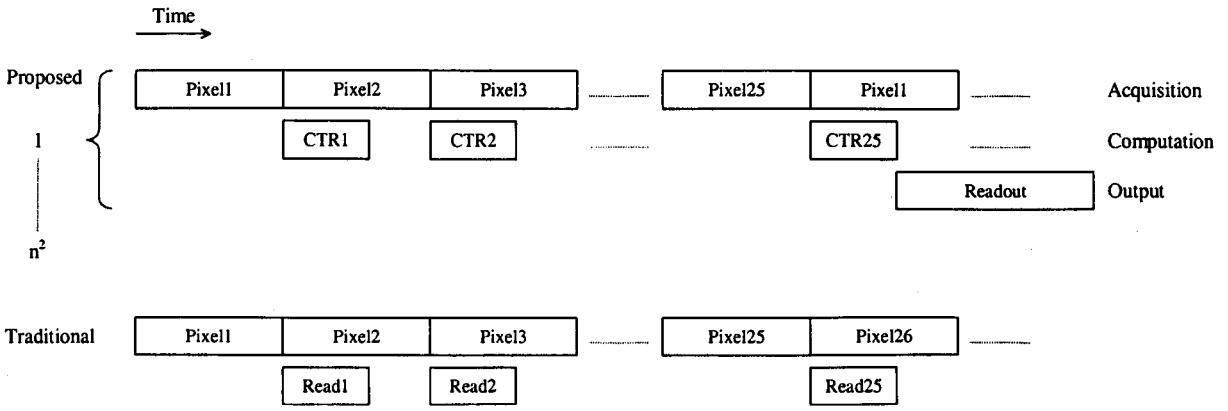


Figure A1.3 Timing diagrams of our proposed system and the traditional system

The acquisition time specified includes the pixel integration period and the ADC acquisition time. For long integration periods, the frame rate is limited by the acquisition time, and in this case the frame period for our system, T_1 , and of the traditional system, T_2 , is given by:

$$T_1 = 25T_{pixel}$$

$$T_2 = 25n^2T_{pixel}$$

For short integration times and fast digitisation, the frame rate is limited by the off-chip readout time⁶², and in this case:

$$T_1 = 2N_{centroid} \times n^2 \times T_{read}$$

$$T_2 = N_{light} \times 25n^2 \times T_{read}$$

In summary, for long integration times, our system removes the data bottleneck by allowing parallel acquisition of the raw data, while for short integration times, the bottleneck is removed by processing the raw data on-chip and only transmitting reduced bandwidth data off-chip. Also, the speed advantage offered increases with the square of the number of subapertures, n , in the system.

⁶² Typically guard row and column pixels are needed to avoid optical crosstalk when a CCD is used. As such the array size and hence, the readout time of the CCD system is larger than that assumed.

Appendix A2.1: Alcatel Microelectronics (Mietec) 0.7 μ m CMOS Process Parameters

This process is a self-aligned twin-well CMOS process with n+doped polysilicon gate. Several key process and electrical parameters are highlighted here.

Electrical Parameters:

Layer Thickness

Layer	Thickness (μ m)
n-well	2.0
p-epilayer	15.0 - 18.7
field oxide	0.45
gate oxide	0.0175

Resistivity

Layer	Resistivity/Doping levels	
	Sheet resistance (Ω /sq) ⁶³	Resistivity (Ω cm)
n-well	1300	-
p+	96	-
n+	67.5	-
Poly	27	-
Metal1	0.050	-
Metal2	0.035	-
p-epilayer	-	27.2 - 40.8
substrate ⁶⁴	-	0.01 - 0.02

⁶³ Sheet resistance, $R_s = \frac{\rho}{t}$ (Ω /sq), where ρ is the resistivity (Ω cm) and t is the thickness (cm)

⁶⁴ p-epilayer is also denoted as p-substrate.

Junction diodes

Junction type	Junction capacitance		Leakage current		Breakdown voltage (V)
	C_j (pF/ μm^2)	i_a (fA/ μm^2)	i_{pf} (fA/ μm)	i_{pf} (fA/ μm)	
p+/n-well	6.0×10^{-4}	3.6×10^{-10}	1.1	0.04	13.3
n+/p-well	5.0×10^{-4}	2.8×10^{-10}	0.13	0.37	14
n-well/p-substrate	7.8901×10^{-5}	7.3315×10^{-10}	1.1	4.1	59

Transistor Parameters

Parameters	NMOS	PMOS
Gate oxide thickness (nm)	17.5	17.5
Threshold voltage (V)	0.75	-1.0
Transconductance ($\mu\text{A}/\text{V}^2$)	95	30

Diode Models:

Shallow n+/p-well junction

.MODEL DNPLUS D IS=3E-7 ISW=6E-11 CJO=5E-4 M=0.35 CSO=2.8E-10
MS=0.21 VJ=0.8

Shallow p+/n-well junction

.MODEL DPPLUS D IS=2E-8 ISW=7E-11 CJO=6.0E-4 M=0.51 CSO=3.6E-10
MS=0.35 VJ=0.8

Deep n-well/p-substrate junction

.MODEL Djunc D IS=1E-15 CJ=7.8901E-5 MJ=0.27412 PB=0.42842
CJSW=7.3315E-10 MJSW=0.25301 +FC=0.99232

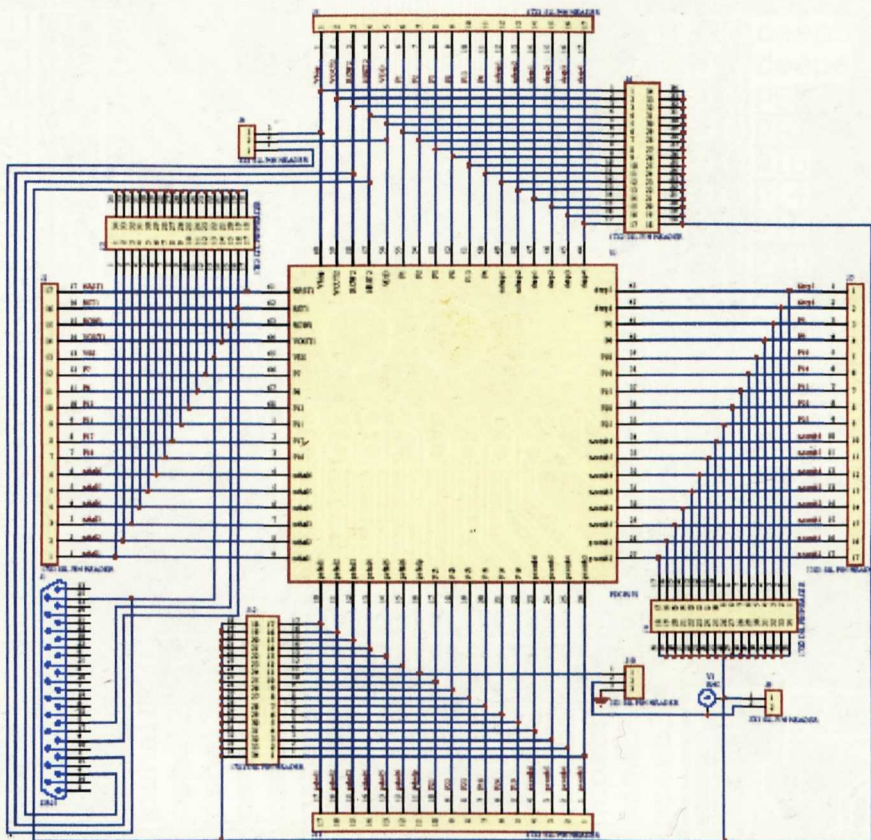
Digital Logic:

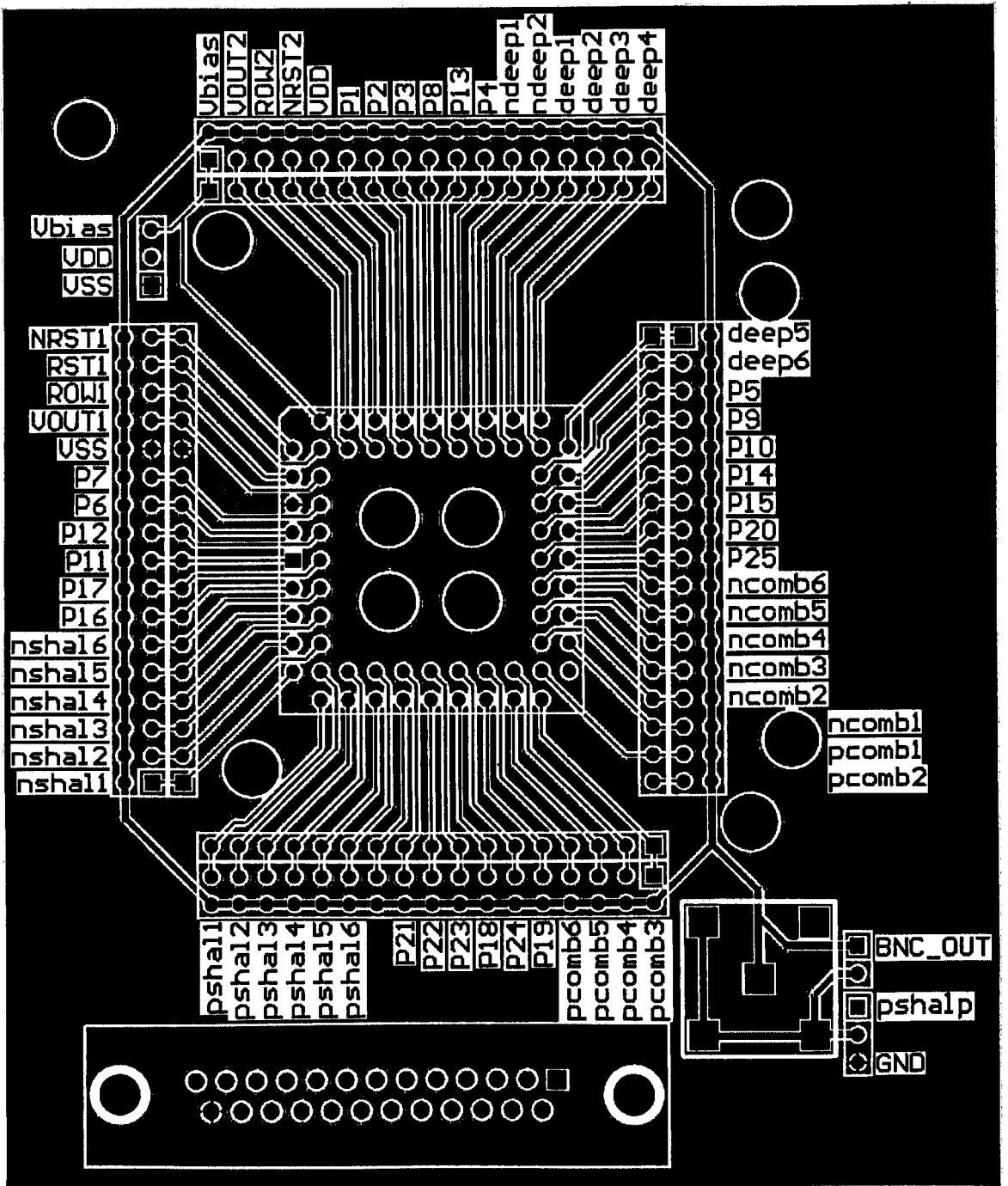
System speed up to 80 MHz

Power: 3.2 μW /gate/MHz at 5 V

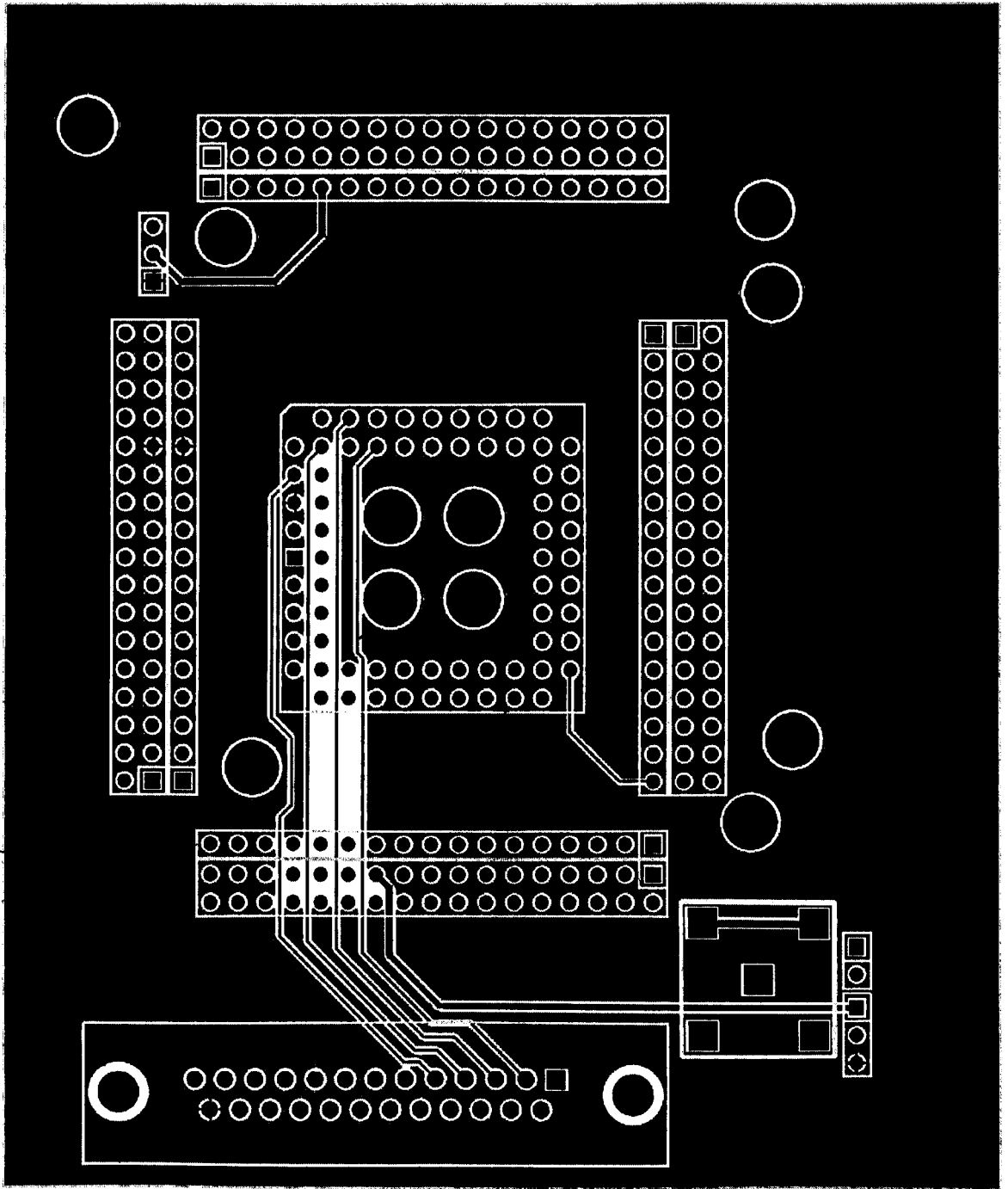
Density: 1250 gates/ mm^2 (incl. routing, typical density for 20,000 gates design)

Appendix A2.2: Schematic and PCB layout of test board for laser scanning experiment

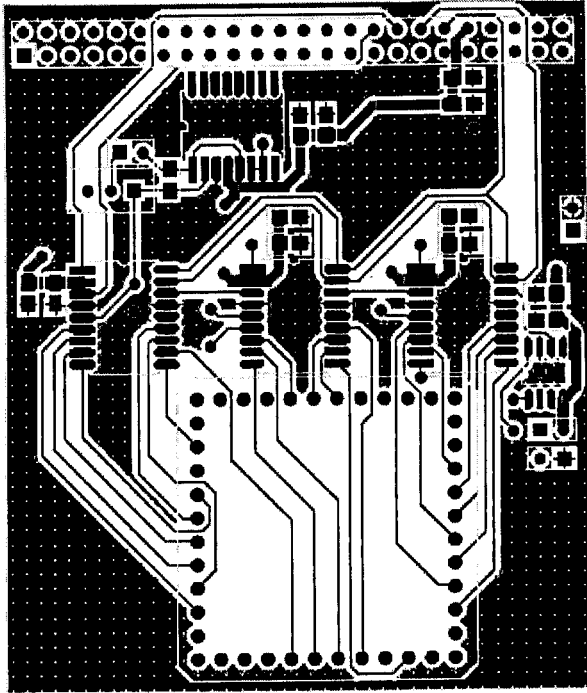




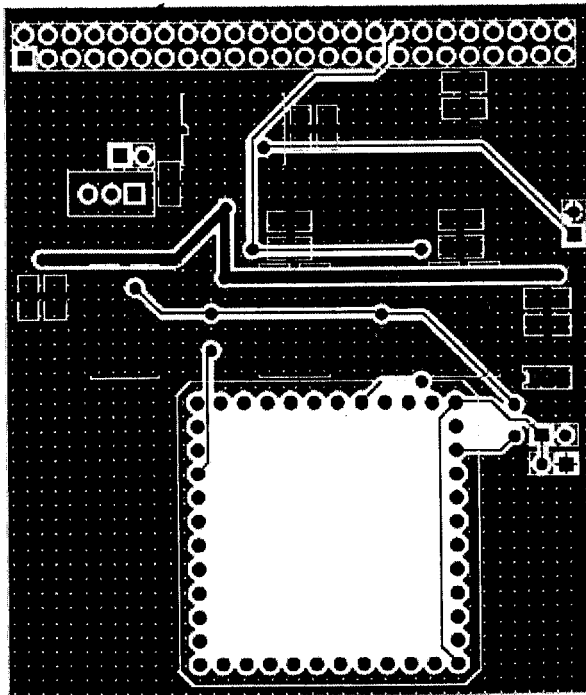
Top layer



Bottom layer

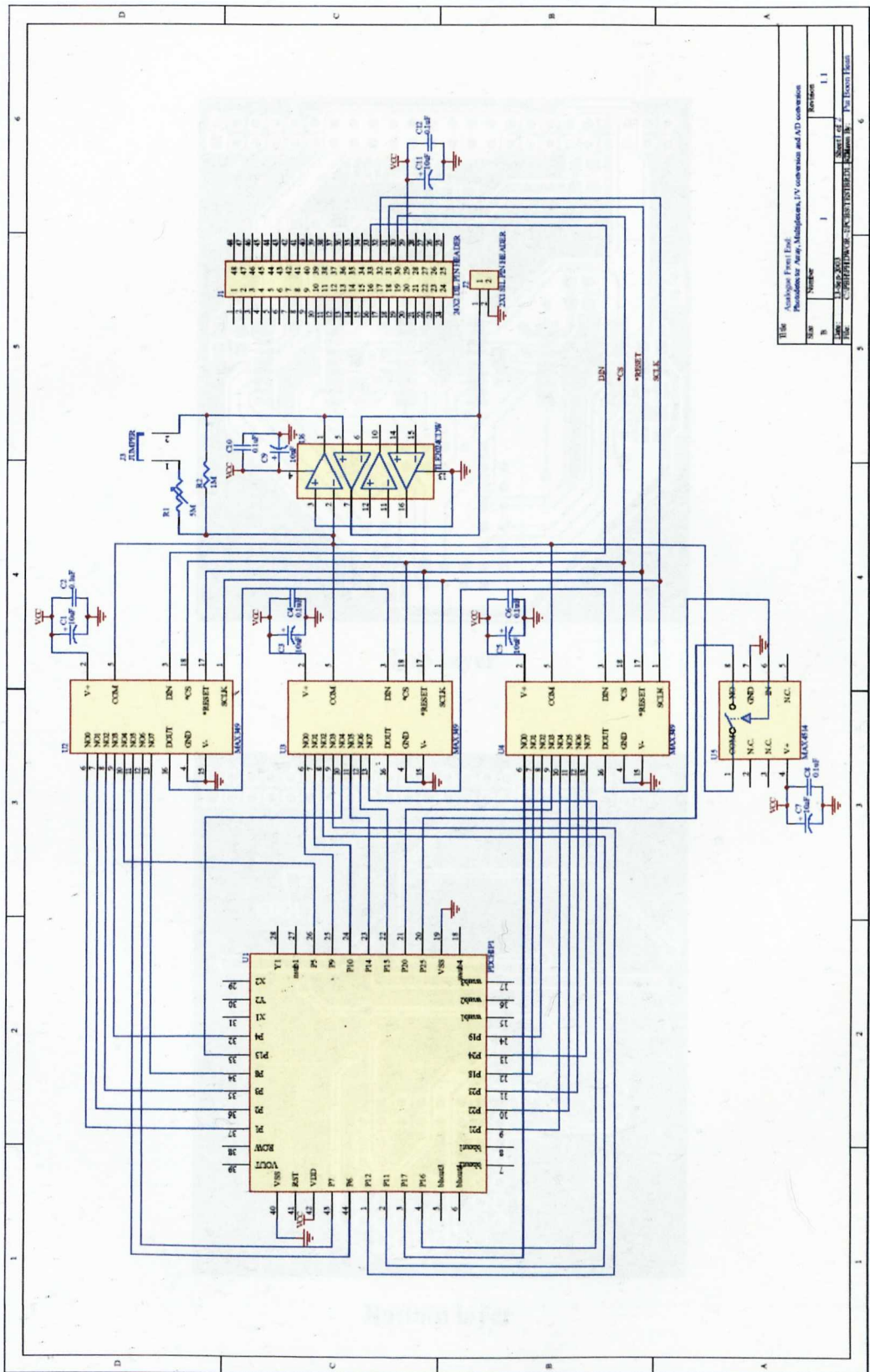


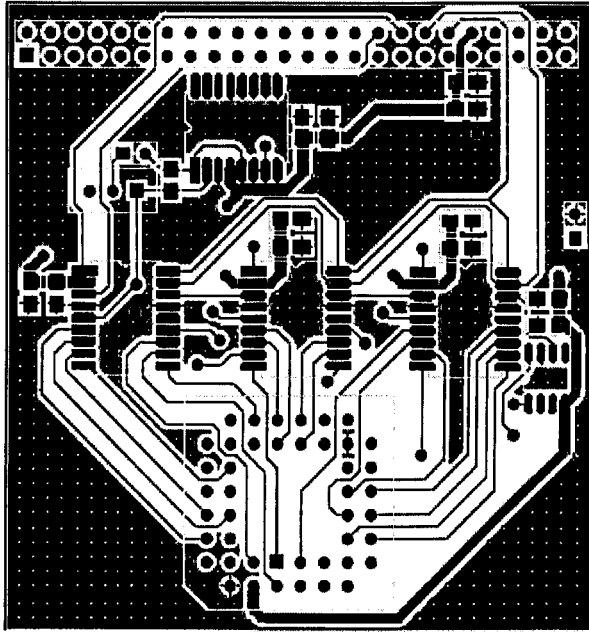
Top layer



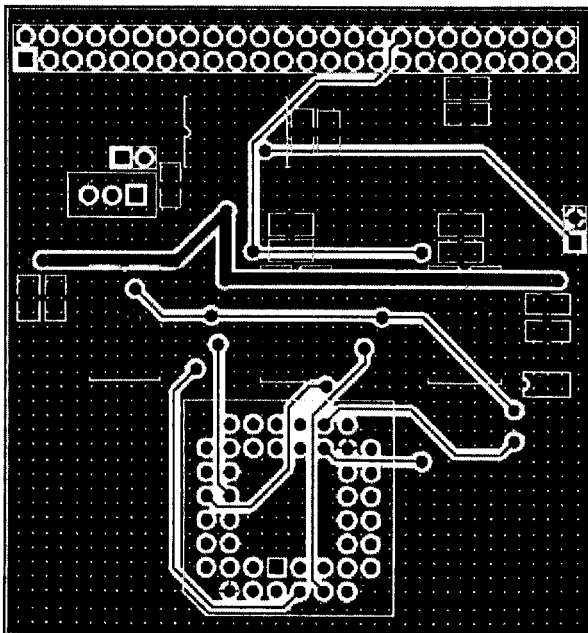
Bottom layer

Appendix A3.2: Schematic and layout of optical front end of the hardware emulation system with a full custom photodetector array

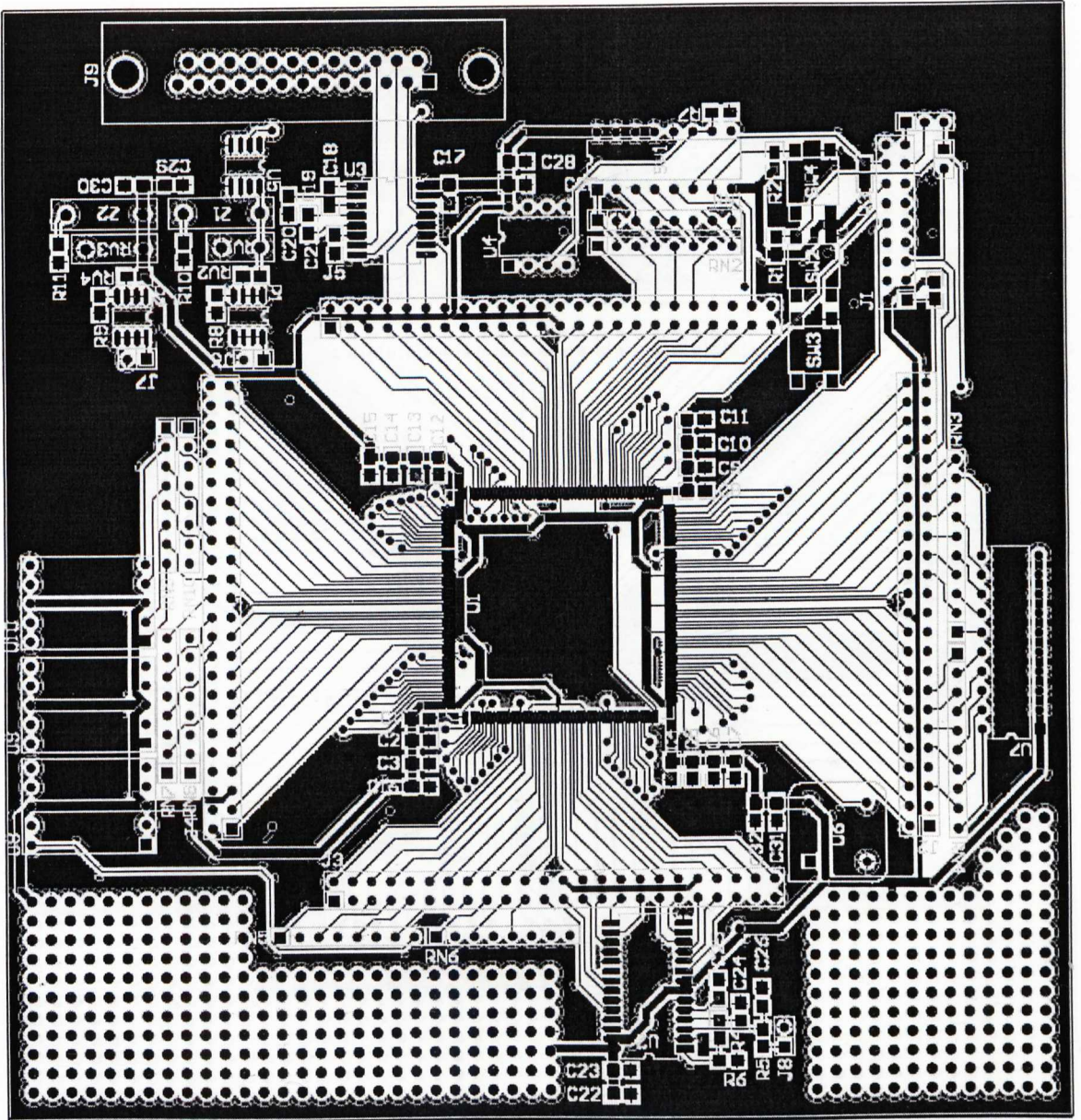




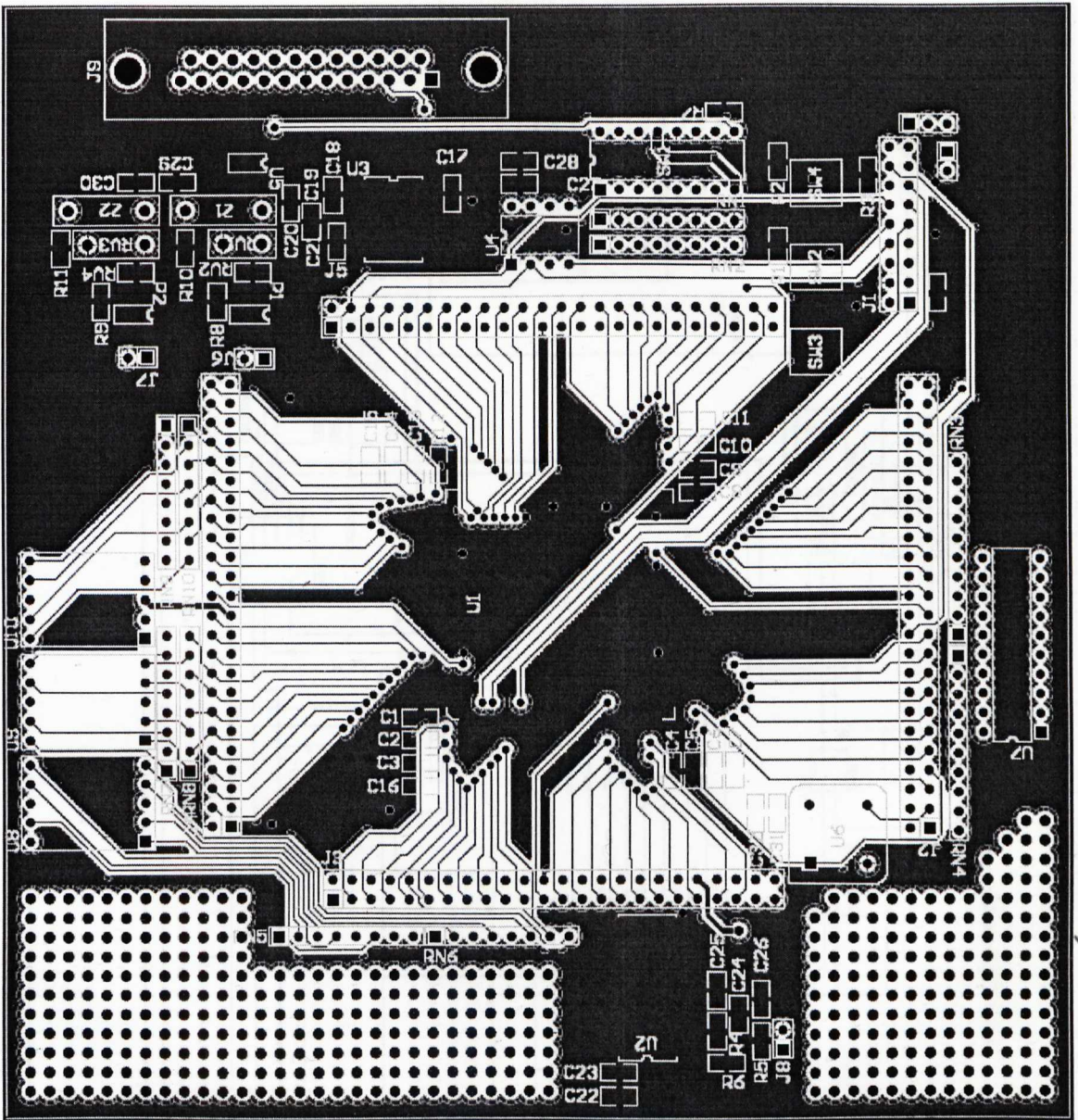
Top layer



Bottom layer

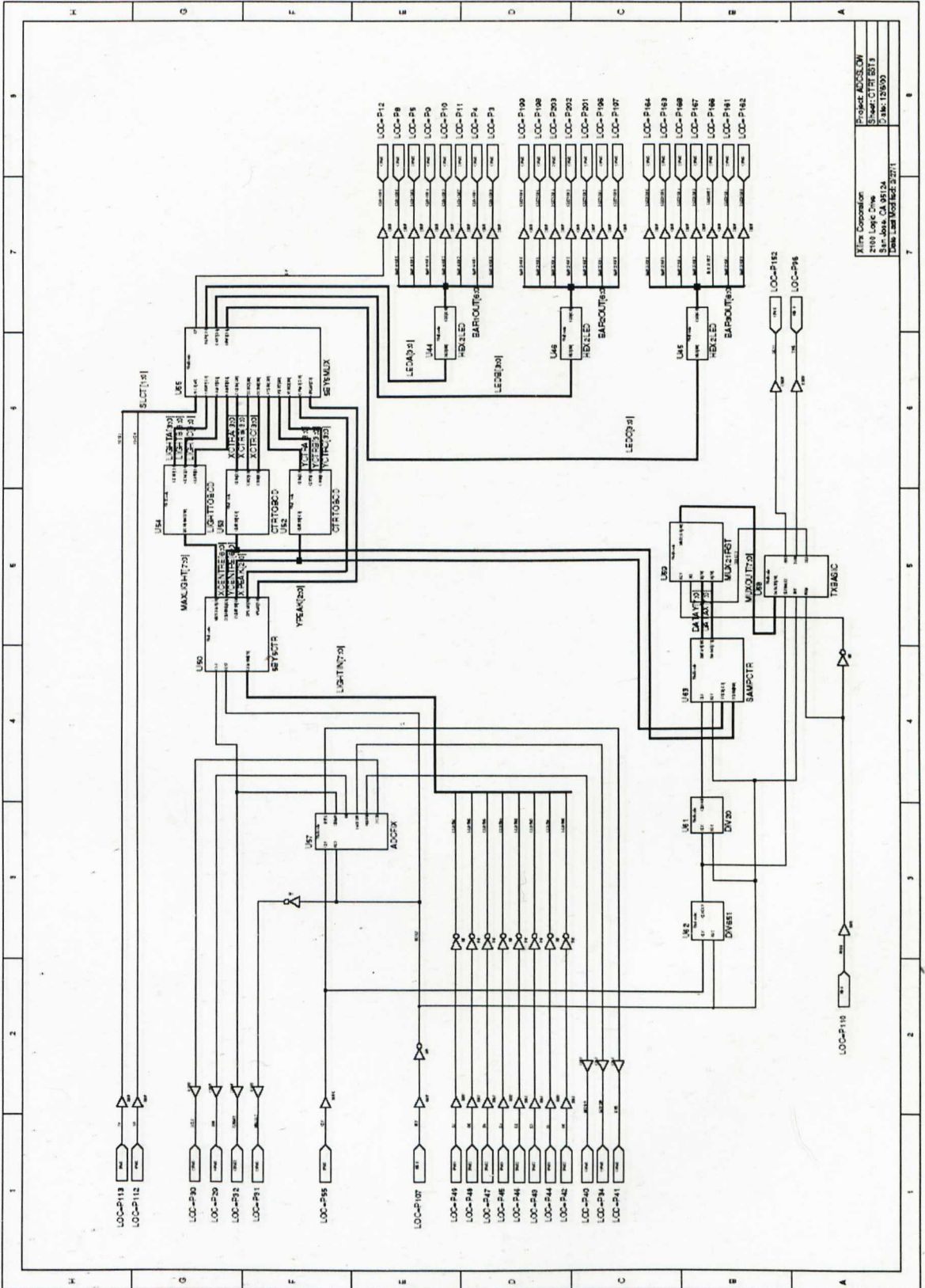


Top layer

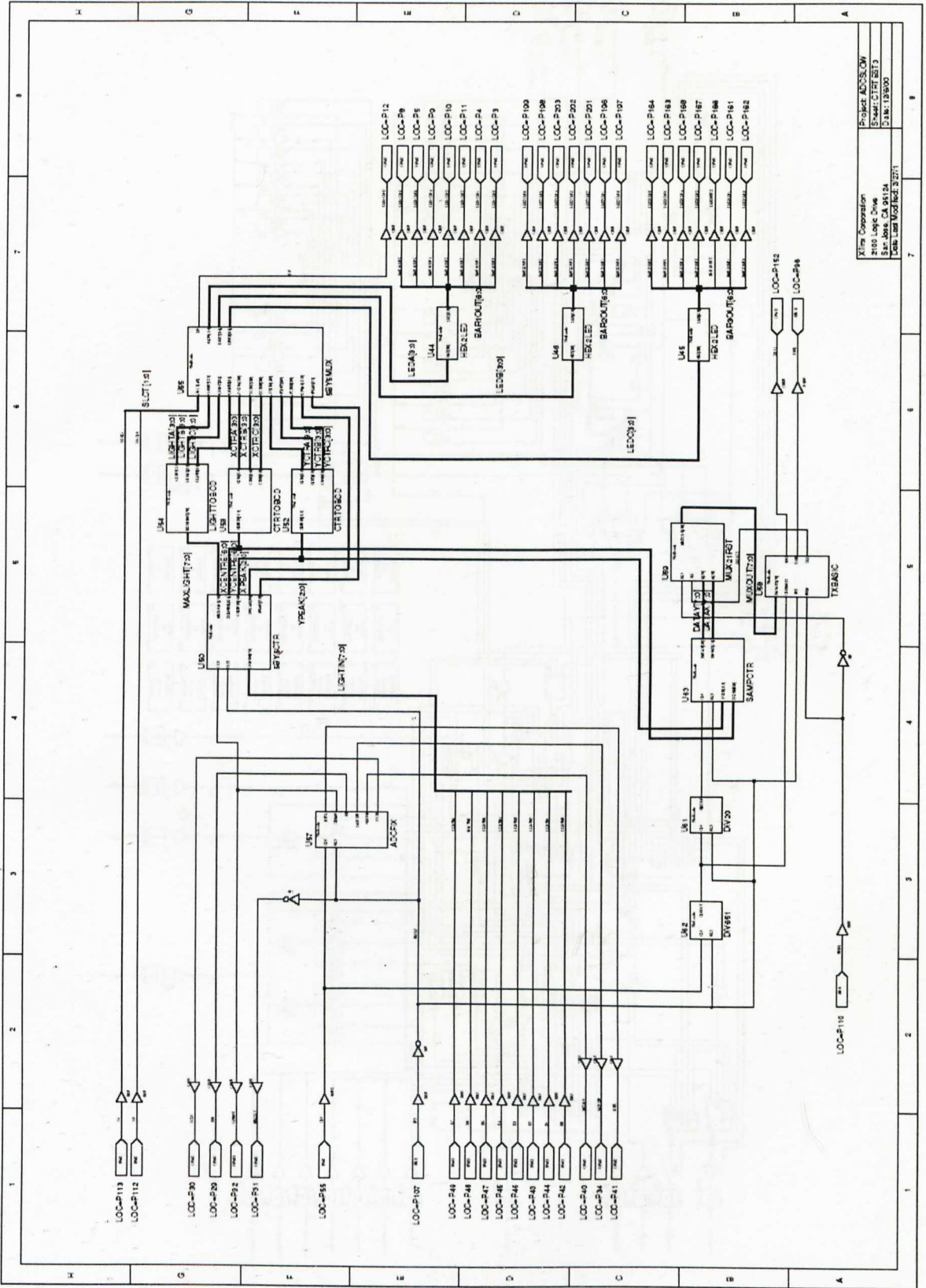


Bottom layer

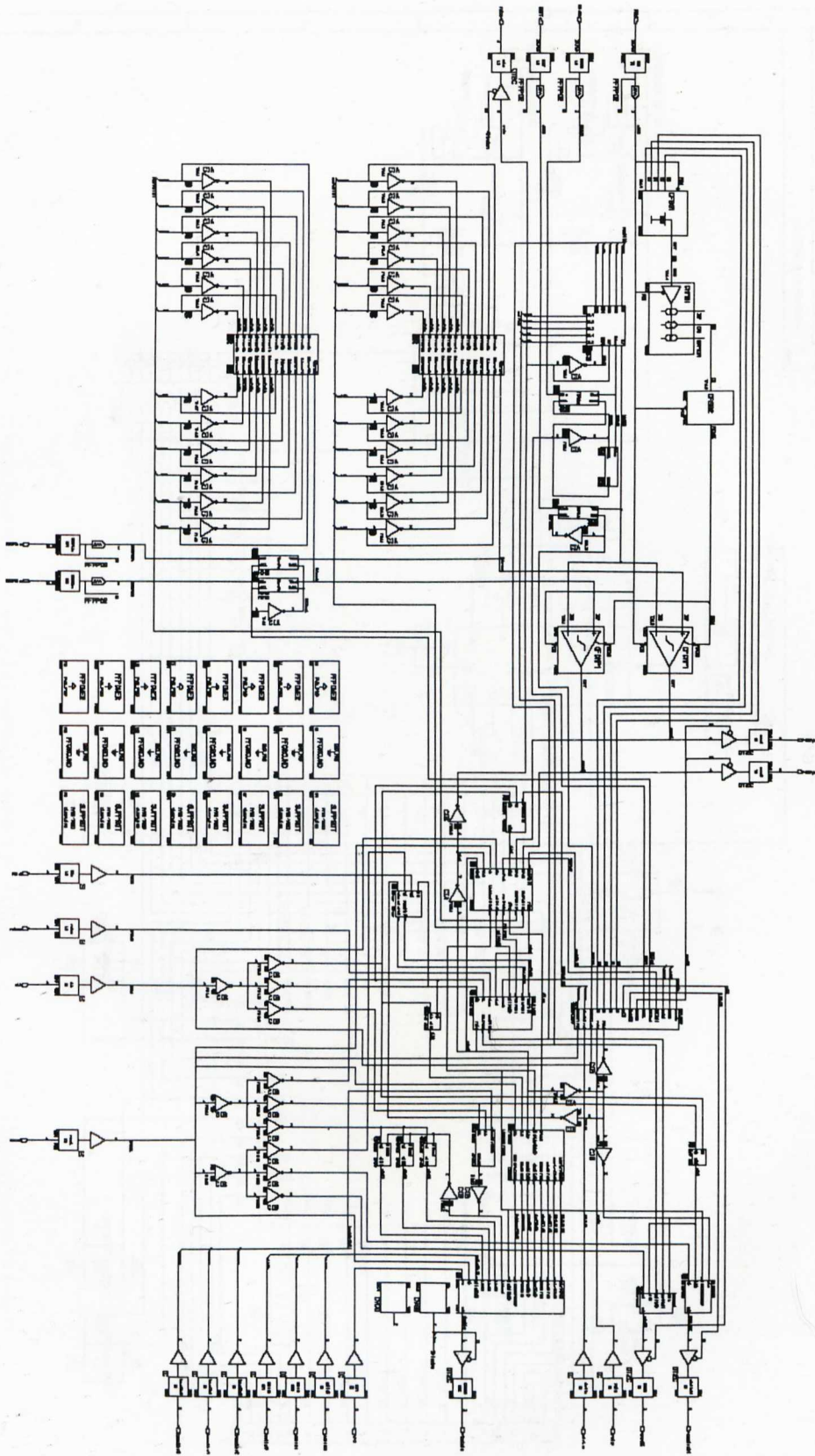
Appendix A3.4: Schematic of FPGA centroid processor for commercial photodetector array front end



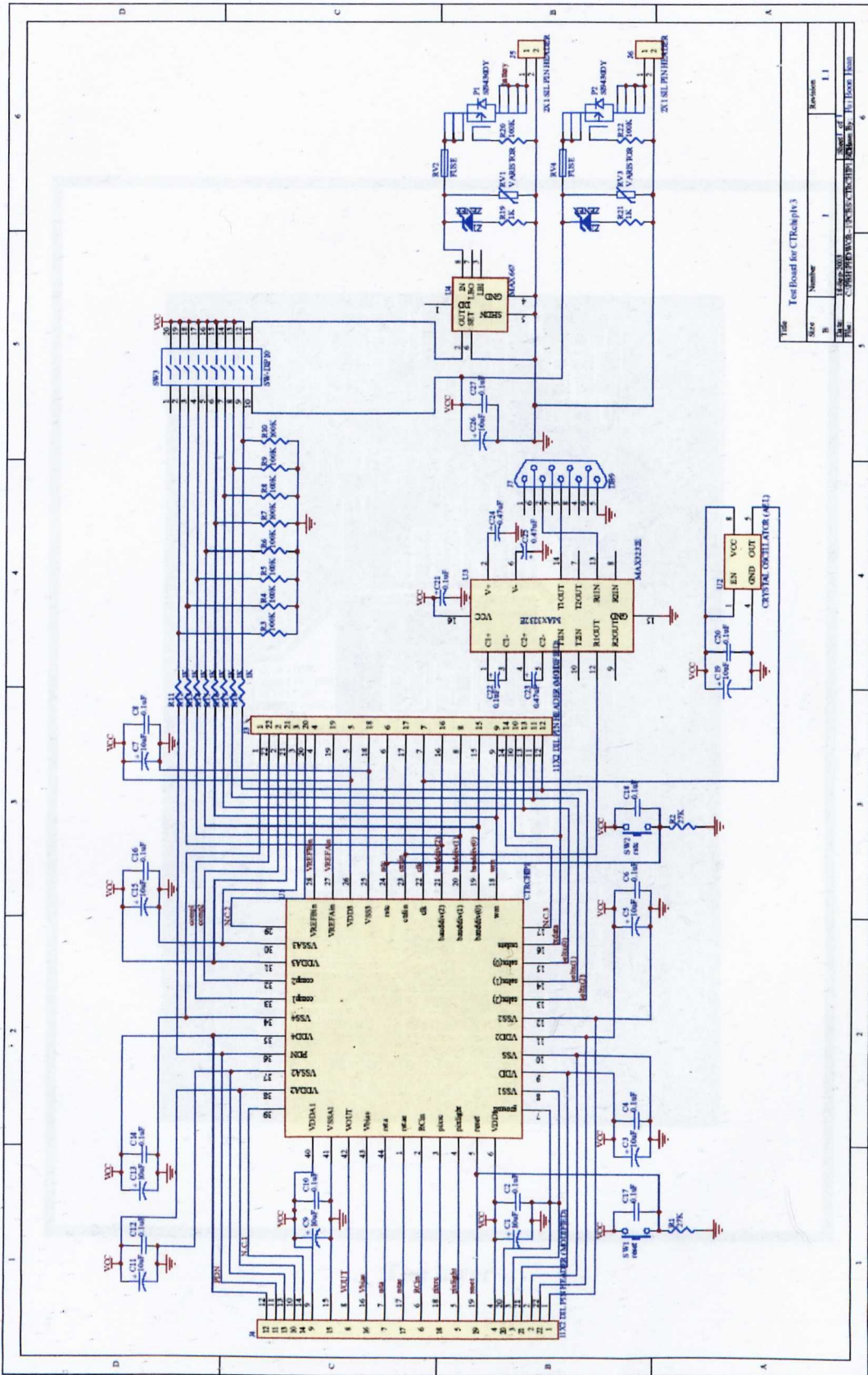
Appendix A3.5: Schematic of FPGA centroid processor for full custom photodetector array front end



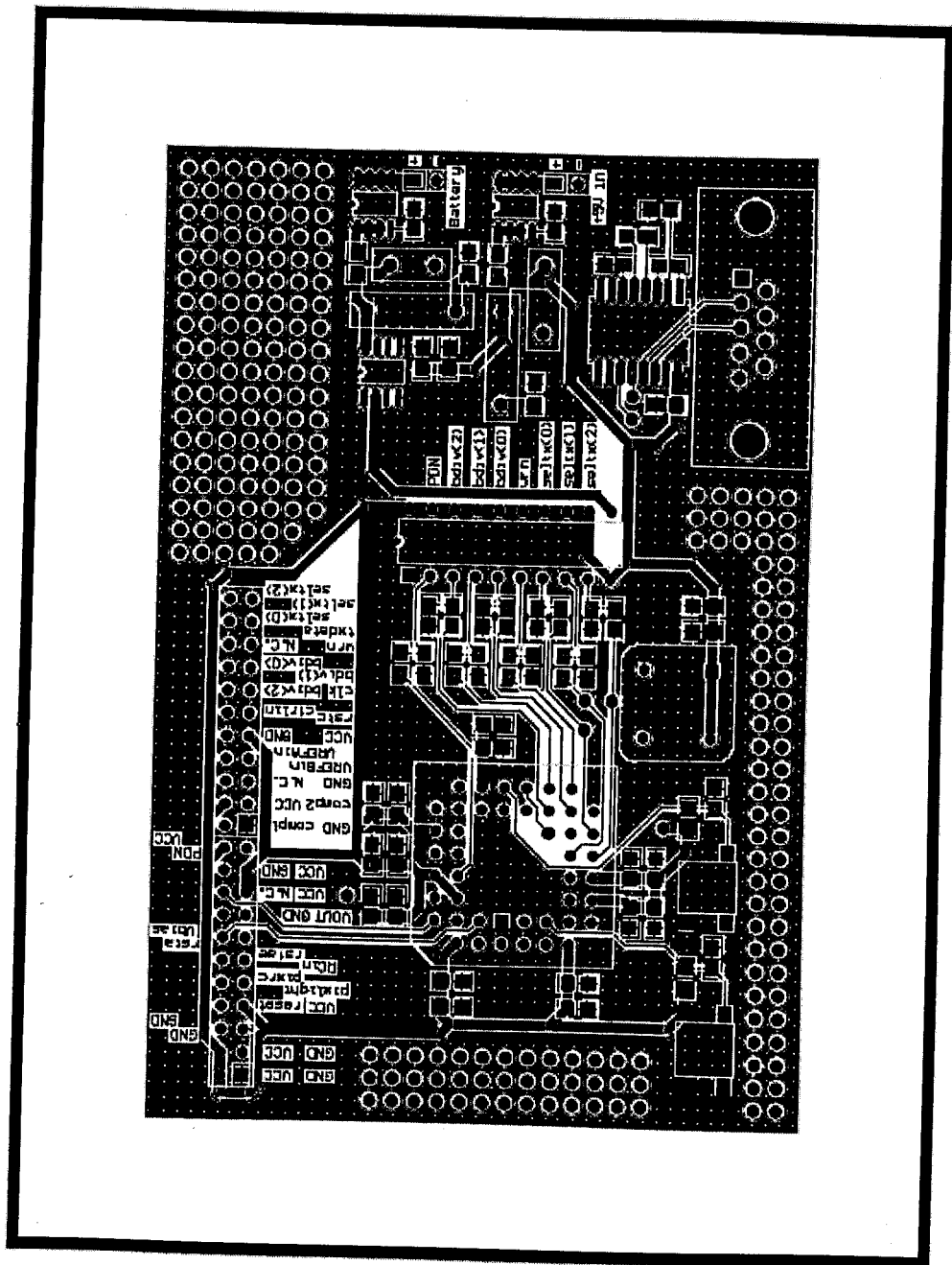
Appendix A4.1: Top level schematic of ASIC centroid processor



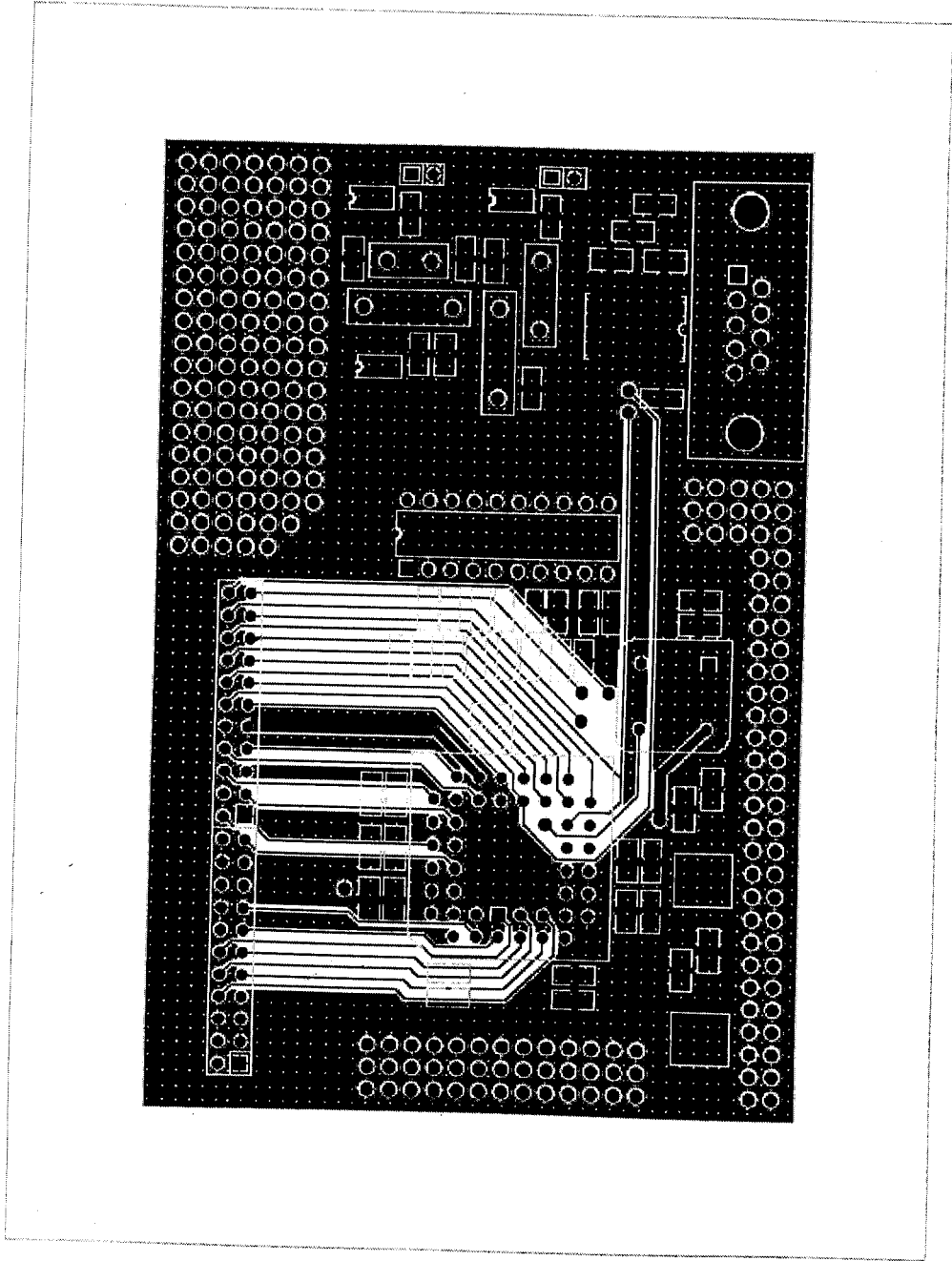
Appendix A4.2: Schematic and layout of ASIC centroid processor test board



Title				Test Board for CTRchip3			
Size	Number	Revision					
B	1	1.1					
Date	15 May 2003	Sheet of	6				
File	C:\PROJECTS\TR-PCB\TR-PCB.PCB	Drawn by	JL				



Top layer



Bottom layer

Appendix A4.3: Power consumption calculation of digital circuitry in ASIC

Block name	Component types												
	CAND2	CAO121	CAO1211	CAO122	CFD2	CFD2S	CFD3	CFD3S	CIA	CM2I11	CNAND2	CNAND3	CNAND4
POW(μ W/MHz)	3.4614	1.4067	3.6269	1.6254	11.7442	13.209	12.1246	13.5925	0.8878	1.726	1.3169	1.5593	1.9488
clkdiv1		2			4				5	1		1	
CTR5by5a	7	11	17	71	4	242		3	364	21	114	28	4
CTRa2d	3	12	2	46	32	37	3	12	98	8	38	10	4
ctrlrevr	1	1	7		13	49	4	3	49		8	2	1
div2					1				1				
div8					4				1		1		
div10				1	4				2		2	1	
div20		2	1		5				4		1	1	1
div30					5				3		3	1	
divbaud	2	4	10		14				21		6	2	1
opctrlsyn						8		3	4				
RCrevr		2	1		13	21	4	2	21		2		
txall	1	5	2	49	13	30	2		47	2	31	8	12
txlightout		4	1	1	9	12	2		20	3	14	3	1
txrowcol		3	2	1	10	10	2		20	2	12	2	2

Component types

Block name	1.933	CNAND5	1.0931	CNOR2	1.5292	CNOR3	1.6801	CNOR4	1.9282	CNOR5	1.8172	COA121	1.8823	COA1211	1.3916	COA122	2.9875	COR2	1.2178	CXNOR2	3.7884	CXOR2	CVDD	CVSS
clkdiv1		2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CTR5by5a	1	119	4	2	1	76	2	77	28	232	82	1												
CTRa2d		12	15	3	2	36	23	7	4	4	7	1												1
ctrlrcvr		5	3	3	3	3	1	1	5	3	2	1												
div2																								
div8		1	1																					
div10		1																						
div20		1																						
div30																								
divbaud		8	1	3	12	2																		
opctrlsyn																								
RCrcvr		5	1	1		2		1																
txall		19	7	2	2	10	1	2	1	1	3	1												
txlightout		1	2	1	2	16		1	1	2	1	1												
txrowcol		2	2	1	1	16		1	1	1	3	1												

Block name	POW(μ W/MHz)	Frequency, f (MHz)	Power Consumption (mW) at frequency, f (MHz)	No. of Instances	No. of Gates	Critical delay (ns)	path
clkdiv1	72.9231	32	2.3335392	21	47	6.46	
CTR5by5a	5131.3544	16	82.1016704	1511	3641	31.37	
CTRa2d	1554.306	32	49.737792	420	1028	17.45	
ctrlcrevr	1029.2836	32	32.9370752	165	625	13.2	
div2	12.632	32	0.404224	2	7	2.37	
div8	58.0278	32	1.8568896	11	35	5.02	
div10	60.6702	0.1152	0.006989207	13	37	4.86	
div20	79.6368	0.1152	0.009174159	18	48	5.51	
div30	74.3362	0.1152	0.00856353	16	46	5.69	
divbaud	307.2977	32	9.8335264	94	191	13.76	
opctrlsyn	150.0007	32	4.8000224	16	92	2.05	
RCcrevr	554.6075	32	17.74744	81	335	7.56	
txall	869.5475	0.1152	0.100171872	251	590	11.02	
txlightout	399.6852	4	1.5987408	99	256	8.84	
txrowcol	381.9001	4	1.5276004	93	239	8.5	
Total				2811	7217		