The University of Nottingham

Design and Implementation of an Integrating Modulated Light Camera

Roger Alan Light, M.Eng (Hons)

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

June 2008

Abstract

In optical experiments, replacing the usual constant intensity light source with one that has its intensity modulated at a fixed frequency gives a number of advantages, namely moving the signal of interest away from low frequency noise, allowing the signal to be detected even in the presence of background illumination, as well as being able to gain more information about the experiment by measuring any change in the phase of the modulation.

A number of different types of integrating pixels have been simulated in Matlab with regard to their suitability for use as a modulated light detector, along with different methods for reconstructing the signal. The quadrature method of reconstruction was chosen as the best technique, where four samples are taken per modulated time period then all four samples are used in the I and Q channels of the demodulator.

A modulated light detector has been implemented using a custom integrated circuit in a standard 0.35 μm CMOS process, linked to a field programmable gate array. The custom circuits consists of a photodiode configured as an integrating type pixel, where the output of the pixel is connected to a comparator so that when coupled with a counter, the pixel can measure the amount of time taken for the photodiode to reach a particular voltage.

A camera with 128 x 128 pixels has been implemented and characterised using a modulated laser as the input, with the results obtained compared to the Matlab simulations carried out. Images have been taken with the camera that show the

camera is able to detect modulated light signals at a modulation depth of 1.39% and modulation frequency of 420 Hz. The camera has also been operated successfully for a modulation frequency of 2.5 kHz.

Acknowledgements

I would like to sincerely thank my supervisors Dr. Barrie Hayes-Gill and Dr. John Crowe for their continued support and guidance along the course of this work.

I would also like to thank Ian Stockford for his invaluable help on setting up optical experiments, John Himsworth for his help in obtaining some of the comparator results in Chapter 6 and all of my other colleagues who have helped in other ways, especially those who have hassled me about getting this finished.

This work has been financially supported by the Engineering and Physical Sciences Research Council (EPSRC), UK and the School of Electrical and Electronic Engineering at the University of Nottingham.

Finally, I would like to thank my friends for their support and for putting up with me during the more difficult parts of this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

This chapter describes the beginnings of photography and how it has developed to the present day whilst also introducing the idea of a camera that can measure modulated light. It also discusses the way in which optical detection is achieved on electronic chips with the advantages and disadvantages of each method. Lastly, the different noise sources that affect cameras are considered.

1.2 Project Aims

The motivation behind this work was to create a new type of modulated light camera using digital processing, capable of distinguishing both the amplitude and phase of an input light signal independently of any other light present. The focus was on having a fairly simple pixel which could be kept small so that the camera would have a high spatial resolution and a large array could be fabricated without requiring a large and expensive chip.

1.3 A Brief History of Photography

Since Louis Daguerre invented the first practical process for capturing an image in the late 1830s, the ability to record the visual world has had an enormous impact in every walk of life. Photography advanced quickly, with the wet plate process developed in 1850 by Frederick Scott Archer allowing exposure times to be reduced significantly, although with the disadvantage that the plates had to be created by the photographer. In 1855 the dry plate was invented, allowing plates to be commercially manufactured instead and then in 1884 the first film negatives were developed, simplifying the photographer's job even further.

The first colour photograph was made in 1861 by James Clerk Maxwell of a tartan ribbon, as can be seen in Figure 1.1, although it would not be until 1907 that colour plates would be commercially available.

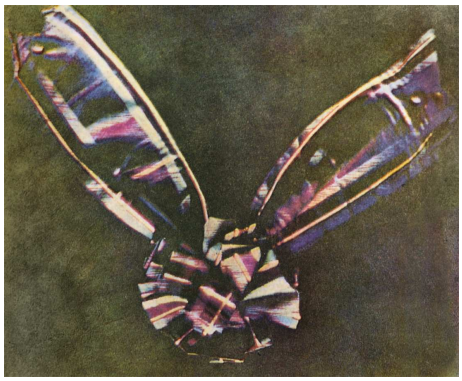


Figure 1.1: First Colour Photo, by James Clerk Maxwell

In 1935, Kodak released its first colour film which used essentially the same technology in use in photographic films today. After these developments, cameras and film technology continued using the same ideas until 1981, when Sony developed a consumer camera using a charge coupled device (CCD) instead of film. This heralded the age of the digital camera.

Digital cameras use a CCD or photodiode array to detect light, whereas black and white film photography uses silver-halide crystals and colour film photography relies on coloured emulsions. Digital cameras offer the advantage over film cameras

that the image is immediately available, can be easily transmitted and can be post processed with ease. As they do not require film, digital cameras can be made to be very compact if required and are now largely ubiquitous, with most mobile phones including a camera of some description and dedicated digital cameras taking over from film cameras in most applications.

Cameras are not limited to taking single still images - films have been made since celluloid replaced the photographic plate. As well as capturing everyday life, films allow the world to be viewed differently through the use of high speed photography and time lapse photography. In the former, film is shot at a high rate and then played back at a much lower rate which allows details to be seen that would otherwise not be possible. One simple application of this was first used in 1878 to determine whether a horse had all of its feet off the ground during a gallop. Time lapse photography on the other hand takes pictures at a very slow rate and then plays them back at a normal rate to show the passage of time as a plant grows for instance.

In all of these cases, the camera is recording the instantaneous intensity of the light of a scene as we see it with our own eyes, even if it does give us more information than we would otherwise see. Making cameras with integrated circuits allows processing to happen at or very close to where the light is detected and so measure effects that would not be able to be seen in any other way. For example, in a typical experiment, light will be shone onto a sample which affects the light in some way and it is this change in the light that should be measured. Unfortunately, the signal is often so small that it is swamped by background light. In some cases the experiment could be carried out in the dark, but for the most part this is undesirable, and in many cases it is not possible to remove the background light as it is inherent to the experiment. However, if the light that is input to the experiment is modulated then it is possible to extract the signal by locking in to the frequency of the input and only considering frequencies that are near by. An image taken by a normal camera has only one parameter which is intensity. A modulated light signal at particular frequency has three quantities that can be measured, the offset (which equates to

the light intensity as measured by a normal camera), and the amplitude and the phase of the modulated light component. These are illustrated in Figure 1.2.

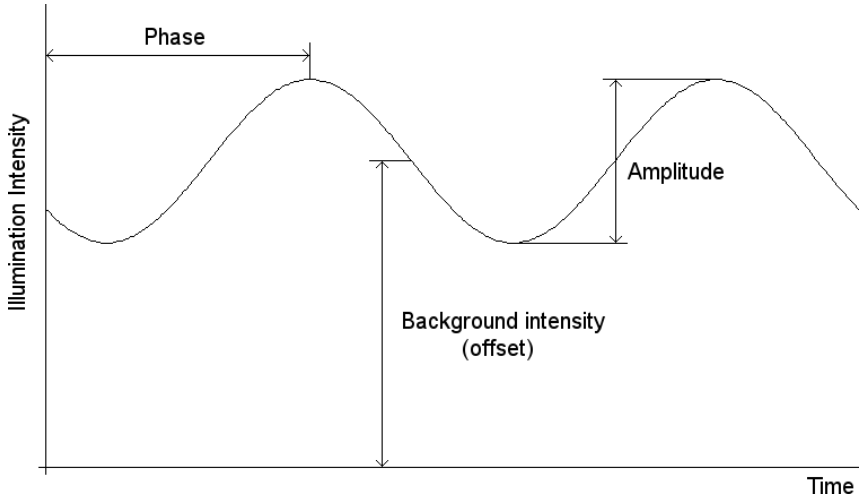


Figure 1.2: Modulated Light Parameters

How this signal is recovered is considered in Chapter 3. Other types of modulated signal exist, namely those where a range of modulation frequencies may be present and the intent is to measure the intensity of all of those frequencies, in other words a basic Fourier transform is performed on the detected light signal [Kongsavatsak, 2005].

A camera that can measure the modulation amplitude or phase is known as a modulated light camera, or MLC. Where the camera is designed to measure the amplitude and/or phase of a known fixed frequency, the camera may also be referred to as a lock-in camera.

There are many situations where using modulated light is useful. In the simplest case, detecting the presence of modulation without needing to know the modulation amplitude or phase is sufficient. One such example is the measurement of sperm motility. Viable sperm produce a range of modulation frequencies up to around 1 kHz [Woolhouse, 1986]. By measuring the presence of modulated signals in this range, the number of potentially viable sperm can be counted and the paths which they take recorded, which is also a way of assessing viability [Comhaire et al., 1992]. It is also possible to envisage modulated light being used to help measure athlete

movements by placing small modulated light sources at positions on their bodies and recording both a standard image and a modulated light image. Doing so allows the movement to be isolated even in ordinary lighting. This is a very straightforward application because there are no real limitations on the frequency of the modulation and again, all that is being measured is the presence (and position) of the modulated sources.

Modulated light is often used in experiments to isolate the signal of interest from any background light. An example of an application where this is the case is in imaging the retina. The lens of the eye is imperfect and there are other factors which contribute to aberrations and mean that any potential image is of much lower quality than desired. Using a modulated light input the signal can be isolated from the background light and part of it then measured with a wavefront sensor to calculate the aberrations and correct for them using a deformable mirror [Stockford et al., 2007]. This application does not require high frequency modulation and indeed benefits from a lower frequency because the light levels are low and the deformable mirrors require a certain amount of time to move and produce the corrections. A modulation frequency of a few hundred Hz should therefore be sufficient.

It is possible to use modulated light detectors to measure range by calculating the phase between the modulation and the local oscillator, as described by many authors [Gruss et al., 1991, Gulden et al., 2001, Lange and Seitz, 2001] and as will be discussed in Chapter 2. This application requires high frequency modulation to obtain good range resolution. With a 100 MHz modulation frequency and a system capable of resolving phase within 1°, it would be possible to measure range with an accuracy of approximately 4 mm.

A final type of modulated detection is in measuring the frequency content of a signal. An example of this is described by Kongsavatsak [Kongsavatsak, 2005] for measuring the blood flow and concentration in skin. This type of camera is quite different to that proposed in this work because it does not work on lock-in techniques.

1.4 Electronic Photodetection

The photodetector in a camera takes photons and converts them into something else that is more permanent. In a traditional black and white camera film, the light sensitive silver halide grains are the photodetectors. If sufficient light is absorbed by a grain, it is converted to metallic silver and can be fixed in that state when developed, thus forming an image. In digital cameras, the photodetectors are a uniform array of pixels which are typically made up of charge coupled devices (CCDs) or photodiodes.

The design choices made for the photodetector are potentially the most important of the entire design; as they are the first stage of the device, any decisions here will impact on the whole of the design. The application that the camera will be used in is also important, for example if we want to create a device capable of the same resolution as a film camera then a small pixel size is needed, otherwise the device will become undesirably large. The equivalent pixel resolution of film is a hotly debated topic. Kodak use 6.8 μm pixels in many of their CCD sensors. Professional quality film cameras can resolve at least 100 lines per mm, which equates to 200 pixels per mm or 5 μm, so it appears as though CCD cameras are not that far behind film cameras in terms of feature size, but with 35 mm film having an image size of 35 x 24 mm, 5 μm pixels equate to a camera with 34.6 million pixels which is a long way beyond current consumer equipment.

Small pixels tend to suffer from worse noise performance (see Sections 1.6.2 and 1.6.6), and they leave little space for ancillary circuits alongside each pixel. Such ancillary circuits are desirable since including more circuitry in the pixel allows for more complex operations but of course reduces the area of the pixel that is light sensitive, also known as the fill factor. The fill factor is calculated as shown in Equation 1.1.

$$Fill\ factor = \frac{Light\ sensitive\ area\ of\ pixel}{Total\ area\ of\ pixel} \tag{1.1}$$

A high fill factor is desirable, because any light falling onto an area of the pixel that is not sensitive to light is wasted. It is possible to improve the effective fill factor significantly with the use of a micro lens array. The micro lenses are created with the same pitch as the pixels and focus the incoming light onto the photosensitive area of the pixel.

This section considers the different photodetectors available in standard CMOS processes and compares them to the charge coupled device (CCD). There are three types of photodetectors available in CMOS; the photodiode, the photogate and the phototransistor. First the process by which optical absorption occurs in semiconductors will be considered.

1.4.1 Optical Absorption in Semiconductors

In order to be able to measure light in an electronic device, it must be converted into electrons. Semiconductors can absorb photons and convert them into electron-hole pairs as follows.

In a semiconductor, the carriers that contribute to current flow exist in the conduction and valence bands. The conduction occurs in the conduction band as electron flow and also occurs in the valence band in the form of so called "hole" flow, where electrons moving within the valence band make it appear as though the hole, or lack of an electron, is moving in the opposite direction. To make an electron move from the valence band to the conduction band, it must be supplied with energy. For an optical detector, this comes in the form of light, or photons. A single photon has energy $E = hv$ in electron volts (eV), where h is Planck's constant and v is the frequency of the light. For a photon to move an electron to the conduction band, it must have an energy of at least the bandgap (that is, the energy gap between the

conduction and valence bands) of the semiconductor, E_g. Assuming that this is the case, an electron-hole pair will be generated.

Semiconductors are split into two groups - those that have a direct bandgap and those that have an indirect bandgap. Direct bandgap semiconductors need only rely on a photon being absorbed to generate an electron-hole pair, whereas indirect bandgap semiconductors require extra help in the form of a phonon, or lattice vibration. This means that optical absorption in indirect bandgap semiconductors is much less effective than in direct bandgap semiconductors.

Knowing the bandgap of a semiconductor means it is possible to calculate what wavelengths of light it is sensitive to. For example, gallium arsenide (GaAs), a semiconductor often used in low noise, high speed applications has a (direct) bandgap of 1.43 eV. The energy of a photon has already been mentioned and can be rearranged to calculate the wavelength of light that can be detected by a given semiconductor:

$$\begin{aligned} E &= hv \\ &= \frac{hc}{\lambda} \\ \therefore \lambda &= \frac{hc}{E_g} \end{aligned} \tag{1.2}$$

Where c is the speed of light. With $E_g = 1.43\,\text{eV}$, this gives $\lambda = 868\,\text{nm}$ as the longest wavelength of light that can be detected using GaAs.

Silicon has a bandgap energy of $E_g = 1.11\,\text{eV}$, so it can detect light of $\lambda = 1.12\,\mu\text{m}$ and below, meaning both it and GaAs can detect visible light. Silicon is an indirect bandgap semiconductor however, so GaAs is a better optical detector in that a higher proportion of photons are converted into electron-hole pairs. A final example is gallium phosphide (GaP), that has a bandgap of $E_g = 2.26\,\text{eV}$, meaning it can detect light of $\lambda = 550\,\text{nm}$ and below, so it is not sensitive to red or yellow light. GaP is used to make light emitting diodes, so not absorbing red or yellow photons is a useful feature.

The sensitivity of the semiconductor to light has a strong wavelength dependence. The material properties play a large role in how deep the light can penetrate and this directly affects how well the light can be absorbed. Figure 1.3 shows the depths to which light can penetrate in different semiconductors.

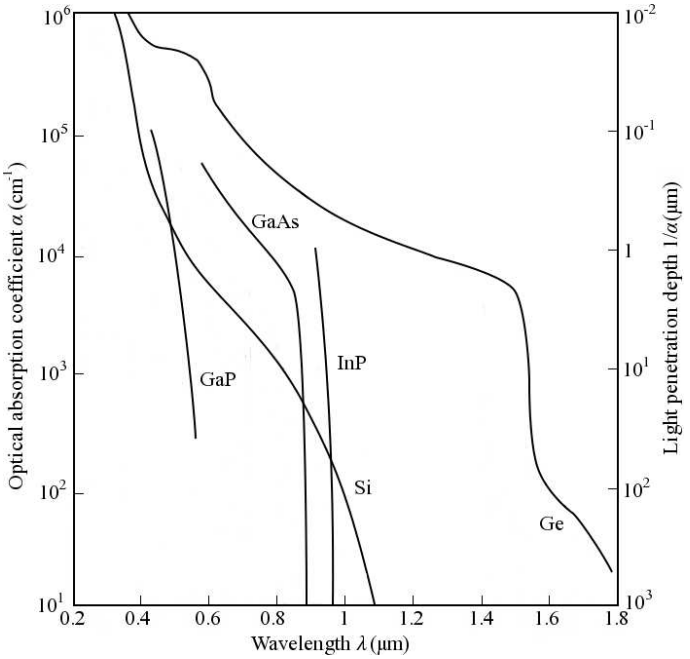


Figure 1.3: Semiconductor Absorption Coefficients [Sze and Ng, 2007]

Even when a photon of the correct energy strikes the detector, it may not cause an electron-hole pair to be generated. The ratio of the electron-hole pairs generated to the number of incident photons is known as the quantum efficiency and is wavelength dependent. Factors other than the material itself affect the number of photons converted into electron-hole pairs. For example, when a detector is fabricated on a chip, it will be covered by a number of transparent layers including the passivation layers that are placed over the entire chip for protection. These layers can absorb light and cause it to be reflected away due to the boundaries between the different layers, which means that the light reaching the detector is of a lower intensity than that which is incident upon the chip. The AMS C35 process has a manufacturing option to place a final layer on the chip that acts as an anti-reflection coating to increase the amount of light reaching the detectors and hence increase the responsivity.

In all cases, the photon being absorbed and an electron-hole pair being generated is not enough. Left to their own devices, the electron and hole will recombine in a short amount of time and there will be no useful effect. To generate a current flow, the electron and hole must be kept separated. This can be achieved by using a *pn* junction as the detector rather than just an intrinsic semiconductor. When the *pn* junction is reverse biased, the electric field across the junction will sweep any electrons generated near the junction into the *n* side and any holes generated into the *p* side, thus contributing to current flow.

The I-V characteristics of an illuminated *pn* junction are shown in Figure 1.4.

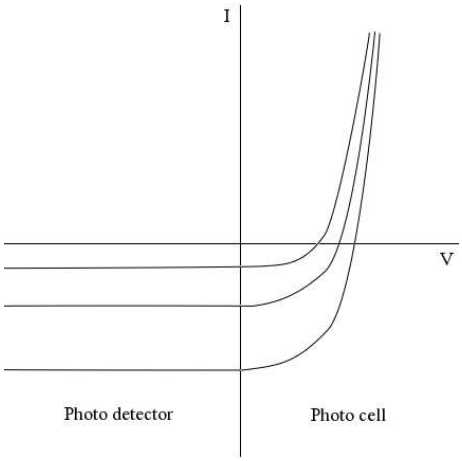


Figure 1.4: Illuminated *pn* Junction I-V Characteristics

When the junction is operating in the third quadrant it is reverse biased and a reverse current flows that consists of the dark leakage current and the photon generated current, also known as the photocurrent. This is the mode of operation when using the junction as a photo detector and the current flow is independent of voltage as long as the diode remains reverse biased. As the intensity of illumination changes, the curve moves up and down the y axis. This photodetector mode is also known as the photoconductive mode.

The *pn* junction can also operate as a photo cell. In the third quadrant, the junction is provided power by the external circuit whereas in the fourth quadrant, the junction provides power to the external circuit. This is also known as the photovoltaic mode and is the region used for solar panels.

The final quadrant of operation for a diode, the first quadrant, has the junction forward biased. This quadrant is used in rectifying circuits but is of no real interest in optical detection.

The following sections (1.4.2 to 1.4.5) describe the different ways in which the *pn* junction is utilised in devices for photo detection.

1.4.2 CMOS Photodiode

The photodiode is probably the most commonly used CMOS detector, being made of the simple pn junction. Every silicon process has a number of different silicon layers (as well as the metal and poly-silicon layers) that can be different thicknesses and different doping levels and types. This means that there are lots of possibilities for forming pn junctions that can then be used as a photodiode - in essence the only difference between a diode and a photodiode is that the photodiode is exposed to light. This further means that any *pn* junctions that are not intended for light detection, such as those in the majority of the transistors in a design, must have a light shield. As an example, the AMS C35 process has a p type substrate and a lightly doped n-well, as well as both $p+$ and $n+$ diffusions, which is the typical configuration for a single well CMOS process. In this case, there are three different photodiodes available, created through the combination of the different p and n layers. More complicated processes have extra layers for the creation of double or triple well structures and so allow for further *pn* diodes and the possibility of making more complicated diodes by using structures similar to *p-i-n* diodes.

The three diodes available in a single well process are all slightly different and offer individual advantages and disadvantages. They are the n-diffusion to p-substrate diode, the n-well to p-substrate diode and the n-well to p-diffusion diode.

1.4.2.1 *N*-Diffusion to *P*-Substrate

The n-diffusion to p-substrate diode, as shown in Figure 1.6, is sometimes referred to as a shallow diode due to its junction depth, which is $0.2\,\mu$m in the AMS C35 process. The shallow junction depth means that its sensitivity tends toward light with shorter wavelengths which is caused by photons penetrating further into silicon at higher wavelengths and so not being captured by the junction. Because the n-diffusion has high doping concentration the doping profile is relatively steep and so the depletion region at the junction is narrow, meaning that the diode has a high capacitance.

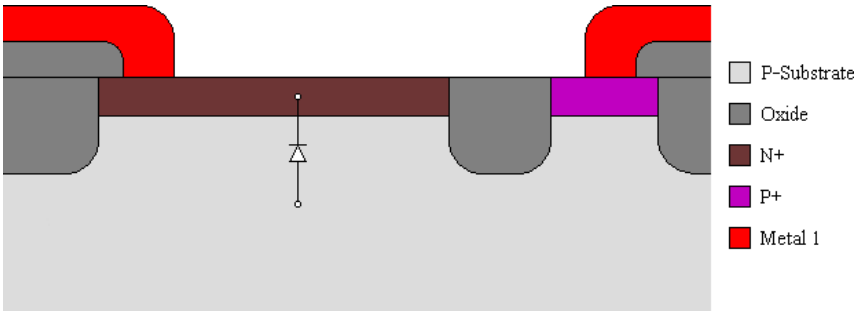


Figure 1.5: N-Diffusion to P-Substrate Diode

1.4.2.2 *N*-Well to *P*-Substrate

The n-well to p-substrate diode, also known as the deep diode, has almost exactly the opposite characteristics of the shallow diode, with relatively low capacitance and better responsivity at higher wavelengths, as well as having a higher leakage current in reverse bias than the shallow diode. In the AMS C35 process the junction is $3.5\,\mu$m deep. The cross section of the diode is shown in Figure 1.6.

1.4.2.3 *N*-Well to *P*-Diffusion / Buried Double Junction

The n-well to p-diffusion diode, as shown in Figure 1.7, offers an attractive alternative to the previous two diodes as both the anode (A) and cathode (C) are able

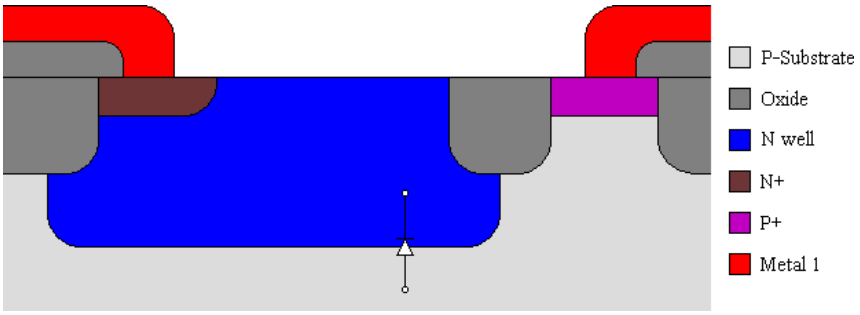


Figure 1.6: N-Well to P-Substrate Diode

to be set at any voltage whereas one of the terminals in each of the previous two diodes was the substrate, which is connected to ground to ensure that the associated pn junctions remain reverse biased with no risk of latch-up occurring. Being able to connect each terminal to any point makes it possible to use the diode in less conventional setups (as both inputs to a differential mixer, or in the feedback path of an amplifier for example). In general this diode offers a similar performance to the n-diffusion to p-substrate diode however the additional deep junction formed by the n-well to p-substrate boundary means that it exhibits both long and short wavelength sensitivity [Pui, 2004], meaning an overall wider wavelength response and the possibility of colour detection [Lu et al., 1996].

As the primary junction is isolated from the substrate by the n-well, this device tends to have better immunity to cross talk from adjacent photodiodes or other devices nearby that are injecting charge into the substrate.

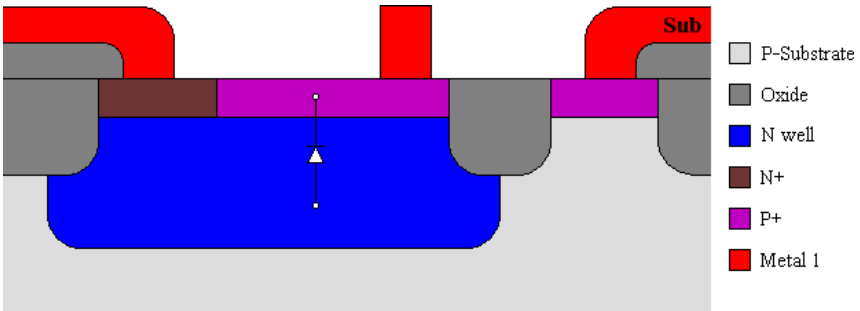
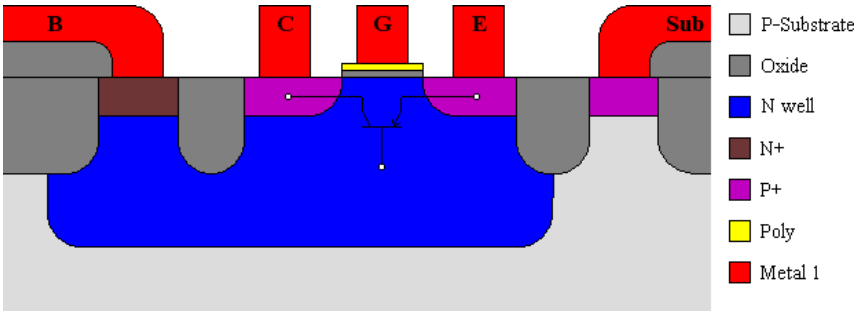


Figure 1.7: N-Well to P-Diffusion Diode

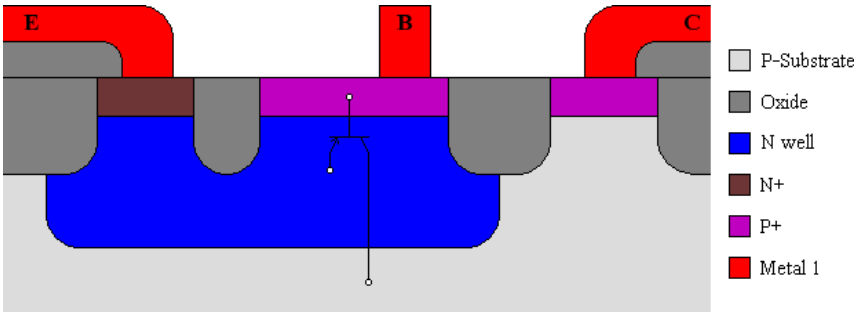
1.4.3 CMOS Phototransistor

An alternative to using a photodiode is to use a bipolar transistor as a photosensitive device. By arranging it so that the photons are absorbed near the base-collector junction, the device operates just as it would if a base current was applied directly. There are two types of bipolar transistors in a single well CMOS process; the lateral transistor, which is formed as a *pnp* device along the x-y plane of the chip and the vertical transistor which is formed as a *pnp* device with layers in the z plane of the chip. Both of these devices are typically treated as "parasitic", that is to say that they are not intentionally included but are only there because *pnp* structures occur anyway. They are modelled, but not as well as the FET devices and they tend to have relatively poor performance. The vertical transistor layout provided by AMS in the C35 process in particular has a very small current gain of around 6 at 100 nA compared to around 100 in a typical bipolar transistor and is also limited by having the collector tied to the substrate whereas the horizontal transistor can connect all terminals independently. As current gain is one of the primary reasons for using a phototransistor, the vertical device is of little interest in practical sensors. The horizontal transistor fares much better with a current gain of 76 at 100 nA whilst also occupying a smaller area.

The cross sections of the two configurations are shown in Figure 1.8. The lateral transistor has a fifth terminal which is very similar to the gate in a MOSFET. Its purpose is to separate the collector and the emitter diffusion during manufacturing by blocking the $p+$ implant which avoids having to resort to using field oxide. This keeps the effective base length as short as possible which helps increase its gain. When the device is being used, the gate is kept at 0 V so that it does not cause inadvertent conduction between the collector and emitter.



(a) Lateral



(b) Vertical

Figure 1.8: Bipolar Photo Transistor Cross Sections

1.4.4 CMOS Photogate

The photogate consists of an area of diffusion covered in a gate and a read out transistor, as shown in Figure 1.9. The operation of the photogate differs from the diodes and transistors already mentioned in that it is always used in a charge integration mode. The first stage of the measurement involves the $n+$ diffusion between the PG and TX terminals being emptied of charge, followed by a high voltage being applied to the diffusion gate, PG. The high voltage creates a strong field which collects any electrons generated by the illumination and accumulates them in the $n+$ diffusion. The photogate PG is held high for the integration period and then the charge is read out by activating the read out transistor to transfer the accumulated charge to a charge amplifier.

The photogate has the disadvantage that the light sensitive area is covered by the gate. Although this is transparent, it still reduces the amount of light reaching the well.

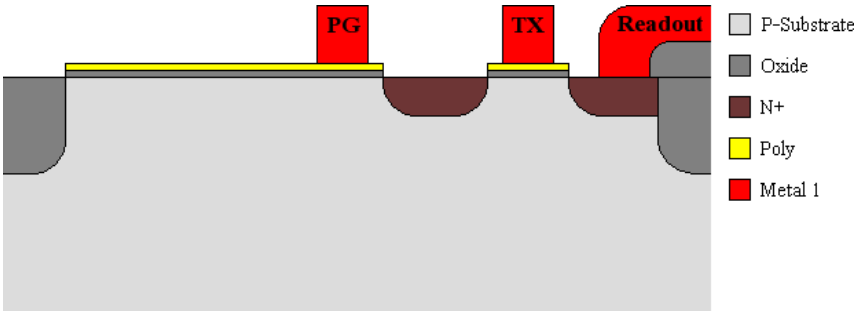


Figure 1.9: Photogate

1.4.5 Charge Coupled Device (CCD)

It is not possible to discuss photodetectors without mentioning the charge coupled device (CCD). First invented by George Smith and Willard Boyle in 1969 CCDs rapidly overtook NMOS sensors (and CMOS sensors when they became available) due to their better sensitivity. Since then CCDs have been further optimised and can achieve noise levels of just a few electrons over long integration times. CCDs operate in a similar way to the photogate in that electrons are collected in a well and then transferred out to be read. A CCD array will transfer the charge from a given pixel across the array from pixel to pixel until it reaches the output stage and so they are manufactured on specialised processes which is optimised for this charge transfer. The disadvantage of this type of process is that the MOS transistors perform poorly so cannot be used for constructing ordinary circuits and there is little scope for combining a photodetector and processing circuitry into each pixel.

1.5 CMOS Pixel Sensors

1.5.1 Passive Pixel Sensor

The passive pixel sensor (PPS) can be regarded as the first MOS pixel. Introduced in 1967 [Weckler, 1967] each pixel consists of a photodiode and a single transistor which is used as a switch to select the pixel. Figure 1.10 shows two pixels, which

consist of D0 and MN0, and D1 and MN1. In addition to the array circuitry, charge amplifiers are included at either the column or array level (shown here as I0).

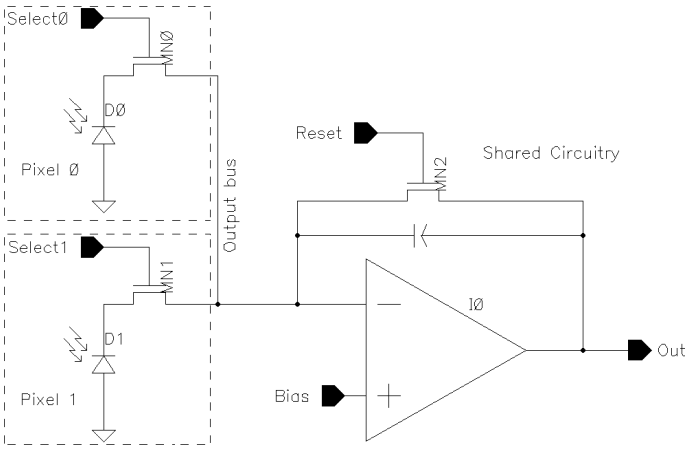


Figure 1.10: PPS Schematic

The PPS operates by making use of the inherent capacitance of the photodiode as a storage element. The selected pixel is first primed by using MN2 to "reset" the selected photodiode and the charge amplifier capacitor. This is done by turning on MN2 when the output is connected to V_{DD}, whereby the voltage across the photodiode will be charged to V_{DD}. When MN2 is turned off, the capacitance will discharge due to the current caused by the incident illumination. Sampling the output a fixed duration after the reset allows the light intensity to be measured.

As the PPS only has a single transistor at the pixel it is very attractive for making cameras; even with small pixels the area of the pixel that is not light sensitive will be minimised and so lead to a more efficient use of the available light. However, it is not possible to access each pixel at once so an array using PPS pixels will suffer from a rolling shutter effect where there is a finite amount of time between each pixel being sampled and so anything in the scene that is moving quickly will be distorted. A more substantial problem with the PPS is that when a given pixel is selected, the photodiode is connected to the output bus directly and so it must drive the bus capacitance as well. This reduces the speed at which a PPS can operate. As already mentioned, the signal is caused by the photodiode capacitance discharging due to the photocurrent, but now the photodiode is connected to a large additional

capacitance and as a result the pixel will take more time to discharge to the same point than without the additional capacitance, given the same photocurrent.

1.5.2 Active Pixel Sensor

Strictly speaking any pixel that has a buffer, amplifier or other device intended to improve the pixel performance, where that device is integrated directly at the sensor is an active pixel sensor (APS), however the term is commonly used to describe an integrating pixel as described below. Adding a buffer means that the reset device must also be included at the pixel. This increases the in-pixel device count from a single transistor to three transistors, but removes the problem of the high column bus capacitance exhibited by the PPS. The APS was first proposed by Chamberlain in 1969 [Chamberlain, 1969].

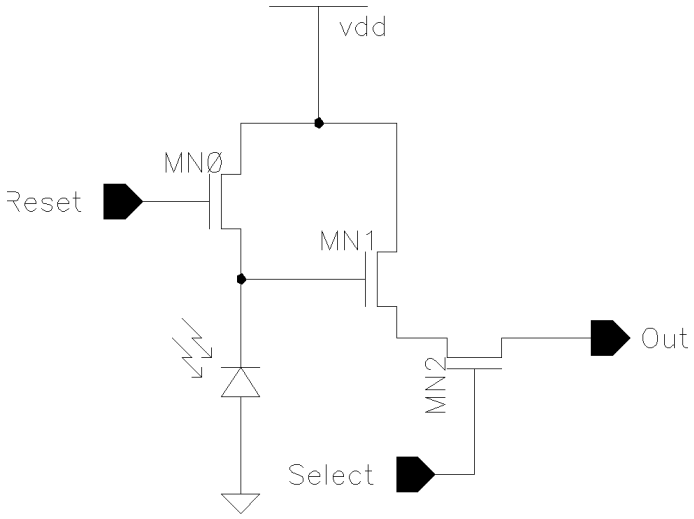


Figure 1.11: Three Transistor APS Schematic

Figure 1.11 shows the in-pixel circuitry of the three transistor APS. MN0 is the reset transistor that is equivalent to transistor MN2 in the PPS. MN2 is the select transistor and MN1 is the top half of a source follower buffer, where the bottom transistor of the follower is common for the entire column and is not shown in this figure. Splitting the follower in this way has a number of advantages: it keeps the fill factor high by removing the need for an additional transistor and the associated

bias wire. A further advantage of having the bottom half of the follower shared per column is that by definition each pixel in the column must use that device. In the case where each pixel had a complete buffer, although the drawn size of the transistors would be identical there would still be mismatch. It also lessens matching problems as the bottom transistor is common to all pixels in a column. Using a shared transistor like this reduces mismatch between pixels; if each pixel contained a complete buffer then the slight mismatch between the size and other parameters of the buffer transistors from pixel to pixel would cause a different voltage offset at each pixel, manifesting itself as increased fixed pattern noise (see Section 1.6.5). With a shared transistor this mismatch is removed for a column, although the other buffer transistor in the pixel still has the same effect, and there is still column to column mismatch. Taking this technique to its extreme would result in a single transistor shared as the bottom half of the source follower for the entire array, which would result in a reduced fixed pattern noise at the expense of a greatly reduced frame rate caused by each pixel having to be integrated and sampled individually rather than a row at once.

A camera using the three transistor APS suffers from the same problem as one using the PPS, namely that only a single pixel (or row of pixels) can be accessed at once and these is no storage at the pixel so it has a rolling shutter. Unlike the PPS, it is possible to circumvent this problem by adding a further transistor to the pixel, at the trade off of fill factor, as shown in Figure 1.12. The extra transistor is used as a shutter to sample the photodiode voltage onto the small gate capacitance of the output buffer. Applying the shutter to each pixel at the same time removes the rolling action whilst still allowing the results to be read serially.

1.5.3 Linear Pixel

It is often desirable to be able to make continuous time measurements of light level, something not possible with the PPS or APS due to their integrating behaviour.

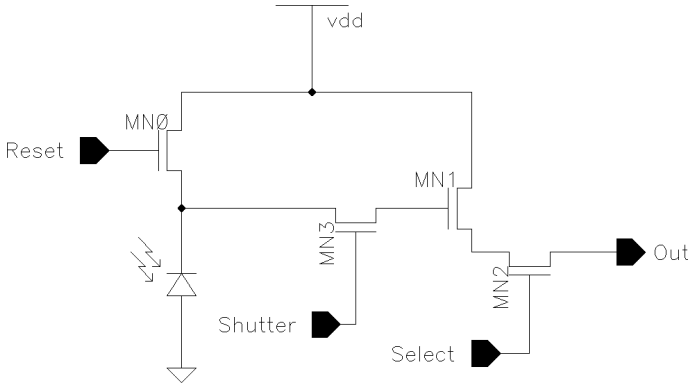


Figure 1.12: Four Transistor Shutter APS Schematic

In fact, the very first solid-state image sensors developed in the 1960s produced continuous time measures of illuminance. For example, Schuster and Strull produced a 50 x 50 array of phototransistors in 1966 [Schuster and Strull, 1966]. To amplify the signal to a useful level, a gain element must be included in non-integrating detectors. By placing a resistor in series with the photodiode, as shown in Figure 1.13, the photocurrent will be converted into a voltage at the Out terminal as per Ohm's law. This means that the gain, or transimpedance, of the pixel is determined by the value of the resistor. Assuming that the resistor is linear, the output voltage will also be linear with photocurrent.

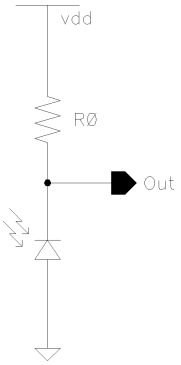


Figure 1.13: Linear Pixel Schematic

The fact that the gain is determined by the resistor can also be a disadvantage; to convert typical photocurrents in the pA - nA range to useful voltages of at least mV means that the resistor must be of the order of $1\,\mathrm{G}\Omega - 1\,\mathrm{M}\Omega$. By examining minimum width resistors made of high-resistance polysilicon using the manufacturer provided design kit it was found that a $1\,\mathrm{M}\Omega$ resistor would consume a silicon square

35 μm x 35 μm and a 200 MΩ resistor, would be 565 μm x 565 μm. Clearly fabricating a 200 MΩ resistor in an array would be extremely impractical, but in some cases a 35 μm square might be feasible. Of course, this is using a minimum width device which means that the pixel to pixel variation in resistance could be quite significant, with a corresponding increase in fixed pattern noise. Using a wider resistor would reduce this variation at the expense of increasing the area required.

A further problem with using a device with such a large area is that in addition to the desired resistance there is also an undesired capacitance. As both parameters are increasing at the same time they have a negative effect on the cut off frequency of the pixel. For example, the 1 MΩ resistor already described has an associated parasitic capacitance of 93 fF, resulting in a cut off frequency of around 1.7 MHz. Simply changing the width of the resistor from 0.8 μm to 2 μm to reduce mismatch, whilst leaving the resistance the same gives a parasitic capacitance of 470 fF and a cut off frequency of around 340 kHz.

The linearity of the output is both an advantage and a disadvantage. On one hand, a linear response is almost always easier to work with, but on the other hand, the dynamic range is limited directly by the available voltage range.

Because of these reasons using linear resistors in a pixel using the typical fabrication processes available today is not particularly feasible.

1.5.4 Logarithmic Pixel

An alternative to the integrating APS is to use a transistor in place of the resistor in the linear pixel and this means that a large transimpedance can be obtained in a relatively small area at the expense of sacrificing the linearity of the response to photocurrent. Because the transistor is operating in the sub-threshold region, the voltage output has a response proportional to the logarithm of the photocurrent [Chamberlain and Lee, 1984]. Although this is a disadvantage in many situations,

it can also be of benefit as well; a logarithmic response increases the dynamic range of the sensor and more closely matches the response of the human eye which are both useful features in real life scenarios. The logarithmic pixel is shown in Figure 1.14.

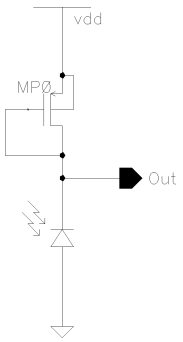


Figure 1.14: Logarithmic Pixel Schematic

1.5.5 Feedback Pixel

The feedback pixel (shown in Figure 1.15) is an improvement on the logarithmic pixel presented by Perelman [Perelman and Ginosar, 2001]. As the name suggests, a feedback amplifier is added to the pixel. The feedback is used to drive the gate of the load transistor with the aim of improving the frequency response of the pixel. The load transistor still works in the sub-threshold region so the pixel is still a log pixel, but the feedback works to ensure that any voltage swing appears at the output of the feedback amplifier. This corresponds to an increase in the usable frequency of the pixel as the capacitance being driven by the feedback amplifier, namely the gate of the load transistor and the drain of the select transistor, is much smaller than the capacitance of the photodiode.

1.6 Noise

Noise in all its forms degrades the performance of a system and so it is important to know about its causes and what can be done to minimise it. Temporal noise

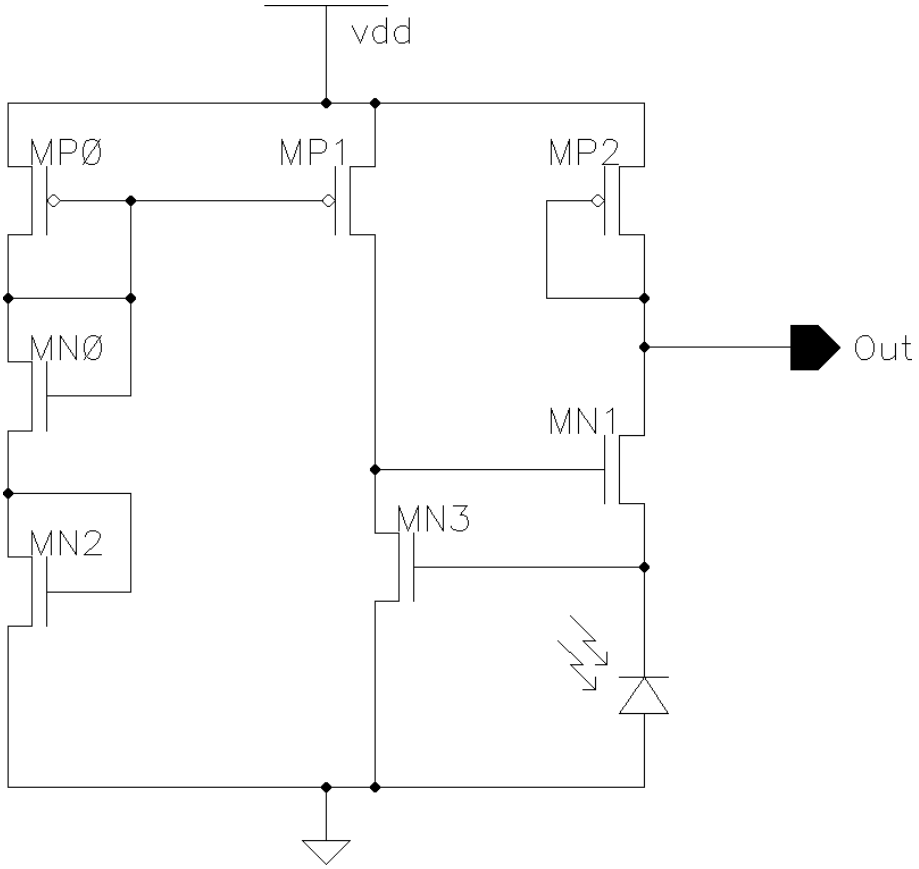


Figure 1.15: Feedback Pixel Schematic

components are thermal noise, shot noise and flicker noise. Fixed pattern noise is a spatial noise component. Each of these noise components will be discussed in more detail now.

1.6.1 Thermal Noise

The random motion of electrons in resistive material causes small voltage fluctuations. As would be expected, higher temperature (causing more active motion) and higher resistance both result in higher thermal noise. A resistive material will experience thermal noise on the voltage at its terminals according to Equation 1.3 .

$$V_{thermal(rms)} = \sqrt{4kT\Delta fR} \tag{1.3}$$

Where k is Boltzmann's constant, $1.38\,\mathrm{x}\,10^{-23}$ J/K , T is the temperature in Kelvins,

Δf is the bandwidth (Hz) under consideration and R is the resistance of the material in Ohms. As there is no current term, it can be seen that thermal noise is present even if there is no current flowing through the conductor. As k and R are constant for a given piece of conductor, the important factors here are the temperature and the bandwidth of the system and by reducing them the thermal noise contribution can be reduced. It is important to note that it is the bandwidth of the system that is important, not the frequency of a signal that it is trying to measure. This is because thermal noise is white noise, that is to say that it has uniform power contribution at all frequencies (up to around 6000 GHz [Maas, 2005]).

A good example of an application where thermal noise reduction is essential is found in astronomical imaging, where the image sensor is typically cooled to cryogenic temperatures and the images are acquired over a long period to help reduce the contribution of thermal noise.

1.6.2 Shot Noise

Shot noise occurs in both electronics and optics. In both cases it is caused by the fact that the energy is being carried by discrete packets (electrons and photons respectively) and that the arrival of these packets is random. In a pn junction, the DC current flow consists of the electrons and holes moving across the junction. As the holes and electrons move in a random direction, they will only cross over the junction if they have enough energy to do so for the direction that they are travelling, which is completely random. Equation 1.4 gives the shot noise magnitude for a DC current I and and bandwidth Δf, where q is the electron charge.

$$i_{rms} = \sqrt{2qI_{DC}\Delta f} \tag{1.4}$$

Shot noise is the often the greatest contributor to noise in a semiconductor device and as it is caused by the current flow rather than the device itself there is not much

that can be done to reduce it apart from reducing I_{DC}.

The shot noise in an integrating detector such as the PPS or APS is proportional to $\sqrt{N}$, where N is the number of electrons captured (with a maximum determined by the well capacity of the detector) and is given in Equation 1.5, where C is the capacitance of the photodiode, V is the voltage across the device at full capacity and q is the electron charge.

$$N = \frac{CV}{q} \tag{1.5}$$

This allows the dynamic range of the pixel to be calculated:

$$\begin{aligned} \text{Dynamic range} &= N/\sqrt{N} \\ &= \sqrt{N} \end{aligned} \tag{1.6}$$

So to increase the dynamic range of a detector, the well capacity must be increased. Although this can be done by increasing the maximum voltage used across the detector, this does complicate the design due to the need to cope with an extra supply voltage. Processes with the main supply voltage beneath 5 V often offer a "thick oxide" transistor that can be used at a higher voltage and are intended for interfacing with the outside world. For example, 3.3 V processes may offer a 5 V option and 1.8 V processes may offer 3.3 V. Both of these possibilities give the option of a reasonable increase in well capacity at the expense of fill factor (the thick oxide transistors occupy more space and require extra control lines) and complexity. Even higher voltages are possible using a dedicated high voltage process rated to 50 V for instance, but this is a move away from the common and inexpensive CMOS process and high voltages are more difficult to deal with.

The other way to increase the well capacity is of course to increase the capacitance of the detector which can be done by either using a device with a greater capacitance per unit area or else increasing the area of the device. Increasing the capacitance

per unit area implies decreasing the junction depth (so changing from an n-well to p-substrate diode to a n-diff to p-substrate diode for example), but this has implications for the other characteristics of the device and so may not be an option. So the simple solution is to increase the area of the device. In fact there are no real disadvantages to this from a sensor point of view since both the dynamic range and the fill factor will increase. The disadvantages are cost, due to the increased area required, and a larger chip may not be compatible with existing optical systems.

With the desire for larger numbers of pixels per camera in a standard size device, the pixel size decreases and hence the well capacity decreases as well. Typical CCDs have a well capacity on the order of $10^5 - 10^6$ electrons. The Kodak KAF-3200 sensor, for example, has a limit of 55 000 electrons, giving a shot noise limited dynamic range of 235, or 7-8 bits resolution [Kodak, 2007].

1.6.3 Flicker Noise (1/f)

Flicker noise is present in all semiconductor devices. Although it is not fully understood it is related to carrier recombinations at the surface and the interface between the silicon and the oxide above it [Sze and Ng, 2007]. The atoms at the edge of the silicon do not have a full set of bonds by definition. The remaining bonds are left dangling (see Figure 1.16) and these bonds mean that electrons are more likely to be trapped randomly at the interface.

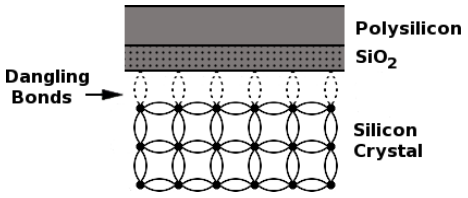


Figure 1.16: Oxide-Silicon Interface [Razavi, 2001]

The electrons being trapped and released manifest themselves as noise with an unusual spectral response in the form of $1/f^{\alpha}$, where $\alpha \approx 1$, in other words flicker noise

is dominant at low frequencies. The concept of the flicker noise corner becomes useful here. As the contribution to the overall noise content from flicker noise decreases with frequency, there will come a point at which it is no longer a significant factor. This is illustrated in Figure 1.17, where f_C is the corner frequency.

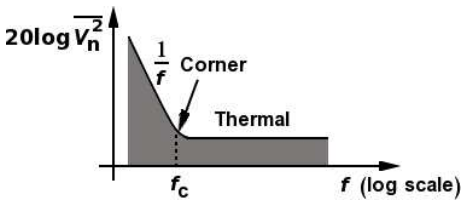


Figure 1.17: Flicker Noise Corner Frequency [Razavi, 2001]

The corner frequency can be calculated with Equation 1.7 [Razavi, 2001], where K is a process parameter close to $10^{-25}\,\mathrm{V^2F}$, k is Boltzmann's constant and T is the temperature in Kelvins.

$$f_C = \frac{K}{C_{ox}WL} \cdot g_m \cdot \frac{3}{8kT} \tag{1.7}$$

Taking $C_{ox} = 4.54\,\mathrm{fF}/\mu\mathrm{m}^2$ [AMS, 2004] with a NMOS transistor $1/1\,\mu$m, f_C can be as high as 158 kHz.

This is particularly important when taking many types of measurement since the rate of change of the parameter being measured can be very low. By using a modulated light input as a carrier to shift the signal away from the flicker noise, the noise performance of the system can be greatly improved. To reduce the effect of flicker noise, the simple solution is to use as large a device as possible within the parameters of how the circuit should work.

1.6.4 Dark Current

Although diodes are often treated as having no current flow when in reverse bias, they do actually suffer from a small leakage current that depends on the type of the diode, its area, the length of its periphery and its reverse bias voltage. This means

that a photodiode will still have a small current flow, even when not illuminated at all (hence "dark current"). When there is photo current flowing in the device the dark current manifests itself as an additional current component. The values of current involved are very small, for example a $50 \times 50\,\mu$m n-well to p-substrate diode in the AMS C35 process reverse biased at 3.3 V has a leakage current of 8.5 fA and an identically sized n-diffusion to p-substrate diode has a leakage current of 1.4 fA. The dark current itself is constant and so will be removed by the demodulation process, but it will provide an additional contribution to the shot noise of the diode. The dark current is also dependent on temperature and an increase can have very noticeable effects. Schanz *et al.* reported a decrease in dynamic range of 10 dB at 85°C caused by the increase in leakage currents [Schanz et al., 2000].

1.6.5 Fixed Pattern Noise

Fixed pattern noise (FPN) differs from the noise sources described so far in that it is a spatial variation on a chip. Fixed pattern noise occurs because not all devices are identical, even when they are designed to be so. CMOS in particular has problems with supposedly identical transistors having voltage offsets of the order of tens of millivolts when under the same stimulus. These offsets are a particularly significant problem in symmetrical circuits such as differential pair amplifiers or Gilbert cell mixers but are a problem wherever two or more devices should be identical, such as the transistors in a camera array. The offsets can be reduced with careful layout techniques, with moderate matching producing offset voltages of around ± 5 mV [Hastings, 2001], but it is not possible to remove them completely. The use of laser trimming of the device after manufacture allows much better matching to be achieved, on the order of ± 1 mV, but this is only really suitable where a small number of devices need correcting. All types of devices have similar problems, not just transistors. Essentially, it is not possible to ensure that any two devices are exactly identical and so each stage of circuitry adds a further contribution to FPN

simply by adding a greater number of devices.

A example of how FPN can appear in a camera is shown in Figure 1.18. The figure shows a part of a uniformly illuminated camera (the design from Chapter 5) that has a so-called column based design, where the readout and amplification circuitry for a particular column is shared for all of the pixels in that column. This means that any variations in the readout circuitry manifests itself as vertical stripes in the image. The variation of the circuitry within the pixel itself is also shown as the difference in intensity moving up a column.

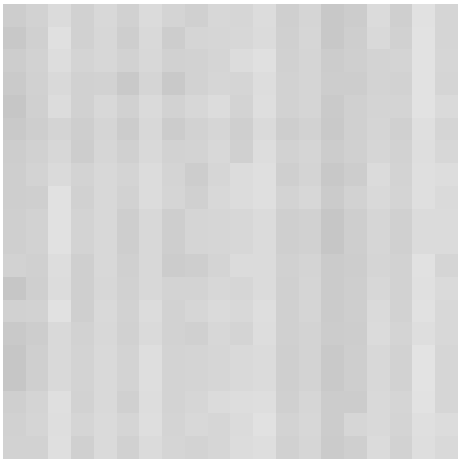


Figure 1.18: Example of FPN

Process variations are typically considered in two different scenarios, the variation across a chip and the chip to chip variation, also known as mismatch and process variations respectively. The process variation is caused by the relatively large difference in the environment for chips in different positions on a wafer whereas the mismatch is the random changes that lie around the process variation. A simple way to view this is as though the process variation is the mean of the mismatch variation, where the mean changes from chip to chip. The process variation has a much greater effect than the mismatch variation.

Because FPN produces a static variation in time, it is often possible to calibrate the camera by illuminating it with uniform intensity and subtracting the mismatched values from any subsequent images. This also means that the effects of FPN will be

usually reduced when measuring the amplitude or phase of a modulated signal with a modulated light camera. Since modulated light cameras are essentially looking at the variation between subsequent signals, a constant offset should have little effect on their results. A case where fixed pattern noise could be a problem with modulated light cameras is with the logarithmic pixel design. If the FPN manifests itself as a change in the amount of DC current flowing through the pixel then the AC gain of the device would also be changed.

1.6.6 Reset Noise

Reset noise affects the integrating type of pixel sensors, where the photodiode capacitance is charged up through a switch. As the switch is resistive, thermal noise will be introduced on to the capacitance and when the switch is opened this noise will remain present along with the desired voltage. The level of the noise is independent of the value of the switch resistance and is actually mainly determined by the value of the photodiode capacitance, C, as shown in Equation 1.8, where k is Boltzmann's constant and T is the temperature in Kelvin [Razavi, 2001].

$$v_{rms} = \sqrt{kT/C} \tag{1.8}$$

Reset noise is often the limiting noise for integrating sensors, especially when the photodiode is small.

1.6.7 Noise Reduction

In addition to the general ways of reducing noise described, there are other techniques more specific to sensor design. Integrating pixels can utilise a number of techniques which can help to reduce the noise contributed from different sources. A technique called data double sampling (DDS) simply takes the results of two consecutive results and averages them which helps reduce the effect of any random noise,

but does assume that the illumination on the detector has not changed between the two samples and does not remove the effect of reset noise. A more complicated technique is correlated double sampling (CDS). This involves taking two samples per integration, where the additional sample is taken just after the reset period. The values are then subtracted from each other to remove the unknown reset noise. CDS can also help reduce FPN and is a more desirable solution than calibration because it does not require any additional storage or post-processing. Some techniques for reducing reset noise are available [Fowler et al., 2006], whereby a greater degree of control is used for the reset to control either the capacitance, bandwidth or the charge on the sensor capacitance, but all of the methods require more circuitry in the pixel and more complicated control which both contribute to reducing the fill factor.

1.7 Summary and Thesis Layout

This chapter has introduced the background to the project and provided a grounding in the basics of electronic photodetection by describing how semiconductors can be used as light sensors by absorbing light as photons and converting the photons into electron-hole pairs. These electron-hole pairs are the photocurrent which can be converted into a voltage using one of the pixel designs described in Section 1.5. Sources of noise which affect electronic circuits have been introduced, along with some ways in which they can be reduced. The idea that modulating the amplitude of a light signal is beneficial for being able to detect small signals, particularly in the presence of high background light has also been introduced.

Chapter 2 provides an overview of previous work carried out in the area of modulated light cameras. The different implementations of modulated light cameras are discussed as are their applications. Some cameras not originally designed for use as a modulated camera are also presented as being suitable for that purpose. Chapter

3 describes the theory and operation of lock-ins and modulation in general terms and more specifically when used in optical systems. It considers the different ways that it is possible to implement a lock-in and what the important characteristics to take into account when designing a lock-in are. It also applies this theory to the designs already mentioned in Chapter 2. Chapter 4 introduces the new lock-in pixel design and shows the designs fabricated on three prototype chips. Results are presented which demonstrate the pixel designs successfully operating as a lock-in. Chapter 5 discusses the final design fabricated, a 128 x 128 pixel camera and the specific issues with scaling up to a large array. Chapter 6 contains the experimental results obtained from the camera chip and Chapter 7 provides the conclusion to the thesis.

Chapter 2

Modulated Light Camera Review

2.1 Introduction

This chapter provides a review of previous work concerned with modulated light cameras. In keeping with the score of this project, this review only considers lock-in type cameras that are used to measure the amplitude, phase and DC offset of a modulated light signal rather than the the type of modulated light camera that is used to measure the frequency content of a light signal. Some cameras that could be used to measure modulated light but were not originally designed to do are also examined.

2.2 CCD Lock-In Cameras

2.2.1 Stenflo

One of the first mentions of lock-in cameras was by Stenflo and Povel [Stenflo and Povel, 1985] when they expressed a desire for a lock-in camera for use in polarimetry. They noted that whilst it is quite simple to attach a lock-in amplifier

to a single pixel system, it is somewhat more difficult in a complete pixel array. At that time, the lock-in amplifier would have been a large and completely separate device and connecting even a small number of them to a chip to create a camera would be very unwieldy. As an alternative they proposed simply reading each pixel value at the highest rate possible and using these values to do off-chip demodulation. This solution works but is not very satisfactory because, limited by the then maximum CCD readout rate of around 5 MHz, they would either be limited to a small array size of around 32 x 32, or else be limited to low modulation frequencies. For example, with a 385 x 288 array and taking two samples per modulation cycle, the maximum modulation frequency detectable would be approximately 22.5 Hz. Furthermore, with a high sampling frequency, the readout noise would likely dominate over any signal as discussed in Section 1.6. Clearly a better solution was required and in 1990 Povel [Povel et al., 1990] demonstrated how a CCD could be used directly as a demodulator. This new design used a commercially available CCD sensor with 385 x 288 pixels illuminated through a slit mask so that every other row was shielded from the light. In the first half cycle of the modulating period the light is integrated onto the exposed rows in the same way as a CCD functions normally, the charge in every row is then quickly shifted one row down. This means that the charge from the first half cycle is now in the shielded rows and the empty rows are exposed. The second half cycle is integrated onto the exposed rows and then all of the charges are quickly shifted up one row. The charge that is now exposed is that from the first half cycle and so the cycle begins again. This technique is illustrated in Figure 2.1.

The two results do not need to be read immediately; it is possible and even desirable to continue integrating each half cycle into the appropriate container to increase the signal until the signal level is satisfactory. The two results obtained per pixel can be used to calculate the amplitude of the modulation as will be discussed in section 3.5.6. The sensor was used successfully to measure modulation frequencies of 50 kHz and 100 kHz. Overall, the sensor is a simple design that works well but it does suffer

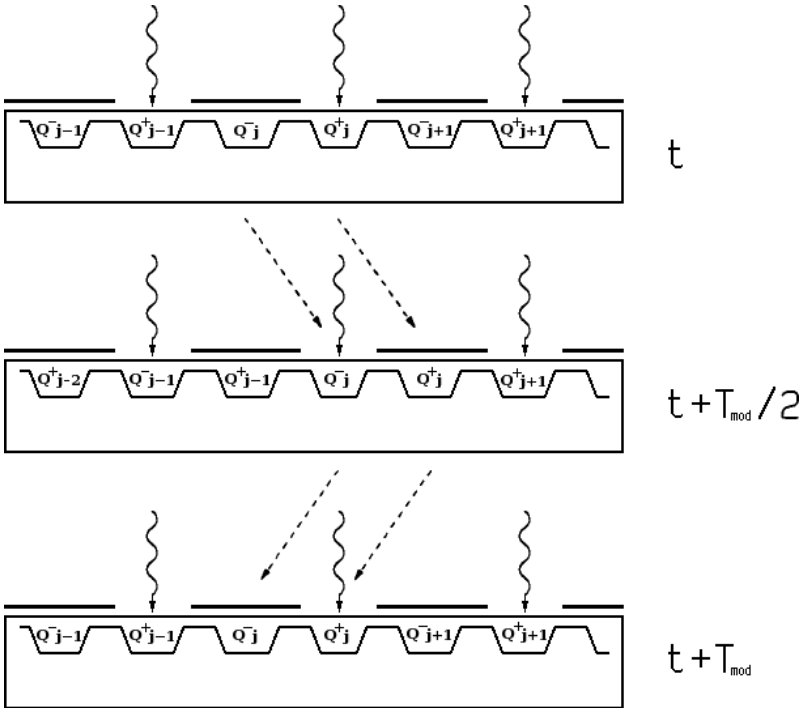


Figure 2.1: Two Phase CCD Demodulation [Povel et al., 1990]

from a number of disadvantages. The need for a shielded storage area creates a few problems, namely that using a slit mask to produce a set of shadows on the array is difficult to set up accurately and is prone to misalignment, possibly due to vibration. This is only a problem with the particular implementation described as it would be perfectly possible to create a custom sensor where every other row was shielded by metal on the die, which would have the added benefit of making the sensor more resistant to background light. The real disadvantage of the shielded storage area is the loss of half of the available light sensitive area and the corresponding loss of spatial resolution. Again, this could be improved by using a custom solution that had smaller storage areas. A further problem with this design is that the CCD remains exposed to light at all times and so will integrate charge constantly, even during the finite amount of time during which the charge from one period is being transferred to the shielded storage area and, during that period, a small amount of mixing will occur between the signal that is being transferred away and the signal that is currently exposed to the CCD.

2.2.2 Spirig

In many cases the phase of a modulated signal is also important. With only two samples per period, Povel's sensor is only capable of calculating the amplitude and offset of the signal, and only if the phase between the signal and the charge transfer is both constant and known; and such that the samples occur at or close to the peaks and troughs of the signal. In 1995 Spirig [Spirig et al., 1995] attempted to build on Povel's work by increasing the number of samples per period to four and so making measurement of the phase of the modulation possible as well as the amplitude and intensity (offset). The new device was custom fabricated as a 3 x 3 array, with each pixel consisting of a single photogate connected to four adjacent storage gates. They called this new device the lock-in CCD. By controlling transmission gates between the photogate and the storage gates it is possible to direct the integrated charge to the correct storage gate for the sample. By using a custom design the problems associated with the previous design could be overcome and the layout of the device optimised to make most efficient use of the available area. Despite this, the pixel is still rather large at $87.5 \times 80\,\mu$m with a fill factor of only 17.1%. The device was capable of measuring modulations at more than 100 kHz although the authors postulated a maximum modulation frequency of 30 MHz if manufactured in a then standard CMOS process. Further work [Spirig et al., 1997] goes on to explore multitap lock-in CCDs where there are even more storage bins, in this case eight, which allow the amplitude, phase and offset to be calculated with greater certainty.

2.2.3 Lange

A large proportion of the research into modulated light cameras is focused on range finding. There are a number of different ways of measuring range, such as the light-stripe method [Sato et al., 1987] and the parallel light-stripe method [Gruss et al., 1991]. Both of these methods operate in roughly the same way, namely illuminating a scene with a vertical stripe of light which can be seen to follow the

contour of any objects in the scene at the receiver. Using trigonometry, the position of each point along the light stripe can be determined and then by stepping the stripe, the range information for the entire scene can be gradually obtained. A simple variant of this technique is used by the company Orthotech, who create custom shoe inserts and other devices intended to help with foot problems. The patient's foot is imaged when they are wearing a special sock with a black and white stripe pattern which captures the contour of the foot in the same way as the light stripe method, but is applicable even in the presence of high background light [Orthotech, 2007]

The light stripe method requires that the stripe can be easily detected and distinguished from any other light in the scene, meaning that the background light must be carefully controlled. An alternative is of course to use a modulated light source to provide immunity from the background light. Using modulated light also means that a new way of measuring range becomes possible, known as the time-of-flight principle. Although already suggested by Spirig in 1997 [Spirig et al., 1997], the principle was not realised as a full field camera until 2001 [Lange and Seitz, 2001]. As the speed of light is known, it is possible to measure distance by measuring the amount of time taken for a pulse of light to travel from a source to the target and then to the receiver. A similar principle applies when using a continuous wave modulated light source rather than a single pulse. To measure the distance to the target the phase of the signal received relative to the source signal must be calculated. Knowing the frequency of the modulation then allows the phase change to be matched to a distance. When the phase change is known, the time of flight can also be calculated as the frequency of modulation is also known and from that the distance of the object calculated as shown in equation 2.1 [Gulden et al., 2002].

$$d = \frac{c}{2} \cdot \tau_{tof} = \frac{c}{2} \cdot \frac{\Delta\phi_{tof}}{2\pi \cdot f_{mod}} \tag{2.1}$$

In such a system, the maximum distance that can be measured for a given modulation frequency corresponds to a phase shift of one wavelength. If the object is any

further away, the phase wraps around and makes it appear much closer than it is. For example, with a distance equivalent to a phase shift of 370° the object would appear as though it was a distance equivalent to just 10° away. To get around this, multiple modulation frequencies can be used - a low frequency to allow for longer range measurement and a high frequency to increase the resolution. Using both frequencies together allows for high resolution distance measurement even for long distances. Lange's work considers a single frequency. His CCD pixels integrate four samples per modulation period, but with only a single storage gate per pixel and with multiple integrations being taken for each of the sample points there can be a reasonable amount of time between the first sample point being measured and the four sample points being measured. At a modulation frequency of 20 MHz, this camera operates at substantially higher frequencies than the others already considered although it is well within the theoretical maximum proposed by Spirig.

2.3 CMOS Integrating Cameras

2.3.1 Ando

Ando and Kimachi claimed one of the first CMOS implementations of a lock-in camera [Ando and Kimachi, 1999] in a device functionally very similar to that proposed earlier [Povel et al., 1990], in that it integrated the photocurrent onto two capacitors, one for each half of the signal cycle. They followed up the design with an improved 100 x 100 pixel single phase correlation image sensor (CIS) and a 64 x 64 pixel three phase CIS [Ando and Kimachi, 2003]. The schematic for the three phase pixel is shown in Figure 2.2, where "PD" is a current source used to indicate the photodiode.

In both of the devices, the overall function of the pixels is the same, but the additional phase of the three phase pixel allows the phase of the modulated signal to be

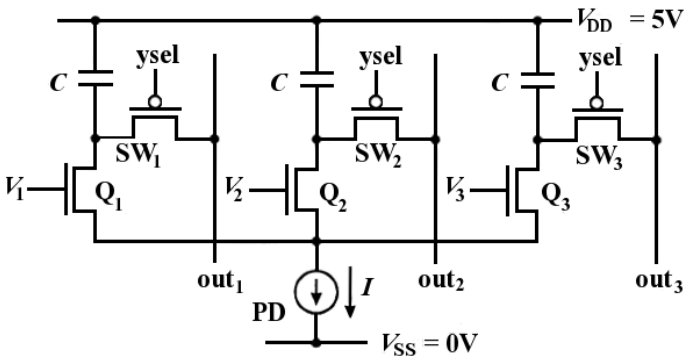


Figure 2.2: Three Phase Correlation Image Sensor [Ando and Kimachi, 2003]

measured as well as the amplitude. The pixels have two modes of operation. The first mode draws a direct analogy to the lock-in CCD [Spirig et al., 1995]: the voltage across the capacitance C is set to zero by closing the switch SW_i and connecting out_i to $V_{DD} = 5\,\mathrm{V}$. The switch is then opened again until the next reset. The control voltages V_i are then operated as non-overlapping pulses so that only one of the phases is in operation at once. The charge at the bottom plate of the capacitor discharges to gives three results (in the case of the three phase CIS), that can be used to calculate the DC offset, amplitude and phase of the modulated signal.

The second mode of operation is more generally interesting and makes more efficient use of the light by integrating with all capacitors at once, albeit still with dead time for readout purposes. In this case, the three control signals V_1, V_2 and V_3 have an identical shape but with a phase shift of 120° between them. In contrast to the first mode, the control signals can be continuous wave voltages. As all the control voltages may be partially on at a given point in time, the photodiode current will be split between the three branches as a function of the different voltage ratios. Using a continuous sine wave as the control voltage is advantageous since the mixing is being done with only a single frequency, in contrast to square wave mixing where an infinite number of frequencies are present. This makes the reconstruction of the signal much simpler, in particular where filtering must be carried out to remove unwanted frequency components.

2.3.2 Schwarte

A similar approach is taken by Schwarte [Schwarte et al., 1997] in his photonic mixer device (PMD) which performs demodulation directly in the sensing device, which is a type of electro-optical mixer (EOM). EOM devices are not a new idea, but they are often fairly complicated devices that are separate to the camera, such as a spatial light modulator, which modifies the light passing through it. The PMD pixel consists of two reverse-biased diodes formed between the p-type substrate and an $n+$ diffusion. The junction is not directly visible to light, instead the substrate region between the two diodes is left exposed through two transparent electrodes as shown in Figure 2.3.

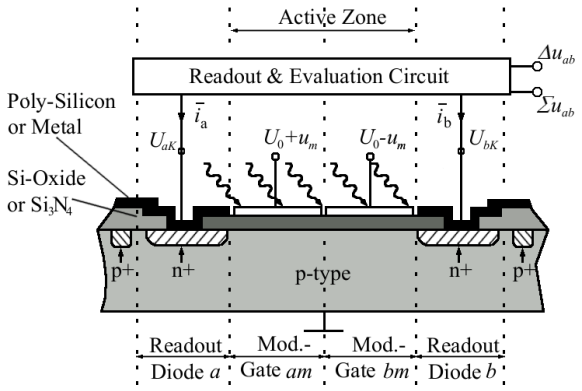


Figure 2.3: PMD Pixel Cross Section [Schwarte et al., 1997]

Each of the photodiodes A and B can be read as any normal diode. With the transparent control gates set at a common voltage, the photocurrent will be split equally between the two diodes. By using different voltages on the control gates it is possible to direct the photocurrent to one or other of the diodes. Typically the control voltages would be signals 180° out of phase with each other, i.e. a push-pull type signal as in a normal differential mixer. Taking the potential difference between gate A and gate B as V_{ab} the relationship between the gate voltages and how the current is divided can be explained. As already mentioned, with $V_{ab} = 0\,\mathrm{V}$, i.e. the gates at the same voltage, the charge flow is split equally between the two diodes. As V_{ab} becomes more negative, a potential difference difference is created across the

surface of the device, causing the drift of the electrons to favour heading toward diode A. The limiting case of this is show in Figure 2.4, where all of the electrons are flowing toward diode A. In practice, the split is never perfect and some electrons do actually flow to the diode B.

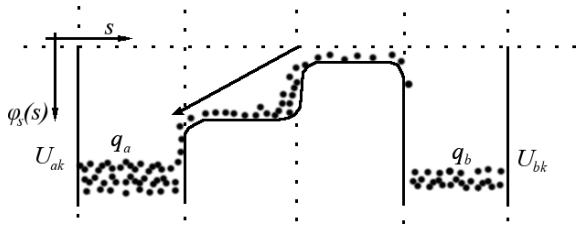


Figure 2.4: PMD Pixel Operation [Schwarte et al., 1997]

If V_{ab} now becomes more positive, then the increasing majority of the electrons will flow toward diode B.

To detect the currents, each diode is connected to a current readout circuit which integrates the current onto a capacitor and provides a voltage output, with a reset transistor to start the integration period. This is in essence a very similar principle to that used in the CCD domain [Povel et al., 1990, Spirig et al., 1995, Spirig et al., 1997] where the charge is integrated in the CCD and then stored in different regions.

Schwarte also proposes a second similar device which has three control gates instead of two, but still with just two diodes. In this device, the outer two control gates are operated as previously, but with the middle gate being held at the voltage mid way between the extremes that are applied to the other gates. The effect of the middle gate is to make the potential distribution across the device smoother which results in a more efficient transfer of the electrons. The charge transfer characteristics of the two devices are compared in Figure 2.5. The 3 gate device is better in all regards - it is more linear, has a higher modulation range and works over a smaller gate voltage range.

The transfer characteristics of the two PMD types can be seen in Figure 2.5, with both devices having the same area (hence the two gate device has gates of length

15 μm and the three gate device has gates of length 10 μm)

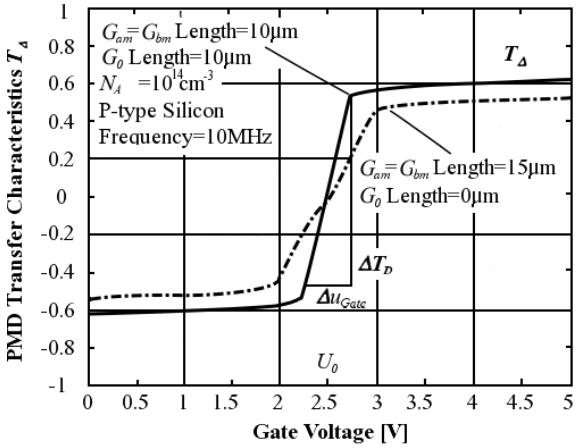


Figure 2.5: 2 Gate PMD (dashed line) and 3 Gate PMD (solid line) Charge Transfer Characteristics [Schwarte et al., 1997]

The operation of the 2 gate device is experimentally demonstrated by illuminating the PMD with a constant amount of light, with the voltage across the control gates being a square wave at the extremes of the allowed modulation to demonstrate the limits of the change in charge transfer which is clear to see in the differing slopes of the capacitor discharge.

The inventors of the PMD are most interested in using it in range finding applications so they have considered its operation almost exclusively for that case. They discuss the use of the device for optical level gauging and 3D-imaging [Gulden et al., 2001, Gulden et al., 2002]. In this case the modulation on the control gates is a sine wave rather than a square wave that makes for cleaner mixing which is much more directly analogous to "normal" continuous wave mixing.

Under the appropriate modulation, that is when both the light and the control gates are controlled by the same modulation source, the light received at the detector having been reflected from an object will have undergone a phase shift proportional to the distance of the object, as discussed in section 2.2.3. This phase shift can be calculated by using the difference between the two photodiode voltages after integration as in Equation 2.2 below:

$$V_d(t) = \kappa cos(\Delta\phi_{tof}) \tag{2.2}$$

Where $V_d(t)$ is the difference voltage, κ is dependent upon the device and optical path characteristics, as well as the intensity of the modulation and the integration time, and $\Delta\phi_{tof}$ is the phase change due to the time of flight. The time of flight and the distance can be calculated using Equation 2.1.

Gulden was able to measure the distance of two objects at the same time to an accuracy of 2.5 cm using a technique known as frequency stepped continuous wave (FSCW) which involves taking distance measurements over a stepped range of frequencies and then using spectrum analysis to find the distances.

Schwarte holds a patent on the PMD as well as its use in various applications [Schwarte, 1999].

2.3.3 Ohta

The simple two phase integration idea was taken up by [Ohta et al., 2003] who designed a 64 x 64 pixel sensor. Each pixel consists of a photogate and two separately controlled floating diffusions (FD1 and FD2) used to accumulate charge. The pixel schematic is shown in Figure 2.6.

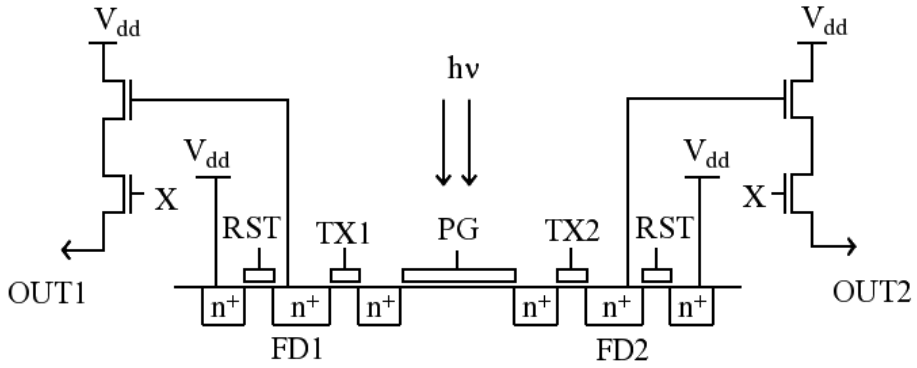


Figure 2.6: Demodulation Pixel Circuit [Ohta et al., 2003]

This sensor was designed for use in detecting modulated light, but specifically for the limited case where the modulated signal is a square wave and either totally off

or a given value of "on". By integrating first when the modulated signal is turned on and second when it is off, the voltage at FD1 will consist of the background and the modulated component whereas the voltage at FD2 will just be caused by the background component. The two outputs are multiplexed off chip, amplified and then read by an analogue to digital converter (ADC) connected to a computer, where they are subtracted from one another to give the intensity of the modulated signal. By choosing the number of times that the signal is integrated into FD1 and FD2, a trade off can be made between frame rate and sensitivity. There is no particular reason why this detector couldn't be used as a more general modulated light camera for detecting a sine wave rather than an on-off modulated wave.

2.3.4 Oike

In most modulated light applications, the background DC light is significantly larger than the modulated signal. Indeed, this is often one of the reasons that the modulation is introduced in the first place namely to make it easier to extract the signal. One problem with trying to use ordinary cameras in this situation is that the dynamic range may not be great enough to cope with the high level of background light whilst still being able to detect a small signal effectively. One technique for increasing the dynamic range, the logarithmic pixel, has already been introduced which produces a voltage output proportional to the log of the input current. Another way of dealing with this problem is to reduce the amount of background light. This is sometimes possible within the scope of the experiment by enclosing everything within a light protective box, but this is not possible for all applications and even where it is applicable the background light level may still be appreciable compared to the signal. A more general solution was proposed by Oike [Oike et al., 2002, Oike et al., 2003]. In this design the current flowing through the photodiode is not immediately converted into a voltage but first mirrored using a cascode mirror. The output of this mirror is then mirrored again into two channels. The first channel acts as a current sink and

has a low pass filter at its input. The second channel is mirrored a final time so that it acts as a current source. At this point the two channels are connected together which has the result that the current flowing into the I-V converter is the difference between the two currents. To a rough approximation this means the current flowing is

$$\begin{aligned} I_1 &= DC + AC \\ I_2 &= DC \\ I_1 - I_2 &= (DC + AC) - DC \\ &= AC \end{aligned}$$

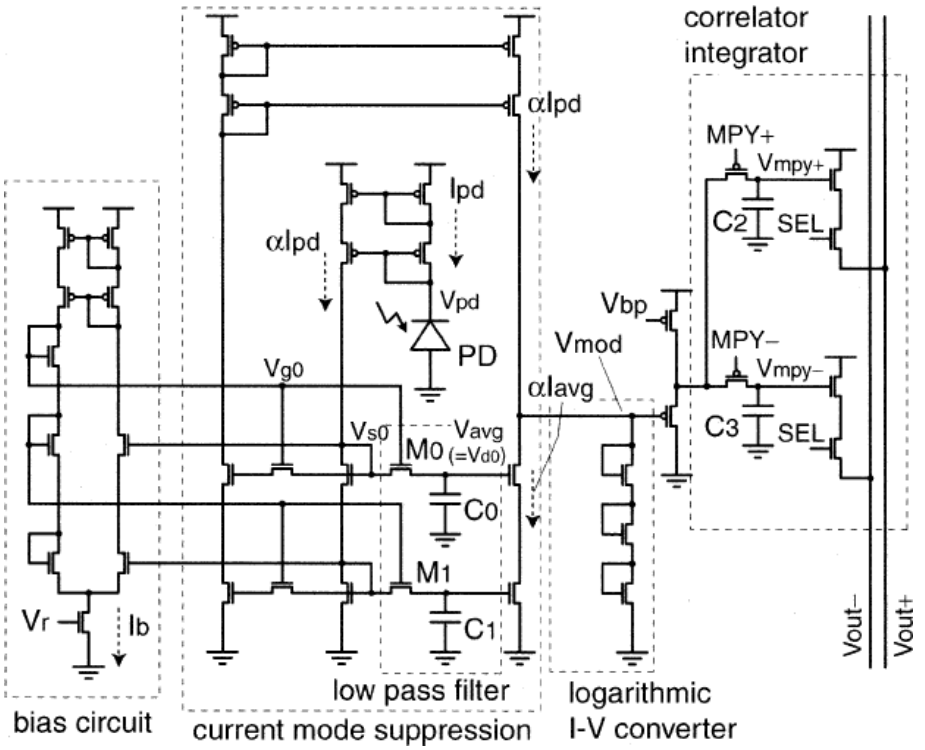


Figure 2.7: Current Suppression Pixel Sensor [Oike et al., 2002]

The pixel design is shown in Figure 2.7. After the DC suppression the AC current is converted into a voltage with a log frontend as with other designs. To carry out demodulation, MPY+ and MPY- are clocked with non-overlapping signals that are at half the frequency of the modulated light and 180° degrees out of phase with each other. In the presence of modulated light of the correct frequency and given

the correct phase between the light and the clocking, the charge on MPY+ will increase more rapidly than on MPY- hence producing a higher output voltage on Vout+ than Vout-. Measuring the difference between Vout+ and Vout- gives an indication of the amplitude of the AC light. The camera was developed for making range measurements based on the light-stripe method as already described. With the use of both modulated light and background light suppression, the ability of the camera to work in bright environments is enhanced. The end result for each measurement is an image where each pixel has a binary state - either the stripe appears at a particular pixel or it doesn't. The difference between Vout+ and Vout- is fed into a comparator, where it is compared with a predefined reference. If the difference is greater than the reference, then modulated light is defined to be present (the stripe exists) and if not, then no modulated light is present. This simple end result means that the output bandwidth of the chip is greatly reduced and is easy to work with. As it stands, however, the camera design is not easily usable in other applications, but by taking the difference between Vout+ and Vout- after a set time as the result then it could function as a modulated light camera capable of measuring the amplitude of the modulated light rather than just its presence or lack thereof. There is of course no reason why both modes of operation couldn't be included on the same device, allowing textured range data if so desired.

2.4 Continuous Time CMOS Cameras

2.4.1 Benten

The cameras described so far have all measured the light signal by integrating the electrons produced at the photosensor. Although this fits in well with existing camera techniques, traditional lock-in techniques tend to work on a continuous wave signal. It is therefore natural for authors to have considered the same principle for cameras. Such a device is proposed by Benten *et al.*, who described a method

for demodulating low frequency light signals for use in measuring arterial oxygen saturation and dental pulp vitality [Benten et al., 1997]. The interesting part of these two applications is that in both cases the light striking the photodiode consists of two different modulation frequencies to be measured at the same time. The sensor consists of two channels, one for detecting each of the two different signals. Each channel is a switching multiplier (square wave mixer), where the voltage input is either multiplied by +1 or by -1 by switching between two amplifiers. The next stage is a low pass filter, the input of which is switched between the +1 and -1 amplifiers. The waveforms for this process are shown in Figure 2.8, where the reference signal controls the switch and chooses the +1 output when it is high, the -1 output when it is low. As such, the detector can be thought of as a standard lock-in.

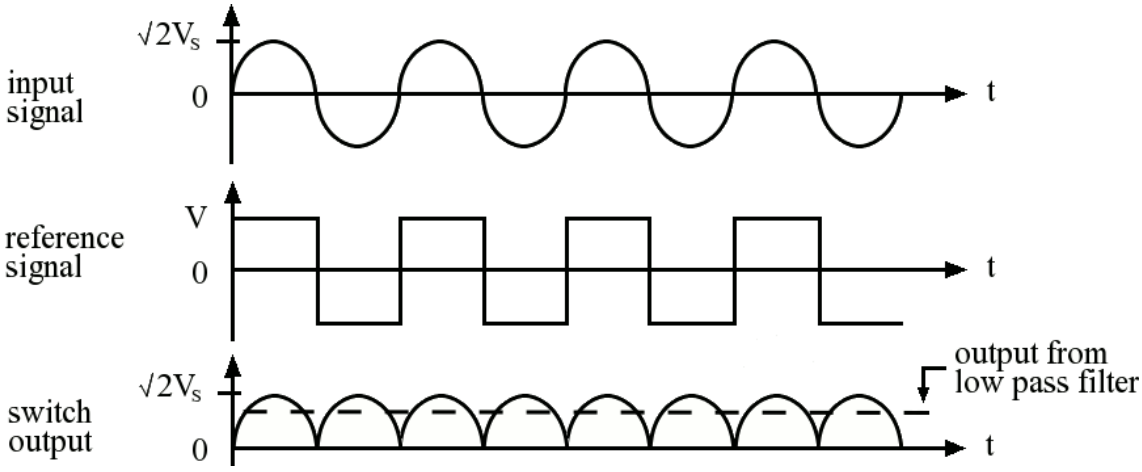


Figure 2.8: Switching Multiplier Operation [Benten et al., 1997]

The output of the low pass filter is proportional to the amplitude of the signal. Using a square wave reference signal for a mixer produces additional frequency sensitivities at the odd harmonics of the reference, but zero sensitivity at the even harmonics. For this reason, the modulation frequencies chosen for the two channels were f_s and $2f_s$. Providing that the bandwidth of the two channels do not overlap there should be no cross contamination of the signals. As the application involves making measurements within the mouth, the sensor must be very compact and this is achieved by using optical fibres to deliver and sense the signal; so using two fibres to transmit the two different modulated light signals and a third fibre to receive the

reflected signal. Using a fibre means that only a single pixel is required which does mean that fewer compromises need to be taken in the design of the sensor, so both the signal amplification and the noise performance can be optimised to a degree not normally possible in a camera array. The transimpedance amplifier demonstrates this principle well; it consists of a BiCMOS amplifier, chosen to minimise 1/f noise, with a 2 MΩ resistor in the feedback loop. Both of these choices provide benefits over the more usual transimpedance amplifiers, namely the linearity of the response and the low noise performance, but they are also wholly impractical in the context of a camera since the full two channel device occupies an area 2.9 x 0.45 mm.

The detector as described is a very specialised device; it has only a single detection element and that one detector is used to detect multiple frequencies that are quite low at just 5 and 10 kHz, although this is quite sufficient for the application. Additionally, it is unusual in being a BiCMOS device rather than the more common CMOS or CCD technologies. However it has been manufactured and shown to work with minimal cross talk between the two detection channels.

2.4.2 Bourquin

A similar continuous wave detector is described by Bourquin [Bourquin et al., 2001]. Each of their 58 x 58 pixels consists of a photodiode connected in series with a voltage controlled current source. The output of the photodiode is fed into an amplifier which is then fed into a low pass filter, as shown in Figure 2.9.

The amplifier and low pass filter feedback loop ensure that the amplifier is kept at the best operating point whilst also tuning the frequency response; when the modulation frequency of the light is higher than the cut off frequency of the filter, the signal is not compensated for at the voltage controlled current source and so the gain of the amplifier is higher than otherwise. The output of the amplifier is then fed to a square wave mixer and low-pass filter for standard lock-in detection. The filter

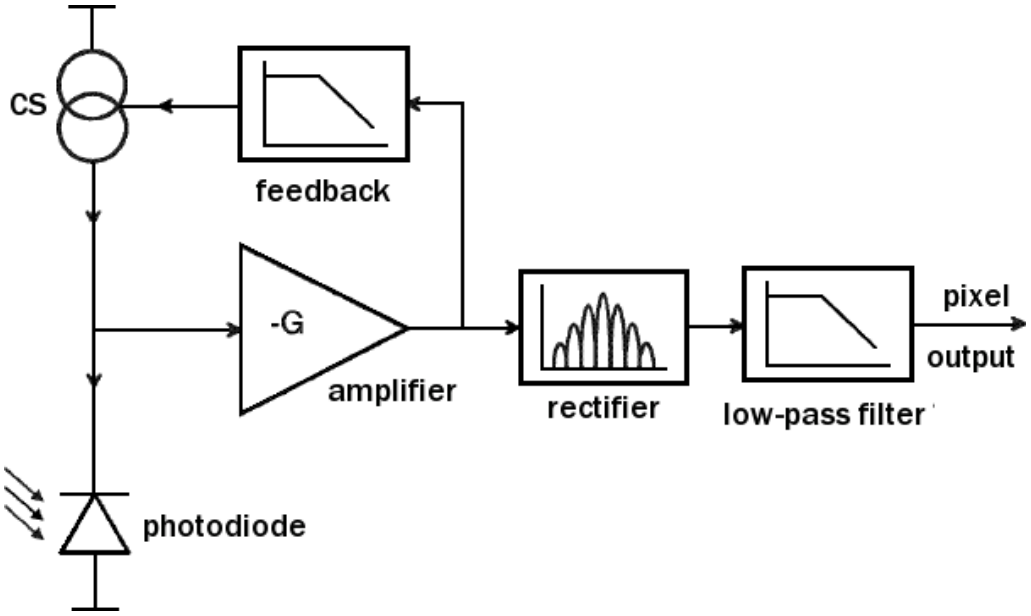


Figure 2.9: Smart Detector Pixel Schematic [Bourquin et al., 2001]

in the feedback loop is adjustable within each pixel so the response of the camera can be controlled and is capable of detecting signals in the range of modulation of 1 kHz to 1 MHz based on this adjustment.

2.4.3 Ducros

Bourquin's camera was used by Ducros *et. al.* to perform parallel optical coherence tomography measurements [Ducros et al., 2002]. Optical coherence tomography (OCT) is a technique for measuring the coherent backscatter from scattering samples at different depths and it typically suffers from having the signal of interest buried within a very high DC signal, so a lock-in camera is ideal for providing a better result. In OCT a wave is split into a reference beam and the sample beam, which is reflected off a sample. The two beams are interfered with each other at the detector. The path length of the reference beam is scanned and this causes the interference at the detector to move between being constructive and destructive so given a constant reflectivity at the sample, the resulting signal at the detector will be a sine wave with frequency determined by the rate at which the path length of the reference beam changes. OCT is typically used for measuring biological samples

and the technique described allows depth measurements to be made, with the amplitude of the sine wave at the receiver indicating the reflectivity of the sample at the depth corresponding the the reference path length. One of the examples provided by Ducros was of an onion imaged at the surface and at a depth of 100 μm as shown in figure 2.10.

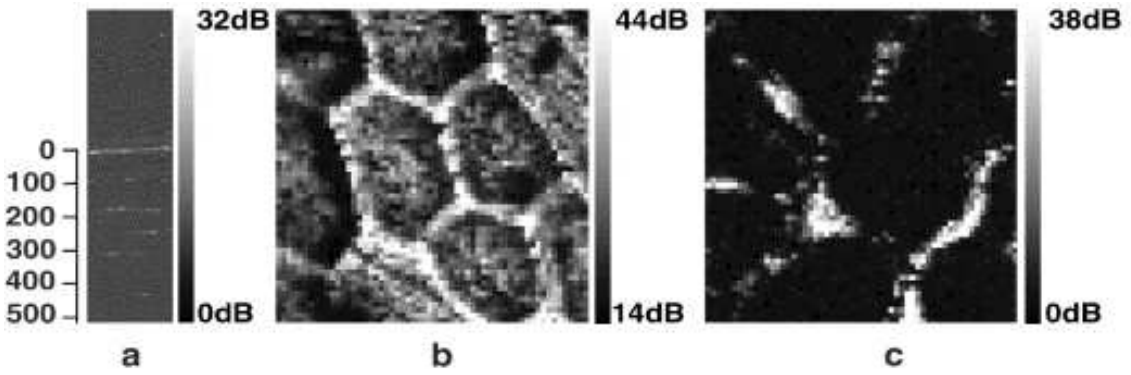


Figure 2.10: OCT image of an onion cross section (a) depth scan, (b) *en face* at depth 0 and (c) at 100 μm [Ducros et al., 2002]

2.4.4 Lu

Another more conventional approach for using a continuous wave signal is to use a mixer with a reference signal to do demodulation as in ordinary analogue communications. A number of people have used this technique in slightly different ways [Lu et al., 2002]. The essential architecture consists of a transimpedance amplifier, differential amplifier, mixer and finally a low pass filter, however the overall design is substantially different from the others considered so far in that it uses a buried double *pn* junction (BDJ) detector. The BDJ, as discussed in section 1.4.2.3, has two outputs that are sensitive to different ranges of wavelengths due to their different junction depths. The two outputs of the BDJ are demodulated with two identical channels as shown in Figure 2.11.

The focus of the work is on minimising noise and increasing sensitivity rather than trying to achieve a high speed and small size. The same design of amplifier used in the transimpedance amplifier is also reused in the differential amplifier. With each of

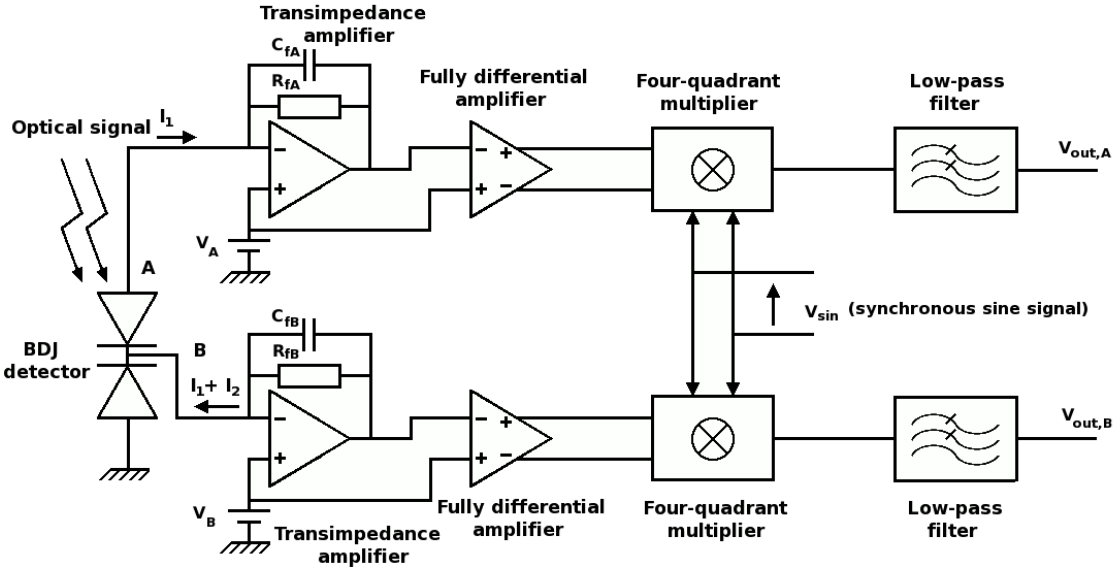


Figure 2.11: BDJ Pixel Sensor[Lu et al., 2002]

these amplifiers occupying 0.2 mm^2 and consuming 1.3 mW , it can be seen that the design will take up a substantial area. The overall area and power consumption is not mentioned, but estimates can be made; the photo detector is 1 mm^2 whilst each mixer occupies 0.15 mm^2 and consumes 6 mW. Even without taking into account the two low pass filters, which must occupy a not insignificant silicon area having two 10 pF capacitors each, the total area of the pixel is at least 2.5 mm^2 and the power consumption is 19.8 mW, making it not a particularly practical device for integrating into a full field camera. The functionality of the detector is quite interesting however. By using the BDJ it can potentially measure two different wavelengths of light at the same time, if their spectra are sufficiently well separated, or alternatively it can measure the wavelength of the light striking the detector by calculating the ratio of the two low pass filter outputs. Both of these properties are useful in fluorescence imaging, where dyes or other markers that fluoresce are added to a biological sample. When a dye molecule is illuminated with light of an appropriate wavelength, typically in the ultraviolet range, it will absorb the energy and be stimulated into an excited state that lasts on the order of a few nano seconds. Whilst the molecule is in the excited state, it loses energy that is dissipated as heat and will then decay back to the ground state for the molecule, with the remaining

energy being emitted as a photon. The emitted photon has less energy than the incoming photon because of the energy lost to heat whilst in the excited state, causing the photon to be in the visible band. The emitted wavelength of light will be constant for a given dye and excitation wavelength. With the short time between excitation and emission, stimulating the dye with a laser passing through a beam chopper, which produces a square modulation, allows the lock-in detectors to be used. The advantage of Lu's proposed detector is that it could be used in fluorescence experiments where two dyes with different emission wavelengths were in use, without the need to split the different wavelengths to separate detectors. Likewise, although the wavelength of the emission is generally known, local effects on the sample can change the emitted wavelength and Lu's detector could potentially measure this change.

2.4.5 Pitter

Pitter [Pitter et al., 2003] fabricated a modulated light camera for measuring modulated thermoreflectance. The 4 x 4 pixel camera converts the incident light into a current and then a voltage in continuous time using a logarithmic front end. To perform the demodulation the signal is amplified and bandpass filtered and finally passed into two channels of a switched capacitor and integrating capacitor demodulator, where the different channels are for the I and Q phases. The circuit for each pixel is shown in Figure 2.12.

Considering just the I channel, C1 is the switched capacitor and C2 is the integrating capacitor. Phi1 and and phi2 are non-overlapping clocks which are at the same frequency as the modulated light signal to be detected. The signal is sampled onto C1 by phi1 and then integrated onto C2 by phi2. The Q channel is identical except that the clock signals phi1' and phi2' are 90° out of phase with phi1 and phi2. For each channel the signal is sampled once per signal cycle at the same phase each time which allows the simple sample and integrate method of demodulation but does bring

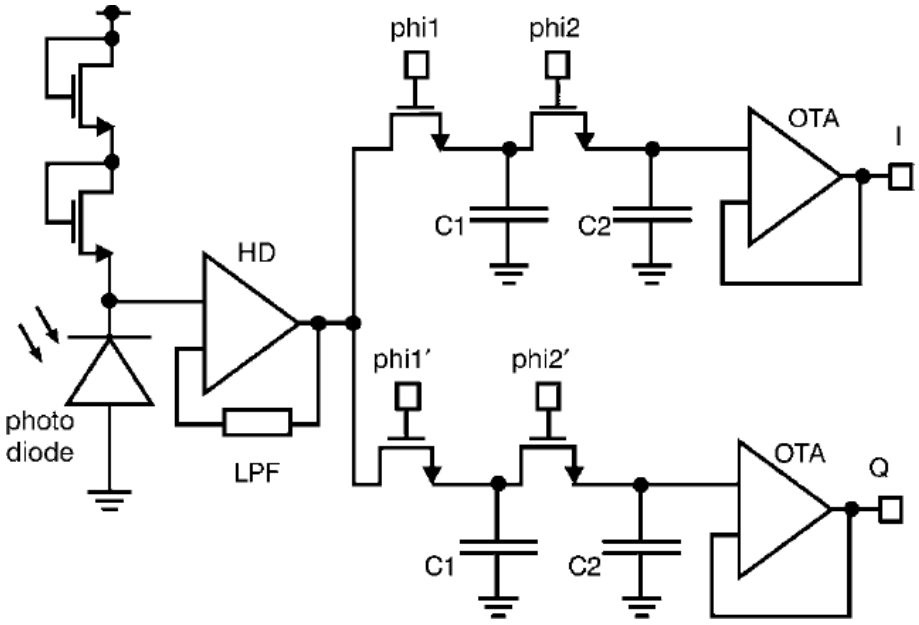


Figure 2.12: Modulated Thermoreflectance Pixel Circuit [Pitter et al., 2003]

with it a number of disadvantages; although the front end of the pixel is continuous time, the mixing is not so it wastes the rest of the signal. More significantly, the frequency selectivity of the pixel is quite wide and as well as being sensitive to the odd signal harmonics it is also sensitive to the even harmonics. These problems can be reduced by averaging over multiple samples, but is only a reduction rather than an elimination. In the case of the thermoreflectance experiments the camera being sensitive to other frequencies is not really an issue because the measurements are carried out in a controlled environment. The pixels as manufactured are $200\,\mathrm{x}\,200\,\mu$m with $60\,\mathrm{x}\,60\,\mu$m photodiodes for a fill factor of 9%.

The prototype camera was later followed up with a 64 x 64 pixel camera [Pitter et al., 2004]. The design is much the same as the prototype except that the front end is slightly modified and the camera is a "column" based design rather than a pixel based design - that is to say that although the pixel contains a current to voltage converter, buffer and a select switch, the bulk of the processing performed, i.e. the amplification, filtering and demodulation, is done outside of the pixel. A column of 64 pixels share the same signal processing circuitry. Although this means that the individual pixels can be much smaller at $25\,\mu$m square, with a fill factor of

56%, only one row of pixels can be used for detection at any one time. This does negate the benefit of the improved fill factor somewhat because 63 out of the 64 rows are not in use; an apparent waste of space. However, by having a complete array of pixels rather than just a single row it is possible to electronically scan through the rows rather than having to physically move the device or the laser to take a two dimensional image. The pixel front end used is again the logarithmic design, but with the addition of feedback to reduce the effect of the diode capacitance and so increase the bandwidth.

2.4.6 Dmochowski

Similar to Lu (section 2.4.4), Dmochowski [Dmochowski et al., 2004] created a pixel using a continuous wave mixer, but with two channels for I and Q instead of different wavelengths, and an emphasis on speed. The pixel consists of a log-pixel frontend with feedback to provide an increased bandwidth, a differential amplifier and a four-quadrant mixer to perform the demodulation with the mixer output then low pass filtered before being sent off pixel. The pixel can detect modulated light from 500 kHz to 25 MHz. This type of design is very attractive since the pixel provides all of the required demodulation functionality over a wide range of frequencies whilst having a very narrow frequency sensitivity as shown in Figure 2.13.

As the circuit operates in continuous time, all of the light that strikes the photodetector is used unlike in the integrating type of pixels. There are a number of unattractive features though; at $250\,\mathrm{x}\,250\,\mu$m, the pixel is very large. Even with a fairly large photodiode at $50\,\mathrm{x}\,50\,\mu$m the fill factor is very low at just 4%. Having a large amount of circuitry in a single pixel also leads to a high current consumption. In this case a single pixel uses $180\,\mu$A which is significantly more than the pixels described already. Both of these factors make scaling the pixel up to a camera a challenge.

Dmochoswki followed this work with a much more compact pixel design in a $24\,\mathrm{x}\,32$

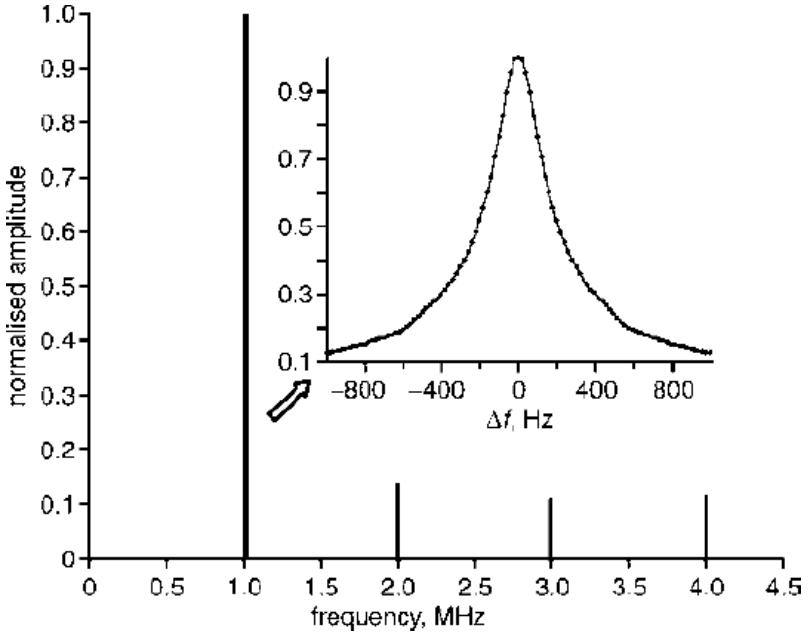


Figure 2.13: Synchronous Pixel Frequency Response [Dmochowski et al., 2004]

array [Dmochowski, 2006]. The pixels are almost a quarter of the area of the previous design, being 130 x 130 μm, with the photodiode also shrinking to 30 x 30 μm which results in a slight improvement in fill factor to 5.3%. The range of frequencies detectable is increased to the range 10 kHz to 30 MHz, although this is at the cost of a higher current consumption, which increased to 200 μA per pixel, leading to a power consumption of 180 mA for the entire chip with output buffers and control circuitry included.

2.5 Other Cameras

Some cameras that have been designed for other purposes may also be applied to the task of measuring modulated light. Although they may not perform as well as cameras designed specifically for the purpose it is still useful to consider them, especially as they can highlight some of the applications where custom lock-in cameras are required.

2.5.1 Ma

Ma presents a CMOS APS camera that could certainly be used as a simple lock-in camera [Ma and Chen, 1999]. The sensor is a 128 x 96 pixel direct frame difference camera manufactured on a 0.5 μm CMOS process. It operates by taking one sample and storing it at the pixel and then taking a second sample. When both samples have been taken, the difference between the two signals is output from the pixel. The camera is intended for use as a motion detector as only changes in the scene will be recorded, but it resembles the device already mentioned by Povel [Povel et al., 1990]. Again, by choosing the reset frequency to be twice the frequency of the modulation frequency it would be possible to demodulate a signal. As the subtraction is done at the pixel level it would be possible to get a value for the amplitude although measuring the DC offset or phase of the light is not possible from those samples. The camera does offer a different mode of operation to read out "normal" images which could potentially be used to measure the signal DC offset if used correctly. The big advantage of using CMOS sensors over CCD sensors is that CMOS sensors allow the integration of processing circuitry as well as the sensors. This advantage is exploited on this chip by integrating an analogue to digital converter. This ADC appears to be the bottleneck in the design; with a maximum speed of 3 MSamples/s the frame rate is limited to roughly 250 frames/s. Unfortunately the authors do not mention the minimum time between the first and second samples (understandably, as it is not particularly relevant to their work), but it can be assumed that it could detect modulated signals of at least 250 Hz. Faster operation may be possible although the frame rate would only ever be as high as 250 frames/s. With only a single modulated cycle being considered, it is quite likely that the output would be fairly poor unless a large level of modulation was present. Either way, this is one of the first examples of a CMOS camera similar in operation to the existing lock-in CCD cameras. Other examples are [Dickinson et al., 1995, Nixon et al., 1996, Nomura et al., 1998].

2.5.2 Kleinfelder

A second example of a camera not originally designed for use as a modulated camera is presented in [Kleinfelder et al., 2001]. This work discusses the design of a 10 000 frames/s CMOS camera. This frame rate is enough to measure modulated light signals at 5 kHz if only amplitude measurement is required or at 2.5 kHz if all of amplitude, phase and intensity are required. This range of modulation frequencies is far enough away from 1/f noise to be attractive yet without being high enough that the design becomes more complicated or power hungry as would be required for radio-frequency design. This design also illustrates another type of pixel - the digital pixel sensor (DPS). The DPS differs from the CCD and the APS in that it directly produces an output that is digital. This means that there must effectively be an ADC at every pixel. Bearing in mind that the ADC used in [Ma and Chen, 1999] consumes 1 mm^2 of silicon, this appears to be a significant task. In reality a traditional ADC is not used, instead, each pixel consists of a photogate connected through a switch to a storage capacitor which is in turn connected to a comparator, with the output of the comparator being connected to an 8-bit memory. To make a measurement, the light is integrated onto the photogate. After the integration is finished, the charge on the photogate is transferred to the storage capacitor. At this point, a voltage ramp is introduced on the second input to the comparator. The comparator output triggers when the voltage on the capacitor and on the ramp are equal and this causes the memory to store the current value of a grey code count, which is available on a bus at each pixel. The memory values can then be read out at high speed. An advantage of the DPS is that stored values will not degrade over time as analogue voltages can. The pixels presented are just 9.4 μm square with a fill factor of 15%, as shown in Figure 2.14, an impressive design which really highlights the benefit of using CMOS processes (this kind of integration is only really possible with a small feature process such as the 0.18 μm 5 metal CMOS process used in their design).

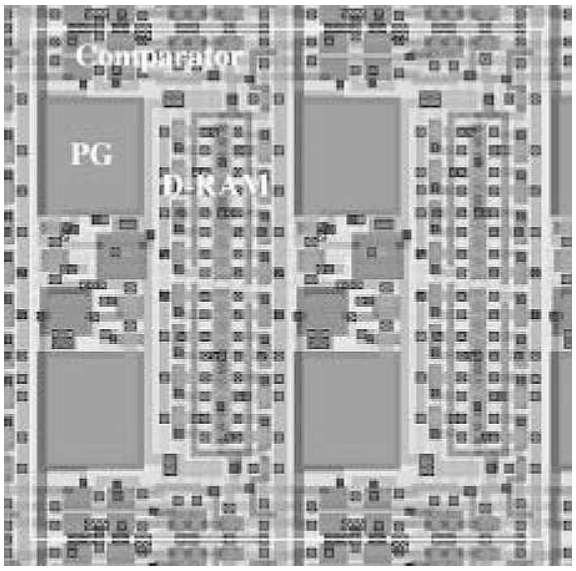


Figure 2.14: 2 x 2 DPS Layout [Kleinfelder et al., 2001]

As a comparison, if this design were implemented in a typical 0.35 μm 4 metal CMOS process, the area taken by the transistor circuitry would be at least 5 times greater due to the increased feature size and the loss of a metal layer. Assuming the photogate remained the same size, this would lead to a pixel size of approximately 19.7 μm square and a fill factor of 3.4%. Whilst the overall pixel size is still not unreasonably large, the fill factor would be very poor. Likewise, as a 0.18 μm process would now be considered large in the digital realm having been first used around the year 2000 [ITRS, 2007], it would be straightforward to shrink this design and achieve a much better fill factor without any further effort. Scaling to a now current 65 nm process would potentially mean a pixel 4.8 x 4.8 μm, with a fill factor of 57.5%. Unfortunately, changing to a process with a smaller feature size does not provide only advantages; as the lateral feature size decreases, so does the vertical feature size, meaning that all of the junctions are shallower, making them sensitive to shorter wavelengths. So why isn't this architecture used for modulated light cameras? There are a number of reasons; firstly, the voltage resolution of the output is rather low; in the case where the modulated signal is 1% of the background light it would probably only be represented by a single bit. It is of course possible to increase the signal information by using multiple frames of data but this does require further off-chip processing. The second reason is the data rate, this camera has 352 x 288 pixels,

each producing an 8-bit value at 10 kHz which leads to an off chip data rate of greater than 1 GB/s. The advantage of modulated light cameras is that a lot of the processing is carried out on chip and so the off-chip data rate is much reduced.

2.6 Comparison of Cameras

Author	Type	Application	Parameters Measured	Array Size	Pixel Size	Modulation Frequency
Povel et al.	CCD integrating	Polarimetry	Amplitude	352 x 288	22 μm	100 kHz
Spirig et al.	CCD integrating	Unspecified	Amplitude, phase	3 x 3	87.75 x 80 μm	100 kHz
Lange et al.	CCD integrating	Range finding	Amplitude	64 x 25	21 x 65 μm	20 MHz
Ando et al.	CMOS integrating	Unspecified	Amplitude, phase	64 x 64, 100 x 100	62 μm, 96 μm	40 kHz
Gulden et al.	CMOS integrating	Range finding	Phase	Unspecified	Unspecified	Unspecified
Ohta et al.	CMOS integrating	Motion capture	Amplitude	64 x 64	42 μm	2.5 kHz
Oike et al.	CMOS continuous time	Range Finding	Amplitude	64 x 64	40 μm	10 kHz
Benten et al.	BiCMOS continuous time	Arterial oxygen saturation	Amplitude	Single pixel	Unspecified	5 kHz, 10 kHz
Ducros et al.	CMOS continuous time	Optical coherence tomography	Amplitude	58 x 58	110 μm	1 MHz
Lu et al.	CMOS continuous time	Fluorescence imaging	Amplitude	Single pixel	1 mm	10 Hz
Pitter et al.	CMOS continuous time	Modulated thermoreflectance	Amplitude	4 x 4, 64 x 64	25 μm	2 kHz
Dmochowski et al.	CMOS continuous time	Unspecified	Amplitude, phase	24 x 32	250 μm, 130 μm	10 kHz - 30 MHz

Table 2.1: Modulated Camera Comparison

2.7 Summary

A variety of existing lock-in camera designs have been presented and compared, showing the wide variety of different implementations possible. Modulated cameras are used for measuring the phase of the modulated signal for range finding, or for measuring the amplitude for any applications where the signal of interest is small compared to noise or ambient light. It can be seen that the trend in lock-in camera design is toward using standard CMOS processes, or even slightly more exotic BiCMOS processes, rather than the more traditional CCD based sensors. CMOS offers the possibility of including on chip processing at both the pixel and column level which reduces the data I/O bottleneck. CMOS modulated light cameras are either continuous time (usually employing a logarithmic pixel or variant) or integrating (with a photodiode or photogate).

The modulated light cameras presented all use analogue techniques to carry out measurement and processing of their signals. This means that they rely on a separate ADC to convert the signal into digital form so that it can be displayed or analysed further. This work goes on to present a pixel design for use in modulated light cameras that provides a digital value without the need for a conventional ADC as well an improved algorithm for demodulation.

Chapter 3

Lock-In Theory

3.1 Introduction

This chapter introduces general signal modulation theory, explaining why it is desirable to have a modulated signal. Section 3.5 details the different ways in which amplitude demodulation can be carried out by using a lock-in sensor, which is a detector that is tuned to capture a narrow band of frequencies.

Modulation refers to the act of taking a signal and changing it so that it can be transmitted more efficiently. There are three modulation techniques that can be applied to analogue signals: amplitude modulation, frequency modulation and phase modulation. Each of these techniques uses a high frequency sine wave as a carrier and varies one parameter of the carrier in proportion to the amplitude of the signal to be transmitted, also known as the baseband signal. Varying the amplitude, frequency or phase of the carrier leads to amplitude, frequency and phase modulation respectively. Collectively these modulation techniques are referred to as continuous wave modulation. In each case, the baseband signal is shifted in the frequency domain so that it is centred around the carrier frequency rather than around DC.

Many other types of modulation exist but are primarily intended for digital communications. They are not relevant to this work and so are not considered here.

3.2 Rationale Behind Modulation

3.2.1 Modulation in Voice Communications

As the most well known use of modulation is in communications it is of no surprise that the benefits of modulation are of particular use in that area. In voice communications the baseband signal occupies the range of frequencies from about 300 Hz to 3 kHz. In order to transmit a signal effectively as an electromagnetic wave, the antenna should be one half the wavelength of the signal. This would mean an antenna of between $50 - 500$ km in length would be required to transmit a voice signal effectively, which is clearly impossible. By shifting the baseband up to a high carrier frequency, the size of the antenna can be suitably reduced; at a carrier frequency of 30 MHz for example, the antenna could be reduced to just 5 m. Even ignoring the problem of transmitting at baseband frequencies, modulation is still desirable. Without modulation, all voice transmissions would take place in the same frequency band and would hence interfere with each other, making it impossible to transmit more than one signal at once. With modulation, different transmissions can be shifted to different carrier frequencies and so many channels can be used at once. Many types of transmission are restricted to particular frequency bands and so must be shifted into those bands. For example, the FM radio broadcast band in the UK is $87.5 - 108$ MHz[OFCOM, 2007].

3.2.2 Modulation in Optical Experiments

In optical experiments, modulation is either used when the baseband signal is very small so that it can be more readily separated from any other light and noise sources in the system, or else the experiment itself produces the modulation, perhaps as a vibration, a fluid flow or some other process. In particular, using modulation allows the baseband signal to be moved away from low frequency (1/f) noise and also the "background" light.

When the experiment itself produces the modulation, the parameter that is most often being measured is the frequency content of the signal as is the case when imaging blood flow[Kongsavatsak, 2005].

3.3 Types of Modulation

3.3.1 Amplitude Modulation (AM)

Amplitude modulation is the simplest type of modulation to understand. The amplitude of a high frequency carrier is modulated by being multiplied by the lower frequency input signal, $I_S(t)$, that is to be transmitted. This is shown in Equation 3.1.

$$I_M(t) = I_S(t)[A + B\cos(\omega t)] \tag{3.1}$$

With $A > 0$, this form of AM is known as "double side band" amplitude modulation because in the frequency domain there is a spike at the carrier frequency which has peaks either side of it that consist of the signal $I_S(t)$. If $A = 0$, the spike at the carrier no longer exists and so this form is known as "double sideband suppressed carrier".

AM was the original means of transmitting audio signals using radio waves as the carrier. It remained the dominant means of modulation until solid state devices became commonplace because AM radio receivers can be constructed with very few simple components. Amplitude modulation is not particularly robust since the amplitude of a transmitted signal can be perturbed relatively easily as it is distorted by the atmosphere and as it is reflected off different surfaces such as buildings and produces varying amounts of interference with itself at the receiver.

3.3.2 Phase Modulation (PM)

Phase modulation keeps the amplitude of the carrier constant whilst its phase is changed proportionally to the amplitude of the input signal, as shown in Equation 3.2.

$$I_M(t) = A + B\cos(\omega t + I_S(t)) \tag{3.2}$$

Phase modulation in an optical experiment is often caused when the carrier signal is reflected by an object and the object moves, so inducing a phase change. A full 360° phase change corresponds to the object moving by a distance equivalent to the wavelength of the carrier. This means that changes in the distance to an object can be measured by calculating the phase (with phase "wrapping" occurring if the movement is greater than the wavelength).

An example of a where physical effects cause phase changes is described by Sharples [Sharples et al., 2007], where a sample is excited by a laser causing surface acoustic waves to propagate through the material. These waves can be imaged with a probe laser and a detector. Both the amplitude and phase measured can give insights into the properties of the material being measured; specifically whether any defects are present.

3.3.3 Frequency Modulation (FM)

Frequency modulation produces an output close in appearance to PM but the input signal is used to vary the frequency of the carrier as in Equation 3.3. FM is not directly relevant to this work but is mentioned for completeness.

$$I_M(t) = A + B\cos([\omega + I_S(t)]t) \tag{3.3}$$

3.4 Amplitude Modulation Theory

This section considers the theory of taking a signal, using it to amplitude modulate a carrier and then demodulating and recovering the signal.

3.4.1 Modulation

The signal of interest is defined here as $I_S(t)$. It can be any time varying signal, whose bandwidth is assumed to be relatively low and centred at 0 Hz. This signal is the actual quantity being measured or transmitted, such as the slowly changing reflectance of an object or an audio signal. For simplicity of analysis, this is taken as a single frequency, δ, sine wave as shown in Equation 3.4 although in reality it can have any form. The carrier signal, at a frequency ω, is defined as $I_C(t)$, where $\omega \gg \delta$. The carrier is also usually a single sine wave as in Equation 3.5, but can take other forms such as a square wave in an optical experiment using a chopper for example. These two equations refer to modulation that is done to the light, they are unrelated to the optical frequency of the light.

$$I_S(t) = J + K\cos(\delta t) \tag{3.4}$$

$$I_C(t) = A + B\cos(\omega t) \tag{3.5}$$

The signals have 'DC' offsets of J and A respectively.

Using Equation 3.1, the modulated signal is given by the product of $I_S(t)$ and $I_C(t)$:

$$\begin{aligned} I_M(t) &= I_S(t) \cdot I_C(t) \\ &= [J + K\cos(\delta t)] \cdot [A + B \cdot \cos(\omega t)] \\ &= AJ + AK\cos(\delta t) + BJ\cos(\omega t) + BK\cos(\delta t)\cos(\omega t) \\ &= AJ + AK\cos(\delta t) + BJ\cos(\omega t) \\ &\quad + \frac{BK}{2}\cos(\omega + \delta)t + \frac{BK}{2}\cos(\omega - \delta)t \end{aligned} \tag{3.6}$$

This results in the frequency spectrum as shown in Figure 3.1, where the information that is required (i.e. J and K), is stored entirely in the carrier at ω (as BJ), and the side bands at $\omega \pm \delta$ (as $\frac{BK}{2}$, shown in red). The other components at DC and δ also contain the same information, but attempting to use them negates the whole point of the modulation since these are at baseband.

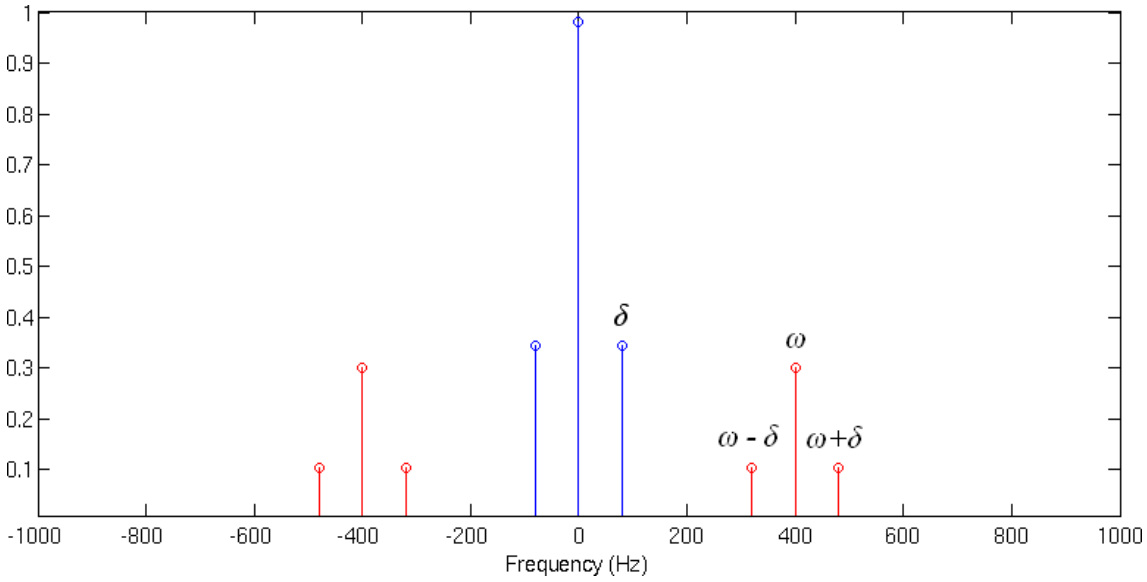


Figure 3.1: Amplitude Modulated Frequency Spectrum

In voice communications, the components at δ and DC must not be present so that the transmitted signal only occupies its allocated bandwidth. By using a carrier with zero DC offset (known as double sideband suppressed carrier), i.e. $A = 0$, the δ and DC components disappear from Equation 3.6 to give Equation 3.7, which consists of the carrier, ω, and two sidebands containing the signal at $\omega \pm \delta$.

$$I_M(t) = BJ\cos(\omega t) + \frac{BK}{2}\cos(\omega + \delta)t + \frac{BK}{2}\cos(\omega - \delta)t \tag{3.7}$$

In optics, creating a signal with zero DC offset (i.e. zero mean), would require the presence of "negative light", hence there must always be a DC offset on the carrier.

3.4.2 Demodulation

After the combined signal and carrier has been received, the signal, $I_S(t)$, must be extracted. One way would be to build a band pass filter centred around ω with bandwidth 2δ. Building a bandpass filter with these specifications is difficult, so it is common to demodulate the signal by mixing it with a sine wave at the same frequency and phase as the carrier, which results in the carrier and side bands being shifted in the frequency domain so that they are centred around DC as they were before being modulated.

The signal used to carry out the demodulation is known as the local oscillator (LO) and is shown in Equation 3.8.

$$I_{LO}(t) = M + N\cos(\omega t) \tag{3.8}$$

Multiplying the two signals $I_M(t)$ and $I_{LO}(t)$ together produces $I_D(t)$, the demodulated signal:

$$\begin{aligned} I_D(t) &= I_M(t) \cdot I_{LO}(t) && (3.9)\\ &= [AJ + AK\cos(\delta t) + BJ\cos(\omega t) \\ &\quad + \frac{BK}{2}\cos(\omega+\delta)t + \frac{BK}{2}\cos(\omega-\delta)t] \cdot [M + N\cdot\cos(\omega t)] \\ &= AJM + \frac{BJN}{2} + (AKM + \frac{BKN}{2})\cos(\delta t) && (3.10)\\ &\quad + (AJN + BJM)\cos(\omega t) \\ &\quad + \frac{AKN + BKM}{2}\cos(\omega+\delta)t + \frac{AKN + BKM}{2}\cos(\omega-\delta)t \\ &\quad + \frac{BJN}{2}\cos(2\omega t) + \frac{BKN}{4}\cos(2\omega+\delta)t + \frac{BKN}{4}\cos(2\omega-\delta)t \end{aligned}$$

This results in the frequency spectrum shown in Figure 3.2. The signal of interest, previously at ω and $\omega \pm \delta$ has been shifted to DC and δ (in red), as well as 2ω and $2\omega \pm \delta$. The low frequency components at DC and δ are the ones that will be used

to reconstruct the original signal - the high frequency components around ω and 2ω are mixing artifacts and are of no interest.

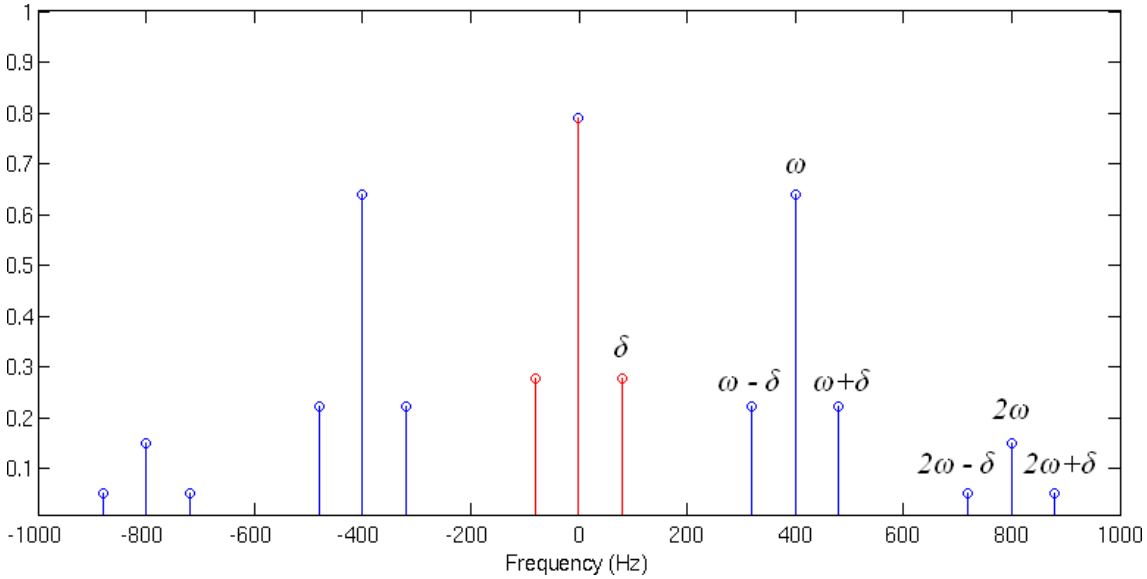


Figure 3.2: Demodulated AM Frequency Spectrum

As we can usually control the LO, it makes sense to simplify $I_D(t)$ by setting its DC offset, M, to zero. This causes all of the terms with an M in to drop out of the equation, making it much simpler whilst not losing any information; to recover the signal offset and amplitude a J and a K term is needed, both of which are in the DC and δ term even with $M = 0$. This change results in Equation 3.11.

$$\begin{aligned} I_D(t) \;=\; & \frac{BJN}{2} + \frac{BKN}{2}\cos(\delta t) + AJN\cos(\omega t) \\ & + \frac{AKN}{2}\cos(\omega+\delta)t + \frac{AKN}{2}\cos(\omega-\delta)t \\ & + \frac{BJN}{2}\cos(2\omega t) + \frac{BKN}{4}\cos(2\omega+\delta)t + \frac{BKN}{4}\cos(2\omega-\delta)t \end{aligned} \tag{3.11}$$

Likewise, setting $N = 1$ simplifies things as well. To recover the original signal $I_S(t)$, $I_D(t)$ must be low pass filtered so that the components around ω and 2ω are suppressed. The bandwidth of a filter is expressed as the point at which an input signal amplitude is halved, i.e. the -3 dB point. Setting the low pass filter bandwidth to $\omega-\delta$ therefore does not provide sufficient suppression of the harmonics so it must be set lower. Setting the filter bandwidth such that the attenuation at

$\omega - \delta$ is -20 dB will result in the harmonics being attenuated to at most 1% of their level and so significantly reduce the amount of harmonic distortion in the recovered signal. In the ideal case, the recovered signal, $I_R(t)$, is simply a scaled version of $I_S(t)$ with the scale factor defined by the carrier and LO amplitudes, B and N respectively, both of which are known and fixed.

$$I_R(t) = \frac{BN}{2}\left(J + K\cos(\delta t)\right) \tag{3.12}$$

3.4.3 Quadrature Demodulation

Section 3.4.2 assumes that the carrier and LO are in phase with each other. This can make the technique difficult to use except in controlled situations where the two signals can be guaranteed to be in phase. In radio transmissions, for instance, the phase of the carrier is not known at the receiver but correct detection is still possible. As the phase between the carrier and the LO changes, the amplitude of the resultant signal $I_D(t)$ also changes in a straightforward sinusoidal response, with zero output when the difference is at 90° or 270°. As the amplitude contains the useful information, a phase offset will result in distortion of the signal or, in the worst case, no signal at all.

In reality the carrier component can have a time varying phase offset to the LO. This is equivalent to the original carrier having the form shown in Equation 3.13. This is identical in form to Equation 3.2, meaning that a real AM system also experiences PM to a certain degree.

$$I_C(t) = A + B\cos(\omega t + \phi(t)) \tag{3.13}$$

A phase sensitive detector is required that can work regardless of the phase difference between the carrier and the LO. This can be achieved by mixing $I_M(t)$ with two

local oscillators I and Q both at the carrier frequency, where Q has a 90° phase shift compared to I (i.e. in-phase and quadrature). This arrangement is shown in Figure 3.3.

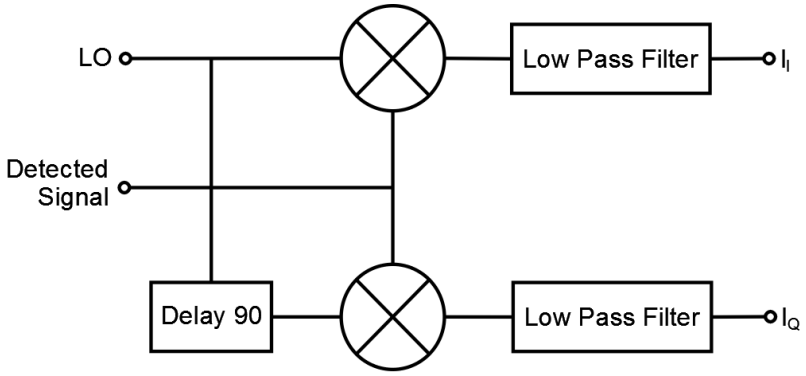


Figure 3.3: Quadrature Demodulation

The resulting mixed signals are low pass filtered exactly in the same way as "normal" AM, so that the high frequency components near ω and 2ω are removed. After the filtering, two signals $I_I(t)$ and $I_Q(t)$ are obtained which consist of the orignal signal each multiplied by a carrier where the two carriers are 90° out of phase with each other. The original signal can be reconstructed using Equation 3.14. The phase between the carrier and the LO can be calculated using Equation 3.15.

$$
\begin{aligned}
I_I(t) &= I_s(t) \cdot \sin(\omega t + \phi) \\
I_Q(t) &= I_S(t) \cdot \cos(\omega t + \phi)
\end{aligned}
$$

$$
\begin{aligned}
I_I^2(t) &= I_S^2(t) \cdot \sin^2(\omega t + \phi) \\
I_Q^2(t) &= I_S^2(t) \cdot \cos^2(\omega t + \phi) \\
I_I^2(t) + I_Q^2(t) &= I_S^2(t) \cdot \sin^2(\omega t + \phi) + I_S^2(t) \cdot \cos^2(\omega t + \phi) \\
&= I_S^2(t) \cdot \left(\sin^2(\omega t + \phi) + \cos^2(\omega t + \phi)\right) \\
&= I_S^2(t) \\
\therefore I_S(t) &= \sqrt{I_I^2(t) + I_Q^2(t)}
\end{aligned} \tag{3.14}
$$

$$
\begin{aligned}
\frac{I_Q(t)}{I_I(t)} &= \frac{I_S(t) \cdot \cos(\omega t + \phi)}{I_S(t) \cdot \sin(\omega t + \phi)} \\
&= \frac{\cos(\omega t + \phi)}{\sin(\omega t + \phi)} \\
&= \tan(\omega t + \phi) \\
\therefore \phi &= \tan^{-1}\left(\frac{I_Q(t)}{I_I(t)}\right)
\end{aligned} \tag{3.15}
$$

3.5 Demodulation in Practice

The previous sections considered demodulation in purely mathematical form and as such resemble a system made entirely of analogue components. In practice the signal is usually sampled and digitised so that it can be displayed as an image on a computer or other display device. The signal being detected, $I_D(t)$, is continuous in both time and value. Sampling refers to changing the signal so that it is no longer continuous in the time domain but it is only defined at discrete time points. Digitisation on the other hand makes the value of the signal discrete. Sampling and digitisation can exist independently of each other - the output of a switched capacitor circuit is an example of a discrete time and continuous value signal for

example. From this point on, any mention of a digitised signal refers to a signal that is discrete in both time and value, that is to say a signal that is suitable for processing with synchronous digital circuitry.

The digitisation of the signal can happen at many different points in the design and the choice of where it takes place has a fundamental affect on the way the design is constructed. The basic structure of a lock-in sensor is shown in Figure 3.4.

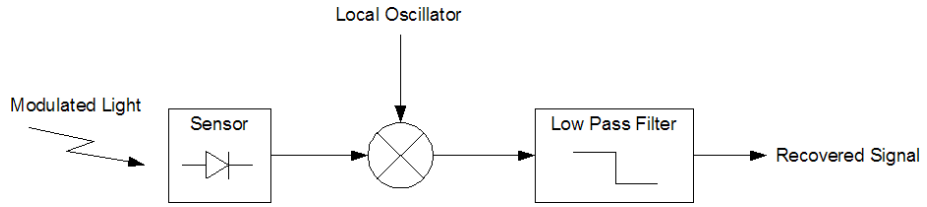


Figure 3.4: Simplified Lock-In Sensor

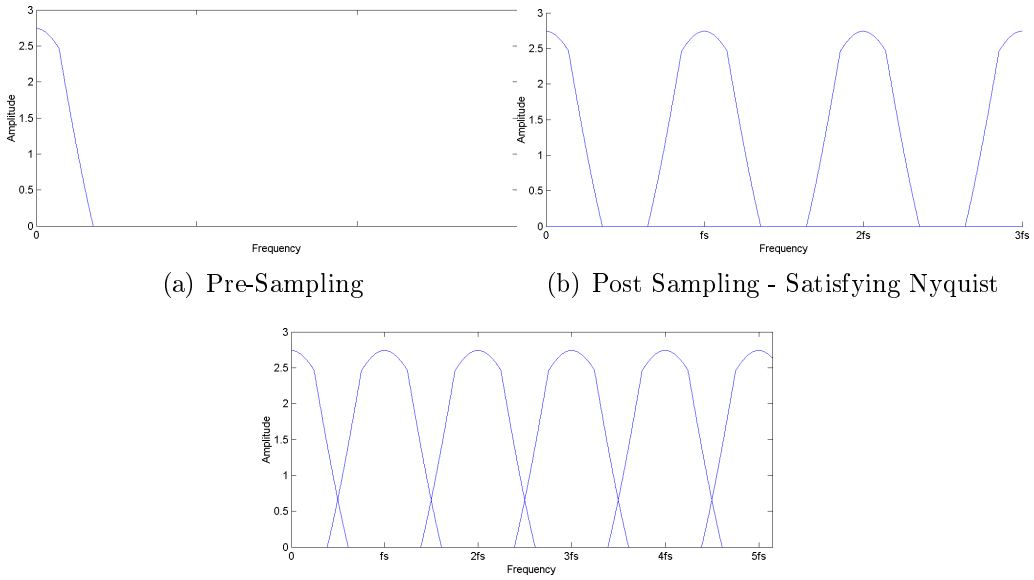
Based on that setup, it can be seen that there are three places where the signal can be digitised, namely after the sensor block, after the mixer or after the low pass filter.

3.5.1 Sampling

The Nyquist theorem states that to be able to perfectly reconstruct a signal after it has been sampled, the sample rate must be greater than twice the bandwidth of the signal. So for a signal with frequency components from 0 to ω, the minimum sampling frequency is 2ω. The maximum frequency that can be reconstructed after sampling at a given frequency is referred to as the Nyquist frequency. If a signal contains frequency components higher than the Nyquist frequency then those frequencies will be aliased into the sampled signal bandwidth. This means that they will appear in the signal bandwidth as false representations of the original high frequency signal.

When a signal is sampled at sampling frequency ω_s, the effect in the frequency domain is to produce an infinite set of copies of the signal spectrum, with the

copies centred around ω_s, $2\omega_s$, $3\omega_s$ and so on, with each copy consisting of an upper sideband and a lower sideband. This can be seen in Figure 3.5 (a) and (b) which also explains why aliasing occurs. If ω_s is too low for a given signal bandwidth, the duplicated signal spectra will overlap with one another as in Figure 3.5 (c). This overlap causes aliasing and prevents the signal from being properly reconstructed.



(a) Pre-Sampling

(b) Post Sampling - Satisfying Nyquist

(c) Post Sampling - Not Satisfying Nyquist

Figure 3.5: Sampled Frequency Spectra (Positive Frequencies Only)

To help prevent aliasing, the signal to be sampled must be truly band limited - any signal outside of the Nyquist frequency will be aliased into the signal band. This is usually achieved by applying an anti-aliasing filter to the signal immediately before sampling to remove any frequency components that lie outside of the desired bandwidth.

3.5.2 Digitisation

As already mentioned, the process of taking a continuous value, in our case voltage, and converting it into discrete values is called digitisation or quantisation. The result of digitisation is a number ranging from 0 to $2^n - 1$, where n is the number of bits in the analogue to digital converter (ADC), typically ranging from 8 to 24

bits. The number of bits and the input range of the ADC defines the smallest unit measurable; if the range of an 8-bit ADC is $0 - 3.3\,\mathrm{V}$, the smallest voltage change that can be measured is $3.3\,\mathrm{V}/256 = 12.9\,\mathrm{mV}$. Changes smaller than this cannot be detected, meaning that any given sample can have an error of $\pm\frac{1}{2}\,\mathrm{LSB}$, where an LSB is the "least significant bit" or distance between two ADC levels. In the example given, one LSB is $12.9\,\mathrm{mV}$.

3.5.3 Sampling After Filtering

Obeying the Nyquist sampling theory, the sample rate must be at least twice the bandwidth of the signal being sampled. This usually means a sample rate of $\omega_s = k\omega$, where ω is the bandwidth and $k \geq 2$.

Sampling the signal after the low pass filter in Figure 3.4 means that the entire demodulation process is analogue and assuming a signal bandwidth of δ means that the sampling frequency must be $\geq 2\delta$. In this case, the sampling frequency also defines the frame rate assuming that all pixels work in parallel with pixel based processing. In a column based design, the frame rate will be much lower because the time required for each row is determined by the settling time of the low pass filter and is likely to be fairly slow.

Any digital circuitry in this type of design will be devoted to controlling the chip and the ADC and sending the results to wherever they are needed rather than playing a direct role in the demodulation.

3.5.4 Sampling After Mixing

Sampling the signal after it has been partially demodulated (i.e. after the mixer), but not filtered as in Equation 3.10, gives a significantly different situation. The frequency components related to the carrier are still present with the highest at

$\omega_{max} = 2\omega + \delta$, meaning that the sampling frequency must be $\omega_s \geq k \cdot (2\omega + \delta)$, where $k \geq 2$. This is substantially higher than in the case described in Section 3.5.3 and puts a much greater demand on the ADC. When sampling, any noise components above $0.5\omega_s$ will be aliased into the sampled bandwidth and this is normally countered by including an anti-aliasing filter directly before the sampling occurs. This means that two filters are needed here; one relatively relaxed filter before the sampling to remove the noise above $0.5\omega_s$ and one after the sampling to remove the unwanted high frequencies around 2ω and ω. The sampling frequency ω_s is the required rate per pixel, so although using a carrier frequency of $\omega = 10\,\text{kHz}$ and a signal frequency of $\delta = 100\,\text{Hz}$ gives a minimum sampling frequency of $\omega_s = 2 \cdot (2 * 10\,\text{kHz} + 100\,\text{Hz}) = 40.2\,\text{kHz}$, the rate for the whole camera must be much higher. With a 64 x 64 camera, the overall sample rate of the ADC must be at least $\omega_s = 40.2\,\text{kHz} * 4096 = 164.7\,\text{MHz}$. This is a high rate for an ADC if a fine resolution is required. This assumes a single ADC per camera so increasing the number of ADCs would allow a corresponding decrease in the sampling frequency required although at an increased cost due to the greater silicon area required by the additional ADCs.

The low pass filter must be implemented in digital logic. A filter that is very simple to implement (and hence requires relative little space) is the moving average filter. In this filter, the previous n samples of the signal are averaged as in Equation 3.16.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3.16}$$

Where n is the size of the filter. Each time a new sample is given to the filter, the oldest sample is replaced with the new, so the original sum of samples from x_0 to x_n becomes x_1 to x_{n+1}. This technique requires only a few operations, the subtraction of the first sample and then the addition of the new sample. The filter must receive n samples before it produces a useful output, meaning that the output has a delay. By keeping n a power of 2, the division can be achieved through a simple logical shift

rather than needing a dedicated division block. The storage requirements are fairly straightforward as well, with just n samples needing storing. A disadvantage of the moving average filter is its frequency response, which is not particularly sharp. The frequency response of the moving average filter, $H(f)$, is shown in Equation 3.17, where f is the frequency and is in the range $0 \leq f \leq 0.5$, with the response at $f = 0$ being 1.

$$H(f) = \frac{\sin(n\pi f)}{n\sin(\pi f)} \tag{3.17}$$

As the input to the moving average filter is sampled, the maximum frequency that can be represented is half the sampling frequency, hence the limitation on f. Figure 3.6 shows the frequency response of a moving average filter for different number of points, where the frequency on the x-axis is normalised to the sampling frequency f_s and is mirrored for values above $f = 0.5$.

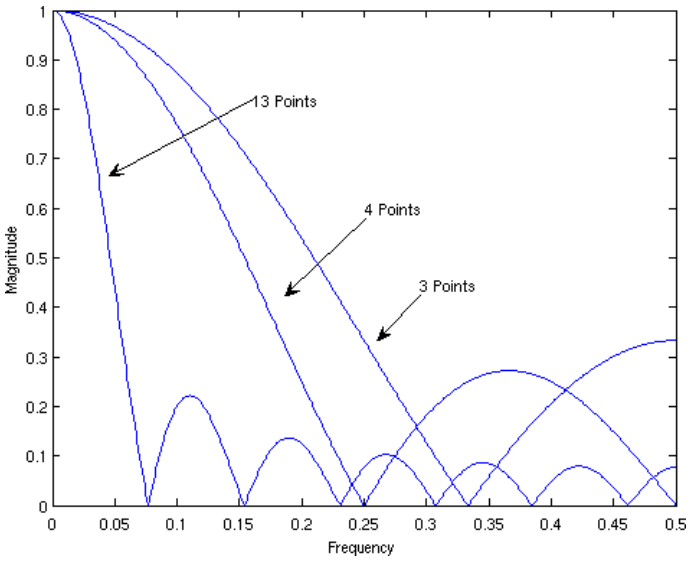


Figure 3.6: Moving Average Filter Response

Although the frequency response of the moving average filter is not optimal for low pass filtering, it does have the advantage of having a lot of zeros that can be placed with relative ease. As the biggest frequency components that need to be filtered are caused by the mixing process rather than general noise sources, it is possible to choose the number of points in the filter to ensure that the response is zero at

these frequencies. There are $n-1$ nulls in the range $0 \leq f \leq 1$ and they are equally spaced across the whole range. So assuming that the sampling frequency is chosen to be $\omega_s = 4\omega$, then the 2ω component will appear at $f = 0.5$ in the moving average filter response. To make a null at this point, there must be an odd number of nulls in total, meaning an even number of points in the filter. Equally, to filter out the components at ω, a null must be placed at $f = 0.25$. To do this, the number of points in the filter must be a multiple of 4. By choosing this as a power of 2 as well being a multiple of 4, the division part of the calculation is made much simpler as described earlier. The four point response in Figure 3.6 shows this scenario well.

In addition to filtering the unwanted components, the filter must also pass the frequencies of interest. This is done by choosing the location of the first null, which is located at ω_s/n. The cut off frequency of the filter should be chosen so it is at least 2δ, so:

$$\begin{aligned} \frac{\omega_s}{n} &= 2\delta \\ \therefore\ n &= \frac{\omega_s}{2\delta} \\ &= \frac{2k\omega}{2\delta} \\ &= \frac{k\omega}{\delta} \end{aligned} \tag{3.18}$$

After filtering, the signal still has a frequency of ω_s, which is much greater than the bandwidth of the data it contains. This means that it can be subsampled to reduce the data rate whilst still maintaining all of the information. To do this, every g^{th} sample can be taken such that:

$$\begin{aligned} \frac{\omega_s}{g} &> 2\delta \\ g &< \frac{\omega_s}{2\delta} \end{aligned} \tag{3.19}$$

This leads to the consideration of a different filter design, namely the Accumulate and Dump filter. This filter is very similar to the Moving Average filter because

it also accumulates a certain number of samples and produces an average value as its output (and so has the same frequency response), but after a set of values have been used they are dumped and the process starts again. This is advantageous in this case because it can carry out the decimation described above with the added advantage that the Accumulate and Dump is simpler to implement as it requires fewer storage elements.

3.5.5 Sampling Before Mixing

Sampling the signal before it has been demodulated at all is the final possibility. Figure 3.4 represents the circuitry before this point as a single block (the "Sensor"), but in reality it can contain any number of circuits. Typically the current to voltage conversion will take place here, as well as any amplification required. These operations do not affect the sample rate at all. The highest frequency present, assuming that there is no high frequency background illumination, is $\omega_{max} = \omega + \delta$. This means that at first glance the sampling frequency must be $\omega_s \geq 2 \cdot (\omega + \delta)$, however after the signal has been sampled, the digitised signal will be mixed to carry out demodulation. This mixing introduces $2\omega \pm \delta$ components in just the same way as the analogue mixing. If the sampling frequency is insufficient to cope with this bandwidth then the high frequency components will alias into the signal band. This means that the sampling frequency for this case must also be $\omega_s \geq 2 \cdot (2\omega + \delta)$.

So as far as sampling goes, digitising immediately after detection is much the same as digitising after the mixing has taken place, however it does place a much higher demand on the digital circuitry. As well as having to filter the signal, it must also be mixed which means multiplication, a reasonably involved operation. The problem is compounded by two factors, namely the speed and resolution of the digitisation. The speed at which the multiplication must be carried out is proportional to the speed of the digitisation whilst an increase in the resolution of the digitisation by one bit means that the multiplication will either take longer to perform or else require

a greater silicon area.

The signal used for demodulation as described in Section 3.4.2 is a sine wave and this needs representing in a digital form, which can either be as a decimal in fixed or floating point form, or an integer. Both fixed and floating point calculations are computationally demanding, with floating point being more flexible but also more complicated. As the sampled value is likely to be an integer in common binary twos complement format, it is reasonable to keep both signals as integers to simplify the calculations. A sine wave that ranges from -1 to 1 must be scaled to be represented as an meaningful integer. For example, using a 16-bit signed integer (with range -32,768 to 32,767) to hold the full range means that all of the sinusoid values are multiplied by 32,767 to calculate the scaled integer values. Any size integer can be used, with larger ones offering better representations of the true value at the cost of requiring more time or area for the multiplication. A further consideration is that when multiplying two integers together, the size of the output is potentially the sum of the number of bits of the input, so for example 32,767 (15 bits) multiplied by 255 (8 bits) is 8,355,585 (23 bits).

Using integer multiplication is the least demanding type of multiplication and likewise the least demanding for addition and subtraction, both of which become an issue in the filter, however whatever type is used the numbers must be stored in a lookup table so that they are immediately available. This does not require as much storage as it might first appear because of the symmetrical nature of the sine wave. Instead of storing data for all 360°, only the first 90° need be stored and this can be used to calculate the remaining values in a very simple manner. Hence using look up table with with one 16-bit integer value per 1°, for a total of 90 values stored, would require a total of 1,440 bits of storage.

An alternative is to use what is known as square wave mixing. The advantage here is that no storage is required for the local oscillator as only two values are required, namely 1 and -1. However, using square wave mixing does bring about its own

problems since a perfect square wave is made up of an infinite number of sine waves that are all odd harmonics of the fundamental frequency and contribute smaller and smaller components with increasing frequency, as shown in Equation 3.20.

$$I_{LO}(t) = \cos(\omega t) + \frac{1}{3}\cos(3\omega t) + \frac{1}{5}\cos(5\omega t) + \cdots + \frac{1}{n}\cos(n\omega t) + \cdots \tag{3.20}$$

When mixing with a square wave, all of these frequency components generate harmonics. Considering Equation 3.9 with the first three terms of Equation 3.20 gives the result in Equation 3.21:

$$\begin{aligned}
I_D(t) &= I_M(t) \cdot I_{LO}(t) \\
&= [AJ + AK\cos(\delta t) + BJ\cos(\omega t) \\
&\quad + \frac{BK}{2}\cos(\omega \pm \delta)t] \cdot [\cos(\omega t) + \frac{1}{3}\cos(3\omega t) + \frac{1}{5}\cos(5\omega t)] \\
&= \frac{1}{2}BJ + \frac{1}{2}BK\cos(\delta t) \\
&\quad + AJ\cos(\omega t) + \frac{1}{2}AK\cos(\omega + \delta)t + \frac{1}{3}BK\cos(2\omega - \delta)t \\
&\quad + \frac{2}{3}BJ\cos(2\omega t) + \frac{1}{3}BK\cos(2\omega + \delta)t + \frac{1}{3}BK\cos(2\omega - \delta)t \\
&\quad + \frac{1}{3}AJ\cos(3\omega t) + \frac{1}{6}AK\cos(3\omega + \delta)t + \frac{1}{6}AK\cos(3\omega - \delta)t \\
&\quad + \frac{4}{15}BJ\cos(4\omega t) + \frac{2}{15}BK\cos(4\omega + \delta)t + \frac{2}{15}BK\cos(4\omega - \delta)t \\
&\quad + \frac{1}{5}AJ\cos(5\omega t) + \frac{1}{10}AK\cos(5\omega + \delta)t + \frac{1}{10}AK\cos(5\omega - \delta)t \\
&\quad + \frac{1}{10}BJ\cos(6\omega t) + \frac{1}{20}BK\cos(6\omega + \delta)t + \frac{1}{20}BK\cos(6\omega - \delta)t
\end{aligned} \tag{3.21}$$

The frequency spectrum for Equation 3.21 is shown in Figure 3.7. As can be seen, the number of terms has increased greatly compared to demodulating with a single sine wave (see Equation 3.10 and Figure 3.2).

In itself this is not a large problem since the filtering in the next stage will suppress the high frequency components, especially since the higher harmonics have only a small contribution, however this does presume a much higher sampling frequency. As an ideal square wave contains an infinite number of harmonics, the result of

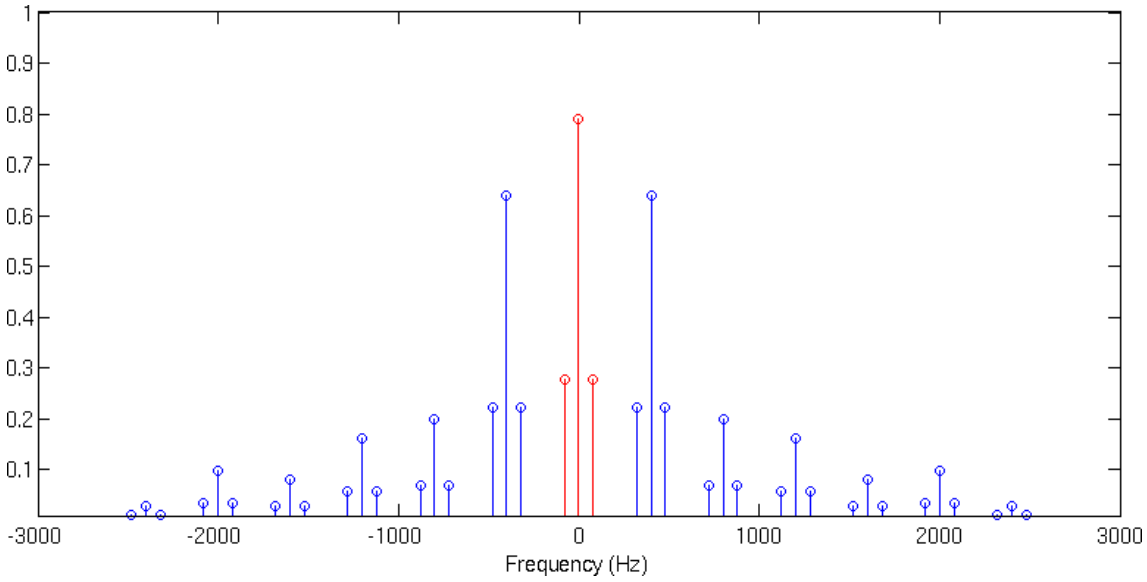


Figure 3.7: Square Wave Demodulation Frequency Spectrum

mixing with it will also contain an infinite number of harmonics, which is an important consideration when sampling because sampling to strictly satisfy the Nyquist criterion is not possible.

If sampling is carried out after square wave mixing then an anti-aliasing filter can be applied to the output before sampling to remove the high frequency components (although this may make a further digital low-pass filter redundant). With sampling done before the mixing, the sample rate should be high enough to correctly sample the entire bandwidth which is obviously impossible. In practice, the mixing components caused by the 7^{th} and higher harmonics are often negligible, although sampling them to satisfy the Nyquist criterion still reflects a considerable increase in the sampling overhead.

3.5.6 Sampling to Simplify Mixing

As noted, the scheme described in Section 3.5.5 suffers because of the computational requirements of the mixing. However, choosing the sampling frequency carefully can mean the mixing becomes much simpler and can be done using integer maths without the need for a lookup table. As the point of this technique is to simplify the mixing

process, it is only possible to apply it when the digitisation occurs immediately after detection, before the mixing has taken place.

Sampling at $\omega_s = 2\omega$ (referred to here as two phase mixing) means sampling in the peaks and troughs of the carrier, assuming the carrier and sampling clock are in phase. The resulting signal can be mixed by multiplying with a square wave; in the simplest case this means multiplying alternate samples by +1 and -1. Multiplying by +1 is the same as doing nothing and multiplying by -1 is just changing the sign of the sample which is a very simple operation. To obtain the peak to peak amplitude of the carrier, the alternate samples are added together; as the square wave mixing has taken place, this is directly equivalent to subtracting the second sample from the first.

Section 3.5.5 showed that square wave mixing produced extra harmonics that mean the sampling frequency must be increased to ensure that aliasing does not occur, yet this simplified mixing uses a sampling frequency that is too low to sample these extra harmonics. In fact, although it appears as though this is square wave mixing from the description, it is actually indistinguishable from sine wave mixing, since the samples are being taken at the peaks of the local oscillator and at that point it is impossible to see any difference between a sine wave and a square wave, as can be seen in Figure 3.8 where the sample points are shown as blue circles. The key difference compared to the square wave mixing in Section 3.5.5 is that here only those points identical between the two types of wave are used whereas the other method uses points across the entire wave form.

Taking the simple case where the signal from the experiment is not changing, then $I_S(t) = 1$ and the carrier is unmodulated. The signal received at the sensor is $I_C(t) = A + B\cos(\omega t)$ as in Equation 3.5. Applying square wave mixing to this case and taking two samples, S_1 and S_2, that occur at the peak and trough of $I_C(t)$ gives:

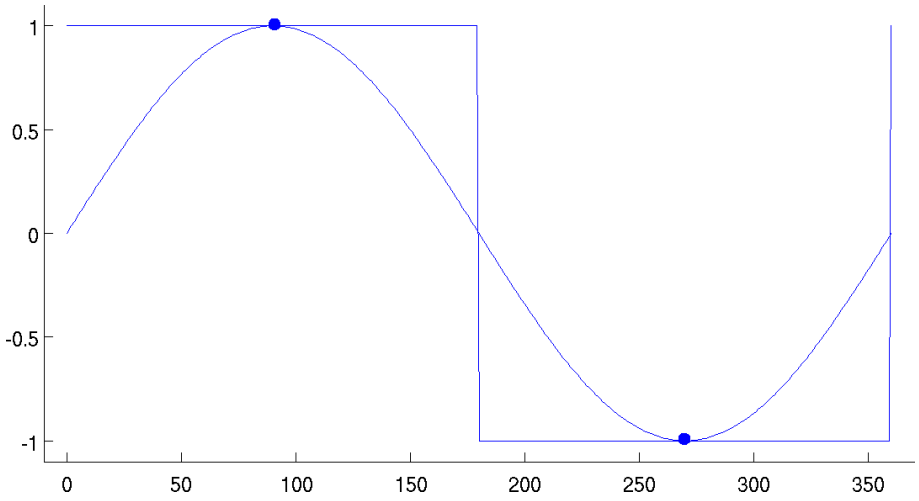


Figure 3.8: Local Oscillator Square/Sine Equivalence

$$\begin{aligned} S_1 &= A + B \\ S_2 &= A - B \\ S &= (1 * S_1) + (-1 * S_2) \\ &= (A + B) + (-A + B) \\ &= 2B \end{aligned} \tag{3.22}$$

Hence it is possible to measure the amplitude of the carrier. This result is still applicable when $I_S(t)$ varies with time because we have already stated that $\omega \gg \delta$ and so between two samples $I_S(t)$ will be close to being stationary.

As stated, this result is heavily dependant on the phase between the carrier and the sampling clock being zero, exactly the same requirements as for analogue mixing as described in Section 3.4.2. This means that there must be some connection between the device producing the carrier and the sensor or else other techniques, such as a phase locked loop, must be employed on the sensor to ensure the phase offset remains both low and constant.

It is also possible to calculate the DC offset (in other words the image as taken by an ordinary camera) by using Equation 3.23.

$$\begin{aligned} DC &= S_1 + S_2 \\ &= (A + B) + (A - B) \\ &= 2A \end{aligned} \tag{3.23}$$

Invariance to the phase between the carrier and the local oscillator can be achieved by using quadrature mixing as described in Section 3.4.3. This is done by using two square waves that are 90° out of phase with each other. To be able to mix the detected signal with these square waves, the sample rate must be increased since sampling at 2ω only gives a sample every 180°. To make mixing with two square waves possible, the sample rate must be increased to $\omega_s = 4\omega$ (four phase mixing). This satisfies Nyquist for the detected signal; the highest frequency component of the detected signal is $\omega + \delta$ and $\delta \ll \omega$.

Given four samples per carrier wave, this results in:

$$\begin{aligned} S_1 &= A + B \cdot \sin(\phi) \\ S_2 &= A + B \cdot \sin(\phi + 90°) \\ S_3 &= A + B \cdot \sin(\phi + 180°) \\ S_4 &= A + B \cdot \sin(\phi + 270°) \end{aligned}$$

Where ϕ is the phase between the carrier and the sampling clock (which is effectively the local oscillator). The four equations above can be simplified:

$$S_1 = A + B \cdot \sin(\phi) \tag{3.24}$$

$$S_2 = A + B \cdot \cos(\phi) \tag{3.25}$$

$$S_3 = A - B \cdot \sin(\phi) \tag{3.26}$$

$$S_4 = A - B \cdot \cos(\phi) \tag{3.27}$$

With four results it is straightforward to calculate the three unknowns as follows,

as described by Lange [Lange and Seitz, 2001]:

$$\begin{aligned} S_1 + S_2 + S_3 + S_4 &= A + B \cdot \sin(\phi) + A + B \cdot \cos(\phi) \\ &\quad + A - B \cdot \sin(\phi) + A - B \cdot \cos(\phi) \qquad (3.28) \\ &= 4A \\ \therefore A &= \frac{S_1 + S_2 + S_3 + S_4}{4} \qquad (3.29) \end{aligned}$$

$$\begin{aligned} S_I = S_1 - S_3 &= A + B \cdot \sin(\phi) - A + B \cdot \sin(\phi) \\ &= 2B \cdot \sin(\phi) \qquad (3.30) \\ S_Q = S_2 - S_4 &= A + B \cdot \cos(\phi) - A + B \cdot \cos(\phi) \\ &= 2B \cdot \cos(\phi) \qquad (3.31) \\ S_I^2 + S_Q^2 &= 4B^2 \cdot \left(\sin^2(\phi) + \cos^2(\phi)\right) \\ &= 4B^2 \\ \therefore B &= \sqrt{\frac{S_I^2 + S_Q^2}{4}} \qquad (3.32) \end{aligned}$$

$$\begin{aligned} \frac{S_I}{S_Q} &= \frac{2B \cdot \sin(\phi)}{2B \cdot \cos(\phi)} \\ &= \tan(\phi) \\ \therefore \phi &= \tan^{-1}\left(\frac{S_I}{S_Q}\right) \qquad (3.33) \end{aligned}$$

Of course, with only three unknown values it is also possible to use just three samples per carrier wave, with the samples spread out in 120° intervals. This is more problematic than using four samples because the mixing is slightly more complicated since the values to be multiplied by are 1, -0.5 and -0.5 at the 0°, 120° and 240° samples respectively. The reconstruction of the unknown parameters is more complicated as well, as shown in Equation 3.34 for the phase:

$$\phi = \tan^{-1}\left(-\sqrt{3}\frac{(S_1 - S_2) + (S_1 - S_3)}{S_2 - S_3}\right) \qquad (3.34)$$

In both of these cases, the final signal rate is at ω and still requires filtering, which can be done with the moving average filter as described in Section 3.5.4.

3.6 Modulated Light Camera Parameters

Given the above, it becomes clear that it is possible to describe the operation of any modulated light cameras in terms of a limited number of parameters that are key to its operation.

Modulation Frequency The frequency at which the light is modulated is often not that crucial, where the desire is simply to benefit from the lock-in nature of the system and move the signal away from low-frequency noise (such as flicker noise). Exceptions where the frequency is important are range measurements (since the frequency determines the wavelength and hence the resolution of the range detection) and where the experiment itself is the cause of the modulation. Consequently the modulation frequency can be chosen to help in simplifying camera design.

Sample Frequency The architecture of the pixel and lock-in circuitry has an important role in determining the sample frequency, in parallel with the modulation frequency. The point at which sampling is carried out has a significant effect as discussed in sections 3.5.3 to 3.5.5.

Frame Rate The frame rate is strictly determined by the rate of change of the experiment, i.e. the frequency δ, however in practice, for example in a microscope system, a faster frame rate may be desirable to aid visualisation (e.g. focusing).

The Equations 3.18 and 3.19 can be used to set the size of the filter, n, required and the rate at which the output from the filter can be subsampled at, g, which is also the minimum frame rate.

3.7 Comparison of Existing Work

This section relates the work described in Chapter 2 with regards to the theory presented in the previous section.

With completely analogue systems, both Bourquin and Dmochowski provide straightforward examples [Bourquin et al., 2001, Dmochowski et al., 2004]. As the output of the devices is analogue and has already been filtered, the sample frequency (for a given pixel rather than for the array) and the frame rate are identical. Bourquin quotes a maximum frame rate of 890 Hz based on the maximum pixel readout rate but does not mention the filter bandwidth, so it is assumed that it around the same frequency. Dmochowski uses a low-pass filter of bandwidth 150 Hz which gives the maximum frame rate.

The different designs presented by Povel, Spirig and Lange all use the simplified mixing techniques, meaning that the sample rate is set at either twice or four times the modulation frequency [Povel et al., 1990, Spirig et al., 1995, Lange and Seitz, 2001]. Filtering can either take place by averaging multiple frames on chip if possible, or by digitising each frame and performing the filtering with a digital filter. Povel performs filtering on chip with a bandwidth of between 1 Hz and 50 Hz, Spirig uses the same with a bandwidth of 695 Hz and Lange with a bandwidth of between 10 Hz and 10 kHz, but they consider the bandwidth (which is in this case the combined integration time of each of the samples) with regard to the signal quality rather than the bandwidth of the experiment, an approach which is quite common.

The design by Pitter [Pitter et al., 2004] is an interesting case as it is not strictly speaking a lock-in. Samples are taken only once per modulation cycle, so extreme care has to be taken to ensure that the samples occur at the peak of the cycle. Likewise, the environment in which the experiment occurs must be more tightly controlled because as well as being sensitive to the odd harmonics of the modulation frequency (as is often the case) it will also show a strong sensitivity to the even

harmonics as well. No mention is made of the bandwidth of the experiment or the amount of filtering undertaken.

3.8 Integrating Lock-In

This section proposes a new way of using the integrating APS type pixel as well as a different method of recovering the amplitude and phase of a modulated signal. The traditional APS, as described in Section 1.5.2, charges the capacitance associated with the photodiode in the pixel up to a known voltage and then allows the charge to discharge due to the photocurrent (and leakage current). After a fixed amount of time, the voltage across the photodiode is measured. This is a simple application of the current-voltage operation of a capacitor, as in $i = C\frac{dV}{dt}$, where the capacitance, C and the integration time, dt, are both known, and the change in voltage, dV, is measured to allow the photocurrent, i, to be calculated and hence the intensity of light determined. Another way of using the same behaviour is to measure the amount of time taken for the voltage across the photodiode to discharge to a certain value. This can be done quite simply by feeding the output of the photodiode to a voltage comparator and using a counter to record the time taken. This is discussed in more detail in Chapter 4. Work done by Stoppa [Stoppa et al., 2002] uses a similar technique where voltage comparators are used to sample two analogue voltage values onto separate capacitors to provide a higher dynamic range.

The different technique for calculating the amplitude and the phase of the modulated signal comes from observing that although four samples are taken per modulated cycle, only two of them are used for calculating each of S_I and S_Q (see Equations 3.30 and 3.31). As many of the noise sources discussed in Section 1.6 are from random processes, it makes sense that using more of the results will help to reduce the effect of these noise sources. To that end, using all four results in each channel was explored. The Equations 3.24 to 3.27 for S_1 to S_4 are still used as the samples themselves are still the same. S_I and S_Q now become:

$$
\begin{aligned}
S_I = S_1 + S_2 - S_3 - S_4 &= A + B \cdot \sin(\phi) + A + B \cdot \cos(\phi) \\
&\quad -A + B \cdot \sin(\phi) - A + B \cdot \cos(\phi) \\
&= 2B \cdot \sin(\phi) + 2B \cdot \cos(\phi) \qquad (3.35) \\
S_Q = S_1 - S_2 - S_3 + S_4 &= A + B \cdot \sin(\phi) - A - B \cdot \cos(\phi) \\
&\quad -A + B \cdot \sin(\phi) + A - B \cdot \cos(\phi) \\
&= 2B \cdot \sin(\phi) - 2B \cdot \cos(\phi) \qquad (3.36)
\end{aligned}
$$

Equations 3.35 and 3.36 can be used to calculate the amplitude of the modulated signal in a similar manner to previously employed:

$$
\begin{aligned}
S_I^2 &= 4B^2 \cdot \sin^2(\phi) + 4B^2 \cdot \cos^2(\phi) - 8B^2 \cdot \sin(\phi)\cos(\phi) \\
&= 4B^2 \left(1 - 2 \cdot \sin(\phi)\cos(\phi)\right) \qquad (3.37) \\
S_Q^2 &= 4B^2 \cdot \sin^2(\phi) + 4B^2 \cdot \cos^2(\phi) + 8B^2 \cdot \sin(\phi)\cos(\phi) \\
&= 4B^2 \left(1 + 2 \cdot \sin(\phi)\cos(\phi)\right) \qquad (3.38) \\
S_I^2 + S_Q^2 &= 4B^2 \left(1 - 2 \cdot \sin(\phi)\cos(\phi)\right) + 4B^2 \left(1 + 2 \cdot \sin(\phi)\cos(\phi)\right) \\
&= 8B^2 \qquad (3.39) \\
\therefore B &= \sqrt{\frac{S_I^2 + S_Q^2}{8}} \qquad (3.40)
\end{aligned}
$$

Likewise for the phase offset:

$$
\begin{aligned}
\frac{S_Q}{S_I} &= \frac{2B \cdot \sin(\phi) - 2B \cdot \cos(\phi)}{2B \cdot \sin(\phi) + 2B \cdot \cos(\phi)} \\
&= \frac{\sin(\phi) - \cos(\phi)}{\sin(\phi) + \cos(\phi)} \\
&= \frac{\sqrt{2}\sin(\phi - 45^\circ)}{\sqrt{2}\cos(\phi - 45^\circ)} \\
&= \tan(\phi - 45^\circ) \\
\therefore \phi &= \tan^{-1}\left(\frac{S_Q}{S_I}\right) + 45^\circ \qquad (3.41)
\end{aligned}
$$

These equations have a very similar form to Equations 3.32 and 3.33, differing only

in the fixed divisor for the amplitude and the 45° offset for the phase. The main difference between the two sets of equations is the differing values of S_I and S_Q. The next section demonstrates the differences between the two phase technique and both of the four phase techniques by simulating them with a Matlab model of an integrating pixel.

3.9 Comparison of Techniques

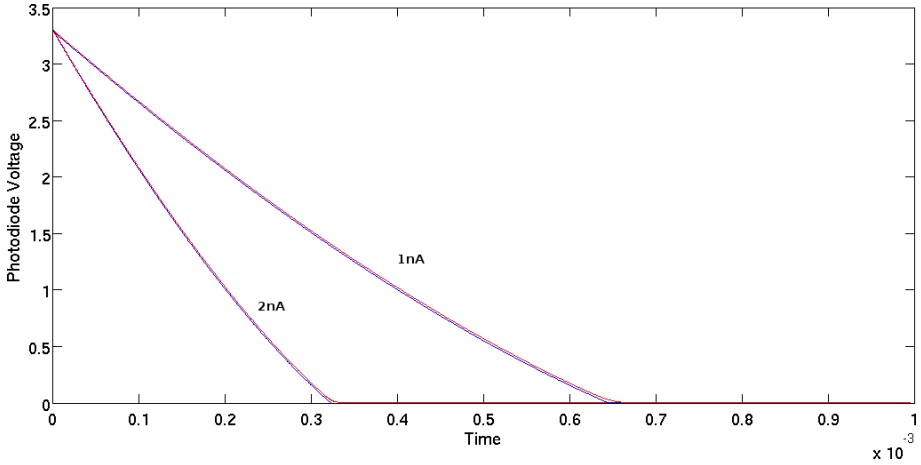
It is not immediately obvious which of the time or voltage pixel designs or calculation techniques is better, so this section compares them. The two designs will be referred to as the "voltage pixel" for the traditional APS and the "time pixel" for the other design. A third type of pixel is also considered, where the values obtained from the time pixel are used as a reciprocal. This is a new type of pixel and is referred to as the "inverse time pixel". The calculation technique described in Section 3.5.6 will be referred to as Lange's method and the technique described in Section 3.8 will be referred to as the Quadrature method.

3.9.1 Matlab Model

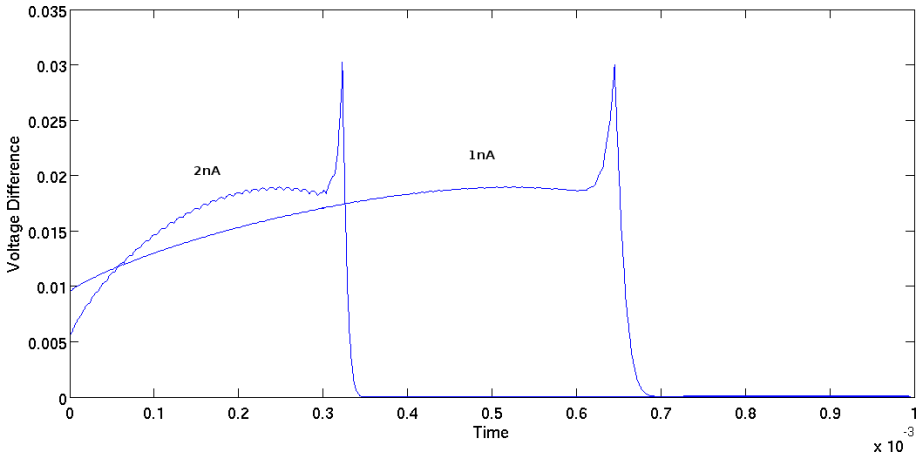
To help facilitate the comparison between the different techniques, a model of the n-well to p-substrate diode in the C35 process was developed in Matlab, using the parameters and equations as provided by AMS[AMS, 2004]. This was used to create a model of the voltage and the time pixel as shown in Appendix A. The advantage of using a Matlab model rather than the more usual Cadence schematic simulation is that the post processing for the calculation of amplitude and phase becomes much more straightforward, especially when there are many simulations to consider.

Figure 3.9 shows a comparison of the Matlab model and the Cadence SPICE model provided by AMS for a $50\,\mathrm{x}\,50\,\mu$m diode. The simulations shown are for a photodiode

used as an integrating pixel that has just been reset and is discharging due to photocurrents of both 1 nA and 2 nA. Figure 3.9 (a) shows the actual response in time, where the red lines show the discharge from the Cadence model and the blue lines show the discharge from the Matlab model.



(a) Photodiode discharge at 1 nA and 2 nA photocurrent (Red curves: Cadence, blue curves: Matlab)



(b) Difference Between the Matlab and Spice Models

Figure 3.9: Matlab / Cadence Diode Model Comparison

The difference in the curves between the two models is shown in Figure 3.9 (b), which shows that the difference between the two models is typically beneath 20 mV and only increases beyond this point when the photodiode voltage approaches 0 V. This difference is due to the Matlab model only considering the capacitance and leakage current of the diode whereas the Cadence model will also consider other factors.

3.9.2 DC Response

The first observation to make is that the measurements made by the three types of pixels (voltage, time and inverse time) have different dependency on the light intensity, as can be seen in Figure 3.10 where the response of the pixels to a range of DC photocurrent (i.e. light) intensities is shown. As the light intensity increases, the output of the voltage pixel decreases. In an ideal case this would be a linear response, however since the capacitance of a photodiode varies with the voltage across it, the discharge rate is non-linear. The time pixel produces a substantially different output because it is inversely proportional to the light intensity. The blue lines show the case where the photodiode capacitance remains constant with voltage and the red lines are the realistic case.

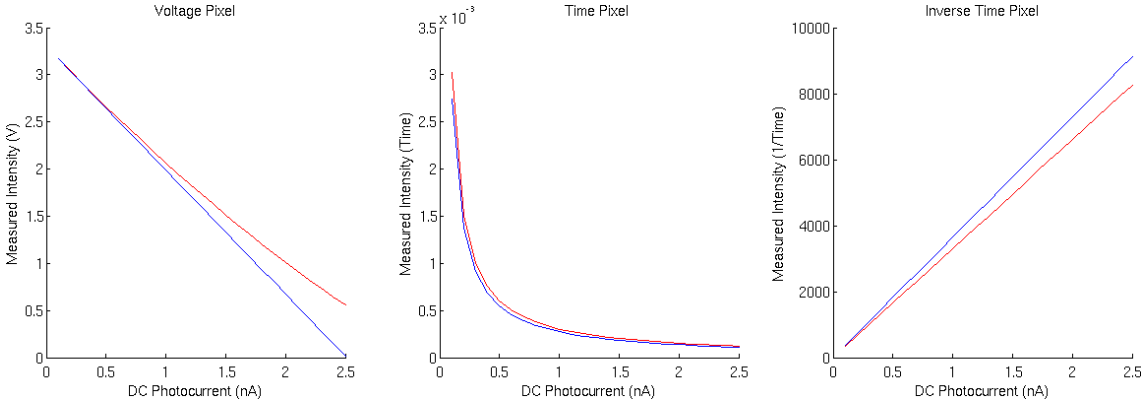


Figure 3.10: DC Pixel Response
(Blue: ideal case, Red: realistic case)

3.9.3 Two Phase Pixel DC Response

With two phase mixing it is possible to measure both the amplitude and DC offset of a locked in signal. The results in Figure 3.11 show the DC response for the three different pixel types when locked into a modulated signal of frequency 1.25 kHz. The three visible lines in each plot are for different carrier amplitudes (i.e. B in $I_C(t)$) of 0.1 nA, 0.5 nA and 1.0 nA. In this and subsequent simulations the photodiode is 50 x 50 μm with the voltage pixel measured after 100 μs and the time pixel measured

at a voltage of 2.5 V. These two values occur at roughly the same time when the diode has 1 nA of DC photocurrent flowing.

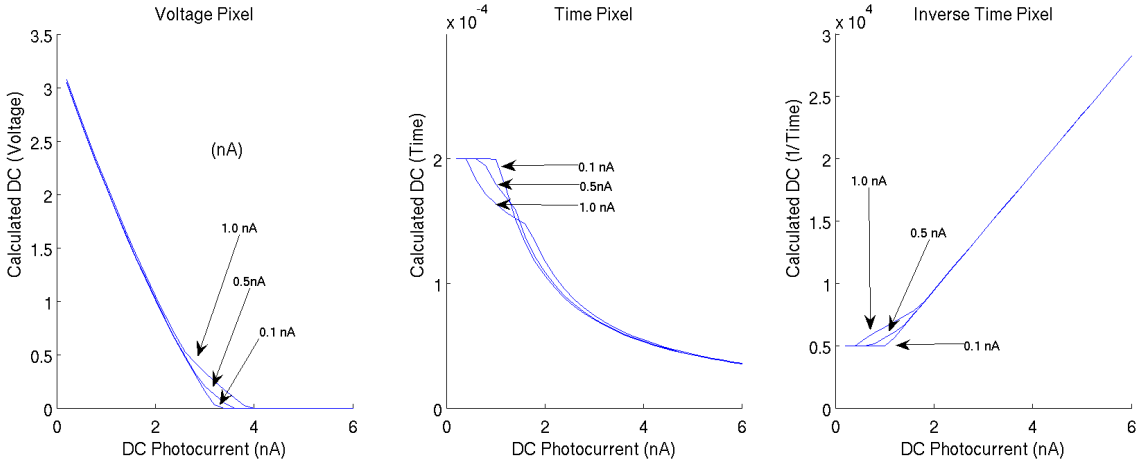


Figure 3.11: Two Phase Pixel DC Response

The results are much as expected, with the voltage and inverse time pixels showing a fairly linear response and the time pixel showing a 1/x response. All three pixel results show a certain dependence on the amplitude of the signal being measured, with the inverse time pixel showing the greatest dependence. Additionally, distortion can be observed at high DC levels for the voltage pixel and low DC levels for the time and inverse time pixels. It is useful to define the term "modulation depth" here. The modulation depth refers to the ratio of the amplitude of the modulated signal to its DC offset and in Figure 3.11 the higher distortion is caused by higher modulation depths. For the voltage pixel, once the photodiode voltage has discharged to zero it cannot discharge any further. If the photodiode voltage has not been sampled by that point then when it is finally sampled, the result will be distorted. This is more noticeable at high modulation depths because the clipping occurs more quickly. For the time and inverse time pixel a similar effect occurs, but for low DC levels coupled with high modulation depths. If the intensity of light is not sufficient to discharge the photodiode to the threshold voltage, then distortion will occur as the measurement will be incorrect.

3.9.4 Two Phase Pixel Amplitude Response

Figure 3.12 shows the results of sweeping the photocurrent amplitude from 0.05 nA to 1 nA for different DC levels. Similar to the DC level being dependent on the AC amplitude as shown in Section 3.9.3, the amplitude is also dependent on the DC level. The voltage pixel response has a fairly linear response with the DC, so measuring both allows the amplitude to be calculated correctly. The same is true of the inverse time pixel for low modulation depths, but using the time pixel means that the amplitude can not be easily reconstructed even knowing the DC, due to the non-linear relationship between the two parameters.

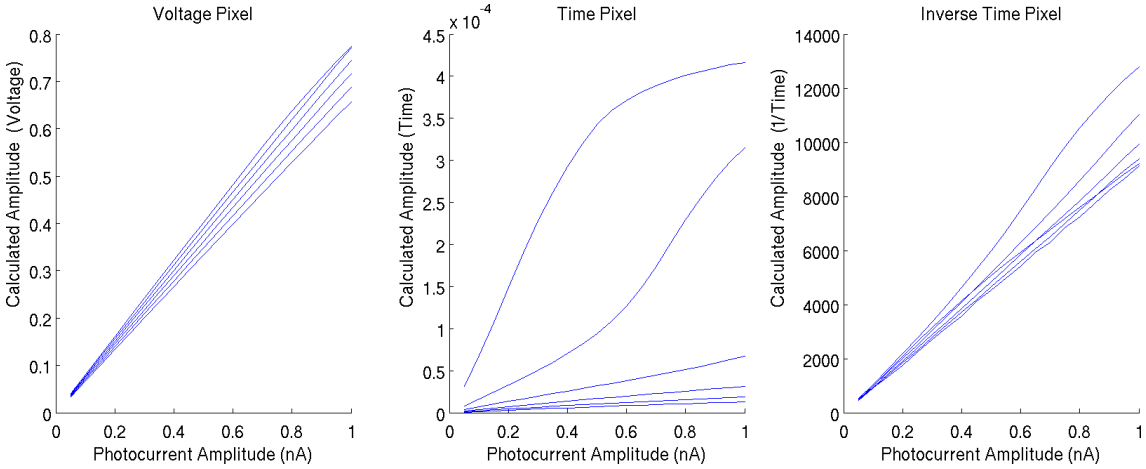
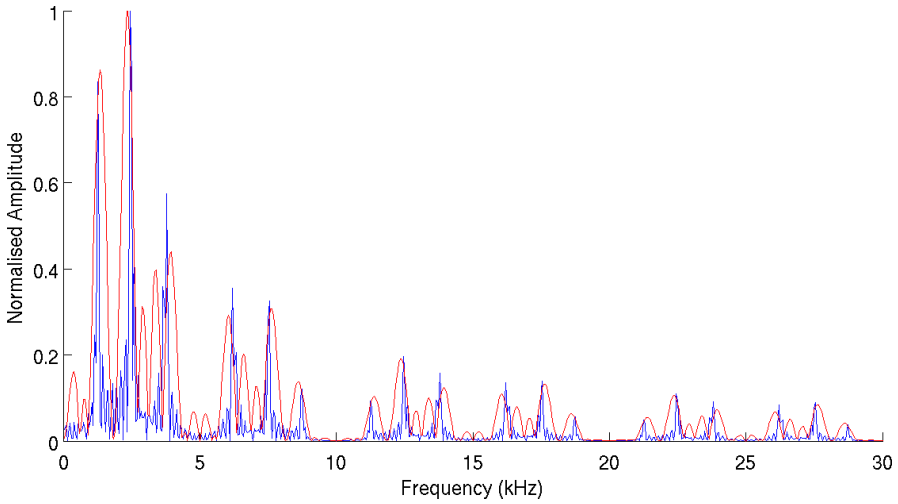


Figure 3.12: Two Phase Pixel Amplitude Response

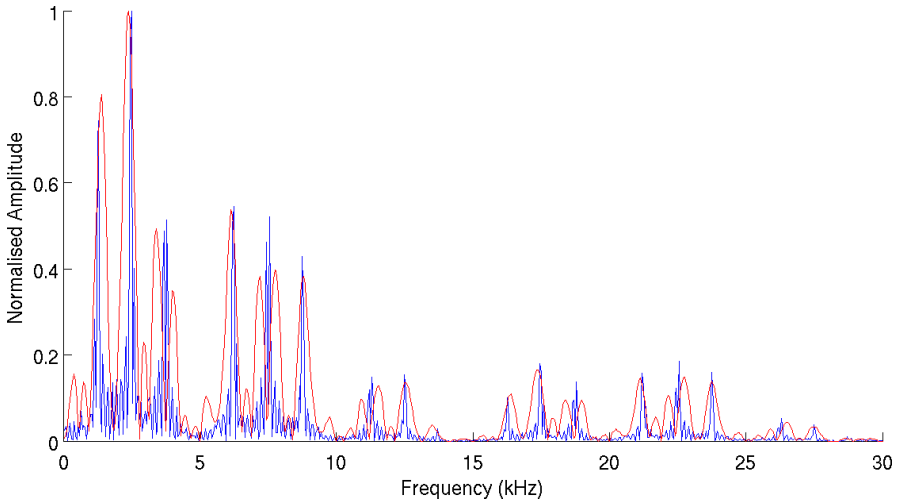
3.9.5 Two Phase Pixel Frequency Response

The frequency response to different modulation frequencies for a fixed local oscillator frequency for the three types of pixel is shown in Figure 3.13. The different plots represent the response for a single set of two samples and for moving average filters of length 2 (red) and 8 (blue). These filter lengths were chosen to allow good comparision between the results. All of the results are normalised to the peak value for easier comparison.

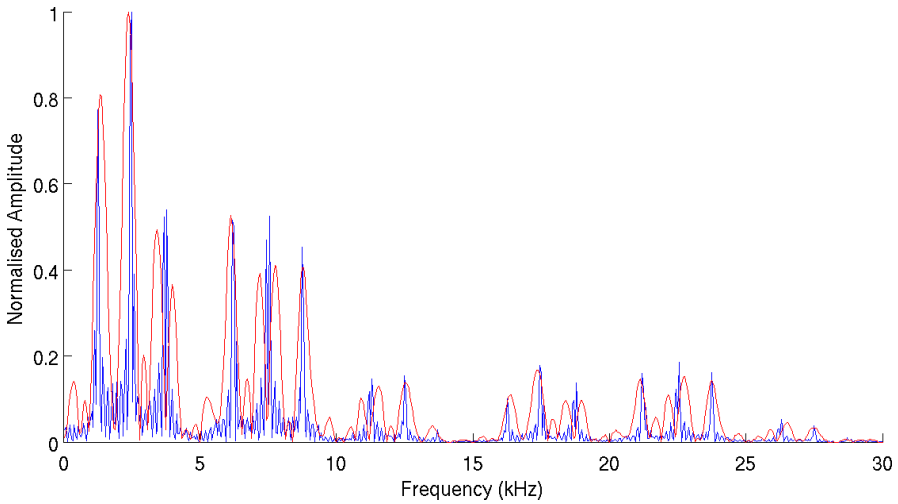
The results show that the lock-in ability of the two phase mixing technique is not particularly good, since it has significant sensitivity to frequencies other than the



(a) Voltage Pixel



(b) Time Pixel



(c) Inverse Time Pixel

Figure 3.13: Two Phase Pixel Frequency Response
(Blue: Filter length 8, Red: filter length 2)

lock-in frequency at 2.5 kHz. It can be seen that increasing the length of the moving average filter helps to make the frequency response narrower at the lock-in frequency and the harmonics, so improving the signal to noise ratio.

3.9.6 Four Phase Pixel DC Response

As should be expected, the DC response for the four phase pixels is practically indistinguishable from the two phase pixel response, so it is not repeated here.

3.9.7 Four Phase Pixel Amplitude Response

Although it is useful to be able to measure the DC intensity in the presence of a modulated light signal, the amplitude is usually the parameter that is of most interest in this work. Figure 3.14 shows the results of sweeping the photocurrent amplitude from 0.1 nA to 1.0 nA with a DC offset of 3 nA. The results show that in this regard there doesn't appear to be a great deal of difference between the different pixels, with only the time pixel showing any signs of non-linearity. The results in Figure 3.14 are for a moving average filter of length four, but as expected the length of the filter has no effect on the amplitude here since this frequency is in the pass band.

This is not the complete picture of course. As with the two phase technique, the amplitude shows a dependence on the DC photocurrent offset. Figure 3.15 shows the effect of sweeping the DC offset for amplitudes between 0.2 nA and 3.0 nA in steps of 0.2 nA and they all show some dependence on the DC. Starting with the voltage pixel results, the bulk of the results in the region where the DC photocurrent is between 2 nA and 4 nA at least show a linear dependence on the DC photocurrent meaning that it could be corrected for if necessary. Outside of this region it is again fairly linear, however with a marked difference in slope and direction. The change in slope at low DC currents is caused by the photocurrent being clipped to zero

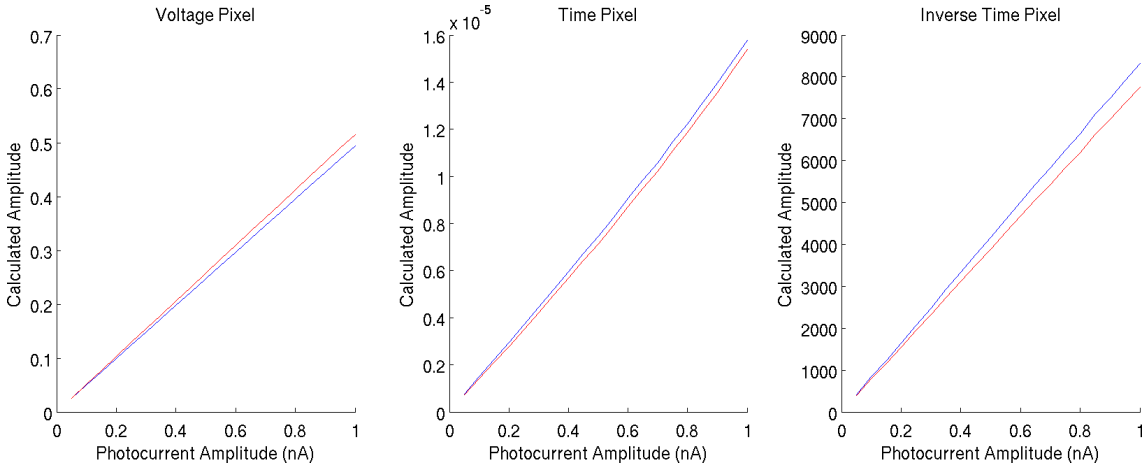


Figure 3.14: Four Phase Pixel Amplitude Response
(Blue: Lange, Red: Quadrature)

where the amplitude is relatively large compared to the DC offset. As such the results make sense, with a larger amplitude, the clipping will remain present until the DC current component becomes large enough to prevent the clipping, just as is shown on the plots. If negative currents are allowed in the simulation, then the slope marked as "a" remains at a fixed gradient all the way to the beginning of the plot at 0.2 nA. The change in slope at high DC currents (marked as "b") is again caused by clipping, but this time when the photodiode voltage is clipped at zero volts, with a similar effect on the plot.

The results for the time pixel are rather discouraging because of the relatively high dependence on the DC current. Although the amplitude response is reasonably linear at a given DC current, changing the value of the DC produces dramatic changes in the calculated amplitude. Post-processing the results further in an attempt to remove the DC dependency can be achieved using Equation 3.42 and does provide passable independence from the DC value for the Lange time pixel, yet with a substantial increase in the computational requirements.

$$Time\,Pixel\,Amplitude\,(Lange) \;=\; \frac{DC}{\sqrt[4]{\left(\frac{1}{S_I}\right)^2+\left(\frac{1}{S_Q}\right)^2}} \tag{3.42}$$

The inverse time results are quite favourable, displaying reasonable linearity along

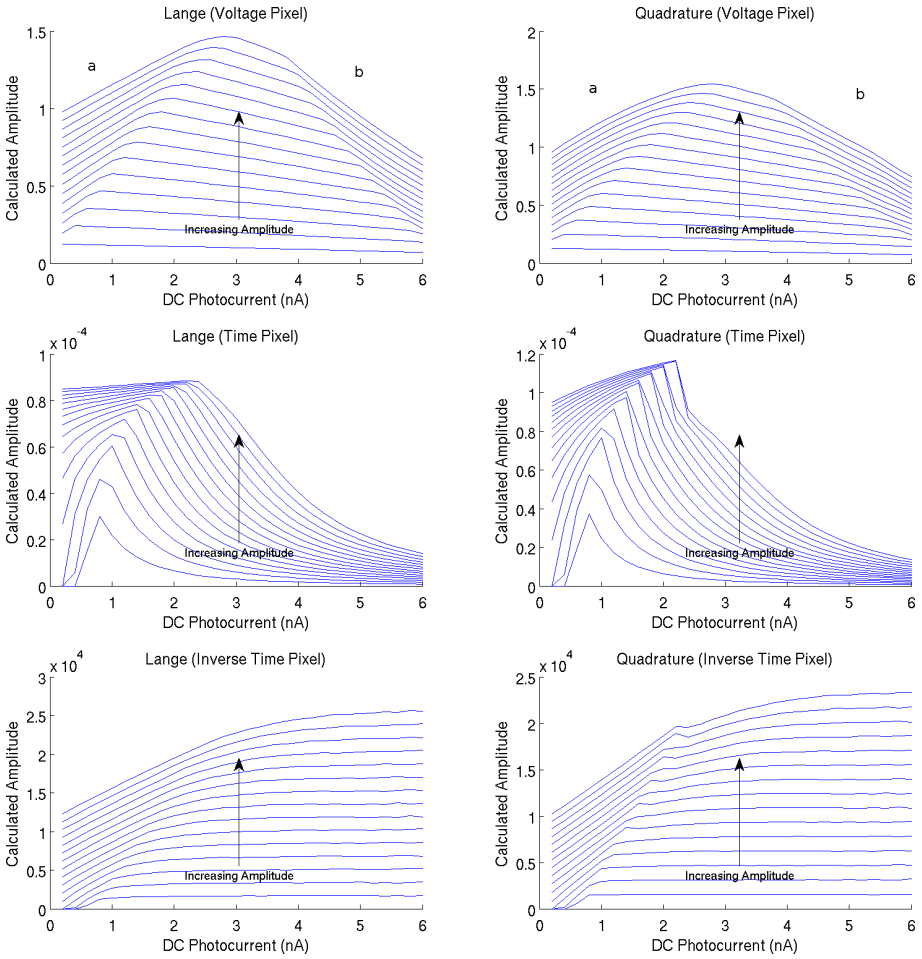


Figure 3.15: Four Phase Pixel Amplitude / DC Response

much of the range. Clipping occurs again when the amplitude is too high for a given DC offset. The bulk of this is caused by the samples taken in the negative going part of the cycle not reaching the threshold, with an additional small effect caused by the photocurrent being clipped to zero in the simulation (so there is no negative light in the simulation).

3.9.8 Four Phase Pixel Phase Response

Figure 3.16 shows the error in the phase response of the pixels as the phase of the carrier is swept, where the blue lines show the results for a DC offset of 3.0 nA and the red lines are for a DC offset of 1.0 nA. The different lines in each plot correspond to amplitudes of 0.1 nA, 0.5 nA and 1.0 nA. The vertical lines on some of the plots are caused by phase wrapping and are of no concern. The results for the Lange method are quite consistent over the different types of pixel but with an error of up to 90° it is far worse than the Quadrature method. For the Quadrature method, after experiencing the results for the amplitude response with respect to the DC offset, it is no great surprise that the inverse time pixel behaves better at low modulation depths. Likewise at higher DC offsets the voltage pixel results become much less favourable. The voltage pixel results at an offset of 1 nA DC are the most consistent results of the set, with a fixed error of 1.98° and a maximum variation of ±0.04°, with the next best result being for the inverse time pixel which has an error of $2.5 \pm 1.5°$. These variations look quite extreme when reports of phase accuracy of less than 1° are made for other pixels [Ando and Kimachi, 2003] and much lower for commercial single point lock-ins. If these pixels were to be used in range finding or other phase based measurement systems the phase error would have to be improved.

As with the amplitude response, the effect of the DC offset on the phase response must be considered. This is shown in Figure 3.17. The sets of curves in each plot are for amplitudes between 0.2 nA and 3.0 nA in steps of 0.2 nA. Despite looking very dramatic, none of the results are necessarily that bad. In the offset range of

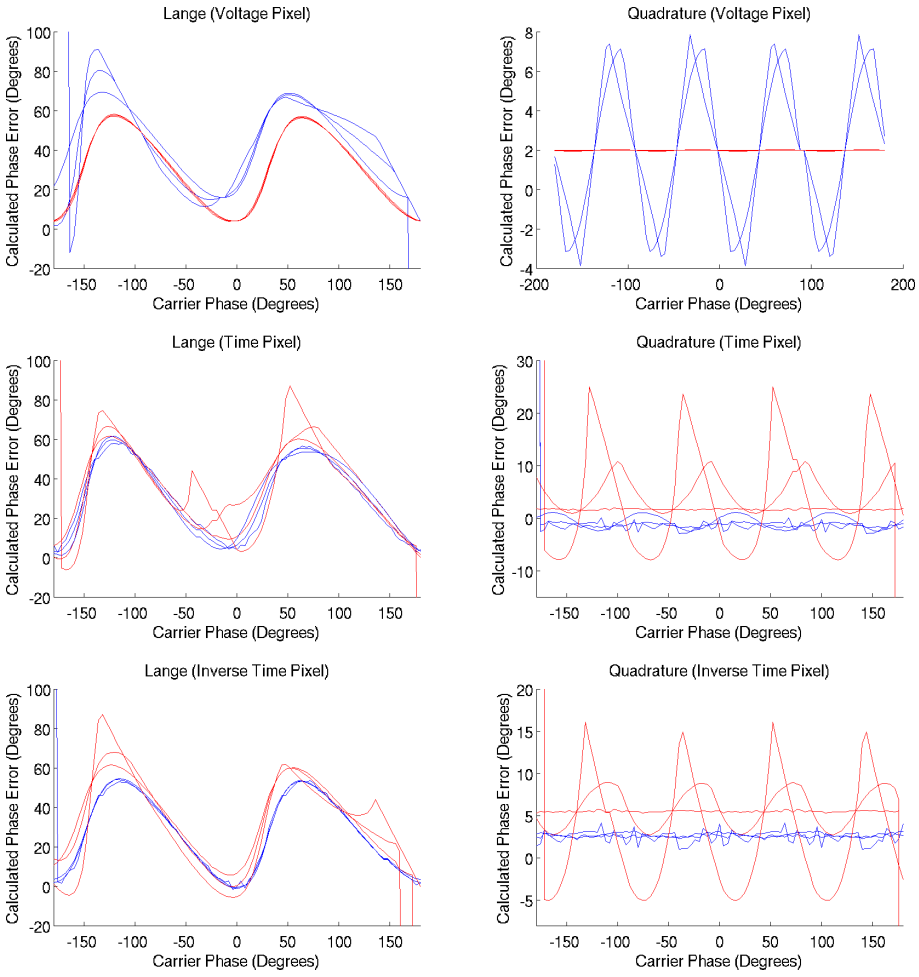


Figure 3.16: Four Phase Pixel Phase Error
(Blue: 3 nA DC photocurrent, Red: 1 nA DC photocurrent)

approximately 2.0 nA to 4.0 nA, the Lange voltage pixel produces results that vary by at most 0.5° and, although being dependent on the amplitude, it is a fairly linear relationship. The results for the Quadrature method voltage pixel are also quite reasonable in the same range, although there is no longer the linear relationship with the amplitude.

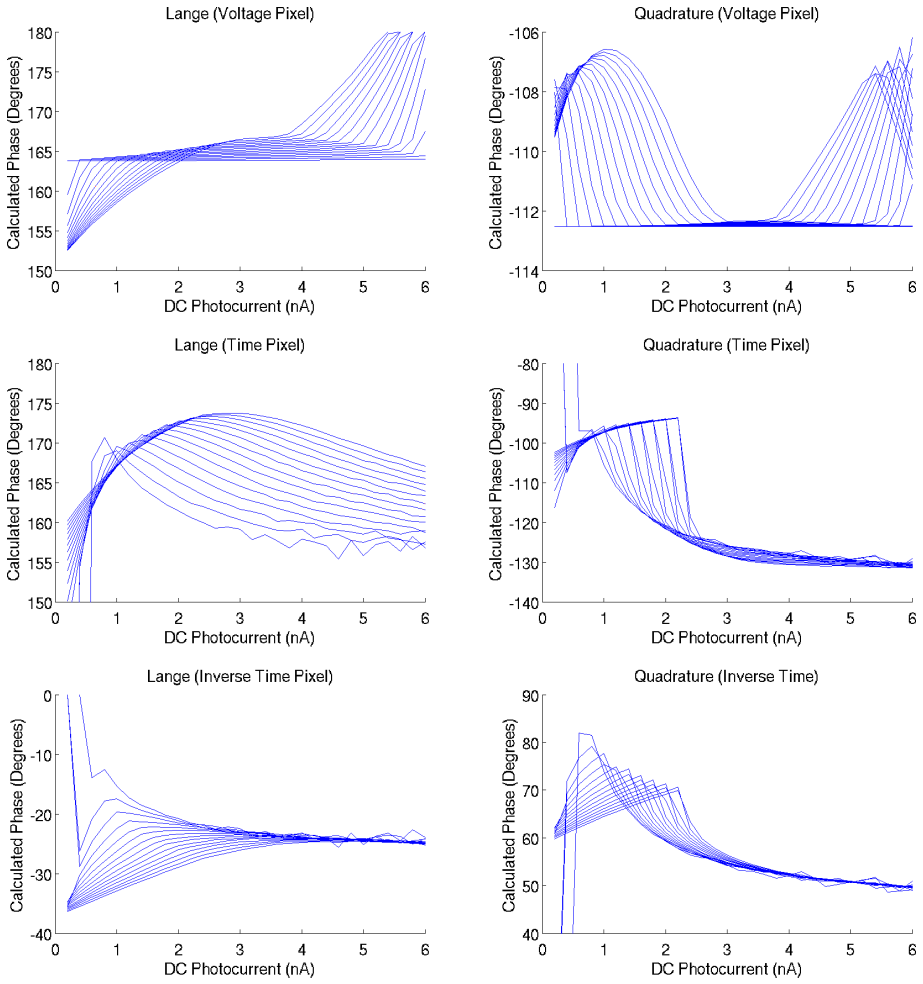


Figure 3.17: Phase Response to DC Offset

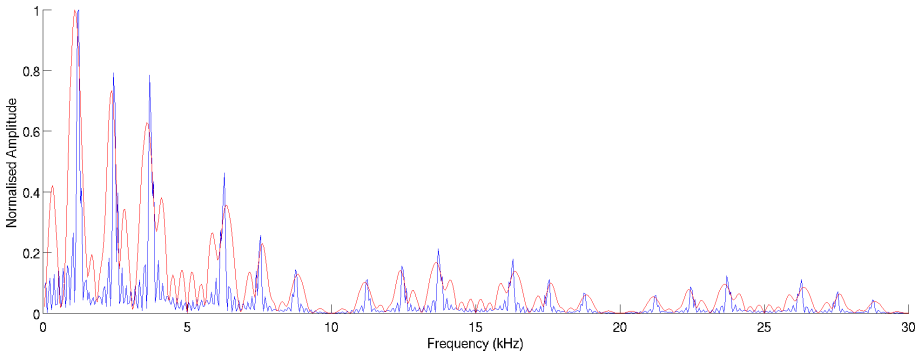
3.9.9 Four Phase Pixel Frequency Response

The ability to only be sensitive to a very narrow range of frequencies is an important feature of a lock-in. By rejecting every frequency other than that which is of interest to us, the amount of noise moving through the system can be reduced and increase the signal to noise ratio. To test the frequency sensitivity of the different pixels

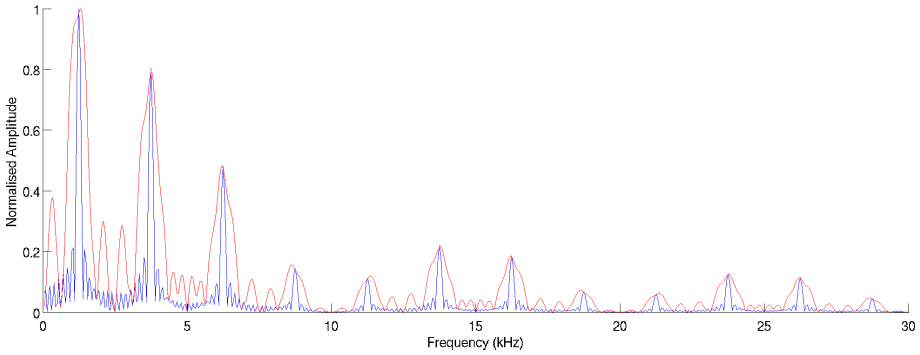
the LO frequency was locked and the carrier frequency was swept between 50 Hz and 30 kHz in 50 Hz steps. The response is shown in Figure 3.18, where all of the amplitudes are normalised to their peak value, the red plots are for a moving average filter of length 2 and the blue plots are for a moving average filter of length 8. The time pixel is omitted in this figure since it is very similar to the inverse pixel response. The peak at a value of 1 is the frequency of interest.

It is quite clear to see that increasing the length of the moving average filter narrows the sensitive frequency bands even with the relatively small difference in the two filters. The Quadrature and Lange technique are very similar for both types of pixel with a very notable exception that there are additional frequencies at the second, sixth, tenth harmonics and so on at every other even harmonic of the frequency of interest. Comparing the results between the two pixels makes the inverse time pixel look more favourable, but in reality due to the different threshold points that they are using they are actually very similar.

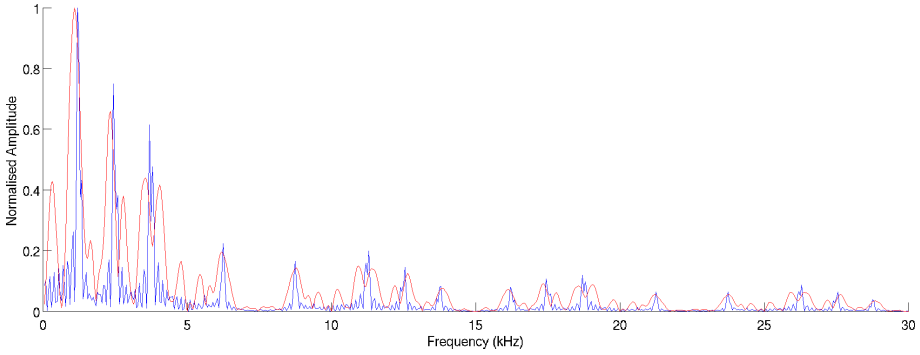
The overall frequency response consists of peaks at the odd harmonics of the frequency of interest, with the height of the peaks following a sinc function, with the sinc reaching zero at 10 kHz and 20 kHz for the voltage pixel and approximately 7.5 kHz and 15 kHz for the inverse time pixel in Figure 3.18. The response of this sinc function is determined by the duration of the integration periods relative to the overall length of the possible integration period (which is slightly less than one quarter of the modulation time period). In Figure 3.18 the voltage pixel was set to measure the voltage after 100 μs or one half of the possible duration. The time pixel was set to take a measurement when the photodiode voltage decayed to 2.5 V. By changing these two threshold values, the frequency and amplitude response of the sinc function can be controlled. By decreasing the integration period the sinc function response expands and by increasing it the sinc function contracts. Both of these cases are shown in Figure 3.19 for the Quadrature voltage pixel, where the red plots are for an integration period of 50 μs and the blue plots are for 150 μs. Clearly integrating for as long a period as possible is the most desirable from the point of



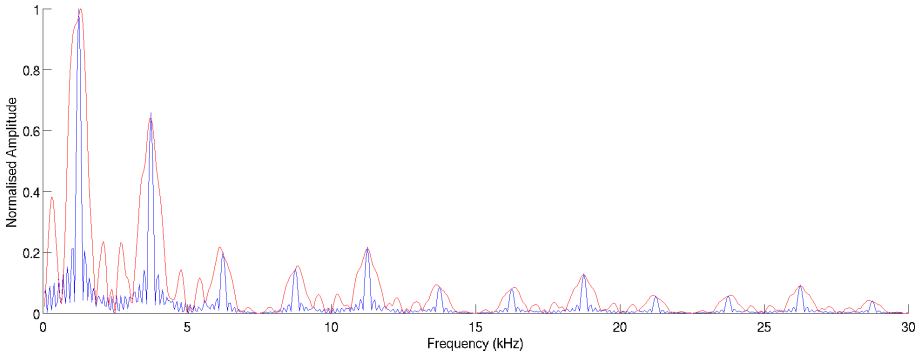
(a) Lange Method Voltage Pixel



(b) Quadrature Method Voltage Pixel



(c) Lange Method Inverse Time Pixel



(d) Quadrature Method Inverse Time Pixel

Figure 3.18: Four Phase Pixel Frequency Response
(Blue: filter length of 8, Red: filter length of 2)

view of the frequency response.

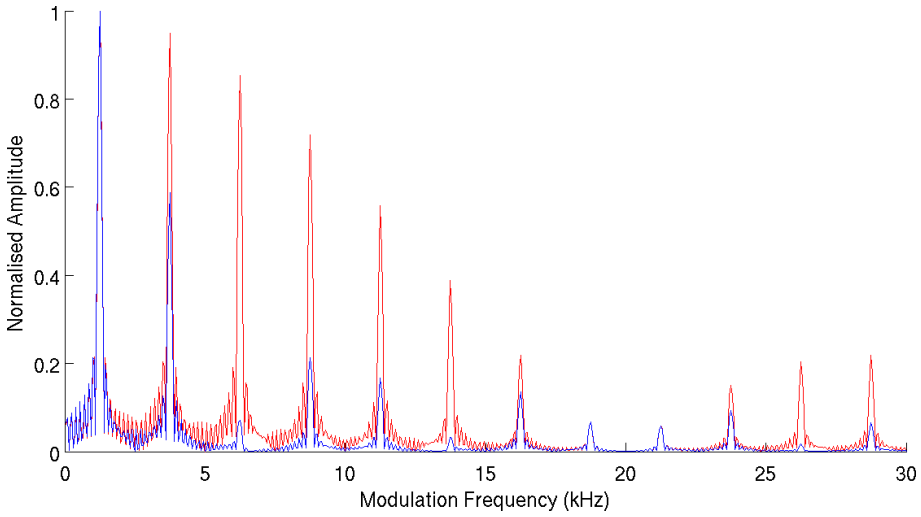


Figure 3.19: Quadrature Voltage Pixel Frequency Response
(Blue: 150 μs integration length, Red: 50 μs integration length)

3.10 Summary

This chapter has introduced the idea of modulation and why it is necessary. The different types of continuous wave analogue modulation have been introduced and the theory of amplitude modulation discussed. The sampling theory was raised in conjunction with the consideration of where in a camera system the sampling takes place and the practical effects that this has on the sample rate and the overall partitioning of the system. Some practical methods for simplifying mixing were introduced and simulated to test their effectiveness with three different integrating pixel types and three different methods of reconstructing the modulated signal (the two phase method, the Lange method and the Quadrature method, where the latter two both take four samples per modulated clock cycle).

Of the three methods the proposed Quadrature method provides the best results in every regard, giving a more linear amplitude response, lower phase error and reduced sensitivity to harmonics than the other methods. Because the two phase method needs a set phase between the modulated signal and the local-oscillator, it

is much less generally useful than the four phase methods, however it can still be used in experiments where this phase relationship can be controlled.

Now considering the differences between the three pixel types (for the Quadrature method); from the point of view of the frequency response, all of the pixel types are all but identical (the change in the responses shown from pixel type to pixel type is caused by the integration periods being different). This leaves the amplitude and the phase response to consider. In terms of the amplitude response itself, there is not a great deal of difference between the pixels, however they do show varying degrees of dependence on the amount of DC light. The inverse time pixel produces the best overall response and only begins to suffer problems when the modulation depth becomes high. The voltage pixel has similar problems, but also suffers with distortion at high overall light levels, meaning that it is only working at its best for a range of values. The time pixel amplitude response does show a strong 1/x dependence on the DC light level, but this can be countered for (at least for the relatively low modulation depths) by normalising for the DC. For the phase response the voltage pixel is the clear winner in terms of the phase error, however at low modulation depths both the time and inverse time pixel have a relatively low error rate as well.

The conclusion of the theory work is that designing a modulated light camera relies on three different parameters being set, namely the modulation frequency (as appropriate for the experiment, to move away from low frequency noise for example, this also sets the required sampling frequency), the filter bandwidth (set to accommodate the range of possible modulating frequencies caused by the experiment) and the frame rate, since the sampling frequency after the filter may still be much higher than its bandwidth.

Chapter 4

Pixel Designs

4.1 Introduction

This chapter describes the design of the test chips that were built as prototypes to the final camera chip. The designs use the integrating pixel as described in Chapter 1, but implemented as the time pixel as described in Chapter 3. Three designs were fabricated in all, with each being part of a multiple project chip shared between many other designs. The first design was a single pixel intended as a simple proof of concept design that would also give an opportunity to evaluate the CMOS process being used. The second was a scalable version of the first pixel and consisted of a 2 x 2 array used to gain some preliminary information on pixel to pixel variations across a chip and from chip-to-chip. The third design was a standalone comparator used to assess chip-to-chip variation when advanced layout matching techniques were not used.

The first chip was manufactured on the AMS CSI process and the latter two on the AMS C35 process. The two processes are roughly compatible from a layout design perspective, with the C35 process having an additional metal layer (four instead of three) and more favourable design rules. However there are differences between the

two processes in terms of device performance, meaning that the second chip was also used to evaluate the C35 process.

4.2 Architecture

As already introduced in Chapter 3, the sensor design is very similar to the traditional APS design as both are based on a photodiode being charged to a known voltage and then allowed to discharge at a rate proportional to the light intensity striking the detector. The traditional APS measures the voltage across the photodiode after a fixed time whilst an alternative approach is to measure the amount of time that it takes the photodiode to discharge to a fixed voltage. By starting a counter when the reset is released and then taking the value of the counter when the photodiode voltage passes the fixed value, the time of the discharge can be calculated [Pui et al., 2004]. A voltage comparator can be used to detect the point at which the photodiode voltage has reached the threshold and so halt the counter. This setup is shown in Figure 4.1.

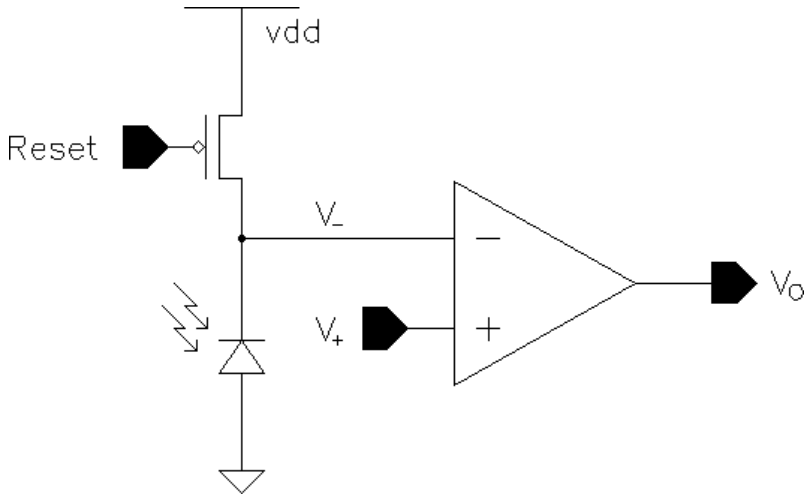


Figure 4.1: Pixel Architecture

4.3 Comparator Design

4.3.1 Operation and Non-Idealities

A comparator is a circuit used to compare two values. Digital comparators compare two numbers whereas analogue comparators compare two voltages or currents. This work is only concerned with analogue voltage comparators so the following description is written from that perspective.

In the ideal case, the operation of the comparator is as follows, assuming that the output can change between V_{OH} and V_{OL}:

$$V_O = \begin{cases} V_{OH} & \text{for } (V_+ - V_-) > 0 \\ V_{OL} < V_O < V_{OH} & \text{for } (V_+ - V_-) = 0 \\ V_{OL} & \text{for } (V_+ - V_-) < 0 \end{cases} \tag{4.1}$$

Where V_+, V_- and V_O are as defined in Figure 4.1.

In reality, of course, any real comparator will be non-ideal. The most obvious effect is the finite gain; in Equation 4.1 V_O can change from V_{OH} to V_{OL} for an infinitesimally small change in the input, implying an infinite gain. A finite gain means that the change from V_{OH} to V_{OL} occurs over a wider input range, specifically V_{IH} to V_{IL}, which are the input voltages required to set the output to just reach V_{OH} and V_{OL} respectively. This allows the equation for the operation of the comparator to be written more correctly as:

$$V_O = \begin{cases} V_{OH} & for\ (V_+ - V_-) > V_{IH} \\ A(V_+ - V_-) & for\ V_{IL} < (V_+ - V_-) < V_{IH} \\ V_{OL} & for\ (V_+ - V_-) < V_{IL} \end{cases} \tag{4.2}$$

Where A is the gain of the comparator, defined as

$$A = \frac{V_{OH} - V_{OL}}{V_{IH} - V_{IL}} \tag{4.3}$$

Another non-ideal effect observed in comparators is known as input offset voltage. Input offset consists of two components; a systematic offset and a random offset. The systematic offset is caused by incorrect design, such as having two transistors sized differently when they should be the same size. The random offset is caused by random process variations in device sizes that are a natural part of any design. These offsets are fixed when the device is fabricated. Input offset manifests itself as an additional voltage on one of the inputs, so the output no longer switches at $V_+ - V_- = 0$, rather at $V_+ - V_- = V_{IO}$.

4.3.2 Simple Comparator

A basic comparator can be built out of a differential pair and a current sink inverter as an output stage, as shown in Figure 4.2. This circuit is basically the same as an operational amplifier so the same triangular symbol is used to represent this block.

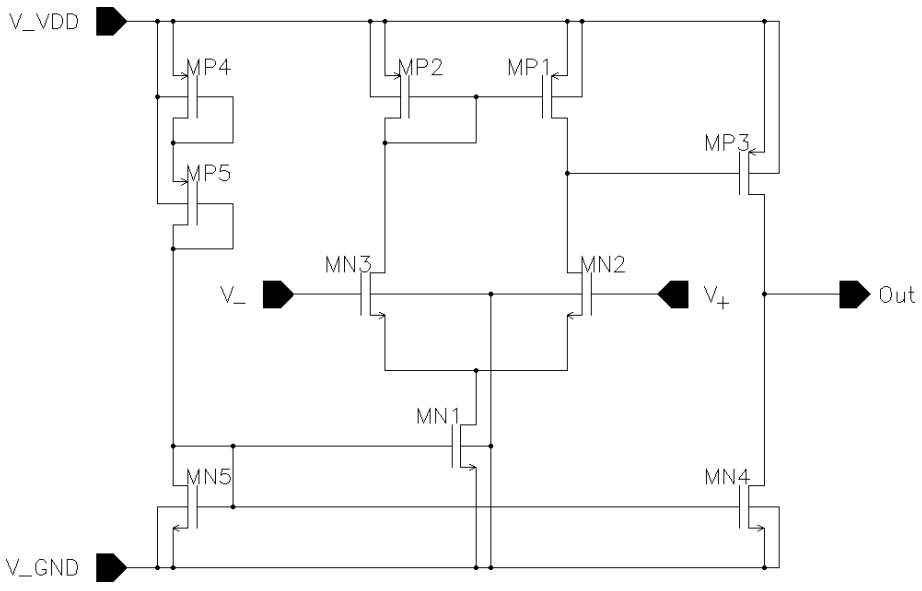


Figure 4.2: Differential Pair Comparator

4.3.3 Comparator with Hysteresis

Although the circuit shown in Figure 4.2 does function as a comparator, a comparator has different requirements to an operational amplifier, so specialised circuits have been created to fulfill these requirements. When the input to a comparator is noisy and the propagation time of the comparator is faster than the noise then the input may pass the threshold many times in quick succession and as a result the output will be likely to have spurious transitions. This is demonstrated in Figure 4.3, where the results of a simulation with the comparator from Figure 4.2 are shown. One input to the comparator, V_+, is held at a constant voltage and the other input, V_-, is falling but with a small, high frequency sine wave superimposed upon it that is large enough to cause V_- to cross V_+ multiple times as the "proper" signal component of V_- decreases. These multiple crossings cause the output of the comparator to switch multiple times as can clearly be seen. Any circuit relying on the value of the comparator in this case is liable to produce incorrect values, especially if the comparator is being used as a clock input to a logic signal as this would cause multiple clock events to occur.

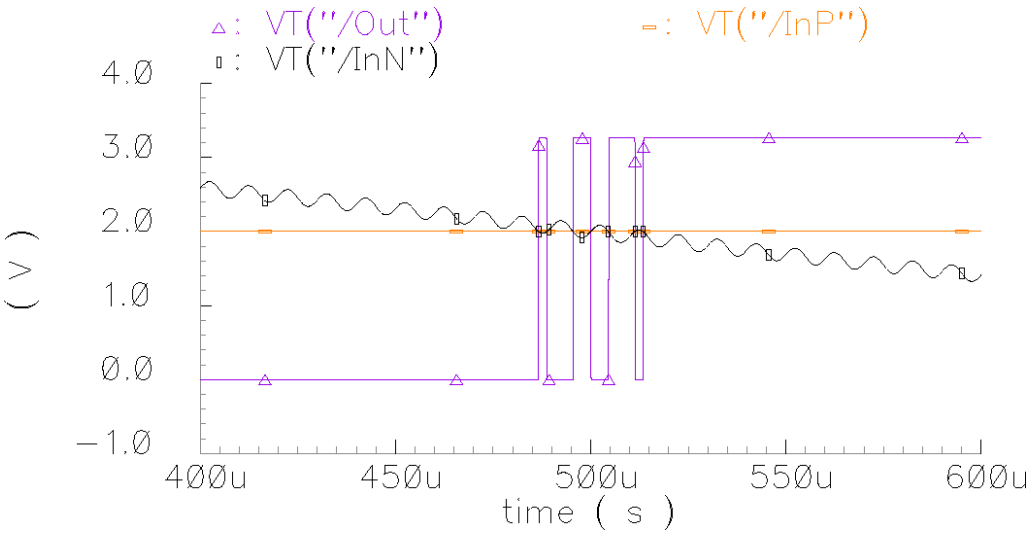


Figure 4.3: Comparator Response in Noisy Environment

To prevent this happening, hysteresis must be added to the input of the comparator. Hysteresis is a feature, not limited to comparators, whereby the threshold condition

changes after it has been exceeded. In the previous example, this would mean that as the value of V_- passed the value of V_+ for the first time, the threshold which was previously at V_+ would shift to being above V_+. As long as this shift is greater than the amount of noise on the input then the output of the comparator will remain glitch free. Figure 4.4 shows how the transfer characteristic appears in practice; as the input increases it follows the right hand curve and the output changes at the high threshold. For the output to change back to the low state, the input must now decrease to the lower threshold shown by the left hand curve.

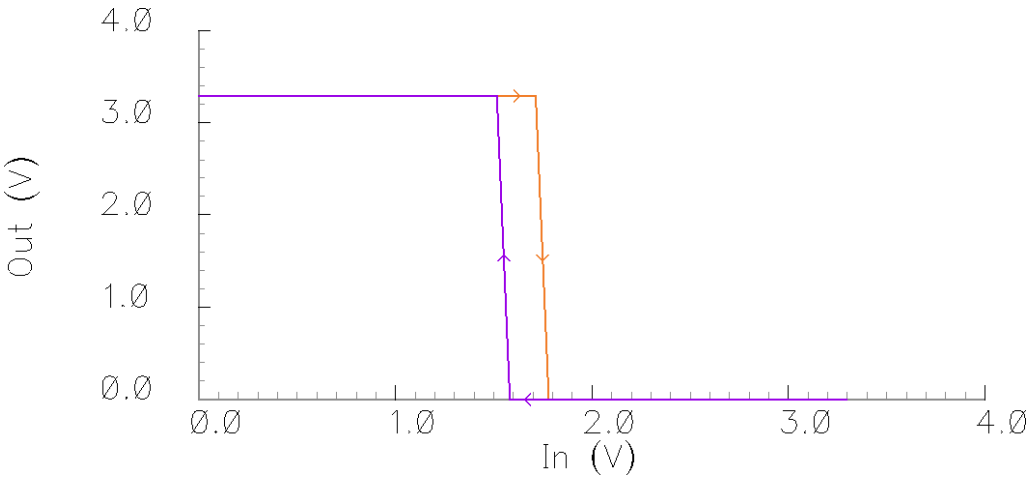


Figure 4.4: Hysteresis Curve

A design of a comparator with hysteresis and output stage is shown in Figure 4.5 [Allen and Holberg, 1987].

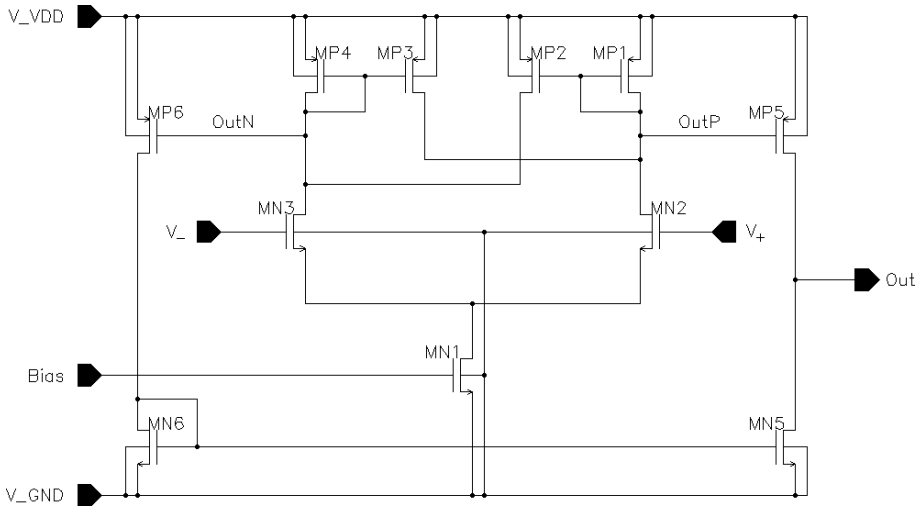


Figure 4.5: Comparator with Hysteresis

Initially consider the first stage of the comparator. MN1, MN2 and MN3 (the differential pair and tail) are identical as in a usual comparator, but the active loads (a differential to single ended converter in the standard comparator) have been replaced with cross coupled loads MP1 to MP4. To explain how the comparator works, consider V_+ as a fixed positive voltage half way between V_{DD} and ground, and V_- as starting at ground, then increasing steadily to V_{DD}.

In the initial state, MN2 is on and MN3 is off. This means that all of the current flowing through MN1 is also flowing through MN2. To resolve the state of the remaining transistors, suppose that the net labeled as OutP is high. This means that both MP1 and MP2 are off. Because it has already been shown that current is flowing through MN2 and MP1 is off, there must be the same current flow through MP3. For this to happen, OutN must be low and for that to happen, MN3 must have current flow. As this is not the case, OutP cannot be high.

With OutP low, MP1 is on and conducts all of the current through MN1 and MN2. Although V_{GS} for MP2 is sufficient to turn the device on, with OutN high the V_{DS} of MP2 is too low to allow MP2 to conduct. If for some reason this is not the case then MP2 conducts and charges OutN to close to V_{DD}.

As V_- increases and approaches V_+, MN3 becomes increasingly conductive and so OutN will decrease; MP2 will start to conduct and so there will be current flow through MP2 and MN3. As there is a fixed amount of current flowing through MN1, the current in MN2 must therefore decrease and so OutP increases. When V_+ and V_- are equal, the current flowing through MN2 and MN3 will be roughly equal. However, if MP2 and MP3 are wider than MP1 and MP4, then the same current through each of the two differential pair transistors will mean that OutP and OutN will not be equal since in this instance, OutN is higher. This also gives the reason why the currents between MN2 and MN3 are not exactly equal (as the two devices have different V_{DS} conditions). At this stage, all of the current is flowing through MP1 and MP2, with the majority through MP2.

The trip point occurs when OutP equals OutN. At this point, the V_{GS} and V_{DS} of the devices MP1-MP4 are identical so all of them are equally conductive. By designing MP3 to have a larger width to length ratio than MP1 (and MP2 a larger width to length ratio than MP4), the effective resistance between OutP and V_{DD} will drop dramatically at this point (in reality it is of course a smooth transition). This means that OutP charges toward V_{DD} through MP3, with the current through MP1 and MP2 dropping at the same time due to the rise in OutP. With MP2 turning off, the current flowing through it switches instead to flow through MP4. With no current flowing through either MP2 or MN2, there is only current flowing in MN1, MN3 and MP4 and so OutN will drop to the steady state value as dictated by the sizes of the transistors and the value of the bias current.

The comparator is now in the same state as at the beginning, but the currents and voltages are switched over to the transistors in the left hand branch of the circuit. To trigger the comparator again, V_- must drop to below V_+ so that OutP and OutN are equal.

4.3.3.1 Output Stage

Since digital circuits expect signals close to the power supply rails, the above comparator requires an output stage. OutP and OutN can both provide V_{DD} quite easily as a maximum voltage, but the minimum voltage they can provide is determined by the bias current set through MN1 and, to a lesser degree, by the size of the pair transistors, MN2 and MN3. As there is always current flowing through MN1 and MN2/MN3, there will always be an associated voltage drop across the devices. The output stage eliminates this problem and helps provide better drive capabilities. MN5, MN6, MP5 and MP6 make up the output stage, which is a push-pull type circuit. When OutP is high, MP5 will be turned hard off. At the same time, OutN will be well above the threshold voltage of MP6, so it will be turned on. With MP6 turned on, the gates of both MN6 and MN5 will be at roughly V_{DD}, so they will

be turned on and so Out will have a strong connection to ground. When OutP is low, MP5 ends up turned completely on and MN5 is turned completely off. This arrangement gives rail to rail outputs with a low output impedance.

4.4 Chip One - MLCv2

4.4.1 Overall Design

The design submitted on the MLCv2 chip is shown in Figure 4.6, where the IO pads have been removed for clarity.

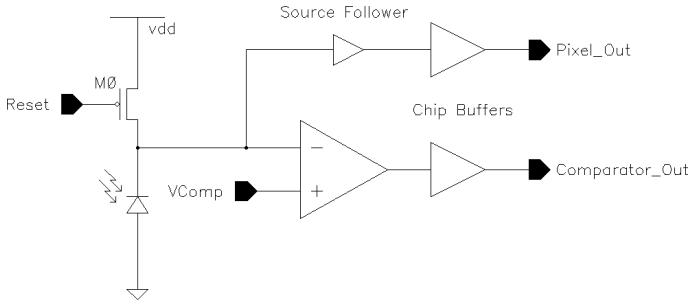


Figure 4.6: MLCv2 Schematic

4.4.2 Front End Circuit

The photodiode and the reset switch are referred to as the front end circuit. With only a single photodiode and transistor to specify there are only a few design decisions to be made, however they are important since any poor choices will impact significantly on the performance of the whole design. The size and type of both the photodiode and the reset switch must be chosen carefully.

The reset switch can be either an NMOS, PMOS or CMOS device. When the photodiode is reset it is charged to V_{DD} and hence the switch must be able to pass a voltage up to V_{DD} successfully. To evaluate which device is best to use, consider where the photodiode is just about to be reset from a low voltage to a high voltage.

4.4.2.1 NMOS Switch

With an NMOS switch the reset voltage must be taken high so that $V_{GS} > V_T$. The photodiode voltage corresponds to the source of the NMOS switch, so as it charges both V_{GS} and V_{DS} of the switch decrease. With the normal case of $V_G = V_D = V_{DD}$, the transistor is in the active load or diode connected configuration. In this configuration, $V_{GS} = V_{DS}$ and so the device will always be in saturation. Using this information, the voltage across the transistor can be obtained from the I/V equation of a transistor in saturation as shown in Equation 4.4.

$$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}(V_{GS} - V_{T_N})^2 \tag{4.4}$$

$$V_{GS} = \sqrt{I_D/(\frac{\mu C_{ox}}{2}\frac{W}{L})} + V_{T_N} = V_{DS} \tag{4.5}$$

Where μ is the mobility of the charge carriers, C_{ox} is the oxide capacitance, W and L are the width and length of the transistor respectively and V_{T_N} is the threshold voltage of an NMOS transistor. Equation 4.5 shows that the voltage across the switch has a square root dependency on the current flowing through it (i.e. the photocurrent). This means that the voltage to which the photodiode eventually charges ($V_{DD} - V_{switch}$), has the same dependency on the photocurrent. As the intensity of light is not known at the reset point, it is not possible to determine the voltage at which the photodiode started to discharge and so measurements of the time taken to reach the reference voltage will be of little use. The same problem exists when measuring the voltage in a traditional APS. The results of a simulation that swept the photodiode current for a frontend circuit with a 50 x 50 μm photodiode and a 2/1 μm NMOS transistor showed that there is a 230 mV change in output voltage when the photocurrent changes from $0-20$ nA and a 52 mV change for the range $20-100$ nA. As photocurrents tend to be quite low, especially for small photodiode sizes, this represents a significant problem. Furthermore as the parameters of the transistors in different pixels will not be identical there will be even more uncertainty in the measurement. Based on 1000 Monté Carlo runs of the

same current sweep simulation, the pixels on a single camera should be expected to have a variation of 21 mV when under uniform illumination. When considering devices on different dies, this variation could rise to 190 mV.

It is possible to remove the effect of the mismatch and prior light level by using correlated double sampling, where the voltage across the photodiode is sampled twice per integration, but this increases the complexity of the sampling circuitry. Even using correlated double sampling, the NMOS is still a poor choice because it is not able to efficiently transfer a high voltage (the 2/1 μm device already mentioned can provide a maximum of 2.6 V across the photodiode based on simulations with zero photocurrent) and it also takes a long time to reach this value. Reducing the maximum voltage to which the photodiode can be reset in this way reduces the dynamic range of the device.

By using a voltage for the reset line that is higher than V_{DD} it is possible to increase the output voltage level of the photodiode as well as reducing the amount of time taken for it to charge, but this is at the expense of additional complexity in powering the higher reset voltage and still does not remove the strong dependency of the photodiode reset voltage on the current light level.

4.4.2.2 PMOS Switch

Using a PMOS transistor as the switch removes many of the problems associated with the NMOS device. As the source contact of the switch is now connected to a fixed point, V_{DD}, then V_{GS} will no longer change as the photodiode charges. PMOS transistors in this configuration are able to pass a high voltage and so the charging time is very short. The photodiode voltage during reset is still dependent on the photocurrent because the switch has a finite resistance and must continually supply current to the photodiode to keep it charged. Because the current flow through the transistor is small (the same value as the photocurrent) and the resistance of the transistor is relatively small, then the voltage drop is also small. Sweeping

the photocurrent over the range $0 - 100\,\text{nA}$ results in a change in the photodiode voltage during reset of approximately $630\,\mu\text{V}$, with a $245\,\mu\text{V}$ chip-to-chip variation at 100 nA. The on-resistance of the switch will always be present regardless of the type of switch, although its effect can be minimised. The compromise is of course between a larger resistance and a large device.

The disadvantage of using a PMOS transistor as the reset switch is its size. As the substrate of the chip is p-type, PMOS devices must be created in an n-well. The photodiode is also an n-well and the AMS C35 process design rules state that the minimum spacing between n-wells that are at different potentials is $3\,\mu\text{m}$. Using an NMOS switch means that the minimum the spacing is reduced to $1.2\,\mu\text{m}$.

4.4.2.3 CMOS Switch

The CMOS switch, or transmission gate, is often used in digital circuits because of its primary advantage in being able to pass both high and low voltages equally well. In the case of trying to charge a photodiode to V_{DD}, this is of no real advantage as only a high voltage needs to be transmitted. In addition, the transmission gate requires complementary control lines, meaning that each pixel must either include an inverter or else an additional control line. This is not taking into account the fact that the switch needs two transistors as well. Both of these points mean that the fill factor of the pixel would be reduced to no great advantage.

4.4.2.4 Conclusion

Taking all of these factors into account, the best type of reset switch to use is the PMOS transistor. It can pass a high voltage successfully and charges the photodiode quickly. Although it consumes more area than the NMOS transistor because of the need for a n-well and the separation required between itself and the photodiode well, it has none of its other more significant disadvantages.

4.4.3 Photodiode

The photodiode used was the n-well to substrate diode due to its lower capacitance and its wavelength sensitivity being toward light of longer wavelengths [Pui, 2004] that makes it more suited for use with red lasers. The size of the photodiode was chosen as $50 \times 50\,\mu$m. This size means that it can capture significant light levels and is not too large to limit the chip size in relatively large arrays. For an integrating pixel, the capacitance of the photodiode is an important factor in determining the rate at which it will discharge for a given light level. The capacitance associated with the photodiode consists of two components; that coming from the horizontal junction at the bottom of the diode (calculated from the area of the diode), and that coming from the vertical junction at the edges of the diode (calculated from the length of the periphery of the diode). If the peripheral capacitance is ignored and the photodiode is uniformly illuminated, then changing the size of the photodiode has no effect on the discharge rate. By doubling the area of the photodiode the capacitance will double, yet so will the light sensitive area and hence the photocurrent. The peripheral capacitance complicates this since as the size of the photodiode decreases, the periphery capacitance does not decrease as quickly as the area capacitance (or the light sensitivity) and so the discharge rate becomes slower (this depends on the shape of the photodiode since periphery required to enclose a fixed area can vary).

4.4.4 Comparator

The comparator used in the pixel design of MLCv2 is a comparator with hysteresis as described in Section 4.3.3. The parameters that are most important for a comparator are gain, input offset voltage, output amplitude and propagation delay, as well as the amount of hysteresis.

The two differential input transistors are critical for ensuring that the voltage at which the comparator triggers is correct. The mismatch between these devices

contributes to the input offset voltage and so they should be the same size and have as good matching as possible in the layout. Two devices with minimal matching will suffer from ±10 mV of offset. With moderate matching this can be reduced to ±5 mV and with precise matching, usually achieved through trimming, the offset can be further reduced to ±1 mV [Hastings, 2001]. Laser trimming is not really an option for cheap devices, especially when there are potentially hundreds of devices to trim. Using good layout techniques reduces the offset and doesn't require any special considerations other than being slightly more complicated to lay out and taking more silicon space.

Device mismatch is caused by geometric, diffusion and gradient effects. The fundamental basis of most matching is that devices are affected by their environment and that to get the best matching between devices, that environment must also be matched. In the simplest case, this means that all devices to be matched must be the same size and have the same spacing and orientation. Both the spacing and orientation affect, for example, how the polysilicon gate of a transistor forms. This is because the rate at which the polysilicon is etched is dependent on the area to be etched; large areas are etched more quickly than small areas so they may end up with more etching than was intended, leading to smaller gate sizes. Conversely, the smaller areas may be under etched. Consider then the devices in Figure 4.7 which are to be matched to each other.

The devices shown are resistors made of polysilicon, with metal contacts at the top and bottom. The spacing between all devices is equal, apart from between devices 3 and 4. At first glance, it appears as though the resistors will be reasonably well matched, with the exception of number 4, but the reality is much worse. The spacing for all of the devices is different as summarised in Table 4.1 and it shows that every single device is in a different environment.

Even when device number 4 has the correct spacing the problem is not solved since 1 and 4 still have different spacings on their outer edges. The solution is to add

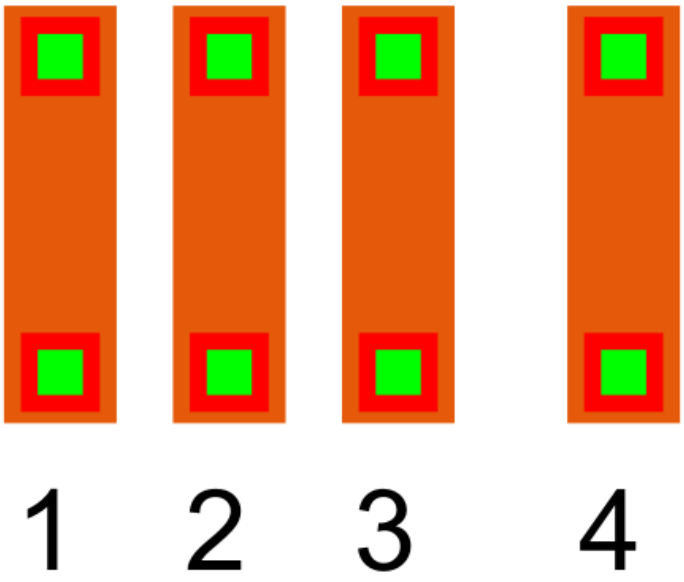


Figure 4.7: Polysilicon Resistor Spacing

Device	Left Spacing	Right Spacing
1	'infinite'	1 μm
2	1 μm	1 μm
3	1 μm	2 μm
4	2 μm	'infinite'

Table 4.1: Polysilicon Resistor Spacings

dummy resistors to each end of the array as shown in Figure 4.8. The devices 1 to 4 will now be better matched at the expense of using more space.

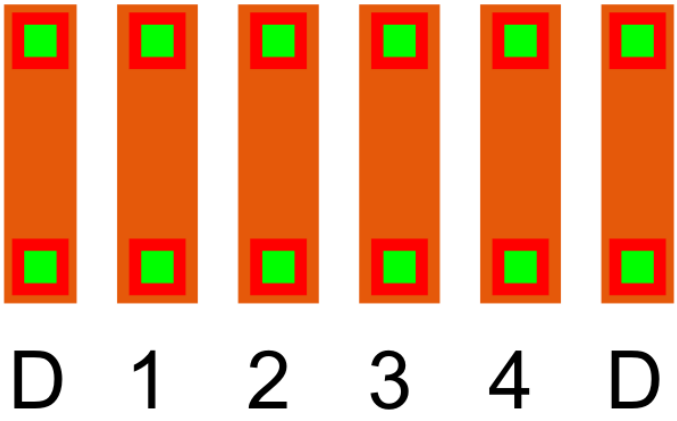


Figure 4.8: Polysilicon Resistors with Dummies (D)

Keeping the spacing equal is a good reason for keeping the orientation of devices consistent, but not the primary reason; some properties of silicon wafers have an orientation dependency, noticeably the carrier mobility. By having matched devices at different orientations to each other, the mobility of the carriers will be different in the different devices and so cause mismatch.

Gradient effects are caused by slight temperature, dopant or other variations across the wafer during fabrication as well as the stresses caused by packaging. As such, the effects tend to be quite small because the variation is across a whole wafer. To minimise gradient effects, devices should be placed as close together as possible. A further technique to minimise gradient effects is to use common centroiding. This involves splitting a device into an equal number of identical parts and then arranging them so that any gradient will be cancelled out. The different devices to be matched must have a common centroid or centre of gravity. A simple example of two transistors each split into two parts is shown in Figure 4.9, where A_1 and A_2 are two halves of the same device connected in parallel with each other (the same being true for B_1 and B_2). It is worth pointing out that whilst matching is usually done between two devices to ensure that they are the same, it is perfectly valid to use matching techniques to ensure that a device properties are as close to those

intended as possible. This is often used when trying to produce more precise passive components such as capacitors and resistors.

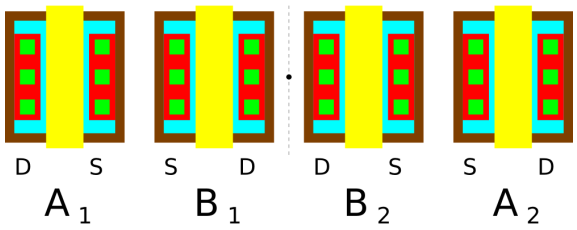


Figure 4.9: One Dimensional Common Centroid Layout

The group of devices in Figure 4.9 has rotational symmetry about the filled black circle and reflectional symmetry about the dashed line which means that any gradient induced effect in the x-axis will be cancelled out. Assuming a linear gradient $g = ax + c$ with the centroid as $x = 0$ then at the sources of both devices, the gradient will have values $g_1 = as + c$ and $g_2 = -as + c$, where s is the position of the source of either device A or B. To cancel gradients in both the x and y-axes at once requires a more complicated layout as shown in Figure 4.10 (where devices A and B are directly equivalent to those in Figure 4.9). In both Figure 4.9 and Figure 4.10, the centroid of the devices is shown by the filled black circle.

4.4.5 Source Follower

In prototype devices it is almost always desirable to have as much controllability and as much visibility as possible. For this reason, the photodiode voltage is taken off chip as well as the comparator output. The photodiode output must be buffered carefully because any change in the load at the photodiode such as increased capacitance or resistance to ground will affect the performance of the pixel. To keep parasitic capacitance and resistance on the output of the photodiode low, the buffer must be close to it; this also sets the requirement that the buffer must be reasonably compact to allow for a good fill factor in pixel arrays.

The source follower schematic is shown in Figure 4.11. MP0, MP1 and MN0 form the bias for the current source load, MN1 and MN2. The addition of MN2 to the

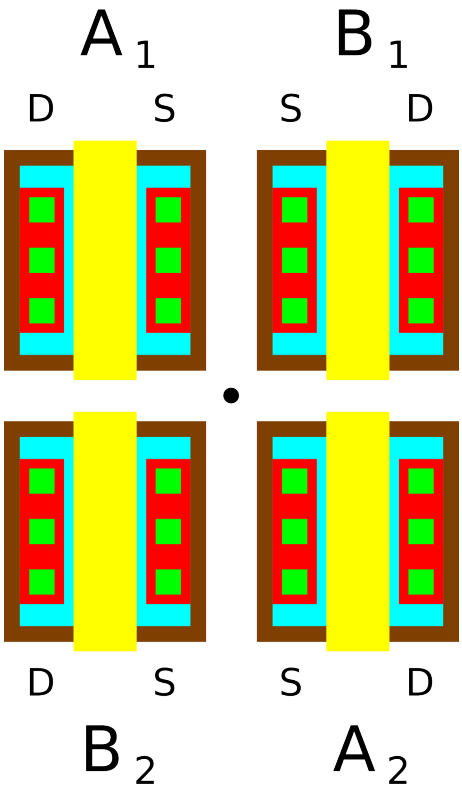


Figure 4.10: Two Dimensional Common Centroid Layout

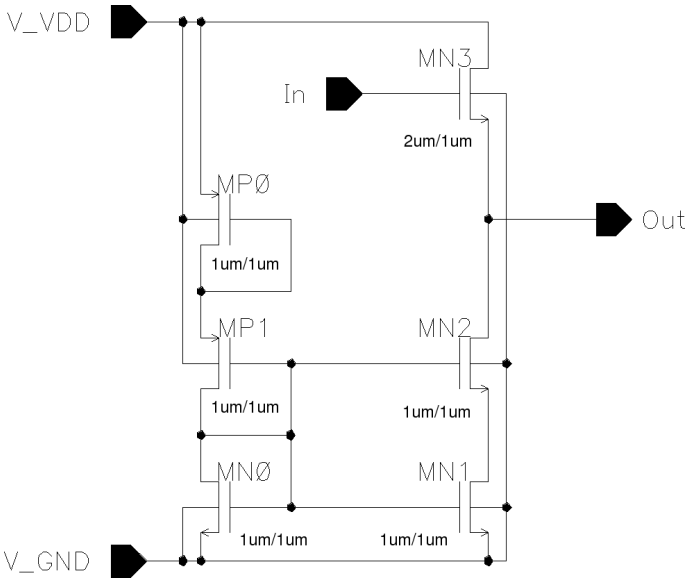
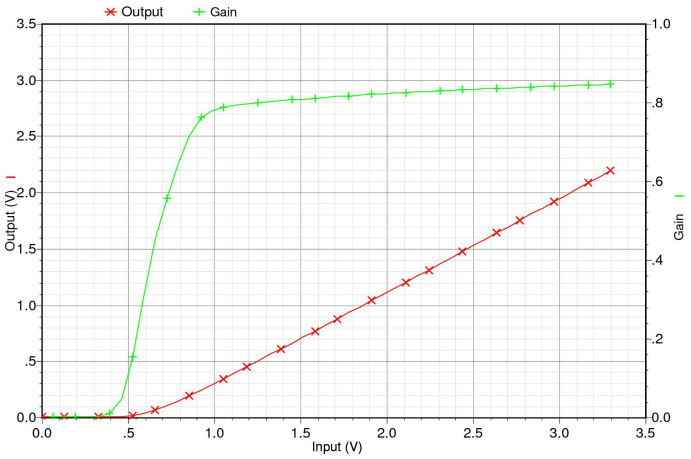


Figure 4.11: MLCv2 Source Follower Schematic

usual source follower design is a simple method of reducing the current used by the follower whilst keeping MN0, MN1, MN2, MP0 and MP1 the same size and at the same time ensuring no performance loss other than would be expected from the reduced current. With MN2 included in the design, there is a maximum flow of 1.7 μA through MN3 under typical device conditions compared to 3.5 μA when MN2 is not present. This design is less practical for real use due to the extra bias line and extra transistor required in the source follower. Although at the time of the design, it was hoped that by keeping all of the devices the same size, the random mismatch between them would be reduced, when it was possible to simulate this scenario (at the time the chip was being designed the tools to simulate process variation were unavailable), it was found that there was no significant advantage to either approach.



(a) Transfer Function

Figure 4.12: Source Follower Transfer Function and Gain

The transfer response and gain of the source follower are shown in Figure 4.12. As can be seen, the useful range of the follower where the gain is roughly linear extends from 3.3V to approximately 1V. The gain appears to remain linear to 780 mV, but is actually decreasing fairly rapidly at that point. Simulating the design over the process corners results in a change in the level shift, up by 200 mV and down by 150 mV from the typical mean value for the worse one and worse zero conditions respectively. This gives a total possible change in the offset voltage of 350 mV, whereas over a more realistic 1000 Monte Carlo runs of the same simulation the variation was only 200 mV. In both cases, the variation only really changes the

offset of the output; neither the gain nor the useful range are affected significantly.

The $-3\,\mathrm{dB}$ upper cut off of the follower is $50.8\,\mathrm{MHz}$ at the typical mean corner, with the minimum and maximum upper cut off frequencies (i.e. the worst speed and worst power respectively) at $24.0\,\mathrm{MHz}$ and $128.7\,\mathrm{MHz}$ respectively. The sharp edge when the photodiode recharges is the fastest event that the follower has to deal with. Although the edge is typically faster than the follower can pass unattenuated, it is not a critical factor in the design of the follower as we are not strictly interested in that edge. It does mean the size of the reset pulse should be chosen to take the worst case cut off into account; that is, the pulse should be no shorter than $42\,\mathrm{ns}$ to allow the follower to reach the correct voltage at the start of the discharge. The discharge curve is much slower, even under high light levels and the follower can easily pass it in its flat band response which extends to at least $2\,\mathrm{MHz}$.

It is pertinent to consider how much light it would take to exceed this limit. Assuming that the reference voltage is $1.5\,\mathrm{V}$ and the start voltage is $3.3\,\mathrm{V}$, the discharge must happen in $0.4\,\mu\mathrm{s}$ ($2.5\,\mathrm{MHz}$), if the photodiode is a square of $50\,\mathrm{x}\,50\,\mu\mathrm{m}$ with capacitance $225\,\mathrm{fF}$ then using the capacitor I/V characteristics:

$$\begin{aligned} i &= C\frac{dV}{dt} \\ &= 225\,\mathrm{fF} \cdot \frac{1.8}{0.4 * 10^{-6}} \\ &= 1.0125\,\mu\mathrm{A} \end{aligned}$$

This corresponds to a total of $3.375\,\mu\mathrm{W}$ of light striking the photodiode, assuming a quantum efficiency of 0.3. This is a substantial amount of light over such a small area. If the illumination was uniform, it would mean there was $1.35\,\mathrm{kW/m^2}$. As a comparison, according to the weather station in Stanwick, UK, a typical peak value of solar intensity for a day in September is around $400\,\mathrm{W/m^2}$ [Stanwick-Weather, 2007].

4.4.6 Final Pixel

The pixel layout is shown in Figure 4.13. The four transistors in the centre of the comparator are the differential pair transistors each split into two devices and laid out using a two dimensional common centroid technique.

4.5 Chip Two - PC2

4.5.1 Pixel

The second chip used the same basic design for the pixel as the first, but addressed the needs of a pixel array. The original layout was not optimised and so was not at all suited for tiling into an array. In addition to more strict layout constraints, designing a pixel for an array has other requirements.

As all of the pixel outputs cannot necessarily be accessed at once, some method of controlling which output is selected must also be included. This is usually achieved through the use of shared output lines for each column of pixels, with row select lines used to control which of the outputs is selected. For an array of $N \times M$ pixels, this results in N outputs at the bottom of the array that can be multiplexed to yet fewer lines if needed.

Due to pad limitations on the shared chip, the array was restricted to a 2×2 size. As with chip one there were two output pads (one for the photodiodes and one for the comparators) that were shared between the pixels, a pixel reset input, a comparator input and the row and column select inputs.

To allow the outputs to be shared among the pixels, the outputs from each pixel in a column were connected together through transmission gates. By ensuring that only one pixel in a column has its transmission gates closed at any given time, the

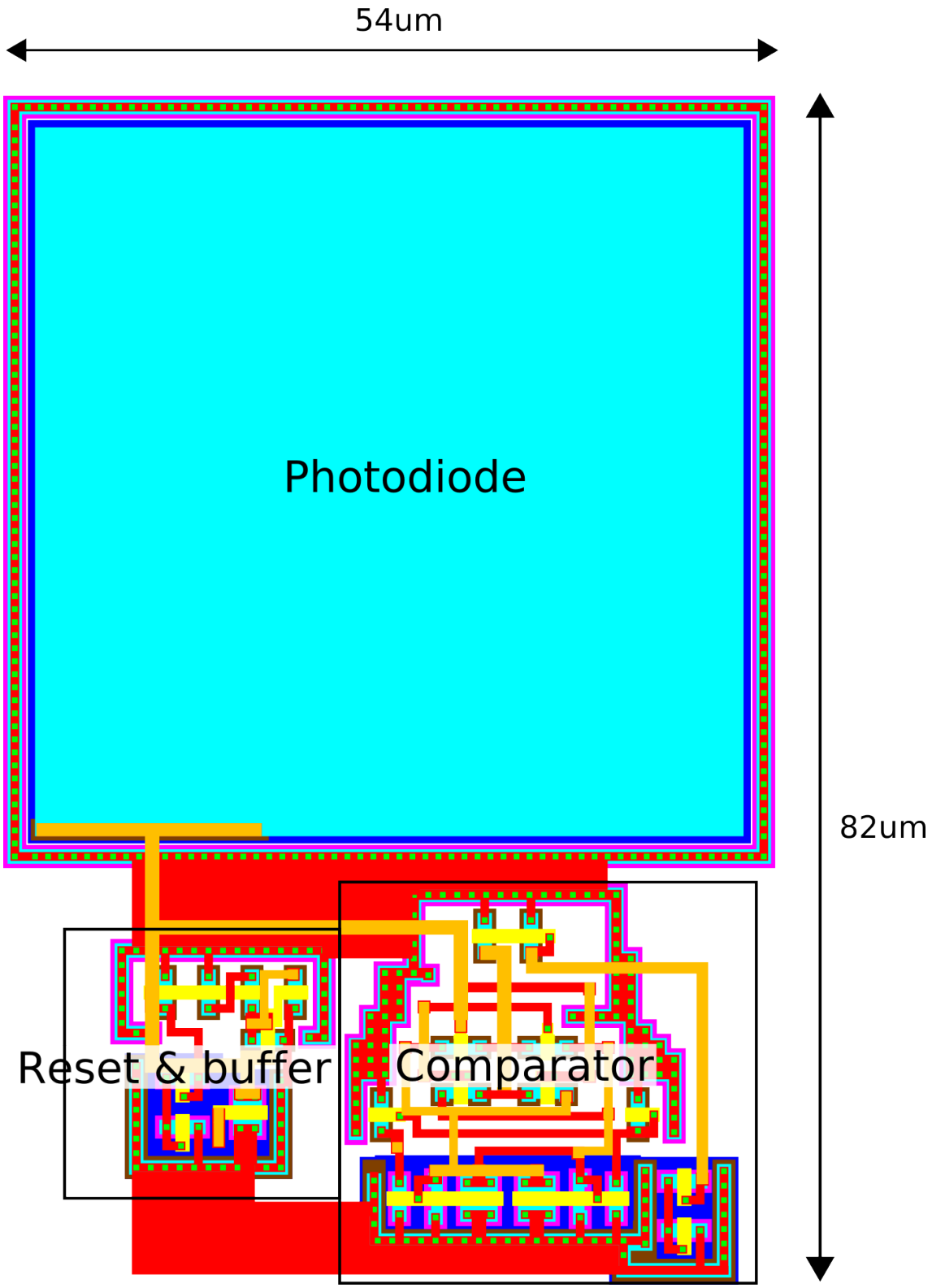


Figure 4.13: MLCv2 Pixel Layout (Chip 1)

outputs will not interfere with each other. The pixel schematic is shown in Figure 4.14.

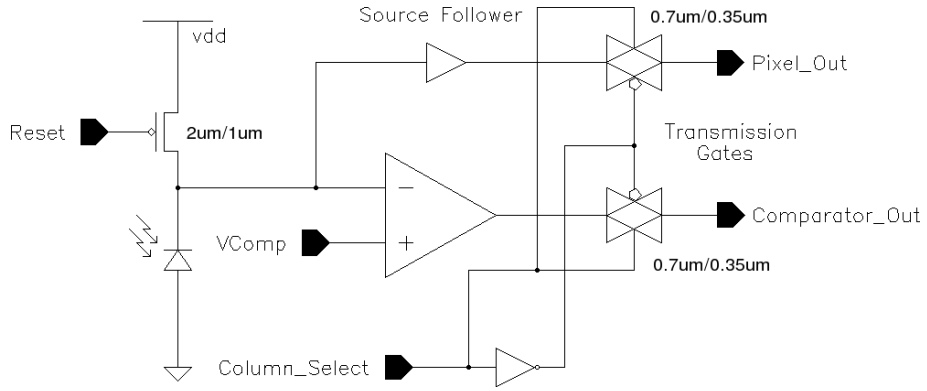


Figure 4.14: PC2 Pixel Schematic

The pixel layout is shown in Figure 4.15. The photodiode was chosen to be the same as in the first chip so a comparison could be made between the CSI and C35 processes. This choice highlights an important factor in pixel design namely that keeping the width and the height of the pixel the same is important for two dimensional arrays. In this case, by keeping the photodiode square, the pixel itself is no longer square due to the front end circuitry. With the x and y pitch being 54.90 μm and 69.25 μm respectively, this puts the pixel 26% taller than it is wide. The result of this difference in pitch makes image reconstruction harder since most display devices assume a square pixel.

4.5.2 Array

The pixel shown in Figure 4.15 tiles perfectly to form a larger array (that is to say that any number of pixels can be placed next to each other without requiring further work to join them together). The inputs in a column all connect together across the chip in the x direction and the outputs in a row all connect together in the y direction. Enabling the select line for a given column means that all of the pixels in that column have their outputs connected to the output bus whilst the rest of the columns are disconnected. The same principle is applied to the outputs that exit

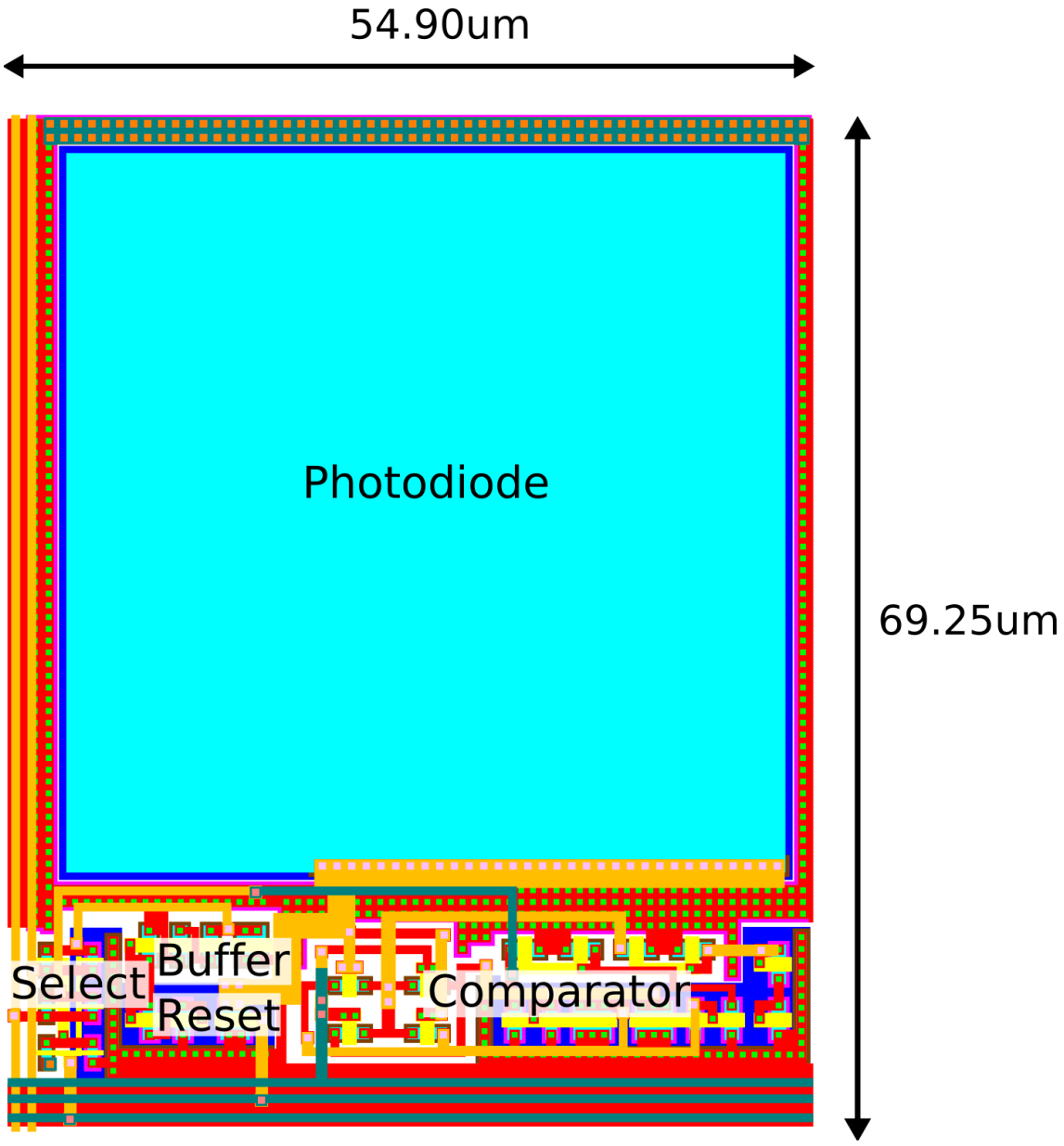


Figure 4.15: PC2 Single Pixel Layout (Chip 2)

the array as they are connected together through the row multiplexers to produce a single output that can go off chip. This is shown in Figure 4.16.

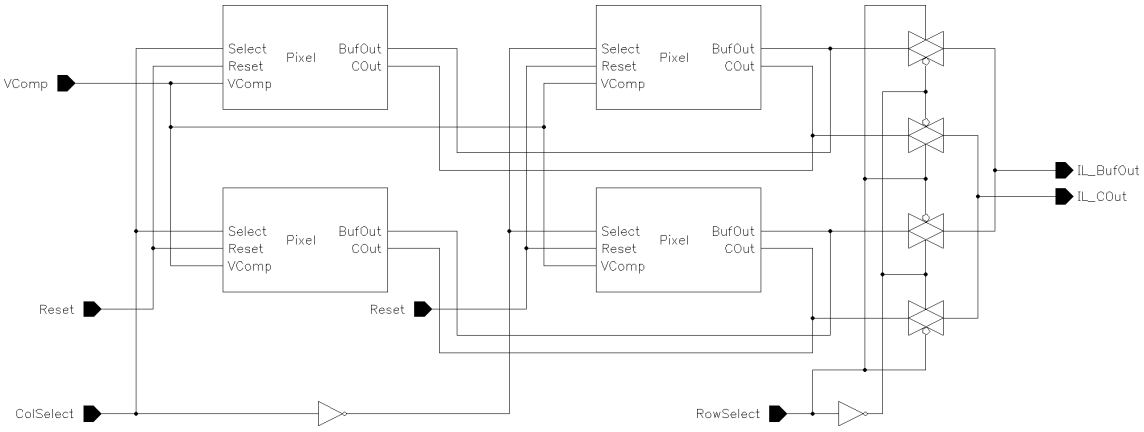


Figure 4.16: PC2 Pixel Array Schematic

As the array is just 2 x 2 the addressing is much less of an issue than it would otherwise be, since all that is needed to decode the two address lines is two inverters, with one each for the row and columns. The fact that the array size was restricted to such a degree by the pads available means that it suffers from similar problems to the poly-silicon resistors described in Section 4.4.4, namely that the pixels (and more importantly the photodiodes) are all in a different environment to each other. This means that results looking at pixel to pixel variation are less valuable than they might have been. The pixel array layout is shown in Figure 4.17.

4.6 Chip Three - PC3

The final prototype design consists solely of a comparator design. At this point, the possibility of using an octagonal photodiode was being considered (this will be discussed further in Chapter 5). By keeping the pitch of the pixels the same in both directions then the pixels will remain "square" in terms of reconstructing the image, even when using octagonal photodiodes. When octagons tile, gaps are formed that can be used for the frontend circuitry. However, this space is limited and to increase it means that the size of the photodiode must be quite dramatically increased. By

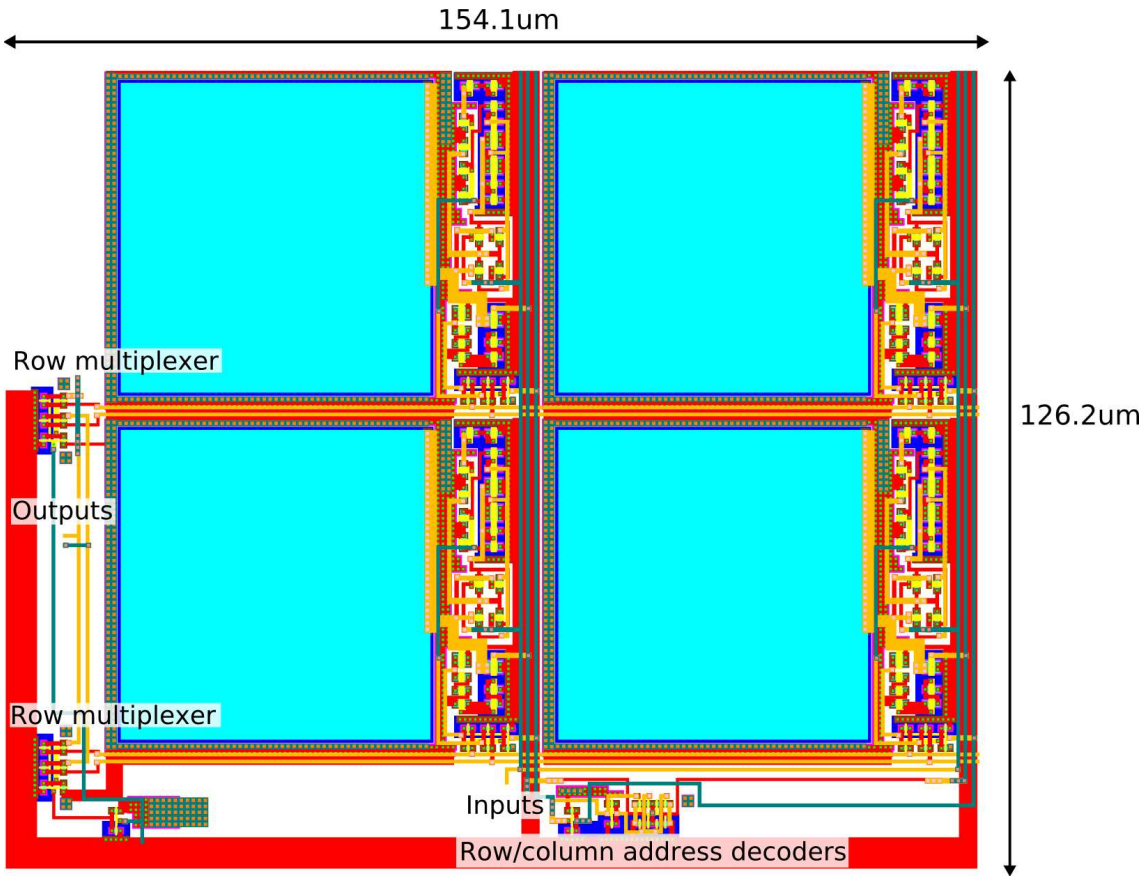


Figure 4.17: PC2 Pixel Array Layout (Chip 2)

making sacrifices in the layout of the front end circuitry, it can be made to fit in a smaller area. The biggest gain is in changing the common centroided transistor layout in the comparator to a plain layout, but this is also potentially the biggest sacrifice in terms of pixel-to-pixel variation, although as this variation is constant after the chip has been fabricated, it should be possible to calibrate the offsets out for each pixel. So chip three contained a pixel layout where the comparator layout within the pixel was created with the intention to minimise the space used rather than to make the most robust layout possible. As with chip two the number of pins available was limited, so only the comparator itself could be used rather than the whole pixel.

In order to make the design as close to a real pixel design as possible, the layout was created exactly as it would be if it were going to be a complete pixel rather than just a comparator. The remaining circuitry that was not in use would be turned permanently off to ensure it did not interfere with the operation of the comparator. The layout is shown in Figure 4.18.

4.7 Results

This section describes results obtained from the comparator on Chip three. The results for the pixels themselves have a great deal of overlap with the camera design (presented in Chapter 5), so to prevent any unnecessary duplication and to make comparison of results simpler, the pixel results are replaced by those from the camera (see Chapter 6).

The third chip manufactured differs from the others in that it only contains a comparator. The comparator was manufactured so that the chip to chip variation in the switching points could be measured in an attempt to obtain an idea of the kind of error that would be introduced. For this experiment, the non-inverting input is held at a reference voltage and the inverting input is gradually swept from ground

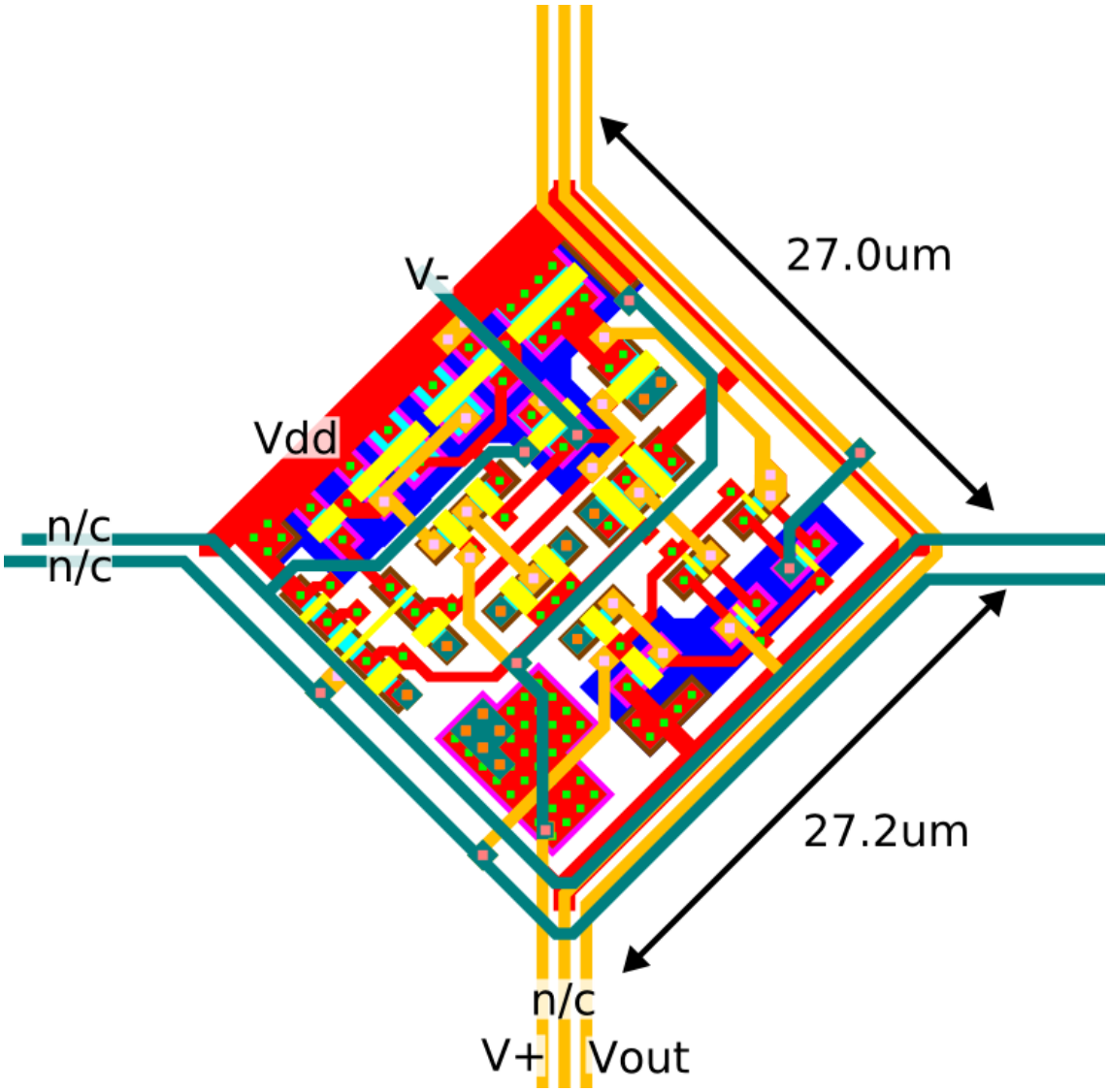
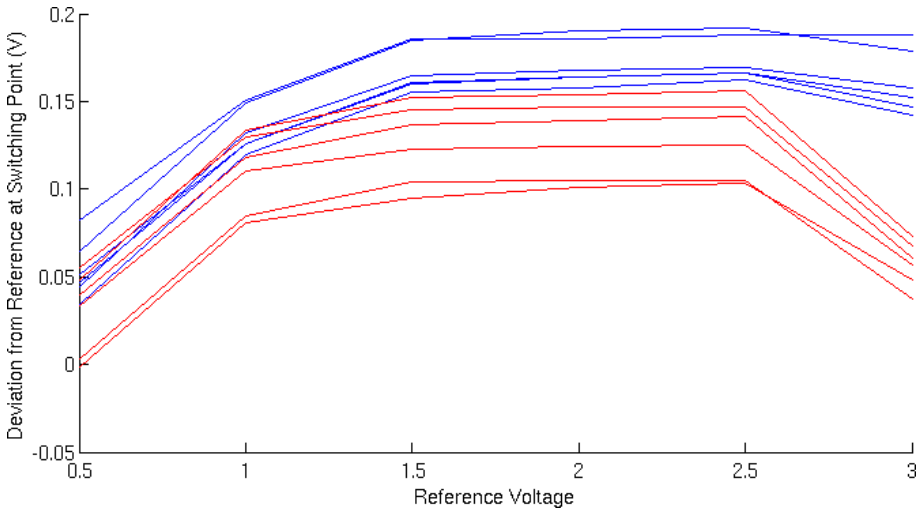


Figure 4.18: PC3 Comparator Layout (Chip 3)

to V_{DD} (i.e. a rising input) and vice-versa (a falling input) until the comparator switches. This setup mimics the way that the comparator is used in the pixel and camera designs and allows the effective offset for both rising and falling inputs to be measured and hence the amount of hysteresis to be calculated.

Figure 4.19 shows the difference between the inverting input and the non-inverting input for both the rising (blue) and the falling (red) cases for six chips, whilst Figure 4.20 shows the corresponding amount of hysteresis. The results for the falling input have had their sign made positive to make comparison easier. The curves follow the same shape as the simulated comparator and are reasonably flat in the middle of the voltage range but less so at the extremes where threshold voltages are likely to come into play. The shape of the curves are almost identical with only a small amount of fixed offset with the exception of a single chip that has noticeably different slopes at both ends of the voltage range. The same chip produces the slightly different results in both the rising and the falling cases and so is very obvious on the hysteresis plot. The variation of the falling input at 1.3 V covers a range of approximately 60 mV. Given a photodiode with capacitance 25 fF (very close to the size of the photodiode in Chapter 5) with 65 pA of photocurrent flowing, it would take 750 μs for the photodiode voltage to fall the 2 V to 1.3 V. The 60 mV variation across the chips means a 30 μs variation in the time taken for the comparator to trigger, or a variation of 1200 in a counter using a 40 MHz clock, compared to the total count value of 30,000.

Comparing the falling and the rising results it can be seen that the spread of values when the inverting input is falling is roughly twice that for when it is rising. This is a disappointing result because when the comparator is in use with a pixel it will be seeing a falling input. It is also rather unexpected because Monté Carlo simulations of the circuit show that the two results are roughly the same. The important part to note with the results is that they remain constant for the different devices, so although there is a reasonable amount of variation it will simply produce a systematic error that will manifest itself as a component of fixed pattern noise and



(a) Rising Input

Figure 4.19: Comparator Voltage Offsets
(Blue: Rising input, Red: Falling input)

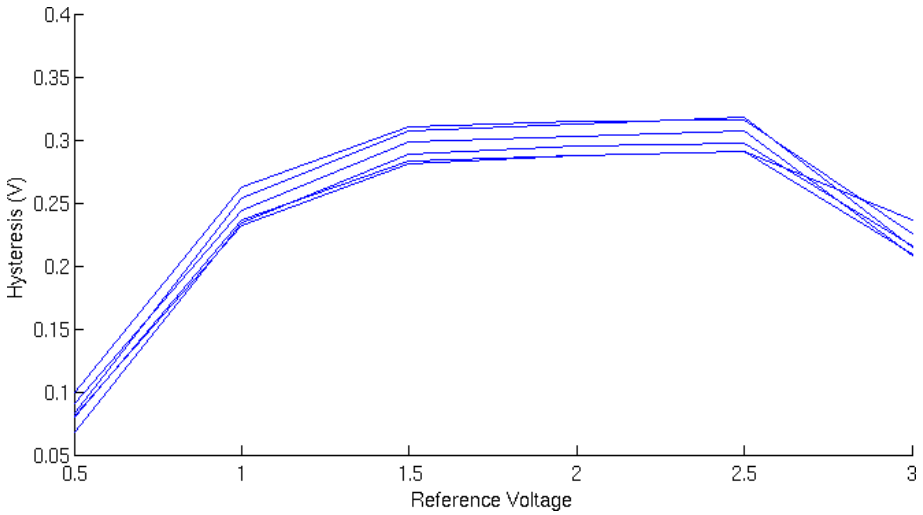


Figure 4.20: Comparator Hysteresis

can be calibrated out. Figure 4.21 shows the output of a single comparator with a triangle wave input at 2 MHz, 3.3 MHz and 4.5 MHz. This gives a pessimistic view of the transient operation because it would never normally drive such a large load.

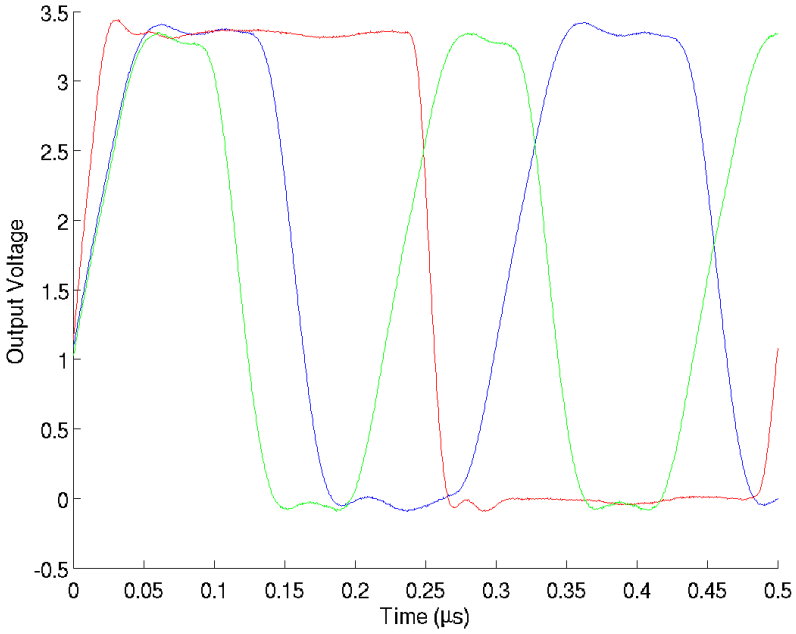


Figure 4.21: Comparator Transient Results
(Red: 2 MHz input, Blue: 3.3 MHz input, Green: 4.5 MHz input)

4.8 Summary

This chapter has introduced the design of the integrating pixel used to measure the time taken for the voltage across a photodiode to discharge to a certain value. The structure and the different components used within the pixel have been described along with the reasons for the way the pixel has been designed. The designs of two prototype pixels manufactured on two separate chips have been shown, with the first chip containing a single pixel and the second chip containing a small 2 x 2 array of completely tileable pixel.

A third chip was manufactured using the same essential comparator design as had been employed in the second chip, but using more relaxed layout techniques that

minimise the amount of space required whilst introducing the possibility of greater voltage offset differences from comparator to comparator. Experimental results of the comparator show that it has hysteresis in the 250 mV to 320 mV range which is more than sufficient for this application and that there is at least a 60 mV variation in the trigger point of the comparator for a falling input.

Chapter 5

Camera Design

5.1 Introduction

Having documented the potential of the integrating lock-in technique, the final chip designed was a large array utilising this approach. Two possibilities were considered for the camera, either including all of the required digital processing on chip with the array, or else having just the pixel array on chip with the processing performed off chip. Both approaches have advantages and disadvantages.

Having the digital processing on chip would mean that the sensor is more complete, possibly with a connection straight to a computer through a universal serial bus (USB) transceiver which would mean a camera could be very compact and easy to use. It has already been mentioned that an advantage of cameras specially designed for modulated light measurement is that the data rate off chip can be greatly reduced compared to using "ordinary" cameras and keeping all of the processing on chip is a major reason why this is true.

Assuming that the chip is a fixed size regardless of whether the digital circuitry is included or not, then a chip without digital processing will have more space available for the actual array, meaning either bigger or more pixels. Processing off chip also

means greater flexibility since it can be changed without having to re manufacture the chips (which would take at least three months), meaning that different unforeseen applications could be implemented. A disadvantage of this approach is that it is much more complicated to create a fully working camera because of the need to join together the chip and the processing component.

Regardless of where the processing is carried out, it makes sense for it to be implemented using a hardware description language such as VHDL. VHDL can be used on FPGA chips such as the Xilinx Spartan 3 that will be introduced in Section 5.9 but it is also possible to synthesis VHDL and use it on a full custom design. This means that prototyping can be carried out with the processing off chip on the FPGA and then if a camera is created with the processing on board then the same VHDL code can be used.

5.2 Camera Architecture

The essential components of the pixel are shown in Figure 5.1, with the dotted lines showing the boundaries at which the design could be partitioned - that is to say, which blocks are included in the pixel.

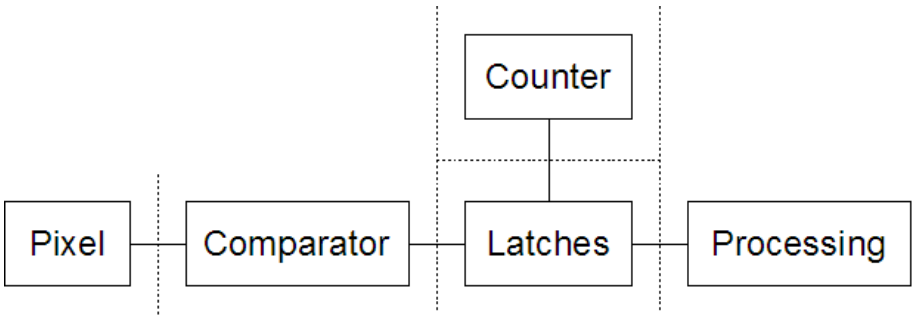


Figure 5.1: Pixel Architecture

In a traditional integrating APS, a shutter switch is often used within the pixel to allow all pixels to sample the image at the same time whilst still being able to be read their values in a serial fashion. This is not possible in the setup shown in Figure

5.1 since it is the time itself that provides a measure of the light intensity. However this does not mean that sampling all pixels at once is impossible when measuring the time; as discussed in Chapter 2, Kleinfelder's design latches a counter value for each pixel that gives a measure of the light intensity into every pixel so that all pixels can be sampled at the same point [Kleinfelder et al., 2001]. In this case to ensure a correct value is obtained for all pixels their resets must happen simultaneously so that the photodiode discharges all start from the same point. This contrasts with the requirement when measuring the photodiode voltage, where both the reset and the shutter or direct sample must be made at the same time for every pixel.

Where a camera is only taking static images, storing the pixel value in an n-bit latch within the pixel is convenient, but where additional processing must be carried out on the resultant values then these must either be read from the pixel extremely quickly or else the processing must happen at the pixel. Processing carried out at the pixel obviously takes up extra silicon area that, even if adequate, will reduce the fill factor considerably. For example, the digital low-pass filter used as part of the processing in a modulated light camera designed for measuring blood flow [Kongsavatsak, 2005] uses approximately 150 gates and is not even all of the digital processing involved. This type of design is much too large to be consistent with what would generally be considered an acceptable fill factor.

In the case where more processing is required than can be reasonably accommodated at the pixel, the design must be partitioned with choices made regarding which parts of the design remain within the pixel and which are moved. Such a type of camera is a "column based" design where a single row of pixels is accessed by reading its outputs at the bottom of each column. The typical arrangement for a column based camera is shown in Figure 5.2.

The outputs from all of the pixels in a column are connected together, with the pixel outputs that are not currently selected being held in a high-impedance state. The row address decoder is used to select a single row to be active. The outputs from

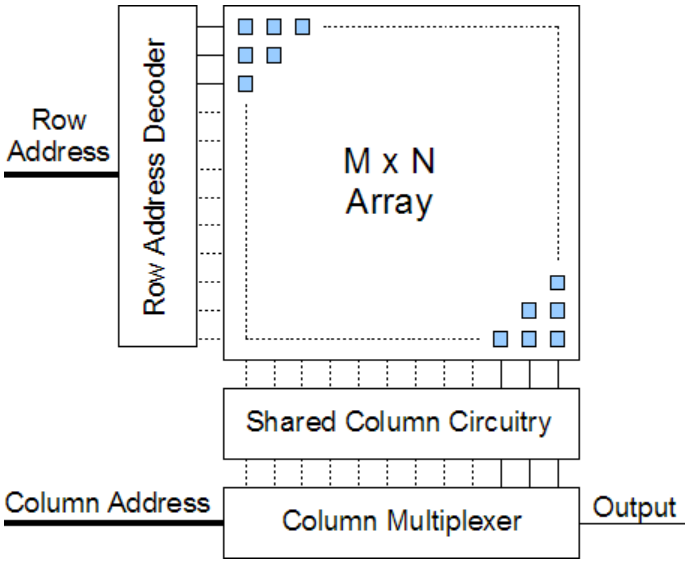


Figure 5.2: Column Based Camera Architecture

the shared column circuitry are multiplexed out in the same way using a column address decoder and multiplexer. Figure 5.2 shows this approach giving a single output, but more outputs are possible if required. This arrangement is exactly the same with pixel based designs, except that then all processing is performed at the pixel so that the block labelled "Shared Column Circuitry" is not needed.

Assuming that all of the blocks shown in Figure 5.1 cannot be included in the pixel, the processing block is the first candidate for removal as it is likely to be not only the largest but also the furthest from the photodiode. With the counter local to the pixel the latches used to store the counter value become redundant and can be removed which means that the overhead of including the counter at the pixel is not as great as it might first appear. Both the counter and the latches, although substantially smaller than any processing circuitry, are likely to take up a sizable area; not only the devices themselves, but also the signal routing wires. Every "bit" of the counter must have a corresponding wire as well as additional switches to allow these to be shared between pixels. This adds further to the area required and so reduces the fill factor of the pixel. The area required is of course less of a problem with shrinking process sizes, especially where greater numbers of metal layers are available for signal routing, but it does remain unfeasible in the larger sub-micron

processes such as 0.35 μm.

Splitting the design between the comparator and the latches means that each pixel has only a single output line. Although it has been assumed that there is only one column output for each output wire that leaves the pixel, it is also possible to have more than one column output for each pixel output. A simple example would be to effectively split the columns in half with shared circuitry at the top and bottom of the array. This leads to three possibilities:

1. Having a wire descending the column for each pixel output (N outputs)
2. Joining the outputs of all of the pixels together and having a single wire descending the column (1 output)
3. Joining a group of pixel outputs together to have a number of wires descending the column (x outputs, where $1 < x < N$).

Having a wire coming out of the array for each individual pixel requires as many output wires as there are pixels in the column, which is not scalable it terms of both space and simplicity of design. Any change in the size of the array requires the pixel layout to be redesigned which is extremely undesirable. Furthermore, this approach is simply translating the space requirements to outside of the pixel array since the processing will still need to consume the same area, just not directly within the pixel. In some cases this may be acceptable to keep the fill factor of the pixels high. Having a single shared output for the entire column (case 2), means that the design is scalable and all pixels can be identical which is an advantage over the previous case. The big disadvantage is that processing an entire frame at once is no longer possible, although this disadvantage is countered by a large reduction in the amount of space required for the processing (for N rows the amount of processing drops by a factor of *N-1)*. The final possibility is a compromise between the two extremes by having a fixed number of outputs in a column that is less than the number of pixels

in each column. Changing the number of these shared outputs then allows a choice to be made between either an increased frame rate or a smaller fill factor with more silicon required for the processing, whilst still utilising identical pixels across the array.

In both of the latter two configurations, the comparators in the pixels not being addressed still draw current. The front end design as used on chip 3 was combined with an octagonal photodiode of area 1919 μm^2 to form a pixel (see Figure 5.3). The pixel achieved a fill factor of 65% which appears respectable until it is noted that the photodiode is fairly large, which is necessary because otherwise the free space between the diodes is insufficient to contain the front end. If the comparator layout was designed with more robust layout techniques, the front end circuitry would take up more space and so the photodiode size would have to be increased further to enable this format to be employed.

The answer to all of these problems is to take the comparator out of the pixel, which is the final option for partitioning. This instantly frees up substantial silicon area and reduces power requirements whilst also providing a number of benefits with regards to the comparator implementation. As the comparator is now outside of the pixel, more space can be used meaning that the more "area hungry" layout techniques that bring benefits to the design can be used, such as common centroiding.

In addition to allowing the comparator implementation itself to be improved, by placing the comparator outside of the pixel there will be fewer comparators overall. In general, increasing the number of shared devices helps reduce the amount of process variation (and hence fixed pattern noise) as there are simply fewer devices to vary. Having a single comparator for each column means that any offsets in that comparator are the same for every pixel in that column, making compensation for such differences simpler (although the logical progression of this argument is to have just a single comparator shared between all pixels, this would mean the maximum frame rate possible being severely reduced).

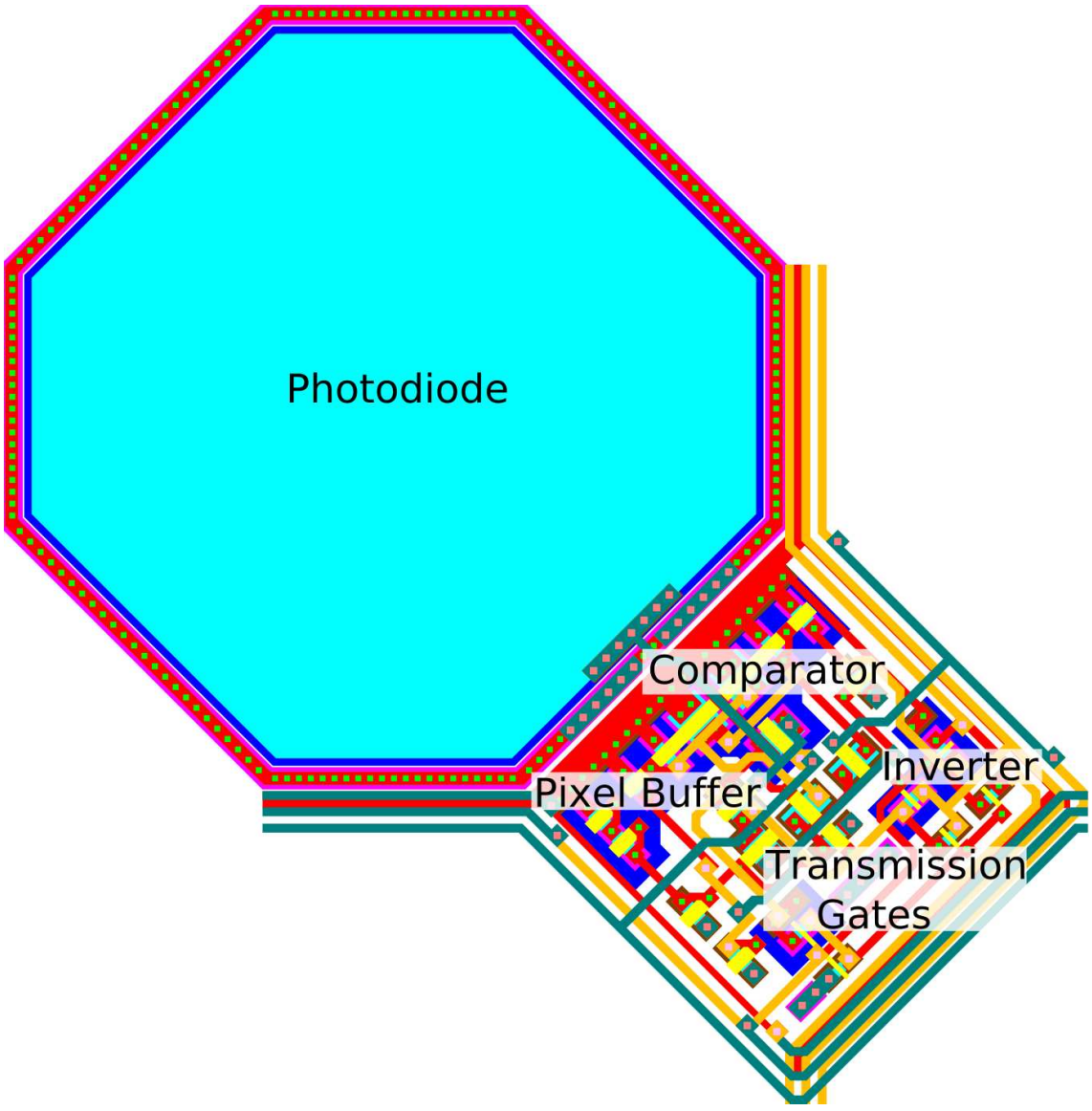


Figure 5.3: Example Tileable Pixel Layout Including Comparator

With the comparator inside the pixel, the photodiode output could be connected directly to the comparator. With the comparator now moved out of the pixel, this is no longer desirable as the capacitive load placed on the photodiode would be considerable (as with the PPS introduced in Chapter 1), meaning that its overall speed would be reduced and the amount of voltage discharge observed across it would also be reduced for a given amount of light, thereby decreasing the sensitivity of the detector (although increasing the range of different light intensities that could be measured). Adding a buffer to the pixel in the same manner as in the traditional APS (and as in the secondary output from the pixel in the first two chips) means that the capacitance at the photodiode is dominated by the photodiode capacitance, with some smaller contributions coming from the wiring and gate capacitance feeding the local pixel buffer.

Now the size of the array must be considered. The array size is most limited by the area required (both in terms of the cost to manufacture and the increased chance of defects in larger chips), the frame rate possible and the power requirements. Each doubling of the number of pixels in the array has an associated doubling in cost, and in power requirements and a halving of the frame rate. In this case, the most significant limitation is the area taken by the array as this correlates directly to the cost of the device.

In order to make the most efficient use of the address pins used to select which pixel should be read, the array size should be chosen so that the number of rows and the number of columns are a power of 2. It is possible to produce recognisable images with arrays of size 16 x 16 and larger. In the fabrication runs available using the AMS C35 process, the minimum size chip available has an area of 10 mm^2 and the desire was to keep the size of the chip reasonably close to this figure to minimise cost. Given that a square chip of 10 mm^2 has sides that are 3160 μm (rounded down to the closest 5 μm based on the manufacturing design rules) and that a border of 300 μm is required around the border of the chip to allow space for pads, this leaves 2560 μm for the actual design. Ignoring the fact that space is required for signal

and power routing, as well as any other circuitry outside of the array, the pixel sizes for different array sizes (assuming the chip is just the pixel array), are 80 μm, 40 μm and 20 μm for 32 x 32, 64 x 64 and 128 x 128 sized arrays respectively. Given these figures and the desire to build as large an array as possible within budget, pixels of around 20 μm are the best choice. This gives the possibility of either creating a 128 x 128 array with no on board processing or a 128 x 64 or 64 x 64 array with on board processing, depending on the amount of space required for the digital processing.

Given the desire to both build as large an array as possible, and for the design to be as flexible as possible, the decision was made to place the processing off chip and focus the on chip design on the array itself.

5.3 Front End and Photodiode

With the comparator placed outside of the pixel, the front end circuit consists only of the reset transistor, a buffer and a select switch. As the photodiode must be made relatively small compared to the previous designs to allow a larger array to be created, the space taken up by the front end circuit must be optimised as far as is possible.

5.3.1 Source Follower and Select Switch

In contrast to chips 1 and 2, the comparator is outside of the pixel so there is only a single output, rather than both the comparator and photodiode output. In the drive to reduce the number of components residing at the pixel, both the source follower buffer and the select switch can be modified to be smaller. In the previous two designs the entire source follower was inside the pixel, but this is not necessary since one of the two transistors in the source follower can be placed at the bottom

of the column and shared among all of the pixels in that column. This approach has a number of advantages, with the most obvious being that the amount of space required within the pixel is reduced. Equally importantly, by sharing a transistor in this way the fixed pattern noise along the column is reduced.

Using a complete transmission gate for the output selection switch as in chip 2 is largely unnecessary with only a buffered pixel output. With a comparator output to contend with, the switch must be able to pass all voltages between ground and V_{DD} equally well, which requires a transmission gate. As the only signal now comes from the buffered photodiode, a similar principles applies to the select switch as does to the reset switch in that a reduced range of voltages must be considered. The output buffer introduces a voltage shift away from the rail, so an NMOS transistor is the correct choice. This situation also allows better integration as the source follower transistor is also NMOS, meaning that they can be placed much nearer to each other than would be possible with mixed transistors.

5.3.2 Photodiode

As was noted in Section 4.5, square photodiodes can be problematic in arrays. The solution proposed there was to use a rectangular photodiode or other shapes such as an "L" which can also be used to keep the pitch of the overall pixel the same in both directions yet still be able to integrate the control circuitry. Another approach is to make use of 45° edges when creating the photodiode and so make an octagonal photodiode. Edges drawn at 45° are commonly the finest angle allowable in IC design and this is the case for the C35 process. Tiling octagons together in an array naturally produces areas that are not in use and so can be filled with the front end circuitry. Using octagonal photodiodes can also lead to improved fill factor (the square photodiode layout in Pitter's work [Pitter et al., 2004] was implemented by this author and was subsequently changed to an octagonal photodiode with the same pitch for another camera, whilst including an additional transistor and still

managing to increase the fill factor from 54.9% to 60.8%). There are other small benefits as well. The electric field induced by the charge on a photodiode becomes intensified at the corners of the device. The sharper the angle, the more intense the field gradient and this contributes to a larger leakage current (i.e. dark current). By widening the internal angles the field intensity is reduced and so the dark current can also be reduced (taking this to the extreme, a circular device should have the lowest leakage current). A similar but smaller effect is also gained due to the reduced length of the periphery of the device for a given area compared to a square device (again, a circular device would be most efficient in this regard). This advantage affects both the leakage current and the capacitance of the photodiode, although the effect is quite small as will be demonstrated in Section 5.3.3. It is worth noting that although a circular photodiode seems like a better choice and that a reasonable approximation of a circle can be manufactured, it would quite probably be a worse choice for an array because of the increased area that was neither light sensitive nor could be used for circuitry. It is extremely likely that the reduced fill factor would outweigh any advantages offered by a smaller leakage current.

5.3.3 Final Pixel

The pixel as shown in Figure 5.4 has a pitch of 20 μm in both directions, which was chosen to accomodate an array 128 pixels square within a reasonable chip area. This resulted in a photodiode area of 221.7 μm^2 and a total area of 397.6 μm^2, giving a fill factor of 55.8%. The photodiode has a perimeter of 54.4 μm, if it was a square device with the same area it would be 14.9 μm x 14.9 μm and have a perimeter of 59.6 μm. The octagonal photodiode has a total capacitance of 24.47 fF and a leakage current of 1.03 fA, the corresponding square photodiode would have a total capacitance of 26.02 fF and a leakage current of 1.07 fA, all calculated at a reverse bias of 3.3 V [AMS, 2004]. These values do not consider the effect of changing the angle of the corners on the leakage current.

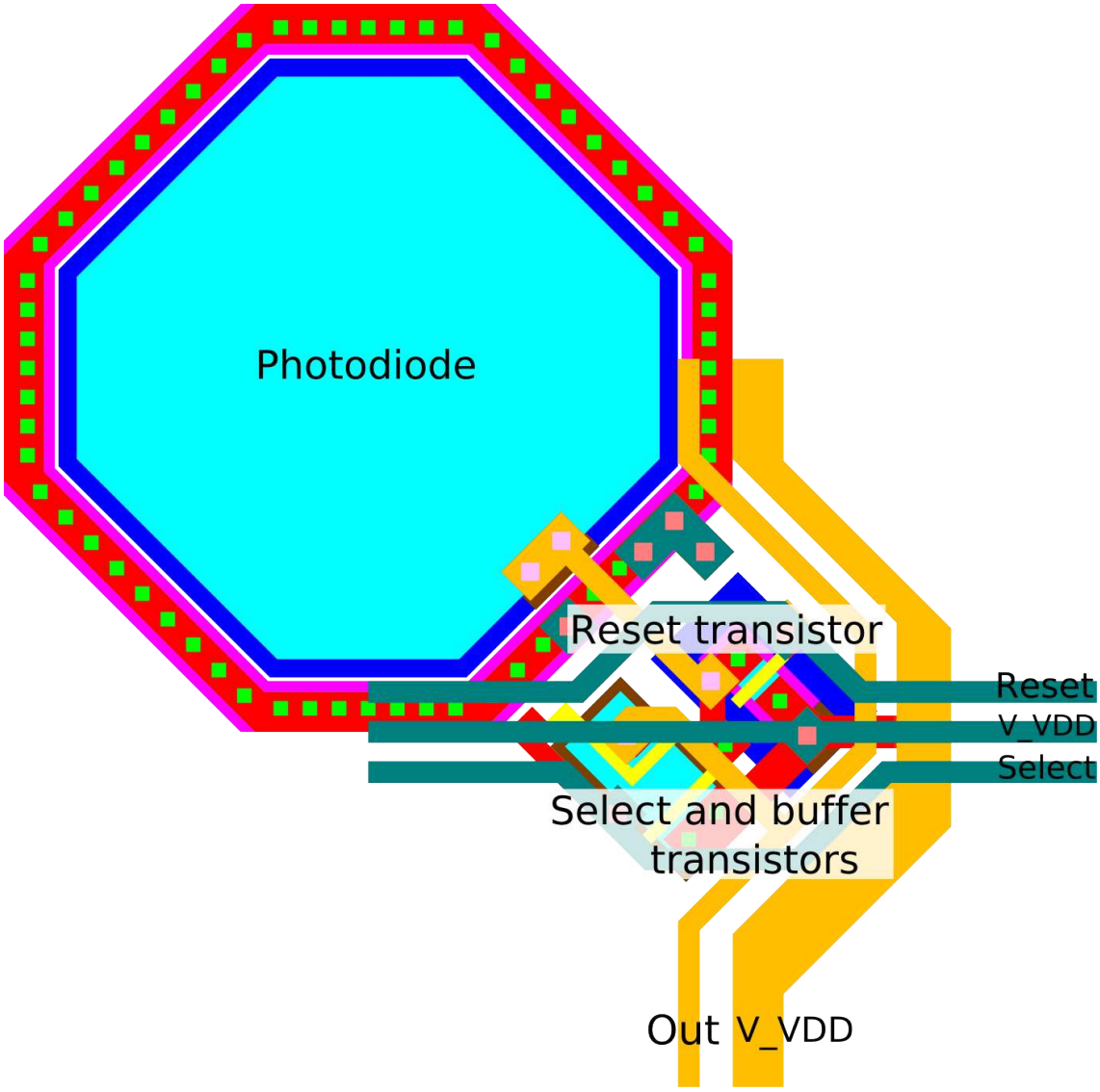


Figure 5.4: Final Pixel Layout

The periphery of the pixel has V_{DD} wires across it both vertically and horizontally to ensure good supply distribution across the array. The reset and select lines run horizontally across the array, which means that a whole row is selected at once and also that the reset lines for the rows can be controlled individually which is an advantage as will be discussed in Section 5.4.

5.4 Array Row Address and Reset Decoders

Having designed the pixel and by extension the array, the next component to consider is the row address decoders. These are used to take a 7-bit digital input in the form of a binary number and convert it into a 128-bit output that ensures only a single row is enabled at once. The reset line for the pixels also lies horizontally across the array in parallel with the select line, so that each row of pixels is reset at the same time. This also allows some control over the reset lines for the rows that are not currently in use by using the address decoders to help control the reset for rows that are not in use. This is very important because if the pixels are left integrating then they will eventually lose all of their charge and then become forward biased. As this happens the electrons are no longer efficiently collected by the photodiode junction and will be injected into the substrate, where they may travel for up to a diffusion length ($300\,\mu$m to $10,000\,\mu$m) [Pui, 2004] and manifest themselves as cross talk at a currently active pixel or disturb the other circuitry in some way. This is also known as "blooming" and is most likely to occur where there are very bright regions in an image. To reduce the possibility of blooming, the pixels that are not in use are held in permanent reset and are so kept reverse biased. The schematic showing the combined row address and the reset decoding is shown in Figure 5.5, where the logic cells used are those provided by AMS as standard cells.

If A0 to A6 are all "high", then the row is selected and ResetIn is passed through to ResetOut unaffected. If the pixel is not selected then ResetOut is "high" regardless of

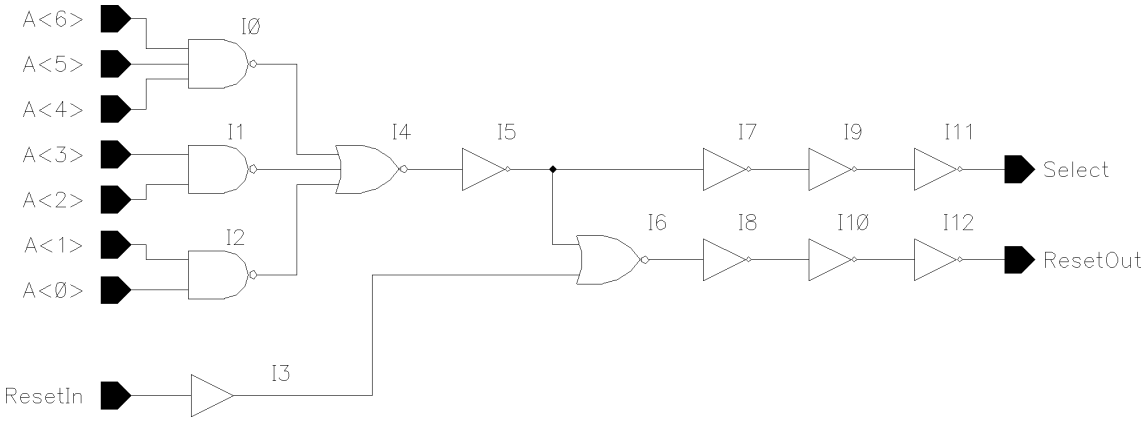


Figure 5.5: Single Row Address and Reset Decoder

the value of ResetIn. The chain of inverters on both of the outputs have increasing drive strength so that a high load can be driven with the minimum delay. The schematic from Figure 5.5 is duplicated for each row, with inverters placed locally to each of them, connected to A0 - A6 as necessary to allow each address to be selected individually. The layout for the address decoder was designed as a parameterised cell (pcell), in other words a cell that can be modified by changing parameters associated with it. A straightforward example of a pcell is a MOS transistor where the width and the length of the gate can be set in the layout directly. For the address decoder, two pcells were created. The first (address_decoder7_element) was for each individual row and allowed the designer to select whether each of the input A0 - A6 should be inverted or not (which allows the address that a row corresponds to to be set) and the height of the row. The second pcell (address_decoder7) automatically creates the entire address decoder which consists of 128 instances of address_decoder7_element with the specified spacing between rows as well as setting the addresses on all 128 instances correctly. This approach helps make more robust designs as the possibility of human error is greatly reduced. In all, address decoders for 2-bit to 7-bit inputs were designed in this manner, with the variable height allowing them to be used in many different designs. These address decoders have now been used in ten other designs on six chips in the School of Electrical and Electronic Engineering at the University of Nottingham [Dmochowski, 2006, He, 2008, Stockford et al., 2007].

5.5 Comparator Design

5.5.1 Final Design

The architecture described in Section 4.3.3 was again used for the design of the final comparator. The important considerations were the slew rate, the propagation delay, the amount of hysteresis and the total current consumption. These factors can be controlled in the actual design, but are all also directly affected by the bias current. As the comparator would only be driving a nearby logic gate, the load would be relatively small. Based on pessimistic test layouts, a load capacitance of 50 fF seemed a reasonable value. In the final layout design, the parasitic capacitances loading the two comparators in the column were 12 fF and 26 fF.

With two comparators per pixel column and 128 columns, even a small change in the current consumption of the comparator produces a significant change in the overall usage. Minimum current consumption was therefore an important part of the design goal and as decreasing the bias current is most likely to have the largest negative impact on the speed of the comparator, this was the main focus. Applying a rail-to-rail triangle wave to the inverting input and keeping the non-inverting input tied to half-rail gives a good method of evaluating the speed performance by observing the shape and parameters of the resulting square wave. The propagation delay from the input crossing the reference voltage to the output switching and the slope of the output during switching are the important factors here; in particular the propagation delay for a rising input effectively defines the amount of time that the pixel must remain reset before integration can start and equally during integration, the pixel output must remain below the reference voltage long enough for the comparator to trigger before being reset again.

Using the simulation setup described, with the input changing from 3.3 V to 0 V in just 1 ns to measure the limits of performance, the rise time of the output from 10% to 90% of the rail was measured for bias currents ranging from 2 μA to 100 μA. The

response is shown in Figure 5.6 and is approximately a 1/x response which shows the expense of aiming for sharper edges.

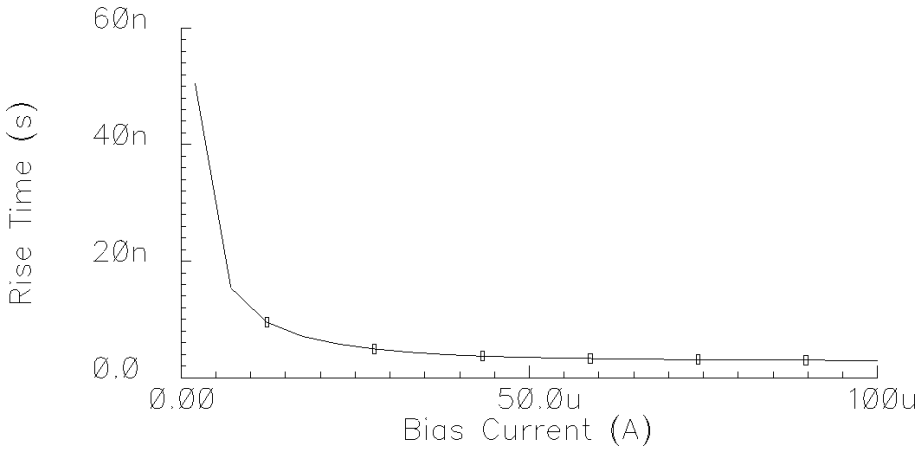


Figure 5.6: Comparator Output Rise Time

The rise time gives some quantitative data; a more qualitative method of evaluating the speed performance is to look at the shape of the output square wave. What this effectively means is setting a limit on the percentage of the period that is taken up by the rise/fall part of the wave. A pure square wave has infinitely short rise and fall times but is a theoretical construct. Deciding, even as an estimate, the percentage of the wave to be taken up by the rise and fall, allows a guide to be obtained on the maximum practical speed attainable. Figure 5.7 shows the comparator output for a bias current of 20 μA at three different frequencies; 1 MHz, 10 MHz and 25 MHz. The response at 1 MHz is very sharp and at 10 MHz still quite acceptable, but at 25 MHz the output barely manages to reach the rail before it begins to fall again.

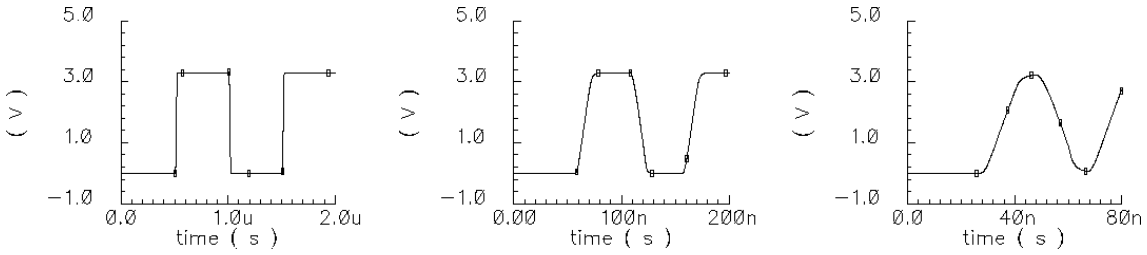


Figure 5.7: Comparator Transient Frequency Comparison

The overall current consumption for a single comparator for the same range of bias currents is shown in Figure 5.8, which shows that it remains linear with the bias current beneath 40 μA and is actually very slightly more than twice the bias

current. The simulations carried out suggest that a bias current of 10 μA provides a reasonable compromise between current consumption and speed which leads to a current draw of 23.7 μA. All 256 comparators combined would use a total of 6.07 mA and have a rise time of approximately 12 ns. If the counter that the comparator is connected to has a clock at 40 MHz then it has a period of 20 ns, so the uncertainty in the value used at the accumulator will be a single bit.

In practice, since there are 120 pins on the package used for the chip, only 64 columns can be accessed simultaneously (again, using a power of two gives the most efficient addressing). Given this situation, half of the comparators could be disabled when they are not in use and so halve the current consumption. Although this is a reasonably attractive proposition especially since the comparators are some of the major consumers of current on the chip, a current draw of 6.07 mA is still relatively low for a chip so all comparators can be left in use to reduce the complexity of the chip. The bias for the comparators is connected to an external pin on the chip so that it can be tuned as desired when being used.

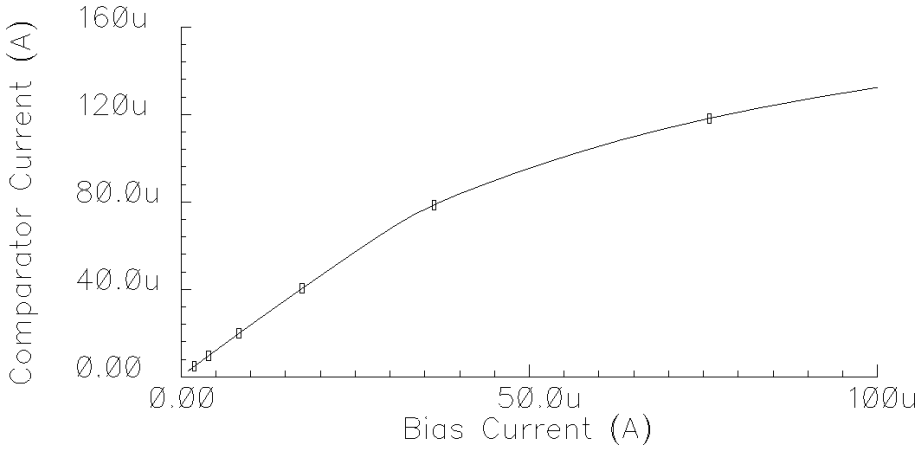


Figure 5.8: Comparator Current Consumption

The next factor to consider is the hysteresis. The ratio of the widths of the two pairs of load transistors MP1, MP4 and MP2, MP3 sets the amount of hysteresis, if the widths are equal then no hysteresis is observed. MP1 and MP4 were chosen to be 0.8 μm and MP2 and MP3 to be 2.6 μm which according to simulation leads to a hysteresis of 198 mV. This level of hysteresis is more than sufficient to protect against

glitches and noise on the input. It compares well to commercially available devices. For example, the ADCMP608 provided by Analog Devices has a programmable hysteresis with a maximum of 160 mV, with other devices often offering much lower protection.

The DC switching operation of the comparator was simulated to obtain a transfer function of the switching voltage, with the output measured at 1.65 V, for different reference voltages. The result of this simulation is shown in Figure 5.9, with results for the cases when the input is rising (horizontal bar) and falling (vertical bar).

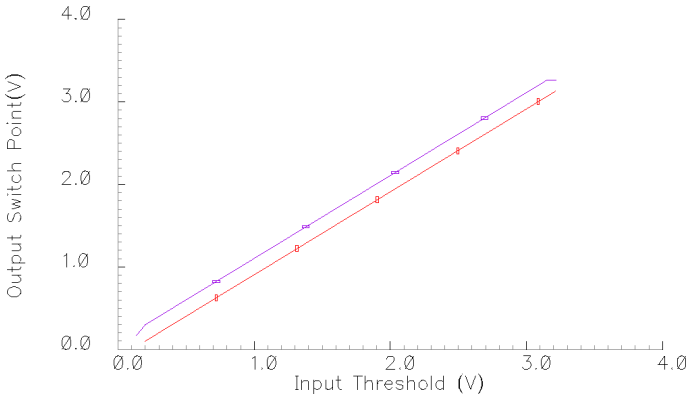


Figure 5.9: Comparator Switch Point Simulation
Rising input: Horizontal bar
Falling input: Vertical bar

In both cases the output is linear, although the rising edge case does exhibit some distortion at the extremes. Both of the outputs also have an offset associated with them. This is not an input offset voltage as usually described because it does not result from the slight variations in transistor parameters when they have been manufactured. Instead, the offset is a direct and fairly obvious result of the hysteresis. The gap between the two lines in Figure 5.9 is the hysteresis. Setting the transistors MP1-MP4 from Figure 4.3.3 to all be the same size removes the hysteresis and the response for both rising and falling inputs are the same and with no offset.

As the comparator only monitors the photodiode discharge, then only the response to a falling input is of interest. The response is given by Equation 5.1 (taken from the simulation results):

$$V_{Switch} = 1.005 * V_{Reference} - 0.099 \tag{5.1}$$

This means that in the context of the whole system, the comparator will trigger at a lower voltage than desired and hence at a later time and higher count value. As the same situation is true for every comparator and every result there should be no problems associated with it except for the systematic offset.

The schematic of the comparator is shown in Figure 5.5.1 and the corresponding layout in Figure 5.5.1 where the top metal light shield has been removed for clarity. The two input transistors have been split into two parts to allow common centroid layout for better matching.

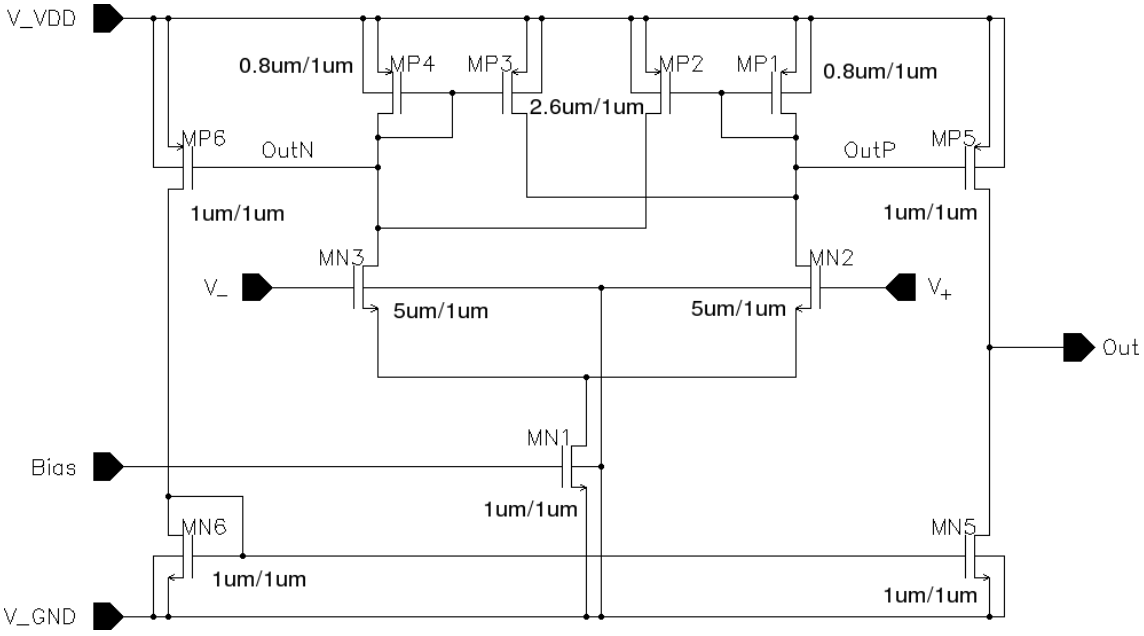


Figure 5.10: Final Comparator Schematic

5.6 Comparator Output Multiplexer

Chapter 1 introduced the idea of correlated double sampling (CDS), that is, measuring the photodiode voltage once at the start of the discharge and once at the end. In the context of measuring time rather than voltage, this is most easily done

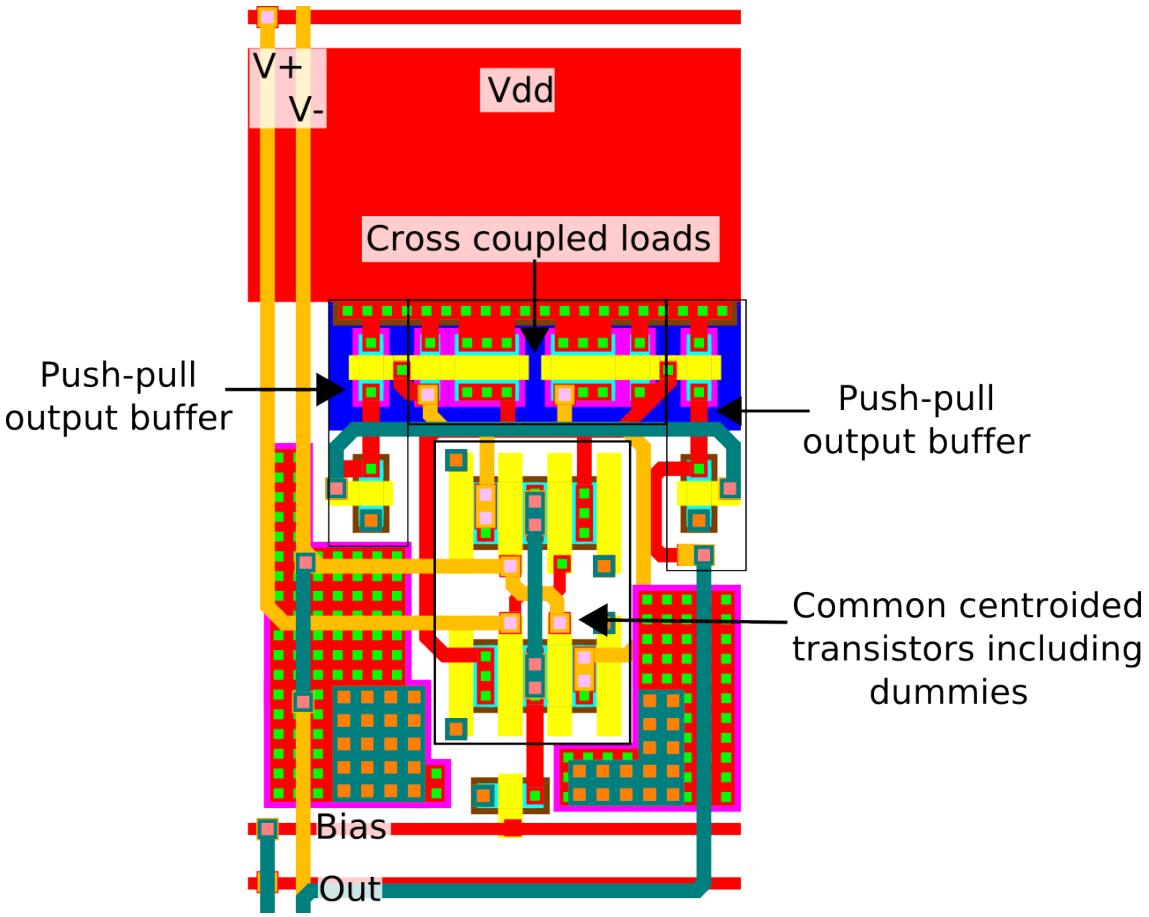


Figure 5.11: Final Comparator Layout

by either waiting for the photodiode voltage to reach a "high" value before starting the counter, or else by latching the value when the photodiode voltage reaches the high value and then subtracting the result from the end value.

The first scenario is more desirable for reducing the processing complexity, but requires a separate counter for each processing block. By calculating the high value as an offset and then subtracting it, the global counter can continue to be used. To be able to measure when the photodiode voltage passes two different thresholds, either the existing comparator must have its reference input changed after the first measurement has been made or a second comparator must be used. There is little to differentiate between the two methods, so two comparators were used due to the relative lack of complexity required.

Using two comparators does present a problem when the processing component is off-chip, namely that it doubles the number of outputs required for the comparators. With 120 pins available on the package and the intention to observe 64 pixels in parallel, using two outputs per comparator is not possible. Whilst moving to a package with a greater number of pins would permit this, it would require an associated increase in the silicon area to accomodate the new pins. With each pin having dimensions $100\,\mu\text{m} \times 300\,\mu\text{m}$, the smallest die that can hold 120 pins is $13\,\text{mm}^2$. With the cost per mm^2 being €580 [Europractice, 2006] and the price per package €60 this gives a total price of €8740 for twenty packaged devices. The next largest package that would accomodate 128 outputs for the comparators is the PGA 208. Assuming that only 56 pins are required on top of the 128 outputs (to remain consistent with the PGA 120 package) this means a die of area $27\,\text{mm}^2$ would be needed. With the new packages costing €105 each, this brings the new cost to €17,760 for twenty packaged devices, or just over a doubling of the cost. This additional cost would be a waste since in the new area available, the number of pixels could be increased to approximately 200 x 200, yet without the pins required to access them.

Consequently simply adding a second comparator could not be achieved without

reducing the number of output pins. The solution is to multiplex the two comparator outputs from one column into a single output pin. It is also useful to be able to choose whether CDS is enabled or not; if CDS is disabled, the output of the "low" comparator should be passed straight off chip whereas if CDS is enabled, the output is a little more complicated. At the start of the measurement, both comparator outputs are at 0. When the photodiode voltage passes the reference of the "high" comparator, its output will go high. This is the start of the measurement and so the output pin should go high. When the photodiode voltage passes the reference of the "low" comparator it is the end of the measurement and the output pin should go low. The end effect is that when CDS is enabled, the output waveform is a pulse whose width encodes time between the high and low comparators being triggered.

The truth table for this circuit is shown in Table 5.1, where having CDS equal to zero means it is disabled.

CDS	High Comparator	Low Comparator	Output
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	0

Table 5.1: Comparator Output Multiplexer Truth Table

The resulting circuit is shown in Figure 5.12, along with a chain of inverters to increase the drive strength and a tristate buffer for column address selection.

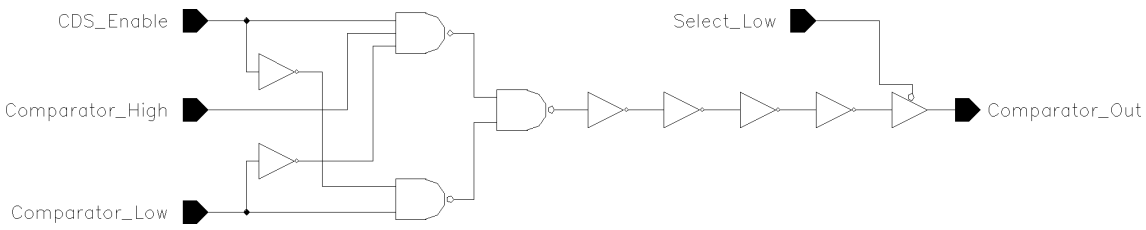


Figure 5.12: Comparator Output Multiplexer Schematic

The multiplexer was constructed using the standard digital logic cells provided by

AMS and added an extra 90 μm to the length of the column. In post layout simulations with a simulated load of 320 fF on the output to represent additional wiring capacitance, the multiplexer was able to successfully transmit pulses representing a difference of just 1 ns between the high and low comparator triggering. This corresponds to a clock frequency of 1 GHz which is much higher than the clock frequency of the digital circuitry that will be introduced in Section 5.9 and so should not limit the operational speed.

To make 64 outputs available at once, a single column address pin was used along with digital tristate buffers to choose between the two halves of the array.

5.7 Analogue Outputs

In addition to the 64 comparator outputs, 16 analogue outputs were made available so that the pixel output could be observed directly. The 16 pixels that were made available were within the 64 pixel block chosen by the row address pins and the single column address pin mentioned in Section 5.6, along with an additional two column address pins to make the final selection. The output from the pixel was taken just before the comparator input, with a further small buffer that was then used to drive an op-amp to take the signal off chip.

5.8 Camera Summary

The camera is an array of 128 x 128 pixels connected in a column based design. Each column has a shared buffer to the pixel, two comparators and a comparator output multiplexer, as well as a further analogue buffer for the pixel output. The 128 digital comparator outputs are multiplexed into 64 digital outputs and the 128 analogue pixel outputs are multiplexed into 16 analogue outputs which are passed through a op-amp buffer and sent off chip. The entire camera apart from the photodiodes has

a covering of the top layer metal to protect the circuitry from stray light. Without this covering, photons falling on the analogue circuits would inject noise into the signals. The metal covering does have a disadvantage in that it introduces an extra capacitance to all wires and so causes a slight reduction in the maximum operating speed of the design. The immunity to stray light is well worth this reduction.

The pin connections are shown in Table 5.3 and the full chip layout in Figure 5.13. The chip was packaged in a pin grid array package with 120 pins.

Pin Number	Pin Name	Purpose
1-30	Comparator Out<0:29>	Digital Output
31, 110, 120	DGND	Digital Ground
32, 109, 119	DVDD	Digital Supply
33-66	Comparator Out<30:63>	Digital Output
67-69	Column Address<0:2>	Digital Input
70	CDS Enable	Digital Input
72	VComp High	Analogue Input
73	Comparator Bias	Analogue Input
74	VComp Low	Analogue Input
75	Pixel N Bias	Analogue Input
76	Pixel P Bias	Analogue Input
77, 98, 100, 102, 104	AVDD (Core)	Analogue Supply
78, 99, 101, 103, 105	AGND (Core)	Analogue Ground
79-94	Pixel<15:0>	Analogue Output
95	AVDD (IO)	Analogue Supply
96	AGND (IO)	Analogue Ground
111	Pixel Reset	Digital Input
112-118	Row Address<0:6>	Digital Input
71, 97, 106-108	n/c	

Table 5.3: Camera Chip Pinout

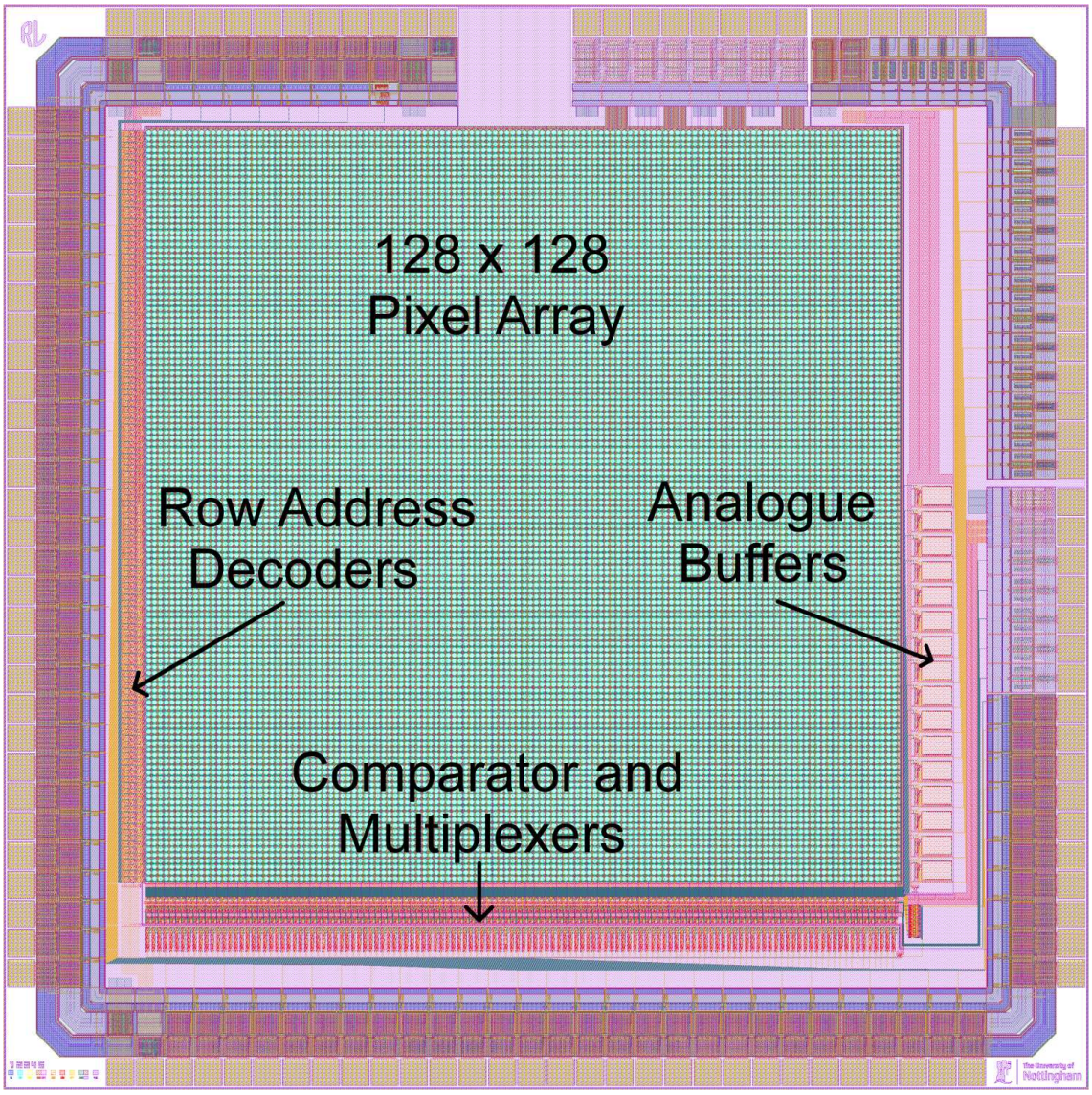


Figure 5.13: 128 x 128 Integrating APS Camera Layout

5.9 Digital Processing

Although the digital processing is not on the camera chip itself, it is still worth considering within the scheme of the camera design as it will be used directly with the camera. This section discusses the digital circuitry used to control the camera chip and perform processing on the data provided by it. This circuitry was designed in VHDL and implemented on an FPGA chip.

5.9.1 Platform

The decision to take the digital processing off the main IC was taken for three reasons; to reduce the risk associated with the chip by making the design less complex; to utilise the available silicon for pixels instead of digital logic thereby increasing the spatial resolution; and, most importantly, to allow the digital algorithm design to be changed after the main IC had been manufactured. Being able to reconfigure the digital logic means that it is possible to change the working of the camera to adapt to new applications that might arise.

The usual way of implementing digital logic, especially when it may be used on a custom IC at a later point, is to do so with a hardware language such as VHDL on a Field Programmable Gate Array (FPGA). The FPGA used was a Spartan 3 with 1.5 million gates on an in-house developed board that has a large number of digital IO pins available as well as an interface to a high speed (480 Mb/s) USB port which can be used for sending data to a computer [Zhu et al., 2006]. The main clock for the board runs at 40 MHz.

5.9.2 Basic Principle

The overall operation of the digital processing is fairly straightforward, as shown in Figure 5.14 for the two phase case. After the USB interface has been initialised,

a processing cycle is started where the relevant internal values are reset and then measurements are made for the first phase and then for the second phase. This process is repeated a number of times specified by the user to choose the overall integration period (i.e. the length of the accumulate and dump filter), after which the data is sent off chip. The row and column addresses are then updated and the whole process begins again.

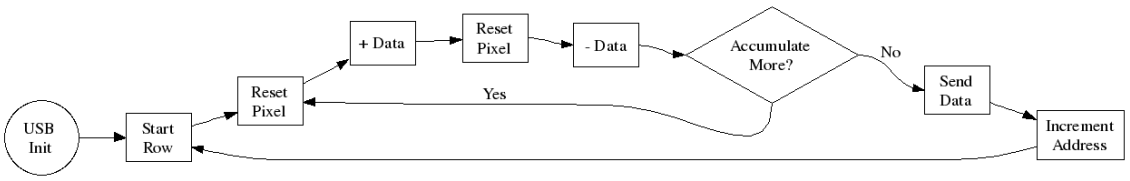


Figure 5.14: Basic Two Phase VHDL Flow

5.9.3 Architecture

5.9.3.1 Top Level Module

The top level VHDL module contains a number of other modules as well as the code for gluing them together and controlling them. A simplified diagram of these modules is shown in Figure 5.15.

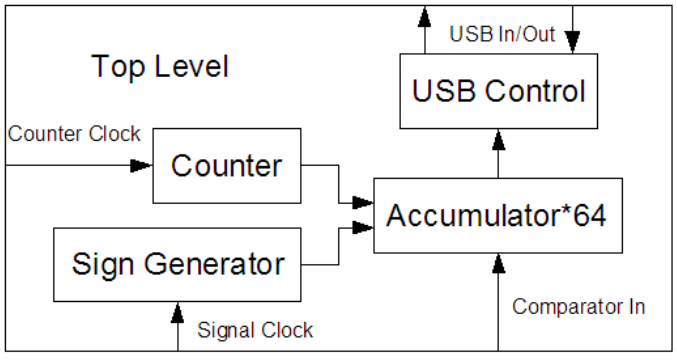


Figure 5.15: Digital Processing Architecture

5.9.3.2 USB Control

The USB control module was provided with the FPGA board [Zhu et al., 2006] and is used to initialise the USB interface as well as to send data over USB to a computer in 512 byte blocks.

5.9.3.3 Counter

The counter is a 16-bit synchronous up counter which will not overflow if it reaches 65,535 (that is, it will remain at 65,535). At 40 MHz this gives the time to maximum count as 1.38 ms, which corresponds to a minimum signal frequency of 305 Hz for the two phase case or 153 Hz for the four phase case. Trying to measure signals beneath these frequencies will cause the counter to reach the maximum value if the overall light level is low, meaning that distortion will occur (if the counter could overflow to zero this would be even greater).

5.9.3.4 Accumulators

The accumulators are 20-bit signed add/subtract devices generated using the Xilinx Coregen tool which allows a wide variety of blocks to be implemented with minimal effort whilst also being well optimised for space. With there being so many accumulators they occupy the bulk of the space in the design, so it is important that they are space efficient. Using Coregen blocks results in a black box with no code, so it is not the best approach to take when the intention is to put the end design on an ASIC, but in this case an accumulator is fairly simple to design. Using a 20-bit accumulator with a 16-bit counter allows multiple accumulations without the risk of overflowing, even at the maximum counter value. It also fits well to the available USB block size. With 512 byte blocks and 64 pixels per block, this gives 8 bytes per pixel. Written with respect to bits rather than bytes, Table 5.4 shows the content of a pixel block containing a header set to FF, the complete address and two values

for each pixel. The space for two values allows the I and Q values to be sent at the same time for the four phase mode and the amplitude and DC values for the two phase mode.

Bits	Purpose
0-7 (8)	Header
8-15 (8)	Column Address
16-23 (8)	Row Address
24-43 (20)	Accumulator Result
44-63 (20)	Accumulator Result

Table 5.4: Pixel USB Block

Having the header and complete address per pixel block means that it is easier to verify at the computer that the block is valid. Three simple errors checks can be made; the block must start with a value of FF, and both the column and row address must be between 0 and 127. If any of these conditions is false, the block is either corrupt or else at an incorrect location.

As the data is read at the computer in blocks of multiples of 512 bytes, the largest concern is that when starting to read a block, the location within the block is not known. The header per pixel block approach allows a known position to be found within one block of data being received.

For the four phase mode, it may also be required to send the DC value off chip. This would require more space than is available in the scheme described above. One way to circumvent this is to remove some of the redundancy so that each pixel block has neither the row address repeated nor has a header. Table 5.5 shows this arrangement, with the accumulators having to be reduced to 18 bits to fit.

Bits	Purpose
0-7 (8)	Column Address
8-25 (18)	Accumulator Result
26-43 (18)	Accumulator Result
44-61 (18)	Accumulator Result

Table 5.5: 62 Bit Three Data Point Pixel USB Block

As each pixel block occupies 62 bits there are 128 bits left to create an overall header for the block as well as to provide the row address. Although appearing a good

solution, having the data for the blocks not aligned to an integer number of bytes means that reconstruction at the computer is more complicated. A much better approach is to increase the block size or to decrease the number of channels being sent at once. Increasing the USB transfer block size to 1024 bytes is straightforward and even makes reconstruction at the computer easier than when sending two values because 32 bits can be devoted to each accumulator result, as shown in Table 5.6

Bits	Purpose
0-15 (16)	Header
16-23 (8)	Column Address
24-31 (8)	Row Address
32-63 (32)	Accumulator Result
64-95 (32)	Accumulator Result
96-127 (32)	Accumulator Result

Table 5.6: 128 Bit Three Data Point Pixel USB Block

5.9.3.5 Sign Generator

The accumulators used can either add or subtract a value from their stored result; the sign generator controls whether an addition or subtraction takes place. When taking DC pictures, the sign generator specifies only additions. When in the two phase AC mode, it alternates between addition and subtraction on every sample whereas in the four phase AC mode its two outputs produce "$++--$" and "$+--+$" for the I and Q accumulators respectively.

5.9.3.6 Inverting Counter

An alternative to using the standard linear counter and then correcting for the $1/x$ effect of the time pixel off chip is to simply invert the counter value before it reaches the accumulators, as discussed in Chapter 3. Three options for implementing this exist; using a linear counter with a divider to invert the value, using a linear counter as an address to a lookup table, or using a counter that has preprogrammed intervals to change its output (effectively a minimised lookup table).

Before considering any of these solutions, there is first the problem of how the signal should be inverted. Simply calculating $1/x$ gives values that are less than one and so are not suited to integer mathematics. Scaling by the maximum value, or using $65,535/x$, means that the output will range from 1 - 65,535 just as before the inversion. Due to the natural compression of the inversion, the dynamic range is now greatly reduced; it is obvious that changing the counter value from 1, with an inverted value of 65,535, to 2, with an inverted value of 32,767 moves the output over half of its range. Overall, due to duplicated values caused by rounding, the 16 bit input range results in 510 different values when inverted which is roughly equivalent to a 9 bit range, or a loss of 7 bits worth of data. This effect can be minimised by replacing the dividend with a larger value equal to 2^n-1, where $n = 16$ in the example above. This technique is simply working around the limitations of the rounding that are caused by using integers. By using a larger dividend, some values that were not previously able to be calculated without rounding can now be calculated correctly. This is at the cost of sacrificing a few values when dividing by very small numbers, since the output of the counter is still limited to 16 bits so $131,071/1$ will be truncated at the maximum value of $65,535$. Each time n is incremented by 2, the range of different values represented in the output is approximately doubled whilst keeping the shape of the overall response the same. Figure 5.16 shows the error response of this type of inverting counter for $n = 16$, 20 and 24. The step type response shows that there are long periods when the output should be increasing when it does not do so in reality, and that the change takes place very suddenly at each step. As can be seen, the higher the value of n, the greater the number of steps and so the closer the actual output remains to the ideal output.

What cannot be seen in Figure 5.16 (due to the scale), is the range of values that are truncated. Each time n is incremented by one, the number of truncated values also doubles, so for $n = 32$, every single value in the range will be truncated to the maximum. To decide on the best value for n, the difference between the actual

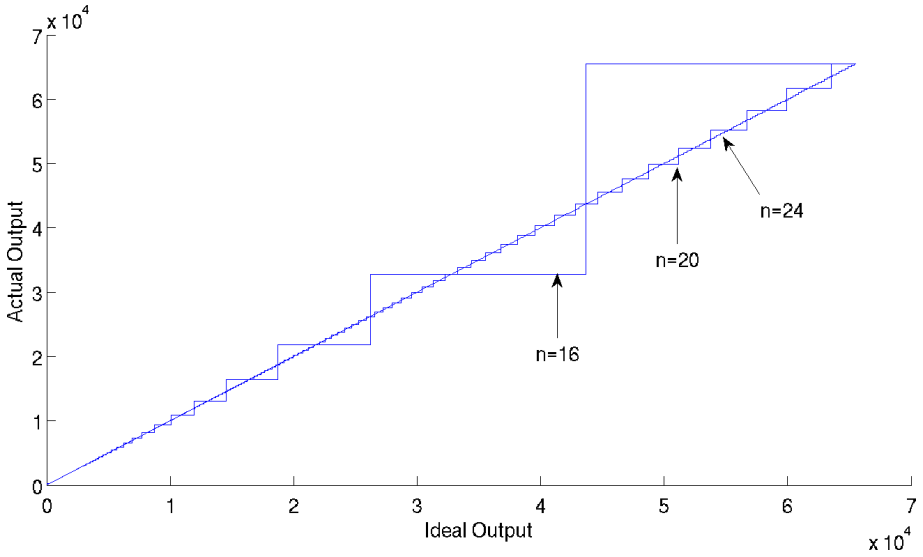


Figure 5.16: Inverting Counter Response

output and the ideal output were taken for all of the values in the range and then summed together. This was repeated for values of n in the range 16 - 32. These values give a measure of how close the actual output is to the ideal output over the entire range, with the lowest value giving the best choice. This turned out to be inbetween $n = 23$ and $n = 24$, where using 24 results in 7680 unique values, or slightly less than 13 bits.

5.9.3.7 Linear Counter with Divider

Of the four basic mathematical operators division is the hardest to implement on chip, with the usual trade off between high speed operation and large silicon usage. The Xilinx Coregen tool offers a very flexible integer division core which can perform division in one, two, four or eight clock cycles and which uses 10,085, 6863, 4089 and 2705 logic gates respectively. As expected, using a single high speed divider at the counter would take much less space than having a divider at each processing block and is also less likely to provide spurious results (the eight clock latency could cause the result to be missed). As the divisor is fixed and known, it should be possible to optimise the division further. The problem with these blocks is that it is not possible to provide a higher bit dividend without also having a higher bit output, so

$65,535$ must be used as the dividend, resulting in a shorter range of output values than desired.

5.9.3.8 Lookup Table

Using lookup tables of precalculated data in place of complex operations is a common technique to improve performance. In this case the counter value is used as an address input to a ROM which stores the inverted values. As the ROM uses a clock to load the data at the selected address, the output on any given clock edge actually represents the inverted value of the previous counter input, not the current one which is the same as with a divider which completes the division in one clock cycle. This is a minor problem compared to the amount of storage required for the lookup table; with 65,536 values of 16 bits each, this means that a total of 1,048,576 bits are required to store all of the data. Even on a custom IC this requires a large area, but on an FPGA it is even more of a problem as memory tends to be harder to implement efficiently. The Spartan 3 FPGA used has some features know as Block RAMs [Xilinx, 2004] which exist in addition to the quoted number of gates available on the chip and can be used to implement memory structures more efficiently. However, even using Block RAMs it was not possible to fit the entire lookup table on the FPGA. By halving the number of stored values to 32,768 it was successfully implemented, but with an equivalent gate count of 1,967,252 including the counter (although the FPGA only has 1.5 million gates, this does not include Block RAM).

However, if using a lookup table in this way, the vast majority of the space is used to store redundant values. As already mentioned, there are only 7680 unique values in the $(2^{24}-1)/x$ scheme meaning that 57,856 values are redundant - or 88.3% of the space. Assuming that the gate count scales linearly, getting rid of the redundant values would mean that the gate count would become 461,075 which is a much more reasonable amount.

5.9.3.9 Reduced Lookup Table

The reduced lookup table attempts to remove this redundant data. With an ordinary lookup table this would not necessarily be possible because they are usually random access read only memory structures. As the counter that is used as the address counts up in single increments, it is possible to write code so that the output value only changes at the required places and at all other times it simply remains unchanged. The VHDL code snippet below demonstrates how this works. The variable "count" is the value to be inverted, with "q_cur" and "q_prev" storing the current and previous output values respectively. On the rising edge of each clock, the value in q_cur is loaded into q_prev. At the start of the operation, count will be equal to "0000000000000000" and so q_cur will be loaded with "1111111111111111". On the next clock edge, the count has been incremented so now equals 1, but there is no explicit output value specified for this count, so q_cur is loaded with q_prev. This continues as count increases until the count reaches "0000000001000001" when the new value will be loaded. A line is required for every unique value in the range.

```
with count select
    q_cur <= ''1111111111111111'' when ''0000000000000000'',
             ''1111110000001111'' when ''0000000001000001'',
             q_prev when others;
```

Table 5.7 shows the space requirements for this style of inverting block (the counter itself is not included in the gate count) for different values of n.

5.9.4 Final Designs

Two versions of the digital code were produced, with the only difference between them being the counter implementation. This gave rise to a version for the time pixel, using an ordinary counter, and a version for the inverse time pixel using a

n	Unique Values	Storage Required	Implemented Gate Count
16	510	8160	10,703
17	721	11,536	15,317
18	1016	16,256	21,899
19	1433	22,928	31,172
20	2016	32,256	44,261
21	2833	45,328	62,717
22	3968	63,488	88,880
23	5537	88,592	127,010
24	7680	122,880	182,240

Table 5.7: Reduced Lookup Table Code Size
as a function of n (number of bits in the counter)

reduced lookup table inverting counter with $n = 24$. Both versions have the option of selecting the mode of operation between either DC, two phase or four phase Quadrature based on the use of switches on the FPGA board. The other options available using other switches are to chose the length of the filter (either 4 or 16) and to control whether the addressing works as normal or is fixed on a single address so that easier characterisation is possible.

The design for the time pixel used 67,390 gates and the design for the inverse time pixel used 257,530 gates.

5.10 Summary

This chapter has introduced the final camera designed, a 128 x 128 pixel design measuring 3.7 mm x 3.7 mm and contained in a PGA120 package. The camera has 64 digital comparator outputs and 16 analogue pixel outputs. There is the option of choosing between the "normal" operation, using a single comparator per pixel column, and a correlated double sampling method which uses two comparators per pixel column. The row addressing circuitry presented also ensures that the pixel rows that are not in use remain in reset so that they contribute as little cross talk as possible to the active pixels. The design differs from existing commercial cameras in the way the light intensity is measured.

The digital processing to be implemented on an FPGA linked to the custom camera has been introduced. The FPGA interfaces to the camera to provide the appropriate control needed, that is the addressing, the reset and the retrieval of results from the comparators. It processes the results based on the mode selected and the length of the filter to be applied and then sends the results to a computer via USB, where they are then recorded and displayed with the final processing applied.

Chapter 6

Results

6.1 Introduction

This chapter details the experimental results obtained from the camera chip discussed in Chapter 5. There is a great deal of overlap between the results obtained from the first two prototype chips and the camera chip so for brevity and to allow better comparison between the results, the results presented here are from the final camera chip. The results consider the two phase method of demodulation for the time and the inverse time pixels, as well as the Quadrature four phase method.

6.2 Camera Experimental Setup

The basic experimental setup is shown in Figure 6.1. This setup allows the camera to be tested in a known environment by using the reference photodiode to determine the amount of light that is striking each pixel on the camera.

The laser is an IQ1A module from Laser 2000. It is a semiconductor laser with wavelength 650 nm that is able to be directly modulated up to 30 MHz which makes it ideal for characterising modulated light cameras. The reference photodiode used

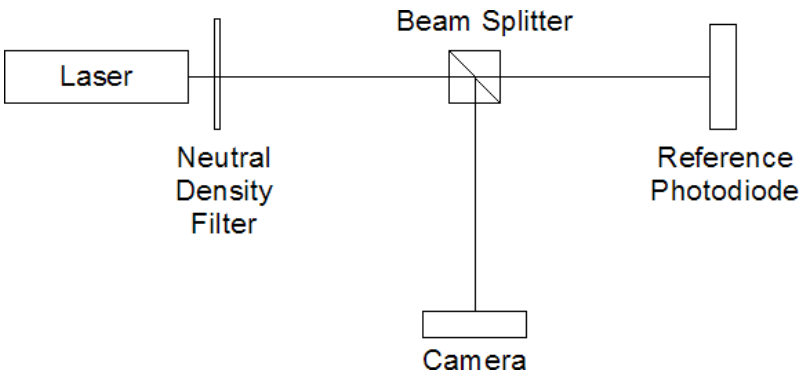


Figure 6.1: Experimental Setup

was a PDA55 from Thorlabs featuring programmable current to voltage conversion gain.

By measuring the light intensity striking the reference diode, the amount of light striking the camera chip could be calculated and hence the current flowing in the photodiodes. Over the linear range of the laser (for a voltage input of 400 mV to 2.5 V), the voltage output from the photodiode varied linearly between 281.8 mV and 3.2 V when used with a gain of 47 kV/A and 150 kV/A respectively. With the photodiode active area being 3.6 mm x 3.6 mm and having a response at 650 nm of approximately 0.4 A/W this corresponds to a power between 362.4 fW/μm^2 and 13.13 pW/μm^2. Given that the photodiodes on the camera have an area of 221.7 μm^2 and a quantum efficiency of approximately 0.3 A/W this gives a photocurrent in the range of 24.1 pA to 873.3 pA per pixel.

Figure 6.2 shows a photograph of the experiment which also includes a diffusing disk placed in front of the chip to remove the laser speckle on the chip for fixed pattern noise measurements. The experiment would normally be covered with a light proof sheet.

6.3 DC Results

As with the simulations in Chapter 3, it is sensible to look at the DC operation of the camera before progressing to the more involved results. The first experiment was to

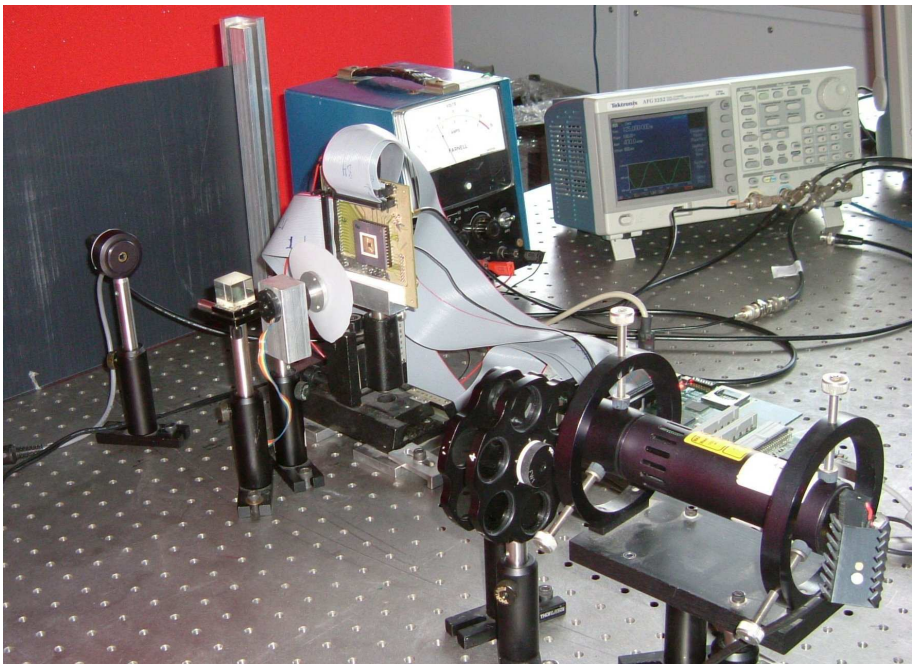


Figure 6.2: Optical Bench Layout

try to measure the amount of dark current. The camera was placed in a completely light-proofed environment and the time for the comparator to trigger (with a reference voltage of 1.25 V) was measured as 1.92 s on an oscilloscope. Simulations of the pixel were carried out to determine the photocurrent which produced the same response in the comparator and this gave a dark current of 7.2 fA, or approximately seven times greater than the predicted value. This difference in results is most likely due to a small amount of light being present in the experiment as it is extremely hard to ensure that no light is present.

The second experiment was to calculate the response of the camera to different DC light levels. In this experiment the camera was illuminated with light of constant intensity and the experiment was performed in an enclosed environment so that no other light would affect it. The results are shown for four different pixels spread evenly across a single chip.

The voltage applied to the laser was swept between 400 mV and 1 V and the corresponding count values recorded for a comparator reference voltage of 0.85 V. The results are shown in Figure 6.3 and show the expected reciprocal relationship to

the photocurrent. The difference in the curves is largely due to the laser speckle producing different levels of brightness at the different pixels, since this was only removed for the fixed pattern noise and image tests.

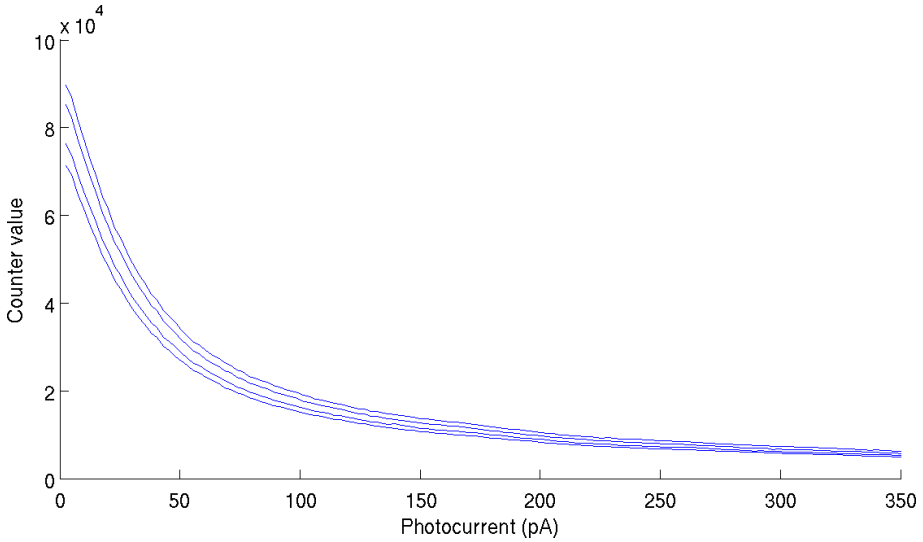


Figure 6.3: DC Response (Time Pixel)

Figure 6.4 shows the pixel and corresponding comparator output when illuminated by a square wave set to produce both the fastest and slowest discharges shown in Figure 6.3. The time between the reset being released and the comparator triggering was measured on an oscilloscope for both cases, with a time of 162 μs for the fast discharge and 2.25 ms for the slow. With the 40 MHz clock for the counter, this results in final count values of 6480 and 90,000 respectively which match with the values recorded above from the FPGA.

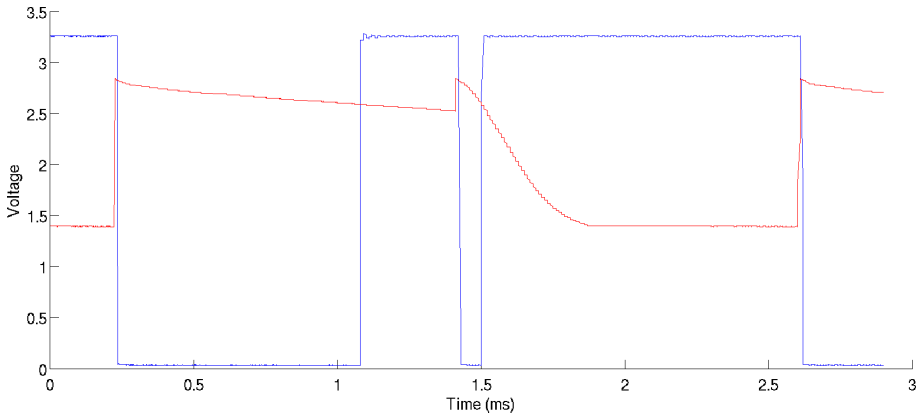
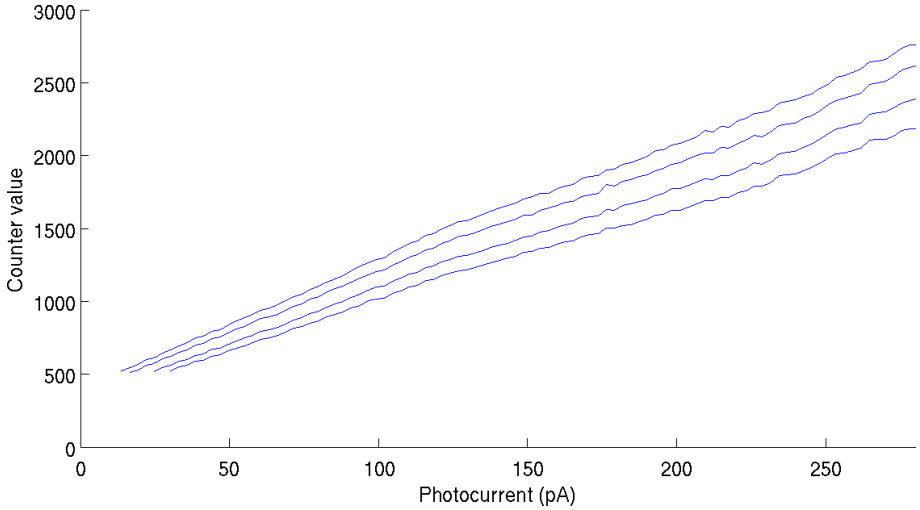


Figure 6.4: Pixel and Comparator Waveforms at DC
(Blue: Comparator, Red: Pixel)

The results for the inverse time pixel are shown in Figure 6.5 and confirm that the inverting counter works.



(a) Time Pixel

Figure 6.5: DC Response (Inverse Time Pixel)

The final test of the DC operation is to look at the variation in time for a constant illumination. Assuming no other environmental factors, this can give an indication of the read noise of the pixel and so determine the smallest signal that can be measured. This was carried out by taking 3000 frames and comparing the different values produced for a pixel. The results for a laser voltage of 0.5 V (photocurrent 135 pA) are plotted in a histogram in Figure 6.6. The results roughly follow a normal distribution with a mean of 16,623 and a standard deviation of 82, meaning that 99.7% of the results lie within ±246 of the mean, or approximately 1.5% of the total value.

6.4 Two Phase Results

6.4.1 Amplitude Response

For the two phase pixel, the amplitude response is the most important result. The phase cannot be calculated using this technique and the frequency response is also

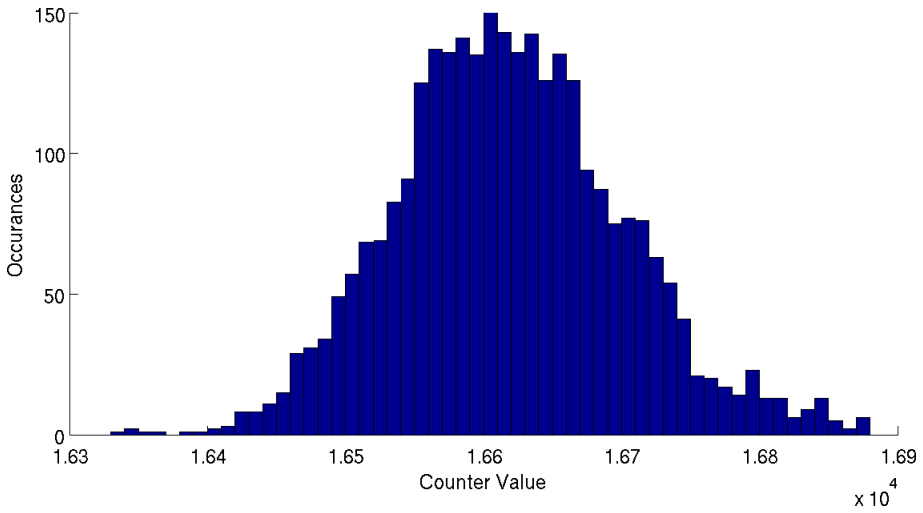


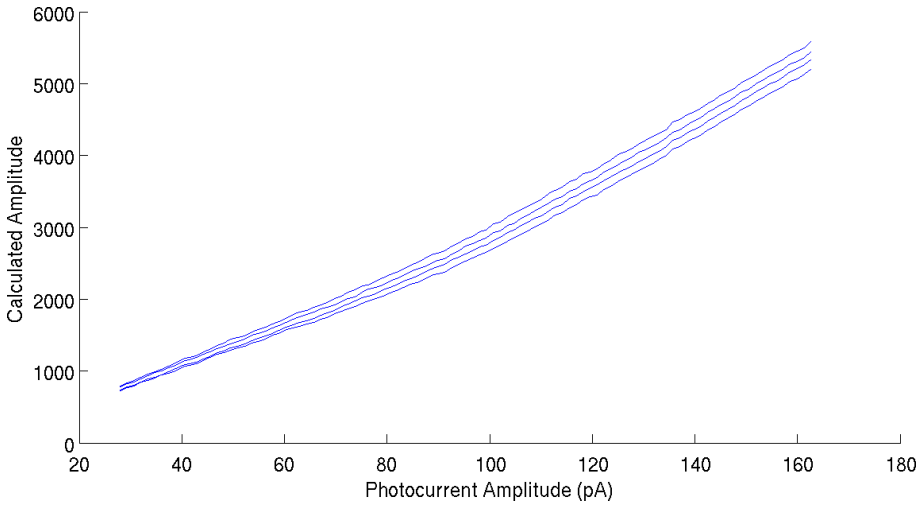
Figure 6.6: Read Noise Results

a measure of the amplitude. Figure 6.7 shows the amplitude response of both the pixels when illuminated with a laser at a fixed DC of 350 pA with an AC amplitude of between 30 pA and 160 pA at a modulation frequency of 500 Hz. The response matches the simulated amplitude response, but with a much more pronounced curve.

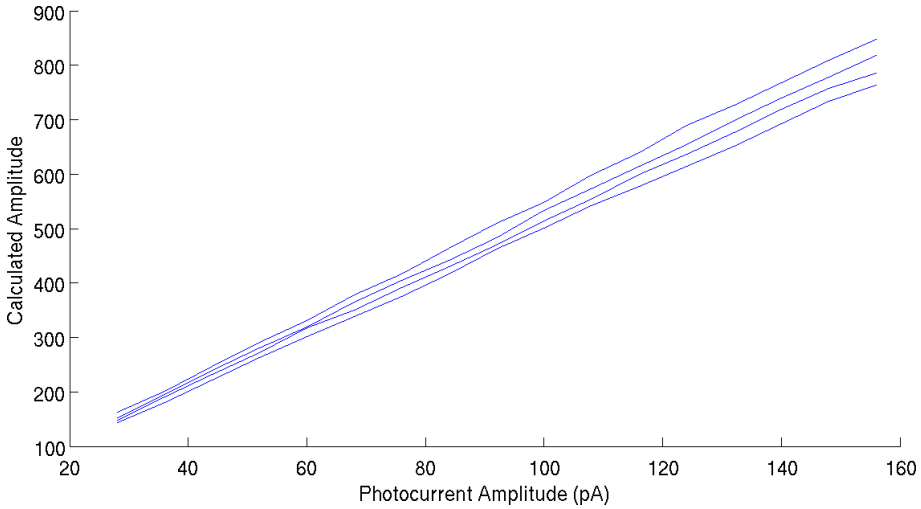
The amplitude response under a changing DC level is shown in Figure 6.8. The response for the time pixel has a very similar shape to the simulation results and the inverse time pixel also has similar components to the simulation but the actual shape is quite different. Considering the plot from the left hand side, there is a linearly increasing segment which is followed by a sharp drop. The same feature appears in the simulation although the drop is not as great. Above this point the response is not as flat as in the simulation, but does remain reasonably constant unlike for the time pixel.

6.4.2 Phase Response

The two phase technique is heavily dependant on the phase between the laser modulation and the local oscillator. Even when using synchronisation signals to lock two signal generators together there can still be a small amount of drift or jitter, causing the phase relationship to be less than ideal. A two channel signal generator was used

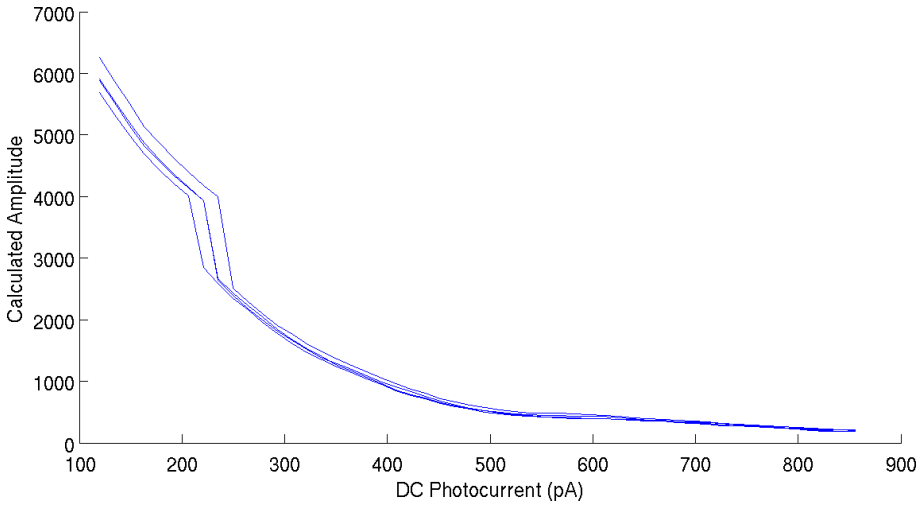


(a) Time Pixel

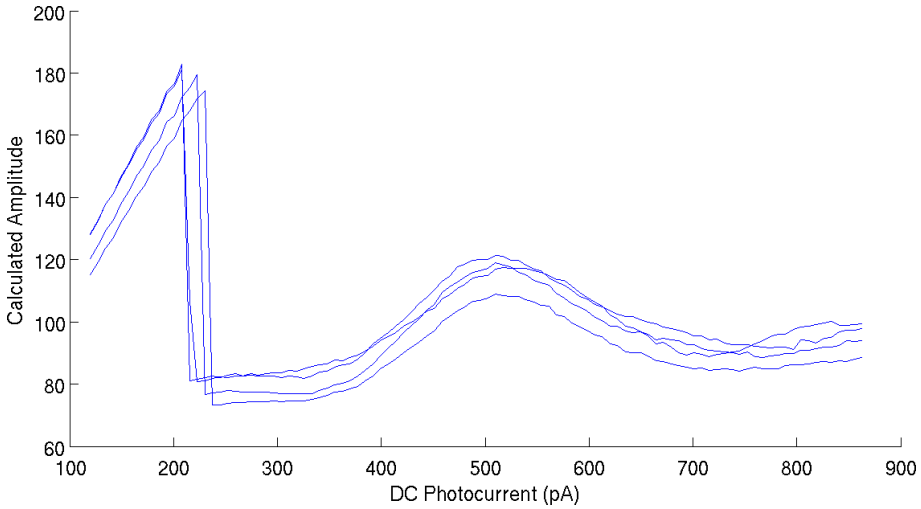


(b) Inverse Time Pixel

Figure 6.7: Two Phase Pixel Amplitude Response (4 Point Filter)



(a) Time Pixel



(b) Inverse Time Pixel

Figure 6.8: Two Phase Pixel Amplitude/DC Response (4 Point Filter)

to get around this problem, since the relationship between the two channels is much more consistent. The signal generator used, a Tektronik AFG3252, has two outputs that can be controlled independently as well as a third output that provides a logic compatible square wave at the same frequency as the first channel. The first channel was used to provide the reset pulse to the camera, with the logic output providing synchronisation for the FPGA. It is perfectly possible to use the FPGA to create the reset and a synchronisation signal (if required), but using a signal generator in this manner gives much greater flexibility when characterising the camera. The second channel was used to control the laser.

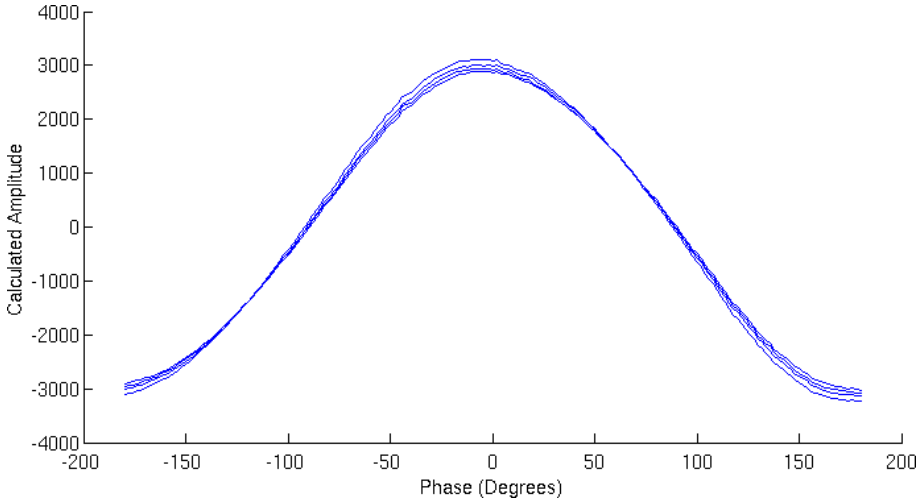


Figure 6.9: Two Phase Pixel Phase Response

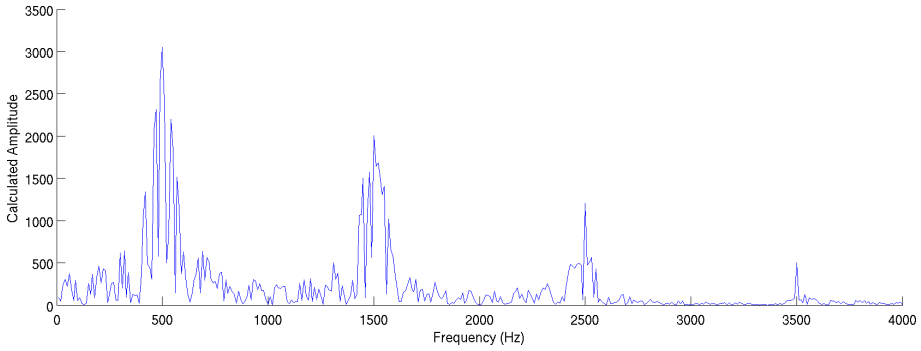
Figure 6.9 shows how the calculated amplitude of the signal varies for the time pixel as the phase relationship between the laser modulation and the local oscillator changes through 360°. This illustrates the point that even if the phase is optimal, the order of the samples may mean that the result is negative, meaning that an extra step is required to ensure that all results have the same sign as each other.

6.4.3 Frequency Response

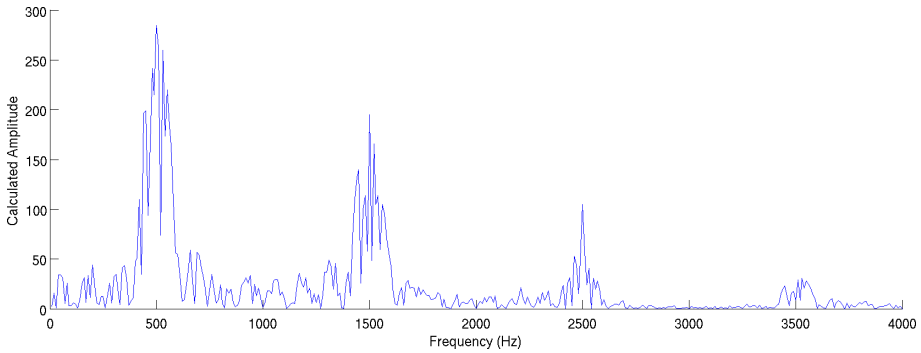
Because the two phase method is so tied to the phase between the modulation and the local oscillator, it is difficult to get a true frequency response. When the modulated signal is at a frequency which is not the lock-in frequency (or harmonic) then

the phase between the modulated signal and the local oscillator will be constantly changing. Due to the way in which the signal generator works, it is not possible to guarantee that the phase between the two signals will produce the maximum possible amplitude when carrying an automated frequency sweep. This means that for any given frequency, the phase will be essentially random and so the result obtained will be unlikely to show the maximum response possible for that frequency.

To determine the frequency response, the sampling frequency was kept constant at 1 kHz (meaning the lock-in frequency is 500 Hz) and the laser frequency was swept from 10 Hz to 4 kHz with a step of 10 Hz. Figure 6.10 shows the frequency response as measured for a single time pixel with a low pass filter of length four.



(a) Time Pixel

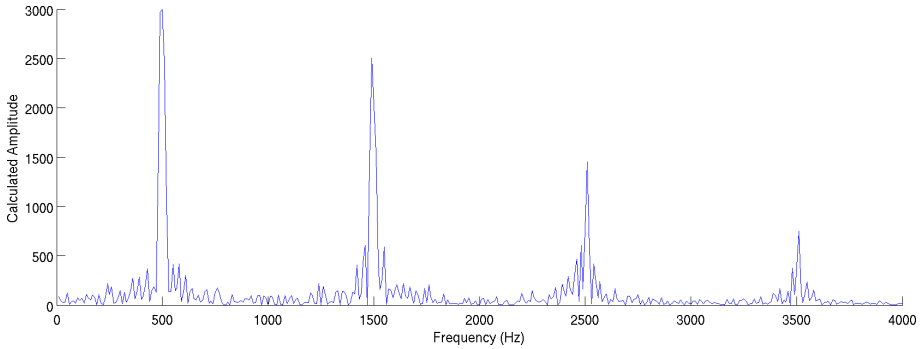


(b) Inverse Time Pixel

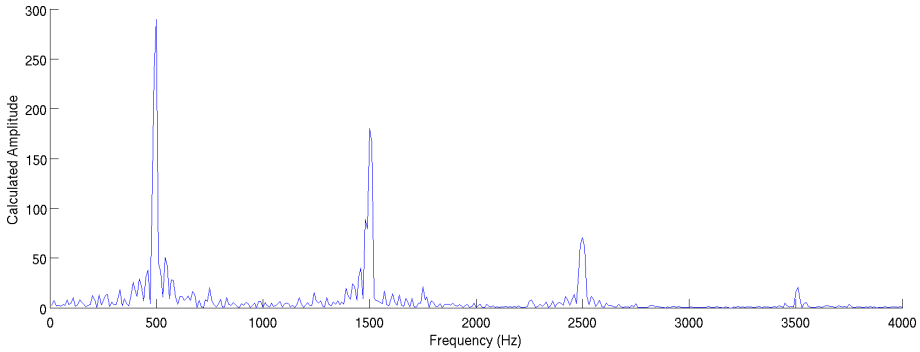
Figure 6.10: Two Phase Pixel Frequency Response (Filter Length 4)

With a correct set of results the response around the peaks (indicating the worst case scenario) would follow a smooth curve but because of the effect described above this is not the case. The measurements at the peaks are all correct, having been manually measured and replaced in the automated results. Figure 6.9 shows the frequency

response for the same arrangement when using a filter with 16 points. The response becomes narrower at the lock-in frequency and at its harmonics exactly as predicted in the simulations.



(a) Time Pixel



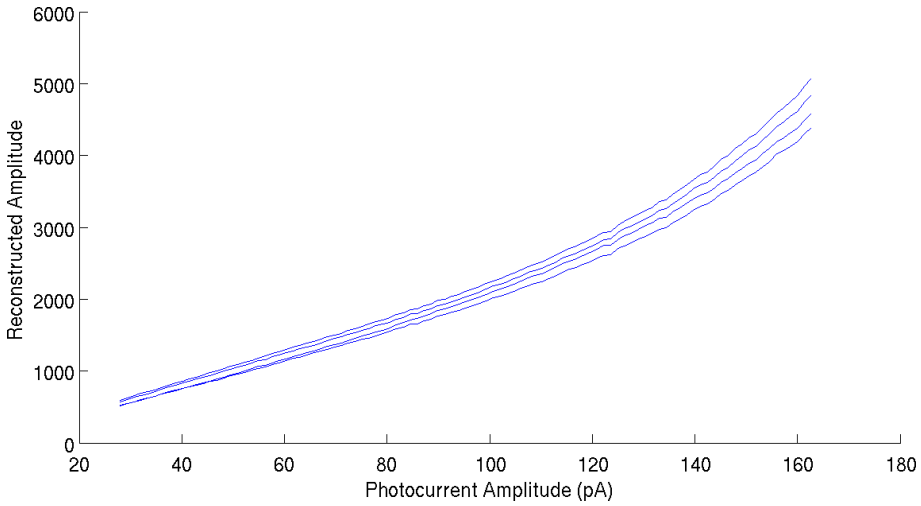
(b) Inverse Time Pixel

Figure 6.11: Two Phase Pixel Frequency Response (Filter Length 16)

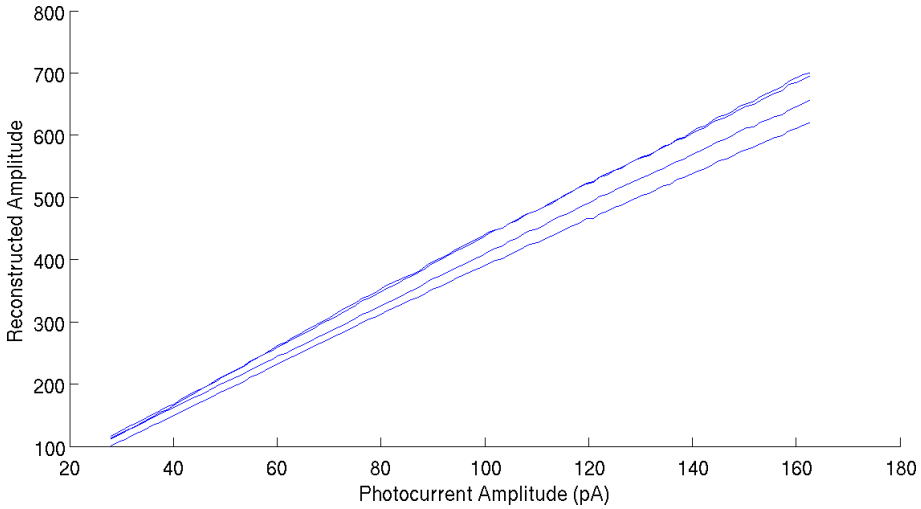
6.5 Four Phase Results

6.5.1 Amplitude Response

The amplitude response for the two types of four phase pixel are shown in Figure 6.12, with their response under different DC light levels shown in Figure 6.13. Both sets of results are much the same as for the two phase pixels, with the response for the inverse time pixel again differing from the simulation results quite obviously.

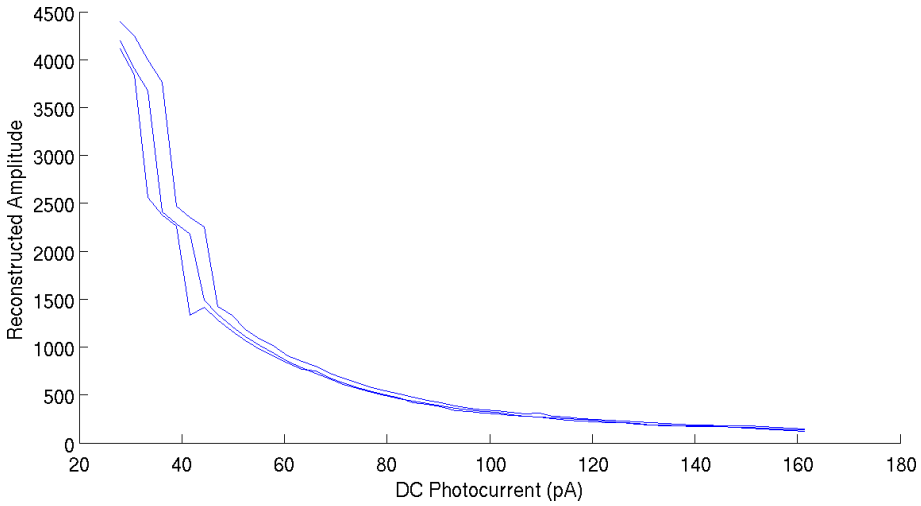


(a) Time Pixel

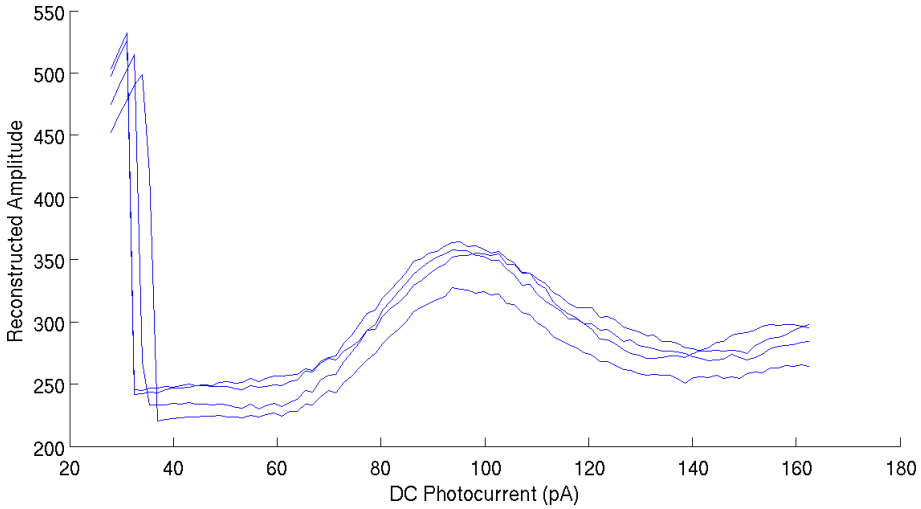


(b) Inverse Time Pixel

Figure 6.12: Four Phase Pixels Amplitude Response



(a) Time Pixel



(b) Inverse Time Pixel

Figure 6.13: Four Phase Pixels Amplitude/DC Response

6.5.2 Phase Response

Figure 6.14 shows the error in the reconstructed phase for the time pixel (blue) and the inverse time pixel (red).

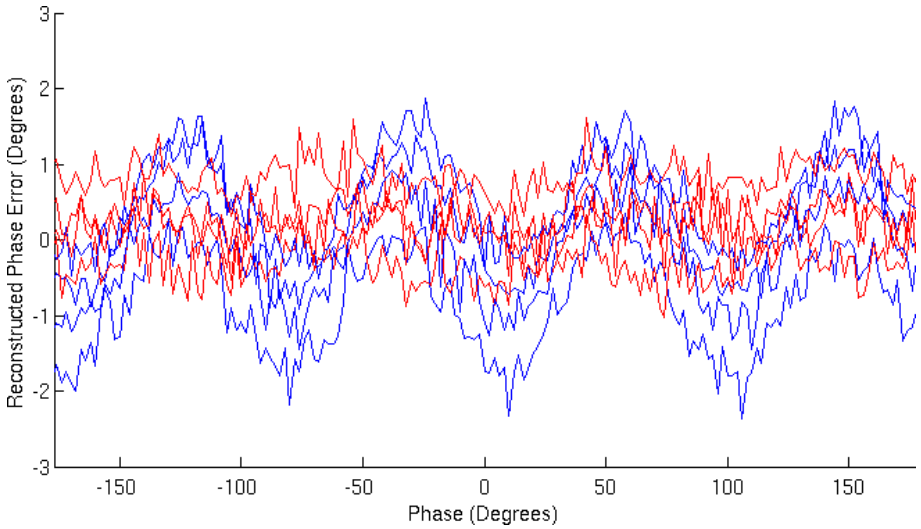
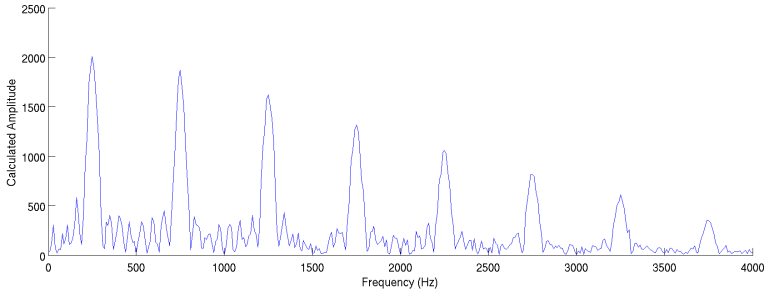


Figure 6.14: Four Phase Pixel Phase Response
(Blue: Time pixel, Red: Inverse time pixel)

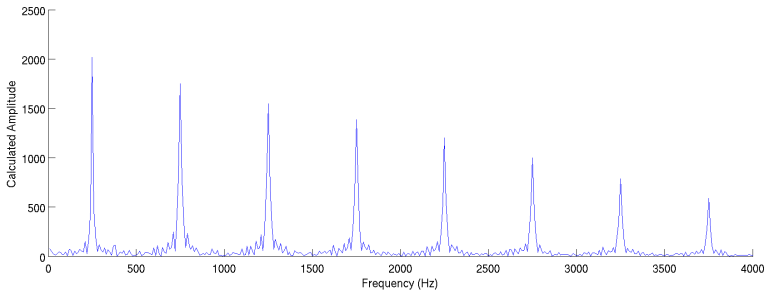
The first thing to note is that the different pixels have produced different offsets, which is caused by the different DC intensity at each of them. When normalising for the DC, this offset is removed. Across the complete range, the time pixel varies by approximately between $\pm 0.5°$ and $\pm 1°$ whereas the inverse time pixel varies by between $\pm 0.25°$ and $\pm 0.5°$. The noise on the waveforms is caused by the variation when reading individual results. By averaging multiple results together the effect of this noise is reduced. The results are better than those predicted by the simulations since the modulation depth is not as high. The results contrast with commercially available lock-in amplifiers where better phase errors are achieved. For example, the Model 7124 lock-in amplifier from Signal Recovery can carry out phase measurements with a resolution of less than $0.01°$ [Signal Recovery, 2008].

6.5.3 Frequency Response

Figure 6.15 shows the frequency response of the four phase time pixel with a reference voltage of 0.85 V, for both 4 and 16 point filters. The results for the inverse time pixel are practically indistinguishable from these results with the exception of a different scale, so they are not repeated here.



(a) 4 Point Filter



(b) 6 Point Filter

Figure 6.15: Four Phase Pixel Frequency Response

Figure 6.16 shows the frequency response for a time pixel using a filter with 4 points where the comparator reference voltage has been increased from 0.85 V to 1.25 V. As shown previously, the relative amplitude of the odd harmonics compared to the fundamental frequency increases because less of the incoming signal is being used and so it becomes closer to single point sampling.

6.5.4 Correlated Double Sampling

Correlated double sampling (CDS) was implemented on the camera as a means of reducing the reset noise at the pixel. The ability exists to enable/disable CDS through

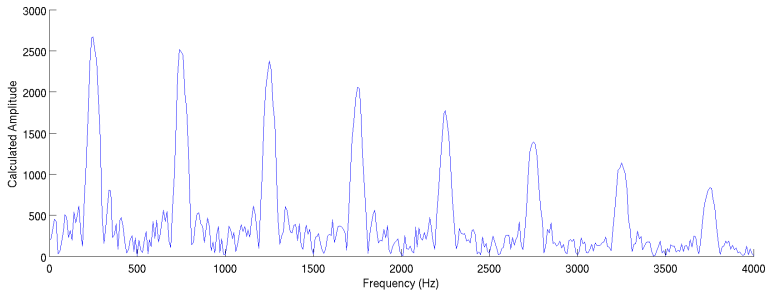


Figure 6.16: Four Phase Pixel Frequency Response (1.25 VReference)

an input select pin, with the default setting being to disable CDS. Unfortunately, when CDS was enabled on the camera the current consumption of the chip went up significantly and no useful output was produced. This was subsequently determined to be caused be a problem with the PCB.

6.6 Image Results

6.6.1 Setup

Having demonstrated that the individual pixels work as required, the next step is to look at the camera as a whole. A slightly different setup was used to simplify the imaging. The setup is shown in Figure 6.17, where the diffuser is a spinning plastic disk used to remove the speckle pattern of the laser. The laser provides the modulated light component and the LED provides static illumination that does not contain any image information. The experiments were carried out with the experiment uncovered and under mains lighting illumination so an additional component at 100 Hz was also present.

The imaging was carried out in transmission mode by printing the graphic to be imaged onto transparent plastic and placing this in the optical path. Imaging in this way was necessary since the modulated laser beam had to be expanded significantly to cover the area being imaged which reduces the light power per unit area, something that is already substantially reduced due to the diffuser. The graphic

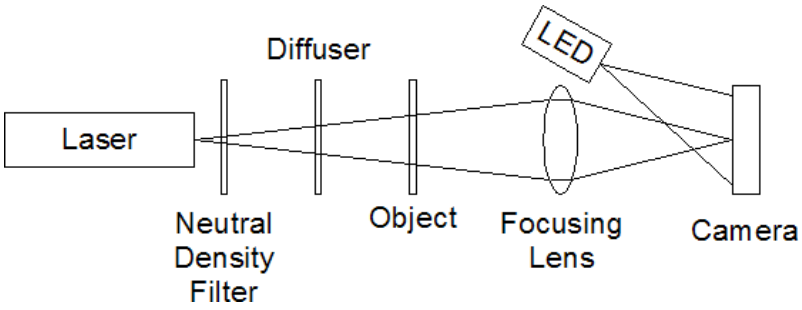


Figure 6.17: Imaging Optical Setup

used for imaging was the Applied Optics Group logo as shown in Figure 6.18. As the camera array is square, only part of the logo could be imaged at once without wasting a large amount of the usable area. The graphic was 8 mm tall and with an array size of 2.56 mm this leads to a magnification of 0.32. A lens with focal 50 mm was used, resulting in a lens to object distance of 206 mm and a lens to image distance of 66 mm to achieve the required magnification. The results were largely the same for both types of pixel (with the exception of the range of values themselves), so only the results for the time pixel are shown.

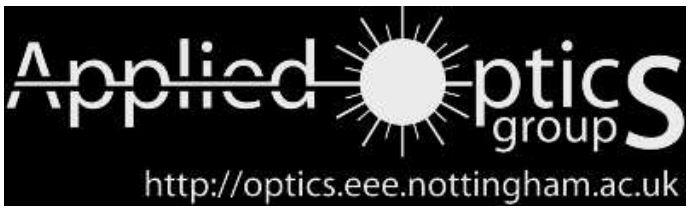


Figure 6.18: Applied Optics Group Logo

Determining the amount of light falling on the sensor is more complicated than in the previous experiment because of the extra light provided by the LED. This is the main reason for changing the setup to remove the beam splitter and the reference photodiode. To estimate the level of the DC and AC components of the light falling on the camera, it was replaced with the reference photodiode and separate measurements made with the LED turned on and the laser turned off (for the DC) and with the LED turned off, the laser turned on and the experiment covered over to prevent other light entering (for the AC). The DC component produced a voltage on the reference of 3.3 V with a gain of 470 kV/A meaning a light power of 17.55 μW, assuming a response of 0.4 A/W across all wavelengths. The AC component was a

sine wave that resulted in a minimum output of 255 mV and a maximum of 558 mV, both with a gain of 1.5 MV/A, resulting in a mean light power at 650 nm of 677.5 nW with an amplitude of 252.5 nW. Using these numbers allows an estimation of the modulation depth to be calculated as 1.39%. The experiments were carried out with a reset frequency of 840 Hz (with modulation frequencies of 420 Hz and 210 Hz for the two and four phase modes respectively).

6.6.2 DC Image Results

Figure 6.19 shows an image taken with the camera in time pixel mode with the laser on but unmodulated, so the only light on the sensor that is modulated is from the mains lighting. The main words in the logo are clear and easy to read however the smaller writing underneath cannot be made out. It is worth noting the two columns of pixels that are on the far left of the image as they are displaying a value consistently around 1000 higher than the next columns along, indicating that the pixels appear to be slightly darker than the others. One possible explanation for this effect is that the light shield around the edge of the array is actually casting a shadow over the rows. This is consistent with the LED illumination coming in at an angle on the correct side of the array however given that the top level metal used as the light shield is approximately 6 μm above the n-well junction the light would have to be arriving at quite a shallow angle to cast a shadow over the first two rows.

6.6.3 Two Phase Image Results

Figure 6.20 shows an image for the time pixel where the laser is turned on but not producing any modulation, exactly as was the case for Figure 6.19. The image intensity shows the calculated amplitude. As is expected, the lack of any pattern can clearly be seen and the range of values present is very small.

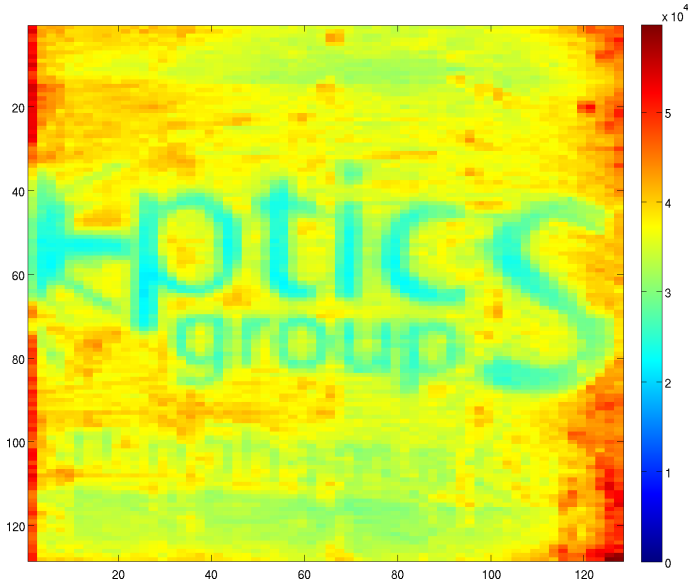


Figure 6.19: DC Image (Time Pixel)

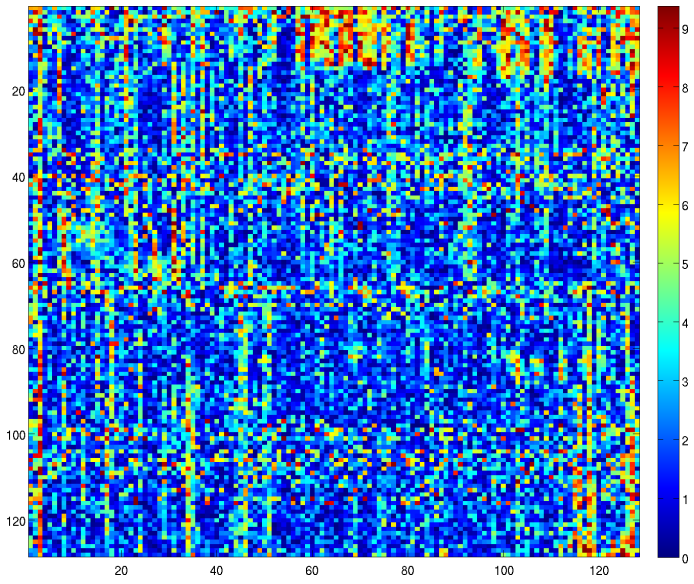


Figure 6.20: Two Phase Image (Time Pixel, No Modulation)

Contrast this with Figure 6.21, where the modulation described in Section 6.6.1 at 420 Hz has been introduced. The logo is clearly visible, with the web address at the bottom being legible and more features on the "sun" on the left becoming clear. Even more obvious is the dark border around the logo which was not visible in the DC image. This shows the large difference between the modulated amplitude caused by the dark area of the logo. It is worth pointing out that the difference in brightness of the main part of the logo compared to the web address and outside the border is caused by the profile of the laser beam being quite wide and narrow.

The same rows that were producing different values in the DC image are again producing seemingly erroneous results, although as there are also similar defects in the image in the lower right hand corner it does seem unlikely that it is caused by shadows.

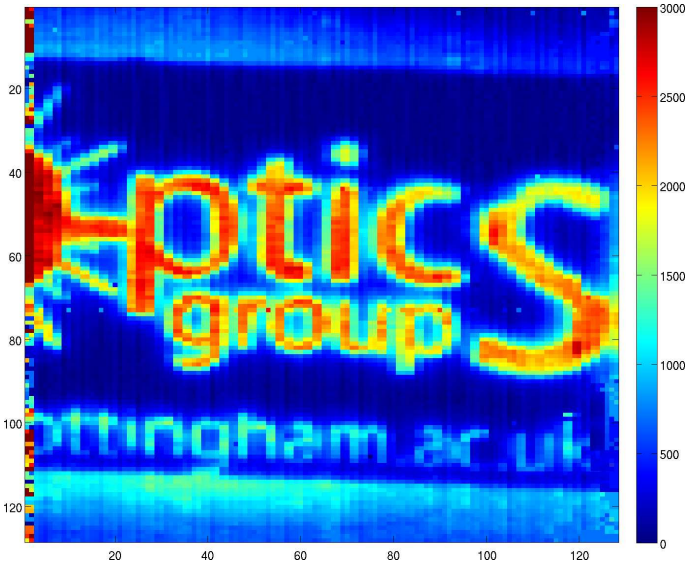
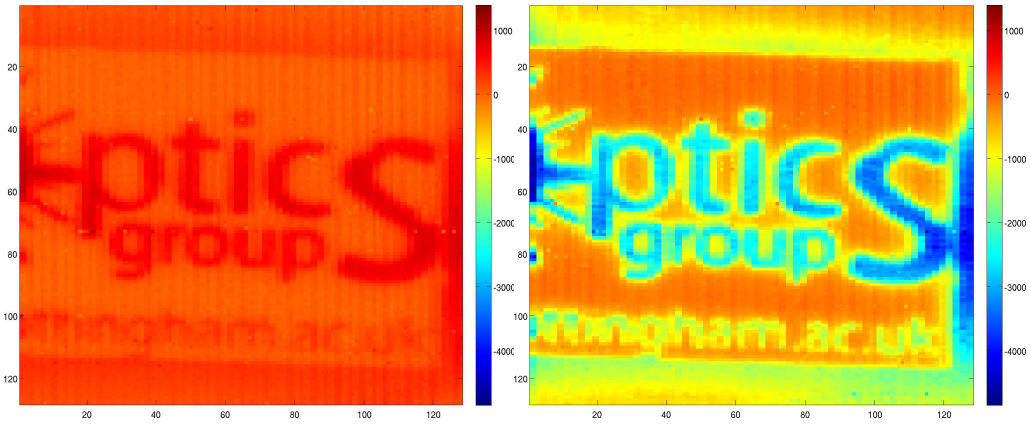


Figure 6.21: Two Phase Image (Time Pixel)

6.6.4 Four Phase Image Results

The four phase results are naturally more complicated than the two phase results. Figure 6.22 shows the "raw" results from the camera, that is the I and the Q results. The images show starkly different values but still give a reasonable idea of the final

image. The difference in the corresponding values between the two results is of course due to the phase difference between the modulation and the local oscillator. When the phase is 0°, the images appear identical. The amplitude response calculated from the I and Q results is shown in Figure 6.23 and displays very similar characteristics to the two phase image, although with a slightly lower measured amplitude. Similar results were also obtained for a modulation frequency of 2.5 kHz (i.e. a sampling frequency of 10 kHz), although this required an increase in the amount of light on the chip to achieve and the contrast of the final image is much lower than in Figure 6.23.



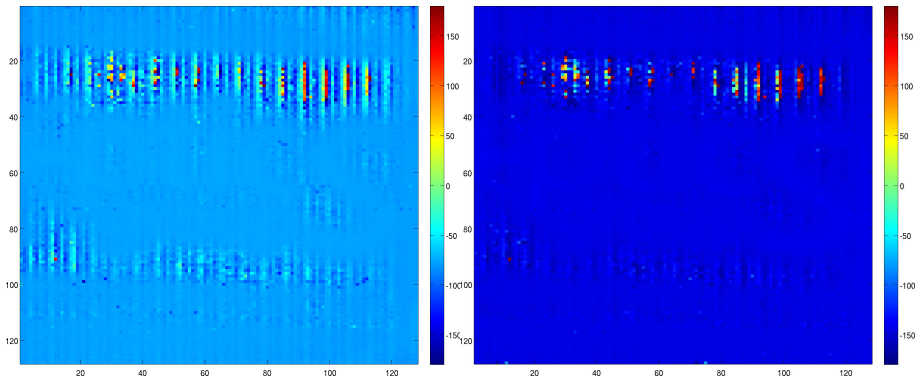
(a) I Channel (b) Q Channel

Figure 6.22: Four Phase Raw Results

Figure 6.23: Four Phase Amplitude Image (Time Pixel)

The next two images show the results of considering the phase. Figure 6.24 (a) uses

the same data as was used for Figure 6.23, but calculates the phase instead. The overall phase is very consistent across the whole image, although some artifacts can be seen. The phase is flattest where the amplitude is highest in Figure 6.23, so the words in the logo can be just about made out. Conversely, the areas where the artifacts are seen are the areas with the lowest amplitude and so even small changes in either of the two values will cause a large change in phase, as is seen (these areas were actually made even darker in the image by colouring them in with permanent marker in an effort to produce a more noticeable effect in the amplitude, something which appeared to make no difference). The second image, Figure 6.24 (b), shows the results for the same setup but with the phase between the modulation and the local oscillator being shifted by 90°.



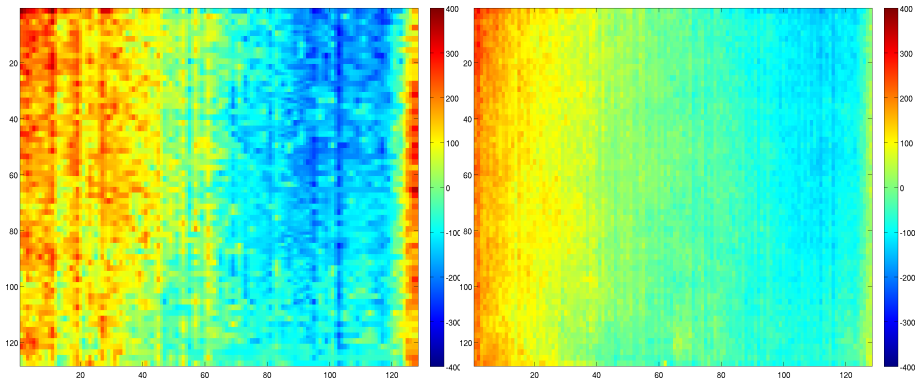
(a) (b)

Figure 6.24: Four Phase Phase Image (Time Pixel)

6.7 Fixed Pattern Noise

Measuring the fixed pattern noise turned out to be extremely difficult. Despite replacing the laser with a higher powered one (that couldn't be modulated) so that the beam could be expanded significantly, it was still possible to detect a gradient along the camera. Approximately 1000 frames were averaged to remove any temporal noise present. The images obtained for two different chips are shown in Figure 6.25. The gradient across the chips is plain to see, as are stripes running vertically (as

could also be seen on some of the previous images). The values shown are not the actual values obtained from the camera, rather they have had the mean of the raw values subtracted from them so that they represent a deviation away from what is hopefully the "true" value.

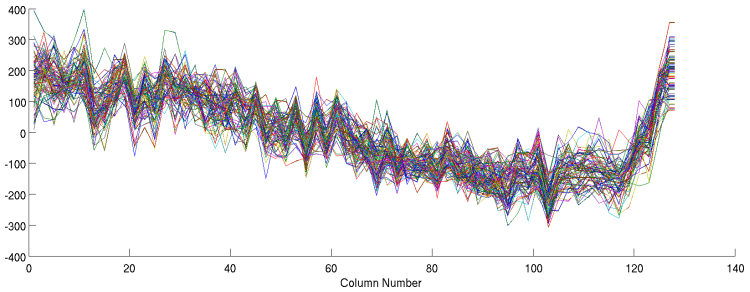


(a) Chip 1 (b) Chip 2

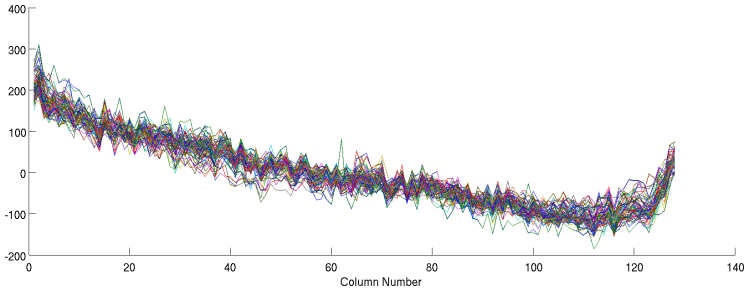
Figure 6.25: Fixed Pattern Noise Image

Figure 6.26 and Figure 6.27 show the data from Figure 6.25 plotted as individual columns and rows respectively. The graphs show nicely the type of behaviour that is expected in this kind of design. Figure 6.26 shows the results along all of the rows for each column. It can be seen that for the most part the results for a particular column have the same bias, with the variation between columns being obvious as well as the gradient across the chip. Figure 6.27 on the other hand shows the results along the columns for each row. Although it is hard to see, this result confirms that the results within a column have a relatively small variation. It also shows some odd behaviour in that the variation down a column tends to change every two pixels, so there are pairs of pixels that are very similar followed by another pair that are close to each other but reasonably different to the previous pair. The most plausible explanation for this is that the least significant bit of the row address is not being set correctly. Although the results shown are not much use for removing the fixed pattern noise on an image, they can give some measure of the performance of the devices compared to those discussed in Chapter 4. Before the raw FPN data had their mean removed, the values were centred around 30,000, which is the same value the comparators in Chip 3 were compared against. In that case, the chip to chip

tests produced a variation in the counter output of 1200. Looking at Figure 6.26 and accounting for the gradient, it can be seen that the variation from column to column is approximately 300 for the first chip and 200 for the second. These are substantially lower than for the comparator test chips and assuming that all of the variation is caused by the comparators it means that the variation in the trigger voltage has been improved from 60 mV to 19.5 mV, validating the idea of using more advanced layout techniques.



(a) Chip 1

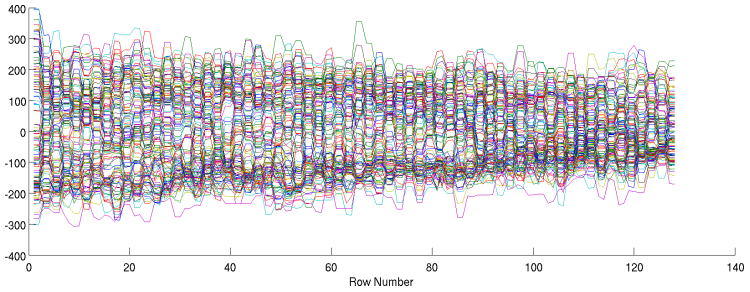


(b) Chip 2

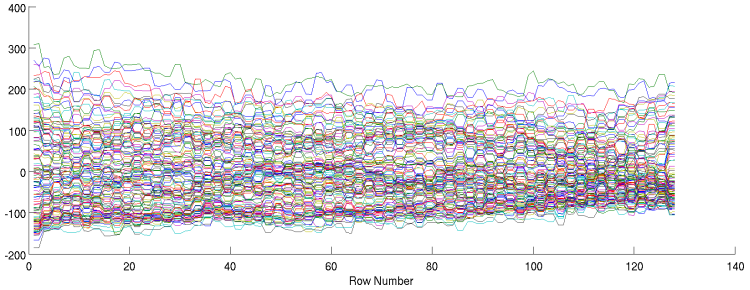
Figure 6.26: Fixed Pattern Noise by Columns

6.8 Summary

This chapter has presented results for the camera chip designed and manufactured. Results for single pixels are given that show that both the time and the inverse time pixels work as predicted in Chapter 3 with regards to the amplitude response, the phase response and the frequency response, although with some changes to the behaviour expected in the case of the changing amplitude response under differing DC light levels. The results show that the pixels have a sharp frequency response, with a full width half maximum of 100 Hz at the modulated light frequency for a



(a) Chip 1



(b) Chip 2

Figure 6.27: Fixed Pattern Noise by Rows

filter length of 4 and a width of 40 Hz for a filter length of 16. Images taken with the camera in DC, two phase pixel and four phase pixel mode have been shown and demonstrate the ability of the camera to pick a small modulated light signal out of a much larger DC background, whilst also being able to show the phase across the image (for the four phase pixel case). It has been demonstrated that the pixels suffer from both temporal and fixed pattern noise, with the fixed pattern noise being most obvious in the columns due to the shared circuitry. Compared to the comparator results in Chapter 4, the fixed variation is greatly reduced. The results show that the camera is a viable choice for modulated light detection.

Chapter 7

Discussion and Conclusions

7.1 Introduction

This chapter provides a summary of the thesis as a whole and provides further discussion of the camera design presented.

7.2 Thesis Summary

Chapter 1 provided an introduction to the area of light detection using CMOS sensors. The different types of detector were introduced along with the different frontend circuits that can be implemented that are used to convert the photocurrent provided by the detector into a voltage for use in the rest of the design. The idea of modulating the intensity of the light that is used as input to an experiment to shift a slowly varying optical signal away from low frequency noise and isolate it from any other signals present such as background light was also presented. The idea of using and efficiently detecting modulated light is the basis of the work presented.

Chapter 2 explored the existing work in the area of modulated light detection and described a number of existing cameras. The cameras described show great variety

in both their techniques for demodulation of the signal but also the types of pixels and frontends used, with the majority of the designs using a logarithmic or traditional integrating frontend and others using slight variants of these designs with, for example the logarithmic pixel modified with feedback or the PMD being two good examples. This chapter also presents some cameras that may be suited to modulated light detection which also serve to highlight why custom modulated light cameras are useful - namely the data rate off chip required to achieve a reasonably high modulation frequency.

Chapter 3 explained why modulation is used in traditional circumstances and introduced the different types of analogue modulation available. The theory of how an amplitude modulated signal is created and demodulated and reconstructed at the receiver was explained. The practicalities of receiving a modulated signal were also considered with regard to ensuring that a signal is correctly sampled with respect to the frequency components that it contains. This was discussed in terms of system design and the advantages and disadvantages of partitioning a camera at different points based on the sampling requirements and the complexity of the subsequent processing. The idea of using a specific sampling frequency that is directly related to the modulation frequency was also presented. This idea allows the sampling of a signal to occur at a frequency that means that the later demodulation is much easier since there is no need to require complicated multiplications any more. The performance of a technique based on these principles that is in use in some of the cameras described in Chapter 2 was carried out and compared to an improved technique. The new technique was shown to be an improvement over the existing technique.

Chapter 4 showed the prototype pixel designs manufactured to explore the realities of the integrating pixel that measures light intensity as a time value rather than a voltage as is the usual method. Three chips were designed, the first including a single pixel, the second including a modified pixel in a 2 x 2 array that could be tiled further as desired. The third chip contained a single comparator design used to evaluate the hysteresis and switching performance of the design when layout techniques were

used to minimise the area of the layout rather than attempt to make each device as well matched to each other as possible.

Chapter 5 introduced the final chip that consisted of a camera with 128 x 128 pixels as well as the digital code that would be used to control the camera on an external FPGA. The implementation of the digital demodulation was discussed as well as the issues that are involved in removing the 1/x nature of the integrating time pixel by using non-linear counter designs.

Chapter 6 presented results obtained from the camera that demonstrate its ability to distinguish modulated light. The characteristics of the pixel were presented along with a number of images taken with the camera, including the image shown in Figure 7.1 that shows a logo illuminated with modulated light in the presence of a strong DC background and the logo remains clear.

Figure 7.1: Modulated Light Image

7.3 Further Discussion

The integrating time pixel has been presented as an alternative to the traditional integrating pixel. The work in Chapter 3 shows the advantage of the time pixel over

the voltage pixel in the range of modulation depths that the pixel is useful over. The Quadrature method of demodulation is superior to the method proposed by Lange, producing a frequency response that is sensitive to fewer of the harmonics of the modulation frequency and has a much lower error when measuring phase.

The use of the comparator to indicate when the photodiode has discharged to the reference voltage means that the period of integration can be very easily changed (by changing the reference voltage) to allow the best response to be achieved. The use of the comparator and counter to carry out the measurement means that the measurements can be carried out in parallel. This is equivalent to using a shutter in a traditional integrating and then quickly sampling all of the pixels with an ADC after the shutter has been disabled. This is both an advantage and a disadvantage in our case since the sampling process itself places the sampled value directly into the demodulation circuit where it needs to be, compared to the traditional case where the ADC must sample the pixel and then transfer the value to the demodulation circuit which is more complicated to achieve. In the traditional case, all pixels in the array can perform their measurements at once and then be sampled afterwards. Each time pixel requires the demodulation circuit and counter information however, so in a column based design as presented it is not possible to sample every pixel at once.

Because the photodiode voltage changes by the same amount for every sample regardless of the light intensity (assuming that the reference voltage is not changed), any non-linearity in the response of the photodiode or the buffer will produce an identical change in every sample. This should mean that the effect will be removed at the demodulation stage since it is effectively a change in the DC. By contrast, the voltage pixel has the advantage of using all of the light across the integration period, whereas the time pixel only uses the light up to the time when the comparator triggers. This is not a significant problem if the DC illumination across the chip is reasonably flat since the reference voltage can be modified so that as much of the integration period is used at once, but this is not really possible if there is a

strong DC gradient. Modifying the reference voltage automatically to produce the best results is also a possibility. This is also possible on a voltage pixel by changing the time at which samples are taken by changing the duration of the shutter, but assuming that both voltage and time pixels would use the same circuit to determine the best integration duration, a simple digital to analogue converter is likely to be much less complicated and prone to errors than changing the timing of the whole array.

In Chapter 3 it was noted that the frame rate of a camera should be chosen to match the bandwidth of the signal being measured, although it may be more convenient to have at least video rate (i.e. around 24 frames per second) to allow easy focusing and arranging of the scene to be photographed.

In the case of the integrating modulated light camera presented, the maximum frame rate is heavily dependent on the modulation frequency because this directly determines the sampling frequency. As the processing for each row is split into two halves (as limited by the number of outputs available on the chip), this means that the maximum frame rate possible is given by Equation 7.1, where ω is the modulation frequency and n is the length of the filter.

$$\text{Maximum frame rate} = \frac{\omega}{256n} \tag{7.1}$$

Assuming a filter of length 4, to achieve a frame rate of 24 frames per second would require a modulation frequency of 24.576 kHz. At low modulation frequencies the frame rate can become very low. For example, in the image shown in Figure 7.1, the modulation frequency used was 210 Hz. This leads to a theoretical maximum frame rate of 0.21 frames per second, which is roughly what was achieved in reality.

By moving to a more advanced (i.e. smaller feature size) process, many of the limitations of the current implementation could be avoided. A smaller feature size allows higher operating frequencies, so a higher speed counter would be possible which allows a greater dynamic range or else a higher maximum operating frequency. A smaller feature size also offers more significant advantages. The work by Kleinfelder

[Kleinfelder et al., 2001] noted in Chapter 2 shows that it is possible to integrate significant amounts of digital circuitry at the pixel level when a sufficiently fine feature size is available. This would result in pixels that are bigger than the current ones, but they would provide the ability to process an entire frame at once, meaning that the frame rate would increase by a factor of 256 (so a frame rate of 24 frames per second with a filter of length 4 could be achieved at just and more importantly the current delay between rows being addressed would be eliminated.

The integrating time pixel has been shown to be a good alternative to the traditional integrating pixel when used for modulated light detection at relatively low modulation frequencies when the modulated light component is very small compared to the DC background. The highest modulation frequency it has been operated at is 2.5 kHz, but as it operates on the same basic principles as the traditional integrating pixel, it should be able to operate at the same modulation frequencies as cameras using that pixel, assuming that a sufficiently high counter frequency is used to be able to provide a useful range of results. For example, the highest frequency CMOS integrating pixel discussed in Chapter 2 works at a modulation frequency of at most 40 kHz. With the 40 MHz clock used for the counter in this work, this would give a maximum counter value of 1000. The Quadrature method of demodulation provides definite advantages over the Lange method of demodulation. The design presented suffers from some limitations that are due to the implementation itself, not the overall idea. Using a more advanced CMOS process would allow these limitations to be removed and a fully integrated camera produced that could output demodulated images directly.

The logical progression of this work is to integrate the whole system on a single chip. This would be best achieved by moving to a finer process as discussed above, so that sufficient digital processing could be integrated to provide a high frame rate, possibly even going as far as placing the digital processing at each pixel or group of pixels, where the processing could be shared between four pixels but be local to the array. An investigation of an automatic or semi-automatic exposure system

may also prove advantageous to this design. By calculating the average DC level of illumination across the chip, it would be possible to determine the proportion of the integration period that was being used. If this was either too high (the pixels are not discharging fully) or too low (the pixels are discharging very quickly), then the voltage trigger level could be adjusted to make better use of the light available.

Bibliography

[Allen and Holberg, 1987] Allen, P. and Holberg, D. (1987). *CMOS Analog Circuit Design*. Oxford University Press. ISBN: 0-19-510720-9.

[AMS, 2004] AMS (2004). 0.35 um CMOS C35 process parameters eng-182. `http://asic.austriamicrosystems.com/`.

[Ando and Kimachi, 1999] Ando, S. and Kimachi, A. (1999). Time-domain correlation image sensor: First CMOS realization of demodulator pixels array. In *Proc. IEEE CCD/AIS*, pages 33–36.

[Ando and Kimachi, 2003] Ando, S. and Kimachi, A. (2003). Correlation image sensor: Two-dimensional matched detection of amplitude-modulated light. *IEEE Trans. Electron Devices*, 50(10):2059–2066.

[Benten et al., 1997] Benten, H. G. P. H., Routsalainen, T., Mäkynen, A., Rahkonen, T. E., and Kopola, H. (1997). Integrated synchronous receiver channel for optical instrumentation applications. In *Sensors, Sensor Systems, and Sensor Data Processing*, volume 3100 of *Proceeding of SPIE EurOpto*, pages 75–88.

[Bourquin et al., 2001] Bourquin, S., Seitz, P., and Salthé, R. P. (2001). Two-dimensional smart detector array for interferometric applications. *Electronics Letters*, 37(15):975–976.

[Chamberlain, 1969] Chamberlain, S. G. (1969). Photosensitivity and scanning of silicon image detector arrays. *IEEE J. Solid-State Circuits*, 4(6):333–342. First APS.

[Chamberlain and Lee, 1984] Chamberlain, S. G. and Lee, J. P. Y. (1984). A novel wide dynamic range silicon photodetector and linear imaging array. *IEEE J. Solid-State Circuits*, 19(1):41–48. First paper on log pixels.

[Comhaire et al., 1992] Comhaire, F., Huysse, S., Hinting, A., Vermeulen, L., and Schoonjans, F. (1992). Objective semen analysis: has the target been reached? *Human Reproduction*, 7(2):237–241.

[Dickinson et al., 1995] Dickinson, A., Ackland, B., Eid, E.-S., Inglis, D., and Fossum, E. R. (1995). A 256 x 256 CMOS active pixel sensor with motion detection. In *Proc. IEEE Int. Solid-State Circuits Conf.*, pages 226–227.

[Dmochowski, 2006] Dmochowski, P. (2006). *CMOS Modulated Light Camera*. PhD thesis, University of Nottingham.

[Dmochowski et al., 2004] Dmochowski, P., Hayes-Gill, B., Clark, M., Crowe, J., Somekh, M., and Morgan, S. (2004). Camera pixel for coherent detection of modulated light. *Electronics Letters*, 40(22):1403–1404.

[Ducros et al., 2002] Ducros, M., Laubscher, M., Bourquin, B. K. S., Lasser, T., and Salathé, R. P. (2002). Parallel optical coherence tomography in scattering samples using a two-dimensional smart-pixel detector array. *Optics Communications*, 202:29–35.

[Europractice, 2006] Europractice (2006). `http://www.europractice-ic.com/`.

[Fowler et al., 2006] Fowler, B., Godrey, M. D., and Mims, S. (2006). Reset noise reduction in capacitive sensors. *IEEE Trans. Circuits and Systems I*, 53(8):1658–1669.

[Gruss et al., 1991] Gruss, A., Carley, L. R., and Kanade, T. (1991). Integrated sensor and range-finding analog signal processor. *IEEE J. Solid-State Circuits*, 26(3):184–191.

[Gulden et al., 2001] Gulden, P., Vossiek, M., Heide, P., and Schwarte, R. (2001). Novel opportunities for optical level gauging and 3-D-imaging with the photoelectronic mixing device. In *IEEE Instrumentation and Measurement Technology Conference*.

[Gulden et al., 2002] Gulden, P., Vossiek, M., Heide, P., and Schwarte, R. (2002). Novel opportunities for optical level gauging and 3-D-imaging with the photoelectronic mixing device. *IEEE Trans. Instrumentation and Measurement*, 51(4):679–684.

[Hastings, 2001] Hastings, A. (2001). *The Art of Analog Layout*. Prentice-Hall, 1 edition. ISBN: 0-13-087061-7.

[He, 2008] He, D. (2008). *Full Field Laser Doppler Blood Flow Sensor*. PhD thesis, University of Nottingham.

[ITRS, 2007] ITRS (2007). *http://www.itrs.net/*. International Technology Roadmap for Semiconductors.

[Kleinfelder et al., 2001] Kleinfelder, S., Lim, S., Liu, X., and Gamal, A. E. (2001). A 10,000 frames/s CMOS digital pixel sensor. *IEEE J. Solid-State Circuits*, 36(12):2049–2059.

[Kodak, 2007] Kodak (2007). Kaf-3200 product specification. `http://www.kodak.com/US/en/dpq/site/SENSORS/name/KAF-3200_product/show/KAF-3200_productSpecifications`.

[Kongsavatsak, 2005] Kongsavatsak, C. (2005). *Full Field Laser Doppler Blood Flow Camera*. PhD thesis, University of Nottingham.

[Lange and Seitz, 2001] Lange, R. and Seitz, P. (2001). Solid-state time-of-flight range camera. *IEEE J. Quantum Electronics*, 37(37):390–397.

[Lu et al., 1996] Lu, G. N., Chouikha, M. B., Sou, G., and Sedjil, M. (1996). Colour detection using a buried double p-n junction structure implemented in the CMOS process. *Electronics Letters*, 32(6):594–596.

[Lu et al., 2002] Lu, G. N., Sou, G., Aubert, A., Carillo, G., and Mourabit, A. E. (2002). On-chip synchronous detection for CMOS BDJ optical detector. In *Proceedings of SPIE Volume: 4755*, pages 208–217.

[Ma and Chen, 1999] Ma, S.-Y. and Chen, L.-G. (1999). A single-chip CMOS APS camera with direct frame difference output. *IEEE J. Solid-State Circuits*, 34(10):1415–1418.

[Maas, 2005] Maas, S. A. (2005). *Noise in Linear and Nonlinear Circuits*. Artech House.

[Nixon et al., 1996] Nixon, R. H., Kemeny, S. E., Staller, C. O., and Fossum, E. R. (1996). 256 x 256 CMOS active pixel sensor camera-on-a-chip. In *Proc. IEEE Int. Solid-State Circuits Conf.*, pages 178–179.

[Nomura et al., 1998] Nomura, H., Shima, T., Kamashita, A., Ishida, T., and Yoneyama, T. (1998). A 256 x 256 BCAST motion detector with simultaneous video output. In *Proc. IEEE Int. Solid-State Circuits Conf.*, page 17.6.

[OFCOM, 2007] OFCOM (2007). *United Kingdom Frequency Allocation Table*. http://www.ofcom.org.uk/radiocomms/isu/ukfat/ukfat07.pdf.

[Ohta et al., 2003] Ohta, J., Yamamoto, K., Hirai, T., Kagawa, K., Nunoshita, M., Yamada, M., Yamasaki, Y., Sugishita, S., and Watanabe, K. (2003). An image sensor with an in-pixel demodulation function for detecting the intensity of a modulated light signal. *IEEE Trans. Electron Devices*, 50(1):166–172.

[Oike et al., 2002] Oike, Y., Ikeda, M., and Asada, K. (2002). High performance photo detector for modulated lighting. In *Proc. of IEEE International Conference of Sensors*, pages 1456–1461.

[Oike et al., 2003] Oike, Y., Ikeda, M., and Asada, K. (2003). High-performance photo detector for correlative feeble lighting using pixel-parallel sensing. *IEEE Sensors Journal*, 3(5):640–645.

[Orthotech, 2007] Orthotech (2007). http://www.orthotech.uk.com/.

[Perelman and Ginosar, 2001] Perelman, Y. and Ginosar, R. (2001). A low-light-level sensor for medical diagnostic applications. *IEEE J. Solid-State Circuits*, 36(2001):1553–1558.

[Pitter et al., 2003] Pitter, M., Goh, J., Somekh, M., Hayes-Gill, B., Clark, M., and Morgan, S. (2003). Phase-sensitive CMOS photo-circuit array for modulted thermoreflectance measurements. *Electronics Letters*, 39(18):1339–1340.

[Pitter et al., 2004] Pitter, M., Light, R., Somekh, M., Clark, M., and Hayes-Gill, B. (2004). Dual-phase synchronous light detection with 64/spl times/64 CMOS modulated light camera. *Electronics Letters*, 40(22):1404–1405.

[Povel et al., 1990] Povel, H., Aebersolf, H., and Stenflo, J. O. (1990). Charge-coupled device image sensor as a demodulator in a 2-d polarimeter with a piezoelastic modulator. *Applied Optics*, 29(8):1186–1190.

[Pui, 2004] Pui, B. H. (2004). *CMOS Optical Centroid Processor for an Integrated Shack-Hartmann Wavefront Sensor*. PhD thesis, University of Nottingham.

[Pui et al., 2004] Pui, B. H., Hayes-Gill, B., Clark, M., Somekh, M. G., See, C. W., Morgan, S., and Ng, A. (2004). Integration of a photodiode array and centroid processing on a single cmos chip for a real-time shack-hartmann wavefront sensor. *IEEE Sensors Journal*, 4(6):787–794.

[Razavi, 2001] Razavi, B. (2001). *Design of Analog CMOS Integrated Circuits*. McGraw-Hill.

[Sato et al., 1987] Sato, Y., Araki, K., and Parthasarathy, S. (1987). High speed rangefinder. *Optics, Illumination and Image Sensing for Machine Vision II, SPIE*, 850:184–188.

[Schanz et al., 2000] Schanz, M., Nitta, C., Bußmann, A., Hosticka, B., and Wertheimer, R. (2000). A high-dynamic-range CMOS image sensor for automotive applications. *IEEE J. Solid-State Circuits*, 35(7):932–938.

[Schuster and Strull, 1966] Schuster, M. A. and Strull, G. (1966). A monolithic mosaic of photon sensors for solid-state imaging applications. *IEEE Trans. Electron Devices*, 13(12):907–912. One of the first silicon sensors.

[Schwarte, 1999] Schwarte, R. (1999). Phase and amplitude detector of electromagnetic waves from pixels operating in visible, UV and IR wavelengths. Patent.

[Schwarte et al., 1997] Schwarte, R., Xu, Z., Olk, J., Klein, R., Buxbaum, B., Fischer, H., and Schulte, J. (1997). A new electrooptical mixing and correlating sensor: Facilities and applications of the photonic mixer device (PMD). In *Sensors, Sensor Systems, and Data Processing*, volume 3100 of *Proceedings of SPIE*, pages 245–253. The International Society for Optical Engineering (SPIE).

[Sharples et al., 2007] Sharples, S. D., Clark, M., and Somekh, M. G. (2007). Adaptive acoustics: correcting for aberration in materials with microstructure. In *Ultrasound and Lasers*, International Congress on Ultrasonics.

[Signal Recovery, 2008] Signal Recovery (2008). Model 7124 specifications. `http://www.signalrecovery.com/7124specifications.htm`.

[Spirig et al., 1997] Spirig, T., Marley, M., and Seitz, P. (1997). The multitap lock-in CCD with offset subtraction. *IEEE Trans. Electron Devices*, 44(10):1643–1647.

[Spirig et al., 1995] Spirig, T., Seitz, P., Vietze, O., and Heitger, F. (1995). The lock-in CCD - two-dimensional synchronous detection of light. *IEEE J. Solid-State Circuits*, 31(9):1705–1708.

[Stanwick-Weather, 2007] Stanwick-Weather (2007). Solar intensity. `http://www.stanwickweather.org.uk/`.

[Stenflo and Povel, 1985] Stenflo, J. O. and Povel, H. (1985). Astronomical polarimeter with 2-d detector arrays. *Applied Optics*, 24(22):3893–3989.

[Stockford et al., 2007] Stockford, I. M., Modha, K. N., Light, R., Clark, M., Pitter, M., and Hayes-Gill, B. (2007). A prototype custom cmos sensor for fast, low cost wavefront sensing. In *6th International Workshop on Adaptive Optics for Industry and Medicine*.

[Stoppa et al., 2002] Stoppa, D., Simoni, A., Gonzo, L., Gottardi, M., and Betta, G.-F. D. (2002). Novel CMOS image sensor with a 132-dB dynamic range. *IEEE J. Solid-State Circuits*, 37(12):1846–1852.

[Sze and Ng, 2007] Sze, S. M. and Ng, K. K. (2007). *Physics of Semiconductor Devices*. John Wiley and Sons, Inc., 3 edition.

[Weckler, 1967] Weckler, G. P. (1967). Operation of p-n junction photodetectors in a photon flux integrating mode. *IEEE J. Solid-State Circuits*, 2:65–73.

[Woolhouse, 1986] Woolhouse, J. (1986). Measurement of sperm motility. US Patent 4,601,578.

[Xilinx, 2004] Xilinx (2004). *http://direct.xilinx.com/bvdocs/publications/ds099.pdf*. Xilinx.

[Zhu et al., 2006] Zhu, Y., Hayes-Gill, B. R., Morgan, S. P., and Hoang, N. C. (2006). An fpga based generic prototyping platform employed in a CMOS laser doppler blood flow camera. In *IEEE International Conference on Field Programmable Technology*, pages 281–284.

Appendix A

Matlab Pixel Models

A.1 Voltage Pixel

```
function pd_voltage=discharge_voltage(start_voltage, end_time,
        pd_width, pd_length, input, T_s)
    %function time=discharge_voltage(start_voltage, end_time,
    %   pd_width, pd_length, input, T_s)
    %
    % Finds the voltage an nwell diode width*length discharges to
    % from start_voltage after end_time amount of time based on
    % input current.
    % T_s is the time step of input.
    % The returned value is a voltage.

    leakage = nwelldiodeleakage(pd_width, pd_length);
    leakage_step = T_s * leakage;

    % 1 Coulomb = 1A/s
    % charge step = i(t) / T_s;
```

```
% Q=CV
l=length(input);
pd_voltage = start_voltage;
t = 0;
for i=1:l,
    charge_step = input(i) * T_s;
    capacitance = nwelldiodecap(pd_width, pd_length, pd_voltage);
    charge = capacitance * pd_voltage;
    charge = charge - charge_step; - leakage_step;
    pd_voltage = charge / capacitance;
    t = t + T_s;
    if(pd_voltage < 0)
        pd_voltage = 0;
        return;
    end
    if(t >= end_time)
        pd_voltage = abs(pd_voltage);
        return;
    end
end
error('discharge_voltage() input not long enough');
```

A.2 Time Pixel

```
function time=discharge_time(start_voltage, end_voltage,
        pd_width, pd_length, input, T_s)
    %function time=discharge_time(start_voltage, end_voltage,
    %   width, length, input, T_s)
    %
    % Finds the time for an nwell diode width*length to discharge
    % from start_voltage to end_voltage based on input current.
    % T_s is the time step of input.
    % The returned time is an actual time (based on T_s).
    leakage_step = T_s * nwelldiodeleakage(pd_width, pd_length);
    % 1 Coulomb = 1A/s
    % charge step = i(t) / T_s;
    % Q=CV
    l=length(input);
    i=1;
    pd_voltage = start_voltage;
    for i=1:l,
        capacitance = nwelldiodecap(pd_width, pd_length, pd_voltage);
        charge = capacitance * pd_voltage;
        charge_step = input(i) * T_s;
        charge = charge - charge_step - leakage_step;
        pd_voltage = charge / capacitance;
        if(pd_voltage <= end_voltage)
            time = T_s *i;
            return;
        end
    end
    time = T_s * i;
```

A.3 N-Well Diode Leakage

```
function cur=nwelldiodeleakage(Width, Length)
    % cap=nwelldiodeleakage( Width, Length )
    % Return NWell diode leakage current at 3.3V for a given width
    % and height
    JSNW = 2.8; % fA/um2
    JSSWNW = 7.6; % fA/um
    cur = JSNW*Width*Length + 2*JSSWNW*(Width+Length);
    cur = cur * 1e-18; % Convert to Amps from atto Amps
```

A.4 N-Well Diode Capacitance

```
function cap=nwelldiodecap( Width, Length, Voltage )
    % cap=nwelldiodecap( Width, Length, Voltage )
    % Return NWell diode capacitance for a given width, height
    % (both in um) and voltage
    CJNW = 0.08; % fF/um2
    MJNW = 0.39;
    PBNW = 0.53; % V
    CJSWNW = 0.51; % fF/um
    MJSWNW = 0.27;
    C_area = (Width*Length*CJNW)/((1+Voltage/PBNW)^MJNW);
    C_periph = (2*CJSWNW*(Width+Length))/((1+Voltage/PBNW)^MJSWNW);
    cap = C_area + C_periph;
    cap = cap * 1e-15; % Convert to farads from fF.
```