Avalanche Photodiode Image Sensing in Standard Silicon BiCMOS Technology

A DISSERTATION

submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH

for the degree of

DOCTOR OF TECHNICAL SCIENCES

presented by

Alice Irene Biber

Master of Science in Electrical Engineering
Massachusetts Institute of Technology

born 11. September 1967
citizen of Zürich, Switzerland

accepted on the recommendation of

Prof. Dr. H. Jäckel, examiner
Prof. Dr. P. Seitz, co-examiner

2000
# Table of contents

Abstract  ........................................................................................................ iv  
Zusammenfassung ......................................................................................... vii  

1 Introduction ............................................................................................... 1  
  1.1 Organization .......................................................................................... 5  

2 Photodiodes and Avalanche Photodiodes .................................................. 7  
  2.1 Introduction to silicon image sensing ...................................................... 7  
  2.2 Photodiodes in silicon .......................................................................... 8  
    2.2.1 Properties of light ........................................................................ 8  
    2.2.2 Photon absorption in silicon ......................................................... 9  
  2.3 Avalanche photodiodes in silicon .......................................................... 11  
    2.3.1 Impact ionization and the breakdown condition ......................... 13  
    2.3.2 Critical field calculation ............................................................... 15  
    2.3.3 Breakdown voltage calculation ..................................................... 18  
    2.3.4 Avalanche multiplication ............................................................. 19  
  2.4 Noise theory in APDs .......................................................................... 21  
  2.5 Conventional APDs ............................................................................ 24  
    2.5.1 Beveled edge diodes .................................................................. 24  
    2.5.2 Reachthrough diodes .................................................................. 25  
    2.5.3 Guard ring diodes ...................................................................... 25  
  2.6 Modes of operation ............................................................................ 26  
    2.6.1 Geiger mode ............................................................................... 26  
    2.6.2 Sub-Geiger mode ...................................................................... 27  
  2.7 Conclusions ......................................................................................... 28  

3 APD Design in BiCMOS Technology ......................................................... 30  
  3.1 Technology considerations ................................................................... 30  
  3.2 Design considerations ......................................................................... 32  
    3.2.1 Geometry and size ..................................................................... 34  
    3.2.2 Quantum efficiency ................................................................... 35  
    3.2.3 High voltage capability ............................................................... 36  
    3.2.4 Capacitance estimate .................................................................. 37  
  3.3 Implemented device designs ................................................................ 37  
    3.3.1 CMOS compatible devices .......................................................... 38  
    3.3.2 BiCMOS or CCD compatible devices ....................................... 40  
  3.4 Conclusions ......................................................................................... 43
4 APD Experimental Results ...................................................... 45
  4.1 Quantum efficiency ............................................................ 45
  4.2 Current vs. voltage measurements ........................................ 46
    4.2.1 Measurement setup ....................................................... 46
    4.2.2 Breakdown measurement conditions .................................. 47
    4.2.3 Dark measurements ...................................................... 47
    4.2.4 Light measurements .................................................... 50
    4.2.5 Gain calculation ......................................................... 52
    4.2.6 Long term operation .................................................... 58
  4.3 Noise of APDs ................................................................. 60
  4.4 Conclusions ...................................................................... 64

5 APD Pixel and Array Design Considerations ............................ 67
  5.1 Introduction ....................................................................... 67
  5.2 APD Control ....................................................................... 68
    5.2.1 Current Control .............................................................. 68
    5.2.2 Voltage Control ............................................................. 69
  5.3 Integrating vs. Non-integrating sensors ................................ 71
  5.4 Array based on universal APS frame ..................................... 73
    5.4.1 Readout ..................................................................... 73
    5.4.2 Addressing .................................................................. 77
  5.5 Pixel design ....................................................................... 78
    5.5.1 APD control and readout ................................................. 79
    5.5.2 Transimpedance amplifier ............................................... 79
    5.5.3 Noise analysis of transimpedance amplifier ....................... 83
    5.5.4 Programmability and biasing .......................................... 87
    5.5.5 Size constraints and layout ............................................. 89
  5.6 Conclusions ..................................................................... 92

6 APD Array Experimental Results ........................................... 93
  6.1 Introduction ....................................................................... 93
  6.2 Pixel characterization .......................................................... 93
    6.2.1 Diode ........................................................................ 93
    6.2.2 Amplifier ..................................................................... 94
    6.2.3 Complete pixel ............................................................. 100
  6.3 Noise analysis ................................................................... 101
    6.3.1 Spectral electronics noise (no avalanche) ......................... 101
    6.3.2 Total noise (electronics and APD) .................................... 103
  6.4 APD camera realization ......................................................... 105
    6.4.1 Operation without extensive calibration ......................... 105
    6.4.2 Operation with extensive calibration ............................... 108
6.4.2.1 Method ................................................................. 109
6.4.2.2 Hardware and software .......................................... 109
6.3.2.3 Results .............................................................. 110
6.5 Dynamic range .......................................................... 110
6.6 Conclusions .............................................................. 111

7 Comparison with other imagers ........................................... 113
  7.1 Noise ................................................................. 113
  7.2 Dynamic range ........................................................ 122
  7.3 Circuit complexity .................................................... 123
  7.4 Applications .......................................................... 124
  7.5 Conclusions ........................................................... 125

8 Improvements and Conclusions .......................................... 128
  8.1 Improvements .......................................................... 128
    8.1.1 Diode improvements .............................................. 128
    8.1.2 General circuit improvements ................................ 128
    8.1.3 High speed improvements ...................................... 129
    8.1.4 Low light level improvements ................................. 129
  8.2 Influence of technology scaling on CMOS APD arrays ......... 131
  8.3 The future of CMOS APD imaging .................................. 134

Appendix A ........................................................................... 137
    References ................................................................. 137
    Acknowledgments ......................................................... 144
    Curriculum vitae ......................................................... 145
Abstract

In recent years, CMOS active pixel sensors (APS) have begun to establish themselves in the imaging market compared to CCDs due to the lower cost of standard technologies, the lower power consumption and the additional combination with already established analog and digital circuitry on the same chip. However, key specifications in state-of-the-art CMOS images have not been improved over CCD image sensors. One of these specifications is the detection limit of these image sensors. Physically, the limiting noise factor should be the Poisson noise distribution of the photon flux, yet in practical devices the limiting noise factor is noise due to the electronics. To improve the situation, several methods are described in literature. Cooling the device, decreasing the readout speed and averaging all improve the noise behavior of the APS circuits. However, the more radical approach is to increase the number of photons at the input by some sort of gain mechanism.

In this thesis, this gain method is investigated as an alternative to the above described noise reduction techniques. To increase the number of photons, a gain mechanism called avalanche multiplication is used. This gain is internal to the photodetecting device and controlled by a higher applied voltage across the photodiode. Avalanche photodiodes (APDs) have been commercially available for 20 years, normally being produced in a specialized process to control the doping concentrations and dimensions very precisely. Since these specialized processes do not allow electronic circuitry to be combined with the APDs monolithically, realizing very sensitive image sensors with APDs has not been possible until now. This work investigates the possibility and conditions using standard commercially available CMOS compatible technologies to realize more sensitive image sensors using avalanche gain. Specific CMOS circuitry within each pixel is used to control the APD and achieve an external, stable programmable avalanche gain.

The first portion of this thesis investigates the design, implementation and characterization of avalanche photodiodes in standard CMOS and BiCMOS technologies. To make sure that the physical phenomenon avalanche dominates over tunneling, the electric field and the applied voltage must be controlled appropriately. The applied voltage of 10-100V guarantees avalanche to dominate. Stabilizing such voltages is possible with CMOS circuitry but requires the avalanche diodes to be implemented
using BiCMOS layers. Thus, the controlling circuitry is standard low voltage circuitry.

Various APD structures were implemented in 1.2 \( \mu \)m and 2.0 \( \mu \)m BiCMOS technologies and characterized in terms of quantum efficiency, breakdown voltage, gain and noise. Multiplication factors of more than 1000 have been achieved. Noise measurements however have shown that the APDs in the BiCMOS process have a high excess noise factor even at small gains (<10). This excess noise is characteristic of the device technology and can not be improved through circuit design or device geometry.

To keep the diodes within the specified operating region, the programmable applied voltage must be stable to less than 30 mV for 10% gain stability. Simultaneously, the photocurrent in each diode must be converted (linearly) into a voltage which can then be read out sequentially. The circuitry designed to fulfill these criteria is a transimpedance amplifier with a programmable feedback resistor. Thus, a minimum sized operational amplifier is implemented in each pixel. The goal of a two dimensional design is to implement the largest number of pixels possible (e.g. small area pixels) and to have the largest fill factor possible (e.g. most light sensitive area). In fulfilling these two requirements, a compromise between circuit complexity, fill factor and area must be achieved. The implemented transimpedance pixel has an area of ca. 50 times the feature size of the technology, yet a fill factor of 20% is achieved.

The electronic characterization of the pixel circuitry shows a -3 dB bandwidth of 30 kHz, a transimpedance of 60 M\( \Omega \) and the measured noise voltage at the output of 1.0 \( \mu \)V/\( \sqrt{\text{Hz}} \).

These results lead to the first monolithic, operational APD image sensor in standard 1.2 \( \mu \)m BiCMOS technology with 12 by 24 pixels on a chip area of 2.4 mm by 2.4 mm. With this APD image sensor, a complete camera with stable and variable gain has been developed. The first images show adjustable, programmable gains of up to 10. Thus, the camera’s images prove that the monolithic integration of stable, functional APD image sensors in standard BiCMOS technology are feasible.

The opto-electronic characterization of the APD image sensor showed that a sensitivity equivalent to state of the art CMOS sensors is achieved. The desired result of improved sensitivity has not been reached due to the high excess noise of the APDs. Reducing this noise through technological
changes is also addressed. Submicron processes will not improve APD image sensor sensitivity. However, process changes (such as N-substrate and modified passivation techniques) would be necessary for noise improvements. As soon as such a process is available, the results of this thesis can be used to implement a monolithic APD image sensor. Such a sensor would exploit the avalanche gain and show its superiority in terms of sensitivity and signal to noise ratio over traditional CCD and CMOS image sensors for low light level image sensing.
Zusammenfassung


Im ersten Teil dieser Arbeit wird untersucht, welche Arten von APDs mit typischen CMOS- und BiCMOS-Prozessen entworfen und realisiert werden können. Um die elektrischen Felder in einer Größenordnung halten zu können, wo der Avalanche Mechanismus nicht durch den Tunneleffekt überdeckt wird, müssen relativ hohe Spannungen von 10-100 V eingesetzt werden. Die Stabilisierung solcher Spannungen mit Standard CMOS-Schaltungstechnik ist möglich, verlangt aber nach APD-Devices, welche nur in BiCMOS-Prozessen realisiert werden können.

Verschiedene solcher APD-Strukturen wurden mit zwei kommerziell erhältlichen BiCMOS-Prozessen in 1.2 µm und 2 µm Technologie realisiert und in Bezug auf ihre spektrale Empfindlichkeit, ihre Multiplikations-Spannungs-Abhängigkeit, ihre langzeit Stabilität und ihr Rauscherhalten charakterisiert. Mit diesen APDs konnten Photostrom-Multiplikationsfaktoren von 1000 und mehr leicht erzielt werden. Rauschmessungen zeigten allerdings, dass die mit den verwendeten BiCMOS-Prozessen realisierten APDs ein sehr hohes Rauschen (excess noise) bereits bei niedrigen Multiplikationsfaktoren von <10 aufweisen, das sich weder durch Design-Verbesserungen noch durch schaltungstechnische Möglichkeiten reduzieren lässt.

Um die Betriebsspannung der einzelnen Pixels im spezifizierten Bereich programmieren und stabil halten zu können, ist eine typische Spannungsconstanz von < 30 mV für 10% Stabilität notwendig. Gleichzeitig muss der (verstärkte) Photostrom linear in eine niederohmige Spannung verwandelt werden, welche dann sequentiell ausgelesen werden kann. Die dazu gewählte Schaltung besteht in jedem Pixel aus einem Transimpedanzverstärker mit einem programmierbaren Feedback-Widerstand. Dazu muss in jedem Pixel ein Operationsverstärker realisiert werden, dessen Fläche möglichst klein sein soll, um die Anzahl Pixel auf dem Bildsensor hoch und die lichtempfindliche Fläche im Pixel möglichst gross halten soll. Es ist uns gelungen, einen schaltungstechnischen Kompromiss zu finden, welcher zu einer Pixelgrösse von etwa 50 Mal der minimalem Featuregrösse bei einem optischen Füllfaktor von immerhin noch 20% führt.

Die elektronische Charakterisierung der Pixelschaltkreise zeigt bei einer -3 dB Bandbreite von 30 kHz, eine Transimpedanz von 60 MΩ und ein Rauschen von 1.0 µV/√Hz, was für den angestrebten Einsatz mit APDs genügend ist.
Aufgrund dieser Resultate konnte der erste monolithische APD-Bildsensor in Standard 1.2 μm BiCMOS Technologie mit 12 x 24 Pixeln auf einer Fläche von 2.4x2.4 mm² integriert werden. Für diesen APD-Bildsensor wurde eine komplette Kameraelektronik gebaut, welche die Programmierung, die Kalibrierung, die Stabilisierung und das Auslesen der Bilddaten ermöglicht. Damit konnten erste Bilder mit wählbaren APD-Verstärkungsfaktoren von < 10 erfolgreich aufgenommen werden, was die grundsätzliche Aussage erlaubt, dass die monolithische Integration von stabil betreibbaren APD-Bildsensoren in Standard BiCMOS-Technologie tatsächlich möglich ist.

1 Introduction

For the past 25 years, solid state image sensing has been practically synonymous with charge coupled device (CCD) imaging technology [Theuwissen95]. Since a CCD process is usually a modification of a MOS or CMOS process, CCD technology can also exploit the excellent properties of silicon in terms of light-to-charge conversion and electronic photocharge detection. Thus, fabricating reliable photosensor lines and matrices with sensitivity curves that encompass the spectral range of human vision, as necessary for good black and white and color cameras, is possible.

Less than a decade ago, the discovery that unmodified CMOS processes can be successfully employed to produce excellent image sensors, rivaling the performance of the best CCDs changed solid state image sensing significantly [Fossum97]. Today, CMOS image sensors not only match the specifications of state-of-the-art CCD image sensors, but do so at a lower cost due to the use of industry standard processes. As a key additional capability, analog and digital circuitry allow the CMOS image sensor to achieve on-chip or intra-pixel functionality, leading to concepts of smart image sensor and smart pixel [Seitz99].

Despite the vast amount of CMOS know-how, improving key specifications in state-of-the-art CMOS images over those of CCD image sensors has not been possible. Sensitivity and dynamic range have remained the same, although users anxiously await substantial improvements for their applications. Inspecting the photodetection chain more closely reveals some reasons for this lack of improvement:

![Photodetection chain diagram]

**Figure 1.1:** Photodetection chain with impinging photon (1) penetrating the semiconductor inactive layers (2), electron hole pair generation (3), electron collection (4), transport (5), storage (6), MOSFET gate (7) and low impedance signal (8).
An incident photon has to be transmitted through protective or inactive layers such as silicon dioxide or silicon nitride before entering the semiconductor bulk. Depending on the wavelength, the photon is absorbed at varying depths and converted into an electron hole pair or photocharge pair. The electron and hole diffuse through the bulk of the semiconductor until they are either recombined or encounter an electric field in which they are swept to a pixel site where they are stored temporarily. From the pixel site, a suitable transport mechanism (often a combination of drift and diffusion in one or more stages) is employed to move the photocharge to a diffusion. The photocharge can then be extracted from the silicon bulk and placed on the gate of a MOSFET for transduction into a low impedance signal, usually a voltage.

In the visible spectral domain, the total quantum efficiency of this photodetection chain is surprisingly high: typically more than half of the incident photons result in a detectable photocharge. In terms of noise the situation is much worse. The inevitable quantum noise of the incident photon flux and the absorption process results in a Poisson noise in the number of temporarily stored photocharges. Although this noise is the physically limiting noise factor in an image sensor, practical devices such as CCDs and CMOS image sensors do not yet achieve this limiting noise. Commercial CCDs achieve ca. 20 electrons rms and CMOS photogate sensors are slightly better, achieving 5-10 electrons rms. Passive photodiodes have noise levels of 250 electrons rms and active pixel sensors have 100 electrons rms. The dominating noise sources are 1/f noise (which can be partially compensated by correlated double sampling) and thermally generated noise. Thermally generated noise in the channel of a MOSFET leads to an input referred charge detection noise at the gate of the MOSFET that is proportional to the gate-source capacitance.

\[
\Delta Q = C \cdot \sqrt{\frac{4kTB}{g_m}}
\]

with absolute temperature \( T \), bandwidth \( B \), Boltzmann’s constant \( k \) and the transconductance \( g_m \) of the MOSFET [Klaassen67].

While other noise sources (reset noise and 1/f noise) can be virtually eliminated by signal processing techniques such as correlated multiple sampling [Hopkinson82], no effective method of reducing thermal channel noise has been found. State-of-the-art CCD and CMOS image sensors
show a typical input referred noise of 30-50 electrons at video frequencies and room temperature. Reducing this limiting noise is only possible with the following measures: cooling (decreasing T), slow readout (decreasing B), averaging (filtering) or lowering the charge storage capacitance (decreasing C).

Several attempts in the literature use one or more of these possibilities to improve the sensitivity of image sensors. Most notable are the Skipper CCD from JPL [Janesick90], the cooled slow scan CCD for astronomy [Blouke83], the double gate FET with sub-fF capacitance [Matsunaga91] or the charge modulation device [Hynecek91]. The charge modulation device (CMD) allows accumulated charge to cause a back gating effect that changes the threshold voltage of the device. Although these devices have found success in niche applications, practicality prevented large volume applications.

Since these measures were not successful in practice, how else can a substantial improvement in image sensor sensitivity be achieved? Inspecting Figure 1.1, contributions (2) to (6) can not be improved since these portions of the chain are already efficient and low noise. The only other possibility is to increase the number of photons at the input by a suitable photon amplification mechanism. In practice, a micro channel plate offers this type of amplification [Saleh91], shown in Figure 1.2.

![Microchannel plate with impinging photon on the photocathode](image)

**Figure 1.2:** Microchannel plate with impinging photon on the photocathode. Microchannels produce a larger number of secondary electrons. These electrons are converted into photons using a scintillating material.
Incident photons are converted into free electrons in a photocathode. The electrons are accelerated in small tubes of a few microns diameter. Each collision with the tube wall produces a larger number of secondary electrons. At the output the multiplied electron number is converted to photons using a suitable scintillating material and the photons are incident on a normal CCD or CMOS photosensor. A similar approach is taken with the silicon intensified targets (SIT) [Saleh91] in which the amplified electrons are incident on the surface of the image sensor.

The major disadvantage of these techniques is the high voltage required (e.g. several hundred volts) to create secondary electron avalanche and the limited quantum efficiency of the photocathode. Typically only 15-20% of the incoming photons are even seen by the photocathode. Thus the total quantum efficiency is reduced by a factor of 5 for the whole process [Saleh91].

Since the reduction of electronic readout noise is difficult, an increased charge signal must be used to improve the signal to noise ratio. External gain mechanisms, as mentioned above, are one option. Internal gain mechanisms are the other option. Such gain mechanisms include avalanching, a form of multiplication of charge carriers within a certain range of electric field, all possible in silicon based diodes [Sze85].

The avalanche effect in silicon has been used extensively in the past 20 years in discrete devices called avalanche photodiodes (APDs) [McIntyre85]. Despite special semiconductor processes to produce these devices, fabrication difficulties concerning device uniformity require proprietary and patented technologies to produce large area (diameter of more than 1 mm) low noise APDs [Koren98]. Still, APDs are used successfully in many low-light applications, offering single photon detection capability and the high total quantum efficiency expected from silicon photodetectors. Unfortunately, APDs are still discrete devices, requiring high performance external electronics to supply the stable high voltage source (typically a few 100 V) and reading out the signal. Although linear arrays of APDs are commercially available, each of the individual APDs must be connected separately to its proper voltage supply and signal readout electronics, resulting in a complex, unwieldy setup, limited (for practical purposes) to a few tens of parallel channels only. This is far from the market’s expectations of a high sensitivity image sensor.
Obviously, to achieve a high sensitivity solid state imaging photosensor a two dimensional array of APDs must be integrated with stabilized high voltage supply and detection electronics. A first attempt by [Mathews95] made use of a CMOS process, modified to meet the requirements of efficient APD integration. This work involved two parts: an array of APDs and the control/photon counting circuit. The APDs were Schottky diodes implemented in a CMOS compatible process with significant process changes. The mode of operation of these diodes was the Geiger mode, allowing pulse detection using a bias voltage above the breakdown voltage. Since our application does not involve pulse detection but focuses more on image detection and includes the restriction of a standard process, only small portions of this work could be utilized for the work presented in this thesis. Though an APD photosensing array was demonstrated, the CMOS process modifications were extensive and his work has not been followed up to date.

Our approach is more extreme, but if successful much more practical and relevant: Use a standard CMOS process, without any modifications to integrate two dimensional photosensing devices that make use of the avalanche effect for increased sensitivity. Each pixel requires apart from the APD itself, proper high voltage stabilization and signal detection circuits. We are aware that APDs integrated in an unmodified CMOS process might have several disadvantages compared to APDs fabricated with an optimized process. Our basic assertion is that these shortcomings (e.g. dark current, gain and homogeneity) will be compensated with suitable intra-pixel and on-chip analog and digital circuitry. A good example of this approach is the APD size: If large area APDs prove to be difficult to integrate due to uniformity problems, we will integrate large arrays of small APDs where uniformity is of no concern, where each APD is operating in its proper optimized working range. The various outputs of the APDs would then be pooled effectively to give a photoresponse that corresponds to one virtually very big APD of unprecedented high uniformity.

### 1.1 Organization

The systematic investigation of CMOS APD arrays is organized as follows: Chapter 2 describes the basics of photodetection and avalanche photodiode operation. Chapter 3 presents BiCMOS avalanche photodiode designs and considerations. Experimental results from these diode designs
are shown in Chapter 4. The pixel and array designs for use with avalanche photodiodes in standard technology are presented in Chapter 5. Chapter 6 shows the pixel and array characterization methods and results. The last chapter, Chapter 7 presents a comparison between this work and other similar sensors in terms of noise, sensitivity, dynamic range and circuit complexity. Due to unfortunate circumstances (no longer offering foundry services) with the foundry producing the chips described in this thesis, not all circuits were implemented. An analysis of the integrating pixel, for example, is presented in Chapter 7, but measurements are not available. Applications, future work and conclusions are in Chapter 8.
2. Photodiodes and Avalanche Photodiodes

The goal of this chapter is to clearly present the theoretical background of avalanche photodiodes. Photodiodes are the basis of image sensing in silicon. Avalanche photodiodes are an extension of traditional pn junctions using much higher reverse bias voltages to produce stronger electric fields within the device. These fields must achieve a critical strength to allow avalanche multiplication to occur. Impact ionization and the breakdown conditions to achieve this multiplication are shown. Additionally, the breakdown voltage can be calculated from the doping concentration and the critical field. Noise theory for avalanche photodiodes is also presented. Conventional diode structures and modes of operation, including advantages and applications complete the background information.

2.1 Introduction to silicon image sensing

A photodetector’s main purpose is to provide an electrical signal proportional to the detected light. The process of providing this electrical signal can be divided into three parts:

- photon to electronic charge conversion
- charge separation, collection and transport
- charge to voltage conversion.

In silicon image sensors, this process begins with a photon impinging on the surface of the semiconductor and penetrating into the semiconductor. If these impinging photons have sufficient energy, electron hole pairs are created. The ratio of the number of electron hole pairs collected to the original number of photons is called the quantum efficiency (QE). For silicon photodiodes with antireflection coating, this quantum efficiency can be close to 100 %. Silicon can be used as an efficient detector of electromagnetic radiation in the wavelength range from 0.1nm to 1100nm.

The electron hole pair generated by the photon is not stationary within the semiconductor. Diffusion allows the pair to move and finally recombine or reach a depletion region. Due to the electric field across this region, the electrons are separated from the holes. The charge is then collected on each side of the depletion region. If necessary, the charge is transported to the conversion node where the next step, charge to voltage conversion takes place.
The charge to voltage conversion is usually accomplished using a high impedance such as a capacitance. The charge is converted into a voltage, a convenient measurement quantity. In terms of sensitivity, noise and dynamic range, this capacitance and the surroundings are critical.

2.2 Photodiodes in silicon

When a semiconductor such as silicon is illuminated, photons are absorbed and electron hole pairs are created. In order to measure how many electron hole pairs are created, the electrons must be separated from the holes before they recombine. An electric field is required to separate these carriers. Such an electric field is always present across a junction’s depletion region. Therefore, a p-n junction is a simple device that may be used for photodetection as shown in Figure 2.1.

![Photodiode p-n junction with a reverse bias voltage.](image)

Figure 2.1: Photodiode p-n junction with a reverse bias voltage.

2.2.1 Properties of light

Light can be described either as an electromagnetic wave, or as a flow of light quanta or photons. The speed of light is constant and the product of the wavelength ($\lambda$) and frequency ($\nu$) as given in Eq. (2.1).

$$c = \lambda \nu$$  \hspace{1cm} (2.1)

The energy of the photon is proportional to its frequency $\nu$ and inversely proportional to its wavelength as shown in Eq. (2.2) with Planck’s constant $h$ (6.623x10^{-34} Js).
\[ E_{ph} = h\nu = \frac{hc}{\lambda} \] (2.2)

The number of photons per unit time \(N_{ph}/t\) is based on the intensity of the light and the energy of each photon \(E_{ph}\). The light intensity is given as an optical power density per area \(P_L\), impinging on a certain area \(A\). Eq. (2.3) shows the relationship between photon flux and power density.

\[ \frac{N_{ph}}{t} = \frac{P_L A}{E_{ph}} = \frac{P_L A \lambda}{hc} \] (2.3)

To determine the number of electrons per second or the current from the photodiode, the quantum efficiency \(\eta\) must be defined. The quantum efficiency expresses the probability that a photon will create an electron hole pair that can be separated and collected before recombination. This factor is a function of the semiconductor, the wavelength as well as the detecting diode. Thus, the electron current is dependent on the photon flux and the quantum efficiency as in Eqs. (2.4) and (2.5). \(N_e/t\) is the number of electrons per unit time. \(I_{ph}\) is the photon current due to the photon flux \((N_{ph}/t)\) and the quantum efficiency.

\[ \frac{N_e}{t} = \eta(\lambda) \frac{N_{ph}}{t} = \eta(\lambda) \frac{\lambda P_L A}{hc} \] (2.4)

\[ I_{ph} = \eta(\lambda) \frac{qN_{ph}}{t} = \eta(\lambda) \frac{q\lambda P_L A}{hc} \] (2.5)

with the electron charge \(q=1.6 \times 10^{-19}\) Coulombs. The quantum efficiency \((\eta(\lambda))\) of a diode is wavelength dependent due to the absorption coefficient and the spectral penetration behavior. Additionally, the surface and the layers covering the diode such as oxide and passivation cause some wavelength dependent interference.

### 2.2.2 Photon absorption in silicon

When light enters the semiconductor, its propagation velocity changes and part of the light is absorbed. The new speed \(c_{mat}\) is given in Eq. (2.6) where \(n\) is the material's refractive index and \(n \geq 1\).
The photon density (as shown in Eq. (2.7)) decreases exponentially with depth in the semiconductor.

\[ N_{ph}(x) = N_{ph0} e^{-\alpha x} \]  

The depth in the semiconductor is represented by \( x \) and \( \alpha \) represents the absorption coefficient which is dependent on wavelength and material as illustrated in Figure 2.3.

The long wavelength cutoff as seen in Figure 2.3 is determined by the bandgap, while the short wavelength cutoff is determined by the charge collection efficiency of electrons and holes near the surface of the semiconductor. For those photons with energies below the bandgap, the material is transparent. Since the penetration depth is inversely related to the absorption coefficient, the higher energy photons do not travel as far as the lower energy photons.

In silicon, penetration depths of light in the visible spectrum range up to about 10 \( \mu \text{m} \). Since CMOS technologies in the range of 0.5 \( \mu \text{m} \) to 5 \( \mu \text{m} \) feature size have junctions within this depth, CMOS can be used conveniently for photosensing. However, the quantum efficiency of the different diodes strongly depends on the depth of the particular photosensing junction and the penetration depth of the light. The theoretical quantum efficiency can be calculated using four material and device parameters: the reflectivity \( R \), a wavelength dependent absorption coefficient \( \alpha \), depletion width \( W \), and diffusion length \( L_p \) of the carriers within the semiconductor device [Sze81].

\[ \eta = (1 - R)(1 - \frac{e^{-\alpha W}}{1 + \alpha L_p}) \]  

The reflectivity is zero for the ideal case. The absorption coefficient values are wavelength dependent and derived from [Spirig97]. The native device uses a depletion width of 1 \( \mu \text{m} \) whereas the well device has a smaller depletion width of 0.5 \( \mu \text{m} \). The diffusion lengths are also different due to the device geometry at 30 \( \mu \text{m} \) and 3 \( \mu \text{m} \) respectively. These values are used for the quantum efficiency shown in Figure 2.2. In reality, the quantum efficiency is only close to 100% when anti-reflective coatings are used. In our case, no special coatings are used. Therefore, a slight
reduction in quantum efficiency is expected. The significant difference between the native diode and the well diode quantum efficiency is due to the loss of carriers created deeper within the semiconductor. This is modeled by the diffusion length of only 3 µm.

Figure 2.2: Theoretical quantum efficiency for two silicon diodes: a native device with depletion width of 7 µm and a diffusion length of 30 µm, and a well device with depletion width of 2 µm and a diffusion length of only 1 µm. The ideal reflectivity of 0 is assumed.

2.3 Avalanche photodiodes in silicon

A photodiode is a p-n junction operated under reverse bias. In such photodiodes, the depletion region’s electric field separates the electrons from the holes, and a current flows. In a p-n junction, the doping profile determines the depth of the junction and the depletion width at a given applied voltage. These two process dependencies can have a large influence on the quantum efficiency and response speed of the photodiode. If the depletion region is very thin, the transit time for electrons and holes is short, thus allowing better high frequency operation. On the other hand, the thin depletion region may have a lower quantum efficiency than a thicker region. These trade-offs should be considered when choosing a specific process.

Avalanche photodiodes are operated under enough reverse bias voltage to enable avalanche multiplication. The internal current gain obtained allows such devices to be useful for low light level applications. The speed of state of the art devices is in the microwave frequency range with dedicated avalanche photodiode technologies [Sze81].
The photo-transduction process in avalanche diodes is the same as in other photodiodes: an impinging photon generates an electron hole pair somewhere in the semiconductor resulting in the conversion of photons to electronic charge. This charge pair must diffuse to the depletion region of a junction before being collected. In the depletion region, there is an electric field that separates the electron from the hole, causing a reverse bias current through the diode. The electric field in an avalanche photodiode is much larger than in a normal p-n photodiode due to the large reverse bias voltage across the junction. The critical field necessary to obtain avalanche multiplication depends on the doping concentration and the geometry of the device, as described in Section 2.3.2.
The carriers entering the depletion region are accelerated by the large electric field. As the carriers are accelerated, a few are able to gain enough energy to cause impact ionization which in turn releases other electron hole pairs to repeat the entire process as can be seen in Figure 2.4. A photo generated electron (labeled 1) is accelerated by the electric field \( E \) and collides with another atom. If a bond is broken upon impact, another electron hole pair (2 and 2’) is created. This process can repeat itself (3 and 3’) and is called avalanche multiplication.

![Energy band diagram under avalanche conditions](image)

**Figure 2.4: Energy band diagram under avalanche conditions**

Therefore, out of one electron, several more electron and hole pairs are split apart, resulting in more electrons or holes to be collected, thus producing the avalanche gain. The output current is \( M \) times larger than the original current that a normal photodiode would create from the same number of incoming photons.

### 2.3.1 Impact ionization and the breakdown condition

When a carrier is accelerated by the electric field, the carrier often collides with the lattice and valence electrons. Such a collision, causing a valence electron to break free of its bond, is called impact ionization. If the ionization coefficients for the material are known, this rate can be predicted. These values represent the ionization probabilities \( \alpha_n \) and \( \alpha_p \) per length for each carrier. The ratio of hole ionization to electron ionization (k) determines whether the avalanche process proceeds in only one direction (k<<1 for electrons in Si), or in both directions (k=1 as in e.g.
GaAs). The carrier with the larger ionization rate achieves a specific multiplication at a smaller electric field, thus at a smaller applied voltage. Physically, if only one carrier ionizes, there is no feedback current from the other carrier to disturb the ionization process by causing unnecessary collisions. Thus, for large but low noise multiplication, it is advantageous to have only one type of carrier that continues avalanching. The reason for this is that carriers flow in one direction only, reducing the number of unwanted, energy wasting collisions. In silicon at room temperature, the primary carrier should be an electron, since the electron ionization probability is larger than that of holes. If both electrons and holes ionize ($k=1$, as can be the case in other materials), device destruction can result from the complete avalanche breakdown [Saleh91].

To calculate the ionization rates the junction width and the maximum field in the junction at breakdown are required. Measuring the junction current as a function of reverse bias voltage is essential to calculate the ionization rates.

According to [Sze66, Grant73], these ionization coefficients can be determined using a set of complementary junctions. In other words, with a wide variety of junctions of different doping concentrations, and junction current vs. reverse bias voltage, as well as maximum field and junction width calculations, the ionization coefficients can be calculated. There are many references to ionization coefficients [Lee64, Sze66, Grant73, Baliga87] and some of the values vary greatly from others due to different measurement methods. The values from [Grant73] are used here because the extraction method considers junctions with larger fields, as those available in industrial CMOS processes due to the doping concentrations. With [Grant73]'s values, shown in Table 2.1, the maximum electric field and junction width at breakdown can be calculated. For doping concentrations $10^{14}$ cm$^{-3} < N_b < 5 \times 10^{18}$ cm$^{-3}$, the ionization coefficients are used to find a junction width that satisfies the breakdown integral condition.

The breakdown voltage is defined as the voltage at which the avalanche multiplication factor ($M$) becomes infinite. To satisfy this condition, Eq. (2.8) must be equal to 1 [Sze81].

$$1 - \frac{1}{M} = \int_0^w \alpha_n(E) \exp[-\int_0^x (\alpha_n(E) - \alpha_p(E)) dx'] dx$$  (2.8)
where $\alpha_n$ and $\alpha_p$ are the ionization rates for electrons and holes respectively, calculated from the ionization coefficients $A$, $b$ and $m$ for silicon as shown in the Table 2.1. $\alpha_n$ is calculated using the empirical fit (e.g. [Chynoweth58, Sze66, Grant73]) of Eq. (2.9) and Eq. (2.10). $E(x)$ is the electric field within the depletion region whose strength is dependent on the background doping $N_b$ and the position $x$ within the depletion region. $\alpha_p$ is calculated similarly, by replacing coefficients for electrons in Table 2.1 with those for holes.

$$\alpha_n = A_n \exp \left\{-b_n / E(x) \right\}^{m_n}$$

$$E(x) = \frac{q}{\varepsilon_s} \cdot N_b \cdot x$$

<table>
<thead>
<tr>
<th>Electrons</th>
<th>Holes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$ ($cm^{-1}$)</td>
<td>$b$ (V/cm)</td>
</tr>
<tr>
<td>6.2x10^5</td>
<td>1.08x10^6</td>
</tr>
</tbody>
</table>

*Table 2.1: Parameters of the Ionization Rate for Silicon [Grant73].*

The values in Table 2.1 are valid for certain electric field strengths only. For electrons, the electric field strength $E$ should be between 2.4x10^5 V/cm and 5.3x10^5 V/cm. For holes, this range is slightly larger, beginning at 2.0x10^5 V/cm and ending at 5.3x10^5 V/cm. Above and below these electric field values, slightly different coefficients are used to calculate the ionization rate $\alpha$ [Grant73]. In the Figure 2.5, ionization rates are shown in comparison to the inverse electric field for silicon, using the ionization rate coefficients from [Grant73].

### 2.3.2 Critical field calculation

One of the most important parameters for many calculations involving APDs is the critical field $E_c$, defined as the electric field at which the APD theoretically shows infinite gain.

The critical field is determined by solving the system of equations (2.8)-(2.10) using empirical values for ionization rates ($\alpha_n$ and $\alpha_p$) derived from measured ionization rate parameters. Since these equations can not be solved analytically, an iterative procedure is required. Similar procedures
have been used by Sze [Sze66] leading to commonly referenced curve in [Sze85]. The ionization rate parameters and measurement structures used in Sze were more appropriate for lower electric field conditions, as typically seen in processes specialized for APD fabrication. Since typical CMOS processes have higher substrate doping concentrations, higher fields are present in APDs fabricated in these processes. As a consequence, ionization rate parameters specifically for higher field APDs developed by [Grant73] are used to determine the critical field behavior for our APDs. The critical field derivation proceeds using the following steps:

![Graph showing ionization rate vs. inverse electric field for Si derived from measurements by [Grant73]. $\alpha_n$ is the ionization coefficient for electrons, $\alpha_p$ is the coefficient for holes. The ionization rate is not constant with electric field strength.]

**Figure 2.5: Ionization rate vs. inverse electric field for Si derived from measurements by [Grant73].** $\alpha_n$ is the ionization coefficient for electrons, $\alpha_p$ is the coefficient for holes. **The ionization rate is not constant with electric field strength.**

The electric field and the depletion region width are calculated by solving Poisson’s equation. To find the depletion width that fulfills the avalanche
breakdown condition, numerical methods have been developed using Mathcad 6.0 [Mathcad].

The goal of this calculation is to solve for the critical field necessary to achieve avalanche multiplication for different doping concentrations. The steps are outlined below:
1. Set doping concentration \( N_b \)
2. Sweep depletion width \( W \) necessary for \( 1/M=1 \)
3. Calculate critical electric field by substituting \( x=W \) into \( E(x) \) equation.
4. Repeat for new doping concentration \( N_b \)

To solve step 2, two integrals are defined in Mathcad 6.0 and shown in Eqs. (2.11) and (2.12).

\[
f(x) = \exp \left( \int_0^y \left( \alpha_n(z) \cdot m_n - \alpha_p(z) \cdot m_p \right) dz \right) \quad (2.11)
\]

\[
g(w) = \int_0^w \alpha_n(x) \cdot m_n \cdot f(x) dx \quad (2.12)
\]

Combining these integrals as in Eq. (2.13), we achieve the integral used in Eq. (2.8).

\[
g(w) = \int_0^w \alpha_n(x) \cdot m_n \cdot \exp \left( \int_0^y \left( \alpha_n(z) \cdot m_n - \alpha_p(z) \cdot m_p \right) dz \right) dx \quad (2.13)
\]

As an example, one calculation is given for a doping value of \( 5 \times 10^{-15} \text{ cm}^{-3} \) and the ionization coefficients shown in Table 2.1. Varying the space charge width from 4 \( \mu \text{m} \) to 5 \( \mu \text{m} \) in 0.1 \( \mu \text{m} \) steps, a graph of the solution to (2.12) is shown in Figure 2.6. In this case, when the width is 4.5276 \( \mu \text{m} \), the integral value is 1.000022. The critical electric field \( E_c \) at this width using Eq. (2.10) is \( 3.44 \times 10^5 \text{ V/cm} \). Such calculations are repeated for many doping concentrations to produce Figure 2.7 of the critical field as a function of the doping level.

This critical field graph is used to estimate the breakdown voltage for a given doping concentration without having to recalculate the integral numerically. The comparison between our calculated values using Grant’s parameters and Sze’s values (estimated from [Sze85]) show a significant difference in critical fields at larger doping concentrations. [Grant73] uses different extraction methods that are more appropriate for higher field diodes than previous works [Grant73]. For this reason, our analyses rely
on our calculated values based on Grant’s higher field parameters instead of Sze’s published graph.

![Graph of g(w) against w(μm)](image)

**Figure 2.6:** Integral for various values of depletion width w.

### 2.3.3 Breakdown voltage calculation

The breakdown voltage is defined as the avalanche voltage for which the gain $M$ tends towards infinity. Calculations with the critical field (Figure 2.7) allow the upper limit of the breakdown voltage to be estimated according to the Eq. (2.14) [Sze81].

$$V_B = \left( -\frac{\varepsilon_s E_c^2}{2q} \right) \left( \frac{1}{N_b} \right)$$  

(2.14)

where $\varepsilon_s$ is the dielectric constant of Silicon (11.9*$\varepsilon_0$), $E_c$ is the critical field, $N_b$ is the background doping concentration (i.e. the lower doping concentration in the abrupt junction) and $q$ is the charge of an electron.

As can be seen from Eq. (2.14), the theoretical breakdown voltage is dependent on the semiconductor doping level as well as the critical field required to reach the avalanche conditions. Microplasma effects, surface leakage, and edge effects like field concentrations can significantly reduce the actual breakdown voltage. Also, as the breakdown voltage decreases below 6 V for silicon and the critical field increases above 10^6 V/cm, the breakdown occurs predominantly by tunneling rather than by avalanche [Sze85]. Without avalanche, no gain can be achieved. These devices involve doping concentrations above 5x10^{17} cm^{-3}. 

The estimated doping concentrations are from a typical 2 μm BiCMOS process and are shown in Table 2.2. For measured breakdown voltages, refer to Chapter 4. The critical field is from Figure 2.7. The breakdown equation is Eq. (2.14).

### 2.3.4 Avalanche multiplication

The multiplication of an avalanche diode can now be calculated using the breakdown voltage as shown in the empirical Eq. (2.15) [Sze81].

\[
M = \frac{I}{I_p} = \frac{1}{1 - \left(\frac{V_R}{V_B}\right)^n}
\]  

(2.15)
where $I$ is the total multiplied current, $I_p$ is the primary (unmultiplied) current, $V_R$ is the reverse bias voltage, and $V_B$ is the breakdown voltage. The exponent $n$ depends on the semiconductor material (for silicon, $2 < n < 6$), the doping profile, device geometry and structure as well as the incident wavelength.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Doping (cm$^{-3}$)</th>
<th>Critical Field (V/cm)</th>
<th>$V_B$ (eq. (2.14))</th>
</tr>
</thead>
<tbody>
<tr>
<td>P sub</td>
<td>$5 \times 10^{14}$</td>
<td>$2.4 \times 10^5$</td>
<td>-380 V</td>
</tr>
<tr>
<td>AntiPunchThrough</td>
<td>$5 \times 10^{16}$</td>
<td>$5.2 \times 10^5$</td>
<td>-17.8 V</td>
</tr>
<tr>
<td>N-well</td>
<td>$1 \times 10^{16}$</td>
<td>$3.9 \times 10^5$</td>
<td>-50 V</td>
</tr>
<tr>
<td>P-base</td>
<td>$1 \times 10^{17}$</td>
<td>$6.3 \times 10^5$</td>
<td>-13 V</td>
</tr>
<tr>
<td>N plus</td>
<td>$1 \times 10^{20}$</td>
<td>$&gt;10 \times 10^5$</td>
<td>-0.03 V</td>
</tr>
</tbody>
</table>

*Table 2.2: Prediction of breakdown voltages using critical field and doping concentration.*

To illustrate this point more precisely, a simulation using the above equation with $V_B=17.8$ V, $n=4.788$ is shown in Figure 2.8. The values presented represent a fit to an actual measurement on a CMOS APD in 1.2 μm technology (see Chapter 6). In Figure 2.9, the slope of this line (dM/dV) is plotted vs. gain M. From this graph, the applied voltage stability can be calculated. At a gain of 10, dM/dV = 30/V meaning that the gain will change by 30 for a one volt change in applied voltage. To achieve a gain variation of less than 10%, the applied voltage must be stabilized to approximately 30 mV.

*Figure 2.8: Simulated gain vs. voltage curve using a breakdown voltage $V_B=17.8$ V and $n=4.788$. The values presented represent a fit to an actual measurement on a CMOS APD in 1.2 μm technology (see Chapter 6).*
Figure 2.9: The slope of Figure 2.8 (dM/dV) vs. gain M. This graph shows how stable the applied voltage must be to achieve a stability of the gain. For example, to achieve 10% gain stability at a gain of 10, the applied voltage must be stable to 30 mV.

2.4 Noise theory in APDs

Before the noise in avalanche photodiodes is analyzed, noise in photodiodes should be considered. In p-n or p-i-n photodiodes the two main sources of noise are dark current noise and quantum noise, both commonly considered shot noise of the photocurrent [Senior85]. The total shot noise current variance is given by Eq. (2.16) where q is the charge of an electron and B is the bandwidth. $I_p$ is the current due to photons and $I_d$ is the dark current due to thermally generated carriers.

$$i_{TS}^2 = 2qB(I_p + I_d) \quad (2.16)$$

When internal gain mechanisms such as avalanche are involved, excess noise due to the randomness of the gain is added to the system. The randomness of this gain exists because not every electron-hole pair generated at a given distance in the depletion region experiences the same gain. The avalanche noise depends on k, the ratio of ionization rate coefficients ($k=\alpha_p/\alpha_n$). If the ratio is small, the avalanche noise should be smaller too. When $\alpha_p=\alpha_n$, each incident photocarrier creates three carriers in the multiplying region: the primary carrier and a secondary hole and electron. Any fluctuation in the number of carriers represents a large percentage change, thus large noise. To minimize noise, semiconductors
with a large difference in $\alpha_p$ and $\alpha_n$ should be used. In silicon, the ratio of $\alpha_p/\alpha_n$ is small but depends strongly on the electric field. For an electric field of $3 \times 10^5$ V/cm the ratio is 0.1. For a field strength of $6 \times 10^5$ V/cm the ratio is 0.5 [Sze81 and Figure 2.5]. To minimize noise in silicon, low electric field at avalanche breakdown to maximize the ratio $k$ and ionization initiated by electrons to minimize the feedback current due to holes are necessary.

The total noise of the avalanche photodiode with gain $M$ is expressed in terms of shot noise and an excess noise factor as shown in Eq. (2.17).

$$\bar{i}_{TS}^2 = 2qB(I_p + I_d) M^2 F(M)$$  \hspace{1cm} (2.17)

The excess avalanche noise factor $F$ is defined as the ratio of the mean square value of the gain $<M^2>$ to the mean value of the gain squared $<M>^2$. Effectively, this is the increase in shot noise over an ideal noiseless multiplier. To compare different APDs, noise and gain measurements are necessary. The noise factor $F$ is calculated as shown in Eq. (2.18) where $<i_m^2>$ is the measured mean square noise current, $<i_{ph}^2>$ is the mean square noise of the primary photocurrent and $M^2$ is the square of the average gain $M$.

$$F(M) = \frac{<i_m^2>}{<i_{ph}^2> \cdot M^2}$$  \hspace{1cm} (2.18)

The value of the excess avalanche noise factor $F$ (Eq. (2.19)) depends on the device material, shape and which carrier initiates the avalanche. When the ionization is initiated by electrons, the excess noise factor is determined by Eq. (2.19).

$$F_e(M) = M \left[ \frac{\alpha_p}{\alpha_n} \right] + \left[ 2 - \frac{1}{M} \right] \left[ 1 - \frac{\alpha_p}{\alpha_n} \right]$$  \hspace{1cm} (2.19)

For pure electron injection, the maximum noise factor is 2 when $\alpha_n >> \alpha_p$ but on the other hand when $\alpha_n = \alpha_p$, the noise factor is proportional to $M$ [Urgell78, McIntyre66, Webb74, Sze85]. For hole injection, Eq. (2.19) can be rewritten using the ratio $\alpha_n/\alpha_p$ instead of $\alpha_p/\alpha_n$ as in Eq. (2.20).

$$F_h(M) = M \left[ \frac{\alpha_n}{\alpha_p} \right] + \left[ 2 - \frac{1}{M} \right] \left[ 1 - \frac{\alpha_n}{\alpha_p} \right]$$  \hspace{1cm} (2.20)
A graph of Eqs. (2.19) and (2.20) is shown in Figure 2.10 for an $\alpha_p/\alpha_n$ ratio (k) of 0.01.

![Graph showing excess noise factor F vs. Gain]

**Figure 2.10**: Excess noise factor $F$ for 100% electron and 100% hole injection in a silicon device with $k=0.01$ calculated with Eqs. (2.19) and (2.20). The significant difference in electron and hole current noise is due to the ionization rate, which is larger for electrons. Thus the electron multiplication occurs at lower electric fields, with less feedback current to decrease the average gain.

This factor $k$ is used as a fit factor for measurement data. The value of $k$ depends on the detailed electric field profile within the avalanche region and the extent to which the avalanche is initiated by holes. The hole initiated gain is much noisier than the electron initiated gain, due to the $k$ ratio in silicon.

When light is absorbed on both sides of the junction, both electron and hole currents are injected into the multiplying region. The effective noise factor is given by Eq. (2.21) [Webb74].
where \( f = \frac{I_{ne}}{I_{ne} + I_{nh}} \) is the ratio of electron to total injected current. \( M_e \) and \( M_h \) are electron and hole multiplication values and \( F_e \) and \( F_h \) are the respective excess noise factors. The physical reason for increased effective excess noise with mixed injection is simple: the hole gain acts as a negative feedback current within the device, decreasing the output current to reduce the average gain. To achieve a certain average gain with both electrons and holes, the electron gain must be considerably higher than the average gain, thus increasing the excess noise factor.

### 2.5 Conventional APDs

Among the commercially available avalanche photodiodes, the three most common structures are [Webb74]: beveled edge diode, reach through diode and guard ring diode, all pictured in Figure 2.11.

#### 2.5.1 Beveled edge diodes

The beveled edge diode in Figure 2.11a is a p+ n junction. In these diodes, the breakdown voltage is very high at 1800-2600 V. Such high breakdown voltage devices have strong electric field values with the large depletion region, allowing gains of up to several hundred. The depletion region is very wide, typically 250 \( \mu m \) and the junction depth is between 50 and 75 \( \mu m \). The dark current generated in the n region is not multiplied significantly since this current is due to holes, which have a lower ionization rate than electrons. Due to its structure however, the response is slow but with high gain for short wavelengths (below 0.9 \( \mu m \)). The same device has a very fast response but low gain for longer wavelengths (over 0.9 \( \mu m \)) [Webb74].

An example of a beveled edge diode is presented as a large area avalanche photodiode from Advanced Photonix [Koren98]. The edge breakdown due to the large electric field at the pn junction is avoided by beveling the edge at a small angle to reduce the field. Large area devices of up to 15 mm in diameter with uniform avalanche breakdown have been achieved by improving process technology. Device inhomogeneities and defects were
reduced. Lower resistivity silicon is used resulting in breakdown voltages of 2 kV and gains of several hundred. Device operation is near the breakdown voltage, but not above it since the large dark current would lead to thermal runaway.

2.5.2 Reach through diodes

The reach through diode (Figure 2.11b) has been developed extensively by RCA [Webb74]. The high speed and high gain are combined with low noise. The trick in this diode is to separate the depletion region into two different areas: a wide drift region for absorption and a narrow multiplying region for multiplication. The applied voltage causes the depletion region to extend rapidly through the lightly doped n region all the way to the backside p+ contact. The field in this device increases only slowly, also due to the lightly doped region. At 100 V reverse bias it is possible to have a 200 μm depletion region. Arrays of multiple devices (e.g. 4 and 32) have also been created [Webb84].

2.5.3 Guard ring diodes

The last type of common diode noted here is the guard ring diode (Figure 2.11c). This type of diode is often used in dedicated APDs but also because the structure is inherently compatible with planar technologies such as CMOS. This device can be made with n+ on p or p+ on n or even a Schottky barrier [Webb74, Mathews95]. The depth of the junction is fairly small, usually less than 2 μm, but the depletion region can be extended to

\[\text{Figure 2.11: Commercially available silicon avalanche photodiodes (a) beveled edge, (b) reach-through, (c) guard ring structures.}\]
as much as 10 μm, depending on the doping concentration. Thus, this silicon guard ring structure works well in the wavelength range of 0.6 to 0.8 μm. Longer wavelengths penetrate through to an undepleted region and must diffuse to a depletion region to be collected. Electron hole pairs generated in the high field region of the diode are detected more quickly than generated pairs which have to diffuse to the depletion region first. The carriers collected by diffusion are slow and contribute to the low frequency response only. As an example, a high frequency response with 30% quantum efficiency is possible at 0.9 μm, whereas at 1.06 μm the response only has 1-2% quantum efficiency [Webb74].

2.6 Modes of operation

Before considering the monolithic integration of APDs, one must consider the different modes of operation of these devices. Avalanche photodiodes must be biased using a reverse bias voltage. There are two specific modes of operation: applied voltage above or below the breakdown voltage. The Geiger mode requires an applied voltage larger than the breakdown voltage and achieves a gain of about $10^8$ [McIntyre85]. The sub-Geiger mode has an applied voltage just below the breakdown voltage. The photons that undergo a specific gain or more are detected using low noise charge sensitive preamplifier. The gain achieved is significantly smaller than in Geiger mode but offers other advantages as described below [Brown86, Brown87, Brown89].

2.6.1 Geiger mode

To operate an avalanche diode in Geiger mode, the applied voltage must be raised above the breakdown voltage. This is only possible when there are no carriers (electrons or holes) within the depletion region to initiate avalanche. In most cases, this involves cooling the device to reduce the dark current to essentially zero. In this mode, a single photon can trigger avalanche. A quenching circuit is needed to prevent the large current from destructing the device and to generate single pulses by ending an avalanche event. Even when the applied voltage is larger than the breakdown voltage, not every electron triggers avalanche [McIntyre85].

The Geiger mode has some drawbacks such as linearity, dead time, dark count rate, afterpulsing, and light emission. If the sum of the dark current and photo-generated current is sufficiently low, the output of the
photodiode is a series of uniform pulses at a rate proportional to the dark and photocurrent, provided an appropriate quenching circuitry is used [McIntyre99]. During the time when the quenching circuitry re-establishes the applied voltage also known as dead time, the avalanche photodiode is not sensitive to incoming light. At applied voltages higher than the breakdown voltage, thermally generated carriers can also cause output pulses which are indistinguishable from photo generated pulses. The rate of these output pulses is called the dark count rate. The only solution is to cool the detectors to minimize thermally generated carriers.

Additionally, when large currents flow, some carriers are trapped and contribute to so-called after pulsing. When the carriers that are trapped are released, they can contribute to another avalanche pulse that is correlated to the original incident photon. To reduce the afterpulsing effect, the number of charge carriers flowing during breakdown should be limited.

Additionally, when the APDs are biased in the Geiger mode, light emission can occur. In one and two dimensional arrays, generated photons can cause other devices in the array to breakdown in a domino effect. Only by isolating devices optically from one another or changing the mode of operation (e.g. sub-Geiger mode) can this effect be remedied [Brown89].

2.6.2 Sub-Geiger mode

As the name indicates, the sub-Geiger mode is the operation regime with the applied voltage just slightly less than the breakdown voltage. The sub-Geiger mode is motivated by a need for larger quantum efficiency, less afterpulsing, higher photon fluxes and the incentive of a two dimensional array of APDs. The gain is not as high as in the Geiger mode but the interelement isolation problem is less severe. Geiger mode gains of $10^5$ are not achieved, making the traps and afterpulsing much less of a problem. However, there are some penalties to operating in this mode: increased demands on the stability of the controlling electronics and low noise electronics for better detection due to the smaller gains achievable. In general the gains in sub-Geiger mode are on the order of 10-100 rather than $10^5$ as in the Geiger mode. The analogy between analog and digital mode is useful. Analog mode similar to sub-Geiger mode requires well controlled gain. Digital operation as in the Geiger mode requires large gain since only the presence or absence of light is detected.
2.7 Conclusions

This chapter summarized the most important properties of silicon diodes employed as photosensors. By increasing the reverse bias voltage, the photodiodes can be driven into the avalanche region, causing the output current to be a multiple of the photocurrent. This physical effect is exploited to overcome noise limitations of the charge detection circuit.

A key parameter for the description of the avalanche process is the critical field, which is primarily a function of doping concentration. An interative mathematical process is used to calculate this critical field using ionization rate parameters from [Grant73]. Critical field values from this work are compared with published values [Sze85]. The differences are attributed to measurement structures used by [Grant73] which are more appropriate for high electric field structures, such as those available in a standard CMOS process. As expected, the critical field values calculated in this work are larger when compared with Sze’s values at increased doping concentrations. This doping region is important since our devices are fabricated with such concentrations.

Unfortunately, the avalanche effect not only multiplies the photocurrent, but also contributes its own noise component which becomes increasingly dominant at larger gains. The APD noise theory synopsis showed that this excess noise factor can be modeled with a gain dependent noise factor, F.

A summary of traditionally employed APD architectures revealed the approaches necessary to optimize performance: wide charge collection area to increase quantum efficiency, electron multiplication instead of hole multiplication to minimize excess noise and a 3D geometry such that uniform breakdown occurs but only in the regions of interest. The silicon material and processing should be gentle to reduce dark current and homogenize the breakdown over large APD areas available today.

APDs can be operated either slightly below the breakdown voltage in the linear or sub-Geiger mode or above breakdown in the Geiger mode. The Geiger mode requires a quenching circuitry which produces a train of pulses, whose average frequency corresponds to the impinging light level. This mode of operation is used for high-speed, low light level and few-photon operation because of dead time considerations. In our case, speed is not a critical issue, photon numbers are not that low and linear performance is required. Hence, the sub-Geiger mode is more appropriate. A gain in the region of 1-20 is reasonable for a linear gain without saturating the rest.
of the circuitry. The target gain should also overcome the noise of the photodetection circuit. As will be shown in this thesis, the avalanche multiplication does not provide a noiseless gain, which ultimately proves to be the limiting factor in applications using avalanche diodes.
3 APD Design in BiCMOS Technology

The goal of this thesis is to investigate a more sensitive low light level detector using avalanche gain. This goal is divided into two sections: the evaluation of avalanche diodes in CMOS and the integration into a two dimensional image sensor. This chapter considers the design requirements of avalanche photodiodes compatible with the planar CMOS process. The most important restriction in this project is in the technology: APDs must be integrated in a standard process. A 2 μm BiCMOS process has been selected for its flexibility in terms of layers and lower doping concentrations necessary to achieve avalanche multiplication instead of tunneling. Diode design considerations in terms of geometry, size, quantum efficiency and high voltage compatibility are presented. Sixteen implemented diode designs are presented in both standard CMOS and BiCMOS technology. Experimental results are presented in the next Chapter.

3.1 Technology considerations

Our past designs have used a 2 μm CMOS process from Orbit Semiconductor in California, USA [Orbit]. This process was chosen for its CCD/CMOS combination which still seems to be unique in the industry offering multi-project wafers. This technology also offered a BiCMOS option at no extra cost. Additionally, a 0.8 μm CMOS process from Austria Mikro Systeme [AMS] is used for comparison.

Before describing the technologies in more detail, important parameters for avalanche diode design need to be defined. The critical parameters for the APD design are:

• doping concentration
• high voltage capability
• integration density

Additionally, all technologies considered must be capable of fabricating CMOS transistors.

The doping concentrations in a typical 0.8 μm technology are higher than in a typical 2.0 μm technology. The doping level influences the transconductance of transistors as well as the lateral breakthrough of
minimum sized devices. As processes move towards smaller technologies, the doping concentration must be increased with an anti-punch through implant to prevent lateral breakthrough. As seen in Chapter 2, the critical field required for avalanche increases with the doping level. Additionally, with larger fields (> $10^6$ V/cm), breakdown occurs predominantly by tunneling instead of avalanche, thus removing the gain. As feature size of technologies decreases, avalanche multiplication in standard technologies will no longer be possible. Lighter doped layers (<$10^{17}$ cm$^{-3}$) in standard technologies are required for both the lightly doped side of the pn junction and guard rings to prevent edge breakdown. Thus, at the moment, the larger technology (2.0 µm) is favored over smaller technologies (0.8 µm). As described in detail in Chapter 8, feature sizes below 0.25 µm have standard doping concentrations which cause tunneling instead of avalanche to dominate the high field effects. Thus, 0.25 µm and smaller feature sizes can not be used as standard processes for APD implementation.

When dealing with avalanche and high electric fields, higher voltages are present than normally encountered in CMOS circuitry. To withstand such voltages, technology considerations are important. The junction depth, for example, influences the breakdown voltage of a particular diffusion significantly. The diffusions in a process are usually rectangular and therefore contain cylindrical and spherical regions as shown in Figure 3.1.

![Figure 3.1: Cylindrical and spherical regions of a rectangular diffusion](image)

Due to the field crowding at the edges, the spherical regions breakdown first. The cylindrical regions are the second regions to breakdown, again
because the field crowds due to the curvature of the diffusion. Such geometrical considerations are important when designing APDs to withstand larger electric fields and applied voltages.

In our case designs have been integrated in a 0.8 μm CMOS, a 2 μm and a 1.2 μm BiCMOS technology, allowing the geometries of the devices to remain physically alike. The typical parameters are shown in Table 3.1 [AMS, Orbit]. These parameters are similar to those obtained from the particular foundries (AMS and Orbit) but exact values are not publishable due to non-disclosure agreements.

<table>
<thead>
<tr>
<th>Technology</th>
<th>2.0 μm</th>
<th>1.2 μm</th>
<th>0.8 μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxide thickness</td>
<td>40 nm</td>
<td>25 nm</td>
<td>15 nm</td>
</tr>
<tr>
<td>n+, p+ depth</td>
<td>0.6 μm</td>
<td>0.3μm</td>
<td>0.3 μm</td>
</tr>
<tr>
<td>N-well depth</td>
<td>3.0 μm</td>
<td>4.0 μm</td>
<td>3.0 μm</td>
</tr>
<tr>
<td>BVDSS n,p</td>
<td>15V, -15V</td>
<td>13V, -13V</td>
<td>&gt;10V, &lt;-10V</td>
</tr>
</tbody>
</table>

*Table 3.1: Comparison of typical technology parameters*

### 3.2 Design considerations

The project goal is to obtain a more sensitive low light level detector using avalanche gain. This goal is divided into two sections: the evaluation of avalanche diodes in CMOS and the integration into a two dimensional image sensor. The most important restriction in this project is in the technology: APDs must be integrated in a standard process. No process changes or extra layers are required to implement the avalanche photodiodes. The standard CMOS processes are significantly cheaper than specialized processes. The combination detector and electronics as a monolithic chip provides an elegant solution to allow APD arrays to become more popular. APDs in special technologies were developed over 30 years ago and are available commercially today [Sze66, Webb74]. Though arrays of APDs have been presented in papers [Webb84, Trakalo87, Komobuchi90, Zappa97], none have integrated amplifiers. They are also all processed in special technologies or technologies with process changes. To make larger area APDs, the process must be very stable with a low defect density so that no current filaments occur. If larger area diodes are split into several diodes, the requirements on the process in terms of uniformity are reduced, but additional spacing between elements and cross talk are potential problems. On the other hand, integrating an array allows positional information to be gained. This, in
turn, allows an image to be created. The motivation towards a two dimensional array is the positional information, whereas the lesser process requirements are added benefit.

To prove the feasibility of APDs in standard technology, a technology with low cost and high flexibility in terms of layers and different diodes available is chosen. The 2 μm BiCMOS process allows several additional diodes to be implemented than in a CMOS process. The lower doping concentrations, the deeper junction depths and larger radii of curvature compared to sub-micron processes are significant features necessary to prove APD feasibility.

As discussed in Chapter 2, guard ring diodes are easily implemented in planar technologies. Guard rings are used to prevent premature breakdown of cylindrical and spherical regions by having a lower doping concentration and a larger radius of curvature than the original diffusion [Melchior66, Mathews95]. The lower doping concentration allows the depletion width to increase for the same applied voltage. Thus, the peak electric field is smaller. The larger radius of curvature decreases crowding at the edge, thus preventing premature breakdown. The avalanche breakdown is then a bulk rather than a surface effect [Melchior78]. In both the CMOS and BiCMOS processes an n+ p diode with guard ring can be integrated. In addition to this diode, a p+ n-well diode with guard ring can be implemented in the BiCMOS process.

In the device in Figure 3.2 the breakdown occurs first vertically within the semiconductor, not at the sides due to the protective features of the guard ring. The guard rings are implemented as n-well regions. The actual high field region is at the n+ p substrate junction.

Figure 3.2: N+ diffusion with n-well guard ring diode showing vertical breakdown region.
The design considerations are divided into four topics:

- geometry and size of devices
- quantum efficiency
- high voltage capability

### 3.2.1 Geometry and size

Geometrical considerations are important to prevent electric field crowding that can cause breakdown prematurely or in a non uniform fashion. In [Mathews95] rectangular diffusions have shown light emission from the corners. Crowding effects are used by [Snyman98] to create silicon LEDs. The electric field is concentrated to a small volume which emits light if biased properly. This light emission can cause problems for arrays of light sensitive APDs. If one region emits light and the others are still sensitive they may be triggered by the emitted light rather than the incoming light. Such cross talk is reduced by either optical isolation [Mathews95] or by preventing the light emission in the first place.

To avoid crowding regions our designs are implemented using discs to produce a uniform curvature along the entire perimeter of the diode. Additionally, guard rings force the breakdown to be vertically within the semiconductor.

If the device is very small in diameter, the diffusion may not have a flat region, where the avalanche breakdown should occur. Though the design rules do not guarantee flat regions, they are a good indication of the smallest reasonable diffusion dimensions. In our case, the diode must be contacted in the middle of the diffusion of the symmetrical device. Since these contacts are always covered with metal and light insensitive, the minimum dimension is slightly larger than the metal contact. In a 2 μm process, the diameter of the diodes must be at least 7 μm. In most cases, larger diodes show an improved response because the metal area above the diffusion no longer dominates. Maximum dimensions are determined by the ultimate pixel size desired. With a 75 μm square pixel pitch, for a p+n diode with guard ring in 2 μm technology the inner diffusion diameter of ca. 25 μm is possible, but allows no area for the amplifier and controlling circuitry. The final dimensions are largely determined by the design rules of the technology. In our case, the largest implemented diodes are 16 μm in diameter with a diode pitch of 32 μm or 61 μm for n+p and p+n diodes respectively. *Figure 3.3* shows the cross section of both of these diodes.
The striation shows the light sensitive regions of the respective avalanche diodes.

3.2.2 Quantum efficiency

The depth of the collecting junction influences the quantum efficiency. Two diodes with guard rings shown in Figure 3.3 have significant differences in the quantum efficiency due to the n-well layer in the upper diode. This n-well creates another reverse biased diode, collecting charge on its own, thus reducing the charge collected by the p+ n-well diode. Since the n-well depth is 3 µm, the photons with longer wavelengths (e.g. 800 nm) are not collected as efficiently by this diode. The theoretical quantum efficiency differences have been discussed in Chapter 2. In infrared applications this can be a very severe restriction. In other cases where the signal information is in the blue/green range of the visible spectrum, this inherent filtering can be useful. One example of such an application is a flame detector where only the blue and ultra-violet light is of interest [Pauchard98].

Figure 3.3: Possible diodes with guard rings: p+ with p-base guard ring in n-well, n+ with n-well guard ring in p-substrate. The innermost diffusion is a disc, the other diffusions are concentric rings around this disc. The striated regions are light sensitive for the respective avalanche diodes.
3.2.3 High voltage capability

In the 2 μm BiCMOS technology chosen, the n+p- device is a native device and the p+n- device is formed in a n-well. Due to this n-well, the p+n- diode is independent of substrate, allowing more freedom in connecting the diode. The electrical signal can be read at both the low and high voltage nodes (anode and cathode) although the high voltage node will also contain charge from the n-p substrate diode. Fortunately this diode's breakdown voltage is significantly higher than the other diode’s. Therefore only unmultiplied current contributes to the signal.

The n+ p substrate diode’s applied voltage for avalanche is referenced to the substrate voltage because the diode is a native device. Floating the substrate to negative voltages does not reduce the problem because NMOS transistors are also native devices in the same substrate. In an array of diodes, the outputs must be at the cathodes because the substrate is common to all devices, thus shorting all anodes. Since the cathode is the high voltage end, problems occur in the readout circuitry. The difficulty is in making amplifiers high voltage compatible at the input gate. Gate oxides are designed to withstand at most 20 V in typical 2.0 μm technologies. High voltage circuits (e.g. 75 V, 50V) have been developed in standard 5 V CMOS technologies but with significant restrictions on the gate source voltage. The high voltage compatibility on the drain source voltage is achieved by either 2-3 additional masks and implantations to a standard CMOS process or a minor change at the mask-shop level [Declerq93]. These circuits allow the drain to source voltage to be high but the gate to source voltage must be limited to 5 V due to the thin gate oxide in the process [Ballan94, Valencic94]. Such a gate voltage would be the input to a differential amplifier necessary for APD readout. Restricting the gate to substrate voltage to only 5 V causes these high voltage circuits to be inappropriate for our high voltage APD biasing configuration. Thus, high voltage readout circuitry cannot be used.

As a consequence in the technology chosen, only p+n diodes are considered for array implementation where close proximity of electronics and APDs are vital. For device characterization, however, both types of devices are considered for comparison.
3.2.4 Capacitance estimate

The estimation of the diode capacitance is important for the signal processing circuitry. The capacitance depends on the voltage bias of the diode, the geometry of the diode and the doping of the semiconductor. The capacitance can strongly influence the bandwidth of the signal processing circuitry.

Assuming a one sided abrupt junction, the parallel plate capacitance can be approximated. The distance between the plates can be calculated using the depletion width, the doping concentration and the applied voltage. The capacitance per unit area would then be as shown in Eq. (3.1).

\[ C_j = \left( \frac{e_s}{x_d(\phi_{bi}+V_{bs})} \right) \left( \frac{q\varepsilon_sN_b}{p} \right) \]

For a reverse bias voltage of 5 V and typical doping concentrations of a p+ diode in n-well in a 2 μm process, the resulting value is 24 fF for a device of 200 μm² area. This value decreases to 14 fF when the applied voltage is 16 V. For a diode of similar size as the n+ diode in p-, the capacitance changes from 5.8 fF to 3.2 fF when the applied voltage changes from 5 V to 16 V.

3.3 Implemented device designs

To test the feasibility of avalanche in CMOS/BiCMOS, 16 diodes of different possible combinations were implemented for measurement. To achieve some symmetry and uniformity of the conditions at the diode edge, all the diodes are round. The layers used in CMOS are n+, n-well, p+ and p substrate. The extra layer called p-base is available in the BiCMOS option of the technology.

According to avalanche photodiode design considerations, guard rings are necessary to prevent premature breakdown. In this technology two diode types with guard rings are implemented and already shown in Figure 3.3. Other possible technology variations such as Schottky diodes were not considered because they have been investigated elsewhere [Mathews95].

The diodes were characterized for breakdown voltage, dark current, light sensitivity, noise and gain. Experimental results are presented in Chapter
4. The geometry of the designs will be described in the next two sections: CMOS compatible and BiCMOS compatible diodes.

In view of the overall goal of a two dimensional sensor using avalanche diodes, the diodes implemented are fairly small in size, allowing a reasonable resolution for the two dimensional array. The round geometry prevents large fill factors, but simultaneously prevents edge breakdown by creating uniform edge effects. These trade-offs are already considered in the device design here.

3.3.1 CMOS compatible devices

Six of the 16 devices implemented in a first run are CMOS compatible devices. These devices are implemented using 4 standard layers, including n+ diffusion, n-well, p+ diffusion, and p-substrate. They all have similar sizes for comparison, differing only slightly to conform to design rules. All devices include a diffusion ring around the diode to provide a good substrate contact. This ring size is dependent on the complexity of the diode, as can be seen in Figure 3.4.

The simplest device (device 1 in Figure 3.4) is an n+ diffusion disc within the p-substrate. The disc’s size is 7.5 μm in diameter allowing the contact and metal covering to reside completely within the disc. Diode size variations were implemented only on the specially selected diodes (e.g. guard ring diodes). Other simple devices include diodes 2 and 3. They consist of n+ diffusions surrounded by an n-well layer. The n+ diffusion is near the edge of the n-well in these two diodes (2 and 3). In diode 2, the n+ diffusion is just a ring within a disc. In diode 3, the n+ diffusion is exactly the same size as the n-well. The breakdown voltage of the n-well layer can be measured using device 4 which contains only a small n+ diffusion to contact the underlying n-well. The guard ring diode (device 5) is very similar to diode 1 except a ring of n-well with a 1μm overlap is around the n+ diffusion. The last CMOS compatible diode (6) is independent of substrate. A p+ diffusion is implemented within the n-well. The n-well is contacted with a n+ diffusion ring to provide good well contacts. The p+ diffusion is also present for substrate contact. This diode is the only diode in CMOS that can be used independently from the p-substrate.

CMOS devices without guard rings are expected to breakdown either near the surface of the device or at the curvature of the junction. These
different structures can help determine whether the guard rings implemented are effective in preventing premature breakdown. The n-well breakdown can also be determined by measuring these devices. Advantages and disadvantages are listed in Table 3.2.

**Figure 3.4:** Cross sectional views of CMOS compatible diodes: Diode 1 shows a simple n+ diffusion in p substrate. Diodes 2 and 3 are n-well diodes with n+ diffusions near the edge of the n-well. Diode 4 is similar, but used to measure the breakdown voltage of the n-well layer. Diode 5 is the guard ring diode using n-well for the guard ring to the n+ diffusion. Diode 6 is a simple p+ diffusion within the n-well, comparable to Diode 1. Diode 6 is the only diode that is independent of the p substrate.

<table>
<thead>
<tr>
<th>Diode</th>
<th>adv./disadv.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3</td>
<td>+</td>
<td>simple structure, large fill factor</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>edge breakdown</td>
</tr>
<tr>
<td>4, 5</td>
<td>+</td>
<td>reduced edge effects, guard ring</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>larger breakdown voltage</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>independent of substrate</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>smaller quantum efficiency</td>
</tr>
</tbody>
</table>

**Table 3.2:** CMOS diode description advantages and disadvantages.

The n+ p- diodes implemented in CMOS technology all have the same substrate. The cathode of the diodes must therefore be used for further signal processing. Unfortunately, the cathode is at higher voltages than the
substrate when the diode is in reverse bias as in the sub-Geiger mode. The high voltage readout node is a problem for the signal processing integrated on the same chip. For simple diode measurements without integrated amplifiers, the high voltage does not present a problem because the substrate node can be negative, while measuring the cathode at virtual ground. The p+ n- diode on the other hand allows the designer the choice of anode or cathode readout. In our case an anode readout is chosen to simplify further signal processing. For experimental results on APDs without controlling circuitry the high voltage presents no problem.

3.3.2 BiCMOS or CCD compatible devices

The additional 10 devices implemented require layers not present in a pure CMOS process. In the particular process used here, there are two additional layers: p-base and n-bccd. The p-base layer is used to create the base of an npn transistor. The n-bccd is a buried channel implant normally used to create charge coupled devices (CCDs). To the best of our knowledge the n-bccd implant combined with the CMOS layers is not available as an MPW from any other foundry. With two additional layers a total of 10 different diodes can be implemented as shown in Figure 3.5. A comparison to the pure CMOS diodes is also possible.

When comparing BiCMOS diodes with CMOS diodes, several diodes are very similar. Diodes 7 and 8 can be compared with diodes 2 and 3 from Figure 3.4. The n-well is replaced by the n-bccd layer. Diode 9 uses an n-well guard ring to protect the n-bccd diffusion from surface or edge breakdown. Since the n-bccd layer is not standard, these devices were not considered for further implementation. Table 3.3 summarizes advantages and disadvantages of the BiCMOS diodes.

Diode 10 is an n+ diffusion within p-base but directly in p substrate. Diode 11 is similar to diode 3 except that the layers used are differently doped: p+ as opposed to n+, p-base as opposed to n-well. Diode 12 and 14 are very similar to diodes 2 and 3. The diodes 14 and 12 include simple structures to measure the p-base breakdown voltage, both with a small p+ contact in the middle of the device and another device with a p+ ring at the edge of the p-base layer. A similar device (diode 11) uses a p+ diffusion with the same size as the p-base layer. Diode 13 is uses all n doped diffusions to increase the doping of the diode. Diode 14 allows the breakdown voltage of the p-base to be determined similarly to diode 4. Diode 15 is the guard ring diode independent from the substrate. The guard
ring device exploits the p-base layer as a guard ring for the p+ diffusion diode. The last diode (16) includes the npn transistor where the n+ diffusion is placed within the p-base layer in an n-well. For a good npn transistor, the base region should be very thin. Punch-through could be a problem when trying to build up a large electric field. The diodes are implemented for breakdown voltage experiments, but the guard ring diodes are expected to be the most suitable for avalanche photodiode operation.

Figure 3.5: Cross sectional view of BiCMOS compatible diodes: Diode 7 is an n+ diffusion ring within nbccd. Diode 8 is similar to 7 and allows the measurement of nbccd breakdown voltage. Diode 9 is a guard ring diode where the diode is the nbccd layer. Guard ring is the lighter doped n-well. Diode 10 is an n+ diffusion within p-base but directly in the p-substrate. Diode 11 is p+ diffusion within p-base, all within the n-well. Diode 12 is similar except the p+ diffusion is a ring instead of the same size as the p-base. Diode 13 includes all n doped diffusions as the diode doping. Diode 14 is a p+ diffusion within p-base in an n-well. The p-base breakdown voltage can be measured with this diode. Diode 15 is the guard ring diode that is independent from substrate. P-base is used as a lighter doped guard ring. Diode 16 is an npn transistor as a diode.
Table 3.3: BiCMOS diode description advantages and disadvantages.

As shown in Chapter 2 the breakdown voltage can be estimated using the doping concentration and the applied voltage. The n-well breakdown voltage should be higher than both the n-bccd and the p-base breakdown voltage. The guard ring diodes are expected to have a higher breakdown voltage, closer to the estimated breakdown voltage than devices without guard rings. The radius of curvature for devices is not to scale in Figure 3.4 and Figure 3.5.

In addition to these diodes, several other experimental structures were implemented. Since previous experiments have shown that the breakdown voltage of guard ring diodes can vary over time (see Figure 4.6 in Chapter 4), structures to influence surface fields (e.g. field plate structures) were designed.

The concept of field plates using metal contacts is not new [Stillman77, Grove67b]. The field plate should be designed to reduce edge effects and influence the spreading of the depletion region near the surface as seen in Figure 3.6. A positive side effect is that the surface leakage current is significantly reduced. The field plate is a metal plate covering the junction region. A voltage applied to this field plate can influence the shape of the depletion region and thereby alter the breakdown voltage. The field plate is implemented as a gate poly and a metal layer covering the poly as shown in Figure 3.7. Due to the gate poly, the n+ region is only protected by the
inherent lateral diffusion of the n-well region. If the metal could be formed on a thin oxide layer, this would provide a better protection, without reducing the overlap of the n-well with the n+ region.

![Figure 3.6: The field plate voltage influences the depletion region form within the semiconductor. The extent of curvature depends on the applied voltage. In the above figure, both decreasing and increasing the depletion region width at the surface is possible [Grove67b].](image)

![Figure 3.7: Field plate implemented in poly and metal above an n+ n-well guard ring diode. Effective n+ and n+ drawn differ significantly, due to poly layer on top.](image)

### 3.4 Conclusions

A 2 μm technology has been selected over sub-micron technologies due to the lower doping concentrations which allow avalanche multiplication instead of tunneling to occur. An additional benefit of this larger technology is the larger radius of curvature for diffusions when compared with sub-micron technologies. The larger radius decreases the field
crowding at higher applied voltages, thus reducing probability of premature breakdown.

We have selected a BiCMOS process for the flexibility in layers and available pn junctions. The BiCMOS process allows two guard ring type APDs to be implemented and compared with several other diode structures. The geometry of the devices is important to achieve uniform breakdown conditions. This uniformity is implemented in a symmetrical structure using round geometries such as discs and rings. These devices have uniform field crowding conditions at all sides, unlike rectangular geometries. High voltage compatibility considerations are also presented. Standard CMOS designs require a high voltage readout circuitry, which, if integrated on the same chip, presents difficulties.

Standard BiCMOS designs on the other hand present a solution to these difficulties. Substrate independent APD designs allow more freedom in selecting anode or cathode readout circuitry. In addition, guard ring structures are also implementable in standard BiCMOS with low voltage readout circuitry. The minor trade-offs include a reduced quantum efficiency and increased pixel pitch both due to the presence of the n-well. Perhaps a significant trade-off, however, is that the carrier initiating the avalanche in the BiCMOS structures is a hole instead of an electron, causing increased noise. Since the ratio of the ionization rates depends on internal field distribution and is not easily predicted, the extent of this trade-off can only be confirmed using measurements as presented in the next chapter.
4 APD Experimental Results

This chapter presents the experimental results to characterize the APD behavior. The breakdown voltages of all devices were measured but special attention is given to the two guard ring devices, whose breakdown occurs deeper in the semiconductor protected from surface effects and field crowding due to the radii of curvature of the diffusions. Dark current, gain and breakdown voltage are investigated. A systematic breakdown voltage shift has been observed. Long term measurements show a stabilization of the breakdown voltage drift. Potential causes have been investigated using specific experiments.

Additionally, excess noise measurements on both p+n and n+p diodes are presented. Because the doping concentrations and depletion widths are not equal in these two devices, a difference in noise factors is expected. A comparison shows that at a gain of 10 the p+n diode is 10 times noisier than the n+p diode of the same dimensions. Conclusions on the feasibility of APDs in standard BiCMOS technology are presented.

4.1 Quantum efficiency

The absolute quantum efficiency is measured using an Optronic radiation measurement system with a light source, a monochromator and a current measurement. The comparison with a standard reference silicon diode allows the calculation to the absolute quantum efficiency. Two diodes with guard rings (p+ n-well with p-base guard ring, n+ p-substrate with n-well guard ring) were measured. The inner diffusions of these diodes are the same size, thus allowing a direct comparison.

The quantum efficiency has been measured from 300 nm to 1100 nm in 10 nm steps as shown in Figure 4.1. The diodes are at zero bias for the measurements. The quantum efficiency is not strongly bias dependent, so the zero bias measurement is a clear indication of the quantum efficiency at larger bias voltages (without avalanche multiplication). The modulation of the curves is due to thin film interference by the covering layers (oxide, nitride) typical of a BiCMOS process. These measurements show a similar response to other larger diodes in the same Orbit Foresight 2 µm n-well process, with the n+p diode peaking at almost 80% quantum efficiency at 640 nm [Vietze97]. As expected due to the limited absorption depth due to
the n-well, the quantum efficiency of the p+ diode within the n-well is significantly smaller especially for longer wavelengths.

![Graph showing quantum efficiency vs. wavelength for n+ and p+ diodes.](image)

**Figure 4.1**: Absolute quantum efficiency for n+ n-well and p+ p-base diodes measured at zero bias from 300 nm to 1100 nm in 10 nm steps.

4.2 Current vs. voltage measurements

4.2.1 Measurement setup

The setup for current vs. voltage measurements consists of a semiconductor parameter analyzer such as the HP4145B or HP4156A, as well as the HP 16058A test fixture. For very small current measurements (10 fA to 100 fA) such a test fixture is essential, providing proper shielding around the device under test. Additionally, triaxial cables are used to reduce the cable leakage current which would otherwise cause a measurement error.

The goal of these measurements is to determine the breakdown voltage of each of the devices. Then, using both dark and light measurements, gain can be determined for selected devices. The diode analysis is fairly straightforward: voltage sweep while measuring the current. With n+ p-
devices, positive voltage is applied at the cathode and ground is at the anode. Current is measured at the cathode. For p+ n- devices, the p substrate must be grounded; a voltage is applied to the cathode. Current can be measured at the anode. Cathode current in this case is the sum of p+ n- current and n-well p substrate current. The breakdown voltage can be determined using only a voltage sweep while measuring the current. The gain requires at least one current measurement while illuminating the diode with a controlled amount of light.

4.2.2 Breakdown measurement conditions

As mentioned in Chapter 2, the breakdown voltage of a device occurs when the gain tends towards infinity. This point has been defined to be when the current limit of 1 μA is first reached. This approximation is compared with Eq. (4.1) and presents a very small error. \( V_B \) is the breakdown voltage, \( V_R \) is the reverse applied voltage and \( n \) is the exponent to fit the multiplication \( M \).

\[
M = \frac{1}{[1 - \left(\frac{V_R}{V_B}\right)^n]} \tag{4.1}
\]

For one particular diode where the 1 μA current is reached at 17.8 V, an experimental fit to Eq. (4.1) results in \( V_B = 17.8001 \pm 0.06682 \) with an \( n = 4.78823 \). In the literature, the current limit of 1 μA for the breakdown voltage is accepted practice [Chang71].

4.2.3 Dark measurements

In the dark measurements, the breakdown voltage of each device is determined in the absence of light. A list of breakdown voltage measurements is shown in Table 4.1. The measured breakdown voltage is a negative diode voltage, i.e. a reverse bias voltage. The predicted breakdown voltage is based on Eq. (4.2) where the doping concentration \( N_b \) and the critical electric field \( E_c \) are used to estimate the breakdown voltage.

\[
V_B = \left(\frac{-\varepsilon_s \times E_c^2}{2 \times q}\right) \times \left(\frac{1}{N_b}\right) \tag{4.2}
\]
These breakdown voltage estimates do not account for surface or microplasma effects or the field crowding due to curvature of the diffusion. Therefore, some discrepancies between predicted and measured values are expected but have to be explained in more detail. Diode cross sections are shown in Figures 3.6 and 3.7 of Chapter 3. For example diode 1 shows fairly good agreement with the predicted value. The doping concentration used for the prediction is the anti-punch through implant, not the substrate doping concentration since the n+ region is surrounded by the anti-punch through implant. Diodes 2, 3, 4 and 6 all have diffusions that are not protected by a lighter doped diffusion. These diffusions are more likely to breakdown near the surface of the semiconductor where the substrate doping is slightly higher than deeper in the semiconductor. These effects are not considered by the prediction equation. Diode 5 on the other hand, has a bulk breakdown. The substrate doping in 10 μm depth is used for the prediction of the breakdown voltage. The discrepancies can only be explained through the radius of curvature and field crowding since the surface effects are prevented by the n-wells whose breakdown voltage is more than 100 V (see Diode 4). C-V measurements have shown that the doping concentration at 6 μm depletion width is 5 x 10¹⁴ cm⁻³. The depletion width shows an insignificant dependence on applied voltage at this depth. Thus, showing that the doping concentration no longer changes significantly at greater depths.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>(V_B) meas</th>
<th>(V_B) pred</th>
<th>Description</th>
<th>Nr.</th>
<th>(V_B) meas</th>
<th>(V_B) pred</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-19.5</td>
<td>-17.8</td>
<td>n+p- surf.</td>
<td>9</td>
<td>-137</td>
<td>-380</td>
<td>nb p- bulk</td>
</tr>
<tr>
<td>2</td>
<td>-60.8</td>
<td>-380</td>
<td>n+n-p- surf.</td>
<td>10</td>
<td>-5</td>
<td>-13</td>
<td>p+ n- surf.</td>
</tr>
<tr>
<td>3</td>
<td>-62</td>
<td>-380</td>
<td>n+n-p- surf.</td>
<td>11</td>
<td>-30.5</td>
<td>-50</td>
<td>p+ n- surf.</td>
</tr>
<tr>
<td>4</td>
<td>-133</td>
<td>-380</td>
<td>n-p- surf.</td>
<td>12</td>
<td>-31.9</td>
<td>-50</td>
<td>p+ n- surf.</td>
</tr>
<tr>
<td>5</td>
<td>-82</td>
<td>-380</td>
<td>n+p- bulk</td>
<td>13</td>
<td>-180</td>
<td>-380</td>
<td>nb n- p- surf.</td>
</tr>
<tr>
<td>6</td>
<td>-29</td>
<td>-50</td>
<td>p+n- surf.</td>
<td>14</td>
<td>-51.7</td>
<td>-50</td>
<td>pb n- surf.</td>
</tr>
<tr>
<td>7</td>
<td>-20</td>
<td>-17.8</td>
<td>n+ nb p- surf.</td>
<td>15</td>
<td>-46</td>
<td>-50</td>
<td>p+ n- bulk</td>
</tr>
<tr>
<td>8</td>
<td>-86</td>
<td>-380</td>
<td>nb p- surf.</td>
<td>16</td>
<td>&gt;-2</td>
<td>-13</td>
<td>n+ pb surf.</td>
</tr>
</tbody>
</table>

Table 4.1: Measured and predicted breakdown voltages for various diodes in the BiCMOS process. Nr. refers to the diode numbers in Figures 3.6 and 3.7 in Chapter 3. The lightly doped n+ buried channel implant is labeled nb, pb is the lightly doped p-base implant. Surface means the breakdown tends to occur near the surface. Bulk means the breakdown occurs in the bulk. Equation (4.2) is used to calculate the predicted breakdown voltage \(V_B\).
The BiCMOS diodes (diodes 7 through 16) correspond to similar structures in the CMOS diodes. For example, diode 7 is similar to diode 1; diode 8 is like number 4; diode 9 resembles 5 but with a deeper diffusion. Diodes 11 and 12 are similar to 2 and 3 but in n-well, so the breakdown voltage is lower than for diodes 2 and 3. The discrepancy between predicted and measured is due to the surface or curvature effects. Diode 10 is an unprotected n+ diffusion in phase where the doping concentrations are fairly large, causing tunneling to be dominant. Diode 13’s measured breakdown voltage is larger due to the additional n-well doping causing the n-well to expand deeper into the semiconductor. Diode 14 is similar to diode 9 though the prediction is better because the breakdown occurs within the n-well. In diode 9, the breakdown occurs due to the curvature of the n-well guard rings, which is difficult to predict. Diode 15 has a guard ring as does diode 5. Diode 16 has two highly doped materials, causing the breakdown to be due to tunneling current. In this case, the breakdown voltage equation is not valid. In conclusion, the breakdown voltage prediction is only an approximate but important indication of the actual breakdown voltage except in a few cases. The simple equation does not account for non-idealities such as surface effects, microplasmas and field crowding which often occur in simple structures.

Measurements in different technologies have shown, as expected, that the breakdown voltage decreases in smaller technologies. For diode 5 in 2.0 μm, breakdown is at 82 V but in a 0.8 μm technology, this breakdown is at 44 V. The doping concentration has increased from $10^{14}$ to $10^{16}$ cm$^{-3}$. Diode 15 in 2.0 μm and 1.2 μm shows a decrease from 46 V to 18 V. Here the doping concentration change is much smaller ($10^{16}$ to $5 \times 10^{16}$ cm$^{-3}$), but the change in the critical field from $3.9 \times 10^5$ to $5.2 \times 10^5$ V/cm is significant. These numbers are visible in Table 2.2 and Figure 2.6 of Chapter 2.

A typical measurement I-V is shown in Figure 4.2. The positive current direction is in the "reverse" direction. The applied voltage is the "reverse" bias voltage. For comparison, two diode current voltage curves are plotted on the same graph. The breakdown voltage of the p+ diode with p-base guard ring is significantly lower than the breakdown voltage of the n+ diode with n-well guard ring.
4.2.4 Light measurements

As opposed to dark measurements, light measurements are more complex. The photodiodes must be illuminated during the measurement with a controlled amount of light, preferably of a particular wavelength. To compare measurements with dark measurements, a very similar setup is used. The only difference is that the test fixture cover is open to allow light onto the chip. A microscope is used to focus a light spot on a particular diode. Using an LED, a 100 μm pin hole and the 20 x lens of the microscope, the light spot on the chip surface is approximately 5 μm in diameter. The LED is a 635 nm LED from HP (HP HLMP 8150) whose current is controlled to allow a linear response with input voltage.

The microscope used is a Mitutoyo FS60 with 3 different objectives. The most commonly used objective is the 20 x with a numerical aperture of 0.42. In Figure 4.3 the setup for measurement is shown in a block diagram. The test fixture (not shown in Figure 4.3) is placed between the microscope lenses and the movable x-y table, allowing the setup to remain the same for both dark and light measurements. During illumination, the x-y table allows adjustment of the spot position.

![Graph](image_url)

**Figure 4.2:** Reverse bias voltage and measured reverse current for two different diodes. Both diodes are guard ring diodes: dark curve is p+ p-base guard ring diode in n-well, light curve is n+ n-well guard ring diode in substrate.
The low voltage current corresponds to the combination of LED light power and quantum efficiency. To calibrate this measurement a NewPort 840 handheld optical power meter is used to measure the light power at a particular wavelength. With a different optical setup, the quantum efficiency has already been measured, allowing us to use the value for 635 nm. The value can be taken from Figure 4.1 and is 25% for the p+ in n-well diode.

Both a light and dark measurement are shown in Figure 4.4. The dark curve has the smallest current, though the difference is difficult to see near the breakdown voltage. The curve labeled 1.0 V has an LED controlling voltage of 1.0 V. The setup uses the 50 x lens on the microscope.

**Figure 4.3:** Measurement setup with LED, pinhole, microscope and XY table. The light spot is approximately 5 μm in diameter on the image sensor chip using the 100 μm pinhole and the 20 x lens of the microscope. The XY stage is to adjust the chip position with micro-manipulators. The eyepiece can be used to view LED spot and chip simultaneously.
4.2.5 Gain calculation

The gain is defined as the ratio between high voltage current and the low voltage current as seen in Eq.(4.3). In most cases, a sweep of the applied voltage with constant light illumination is necessary to find the gain. The current increases as the applied voltage nears the breakdown voltage. If the dark current is examined, one notices a gain in dark current also. The dark current gain and the light voltage gain are not equal. The dark current increases well before the avalanche multiplication occurs. This increase can be attributed to the larger depletion region (due to the larger applied voltage). Thus when comparing the high voltage current to the low voltage current, the so-called gain increases with slowly applied voltage until the real avalanche gain is present. The photocurrent measurement avoids this situation because the light current is significantly larger than the dark current and the dark current increase.

Figure 4.4: Diode characteristics with different illumination levels. 0.25 nW and 0.15 nW correspond to the light intensities.

The gain measurement shown in Figure 4.5 clearly shows the advantage of avalanche diodes over normal pn diodes. Two measurements are performed: the dark measurement to find the dark current and the
breakdown voltage and the light measurement to calculate the gain. The light measurement uses a calibrated light source to illuminate the sensor. The current is measured as the diode voltage is increased. The ratio of the light current at high voltages to the light current at low voltages is the gain that the photocurrent experiences.

The gain equation is shown in Eq. (4.3) where $I_{phv}$ is the high voltage photon current, $I_{dhv}$ is the high voltage dark current, $I_{plv}$ is the low voltage photon current and $I_{dlv}$ is the low voltage dark current.

$$M = \frac{I_{phv} - I_{dhv}}{I_{plv} - I_{dlv}} \approx \frac{I_{phv}}{I_{plv}}$$  \hspace{1cm} (4.3)

Gain for two diodes (p+ and n+ both with guard rings) is shown in Figure 4.5. The gain is calculated using two measurements, one dark and one light.

**Figure 4.5:** Gain for two diodes: p+ and n+ both with guard rings. The gain is calculated using two measurements, one dark and one light. Multiplication gain of more than 1000 is achieved by both diodes.
Significant gains of more than 1000 are achieved with both diodes. In the measurement, the multiplication gain is limited by the 1 μA current limit, corresponding to a gain of more than 10000.

To show the uniformity of the devices, gain vs. position has been measured with a small light spot of 2 μm. The decrease in gain near position 5 is due to the contact metal covering. The decreases in gain at the edges of the device are also due to the metal covering.

![Graph showing gain vs. position with a cross section of the diode. Metal contact causes decrease in gain at center of device.](image)

**Figure 4.6:** Gain vs. position with a cross section of the diode. Metal contact causes decrease in gain at center of device.

When measuring current vs. voltage curves with APDs, a type of memory effect is observed. When sweeping voltage more than once as shown in Figure 4.7, the voltage at which 1μA of current is reached seems to decrease. This result is independent of light power, but reversible if a longer period of zero applied voltage is allowed. The result depends on the type of diode used. A n+ p- diode tends to have the voltage decrease (as in Figure 4.7) whereas the p+ n- diode has a breakdown voltage increase. Long term operation is investigated in the next section.
Additional devices with field plates as described in Chapter 3 are implemented to investigate this breakdown voltage drift. The field plate devices allow the direction of the breakdown voltage shift to be influenced by the gate voltage. The influence is very direct, influencing the depletion region at the surface. When the gate is biased at a positive voltage, the breakdown voltage of n+p devices increases and p+n- decreases. The guard ring protection of n+p devices is not quite optimal, due to the poly which prevents the n+ or p+ implantation. Thus, the effective n+ region does not overlap with the n-well guard rings as shown in Figure 4.8a. The two regions are simply adjacent, thus the guard ring cannot prevent breakdown due to the curvature of the n+ diffusion. The more ideal field plate structure is shown in Figure 4.8b, where the n-well guard ring overlaps the n+ region more, creating better protection against premature breakdown. Though these devices allow the breakdown voltage to be controlled to a certain extent, neither of the devices prevent the breakdown voltage drift. A constant gate voltage allows the change in the breakdown voltage to be more predictable, though the breakdown voltage still changes with each voltage sweep. The experiments indicate that the breakdown
voltage drift is due to charge accumulation near the surface causing a change in the depletion region. To prove the drift is due to surface charge, further experiments are necessary and described below.

Figure 4.8: Field plate effective diffusions showing that the guard ring in (a) is not ideal because there is no overlap between n+ and n-well to reduce the radius of curvature of the n+ diffusion. Structure (b) shows a larger overlap of the n-well region, thus improving the guard ring.

For the explanation of breakdown voltage drift, four causes have been proposed in the literature. We investigated all of these possible causes with additional experiments to eliminate unlikely effects. The four possible causes are:
1) a thermal effect due to the power dissipation during avalanche
2) charge trapping phenomenon within the bulk
3) surface charge trapping in the oxide causing a change in the depletion region
4) dopant compensation in the pn junction by hydrogen

The reverse breakdown voltage of an avalanching junction is temperature dependent [Grant73, Chang71, Chynoweth60]. The breakdown voltage increases with temperature due to the temperature dependence of the ionization rates, which both decrease with temperature. In the investigated n+p diode, the breakdown voltage shift is in the wrong direction for a thermal effect. Decreasing breakdown voltage with temperature is not characteristic of avalanche junctions [Chynoweth60]. Additionally, the power used in these experiments is not sufficient to create a thermal problem since the maximum current of 1 μA flows for a very short period of time during the measurement. In our case, the n+p diode thermal budget consists of a total of 1 μA at 85 V, corresponding to 85 μW. The p+n diode thermal budget is approximately 45 μW. These thermal quantities are too small to cause this breakdown voltage drift. Thus, temperature effects can be ruled out as the cause for the breakdown voltage drift.
Breakdown voltage variation can also be caused by charge trapping deep in the semiconductor [Haitz65]. The charge distribution of the traps during avalanche are different from the distribution before breakdown due to the increased free carrier concentrations as well as hot carriers due to the larger electric field. For example, in a n+p abrupt junction in the p region, holes can be trapped changing the charge by +q or hot holes can knock either holes or electrons from their position causing either a +q or −q charge change. A +q charge decreases the space charge density which results in an increased breakdown voltage. These holes are trapped for an average time before being released. In steady state, the capture and emission rate must be equal. However, the time constants for such trapped charge are on the order of microseconds at experimental temperatures of 77 K using 17 μm diameter guard ring diodes. At room temperature, this time constant decreases such that practically no variation is observed [Haitz65]. Therefore, the observed experimental drift with a significantly longer time constant must be due to some other phenomenon.

Another possible explanation is trapped charge in the oxide of a high resistivity substrate. The time constant for this trapped charge can be several hours. The inversion layer caused by the effect can create more dark current and seemingly decrease the breakdown voltage [Melchior78, Webb84]. This charge can affect the depletion region’s shape and thus the breakdown voltage [Grove67b]. The exposure to UV light anneals the oxide space charge [McDonald70]. In order to investigate the trapped charge, UV exposure experiments have been performed. Diode breakdown voltages were measured prior and after exposure to UV illumination using an EPROM eraser. Results have shown that UV exposure decreases the dark current (in some cases by a factor of 2) but has little influence on the breakdown voltage drift. The drift continues in the same direction and similar proportions as prior to exposure. Thus, the breakdown voltage drift is not due to the charge in the oxide.

The remaining possible cause is dopant compensation in the pn junction due to the hydrogen release from the Si-SiO₂ interface during avalanche. This phenomenon has been proposed in the degradation of base emitter junctions and the emitter base breakdown voltage (V_{cbo}) induced via avalanche stress methods [Gopi93, Quon94]. Hydrogen can passivate silicon dangling bonds at the interface and reduce interface state density. Under avalanche conditions, energetic carriers bombard the Si-SiO₂ interface generate interface states and free bonded hydrogen. The rate of hydrogen release is proportional to the increase of interface state density.
In an npn transistor, the released hydrogen can then diffuse into the base and pair with boron acceptor atoms, compensating the base dopants, thus decreasing the doping concentration of the base [Gopi93]. The hydrogen at the perimeter moves into the bulk, thus substantially increasing the formation of B-H complexes. Both of these effects result in a widened emitter base depletion width, which in the npn bipolar case causes $V_{eb}$ to increase. The p+n avalanche diode used in this work is very similar to the npn transistor available in this BiCMOS technology. The n-well and pbase layers are used in both structures. This compensation occurs in the pbase and p+ layers which are both doped using boron. The hydrogen causes an acceptor concentration decrease in the pbase and p+ region which in turn increases the depletion width for the same applied voltage, thus decreasing the electric field. Such effects cause the breakdown voltage to increase during measurements. The increase does not continue forever though, since these boron-hydrogen (B-H) bonds can dissociate, releasing the hydrogen further into the bulk. Thus, the breakdown voltage drift is dependent on the rate of hydrogen generation, dopant compensation and dissociation.

Avoiding this type of degradation is possible only by inhibiting the development of a critically large peak depletion electric field, the rapid depassivation of the oxide interface by hot carrier bombardment, extensive diffusion of hydrogen into the bulk or reduction in hydrogen surface passivation itself at the oxide interface [Quon94]. Avoiding the large peak electric fields would mean avoiding the avalanche multiplication, thus is not an option for the reliability of avalanche photodiodes. Thus, for avalanche diodes, several technological problems (e.g. in the oxide passivation) must be solved to reduce the breakdown voltage drift.

### 4.2.6 Long term operation

To measure long term stability, the diode is operated in series with a current source delivering ca. 1 $\mu$A of current as shown in Figure 4.9. The current source is realized using an LM334 with a 68 k$\Omega$ resistor. The high voltage source (Keithley 6517A electrometer / high resistance meter) is set to 56 V and the voltage across the current source is measured. The high voltage source can simultaneously measure the current. When the voltage across the current source is measured using a FLUKE voltmeter, the Keithley shows more current than the 1 $\mu$A of the current source. The internal resistance of the voltmeter (labeled measurement resistor in Figure 4.9) is responsible for this increase in current by creating another current.
path. The actual applied voltage can be calculated by taking the difference of the high voltage source and the measured voltage.

Figure 4.9: Measurement setup for long-term testing. High voltage source is set to 56V, constant current source is 1 µA. The voltage across the current source is measured.

The diode used for the long-term measurement is a p+ diffusion with p-base ring within the n-well. The expected breakdown voltage (from earlier measurements) is ca. 46 V. Figure 4.10 shows the increase in bias voltage for the 1 µA constant current. The measurement was extended to over 2000 hours and showed an increase in breakdown voltage (as defined earlier at 1 µA current) of 14%. The measurement was operated in a normal office environment without special temperature control. The breakdown voltage drift is shown to stabilize, thus preventing the diode's gain from running away completely. This measurement also shows the device to be reliable with more than 2000 hours of continuous operation corresponding well with other avalanche photodiodes that have mean time to failure in excess of 1000 hours at 200 °C and 300 V bias [Melchior78].
Figure 4.10: Long term operation of p+ diode with p-base guard ring. Operation over 2000 hours shows the stabilization of the breakdown voltage.

4.3 Noise of APDs

The noise measurements of APDs require both noise and gain measurements simultaneously. The output current of the APD must be measured for its DC level to determine the gain and for its AC noise to determine the noise. These measurements are done using a current amplifier and a spectrum analyzer. As calibration, small resistors with large current noise are used. The current amplifier (Stanford Research Low Noise Current Preamplifier SR570) converts the current into an output voltage. The bandwidth at a sensitivity setting of 100 nA/V is 2 kHz. The spectral noise measurements were obtained using LED light illumination (3 nW at 635 nm). As can be seen in Figure 4.11 and Figure 4.12, the n+p devices show more significant 1/f noise at lower frequencies when compared to the p+n devices. Experiments light intensity have shown that the 1/f noise is light dependent. Results in Figure 4.11 and Figure 4.12 confirm this dependency.
Due to the smaller quantum efficiency of the light in the p+n diode, the 1/f noise of the p+n diode is expected to be smaller than the respective n+p diode. Experiments using laser light (694 nm, 18.4 nW) and a 0.531 transmission filter at similar avalanche gains of 10 show that 1/f noise
increases significantly with light intensity. This result is much more pronounced in the n+p diode than the p+n diode due in part to the quantum efficiency difference (70% vs. 15%) at the 694 nm wavelength.

The excess noise measurements were obtained for a constant frequency of 175 Hz. The diode was also illuminated with 3.6 nW of 635 nm LED light power. The noise factor vs. gain curves for p+n diode (Figure 4.13) are fit using the hole noise equation $F_h(M)$, Eq. (4.4) [McIntyre66,Webb74].

\[
F_h(M) = M \left( \frac{\alpha_n}{\alpha_p} \right) + \left[ 2 - \frac{1}{M} \right] \left[ 1 - \frac{\alpha_n}{\alpha_p} \right]
\]  

(4.4)

The ratio of $\alpha_n/\alpha_p$ is used as a fit factor as suggested in [McIntyre66]. The fit is best using $\alpha_n/\alpha_p$ of 2000. Such a large ratio is expected of a diode with hole initiated avalanche due to the much smaller ionization rate of holes compared to electrons in silicon.

![Figure 4.13: Noise factor F vs. Gain for a p+n diode in 2 μm BiCMOS technology. The ratio of $\alpha_n/\alpha_p$ is 2000 for best fit of Feff. The diode was illuminated with 3.6 nW of 635 nm LED light power.](Figure 4.13)

In this diode structure, most of the light generating electron hole pairs is absorbed on the n-side of the high field junction. Since the ionization rate of holes is much smaller than the electron ionization rate, the electron gain
must be much larger than the hole gain, causing the noise factor to be large. As can be seen from Figure 4.13, the excess noise factor at a gain of 10 is slightly more than 10000. This value compares the noise of the avalanche multiplication with a noiseless multiplication which has not yet been achieved in silicon devices. Thus, the output noise is 10000 times larger than the shot noise of the diode current.

The n+p- diode (Figure 4.14), on the other hand, has the light impinging on the p-side of the high field junction. Thus, Eq. (4.5) representing the electron noise $F_e(M)$ of an avalanche diode is used. The $\alpha_p/\alpha_n$ ratio is considered the adjustable parameter when comparing noise measurements made on a real diode to the theoretical curves [McIntyre66, Webb74].

$$F_e(M) = M \left[ \frac{\alpha_p}{\alpha_n} \right] + \left[ 2 - \frac{1}{M} \right] \left[ 1 - \frac{\alpha_p}{\alpha_n} \right]$$  \hspace{1cm} (4.5)

Fitting Eq. (4.5) to the measurement data results in an $\alpha_p/\alpha_n$ fit ratio of 100. Such a large ratio indicates that the injection is not pure electron injection, as expected from this structure. Typically, Si diodes with pure electron injection show k ratios ($\alpha_p/\alpha_n$) of less than 0.01.

Adjusting for the electron to total injection ratio ($f = I_{po}/(I_{no}+I_{po})$), Eq. (4.6) can be used [Webb74].

$$F_{eff} = \frac{fM_e^2F_e + (1-f)M_h^2F_h}{[fM_e + (1-f)M_h]^2}$$  \hspace{1cm} (4.6)

This equation combines electron and hole noise equations ($F_e$ and $F_h$) with another fit parameter $f$. The fit, shown in Figure 4.14, is best if the ratio of electron current to total current ($f$) is 0.9 and the ratio $\alpha_p/\alpha_n$ ($k$) is 0.001.

As can be seen from Figure 4.14, the excess noise factor at a gain of 10 is slightly less than 1000, indicating that this diode has more electron injection than the p+n diode. This value compares the noise of the avalanche multiplication with a noiseless multiplication which has not yet been achieved in silicon devices. Thus, the output noise is 1000 times larger than the shot noise of the diode current.

Comparison of these two figures (Figure 4.13 and Figure 4.14) shows that the n+p diode's noise at gain of 10 is 10 times smaller than the p+n's
noise. Thus, for minimum noise, the n+p diode should be chosen. However, this diode requires the readout to be on the high voltage side. Thus, such a diode requires a hybrid solution (i.e. two separate substrates, one for the APD, one for the electronics) to combine the diode with readout electronics. Since the advantages of a monolithic solution include the ability to create larger arrays and readout circuitry on-chip, a hybrid solution is not an option. For a 12 by 24 APD array, at least 288 output lines would have to be used, along with 288 high voltage sources and current to voltage converters to have a hybrid solution for a fairly small array. To avoid the hybrid solution, the p+n diode was chosen, taking the noise degradation as a trade-off to system integration in a standard process technology but allowing integrated controlling and readout electronics.

![Figure 4.14: Noise factor F vs. Gain for a n+p diode in 2 μm BiCMOS technology. The ratio of $\alpha_p/\alpha_n$ is 0.001 and $f$ is 0.9 for best fit of $F_{eff}$. The diode was illuminated with 3.6 nW of 635 nm LED light power.](image)

4.4 Conclusions

The avalanche photodiodes fabricated in BiCMOS technology have been characterized in terms of quantum efficiency, breakdown voltage, gain and noise. Two particular diodes have analyzed and compared in detail: a p+n diode with p-base guard ring and an n+p diode with n-well guard ring. The breakdown voltage of these diodes is ca. 46 V and 82 V respectively.
Gains of more than 1000 have been measured with both types of diodes, proving that avalanche gain in standard technology is possible.

In characterizing these diodes, a drift in the breakdown voltage was observed. To investigate this effect field plate diodes were implemented and characterized. Using the field plate voltage, the direction of the breakdown voltage shift can be influenced. This observation leads to the conclusion that a similar effect causes the breakdown voltage drift. Several other probable causes are investigated using UV illumination and time constant observation. The drift is not due to a thermal effect from the power dissipation because the drift is in the wrong direction. The charge trapping in the bulk is also not the main phenomenon at work because the time constant of microseconds is too small to be visible in our measurements. The main time constant visible is on the order of hours, not seconds. The charge trapping in the oxide is possible, but the UV experiments have neglected to reduce the breakdown voltage shift, suggesting that the real cause for the shift is not within the oxide. The only remaining effect is described as a dopant compensation due to hydrogen. This effect is described in the literature when examining base-emitter junctions of npn transistors. The doping in the phase region of the device is compensated by boron hydrogen bonds, thus changing the depletion region width for the same applied voltage. As the depletion width increases, the maximum field present decreases, causing less gain to occur. Our p+n diode, being similar to an npn transistor, shows the same effect causing a breakdown voltage increase.

Despite this breakdown voltage drift, long term operation over 2000 hours has shown the stability of these diodes. Stabilization implies that the operational gain of the diode will not diverge and destruct the device. Pixel circuitry to compensate for variable breakdown voltages is implemented to allow a calibration over the array of diodes. The circuitry and calibration are described in Chapters 5 and 6.

Noise analysis of the two different guard ring diodes shows that the light should be absorbed mostly on the p-side of the high field junction for lower noise operation. In a direct comparison at gain of 10, the n+p diode has 10 times less noise than the similarly dimensions p+n diode. For low noise operation, the n+p diode is the diode of choice in this technology. However, due to the complexity of high voltage readout as described in Chapter 3, the trade-off of increased noise must be accepted. The p+n
diode allows the integration of an array of diodes with readout electronics on the same chip in standard technology.

The feasibility of the avalanche diodes in BiCMOS technology has been proven through extensive design and characterization. The noise behavior is, as expected, worse than specially designed APD processes where noise factors of 3-5 are measured at gains of 100 [Melchior78]. This noise is increased by two factors: 1) increased doping concentration in a typical CMOS process as compared to specialized APD processes, and 2) the generation of holes due to illumination. The increased doping concentrations decrease the operating voltage of the diode and increase the electric fields necessary for avalanche operation. These two factors simultaneously increase the k ratio from a reasonably low value at low fields to a larger value at high fields, thus increasing the noise of the diode. The hole generation decreases the gain by causing a feedback current within the depletion region which counteracts the electron gain in the silicon n+p diode. Pure electron generation is necessary for minimal noise in avalanche multiplication [McIntyre66]. Thus, for an average gain of M, the actual electron gain must be increased to compensate for the hole current. With the increased internal electron gain, the noise component also increases.

The original goal of this thesis was to implement a two dimensional array of APD pixels. With this in mind, implemented device sizes and geometries are selected to allow two dimensional integration. Since the chip with implemented diodes suffered from a very long delay in the foundry, the two dimensional design was implemented before the return of the diodes. Thus, there was no time to react to diode measurement results. Thus, the combination with APDs is analyzed in the next chapter. The comparison of interest, a CMOS APD imaging array with a standard CMOS imager is the crucial test of the CMOS APD array’s practical relevance.
5 APD Pixel and Array Design Considerations

5.1 Introduction

Although arrays of APDs have been commercially available for some time, simply placing APDs in an array fashion on a chip is not practical for arrays of more than a few tens of APDs since each diode would require at least one output line. With a small array of 12 by 24 pixels, this would mean at least 288 pins. Not only are 288 pins required but also a second chip including 288 output amplifiers required to measure the current in the avalanche diodes. Thus, we have opted towards the active pixel CMOS imager structure, using two source followers to select row and column. This architecture allows the 288 outputs to be multiplexed to only one output.

The pixels used here however are much more complex than in traditional CMOS imagers. These pixels must include circuitry to control the applied voltage on the diode and change the input current into an output voltage. Various circuits for APD control are analyzed, all using the sub-Geiger mode with an applied voltage slightly less than the breakdown voltage. The resulting analysis shows that an amplifier is necessary to operate the APD without quenching the gain when light is measured.

The requirements of a two dimensional APD image sensor include the following:

- an amplifier to control the voltage, area as small as possible with an open loop gain of at least 100
- power consumption of amplifier of less than 100 µA since it occurs in each pixel and is running continuously
- square pixels to allow the same resolution in both directions.
- round geometry of the avalanche diodes to prevent edge breakdown, and largest active area possible to allow the best possible fill factor.
- minimum number of pixels (30 x 30) to allow reasonable image
- readout using most sensitive circuitry (i.e. integrating amplifier)

The above requirements were not all met in this work due to significant trade-offs. The square pixels were not possible in this case because layout around the round diode is much less efficient than the rectangular pixels implemented. The minimum number of pixels was also not achieved, since
the area of the chip was restricted to 2.2 by 2.2 mm. Due to the discontinuation of the foundry service by Orbit, the integrating pixel with 30 by 30 pixels was not implemented into hardware.

The circuitry implemented is a transimpedance amplifier using a MOSFET as the feedback resistance. Design and noise analysis of the implemented transimpedance amplifier are presented. Layout considerations and programmability of the pixel are presented. Experimental results are presented in the next Chapter.

5.2 APD Control

In creating an APD image sensor the APD itself is only one consideration. The circuitry around the APD must provide the bias, protection and control to allow the detection of low light level signals. As described in Chapter 2, the APD can be operated either slightly below breakdown in the sub-Geiger mode or above breakdown in the Geiger mode. The Geiger mode requires quenching circuitry to prevent device breakdown and produces a train of pulses, whose frequency corresponds to the impinging light level. For high speed, low light level and few photon detection, this mode is suitable. In our case, speed is not a critical issue, photon numbers are not that low and linear performance is required. Hence, the sub-Geiger mode is more appropriate.

To control the APD two different strategies have been analyzed: current control where the APD’s bias current is set prior to exposure to light and voltage control where the APD’s voltage bias is set.

5.2.1 Current Control

Current control requires the APD’s bias current to be set using a variable current source. A current mirror is a simple method of controlling current in the APD. The current is set externally to bias the APD near the breakdown voltage. Since the current mirror’s output voltage is variable, each APD adjusts for the breakdown voltage variation, even with time and temperature. The original current source would be the same for all devices, yet the applied voltage at the APD would differ for each device, depending on its breakdown voltage as shown in Figure 5.1.
Figure 5.1: Current mirror control of APDs with one current source $I_{\text{ext}}$, current mirror transistors MCM1 and MN1..3, and APDs D1..D3. The high voltage source is combined for all APDs.

Though this method of control seems like the ideal bias situation, the response of the system to light is less than suitable for an avalanche image sensor. The light on the APD increases the current in the diode. Since this current is larger than the current in the mirror, the excess current charges up the node capacitance. Thus, as the output voltage increases, the APD bias voltage is decreased causing a quenching of the avalanche gain. For a single photon detector or a photon counter, such a circuit is useful but not if controllable gain is desired. From the gain vs. voltage curve (as in Figure 4.5), a slight change in the voltage near the breakdown voltage can have a significant influence on the gain. Therefore, for this application other options must be considered.

5.2.2 Voltage Control

Voltage control methods are an alternative to current control. The different methods all set the applied voltage across the diode and measure the current by various circuits. The methods considered include:
- circuitry similar to traditional active pixel circuitry (Figure 5.2a)
- charge sensitive preamplification (Figure 5.2b)
- transimpedance amplifier (Figure 5.2c)

The reset switch applies the bias voltage across the diode. The switch is then opened and allows the APD to integrate charge on the output node. To accommodate the high voltage, the reset switch is placed below the
APD rather than above the photodiode as in traditional APS circuitry [Mendis93]. Unfortunately the situation is similar to the current control mode. The charge integrated decreases the applied voltage, thus quenching the system. For our application this behavior is unsuitable.

The next two options (the charge sensitive preamplifier and the transimpedance amplifier) require a larger effort in terms of circuit complexity. Both options require operational amplifiers (op amp) integrated within the pixel. The op amp separates the varying output voltage from the applied diode bias voltage. These op amps offer certain advantages but increase the pixel layout size significantly.

Figure 5.2: Different voltage control modes for APDs. (a) is similar to traditional active pixel circuitry with a reset switch, (b) is a charge sensitive preamplifier with reset switch for the integration capacitance and (c) is a transimpedance amplifier using a resistor in feedback to convert the current into a voltage.

The charge sensitive amplifier (Figure 5.2b) transfers the charge from the diode directly to the feedback capacitance of an op amp. The gain of this circuit is dependent on the ratio of the diode node capacitance to the feedback capacitance. The diode capacitance is also dependent on the applied voltage. The larger the applied voltage, the smaller the parasitic capacitance due to the diode depletion region. Additionally, a reset switch must be integrated in order to reset the capacitance and dictate the integration time. The parasitic capacitances of this reset transistor also contribute to the feedback capacitance.

In an initial design, individual time continuous observation of the APD over a range of input currents is desirable. Thus, a transimpedance amplifier whose output corresponds to the product of the instantaneous current with the feedback resistance is implemented. The feedback resistance can also be varied to accommodate different input current
ranges. In CMOS technology processes, large resistors of controlled values are difficult to fabricate. Resistors can be implemented as:

- poly resistors where the resistance is approximately 20 $\Omega$/square. These resistors are very area intensive.
- n-well resistors where the resistance is approximately 2000 $\Omega$/square. These values are not easy to control and require large separations in layout to other n-wells.
- MOS transistors where the resistance depends on the operating point. These resistors are voltage dependent, process dependent (threshold voltage) but variable by adjusting the operating point. This adjustment can allow an adaptation of the resistor value but the adjustment voltage must be very stable.

The resistance is implemented as a PMOS transistor to allow the user to modify the transimpedance. Unfortunately, modifying this transimpedance also modifies the gain and bandwidth of the entire circuit. Thus, for very small input currents, a large resistance is used but the bandwidth of the readout amplifier is small also, allowing only low frequency components of the current to be detected. The advantage of the transimpedance is that the circuitry is non-integrating and allows the observation of the time dependence of the response. The transimpedance amplifier has been implemented in the APD pixel.

As mentioned before, both of these options require operational amplifiers to stabilize the applied voltage across the diode. The op amp stabilization also guarantees that the gain remains fairly constant. In terms of biasing and programmability, both these options also allow the greatest flexibility during testing, as is described later in this chapter.

### 5.3 Integrating vs. Non-integrating sensors

Depending on the application of the sensor, an integrating or non-integrating sensor can be advantageous. Since the applications of most of the traditional APS cameras are cameras for relatively controlled light environments, usually integrating sensors are used. These sensors have a minimum exposure time and a maximum exposure time. These times are adjusted to accommodate the readout speed of the array and the amount of light present in the applications environment.

The logarithmic sensor (Fugia5 [Ricquier95]) presented by IMEC Belgium and a commercially available high dynamic range camera based...
on the logarithmic sensor (IMS [IMS Chips]) show that non-integrating sensors are already available. These sensors have no minimum exposure time and readout can be completely random and time asynchronous.

The two main advantages of the non-integrating sensors are:
- The signal is continuous and directly related to the instantaneous light intensity.
- The sensor can be examined in quick succession and not based on a minimum integration time.

The disadvantage is that the beneficial averaging in terms of Signal to Noise Ratio (S/N or SNR) from an integrating sensor cannot be exploited [Dierickx96]. For comparison, two diodes are considered, one with a capacitance at the output, the other with only a resistor at the output. The capacitance is 50 fF; the resistor is 40 MΩ and the diode capacitance is 20 fF. These values are typical of the implemented structures in a 2 µm technology. Both outputs are measured using a high ohmic voltage amplification with equivalent input capacitance of 50 fF. With an integration time of 1 ms, the integrating circuit (with the capacitance) has a bandwidth of 1 kHz. The signal is determined by the integral of the diode current over time and the total capacitance as shown in Eq. (5.1) [Morf96].

\[
V_{\text{cap}} = \frac{1}{C_{\text{tot}}} \int_{t_i}^{t_f} I_{\text{ph}}(t) \, dt
\]

For 20 pA signal current, 1 ms integration time and 120 fF conversion capacitance, the output signal voltage is 166 mV. The shot noise of the diode current of 20 pA DC current is \(\sqrt{2qIB} = 80 \text{ fA}_{\text{rms}}\).

With a non-integrating circuit, the output voltage signal is dependent on the resistor’s value. Using typical values possible in this BiCMOS technology (e.g 40 MΩ resistor and 20 pA signal current), the output signal voltage is 800 µV. Increasing the resistance is not possible due to technological restrictions such as area. Additionally, increasing the resistance also decreases the bandwidth because the RC time constant increases. Using a 40 MΩ resistor and 70 fF diode capacitance, a bandwidth of 360 kHz results. The shot noise of the diode is therefore 19 or \(\sqrt{360}\) times larger than in the integrating case due to the bandwidth difference only. In addition to the shot noise increase, a thermal noise component from the resistor must be added. Increasing the resistance by a factor of 2 increases the noise voltage by \(\sqrt{2}\). Because the signal is directly proportional to the resistance, maximum signal to noise is achieved with the largest resistor possible.
As can be seen from these numbers, the integrating case improves both the signal and noise as compared to the non-integrating case. However, because the time dependent observation of the diode is critical to the understanding of the pixel operation, a non-integrating transimpedance amplifier was implemented in the pixel array.

### 5.4 Array based on universal APS frame

The image sensor architecture is derived from the universal APS array [Mendis93]. The schematic is typical of CMOS imagers including a two stage source follower buffer circuit called pixel buffer and column buffer in Figure 5.3. The actual sensor is an array of light sensitive diodes and controlling circuitry. The light sensitive diodes are APDs and the controlling circuitry is a pixel amplifier in each pixel. The row address decoder selects one entire row at a time, allowing the analog voltage to appear on the output line of each column. The column address decoder then sequentially selects each column. The analog voltage appearing at the output is from the selected row and selected column. The column buffers are slightly larger than the row buffers due to the architecture and loading appearing at the output lines respectively. The speed of readout is limited by the sequential access of the column buffers, since the column address decoder is run at a higher frequency than the row address decoder.

In most APS sensors, the pixel pitch is defined and fixed. In the case of the APD array sensor, the address decoder pitch is adjusted to the slightly larger APD pitch. As a rule of thumb, a simple APS design requires a pitch of 15-20 times the minimum feature width. In [Vietze97], for example, an APS array with 30 \( \mu \text{m} \) pixel pitch fabricated in a 2 \( \mu \text{m} \) process is described. Due to the increased number of transistors (10 vs. 3) in the pixel, the multiple n-wells and the larger APD, the pitch is significantly larger at 154 \( \mu \text{m} \) by 71.5 \( \mu \text{m} \).

#### 5.4.1 Readout

The readout buffer architecture has been described in detail in other works [Vietze97] but will be summarized here. Figure 5.4 shows the readout of a two by two array using two source followers, one for row select and the other for column select. The in-pixel buffer is an NMOS source follower circuit with a row select pass transistor (transistors MNSF, MNCS and
The current source bias (MNCS) for the in-pixel buffer is implemented on a per column basis whereas the column buffer (MPCS) is implemented only once. The advantage of this architecture is that largest transistors are not within the pixel. Therefore the size can be optimally designed and implemented.

**Figure 5.3:** Image sensor architecture with row address and column address decoders. The pixel buffer receives input from the APD and transimpedance amplifier (not shown). The column buffers are larger than the pixel buffers due to the switching speed and the larger load. The analog output buffer is even larger to drive the external load.

The column buffers are PMOS transistors (MPSF, MPCS and MPSW in Figure 5.4) which compensate the voltage level shift of the NMOS row buffers. This type of compensation allows a larger output voltage swing. The level shift of the source followers is given by Eq. (5.2), neglecting the bulk source voltage. $V_{GS}$ is the gate source voltage of the source follower transistor MNSF. $V_T$ is the threshold voltage of this transistor. The current source transistor is MNCS. $K'$ is a device parameter corresponding to $\mu C_{ox}/2$. $W$ and $L$ are width and length of the transistor respectively \cite{Laker94}.

\[
V_{in} - V_{outc0} = V_{GS} = V_T + \frac{I_{MNCS}}{K'(W/L)_{MNSF}} \tag{5.2}
\]

For the PMOS source follower, the bulk source voltage can be neglected since the n-well of MPSF and MPSW is connected to $V_{out}$, causing the bulk source voltage to be zero. For the NMOS source follower, an output
voltage dependent shift occurs in the threshold voltage, causing the new threshold voltage to be as in Eq. (5.3) [Laker94].

\[ V_T = V_{TO} + \gamma (\sqrt{2\phi_f + V_{outc0}} - \sqrt{2\phi_f}) \]  

(5.3)

where \( \gamma \) is the bulk threshold parameter, \( \phi_f \) is the surface potential at the source side of the device and \( V_{TO} \) is the zero bias threshold voltage. This equation can be solved by iteration or roots [Laker94]. Another negative effect of this bulk voltage is the change in the small signal gain \( g_A \) from unity to less than one as shown in Eq. (5.4) [Laker94].

\[ g_A = \frac{\Delta v_{outc0}}{\Delta v_{in}} = \frac{(g_m)_{MNSF}}{(g_{DS})_{MNCS} + (g_m + g_{mb} + g_{DS})_{MNSF}} \]  

(5.4)

Normally the \( g_{DS} \) values can be neglected in comparison to the \( g_m \), allowing the gain to be almost unity, if no bulk source voltage is present. With a bulk source voltage gains normally range between 0.9 and 0.7.

In most APS circuits the pixel buffer is designed for charge to voltage conversion. In this APD pixel, the buffer is used solely to allow the row selection to occur. Because the first stage of the pixel is an operational amplifier, the noise contribution of the source followers is not quite as critical as in the APS circuits where the source follower is the first stage.

The noise contribution of the NMOS SF circuitry is more significant than the noise contribution of the PMOS SF circuitry due to the dimensions of the transistors. The NMOS SF design is restricted by the pixel size. Thus, smaller transistors are used than in the PMOS design. Larger transistors and currents in the second stage reduce the thermal noise contribution significantly.

The thermal noise of a source follower is the input referred noise contribution due to the channel resistance. When operating in saturation, the channel resistance is not \( g_m \) but reduced by a factor 2/3 due to field effects [Laker94]. Thus, the input referred thermal noise contribution is shown in Eq. (5.5).

\[ \frac{\overline{dv_{teq}^2}}{df} = (4kT R_{CH}) df = \left( \frac{8kT}{3} \frac{1}{g_{mMNSF}} \right) df \]  

(5.5)
Figure 5.4: Readout of typical 2 by 2 array, using two source followers. The NMOS source follower (MNSF) is contained within the pixel except for the current source controlled by Vbn. The second source follower (MPSF) is PMOS and exists once for each column. The current source (MPCS) controlled by Vbp is present only once on the entire chip.

As can be seen from this equation, the thermal noise depends only on gm. This gm can be increased by either increasing the W/L or the current through the device. Both these factors are efficient in decreasing the thermal noise. Decreasing temperature T is also effective in decreasing thermal noise.

For the NMOS source follower with dimensions 9 μm/2 μm and 8.3 μA of current, the noise contribution is 12 nV/√Hz. Adding the noise of the current source transistor (30 μm / 15 μm), the total noise at the NMOS
source follower input is $15 \text{nV/√Hz}$. For the PMOS source follower with dimensions $30 \mu\text{m}/2 \mu\text{m}$ and $120 \mu\text{A}$ current, the noise contribution is only $4.6 \text{nV/√Hz}$. Again adding the noise contribution from the current source ($200 \mu\text{m}/4 \mu\text{m}$) the total noise at the PMOS source follower input is $5.6 \text{nV/√Hz}$. To transfer this noise to the input of the NMOS transistor, the voltage gain (0.8) of the NMOS source follower is used. The resulting noise voltage ($7\text{nV/√Hz}$) is added to the NMOS source follower’s noise contribution. The total input referred noise of the source follower combination (NMOS and PMOS) is $16.8 \text{nV/√Hz}$. The noise of the PMOS source follower can almost be neglected when compared to the NMOS source follower noise contribution.

Flicker noise is more difficult to predict since the corresponding model parameters are not included in the SPICE models from the foundry. The corresponding equation is shown in Eq. (5.6).

$$
\frac{dV_{\text{flicker}}}{2} = \left(\frac{KF_f}{WLC_{\text{ox}}^{\frac{1}{2}}}\right) \frac{df}{f}
$$ (5.6)

The flicker noise depends on the area of the device. For low flicker noise, large area devices are best. However, flicker noise can be compensated with an appropriate correlated double sampling mechanism [Hopkinson82], thus is not as important as the thermal noise components.

5.4.2 Addressing

To select the row and column, two address decoders are implemented, allowing complete random addressing if desired. The row and column address decoders must be logically different because the row and column select transistors are NMOS and PMOS respectively. The row select is active “high”. The column select is active “low”.

The address decoder selects only one row and one column at any one time. The NMOS row select transistor requires AND logic while the PMOS column select transistors use NAND logic. As shown in Figure 5.5 an AND gate can easily be created from a NAND gate, by adding a simple inverter. Thus, both address decoders are based on one NAND gate. In view of a larger chip, the decoders are implemented with 6 address bits allowing them to be capable of addressing a 64 by 64 pixel array.
Other simple logic is combined with these decoders to allow special programming of the array and will be described in the next section.

![Address decoder logic using NAND and simple inverter to create AND. s0..s5 are external addresses. NAND is used for the column select transistors while the AND is used for row select transistors.](image)

**Figure 5.5:** Address decoder logic using NAND and simple inverter to create AND. s0..s5 are external addresses. NAND is used for the column select transistors while the AND is used for row select transistors.

### 5.5 Pixel design

The goal of the pixel design is to create a compact layout while still meeting all the significant specifications. This particular pixel design requires a round APD with guard ring, a transimpedance amplifier as well as row and column addressing. The design of this pixel is divided into 5 parts:

1. APD control and readout
2. Transimpedance amplifier design
3. Transimpedance amplifier noise
4. Programming and biasing
5. Size and layout
5.5.1 APD control and readout

The APD chosen for this implementation of the array is a p+ in n-well with p-base guard ring. There are several advantages to this structure:

- smaller breakdown voltage than the other guard ring diode
- independent of substrate, giving a greater freedom in biasing the diode
- low voltage readout node possible

The anode is connected to the readout circuitry while the cathode is connected to the high voltage supply. The lower quantum efficiency is accepted as a trade-off. The negative effects may be balanced by the avalanche gain of the diode.

The controlling circuitry for the avalanche diode includes an operational amplifier and a feedback resistance. This configuration was chosen to allow continuous time observation of the APD. This time observation is limited to a few hundred kilohertz due to the 2 μm process used for amplifier implementation. Thus, pulse detection in the picosecond or even nanosecond range as seen in [Melchior77] is not possible due to the amplifier speed and the relatively short multiplication region, making the carriers diffuse to the depletion region before multiplication.

*Figure 5.6* shows the basic APD control circuitry. The high voltage supply is common to all APD cathodes on chip. The reference voltage to the transimpedance amplifier \( V_{\text{prog}} \) is the virtual voltage for the anode of the APDs. In first experiments the anode voltage is the same for each APD. In further experiments this voltage can be individually programmed, as described in the programming and biasing section.

5.5.2 Transimpedance amplifier

The transimpedance amplifier (TIA) consists of a simple operational amplifier and a feedback resistor. Traditionally the amplifier would be a two stage differential amplifier and the feedback would be an external or integrated resistor. However, due to area and layout constraints, the amplifier was reduced to a minimum single stage amplifier. The resistor was implemented as a small area PMOS transistor.
Figure 5.6: Two by two array of APDs and controlling transimpedance amplifiers. The high voltage line is common for all diodes. The programmable voltage ($V_{prog}$) is the same for all diodes in first experiments. This voltage is programmable in later experiments. The feedback resistor is implemented through a PMOS transistor.

A single stage amplifier uses a 5 transistor differential stage as shown in Figure 5.7a. The simplest amplifier would be an inverter which is much smaller but does not have the capability of adjusting the operating point during operation. The differential stage has a large common mode range, allowing the bias point of the input to be varied while keeping the small signal gain relatively constant. Another advantage of such a simple structure is the stability achieved because only one dominant pole is present. In a two stage amplifier design, a compensation capacitance is needed to split the poles, creating a dominant pole to achieve unity gain feedback stability. In this case, only one dominant pole due to the load capacitance is present and pole splitting is not necessary. The non-dominant pole of the current mirror is at a much higher frequency due to the smaller impedance (ca. 100 times smaller) at the current mirror node.

The goal of the amplifier design is to have a gain of at least 100 and the bandwidth to be as large as possible. Increasing the gain decreases the
bandwidth due to the constant gain-bandwidth product. The amplifier
shown in Figure 5.7a has been simulated using SPICE and models from
the semiconductor foundry. The dimensions of the transistors are shown in
Table 5.1 and have been set to allow a large transconductance $g_m$ at the
input. The total current consumption is regulated by $V_{biasn}$ and must be
kept low for each amp due to the number of amplifiers on chip. Even with
a quiescent current of 10 $\mu$A, the total power consumption for all 288
amplifiers would be 14.4 mW. The dimensions of the PMOS load
transistors are restricted by the layout area. The gain of approximately 500
is simulated in SPICE with a simulated -3 dB bandwidth of 140 kHz. The
load capacitance used in simulations is 100 fF. Unity gain frequency
crossover is above 30 MHz. The open loop gain can only be measured
indirectly as will be described in Chapter 6.

The transimpedance is created by implementing a resistive feedback
around the operational amplifier. In this case a PMOS transistor is used as
the resistor. For a large resistor, the length should be larger than the width.
The transistor’s gate voltage allows the resistance to be varied during
operation. The operating point of the transistor is the linear sub-threshold
region of transistor operation. The equation for the current in this region is
shown in Eq. (5.7) [Foty97]. For small drain source voltages, $V_{ds}$, a first
order approximation shows that the current not only depends on $V_{ds}$ but
also on the thermal voltage $kT/q$ as shown in Eq. (5.8).

$$I_{ext} = I_0 e^{\frac{q(V_{GS}-V_T)}{kT n}} \left(1 - e^{\frac{qV_{DS}}{kT}}\right)$$  \hspace{1cm} (5.7)

$$I_{ext} \approx I_0 \frac{q}{kT} e^{\frac{q(V_{GS}-V_T)}{nT}} V_{DS}$$  \hspace{1cm} (5.8)

As can be seen from the Eq. (5.8), the resistance is dependent on the gate
source voltage and the threshold voltage, as well as the thermal voltage
$kT/q$. The threshold voltage also changes with operating point because the
bulk source voltage of this transistor is not zero. The gate source voltage
is effectively $V_{resist} - V_{o2}$. The drain source voltage is $V_{im} - V_{o2}$. Thus, the
resistance varies slightly with operating point, making the linear
adjustment of the resistance difficult. However, the adjustment of the
resistor gate voltage ($V_{resist}$) and the amplifier reference voltage ($V_{prog}$)
allows the feedback resistance and the dc operating point of the
transimpedance amplifier as well as the avalanche photodiode to be
adjusted. Since these values can also be programmed individually, the
adjustments can be different for each pixel. The programmability is described in more detail in section 5.5.4.

![Diagram of a simple differential amplifier](a)

Figure 5.7: Simple differential amplifier with feedback resistance. (a) is a 5 transistor differential stage with active load. (b) shows the amplifier with the feedback resistor is placed between the negative input Inn and the output vo2. This resistance is implemented as a PMOS transistor.

![Diagram of an amplifier with feedback resistor](b)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>W/L (μm/μm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mi1, Mi2</td>
<td>NMOS differential input</td>
<td>40/4</td>
</tr>
<tr>
<td>Mp1, Mp2</td>
<td>PMOS current mirror load</td>
<td>6/4</td>
</tr>
<tr>
<td>Mbn</td>
<td>current source</td>
<td>16.5/5</td>
</tr>
<tr>
<td>Mpr</td>
<td>PMOS resistor</td>
<td>5/6</td>
</tr>
</tbody>
</table>

Table 5.1: Transistor sizing table for transimpedance amplifier and PMOS resistive feedback. All device sizes are drawn dimensions in μm.

The goal of the transimpedance amplifier with a photodiode (shown in Figure 5.6) is to convert the photocurrent into an output voltage. The load voltage swing is isolated by the transimpedance amplifier from the photodiode capacitance for improved bandwidth. The output voltage \( v_{o2} \) is equal to the diode current including leakage current \( I_p \) times the feedback resistance \( R_f \): \( v_{o2} = I_p R_f \). The voltage at the op amp input is reduced by the op amp’s open loop gain. At low frequencies, this open loop gain is normally fairly large, allowing the error voltage across the op amp input to
be very small. When the open loop gain decreases with frequency, however, this error voltage becomes significant, thus degrading the performance at higher frequencies.

The system response has a constant gain-bandwidth product. If the transimpedance is increased (e.g. the feedback resistor is increased) the bandwidth decreases. Additionally, as the resistance increases, the parasitic capacitance has a larger effect, thus reducing the bandwidth even more. However, for the largest output signal of a DC input current, the feedback resistance should be as large as possible. In our case, this resistance is on the order of 25 MΩ, using a PMOS transistor of size W/L=5 μm /6 μm.

5.5.3 Noise analysis of transimpedance amplifier

Though the transimpedance amplifier with photodiode seems to be a simple structure, it shows a highly complex noise behavior [Graeme96]. The complex behavior comes from diode’s impedance which changes with frequency due to the typically large diode capacitance. In our case, the diode capacitance is extremely small (25 fF) and therefore this high frequency behavior is not seen at the frequencies of interest (below 1 MHz). The noise analysis assumes shot noise for the photodiode and a purely resistive feedback, instead of a MOS transistor. The avalanche noise factor F can be added to the analysis later. The basic noise components of the circuit include the feedback resistor’s noise source and the amplifiers input noise voltage. The MOS amplifier has no equivalent input noise current since no gate current flows. The contribution of the resistor’s noise to the output (enR) is direct as seen in Figure 5.8. The diode’s shot noise is simply multiplied by the transimpedance to get the voltage noise at the output (enoI). The equivalent input noise voltage source transfers to the output (enox) with a gain 1+Rf / RD where Rf is the feedback resistance and RD is the equivalent diode resistance. The diode resistance R_D is very large, causing the transfer to be close to one [Graeme96]. Also, since the diode capacitance is very small, this equivalent resistance is not affected significantly by the capacitance. Therefore the transfer function for the input noise voltage remains constant over the frequency range.
Figure 5.8: Noise sources in a photodiode transimpedance amplifier. $I_B$ is the dc bias current in the photodiode. $i_{nD}$ is the noise current due to the diode, $e_{ni}$ is the input noise voltage to the amplifier. $e_{nR}$ is the resistance’s noise contribution. $R_D$ and $C_D$ are the equivalent diode resistance and capacitance respectively.

The total output noise $e_{no}$ is a combination of these noise sources and their transfer function from the noise source to the output. The spectral density of the resistor $e_{nor}$ is rms summed (as in Eq. (5.12)) to the spectral density of the input current $e_{not}$ and input voltage $e_{noe}$. The output spectral density $e_o$ has two components: a noise free signal $i_pR_f$ and the noise at the output $e_{no}$ as seen in Eq. (5.12). Equations (5.9) through (5.13) show the relationships of the spectral density to the actual circuit components. $R_f$ is the feedback resistance [Graeme96].

\[
\begin{align*}
  e_{nor} &= \sqrt{4kTR_f} \quad (5.9) \\
  e_{noi} &= R_f \sqrt{2qI_B} \quad (5.10) \\
  e_{noe} &= (1 + R_f / R_D)e_{ni} \quad (5.11) \\
  e_{no} &= \sqrt{(e_{nor})^2 + (e_{noi})^2 + (e_{noe})^2} \quad (5.12) \\
  e_o &= i_pR_f + e_{no} \quad (5.13)
\end{align*}
\]

Increasing the feedback resistance improves the signal to noise ratio. Increasing the resistance increases the noise at the output by the square root of the resistance value. The signal increase due to an increase in the resistance is directly proportional to the resistance. Therefore the total improvement in signal to noise is proportional to the square root of the resistance if the noise of the resistor dominates. Should the amplifier noise
also be significant, the noise improvement will not be as large when increasing the resistance. As the resistance increases, the noise of the resistance also tends to dominate.

Assuming a feedback resistance \( R_t = 25 \, \text{M} \Omega \), a dc current \( I_B = 20 \, \text{pA} \) and an input noise source of the amplifier \( e_m = 30 \, \text{nV/\sqrt{Hz}} \), the total noise can be calculated easily (neglecting 1/f noise). These values are realistic for the components available and implemented on chip. The calculated output noise components are: \( e_{noR} = 644 \, \text{nV/\sqrt{Hz}} \), \( e_{noi} = 63 \, \text{nV/\sqrt{Hz}} \), \( e_{noc} = 30 \, \text{nV/\sqrt{Hz}} \) for the resistance, dc current and amplifier noise respectively. Summing these as in Eq. (5.12), the total noise at the output is \( e_{no} = 647 \, \text{nV/\sqrt{Hz}} \). Thus, the resistance produces the significant noise component.

The detection chain implemented includes the following components and is shown in Figure 5.9:

- the photodiode for conversion from light to a current and its Poisson noise \( i_{np} \),
- the avalanche diode for current gain and the excess noise due to the avalanche gain mechanism \( F_{apd} \),
- the current voltage conversion (implemented as a transimpedance amplifier) and the amplifier and feedback resistance noise \( v_{namp+R} \),
- the row and column selection circuit (implemented as a source follower) \( V_{nsf} \).

The signal at the output of the current voltage converter (transimpedance amplifier) is \( M \cdot I_p \cdot R \) where \( M \) is the gain of the avalanche diode, \( I_p \) is the current due to the incoming light and \( R \) is the current voltage converter. The noise at the same point is shown in Eq. (5.14). The first term in the equation is the Poisson noise combined with the avalanche excess noise. The second and third terms correspond to the noise of the transimpedance amplifier. For low frequencies, the amplifier noise is converted directly to the output using a gain of 1. The last component is the source follower noise due to the rest of the detection chain.

\[
\begin{align*}
v_{notot}[\frac{V}{\sqrt{Hz}}] &= \sqrt{2qI_p F_{APD} M^2 R^2 + v_{namp}^2 + 4kTR + v_{nsf}^2} \\
&= (5.14)
\end{align*}
\]

In Figure 5.10, the resistance noise (4kTR) is dominant until the input current is large enough to allow the shot noise to dominate. The amplifier noise and source follower noise (\( v_{namp} \) and \( v_{nsf} \)) do not contribute greatly to the sum unless the feedback resistance value is reduced dramatically.
Figure 5.9: Signal and noise block diagram of detection chain with transimpedance amplifier. The incoming light $P$ is transformed to a current through the diode. This current is multiplied by $M$ in the avalanche portion of the diode. The conversion to a voltage occurs using a transimpedance amplifier. The row and column selection is implemented as a source follower. The respective noise contributions are shown.

To consider avalanche noise, the ideal $F_{apd}$ factor of 2 from Chapter 2 is used for calculations. This factor influences only the shot noise introduced to the system, thus increasing this noise by a factor of 2 when compared with no avalanche gain. The signal, on the other hand is increased by the factor $M$ (previously, in Figure 5.10, $M=1$). The result is that the shot noise component dominates more quickly, as shown in Figure 5.11. In either case, for low currents, the feedback resistance’s noise components dominate. An integrating pixel using a charge sensitive amplifier is analyzed in a similar fashion in Chapter 7.

Avalanche photodiode noise increases the noise contribution of the diode by the factor $F$, as seen in Chapter 2. Simply changing Eq. (5.10) to include this factor shows the increased noise due to avalanche. Measurements of the total noise behavior, including avalanche are shown in the next chapter. Signal to noise analysis is shown in Chapter 7.

The transimpedance amplifier noise performance is always poorer than for example a high impedance front end such as an open loop amplifier [Senior85]. The advantage of the transimpedance amplifier is the significant improvement in bandwidth and dynamic range when compared to the same high impedance front end. These advantages are offset by the noise penalty incurred. The dynamic range achieved by the transimpedance amplifier is also larger than the high impedance front end. The low frequency signal components are amplified by the closed loop
gain rather than the open loop gain. Thus, transimpedance amplifiers are still often preferred in wideband optical receiver front ends [Senior85].

Figure 5.10: Noise components of the transimpedance detection chain. No avalanche excess noise (Fapd=1) is assumed for these calculations. Two feedback resistances (10 and 20 Meg) are used.

Figure 5.11: Noise components with avalanche excess noise for 10 MΩ feedback resistance and F=2.

5.5.4 Programmability and biasing

The flexibility of this pixel design is the ability to adapt to different conditions. The programmability and biasing capabilities described here
contribute to this flexibility. The breakdown voltage of individual pixels can be programmed by the reference voltage $V_{\text{prog}}$ in each pixel. The bias voltage of the diode is the high voltage minus the programmed reference voltage. The programmability occurs through switches which should be complementary for improved operation. To save area, these switches are implemented as single switches only.

The gate voltage of the PMOS resistor ($V_{\text{resist}}$) is also programmable to adjust the feedback resistance on a per pixel basis. These features allow the adjustment of the gain and the transimpedance for each pixel separately. Programming all voltages simultaneously is achieved by two external pins, pr and pc. When pr is logical one and pc is logical zero, all pixels have the same programmed voltages. The logic implemented consists of an additional AND in the column decoder (per column) and an additional NOR in the row decoder (per row). When pr and pc are inactive, the voltages $V_{\text{prog}}$ and $V_{\text{resist}}$ are programmed when the individual pixel is selected. Thus, the external applied voltages can be changed with each address, allowing individual pixels to have different applied voltages. The time constant of the programmed voltages is very small because no additional storage capacitance has been implemented within the pixel. This short time constant is not a restriction in this case because the sensor is a non-integrating sensor. Thus it requires the programmed voltages only during readout (when the voltages are refreshed anyhow), not during the entire integration time like an integrating pixel.

In addition to the reference voltage and the resistance voltage three other bias voltages can be set. These are global voltages used to control the current sources in the transimpedance amplifier and source followers. The transimpedance amplifier current source can be set using voltage $V_{\text{biasn}}$ (see Figure 5.7b). The current in each individual amplifier is kept to a minimum because of the contribution to the static power dissipation since these amps have a quiescent current and are repeated 288 times on this chip. The current for the NMOS source follower (SF) stage and the PMOS source follower stage are set using $V_{\text{bn}}$ and $V_{\text{bp}}$ respectively. These currents do not have to be restricted since the number of current sources operating simultaneously is limited to the one row for NMOS SF (e.g. 12 NMOS current sources) and only one current source for PMOS SF.
5.5.5 Size constraints and layout

The ultimate pixel size and layout are determined by the chip designer. The amplifier required for the pixel is designed using SPICE and the specifications (5 transistor differential stage with gain of at least 100 and a bandwidth as large as possible). The layout is then attempted, with the least area possible. This is an iterative process since the arrangement of the transistors allows many degrees of freedom. Once the size of the amplifier is determined, the peripheral circuitry is added and the diode size is adjusted to allow a reasonable pixel size. The fill factor achieved is only 20%, comparing with APS 256 where the fill factor is 35%. The last step in the pixel design and layout is to form an array with all signal and supply lines connecting through the pixels. After many iterations, the result is shown in Figure 5.13. The result is the best solution achieved in the iterations by the designer. Increasing the fill factor would require much larger avalanche diodes and therefore much larger pixels.
The size of the avalanche photodiode pixel is determined largely by the avalanche photodiode used. In our case 3 different APDs are integrated in the same array. The size of the APDs is measured from the outer most n-well. The smallest APD’s n-well measures 48 µm in diameter. The actual diode (p+) diameter is 16 µm. The largest APD n-well is 54 µm with a 22 µm diameter p+ region. In the third variation a p+ ring was added around the n-well as an additional substrate contact. This APD including the p+ guard ring measures 63 µm in diameter. The inner APD n-well dimensions are the same as the smallest APD (48 µm n-well and 16 µm p+). For comparison, all pixels use the same transimpedance amplifier and the same pixel pitch. The only differences are the APD size or the guard ring presence.

The layout of a complete pixel containing the smallest APD, a transimpedance amplifier and the active pixel readout source follower is shown in Figure 5.13. The complete pixel requires one supply voltage Vdd and ground GND and a high voltage HV for the cathode of the diode. These supply lines must be in metal 1 or 2. Since this technology only has two metal layers and the pixel is very complex, both layers must be used as wiring layers. In remaining areas of the chip, metal 2 is used as a light shield wherever possible. Susceptible nodes include all high impedance diffusions such as those in the address decoders or those when programming signals pc and pr are inactive.

The supply and ground lines are vertical through the pixel. The additional reference and resistance voltages $V_{\text{prog}}$ and $V_{\text{resist}}$ are also vertical in metal 1. The output voltage must also be fed vertically through the pixel to allow the row and column addressing to function properly. The other circuit signals needed in the pixel are rsel1, biasn, prog1 and strb. Biasn is the current source bias for the amplifier. Prog1 selects program reference and resistance voltages. Strb is used to disconnect the amplifier output from the source follower like a sample and hold circuit. These signals are connected through the pixels horizontally using polysilicon lines. Thus, a grid structure evolves with supply lines vertical in metal and digital select signals and the bias current horizontal in polysilicon. The rest of the circuitry including the 5 transistor amplifier, the PMOS resistor, 2 select transistors ($V_{\text{prog}}$) and the complete source follower is placed between this grid. Total number of transistors in the pixel is 5 PMOS, 6 NMOS and an APD. The layout dimensions are 154 µm by 71.5 µm allowing an array of 12 by 24 pixels to be placed on a chip of 2.4 mm by 2.4 mm.
Figure 5.13: Layout of one pixel including the round avalanche diode and the transistors for the transimpedance amplifier and the source follower. Complete pixel schematic is shown in Figure 5.14.

Figure 5.14: Complete circuitry implemented on chip. Within the dashed box is the circuitry in each pixel. Layout of the pixel is shown in Figure 5.13.
5.6 Conclusions

After examining different control methods for the APD, several solutions using operational amplifiers were suggested. Since the APD characterization in the array is also important, time continuous observation is vital to draw conclusions about the gain variation and the influence of the pixel on the diodes. Thus, a transimpedance amplifier was designed and implemented to control the APDs within the array. Implementing a charge sensitive amplifier requires only a few modifications, as described in Chapter 7.

The array frame for traditional CMOS imagers is used to multiplex the 288 outputs to one, thus allowing an array of 12 by 24 pixels to be implemented. Though the readout architecture is well known, the analysis of the thermal noise is specific to this design.

The more important design of this array however is the pixel design. The largest restriction is the total pixel area which is very large when compared with other CMOS imagers but tiny when considering amplifier designs. Effectively, a complete op amp is implemented in each pixel.

In terms of noise, larger area allows improvement but the quiescent current is really limited by the number of pixels on chip and the maximum total power dissipation. Increasing power dissipation also increases chip temperature, causing an unwanted breakdown voltage increase. To compensate for such a breakdown voltage shift, as well as an on-chip variation, programmability of biasing has been implemented. This feature allows each pixel to have 1) a different voltage across the diode and 2) a different transimpedance. The first of these features allows the adjustment of the gain and the second allows the adjustment of the uniformity of the resistances.

As a result the 5 PMOS, 6 NMOS an APD are placed carefully within a 154 μm by 75 μm pixel size which allows the implementation of 12 by 24 pixels on a 2.4 mm by 2.4 mm chip. Measurement on the individual pixels, amplifier, biasing, and array are presented in the next chapter.
6 APD Pixel and Array Experimental Results

6.1 Introduction

The chip characterization involved two major portions: pixel and array operation. Pixel characterization consists of several measurements:

- diode characterization to see how this process compares to others.
- amplifier characterization to confirm design specifications.
- complete pixel measurements (diode, amplifier and source follower readout circuitry) to analyze the system operation.

The pixel characterization is followed by a camera realization using the entire array with 288 pixels. Figure 6.1 shows the entire array and a close up of several pixels. In the left hand image, the metal 2 covering appears as horizontal rows. The right hand image shows the same chip in more detail, where structure of the pixel is visible. The camera realization involves obtaining an image from the sensor array under two conditions: 1) with and 2) without avalanche gain. The avalanche gain can be regulated by controlling and varying the high voltage supply.

6.2 Pixel characterization

6.2.1 Diode

The chip consists of an array of 12 by 24 pixels as well as a separate diode for characterization. This diode is connected to the same high voltage supply as all other diodes. Using this diode the breakdown voltage range of the APDs on this chip is characterized. Since all the diodes within the pixels are biased using \( V_{\text{prog}} = 3 \text{ V} \), they are all still in the reverse bias region, yet not in avalanche. Such diodes present a very sensitive, high ohmic node in terms of noise. An external capacitance stabilizes the high voltage node to reduce the noise on chip.

The current measurement on the individual diode confirms the previous large gain measurements on other chips. The setup includes the HP4145B and the microscope with the LED light spot. Using the current versus voltage measurement of the HP parameter analyzer, the gain can be calculated as described in Chapter 2. Experimental results on this single
separate diode (without amplifier) on chip showed a gain of more than 1000 as shown in Figure 6.2. Two different light levels were used for measurements: 0.81 nW and 1.81 nW respectively. As expected, the gain is independent of light level. The breakdown voltage is close to 17 V. Since this chip was produced in a 1.2 μm technology, this decreased breakdown voltage is expected due to the increased doping concentration in the smaller technology.

Figure 6.1: Photomicrographs of 12 by 24 pixel array (left) and a more detailed view of individual pixels (right). In the left photomicrograph, metal 2 covers a large portion of the chip as a light shield. The address decoders are at the top and left side of the chip. In the right photomicrograph the pixel is shown in more detail. The round structures are the avalanche photodiodes.

6.2.2 Amplifier

To characterize the amplifier used in each pixel, one pixel in the array contains no diode. Its input is connected to another external pad to allow better characterization of the amplifier and additional source followers. The bandwidth of the transimpedance amplifier is influenced by the feedback resistance which, in turn, also depends on the operating point of the amplifier. Additionally, the capacitance at the input is very different when pad and pin are connected or simply a reverse biased APD. A pin capacitance is on the order of 2 pF, whereas the reverse biased APD's capacitance is only ca. 25 fF. Thus, the same bandwidth behavior is not expected if, for example, an external diode is connected when compared to the internal diode. To analyze the amplifier itself, the capacitance is not as important because a voltage source is used to drive the input.
Figure 6.2: Current gain versus applied voltage for single diode on FS45 chip. The two curves represent different light intensities corresponding to 0.8 nW and 1.8 nW on each diode. The setup uses a small light spot focused on the diode and measures the current using the HP4145B semiconductor parameter analyzer. The high voltage current is then divided by the low voltage current to obtain the current gain.

The amplifier and the two source followers are shown in Figure 6.3. Amplifier characterization begins with a DC open loop gain measurement. In this case the feedback resistance is increased to the largest possible value (i.e. $V_{\text{resist}} = 5$ V) and a small signal voltage with the appropriate dc bias is applied to the input of the amplifier. The output of the amplifier corresponds to the ac input times the open loop gain. The two series source followers influence both the dc and ac components of the output signal. The gain of these source followers is actually an attenuation of 0.8 per source follower for low frequency inputs. Cascading the two source followers results in a total attenuation of 0.64. The measured value of the open loop gain is 100. When the source follower attenuation is taken into account, the open loop gain is approximately 150. The simulated open loop gain is 480, corresponding to a difference of a factor of 3. At 7.7 MHz frequency, the measured value for gain is 2. Thus, the unity gain frequency is slightly higher. The unity gain frequency in simulation is 20 MHz. Again, the simulation value is optimistic. These measurements were done using the bias voltages as follows: $V_{\text{resist}}=1.962$ V, $V_{\text{prog}}=2.996$ V, $V_{\text{bn}}=1.241$ V, $V_{\text{biasn}}=1.697$ V, $V_{\text{bp}}=3.52$ V. Comparing
the measured values to the simulated values, we see that the simulation gives a slightly optimistic result in terms of gain and bandwidth but the goal of an open loop gain of 100 and a unity gain bandwidth of more than 1 MHz was achieved in the design.

![Circuit Diagram](image)

**Figure 6.3:** Simplified amplifier and source follower circuitry in one pixel. The input ampin is normally connected to a diode, but for amplifier characterization in one pixel, the diode is missing. The input is connected to an external pad. For open loop gain characterization, V\textsubscript{resist} is set to 5 V. For voltage follower operation, V\textsubscript{resist}= 0 V.

To characterize the amplifier further, a voltage follower is implemented. The resistance (PMOS transistor) is reduced to the smallest value possible by applying V\textsubscript{resist}=0 V. The actual value of the gate voltage at the transimpedance PMOS transistor is not measurable in this configuration but significantly higher than 0 V. The switches to apply V\textsubscript{resist} at the gate are implemented as PMOS transistors which do not pass low voltages as well as high voltages. Thus, according to simulations, the minimum V\textsubscript{resist} value at the gate is approximately 1.4 V. Simulated, the resistance corresponds to 300 kΩ. However, the behavior is still similar to a voltage follower and useful for characterization. The amplifier output range measured in using this configuration is limited to approximately 1.5 V. The output voltage lower level is limited by the input transistors, which are biased at the reference voltage V\textsubscript{prog}. When the output voltage decreases, the input transistor is pushed out of saturation, causing the current distribution in the differential stage to be one-sided. When this happens, the output voltage saturates.
When operating the APD pixel in the normal mode, the resistance is between the open loop and voltage follower mode. To determine the effective resistance, an inverting amplifier is created by using the amplifier without a diode and adding an external resistance. Thus, an inverting amplifier (shown in Figure 6.4) can be used to measure the resistance. The external resistance is known and the externally applied ac-coupled source supplies the ac signal. The dc bias is set by the $V_{\text{prog}}$ reference voltage. The output voltage is dependent on the external resistance, the internal resistance and the amplifier open loop gain.

![Feedback resistance determination using an external resistor](image)

**Figure 6.4:** Feedback resistance determination using an external resistor $R_{\text{ext}}$ to measure the internal resistance $R_{\text{meas}}$. The DC bias is determined by $V_{\text{prog}}$. The AC input is dependent on the open loop gain and the ratio of the feedback resistance to the external resistance. The source follower attenuation is neglected in the equations (6.1) and (6.2).

Since this open loop gain is fairly small (due to the one stage amplifier) the resistance calculation must include the finite open loop gain as shown in Eq. (6.1). The influence of the open loop gain is significant when the open loop gain is of the same order as the resistance ratio $R_{\text{meas}}/R_{\text{ext}}$. Using this method, the measured resistance values vary between 270 kΩ and 40 MΩ for different $V_{\text{resist}}$ voltages. The source follower gain has been included in these calculations (e.g. open loop gain of 150). The equations (Eqs. (6.1) and (6.2)) neglect the attenuation due to the source follower gain.

\[
\frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{AR_{\text{meas}}}{R_{\text{meas}} + R_{\text{ext}} + R_{\text{ext}}A} = -\frac{R_{\text{meas}}}{R_{\text{ext}}}\left(\frac{1}{R_{\text{meas}} + \frac{1}{AR_{\text{ext}}A}} + 1\right) \quad (6.1)
\]
As seen in Chapter 5 (Eq. 5.8), the feedback resistance has an exponential dependence on the gate voltage of the pmos transistor. The characterization of this resistance uses different gate voltages (V_{resist}) and measures the effective feedback resistance. The effective resistance can be derived from the linear sub threshold equation leading to Eq. (6.3).

\[
\frac{1}{R_{\text{meas}}} \approx \frac{q}{kT} \frac{q}{n} e^{\frac{q(V_{GS}-V_T)}{nkT}}
\]  

(6.3)

where \(I_o\) is a technology dependent parameter relating to the mobility, the oxide capacitance and the transistor dimensions, \(n=3\) and \(V_T=1.17\) Volts.

The comparison between measured resistance values and the above equation is shown in Figure 6.5. The effective resistance is plotted vs. the V_{resist} voltage applied externally. As can be seen from the graph and the equation, the resistance is not linearly dependent on the V_{resist} voltage, but has an exponential dependence. Thus, varying the V_{resist} voltage slightly changes the resistance significantly. This adjustment can be used to accommodate a larger range of input currents than if no adjustment of the resistance were possible. The differences in measured vs. equation values of effective resistance near V_{resist}=2 V are due to the transistor no longer operating in sub-threshold. The differences near V_{resist}=1.5 V are due to the PMOS pass gates which do not pass low voltages as well as higher voltages, thus causing the externally applied voltage to be lower than the voltage at the gate of the feedback transistor.

Additionally, the feedback resistance shows a non-linearity due to the operating point of the PMOS transistor. The non-linearity can be improved by implementing the PMOS transistor in a separate well, which removes the threshold voltage variation due to the operating point. The SPICE simulation of the PMOS transistor and measurements are shown in Figure 6.6. The output voltage of the amplifier vs. the photocurrent of the diode are shown in this figure. The nonlinearity is due to the PMOS transistor used as a transimpedance. The improvements (using a separate n-well) allow the resistance to be linear over the entire range but at much smaller output voltage values.
Figure 6.5: Resistance variation with $V_{\text{resist}}$ voltage. Theory of linear sub-threshold operating region is shown as $R_{\text{fsim}}$. Measured resistance values shown as $R_{\text{fout}}$. The differences near $V_{\text{resist}}=2$ V are due to the transistor no longer operating in sub-threshold. The differences near $V_{\text{resist}}=1.5$ V are due to the pmos pass gates which do not pass low voltages as well as higher voltages, thus causing the externally applied voltage to be lower than the voltage at the gate of the feedback transistor.

Figure 6.6: Simulated and measured output voltage vs. input current.
6.2.3 Complete pixel

Exposing one pixel to constant light power, the APD gain is shown by only increasing the high voltage (see Figure 6.7). With a high voltage of 12 V, the output swing for this particular light power and resistance setting is approximately 10 mV. These conditions correspond to a gain of 1. When the high voltage is turned up to 19.03 V (V<sub>prog</sub> = 3 V), the gain is increased and the output voltage swing is approximately 50 mV. The gain achieved using this voltage is 5. The gain is then increased even more by turning up the high voltage to 19.15 V. Here the gain is 15, but as can be seen from the figure, the output voltage is quite noisy.

![Figure 6.7](image)

**Figure 6.7:** Single pixel response to a pulse of light with A) no avalanche (gain 1), B) onset of avalanche (gain 5) and C) avalanche (gain 15). The voltages applied are 12.85 V, 19.07 V and 19.15 V for A, B and C respectively. Curve 1 is the light power. Amount of light is 3 nW during the “on” phase. V<sub>resist</sub> is set to 1.833 V. The effective resistance corresponds to 40 MΩ without avalanche gain when using 25% as the quantum efficiency for the 635 nm wavelength.

The rms noise value of the output voltage can be estimated from Figure 6.7. The peak to peak noise voltage is 110 mV. The rms voltage can be estimated by dividing the observed peak-to-peak voltage by the crest factor.
The crest factor of 4 leads to errors of less than 0.5%. Therefore, the rms voltage becomes 27.5 mVrms. Converting to the input using the 40 MΩ gives an input referred rms noise current of 687 pA rms. Calculating the total shot noise due to the avalanche photodiode (using photocurrent of 383.5 pA, the avalanche gain of 15 and its excess noise factor of 20,000 (from Chapter 4)), we obtain 24 pA/√Hz. Combining these two numbers (687 pA rms and 24 pA/√Hz), a measurement bandwidth of 800 Hz is obtained. Other measurements have shown that significantly larger bandwidths (ca. 10 kHz) are obtained using a 40 MΩ feedback resistance. This discrepancy can be attributed to the estimation of the avalanche noise at gain 15. Since the avalanche gain noise measurement was performed on a different device (integrated in a larger technology) than these measurements, it is quite possible to have such large discrepancies. The increase in noise seen in Figure 6.7, shows that the avalanche excess noise dominates the noise behavior at larger gain values. This is consistent with the large excess noise measured in Chapter 4.

6.3 Noise analysis

The noise analysis of the APD pixel is divided into two parts: diode and electronics noise. Diode noise has been analyzed and measured in Chapters 2 and 4. The electronics noise portion includes noise of the amplifier, transimpedance resistance and readout source followers. The measurement conditions have the diodes in a reverse bias of 5 V. Noise vs. gain measurements require optical input in the form of an LED and a high voltage source to achieve avalanche gain.

For all noise measurements, the row and column address of the pixel is fixed during the measurement. The power supply is a set of batteries to reduce the line noise in the measurement. The output of the chip is buffered using an external source follower (BF244) and 10 kΩ resistor. For measurements with the APD, the reverse bias voltage of the APD is supplied by the Keithley electrometer. In the electronics noise measurement setup, the reverse voltage is 5 V.

6.3.1 Spectral electronics noise (no avalanche)

The measurement board is placed in a metal box, along with the battery pack as a power supply. The only wire exiting the grounded metal box is a
coax cable, kept as short as possible. The output noise is measured using a spectrum analyzer.

To calibrate the noise measurement setup, a resistor whose noise spectrum is well known is used for calibration. When measuring small noise values, optimal shielding is necessary. In this case, such shielding is achieved through a Faraday cage with a separate power supply and filter.

The dynamic signal analyzer (HP89410A) is used to measure the spectrum of the output voltage. Since low frequency signals are most interesting for our applications, the spectral density is measured from 1 Hz to 1 MHz. The spectral response of the noise without avalanche (Vddh = 5 V) is shown on a log-log plot in Figure 6.8. Modifying the feedback resistance changes the noise contribution at the output. Two different resistance values of 10 MΩ and 60 MΩ were measured. The important details seen in this figure are the difference in the noise and the bandwidth change due to the resistance change. Thus, as expected, the largest resistance also causes the largest DC noise and the smallest bandwidth. The resistance values are different from Figure 6.5 due to the change in threshold voltage between different chips. As can be seen from eq. (6.3), the threshold voltage is in the exponent causing a large change in Rt. A threshold voltage change of 100 mV causes a resistance value change of a factor of 3.6.

The low frequency noise in Figure 6.8 can be identified as 1/fα noise where α is an exponent between 0.5 and 1.5 [Voss79]. The low frequency noise can be partially compensated using a correlated double sampling algorithm common to other CMOS sensors.

The transimpedance amplifier’s noise performance is improved when the feedback resistance is large [Senior85]. Though the output noise increases with larger feedback resistance, the transfer function also increases. This combination improves the signal to noise ratio by the square root of the feedback resistance. The disadvantage is the effect of the increased resistance on the bandwidth of the amplifier, which causes a decrease proportional to the resistance change. The noise performance in a transimpedance amplifier will always exceed that incurred by a high impedance front end (e.g. open loop amplifier or common source amplifier) due to the effective resistance noise.
6.3.2 Total noise (electronics and APD)

The total noise of the pixel including avalanche was measured using a spectrum analyzer. The gain and noise are measured simultaneously. Results are shown in Figure 6.9. The measurement is performed while the chip is illuminated using 4 nW of 635 nm LED light. The following voltage values were used for the measurement: $V_{\text{resist}}=1.923$, $V_{\text{pro}}=2.993$, $V_{\text{bn}}=1.240$, $V_{\text{bp}}=3.645$, $V_{\text{bias}}=1.696$. The theoretical noise includes noise contributions of the feedback resistance and the avalanche excess noise only. The avalanche excess noise is calculated by pure hole injection with the equation given in Chapter 2. Additionally, the feedback resistance is output current dependent because the PMOS transistor’s operating point changes with output current. All other contributions are neglected. The measured noise corresponds well to the theoretical noise contributions calculated here. The main noise contribution, especially with gain is the
excess noise of the avalanche diode. Improvements are possible and described in Chapter 8.

![Figure 6.9: Output noise voltage (per root Hz measured at 135 Hz) for different applied high voltage values. The theoretical noise includes the noise due to the feedback resistance as well as the avalanche excess noise factor and gain for pure hole injection. Additionally, the feedback resistance is output current dependent because the PMOS transistor’s operating point changes with output current. From the noise vs. gain curve, the optimal gain region is at the low gain end of the curve due to the excess noise from the avalanche diode. Thus for gains less than 10, the noise contribution of the avalanche diode is not yet dominant, allowing the noise of the feedback resistance to determine the noise performance. In the case of the transimpedance amplifier, significant gains (of more than 10) are dominated by excess noise which increases much more rapidly than the gain. For improved noise performance, an integrating amplifier is therefore the circuit of choice. Such a charge sensitive amplifier is directly comparable to other CMOS image sensors such as active pixel sensors. The advantage of such a circuit is also that the noise contribution is filtered due to the integrating characteristics and thus only measurable as an rms value for a certain]
bandwidth. This idea is pursued experimentally in Chapter 8. However, for characterization of the diode and amplifier, the transimpedance amplifier is preferred because it allows the measurement of the spectral noise characteristics, the gain variation in time as well as the noise vs. gain characterization.

6.4 APD camera realization

With these measurements the pixel operation has been shown with small gain values. Therefore, the array including all 288 diodes can now be used to build a complete camera. The 2D array already includes the two address decoders and the programming logic necessary for the camera.

6.4.1 Operation without extensive calibration

A complete camera was built using the APD array and peripheral circuitry shown in Figure 6.10. The block diagram includes the sensor, the microcontroller Microchip PIC16C55, the bias voltages and the National Instruments Data Acquisition Card (NIDAQ) for a PC. The first images from the camera do not include the extensive calibration possible with this particular pixel design.

The microcontroller supplies all the addresses for the row and column decoders. The addresses are read out sequentially. The frame and pixel clocks are used to trigger the data acquisition. Various other digital signals generated by the NIDAQ card are used for synchronization to the PIC on board. Three different bias voltages are required: Vbn, Vbp and Vbiasn. The Keithley 6517A electrometer is used to generate the programmable high voltage supply (Vddh). A photograph of the first APD camera in standard BiCMOS technology is shown in Figure 6.11.

An f/1.2 lens in front of the pixel array allows an image to be formed on the array. As a first experiment, the high voltage is set to 5 V allowing the circuitry to function without avalanche gain. The output of the sensor is buffered using an external source follower. The output signal is then digitized with a 12-bit ADC, transferred via the parallel port to the PC and displayed using software adapted for this sensor. The light level and focus are adjusted to allow recognition of the “CSEM” placed in front of the sensor (see Figure 6.12).
As expected, we observed non-negligible pixel non-uniformity slightly larger to the non-uniformity observed in APS sensors. The non-uniformity includes the traditional fixed pattern noise due to the source follower threshold voltage and level shifts. In the APD pixel, amplifier offset, gain, bias variation and the PMOS resistor variation contribute to the non-uniformity. With the calibration described in the next section, these effects can be compensated to a certain extent.

The APD gain is only necessary when the amount of light is not sufficient to produce an image. Increasing the high voltage increases the avalanche gain. Unfortunately the avalanche gain cannot be increased indefinitely because of the dark current multiplication and the limited output swing of the operational amplifier. The resulting combination is a pixel that is
“white” despite the lens being covered with a lens cover. The avalanche gain begins at approximately 18.61 V when the first pixel appears as a white spot. All pixels are white at a voltage of 19.51 V. These values are dependent on the actual transimpedance resistance (e.g. 60 MΩ) and the gain and offset settings of the parallel port adapter which define the 8 bit gray levels displayed on the PC. However, this shows the range of variation of the avalanche gain across the entire chip.

Figure 6.11: The first APD camera in standard BiCMOS technology using the 12 x 24 pixel image sensor. DC-DC converter creates the higher voltage (ca. 18 V) necessary for avalanche gain.

Figure 6.12: CSEM image taken by using the 12 x 24 pixel image sensor four times. Light level necessary for these images was ca. 1400 Lux. Bias voltages: Vresist=2.8 V, Vddh=5 V, Vprog=3.7 V, Vbiasn=0.864 V, Vbp=3.476 V, Vbn=1.290 V. No fixed pattern noise correction.

The most important two images are compared: one without avalanche, one with avalanche. The amount of gain achievable is presently limited by the output swing but two images are shown in Figure 6.13. The top image shows the “CSEM” under low volt conditions. With plenty of light, this
image resembles Figure 6.12. The level of light necessary to produce a legible image is approximately 1400 Lux. With avalanche gain, the light level is about 200 Lux. Thus, a gain of 7 is visible in the two images (dark5, csem17) of Figure 6.13. Vresist is 2.8 V. Images with gains of up to 15 have been achieved.

The avalanche gain actually allows an image to become visible though the light conditions do not allow the same sensor without avalanche gain to see the image. The quality of the image is limited by the small number of pixels in this array. Additionally, the noise level and the small output swing of 1 V prevent us from increasing the gain significantly. This particular set of images proves that avalanche gain is possible but there is significant room for improvement. These possibilities are discussed in the Chapter 8.

![Figure 6.13](image-url)

**Figure 6.13:** Two images: top without avalanche, bottom with avalanche. The light level in both cases is approximately 200 Lux. To allow better resolution, each letter image was taken separately. The avalanche voltage on the bottom image is 19.09 V. The avalanche gain is approximately 7. Vresist voltage is 2.8 V. No fixed pattern noise correction is used.

### 6.4.2 Operation with extensive calibration

A calibration scheme to adjust the individual feedback resistances and operating points of each pixel is presented. The goal is to compensate variation across the chip to improve uniformity. In reality, the results show that such a calibration is not necessary for arrays of this size (12 by 24 pixels). In larger arrays however, this calibration is useful to equalize the product of the diode current and the feedback resistance across the entire chip.
6.4.2.1 Method

The calibration program’s flow chart is shown in Figure Z. The program begins with a Vresist calibration where the Vresist voltage is adjusted to allow a homogeneous sensitivity in non-avalanche mode. Using this sensitivity, the avalanche mode is also calibrated by finding the proper operating point. After the high voltage (Vddh) is calibrated, the Vprog voltage can be adjusted to allow a common Vddh yet individually adjusted applied voltage across the diode (not shown in flow chart). These adjustments are repeated for each pixel.

Figure 6.14: Simplified flow chart of calibration program. Vresist adjustment occurs before Vddh adjustment. These adjustments are repeated for each pixel.

6.4.2.2 Hardware and software

The hardware setup necessary to implement this calibration is shown in Figure 6.10. Required programmable voltage sources include:
- high voltage source (Vddh) of up to 25 V (e.g. Keithley 6517A)
- reference voltage source (Vprog)
- resistance voltage source (Vresist)
- LED voltage or current source (VLED) to control the light intensity

Required digital signals include:
- row address
- column address
• pixel and frame clock
• synchronization with NIDAQ and LED

The only measurement equipment used in the calibration is the NIDAQ card. To observe the results of the calibration, a PC and a 12-bit ADC were used.

6.4.2.3 Results

The high voltage source (Vddh) is adjusted to allow the output voltage to be between 100 mV and 300 mV. The gain for each pixel is recorded in two cases: calibrated (adjusted Vddh) and uncalibrated. Figure 6.15 shows the results of the calibration program. The adjusted voltages range between 20.64 and 20.74 V. The uncalibrated high voltage is 20.7 V. The improvement in uniformity is clearly visible.

![Figure 6.15: Gain uniformity after calibration compared with no calibration. The applied voltage for the uncalibrated curve is 20.7 V. The adjusted applied voltage varies from 20.64 to 20.74. The improvement in uniformity is clearly visible.](image)

Though the gain uniformity shown is significant, in evaluating the images in Figure 6.12 and Figure 6.13, the uniformity presented is sufficient to allow operation without adjustment. During pixel design, no indication of uniformity was known, thus the calibration was included as a safety measure for operation improvement.

6.5 Dynamic range

The dynamic range of the sensor is defined as the maximum output swing divided by the input noise voltage in the dark. Using the noise level
measured with the spectrum analyzer as seen in Figure 6.8 (1.0 μV/√Hz), and normalizing to a frequency of 10 Hz, the rms noise level is approximately 3 μV rms. The maximum output swing after the two source followers is 0.8 V. Therefore, the dynamic range can be calculated to 108 dB.

6.6 Conclusions

This chapter has shown the experimental results of an APD array implemented in a 1.2 μm BiCMOS technology. As expected, the breakdown voltage of the individual diodes is significantly lower than the diodes in Chapter 4 due to the increased doping concentration of the n-well, typical of smaller technologies. However, since the avalanche gain has been shown, we know that we have not yet reached the doping concentration where the carriers cross the barrier through tunneling rather than avalanche. Thus, this avalanche gain can still be used in this technology. To the best of our knowledge, this is the first functional avalanche photodiode in standard BiCMOS or CMOS technology. No other APDs in standard technology have been presented in the literature.

The amplifier implemented is also the first amplifier presented of this size. No other similar sized operational amplifiers are found in the literature. The characterization of the amplifier shows the open loop gain and bandwidth to be 150 and more than 7.7 MHz respectively. For a single stage, 5 transistor amplifier in an area of ca. 6000 μm², the achieved gain bandwidth product is quite respectable in a 1.2 μm technology, requiring a very compact layout.

The characterization of the feedback resistance shows a controlled variation of 270 kΩ to 40 MΩ. The resistance is dependent on the current through the resistor (i.e. the operating point of the amplifier) which creates an operating point dependent transfer function. This undesired effect can be practically eliminated by placing the resistor in a separate n-well. The additional area required for the extra well increases the pixel size by approximately 10 %. In a future transimpedance pixel design, the separate n-well should be implemented to remove the resistor’s operating point dependence.

The transimpedance pixel implemented allows the time dependence of the gain and the spectral noise to be measured directly. The output voltage is
directly proportional to the multiplied diode current through the transimpedance resistance. The measurements have shown clearly that noise is essentially due to this resistance and the excess noise of the diode (when gain is involved). The noise performance of the pixel can be improved by implementing an integrating pixel where the resistance noise is replaced by reset noise and the excess noise bandwidth is dictated by the integration time. However, with such an integrating pixel the spectral noise characterization and the observation of the gain variation are no longer possible in such a direct manner.

The camera realization shows that the 12 by 24 pixel array can be used to create APD images. In these images, gains of 15 were utilized. Calibration has been implemented but the resistance and breakdown voltage variation over the entire chip are small enough to allow operation even without extensive calibration.

For a camera implementation with increased sensitivity the pixel needs an integrating architecture, in the APD case involving an entire operational amplifier in each pixel. The amplifier implemented in the transimpedance pixel is an excellent example of area vs. gain and bandwidth. Decreasing the area is only possible by using submicron technologies. This is examined in detail in Chapter 8. In terms of noise, the integrating amplifier is better because the bandwidth of the diode's excess noise is reduced and the large feedback resistance is removed. These are the two dominating noise contributions in the implemented pixel. The dominating effect in the integrating pixel is expected to be reset noise (as shown in Chapter 7). Thus, for the improved sensitivity and noise behavior, the next APD camera implementation would require an integrating pixel. To build an integrating pixel an amplifier, feedback capacitance and reset switch are necessary. The amplifier implemented here can still be used. The feedback resistance in the transimpedance pixel should be changed into a reset switch and an operating point independent feedback capacitance needs to be added. This capacitance should be made using poly 1/poly 2 layers to prevent an operating point dependence.
7 Comparison with other imagers

The goal of this chapter is to compare our APD imager with conventional CMOS imagers. In particular, three areas are important: noise, dynamic range and circuit complexity. The sensor chosen for comparison is an active pixel sensor (256 by 256 pixels) [Vietze97]. One of the difficulties of the comparison is that integrating and non-integrating sensors can not be compared directly. Calculations are necessary to convert the measurement values of the non-integrating sensor, using an integration time, to values comparable with integrating sensors. Because of the implemented transimpedance amplifier, the characterization of the diode and pixel are more direct, but in comparison to standard active pixel sensors (APS) the noise level is significantly higher.

7.1 Noise

The equivalent light induced input power necessary to create a signal as large as the noise in an image sensor is the noise equivalent input power. Because the comparison involves light power, different fill factors and quantum efficiencies are included in the calculations. Thus, the significant variation between image sensors has to be attributed in part to these two factors.

The noise sources present in an APS pixel, an APD with transimpedance amplifier (APD T) and an APD with integrating amplifier (APD I) are listed in Table 7.1. To allow a better comparison, the shot noise of the diode (without avalanche) is assumed to be the same in all pixels. The source follower noise is also assumed identical for all pixels. The amplifier input referred noise is also the same in both types of APD pixels.

The excess noise of the APD is described in Chapter 2. This noise depends on the multiplication factor set by the applied voltage of the diode. The APS pixel has fewer noise components, most of which are also present in the APD pixels. Thus, the noise performance of the APS pixel is the best of the three pixels described here. The APD I pixel is the second best structure because only the amplifier noise is added to the pixel. The APD T pixel noise behavior is usually dominated by the feedback thermal noise (when no APD gain is present) as seen in Chapter 6.

The comparison between integrating and non-integrating pixels requires conversion calculations since the transfer functions are different. Both the
APS and APD I structures convert the photocharge to a voltage using a conversion capacitance $C$. The APD T pixel converts the photocurrent to a voltage using the feedback resistance $R_f$. Though charge is the integral of current over time, the comparison is mathematically slightly more complex because noise current may not be integrated directly to get noise charge.

<table>
<thead>
<tr>
<th></th>
<th>APS</th>
<th>APD T</th>
<th>APD I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diode</td>
<td>Shot</td>
<td>Shot + excess</td>
<td>Shot + excess</td>
</tr>
<tr>
<td>Reset</td>
<td>$kT/C$</td>
<td>None</td>
<td>$kT/C$</td>
</tr>
<tr>
<td>Source follower</td>
<td>Thermal + $1/f$</td>
<td>Thermal + $1/f$</td>
<td>Thermal + $1/f$</td>
</tr>
<tr>
<td>Amplifier</td>
<td>None</td>
<td>Thermal + $1/f$</td>
<td>Thermal + $1/f$</td>
</tr>
<tr>
<td>Feedback</td>
<td>None</td>
<td>Thermal $(4kTR_t)$</td>
<td>See reset</td>
</tr>
</tbody>
</table>

Table 7.1: Noise sources present in different image pixel circuits. Comparison between APS and APD pixels with transimpedance or integrating amplifiers (APD T and APD I respectively).

![Figure 7.1: Figures for conversion of time dependent current into a rms noise charge, as shown in equations (7.1)-(7.5).](image)

In the following, we present a mathematical method for the comparison of integrating and non-integrating photosensor pixels. To calculate the integrated charge $Q(t)$, the time dependent current $I(t)$ must be convolved with a rectangle function of length $T_{int}$, $\text{rect}_{T_{int}}(t)$ (eq. (7.1)) where $T_{int}$ is a measure of the integration time used. A graphical representation is shown in Figure 7.1

$$Q(t) = I(t) * \text{rect}_{T_{int}}(t)$$ (7.1)

Applying the Fourier transform to both sides of the equation results in Eq. (7.2). The sinc function stems from the Fourier transform of the rectangle function $\text{rect}_{T}(t)$. 
To calculate the rms charge noise, the power spectrum (the square of the absolute Fourier transform) must be integrated over frequency as in Eq. (7.3). If the power spectral density $|F_f|^2$ is essentially constant for frequencies up to several times $1/T_{\text{int}}$ (the first zero of the sinc function), we can approximate:

$$\Delta Q_{r,m,s}^2 \equiv \left| F_f \right|^2 \cdot T_{\text{int}}^2 \cdot \int_{-\infty}^{\infty} \frac{\sin^2 \left( T_{\text{int}} \pi f \right)}{(T_{\text{int}} \pi f)^2} df$$  \hspace{1cm} (7.3)

The integral term can be evaluated to yield (Eq. (7.4)):

$$\int_{-\infty}^{\infty} \frac{\sin^2 \left( T_{\text{int}} \pi f \right)}{(T_{\text{int}} \pi f)^2} df = \frac{1}{T_{\text{int}} \pi} \cdot 2 \cdot \frac{1}{2} = \frac{1}{T_{\text{int}} \pi} \cdot \frac{\pi}{2} \hspace{1cm} (7.4)$$

We therefore obtain (Eq. (7.5)):

$$\Delta Q_{r,m,s}^2 \equiv \left| F_f \right|^2 \cdot T_{\text{int}}^2 \cdot \frac{1}{T_{\text{int}}} = \left| F_f \right|^2 \cdot T_{\text{int}}$$ \hspace{1cm} (7.5)

The current spectral density $F_f$ ($[F_f^2]=\text{A}^2/\text{Hz}$) is multiplied by the integration time $T_{\text{int}}$. The square root of this product is the RMS charge ($\Delta Q_{\text{RMS}}$).

The advantage of the integrating pixels (APS and APD I) is that the noise of the diode is bandwidth limited by the integration time. For example, the APD T pixel has an amplifier bandwidth of 10 kHz. The APD I pixel has a bandwidth of 10 Hz for the noise from the diode. Comparing only the shot and excess noise of the diode using the respective bandwidths of the pixel, the noise contribution of the APD T pixel is $\sqrt{1000}$ or 31 times larger than the APD I pixel.

The output noise voltage of the APD T pixel includes the noise due to the feedback resistance which is not present in the APD I pixel. The equivalent APD T noise is expected to be significantly larger (for large resistances) than an integrating pixel’s noise for long integration times.

The APD T pixel is operated without light and at a gain of 1. Using the measured output spectral noise density of 1.0 $\mu\text{Vrms}/\sqrt{\text{Hz}}$ (from Figure 6.8) and a transimpedance resistance of 60 M$\Omega$, the input spectral noise density squared is approximately $277 \times 10^{10}$ $\text{A}^2/\text{Hz}$. Depending on the integration time, the RMS charge is calculated in Table 7.2.
Table 7.2: Calculated equivalent number of noise electrons at input for APD pixel without light and a gain of 1. Transimpedance of 60 MΩ, output noise spectral density of 1.0 µVrms/√Hz and sufficient amplifier bandwidth are assumed.

Typical values for APS structures in terms of rms noise are approximately 110 electrons at 20 ms integration time [Vietze97]. The comparable value for APD T structure is 14700 electrons. The major contribution of the noise is due to the feedback resistance which dominates the noise behavior.

Table 7.3: Minimum detectable light power is measured. Light power is calculated from measured light intensity and filter transmission values. Quantum efficiency (635 nm) is 25%. Amplitude is measured with lock-in amplifier. The effective resistance is the ratio of the output voltage (corrected for the source follower attenuation) to the input current. Effective transimpedance is approx. 40 MΩ without avalanche gain.

The noise floor of the lock-in amplifier and the pixel is measured to be 0.4 µV or three times smaller than the output amplitude at 18.5 V Vddh. Therefore, the minimum detectable signal is three times smaller than 31.8 fA, detectable only with the avalanche gain at 18.5 V. From the
effective resistance calculation, the gain is 1.6. Thus, 10 fA is the minimum detectable signal at a gain of 1.6. The equivalent light power on the diode is 49 fW using a 25% quantum efficiency at a wavelength of 635 nm, and an avalanche gain of 1.6.

The noise equivalent power for the APS256 sensor is given as 7.4 μW/m² [Vietze97]. The area of the pixel (15 μm x 15 μm) and the fill factor (35%) are used to calculate the minimum detectable light power for the APS256 sensor of 0.6 fW. Comparing directly, the APS256 sensor is 20 times more sensitive than the APD pixel. A factor of almost 3 is due to the smaller quantum efficiency of the APD pixel, 25% vs. 70%. The remaining factor of 7 is due to the large transimpedance of 40 MΩ that dominates the noise behavior of the APD pixel and the excess noise contribution of the avalanche diode.

![Figure 7.2: Signal and noise block diagram of detection chain with charge sensitive amplifier. The incoming light P is transformed to a charge which is multiplied by the avalanche gain. The charge to voltage conversion occurs through a charge sensitive amplifier. The row and column selection is implemented as a source follower. The noise contributions are shown.](image)

The best low noise pixel design is an integrating pixel. However, this pixel could not be implemented due to a sudden foundry closure, therefore, the noise analysis is completed using theoretical calculations. In the integrating APD pixel, the feedback resistor of the APD T pixel is replaced by a feedback capacitance. This capacitance needs to be reset by a switch to keep the amplifier within its operating range. This reset switch and the capacitance cause reset noise which corresponds to kT/C (V²). The noise can be decreased by increasing the capacitance, but this also decreases the signal amplitude also. The signal amplitude is the ratio of the input capacitance (including diode capacitance) to the feedback capacitance.
The smaller the feedback capacitance, the larger the signal gain, but the larger the reset noise also. The minimum reasonable capacitance available as a double poly feedback capacitance in 2 μm technology is on the order of 200 fF. In this specific case, the gain from diode to amplifier output is almost one. The noise improvement over the APD T pixel however is significant. The noise contributions are shown in numbers in Table 7.4. The bandwidth decrease (factor of 31) and the removal of the feedback resistance (4kTRfB) replaced only by reset noise (kT/C), lead to significant improvement in noise performance. The amplifier noise components that were neglected in the APD T pixel noise analysis are no longer negligible in the APD I pixel. Eq. (7.6) shows the sum of the thermal noise components ($v_{\text{nosarms}}$) at the output of the operational amplifier.

<table>
<thead>
<tr>
<th>Noise contribution</th>
<th>Bandwidth</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source follower</td>
<td>17 nV/√Hz</td>
<td>1 MHz</td>
</tr>
<tr>
<td>Amplifier</td>
<td>30 nV/√Hz</td>
<td>10 Hz</td>
</tr>
<tr>
<td>Reset</td>
<td>144 µV</td>
<td>-</td>
</tr>
<tr>
<td>Diode (gain=1)</td>
<td>$3.2 \times 10^{-35} \text{A}^2/\text{Hz}$</td>
<td>10 Hz</td>
</tr>
<tr>
<td>Diode (gain=15)</td>
<td>$14.4 \times 10^{-27} \text{A}^2/\text{Hz}$</td>
<td>10 Hz</td>
</tr>
</tbody>
</table>

Table 7.4: Noise contributions and bandwidth for an integrating pixel. Comments show how the values were calculated.

The source follower ($v_{\text{nsf}}$), the amplifier ($v_{\text{namp}}$), reset noise of the 200 fF capacitance ($kT/C$), and the excess and shot noise of the avalanche diode ($i_{\text{nAPD}}$) are summed quadratically as in Eq. (7.6).

$$v_{\text{nosarms}}[V] = \sqrt{\frac{i_{\text{nAPD}}^2}{C^2} \cdot T_{\text{int}} + v_{\text{namp}}^2 \cdot B_2 + \frac{kT}{C} + v_{\text{nsf}}^2 \cdot B_1} \quad (7.6)$$

$$i_{\text{nAPD}}[\sqrt{\text{Hz}}] = \sqrt{2qI_p M^2 F_{\text{APD}}} = \sqrt{S_f} \quad (7.7)$$

The bandwidth used for the source follower is 1 MHz ($B_1$), while the bandwidth of 10 Hz ($B_2$) is used for the noise of the avalanche diode and the amplifier. $T_{\text{int}}$ is the integration time of 0.1 second. The photocurrent used is 10 fA. The result is dominated by the kT/C noise of the 200 fF capacitance. Thus, the thermal noise of gain of 1 is 173 µVrms. In other CMOS active pixel sensors, the kT/C noise also dominates [Vietze97]. The noise of gain of 15 on the other hand is completely dominated by the avalanche diode’s excess noise. The noise total of gain of 15 is 189 mVrms.
Significant 1/f noise is seen in Figure 6.8 but when the feedback resistance is sufficiently large, the resistance noise dominates. In comparison the integrating pixel must show flicker noise, which will increase the noise contributions significantly. In Vietze97, for example, 1/f noise is considered to be on the order of kT/C noise [Vietze97]. 1/f noise can be mostly compensated using a correlated double sampling circuitry for short integration times.

The resulting equivalent noise of the APD I pixel calculated above is 216 electrons. This value is much closer to the APS noise where 115 noise electrons are calculated [Vietze97] but only considers the thermal noise. A graphical representation is shown in Figure 7.3. The figure shows that the kT/C noise dominates total noise behavior. However, the kT/C noise can be compensated. Thus, the dominating noise becomes the Poisson or shot noise of the diode current (depending on the integration time Tint). The source follower thermal noise is non-negligible due to the large source follower bandwidth of 1 MHz. Due to the small amplifier bandwidth the amplifier noise is negligible in this case.

Figure 7.3: Noise components of the integrating pixel. Two integration times are used. $F_{APD} = 1$ for these calculations.

Increasing the Poisson noise by adding the $F_{apd}$ factor causes the dominant noise factor to be the shot and excess noise of the avalanche diode. For the integrating pixel, Eq. (7.6) and the values shown in Figure 7.3 with $I_p=10^{-11}$ A and $T_{int}=0.001$s are used to calculated the noise of 330 $\mu$Vrms. Using this integration time, the Poisson noise alone is 300 $\mu$V. If kT/C noise is cancelled using correlated double sampling, the noise behavior of
the integrating pixel is shot noise limited for diode currents of more than 1 fA.

In comparison, the total rms output noise for the transimpedance amplifier is calculated as shown in Eq. (7.8) and (7.9). The values are taken from Figure 5.10 in Chapter 5 with a bandwidth of $B_{\text{lia}}=1 \text{ MHz}$, $R=10 \text{ M} \Omega$ and $I_p=10^{-11} \text{ A}$. The result is $V_{\text{rms}}=450 \mu \text{Vrms}$.

$$\frac{V}{\sqrt{Hz}} = \sqrt{2qI_pF_{\text{APD}}M^2R^2 + v_{\text{namp}}^2 + 4kTR + v_{\text{nsf}}^2} \quad (7.8)$$

$$v_{\text{notiarms}}[\text{V}] = \sqrt{v_{\text{ntot}}^2 \cdot B_{\text{lia}}} \quad (7.9)$$

The noise comparison between the transimpedance amplifier pixel ($v_{\text{notiarms}}$) and the charge sensitive amplifier pixel ($v_{\text{noesarsms}}$) shows that the integrating pixel has the better noise performance for this integration time of 0.001 s.

For avalanche diodes with $M>1$, the optimal $M$ is achieved by maximizing the signal to noise ratio for the given photocurrent. Figure 7.4 shows an example of the signal to noise ratio for the given photocurrents of $1 \text{nA}$, $0.9 \text{nA}$ and $0.8 \text{nA}$ for the transimpedance amplifier with $10 \text{ M} \Omega$ feedback resistance.

![Figure 7.4: Signal to noise ratio for different gain values at specific photocurrents of $1 \text{nA}$, $0.9 \text{nA}$ and $0.8 \text{nA}$ for the transimpedance amplifier with $10 \text{ M} \Omega$ feedback resistance.](image-url)
Figure 7.5: Signal to noise ratio as a function of current and gain. The smaller photocurrents require larger gain for optimized signal to noise ratio.

In Figure 7.5, a three dimensional plot of the signal to noise ratio vs. current and gain is shown. The graph shows that larger gains are required to achieve the best signal to noise ratio for a particular photocurrent. As the photocurrent increases, there is a break even point where gain =1 achieves the best signal to noise ratio. At this point (ca. 9 nA in this case), the avalanche gain no longer improves the signal to noise ratio. This point is specific to the device and depends critically on the noise behavior of the diode. The following assumptions were used to calculate the plot in Figure 7.5. 1) The dark current is negligible (not necessarily true for CMOS sensors). 2) Amplifier and source follower noise are neglected from Eqs. (7.8) and (7.9). Only excess noise and thermal noise due to the resistance are used. 3) The excess noise factor F is equal to the multiplication factor M (F=M). 4) Resistance R=10Meg, and bandwidth B=100kHz are used for the calculations.

Summarizing: for each input photocurrent, given that the noise behavior of the device is known, an optimal gain can be calculated to achieve the best possible signal to noise ration. Unfortunately, realistically speaking, this involves measuring the photocurrent of each pixel and optimizing the gain
individually. However, if I can measure the photocurrent without gain, then I do not need the APD, which was supposed to improve the detectability. But, theoretically, an optimized signal to noise ratio can be found for each pixel’s photocurrent.

Another method of finding the best signal to noise ratio is to find the maximum mathematically, by taking the derivative of the signal to noise ratio and setting it equal to zero. If the noise factor $F=M^2$, the maximum is achieved as in Eq. (7.10). When $F=M$ rather than $M^2$, the maximum shifts by a factor of 2. This is consistent with literature where the maximum is achieved by splitting the total noise into two approximately equal components, one from the diode, the other from the electronics [Mueller79, Paul92]. Thus, for the transimpedance case where the noise is dominated by the feedback resistance, (Eq. (7.10)) shows the diode noise equal to the resistance noise.

$$2qI_{p}FM^2R^2B_{tia} = 4kTRB_{tia}$$  \hspace{1cm} (7.10)

Solving this equation for $FM^2=1$ (no avalanche gain) and $R=10 \, \text{M}\Omega$ results in the current $I_p$ where the signal to noise ratio is better without avalanche gain than with it. This limit is 5 nA for these conditions. The signal to noise ratio is shown in Figure 7.6 for $M=1,5$ and 10.

![Figure 7.6](image)

**Figure 7.6:** Signal to noise ratio for a transimpedance amplifier with $10 \, \text{M}\Omega$ feedback resistance and gains of 1,5 and 10 from the avalanche diode. Excess noise factor of the pure electron injection diode described in Chapter 2 is used.
For the charge sensitive case, the equation is slightly more complex because one noise source no longer dominates the electronics noise behavior. Thus, as in Eq. (7.11), the diode noise is equated with the sum of the electronic noise components (kT/C, amplifier noise and source follower noise).

For the charge sensitive case the equation is slightly more complex as shown in Eq. (7.11).

\[ \frac{2qI_p FM^2 T_{int}}{C^2} = \frac{kT}{C} + v_{namp}^2 B_2 + v_{nsf}^2 B_1 \]  

(7.11)

Solving this equation using FM^2=1 (no avalanche gain) for an integration time of 0.1 seconds and a capacitance of 200 fF, the signal to noise ratio becomes better for a non-avalanche diode at 30 fA diode current. This value is very dependent on the integration time chosen as seen from Eq. (7.11). A graphical representation of the signal to noise ratio vs. input current is shown in Figure 7.7.

**Figure 7.7:** Signal to noise ratio for a charge sensitive amplifier with an integration time of 0.1 seconds and gains of 1, 5 and 10 from the avalanche diode. Excess noise factor of the pure electron injection diode described in Chapter 2 is used.
7.2 Dynamic range

The dynamic range of an image sensor is defined as the ratio in dB of the maximum output voltage swing to the output noise voltage. The dynamic range of the avalanche diode can be defined similarly using maximum output current and output noise current but the maximum output current is not limited other than by breakdown of the device. Additionally, the dynamic range must be defined at a particular gain. The operating range of the system can be increased by adjusting the gain through the applied bias or the feedback resistance (through \( V_{\text{res}} \)). The output voltage swing is limited by the amplifier’s design and for the current chip is approximately 1.5 volts.

The dynamic range of the pixel is dependent on the noise level. Using the noise level measured with the spectrum analyzer (1.5 μV/√Hz) without avalanche, and normalizing to a frequency of 10 Hz, the rms noise level is approximately 5 μV rms. The maximum output swing after the two source followers is 1 V. Therefore, the dynamic range can be calculated to 106.4 dB. The maximum dynamic range occurs where minimum noise is present. Since the noise increases with gain, but the output swing remains the same, the dynamic range decreases with gain.

The dynamic range of the APS256 sensor is 70 dB. When the output swing size is similar (ca. 1 V), the dynamic range is determined by the input noise level. In the APD sensor case, bandwidth of the system used for noise calculations is 10 Hz, whereas the system bandwidth for integrating sensors is determined by the integration time. The value of 70 dB is achieved using an integration time of 0.16 seconds, corresponding to a bandwidth of approx. 6.25 Hz. The dynamic range of 106 dB of the transimpedance pixel is significantly larger than the APS256 sensor.

7.3 Circuit complexity

Comparing active pixel sensors and photodiode arrays, the active pixel sensor with at least one active element in each pixel was significantly more complex than the photodiode array with the light sensitive elements only. In the implemented avalanche photodiode imaging array, the complexity is increased additionally by the amplifier required to control the avalanche bias voltage. One measurement of the complexity is the fill factor, which is defined as the ratio of light sensitive area to total pixel area. The fill factor
of the low light level pixel is 65%, whereas the fill factor of the APD pixel is 20%.

The complexity of the amplifier is limited by the area of the pixel. In the transimpedance pixel, an amplifier is in each pixel. The thermoelectric infrared imaging microsystem described by [Menolfi97] is an excellent example of an increasingly complex amplifier used to amplify very weak signals in the sub-microvolt range. The complex amplifier is implemented only once on chip, similar to previous photodiode arrays. In such a configuration, the area of the amplifier can be large, thus allowing more freedom in design to improving the noise, offset and drift characteristics. The amplifier presented in [Menolfi97] is the same size as the entire 12 by 24 pixel array. On the other hand, the disadvantages observed in MOS photodiode arrays would also be present in such a solution. The large capacitance of the output line must be driven by the photocurrent. Additionally, the noise level (of integrating pixels) scales with the imager size due to the large capacitance of the output lines. In the implemented transimpedance pixel, the amplifier in each pixel decouples the output line capacitance from the diode capacitance.

7.4 Applications

Avalanche photodiodes on CMOS exploit the combination of current gain with analog electronics. The applications are based on systems where the increased current of the avalanche diode is beneficial. The noise characterization of these applications is not considered in detail.

In time of flight range imaging, light is emitted from a source and received very close to the source after being reflected from the object. The time delay between the source and detection of the light is directly proportional to the distance the light traveled [Spirig97, Lange99]. The object reflects this light with a range dependent phase shift. This phase difference between sent and received signal is measured. The APD sensor could be used in two manners: 1) as a more sensitive detector and 2) as a demodulator and detector combined. The more sensitive detector would require an APD integrator to improve the signal strength and reduce noise. The less cooperative targets or targets further away might still be detectable using an APD. In addition, the power of the active illumination source could be reduced. To operate as a combination demodulator and detector the APD gain would be modulated by varying the applied voltage across the diode. Thus, the sensitivity of the array would be modulated allowing the incoming light to be mixed with this demodulation frequency.
The result is a combined demodulator detector with increased sensitivity. Requirements are limited to a fast variation of the high voltage source and an integrator to collect the increased current.

Another application includes any type of signal observed using a lock-in detector. Since the implemented amplifier is a non-integrating system, the output signal should be proportional to the input signal at lower frequencies. Thus, any slowly modulated signal should be directly detectable with the APD array.

One specific application includes observing fluorescence. The fluorescence signal is related to the excitation signal and the density of fluorescent labels bound at the surface [Voirin98, Zeller98]. In this setup, the fluorescent labels are excited using a laser beam. The illumination depends on the number of labels bound within the excitation region. Thus, if 10 labels are within 1 cm² or 1 mm², the illumination level is the same. One method of detection of the fluorescence is by a PIN diode and a lock in amplifier. With an APD array, the signal can be detected with positional information and a signal gain, thus perhaps detecting even smaller concentrations.

An additional application would be observing the fluorescent decay using an APD array. The fast detection of light within an APD might be able to provide the time resolution required to observe decay. Two examples of decay times show that a wide range of decay times exists. Potassium, for example has a decay time of ca. 30 ns, whereas the oxygen decay time can be as long as 3.1 µs [Wolfbeis98]. Unfortunately, the amplifier and readout stage must also achieve this speed to be useful unless some type of analog storage is available on chip. Such analog storage would allow very fast picture acquisition rate while keeping the readout rate at traditional speeds. However, the combination of analog storage and APDs is far from being implemented. A first step in this direction is the combination of APDs and readout electronics in BiCMOS which has been shown.

7.5 Conclusions

The comparison between APS and APD transimpedance pixels is non-trivial due to the integrating and non-integrating features. Thus, the output noise voltage is not directly comparable. Calculations to normalize the non-integrating measurements to compare to the APS measurements are presented. The transimpedance pixel is not ideal in terms of noise but
the pixel allows a more detailed characterization of the diode and its time-varying features as well as the noise spectrum in the frequency domain. This characterization is the emphasis of the thesis, allowing the final conclusion to determine in which applications such avalanche photodiodes are beneficial to the system. The increased responsivity and reduced noise requires an integrating pixel with a small feedback capacitance. Noise behavior of such a circuit is analyzed, resulting in a similar noise value as the current CMOS APS.

Applications presented in this chapter would benefit from the increased signal level of avalanche photodiodes. However, the technological problems presented in this thesis (including the breakdown voltage shift and the large noise present) must be improved before the benefits of the avalanche gain are clearly visible. The realized camera in Chapter 6 is the first APD camera implemented in a standard BiCMOS technology. This is a prime example of how APDs and CMOS electronics can be combined on one chip to form an APD camera. Simple improvements for APD cameras, derived from the results obtained in this thesis are suggested in the next chapter.
8 Improvements and Conclusions

8.1 Improvements

Specific improvement possibilities for this sensor are divided into a few sections: diode improvements, general improvements, high speed improvements and low light level improvements. General improvements pertain to all types of applications. Specific solutions and improvements for high speed and low light level applications are suggested also.

8.1.1 Diode improvements

Improvements in the diode itself would be beneficial for all applications. The breakdown voltage change is the most disturbing feature of these diodes. Basically, such problems have already been solved in dedicated APD technologies, but normally require a process change involving a better surface passivation. The only other solution is to place a metal layer very close to the surface, to prevent oxide charge build up. Though examples of such devices have been integrated and measurements have shown controllability, more device experiments are necessary. Stabilizing the breakdown voltage using burn-in techniques should also be investigated.

Low noise avalanche gain is best when electrons are the avalanche initiating carriers. Since the diode implemented here is a p+ in n-well diode, holes are the initiating carriers, causing a disadvantage in terms of low noise gain. Implementing the entire circuitry in an n-substrate technology with pnp option would prove advantageous. The pnp option would be necessary to provide guard rings to the n+ regions within the p-well. Thus, the low noise avalanche gain would be possible.

8.1.2 General circuit improvements

When integrating larger sensors, careful consideration of the resistances in supply lines is important. The pass gates used for programmability in this sensor can be improved by using complementary transistors. The controlling digital signal lines should be in a metal layer unless the lines are static. If the lines are to be clocked, the distributed R-C characteristics must be considered when analyzing performance.
In a two dimensional sensor with integrated amplifiers in each pixel, the fill factor is small. To improve the fill factor, smaller amplifiers are necessary. Since the technology minimum size is decreasing consistently with Moore’s Law, smaller technologies are available. As an additional benefit the bandwidth and speed of amplifiers in smaller technologies improves. Furthermore, the fill factor can also be improved by adding a microlens array to the sensor array. A linear array instead of a two dimensional array also reduces the fill factor problem. The resulting sensor would have elongated pixels, where the light is focused on the diodes and the circuitry fills the rest of the pixel.

8.1.3 High speed improvements

For high speed sensing, a transimpedance amplifier can be used. The transistor used in this amplifier to create the feedback resistance can be improved by using a separate well. Though this requires more area, the transistor will have a larger operating range and a smaller, constant threshold voltage. Thus, the resistance is more predictable and less operating point dependent. The amplifier’s output voltage range can also be improved to increase the output swing. One suggestion is to use a slightly different input topology such as a pmos differential input pair. Since the output voltage swing is only important in one direction, such a topology would not restrict the output swing. The other suggestion is to add a second stage to the amplifier. Both suggestions allow the output voltage to decrease further to the lower supply voltage without being restricted by the reference voltage at the input. This output swing improvement is beneficial for all circuits involving an amplifier.

8.1.4 Low light level improvements

For low light level sensing, an integrator should be used because charge is then collected and stored and the noise level is reduced by the decreased bandwidth. Instead of the transimpedance amplifier, a charge sensitive amplifier can be used as the integrator. The output swing improvements are beneficial here also. The charge sensitive amplifier design must consider the ratio of the diode and input capacitance to the feedback capacitance. The feedback capacitance must be large enough to dominate over the voltage dependent parasitic capacitances in order to have a linear, operating point independent output. The input capacitance of the amplifier should also be minimized such that the diode capacitance dominates. Unfortunately with a diode capacitance of 25 fF without avalanche and 15
130

fF in avalanche as in the integrated devices, this requirement is very
difficult to meet. The amplifier’s design trades these input capacitance
requirements with the open loop gain and bandwidth of the amplifier.
However, a two stage amplifier could improve the characteristics including
output swing, open loop gain and bandwidth.

Through some relatively simple changes in design and topology, an
integrator can be fashioned using an amplifier and a feedback capacitance.
With such an integrator, a direct comparison to the integrating APS circuit
can be made. In addition, amplifiers with significantly smaller noise
voltages have been designed and integrated, although the area
consumption is significantly larger than the area used in the APD pixel
[Menolfi97]. In fact, this particular amplifier is as large as the entire APD
array.

To prove the functionality of the integrating pixel, the transistor used as
the feedback resistance is clocked and becomes a reset switch. The
parasitic capacitance between input and output is used as the integrating
capacitance. Figure 8.1 shows the measurement of an integrator where the
top curve is the reset clock pulse, the bottom curve is the LED on pulse.
The middle curve is the output voltage, showing the integration of the
LED’s light power by the ramp. The parasitic capacitance shows a visible
non-linearity due to the capacitance value changing with operating point.
The output voltage change is approximately 180 mV. Scales on the figure
are 5 V/div for the top and bottom curves and 200 mV/div for the output
signal (middle curve). As can be seen from this figure, an integrator
actually exists. To improve the linearity of the response, an operating
point independent capacitance needs to be implemented using poly1/poly2.

Calculations show that the sense capacitance corresponds to approximately
350 fF. The light power used for these measurements was 43.7 pW. The
quantum efficiency at this wavelength is 25% and the corresponding input
is 35 x10^6 electrons/second. Thus, with a measured output of 46 mV, the
capacitance can be calculated. Using this capacitance, the dark current can
also be measured. The output corresponds to 40 mV in 160 ms. The dark
current is 87 fA at the input. Converting to square centimeters, the value is
approximately 5 nA/cm^2. Comparing this to the typical figure of 1 nA/cm^2,
this value seems reasonable in view of the higher doping in the n-well
material.
Most of these specific circuit improvements are easily implemented in a standard technology. Improving device characteristics is more difficult. In a future implementation, a reference diode with similar device characteristics may be necessary to compensate for changes. Such a diode could be placed once on chip and protected from light. The breakdown voltage of this reference diode would then be the reference for all other breakdown voltages because it shows any drift present on chip. In particular monitoring the breakdown voltage to stabilize the gain is important. Such circuitry is a simple current source and avalanche diode where the output voltage can be monitored. This could, at the same time, also compensate temperature drift. In reach through avalanche devices (operating voltages of ca. 100 V), a 1-3 volt bias increase is needed per degree C to keep the gain constant [Melchior77].

![Figure 8.1: Parasitic integrator on-chip.](image)

*Figure 8.1: Parasitic integrator on-chip. The output of the integrator is shown as the middle curve, and the non-linearity of the capacitance is visible with an output voltage change of 180 mV (200 mV/div). The top curve is the reset signal to the feedback MOSFET. The bottom curve corresponds to the LED “on” signal. Both these signals are 5 V/div.*

### 8.2 Influence of technology scaling on CMOS APD arrays

CMOS technologies are continually scaling in feature size such that 0.25μm technologies are commercially available from foundries in 1999. As these technologies scale, the important question is: are better APD imagers possible in these smaller technologies than in the 1.2 μm technology used to implement the APD imager described in this thesis?
To answer this question let us consider technology selection and divide this into three groups:

- for the design of photodiodes with stable operation in the avalanche regime
- for low noise multiplication and high quantum efficiency with APDs
- for standard readout circuitry

For stable APDs in a standard technology we require:

- a reverse biasable junction to separate electron hole pairs and support avalanche multiplication.
- a guard ring of lighter doped material to prevent premature lateral breakdown due to the large applied voltage for avalanche.
- a background doping level of the diode of less than $10^{18}$ cm$^{-3}$ to avoid regions where tunneling dominates over avalanche multiplication [Sze85]. Lower doping concentrations decrease the critical field necessary for breakdown, but increase the applied voltage necessary. Though higher applied voltages may not be desired, exploiting avalanche multiplication is more important. Normally doping levels increase with smaller technologies to increase the standard transistor transconductance desirable for circuit design. Therefore, this is a significant restriction for a standard submicron process.

For low noise multiplication and high quantum efficiency with APDs the necessary points are listed below.

- a p region where most of the electron hole pair generation due to light occurs. The electrons have a larger impact ionization rate (in silicon) than the holes allowing a lower noise multiplication to occur.
- a depletion region which extends to a depth of between 3 and 10 $\mu$m when the avalanche reverse bias voltage is applied. The larger depletion region depth increases the quantum efficiency for longer wavelengths, thus significantly decreasing the lost electron hole pairs due to recombination or drainage to a nearby pn junction. Should this extension not be possible, the quantum efficiency is reduced significantly for longer wavelengths.

For standard readout circuitry we need:

- a low voltage node (e.g. not more than 5 V). The complexity of the circuit design decreases if standard low voltage design can be used. Reliability also increases with low voltage design since these devices are standard.
• an individually adjustable bias voltage of the avalanche photodiodes to compensate for on-chip breakdown voltage variations.
• a second poly layer for voltage independent capacitances. For an integrating amplifier, a linear feedback capacitance around the amplifier is desirable.

Clearly, all of these points are required, yet not available in every technology. In CMOS, APDs are only possible in two cases: 1) n+ diffusion with n-well as guard rings in p substrate or 2) p+ diffusion with p-well as guard rings in n substrate. The first case requires a high voltage readout from the n+ node. Significant circuit complexity and reliability issues prevent such a readout. The second case requires the substrate to be at high voltage, thus causing the native PMOS devices to be unusable. Therefore, in CMOS technologies, no matter what feature size, APDs with low voltage readout are not possible.

In BiCMOS the situation is more promising. There are also two cases: 1) p+ with p-base as guard ring in n-well in p substrate and 2) n+ with n-base as guard ring in p-well in n substrate. Both cases allow substrate independent diodes with guard rings. However, both cases also have significant quantum efficiency reduction due to the depth of the wells. Case 1 collects holes but allows a low voltage readout. Low noise multiplication is not possible with this type of silicon APD because the hole ionization rate is much smaller than that of electrons. Case 2 on the other hand allows low noise multiplication but the readout is the high voltage node. The significant difference is that the p-well can be at a negative voltage, thus allowing the n+ node to be within standard CMOS voltages. The low noise multiplication and low voltage readout required are possible using this structure.

The technology required is a n substrate pnp BiCMOS process with a p-well concentration of less than $10^{18}$ cm$^{-3}$ to allow avalanche to occur. The depth of this well influences the quantum efficiency for longer wavelengths but low noise multiplication is possible due to the generation of electron hole pairs in the p material.

The feature size of the technology plays an important role in determining the pixel dimensions. For similar topologies as described in this thesis, involving a differential amplifier and source follower readout in each pixel, fill factors of 20% are possible in 2 μm technology. Decreasing the feature size to 0.25 μm, the circuit area ideally scales to 1/8 of the 2 μm
technology. Thus, at least 1100 \( \mu m^2 \) are required for circuitry in a 0.25\( \mu m \) process. With a fill factor of at least 30\%, the diode area becomes 472 \( \mu m^2 \) and the total pixel area is ca. 1600 \( \mu m^2 \). Allowing for process dependent spacing (e.g. between p-wells) the minimum pixel pitch is ca. 50 \( \mu m \). Therefore, in a 0.25\( \mu m \) technology, pixel area becomes 2500 \( \mu m^2 \) compared to 11000 \( \mu m^2 \) in a 2 \( \mu m \) technology. The reduction in pixel size is only a factor of 4 though the feature size reduction is 8.

The standard BiCMOS technologies of 0.25\( \mu m \) and smaller will require a lighter doped and deeper p-well to allow avalanche instead of tunneling to occur and for improved quantum efficiency. With a small process change, APDs are possible in significantly smaller technologies.

Now that we have determined that these APD imagers are possible in smaller technologies, what is the performance compared to today’s version? The increased leakage currents due to higher doping concentrations, increased interference due to the large number of inter-layer oxides covering the light sensitive areas and decreased supply voltage reduce image sensor performance in smaller (<0.5\( \mu m \)) technologies. Decreased dynamic range due to reduced output swing and supply voltage restrict image sensor performance in general. The quality of the APD performance also decreases as the doping increases due to increased tunneling and increased leakage currents. The electronic circuitry however profits from the decreased feature size in two significant qualities: speed and noise. The speed and bandwidth of circuitry improves due to the decrease in capacitance and the increase in transconductance. The white noise also improves due to the increase in transconductance. Thus, for high speed detection, submicron processes undoubtedly provide better performance due to the larger circuit bandwidths available. For applications where speed is of lesser importance, performance will remain the same.

8.3 The future of CMOS APD imaging

The thesis goal is to investigate avalanche photodiodes in standard technology (e.g. CMOS, BiCMOS) with the intention of creating more sensitive light detectors. Clearly achieving the most sensitive detection is only possible if APDs in standard technology are feasible and realistic. The approach used in this thesis is aggressive, requiring APDs to be implemented in standard technology without process changes. Thus, the most critical device in an APD imager can only be optimized in terms of
technology selection and geometry and layout within this technology. The optimal standard technology for APD design is a BiCMOS n-substrate with doping concentrations of the p-well of less than $10^{18}$ cm$^{-3}$. This standard technology also requires special surface passivation to minimize the boron-hydrogen doping concentration compensation that causes the breakdown voltage shifts in the implemented diodes. At the moment, this is the most disturbing feature of our standard APDs, hindering widespread use. The excess noise factor of the diode can be improved by using n+p diode rather than the implemented p+n diodes.

To control the APD, an amplifier is necessary. Such analog circuits within digital processes have long been standard designs but not if the area is constrained to a few thousand square microns! Designing low noise amplifiers requires current and area, both of which were severely restricted in the pixel design. In this case, compact layouts as well as low power design for each pixel are very important. The integrated array shows the feasibility of amplifiers in each pixel, however at the cost of an increased pixel size compared to APS pixels.

In terms of increased sensitivity, the goal of the most sensitive CMOS imager has not yet been achieved using APDs. The characterization of APDs in standard technology has shown that avalanche diodes are feasible, but excess noise figures remain large due to the small multiplying regions inherent to the doping profiles of the technology and the hole injection from the front illumination. The technology restriction prevents optimization of the multiplying device, causing both gain variation and noise to be the ultimate obstacles.

In short, for analog applications where the gain should be constant for all photons, the avalanche diodes in CMOS are not ideal due to the gain variation, excess noise and circuit complexity. For applications where constant gain is not necessary, such as black/white cameras, single photon counting or communications, such diodes may be usable, but with significantly different electronics. The integrating APD where the diode capacitance is used as an integrating capacitance and a threshold circuit detects the voltage level is more immune to gain variations.

The APD used is a non-linear device (current vs. voltage) whose operating point is stabilized to allow a linearization around the operating point. This linearization and stabilization are extremely dependent on the device itself and the controlling circuitry. Any instability of the device such as drift or
temperature dependence inherently complicates keeping the device at the same operating point. Calibration and stabilization, however, allow operating point shifts to be compensated by using a reference diode with similar drift properties. Thus, APD applications where the exact value of the gain is not important (e.g. photon counting) are more suited than analog applications where a linear response is expected. Thus independent of technology, APD devices integrated in standard technology are more appropriate for digital applications with yes/no responses. Future research should concentrate on these applications.

This thesis shows the first monolithic APD array in standard BiCMOS technology, which has been used to present low light level images. Though these images do not yet achieve the sensitivity calculated from theory, they do prove the feasibility of avalanche gain in standard technology. The sensitivity shown is, at best, equivalent to early APS sensors. The current APD sensor’s performance is limited by noise contributions from two significant areas: diode excess noise and feedback resistance noise from the transimpedance amplifier. Though the feedback resistance noise contribution is significant, the increased noise with increasing gain proves to be the main noise contribution even at small gains of 10. To decrease this noise contribution one solution is available: changing technologies. A technology change involves using n-substrate BiCMOS technology, preferably a process specifically for photodiodes where the passivation layer is specially developed to minimize the hydrogen content. Ultimately, the APD gain is increased until the noise contribution due to excess noise is equal to the noise contribution due to the amplifier (including reset noise), thus allowing the optimal system noise to be achieved. Making use of this option, not readily available in the scope of this thesis will open the way to APD arrays in improved but otherwise standard BiCMOS technology that are superior in their low light level performance to state-of-the-art APS/CMOS image sensors.
A. References:


[IMS Chips] IMS Chips, HDRC2-EC Datasheet, IMS Chips, Allmandring 30A, D-70569 Stuttgart, Germany


1977, pp.238ff.


Acknowledgments

First I would like to thank my thesis advisors for their support and critical comments: Prof. Dr. Peter Seitz, Image Sensing Group Leader at CSEM in Zürich and Prof. Dr. Heinz Jäckel at the Institute for Electronics at the Swiss Federal Institute of Technology (ETH) in Zürich.

Next, I’d like to thank my officemate Kai Engelhardt for keeping me sane during these past 4 years and for the friendly competition at badminton on Monday nights.

Let me also thank my friends and colleagues at CSEM for the psychological support and the laughter that always made me feel better: Philipp Zeller, Robert Lange, Stephane Bourquin, Vinz Monterosso and Max Wiki. Thanks to Peter Metzler for all the advice on analog circuits, measurements and life in general. Thanks also to the rest of the image sensing group: Graham Lang, Nicolas Blanc, Roger Dangel, Stefan Lauxtermann, Peter Schwider, Martin Wäny, Georg Israel, Gerben Boer and Michel Willemin. Special thanks also to Helmut Teichmann and Susanne Westernhöfer.

Thanks also go to Beat Furrer, for introducing me to IC design and for the welcomed advice on all my projects. For getting me started on the physics of avalanche diodes, I’d like to thank Lukas Snyman. I also don’t want to forget those that have left us (for bigger and better things) such as: Oliver Vietze, Jeff Raynor and Thomas Spirig.

Last but not least I want to thank Charlotte Biber and Thomas Morf for their patience when I lost mine.
Curriculum vitae

I was born on Sept. 11, 1967 in Boston, Massachusetts, USA. From 1973 to 1985, I attended public schools in Needham, Massachusetts, earning a high school diploma in 1985.

From 1985 to 1990, I attended the Massachusetts Institute of Technology (MIT) in Cambridge, MA. I received a Bachelor of Science (BSEE) in 1989 and a Master of Science (MSEE) in 1990, both in electrical engineering. After graduation, I moved to Zürich Switzerland, where I worked as an electrical engineer at Landis and Gyr in Zug, Switzerland from 1990 to 1993. From July 1993 to November 1995, I was employed by the Integrated Systems Laboratory at the Swiss Federal Institute of Technology.

In December of 1995, I joined the image sensing group of the Paul Scherrer Institute in Zürich. In September 1997, the same group was transferred to Centre Suisse d'Electronique et de Microtechnique in Neuchâtel. The projects I worked on during these past 4 years pertain to image sensing, analog IC design and measurements of optical image sensors. During this time, the research for this dissertation was performed.