DISS. ETH NR.

An AER Temporal Contrast Vision Sensor

ABHANDLUNG
zur Erlangung des Titels

DOKTOR DER TECHNISCHEN WISSENSCHAFTEN

der

EIDGENÖSSISCHEN TECHNISCHEN HOCHSCHULE ZÜRICH

vorgelegt von

PATRICK LICHTSTEINER

Dipl. Phys. ETH

geboren am 1. März 1976

auf Antrag von

Rodney Douglas
Tobias Delbrück

2006

Outline

Dear reader, today you have probably been filmed by several electronic cameras. You were in the visual field of a video camera while you stood close to a cash-machine, while you walked through a railway station or an airport, anytime you entered a bank or large public building, during the shopping in the super-market and also on many public spaces.

Electronic eyes became very abundant in our environment in the last few years. But surveillance is just one of many application fields for electronic vision devices. Artificial vision is also used in industrial manufacture, in safety systems in industrial environments, for visual quality control and failure investigation, for visual stock control, for barcode reading, to control automated guided vehicles etc. In these applications, the human workmanship is replaced by an electronic camera paired with sophisticated computer vision software running on the computer to which the camera is attached. The complete system, comprising one or more cameras, computers and sometimes also actuators (i. e. robots), is called a machine vision system.

Although machine vision systems are widely and very successfully used, they are limited to perform narrowly defined tasks under highly controlled and repetitive circumstances. No machine vision system comes even close to the capabilities of human vision in terms of image comprehension, tolerance to lighting variations, image degradation, scene variability etc. Machine vision systems do not *see* in the same way that human beings are able to and the principle differences between machine vision systems and biological vision systems start right at the front end electronic cameras that are not equivalent to human eyes.

Almost all machine vision systems rely on standard electronic imagers that produce images, so called frames, in a regular frequency. Frame-based architectures carry hidden costs because they are based on a stroboscopic series of snapshots taken at a constant rate. All pixels are sampled repetitively even if they have nothing novel to say. The pixel bandwidth is limited by the identical global sampling rate, the frame rate. The dynamic range of frame based cameras is typically limited by the finite pixel capacity for integrated photocharge as well as by the identical pixel gain and integration time. In unsupervised environments with natural lighting, the disadvantages of limited dynamic range and bandwidth can become the limiting factor of the whole machine vision system.

Biological visual systems have no notion of a frame. The retinal photorecepotrs work in continuous time lacking any global reset or readout signal. They are embedded in a neuronal network that does significant focal plane computation such as gain control and redundancy reduction. The retinal outputs are massively parallel and data-driven. In vertebrates, the decisions on when to quantize are made by the ganglion cells that project to the brain. A quantiza-

tion is immediately communicated from the retina to the brain by the axon of the ganglion cell in the form of a digital pulse (this principle is often called *integrate and fire*). The information about the spatial origin of this pulse is given by its unique communication channel, and information about the intensity or intensity differences is coded in the relative timing of a sequence of such pulses. Eyes are optimized for the extraction of relevant information from a visual scene and in this respect outperform any existing camera by far.

These considerations underlie the belief that emulating this asynchronous data driven biological architecture will lead to capturing the power of biological vision into electronic devices. The continuing increase in integration density in silicon integrated circuits means that it is becoming increasingly feasible to develop vision sensor devices with complex pixels that still have acceptable pixel size and fill factor.

In this thesis we will describe the development of a vision sensor that imitates some specific retinal functionality; the sensitivity to local temporal changes in illumination. The vision sensor that responds purely to relative changes in intensity, loosely models the transient pathway in the retina by emulating three key principles of biological vision: its event-based output, its representation of relative luminance change, and its representation of sign into separate positive and negative channels. Our pixel combines an active continuous-time logarithmic photosensor with a well-matched self-timed switched-capacitor amplifier. Each pixel continuously monitors its photocurrent for changes and responds with an ON or OFF event that represents a fractional increase or decrease in intensity that exceeds a tunable threshold. Events are immediately communicated asynchronously off chip on a self-timed bus using *Address Event Representation* (AER).

The latest vision sensor that we designed (chapter 6) contains a 128×128 array of pixels that encode scene reflectance change by outputting asynchronous address-events. These events are triggered by local relative intensity changes that exceed a global threshold. These events typically signify object movement and have sub-millisecond timing precision. An individual pixel bandwidth of above $3kHz$ under $1klux$ scene illumination and an array mismatch of 2.1% in relative intensity is achieved. The chip is built in a $0.35\mu m$ 4M 2P process yielding 40×40 μm^2 pixels with 9.4% fill-factor. The measured dynamic range is $120dB$ and chip power consumption is $23mW$. The response latency varies slowly with illumination and is $15\mu s$ at $> 1klux$ pixel illumination. This combination of characteristics provides unique capabilities for low-latency dynamic vision under wide illumination range with low computational requirements. The sensor has already been used successfully for different experimental applications.

The thesis is structured into 7 chapters. Chapter 1 is an introduction that encompasses a short descriptions of electronic cameras and recent development trends in imager design, relates some common knowledge about eyes, familiarizes the reader with neuromorphic engineering and AER communication and finishes in a discussion of the state of the art of comparable sensors and the justification for this work. The description of stepwise advances in the development of our sensor covers two different pixel designs. The first of this designs existed before we started the this work and was done by Jörg Kramer. In chapter 2 we will describe Jörg Kramer's pixel design, his transient imager and our own implementation of his design. The conclusions we draw after the characterization results of our first transient imager acquaints the reader with the rationale to

switch to a new pixel design. This novel pixel design is described in chapter 3. The description of this pixel design and a discussion of its three different implementations with thorough characterization results are given in chapters 4 to 6. The thesis is ended with a conclusion in chapter 7.

The work and achievements presented in this thesis stem from work done by myself in partnership with my supervisor Tobi Delbrück. We had some help from other people, specifically Christoph Posch who joined our design, implementation and characterization efforts somewhere between chapter 4 and 6. The rest of the people who contributed to this work will be named in the acknowledgments or directly in the text at the appropriate point. Please note, that I use "we" throughout the thesis instead of specifically naming the persons meant. This we includes Tobi Delbrück and myself and in chapter 4 to 6 also Christoph Posch.

Zusammenfassung

das ist die zusammenfassung...

Contents

Chapter 1

Introduction

1.1 Money, Frames and Biology

Money and MegaPixels

Sales of point and shoot electronic still cameras amounted to over 50M units in 2004 and are expected to reach 80M units in 2008. The market leaders in digital point and shoot cameras are Canon, Sony, Olympus, Fujifilm and Kodak with a combined market share of about 75%.

Marketing strategies from this industry led to the emergence of one dominant figure of merit of electronic cameras; the number of megapixels (MP). The megapixel rush has cumulated in standard electronic cameras that make images that have a far better resolution than most people can appreciate viewing their photos on screens and beamers with resolutions of under 2 MP. The number of megapixels recently has become the center of attention in the market of camera cell phones. Since Summer 2006, Sony Ericson, one of the leaders in the market of camera cell phones, is selling in Europe a mobile phone called K800i with a digital 3.2 MP camera with autofocus, built-in xenon flash and image and video stabilizer function.

The huge market potential has driven a fast development in electronic imager design aimed at high quality still imaging suitable for photography during the past decade. The emergence of the cell phones has shifted the attention to the development of cheaper imagers, with very small pixel sizes, suitable for

1

use with small and cheap lenses. At the 2005 *IEEE Workshop on Chanrge-Coupled Devices and Advanced Image Sensors* in Japan, Matsushita, Kodak and Sony presented imager pixels with sizes below the diffraction limit of standard lenses [1–3].

We will give a short review of the development of electronic imagers and latest trends in electronic imager design in section 1.2.

Machine Vision With Images

Computer vision, image processing, image analysis, robot vision and machine vision are closely related fields all trying to solve vision problems based on image data. Automated vision applications, mostly belonging to the field of machine vision, profited a lot from the fast development in electronic imagers and the ever increasing computing power of processors. It is today standard to use electronic imagers in safety systems in industrial environments, inspection of pre-manufactured objects and food products (e.g. quality control, failure investigation), visual stock control and management systems (counting, bar-code reading, store interfaces for dtigital systems), control of automated guided vehicles (AGVs) and automated surveillance.

Products for automated vision applications already constitute a major market with huge potential. The market potential for digital cameras in surveillance alone is estimated to be several billions in a few years profiting from rising security awareness. Electronic imager sensors for machine vision have to meet other specifications than imagers for consumer cameras. Wide dynamic range, high speed capability and high sensitivity are just a few of the features demanded for specific machine vision applications.

The Cost of Basing Vision on Frames

All imagers and almost all automated vision products are based on the principle of *frames*. A frame is an 2D matrix that represents the intensities and colors of the visual scene under observation at a certain time. Video data consists of a stream of frames taken at a regular frequency. The notion of a frame has become so embedded in machine vision that it is usually taken for granted. This is natural given that frame-based devices have been dominant from the days of drum scanners and videon tubes to today's CCDs and CMOS imagers. The principle of a frame is also kind of the logical follower of a painting and a film photograph, it is the representation of a snapshot, totally laking dynamics. Time-domain computation is done by operating on sequences of snapshots.

The undeniable advantage of frame-based imagers is that they have small pixels, leading to high resolution, large fill factor, and low imager cost. The frame as an output format is well understood and is the basis for many years of research in machine vision. Finally, frames are intrinsically compatible with frame-based outputs like video displays.

Frame-based architectures carry hidden costs for vision applications. Frame based vision is centered on a stroboscopic series of snapshots taken at a constant sample rate. Every pixel is sampled repetitively even if its output is always the same. Global and single pixel bandwidths are equal and both are limited by the sampling rate. The dynamic range of frame-based imagers is typically limited by the identical pixel gain, the identical integration time, and the finite

pixel capacity for integrated photo-charge. Machine vision deals with these disadvantages by sophisticated and costly high-performance artificial lighting[1]. For unattended environments with natural lighting, limited dynamic range and bandwidth can be a major disadvantage.

Biology Works Different

Biological vision by far outperforms any of the state of the art machine vision except in very narrowly specified tasks in controlled environments. We heavily rely on our vision system when reading a book with candle light, driving a car in the night, playing soccer under bright sunlight, searching for shells on a beach at sunset or finding the tent at the campsite in the darkness of a moonless night. Biological vision evolved to efficiently perform real time tasks and biological vision systems are quite differently structured than machine vision systems.

Biological visual systems have no notion of a frame, except perhaps in the sense of the data that arrives immediately after the end of a rapid eye movement. The retinal outputs are massively parallel and data-driven. In vertebrates, the decisions on when to quantize are made by the ganglion cells that project to the brain and that send their information in the form of digital events which occur in continuous time. Preprocessing computations such as gain control and redundancy reduction precede this quantization.

Because of the importance of biological vision systems as a justification for our work, we will shortly review some aspects of the human vision system in section 1.3.

Neuromorphic Engineering

In the work we present here, we try to incorporate biological principles on CMOS substrate with the hope to create a competitive alternative to the conventional imagers for machine vision. Our approach belongs to the academic field of neuromorphic engineering. This research discipline, born around 1986 in Carver Mead's lab at Caltech [4], tries to emulate neural architectures with the hope to capture the power of biological neural systems into electronic devices [5]. The chosen substrate for the functional implementation of neural structures is silicon and the method is *analog very large scale integration* (aVLSI), which is the standard method for analog integrated circuit (IC) design.

In section 1.4 we will come back to neuromorphic engineering and how this work contributes to the advance of this field.

1.2 Electronic Photo Sensors

1.2.1 Phototransduction

In film photography, phototransduction is the process that uses the energy of a photon to elicit a chemical reaction on photographic plates, in which a silver halide molecule is split into an atom of metallic silver and a halogen atom. In electronic photo sensors however, phototransduction relies on the effect that the absorption of a photon can happen by elevating an electron from lower

---

[1]It is said that creating a good scene illumination is one of the most important aspects in machine vision

energetic state into higher energetic state. In semiconductors the transition of the electron induced by the absorbed photon is from the valence band the conduction band. An electron elevated into the conduction band leaves behind a hole in the valence band, therefore phototransduction in semiconductors leads to the generation of electron hole pairs.

The material properties of the semiconductor used for phototransduction is critical, because only photons with sufficient energy to excite an electron across the material's bandgap from the valence band to the conduction band will lead to a significant electron-hole pair generation. Silicon has a bandgap that corresponds to the energy of a photon with a wavelength of $1100nm$ and therefore is perfectly suitable for phototransduction in the visible light.

Every photon with energy exceeding the bandgap of the used semiconductor, can generate at most one electron hole pair. The rest of the energy of the photon is absorbed in the lattice structure of the semiconductor. Higher energy photons have a higher probability of being absorbed leading to the fact that longer wavelength photons penetrate the semiconductor deeper. The electron-hole pairs can be separated in a semiconductor by applying an external electrical field. The most simple structure that is built this way is the photo-conductor, which consists of a slab of semiconductor clamped to an external potential that creates an electric field across the semiconductor. The conductance of this photo-conductor depends on the photon-flux hitting the exposed semiconductor. The photo-conductor has a poor signal to noise ratio because of the relatively large current flowing through the semiconductor even in absence of any light.

Three more elaborate different semiconductor structures emerged, that used the generation of electron-hole pairs to measure incident light; photodidodes that are most prominently used in CMOS, photogates, that are used for CCDs and phototransistors that have a larger gain than photodiodes but a relatively poor signal to noise ratio.
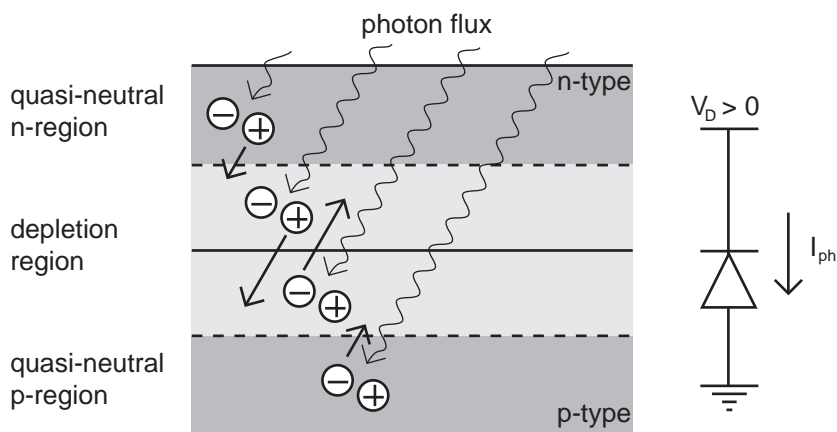

Figure 1.1: **The rate of the generation of electron-hole pairs is used in photodiodes to measure the incident photon flux by separating the electrons form the holes before they recombine. In the photosensing mode, the photodiode is reverse biased, meaning that the n-type side of the photodiode is held at a higher potential than the p-type side.**

A photodiode, as shown in Figure 1.1, comprises a p-n junction built by creating a region in the semiconductor that is doped with donors (n-type) adjacent to a region that is doped with acceptors (p-type). The spatial arrangement of a p-n junction leads to a region in the center that is devoid of mobile charges, called the depletion region. The remaining fixed charges in the depletion region create an electric field pointing from the n-type region to the p-type region. This built in electric field separates the electron-hole pairs leading to a net photo-current flow through the photodiode that is proportional to the incident photon-flux. The probability that a photon reaching the optical sensor generates an electron-hole pair that then contributes to the photocurrent is called the quantum-efficiency. The quantum-efficiency varies with the wavelength.

Photogates are so called *Metal Insulator Semiconductor* (MIS) structures that use a metal gate to produce an electric field in the semiconductor. Electrons that are elevated into the conduction band of the semiconductor are sucked toward the positively charged metal gate. After a certain period of time, called the integration time, the electrons accumulated under the metal gate are read out. The readout of an array of photogates is achieved by cleverly clocking the the photogates to shift the charge from one photogate to the next till it reaches the border of the array. The amount of the electrons in such a charge package is proportional to the average photon-flux during the integration time. Nowadays photogate structures are very optimized; gates are not made of metal but of thin polysilicon, that absorbs and reflects less photons, and the charge is not collected directly underneath the gate but in a so called buried channel to avoid the disturbing effects of lattice defects that are most abundant near the surface of the semiconductor. A concise description of photodiodes, and photogates can be found in the book *Analog VLSI:Circuits and Principles* [6].

Nowadays, photodiodes and photogates belong to the most abundant structures in our world. There are millions of photodiodes in every cell phone that comprises a camera and in every electronic camera.

1.2.2 Charge-Coupled Device

A charge-coupled device (CCD) is an image sensor, consisting of an integrated circuit containing an array of linked, or coupled, capacitors sensitive to light, so called photogates. Under the control of an external circuit, each capacitor can transfer its electric charge to one of its neighbors. CCDs are used in digital photography and astronomy.

The CCD was invented in 1969 by Willard Boyle and George Smith at AT&T Bell Labs (one of many influential inventions of AT&T Bell Labs [7]). The CCD was born in 1970, when Bell researchers were able to capture images with simple linear devices. Several companies, including Fairchild Semiconductor, RCA and Texas Instruments, picked up on the invention leading to the first commercial devices by Fairchild in 1974 (a linear 500 element device and a 2-D 100 x 100 pixel device).

CCD Architecture

CCD image sensors can be implemented in several different architectures. The most common are full-frame, frame-transfer and interline. The distinguishing

characteristic of each of these architectures is their approach to the problem of shuttering.

In a full-frame device, all of the image area is active and there is no electronic shutter. A mechanical shutter must be added to this type of sensor or the image will smear as the device is clocked or read out.

With a frame transfer CCD, half of the silicon area is covered by an opaque mask (typically aluminum). The image can be quickly transferred from the image area to the opaque area or storage region with acceptable smear of a few percent. That image can then be read out slowly from the storage region while a new image is integrating or exposing in the active area. The frame-transfer architecture requires twice the silicon area of an equivalent full-frame device and therefore costs roughly twice as much.

In the interline architecture the frame transfer concept is taken one step further and every other column of the image sensor is masked for storage. In this device, only one pixel shift has to occur to transfer from image area to storage area enabling transfer times in the order of microseconds and essentially eliminating smear. The disadvantage of this approach is that the imaging area is now covered by opaque strips resulting in halve the fill factor and therefore also halve the effective quantum efficiency. By adding microlenses on the surface of the device to direct light away from the opaque regions and on the active area the fill factor is can be increased back up to 90%.

The choice of architecture depends on the application. If an expensive, power hungry and failure prone mechanical shutter can not be tolerated an interline device is the right choice. Consumer CCD snap-shot cameras use interline architectures. The full frame architecture is often the choice in astronomy because it grants the best possible light collection of all three architectures. The frame-transfer, outdated after microlenses were introduced, is today used, where interline architecture is not available, for example in back-illuminated devices.

1.2.3   The Active Pixel Sensor

The active pixel sensor is descended from the original MOS image sensors invented in the late 1960s. The original MOS sensors were passive pixel sensors in which each pixel contained a photodiode and an access transistor. Pixels were implemented in two dimensional arrays, with circuitry shared by pixels in the same row, and other circuitry by the same column. At the end of each column was an integrating amplifier to read out a pixel in a column select by the row circuits. Passive pixel sensors suffered from severe problems, such as slow readout and high noise. The addition of an amplifier to each pixel addressed the problems of the active pixel sensor, and resulted in the creation of the active pixel sensor.

The standard CMOS APS pixel consists of three transistors as well as a photodetector, usually a photodiode. Light causes an accumulation, or integration of charge on the capacitance of the photodiode[2] that is dependent on the power of incident light. The transistor, M_{rst}, acts as a switch that resets the device. When this switch is closed, the photodiode is effectively connected to the power supply, V_{RST}, clearing all integrated charge. The second transistor, M_{sf}, acts as a source follower buffer, that allows the pixel voltage to be observed without

---

[2]The photodiode capacitance is often called the parasitic capacitance of the photodiode.
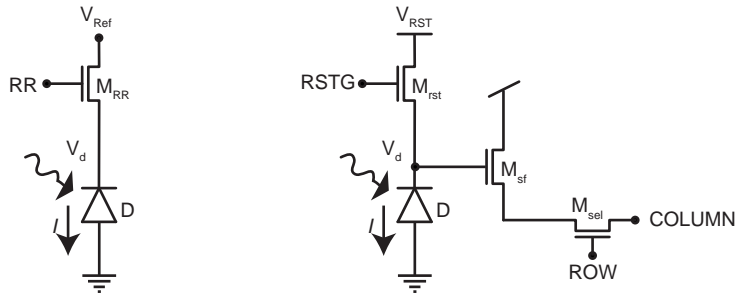
Figure 1.2: **The passive pixel sensor circuit comprising only a photodiode and a single transistor is depicted on the right side of the figure. A three transistor APS pixel circuit is shown on the left side.**

removing the accumulated charge. Typically VDD and V_{RST} are equal. The third transistor, M_{sel}, is a switch that allows a single row of the pixel array to be read by the read-out electronics present in for each column. Typically further amplifier circuitry is present for each column.

1.2.4 CCD or APS?

In 1993, Eric Fossum, published an extensive article [8] predicting the APS sensors as the commercial successors of CCDs. In the mean time it has become practical to create competitive Active Pixel Sensor (APS). Because these sensors are integrated in CMOS technology, which is the dominant technology for all chip-making, active pixel sensors are cheap to make. The reason that makes CMOS sensors very attractive is not just the price but the ability to incorporate signal conditioning circuitry and even image processing circuitry into the same device. CMOS sensors also have the advantage of lower power consumption and lower image lag than CCDs. CCDs still boast higher sensitivity, and higher dynamic range than CMOS sensors, and for these reasons CCDs are preferred in astronomical imaging where these factors are of prime importance.

Today, a number of semiconductor manufacturers offer APS sensors of various types. These include Micron Technology, Inc., Toshiba, Inc., Omnivision Technology, Inc. and Canon, among others.

1.2.5 Color Imagers

Because standard photodiodes are sensitive to the whole visible light spectrum down to the near infrared, color sensitivity is achieved by either applying color filters or using a prism.

The Bayer Pattern

Most single-chip digital image sensors used in digital cameras, camcorders, and scanners are overlaid with a Bayer filter mosaic[3]. In a Bayer pattern, each square of four pixels has one pixel with a red filter, one pixel with a blue filter, and

---

[3]The Bayer filter mosaic is named after its inventor B. E. Bayer who was with Eastman Kodak at the time of the invention.

two pixels with green filters[4]. The result of this is that luminance information is collected at every pixel, but the color resolution is lower than the luminance resolution.

The raw output of Bayer-filter cameras is called a Bayer pattern image. Since each pixel is filtered to receive only one of three colors, two-thirds of the color data is missing in each pixel. To obtain a full-color image, various demosaicing algorithms can be applied to a Bayer pattern image to interpolate a set of complete red, green, and blue values for each pixel. This can be done in-camera, or outside the camera using the raw data directly from the sensor.

3CCD

A better color separation than with a Bayer pattern can be reached by three-CCD devices (3CCD) and a dichroic beam splitter prism, that splits the image into red, green and blue components. The three CCDs are arranged in such a way that each responds to a particular color. All professional and some semi-professional digital video camcorders use this technique.

For photography a cheaper solution than using three CCD chips, which are still expensive in high resolution, is to use a rotating color filter. The advantage of this solution is, that it grants at the same time color-fidelity and high-resolution, the drawback is, that these multi-shot cameras can only be used for scenes without movement.

The Foveon X3 Sensor

In contrast to the mosaic Bayer filter sensor design commonly used in digital camera sensors, the Foveon X3 sensor, produced by Foveon, Inc., is a layered sensor design, in which each location in a grid has layered photosensors sensitive to all three primary colors. This layered photosensors consist of three photodiodes in three different depths, using the fact that the absorption coefficients for photons of different wavelengths differ (as discussed in section 1.2.1).

Because demosaicing of the output of the Foveon X3 sensor is not required to produce a full-color image, the color artifacts associated with that process are not seen. Another advantage of the Foveon sensor is that it does in contrast to a mosaic senor not filter away photons. Therefore, about 3 times more of the photons entering the camera will be detected by the Foveon sensor. Finally, because the Foveon sensor uses the intrinsic properties of silicon for color sensitivity, the need of adding a color filters onto pixels is erased, lowering the production costs of the photosensor.

The question of pixel count has been a point of controversy for the Foveon X3 sensor: For example, the Sigma SD10 digital SLR camera, which produces a native RAW file size of 3.4 Million RGB pixels, is advertised as a 10.2 MP camera. Sometimes the clarification 3.4 MP Red + 3.4 MP Green + 3.4 MP Blue is added. An 8 MP Bayer-mosaic camera would similarly have to be clarified as 2 MP Red + 4 MP Green + 2 MP Blue.

---

[4]It is not possible to distribute 3 colors evenly and with equidistant spacing in a quadratic grid. Green was chosen as the color to be used twice because the human eye is more sensitive to green than either red or blue.

The Foveon X3 sensor is used in the Sigma SD9, SD10 and in the SD14[5] digital SLR cameras and the Polaroid x530 compact digital camera[6].

1.2.6 Trends in Image Sensor Design

The Magic of Miniaturization

Up to now, the most prominent research direction remains the miniaturization of pixels. The main reason for this effort is certainly to make sensors and cameras cheaper. The optics are a dominant part of the cost of a camera and smaller formats of sensors allow to use cheaper optics. Another reason for the miniaturization of the image sensors is, that in consequence also cameras can be miniaturized. This property is most beneficial for cheap cell phone cameras. Finally, miniaturization is trendy in the in the microcircuit industry and has made microcircuits as powerful as they are nowadays.

The shrinking pixel sizes carry along interesting problems. Pixel sizes are already below the diffraction limit of standard lenses [1–3] and the resolution of image sensors has to be redefined. Photodiodes are surrounded with walls consisting of metal interconnect layers that are becoming much higher than the photodiode is wide creating long optical paths and critically reducing the angle of response of a pixel. The charge storing capacity of the small photodiodes is shrinking leading to worse dynamic range and poorer signal to noise ratios.

Sub-Diffraction-Limit Pixel

Sub-diffraction-limit (SDL) pixels[7] could for example be average such that a block of 4 pixels would result in one output pixel. This average would increase signal to noise ratio and in a imager overlaid with a Bayer filter mosaic, color reconstruction would be simplified and color aliasing effects would be reduced. E. R. Fossum proposed to use SDL pixels for a gigapixel digital film sensor [9].

Interconnect Layer Stack Hight, Copper and Back-Illumination

The industry's conventional approach to deal with the problem of the thickness of the metal interconnect layers, is just to try to decrease it. Using optimized copper instead of the conventional aluminum for the metal interconnects allows to use significantly thinner wires that have the same conductance as the relatively thick aluminum wires. A $2.7\mu m$ imager using optimized copper interconnect was presented in 2005 by IBM and Eastmyn Kodak [1] and they achieved a reduction of the thickness of the four layers over the pixel from $3.6\mu m$ for aluminum to $2.1\mu m$ for optimized copper. A schematic drawing of this concept is shown in Figure 1.3a) and b).

An interesting solution to the layer stack problem is to turn the sensor chip upside down. Figure 1.3c) shows a schematic drawing of a back-illuminated image sensor. The back-illumination principle has a number of advantages for

---

[5]http://www.sigma-photo.com

[6]http://www.polaroid.com

[7]Sub-diffraction-limit pixels are pixel with sizes below the Airy disk diamater DA, which is given by $D_A = 2.44\lambda F\#$, where λ is the wavelength of interest (for imagers $\lambda = 550nm$) and $F\#$ is the F-number of the optical Sastem.
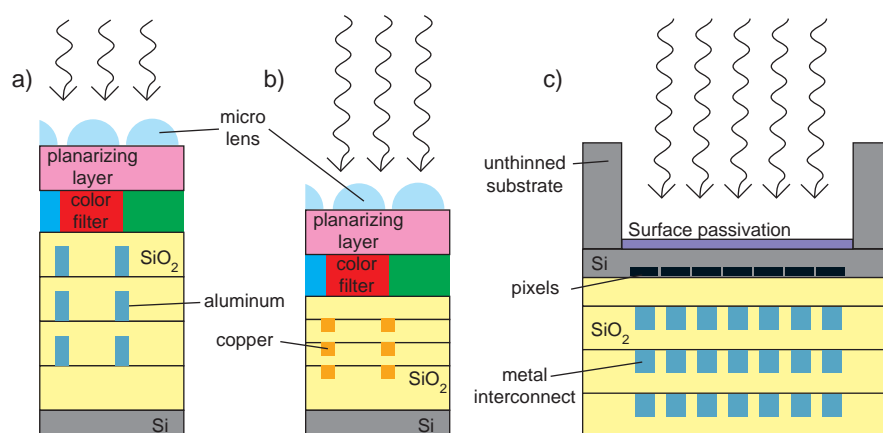
Figure 1.3: **Shows in a) a conventional layer stack for a CMOS imager with aluminum interconnect and in b) a significantly thinner layer stack using optimized copper. Part c) shows the concept of back-illumination.**

small pixel sensors. The advantages include very high (up to 100%) fill factor [10], improved angular response, availability of the frontside of the pixel to incorporate processing circuits in the pixel [11] and last but not least improved quantum efficiency. The problem is, that these wonderful properties are buried under a relativly thick substrate and must be carefully uncovered by complicated backside thinning. Due to the complicated and failure prone process of backside thinning, these sensors have not made it into mass market yet.

High-Speed and HDTV

High-speed imaging[8], is often needed for machine vision tasks in industrial quality control and for scientific purposes. The challenges of short integration time and high readout data loads are solved by using larger and fewer pixels.

A common architecture for "very high-speed" devices with a high resolution is the single line architecture, which allows to make long and narrow pixels with small pixel pitch for high resolution, high fill factor and good quantum efficiency and on the other hand it also enables the placement amplifiers or even *Analog to Digital Converters* ADC directly beside the pixel reducing access delays and noise. Recently Dalsa Corp. reported on linescan CCD sensor with $12k$ pixels and a line scan speed of $23.7kHz$ [12].

Using an ADC in each pixel, Kleinfelder et al. took an approach used in linescan sensors to a 2D sensor and realized a 352×288 APS sensor that reaches a frame-rate of $10kHz$ [13]. The price of placing an ADC in each pixel is larger pixel size and lower fill factor resulting in a larger and more expensive chip and also lower sensitivity. Furthermore the analog to digital conversion time increases exponentially with bit-resolution due to limited room for digitizing circuitry in the pixel, resulting in a relatively small dynamic range.

---

[8]High-speed imaging is not a specifically defined term but refers usually to frame-rates surpassing the $30Hz$ used in standard video cameras; a $60Hz$ frame-rate device is often considered to be high-speed.

With the establishment HDTV[9] and increasing HDTV broadcast, the demand for high-speed high resolution sensors has greatly increased. Shortcomings of earlier designs are overcome by sophisticated voltage mode column buffers that amplify and buffer the column readout voltage of an APS sensor onto a very fast ADC [14] or by specialized *High Definition* CCD (HDCCD) designs that enable fast charge transfer [15].

Range Imaging

Fast 3D acquisition is desirable for a number of machine vision applications, such as obstacle detection, visual surveillance and virtual reality. Conventional range finding approaches are based upon either scanning-laser devices or systems of several conventional cameras. Both approaches are not ideal for real time 3D sensing. The state of the art systems in real time 3D imaging nowadays mostly apply the *Time-of-Flight* (TOF) principle. TOF imager systems always include a light source that emits intensity modulated light, mostly monochromatic near infrared lasers. There are two methods of distance measurements, the direct measurement of the time of flight and the indirect measure of the time of flight, by comparing the reflected light signal with a reference signal. The latter method is more prominent and makes use of either pulses of emitted and reflected light that can be multiplied on the receiver with a delayed signal with nanosecond precision [16] or of sinusoidally modulated light that can be compared to a reference sinus for phase shift extraction [17].

Although some of the range imaging devices based on the TOF principle are quite successful in measuring pixel based ranges, they all share a common drawback, the need of a powerful monochromatic light source. This drawback is a commercial disadvantage in environments with "well-behaved" background lighting conditions, where the modulated source is dominant within the spectral band that passes the filter applied to the sensor. The active illumination becomes a major obstacle in outdoor lighting conditions, because of the competition with the sun, which is a very powerful light source over a wide range of wavelengths, including the near infrared.

Stereo vision, as done by biological vision systems, can work in a passive mode and does not share the problems of TOF sensors, but stereo vision usually involves a lot of processing and so far there exist no artificial real-time stereo vision systems that are competitive.

1.3 Biological Vision

From an engineering point of view, biological vision sensors, we will call them eyes, have astounding capabilities. Human eyes have photoreceptors that are able to react to a single photon catch and on the other end of the scale still perform perfectly well in bright sunlight conditions. The range of usable illumination up to the saturation of the photoreceptors is from a an overcast night sky emitting about $10^{-6}cd/m^2$ up to snow in sunlight emitting between $10^4cd/m^2$ and $10^5cd/m^2$ [18]. This range spans over 10 decades, a huge dynamic range

---

[9]The term *High Definition Television* (HDTV) is used for a collection of new television standards that surpass the *Standard Television* (SDTV) standards (PAL and NTSC) in spatial and some of them also in temporal resolution.

of $200dB$ compared to the dynamic range of a standard electronic imager that is below $70dB$. The human eye has several 10 million photoreceptors and the density of photoreceptors in the center of the visual field is as high as 150000 photoreceptors per square-millimeter, this means that the photoreceptor pitch goes down to about $2.6\mu m$, comparable to the smallest pixel pitch in the latest imagers (see section 1.2.6).

Although eyes have many other magnificent metrics that can be easily compared to electronic imagers, like power efficiency and high angular accuracy, the even more astounding feature of eyes is that they convey preprocessed information to the brain in such an efficient and accurate way, that we can rely on our vision for very demanding real time tasks, that are beyond any of nowadays machine vision systems. Eyes are optimized for efficient information extraction from the visual scenes and they are abviously not optimized for producing images for other eyes to look at, like electronic cameras.

1.3.1   The Human Eye

The eyes of the vertebrates are based upon a common structural plan. Figure 1.4 shows a schematic drawing of the cross-section through the human eye. The light enters the eye from an external medium such as air or water, passes through the cornea, and into the first of two humours, the aqueous humour. The aqueous humor is a clear mass that connects the cornea with the lens. It helps to maintain the shape of the cornea and provides nutrients for the corneal endothelium. Between the aqueous humor and the lens lies a colored ring of muscle fibers called the iris. The iris is blocking light, forming an adjustable hole in the center where light can enter the eye through the lens. This hole, called the pupil, is adjusted actively by the iris muscles to maintain a relatively constant level of light entering the the eye. The light passing through the pupil is focused by the cornea and the lens. While the cornea has a fixed curvature, the lens can be stretched to adjust the focus to distance of the object of the eye is looking at. The lens is hold in place by the suspensory ligaments. On the back side of the lens is the vitreous humor that helps to maintain the shape of the eye. Light entering the pupil forms an inverted image on the innermost tissue layer on the back portion of the eyeball, called the retina. All of the individual components through which light travels within the eye before reaching the retina are transparent, minimizing dimming of the light.

The retina contains the photosensitive rod and cone cells, and neurons. It is a relatively smooth, curved layer with two distinct points; the fovea and optic disc. The fovea is a dip in the retina directly opposite the lens. It is densely packed with cone cells and largely responsible for color vision in humans. The fovea enables high acuity, such as is necessary in reading. The optic disc, also called the anatomical blind spot, is a point on the retina where the axons of the ganglion cells leave the eye to form the optic nerve. No photosensitive cells exist at the optic disc.

The retina lies on a dense network of blood vessels and other tissue that is called the choroid. The choroid supplies the retinal cells with necessary oxygen and remove the waste products of respiration. Additionally, the choroid gives the inner eye a dark colour, which prevents disruptive reflections within the eye. The outer surface of the eyeball is coated with a tough but flexible fibrous tissue called the sclera. The sclera corresponds to the white of the eyeball.
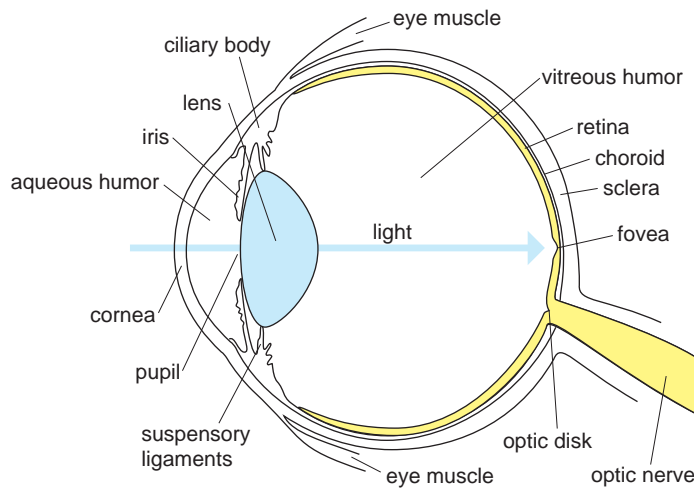
Figure 1.4: **Schematic drawing of the human eye.**

1.3.2   The Human Retina

The retina contains two different forms of photoreceptors, called rods and cones. Although rods and cones are structurally and metabolically similar, their function is quite different. Rod cells are highly sensitive to light allowing them to respond in dim light and dark conditions. Rods do not distinguish between colours, and have low visual acuity. This is why we perceive in darker conditions that objects seem to have less color. Cone cells, conversely, need high light intensities to respond and have high visual acuity. There are three different cone cells in human retinas. This three types of cones are termed L (low frequency), M (middle frequency) and S (supra-frequency) and they are responsive to different portions of the spectrum of light. They allow us to perceive color.
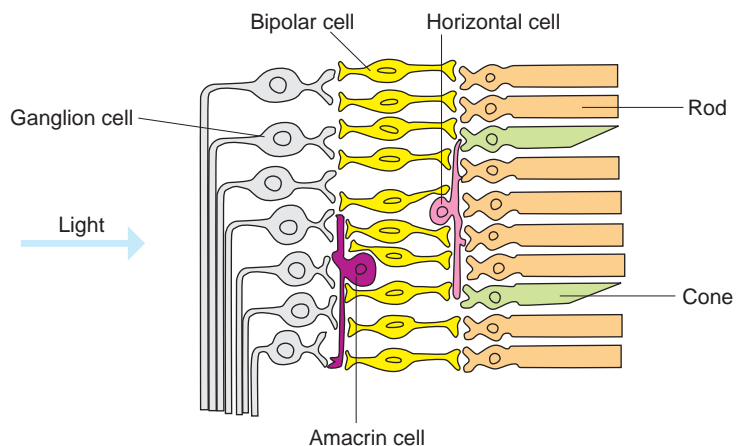
Figure 1.5: **Schematic drawing of the human retina.**

From the photoreceptors, neural responses pass through a series of linking cells, called bipolar, horizontal, and amacrine cells (see Figure 1.5). These cells combine and compare the responses from individual photoreceptors before transmitting the signals to the retinal ganglia cells. The linkages between neighboring cells provide a mechanism for spatial and temporal filtering, facilitating relative, rather than absolute, judgments of intensity, emphasizing edges and temporal changes in the visual field.

It is this layer of massive parallel computation in the retina that makes our eyes efficient. Vision is not about measuring the power and spectral composition of light at a certain very small location in a visual scene, but about using the available information to help an individuum to perform in an optimal way. The network of photoreceptors, horizontal cells, bipolar cells, amacrine cells and ganglion cells discriminate between useful information enhanced and conveyed to the brain and information that is redundant or useless and is better discarded off immediately.

Generalizing a bit, we can say that vision is about objects, their form, their structure and surface as well as their motion. While objects do not change under different lighting conditions, the light reflected by objects is largely dependent on the light source illuminating an object. This means that absolute illumination carries information about the light source and the object illuminated by the light source. Replacing the absolute illumination information with local ratios of illumination differences gets rid of the information about the light source and carries only information about the object illuminated. The local ratios, called contrasts, are independent of the properties of the light source. Biological vision systems make use of this fact and throw away the absolute information right in the front end sensor plane, before any transmission happens.

1.3.3   Transmitting to the Visual Areas in the Brain

Neural responses leave the eye via digital pulses transmitted by the axons of the ganglion cells. The axons of the ganglion cells are bundled together forming the optic nerve, that passes through the optic chiasm where visual responses from each side of the visual field are sent to the opposite side of the brain. From the optic chiasm the reorganized bundles of ganglion cell axons, now called the optic tract, pass to the lateral geniculate nuclei (LGN). Each LGN (one for each side of the visual field) is structured into six layers. Two layers belong to the magnocellular pathway, each receiving input from the achromatic opponent color channel originating in one retina. The remaining four layers belong to the parvocellular pathway and receive input from all three opponent color channels. The LGN also receives a large amount of input (approximately 60 percent) from later stages of the visual pathways in the cortex. These upstream connections appear to provide a mechanism for expectations and previous perception to influence even the early stages of visual processing. From the LGN, visual signals proceed to the primary visual cortex, visual area 2, and then to various areas of higher visual processing.

Two Independent Pathways Leave The Retina

In the retinal ganglia cells the visual system is split into two independent pathways. The magnocellular pathway detects objects, their boundaries and pro-
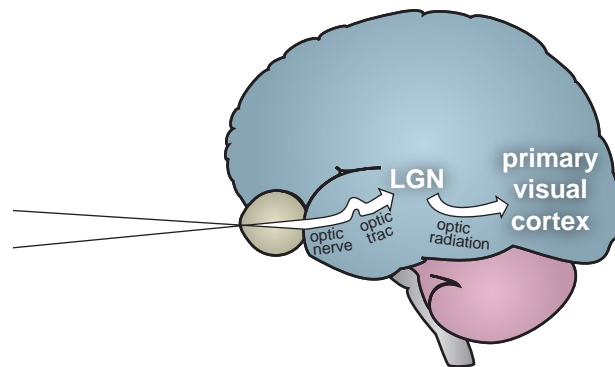
Figure 1.6: **The optic nerve, which is the bundle of all the axons of the retinal ganglion cells, is the output channel of the eye. The optic nerves of both eyes meet in the optic chiasm where the bundles of ganglion cell axons are reorganized. The two bundles of ganglion cell axons that leave the optic chiasm are called the optic tract. The optic tracts pass the bundles of ganglion cell axons to the lateral geniculate nuclei (LGN). The ganglion cell axons terminate in the LGN. From each LGN, visual signals proceed to the primary visual cortex.**

vides a basis for the perception of depth and motion. This pathway begins in a subset of the retinal ganglia cells called the parasol cells. Parasol cells have large receptive fields and receive input that is only depends on light intensity but not on the light wavelength. The parvocellular pathway is responsible for the perception of color and fine detail and receives input from the midget ganglion cells. The midget cells receive input from all three opponent color channels. The judgments of the independent magnocellular and parvocellular pathways are reconciled at a much later stage in visual processing.

Later stages of the magnocellular pathway, particularly MT, have a large number of cells which are selective to motion and direction. This suggests that the magnocellular pathway is also responsible for motion perception.

Motion Provides Important Information

The motion perception pathway is independent of other visual pathways. This was demonstrated by a few people who have no motion perception but have otherwise normal vision [19]. Motion gives valuable clues to our relationship with our environment, like how to steer our eye movements and the time to collision with objects in our environment. Motion is improtant for the perception of relative depths, 3D structures, and for grouping objects that we see.

1.4 Neuromorphic Engineering

A short web research yields one consistent definition of neuromorphic engineering. This definition originates from a careful entry in Wikipedia[10] done by

---

[10]Wikipedia is a big multilingual free-content encyclopedia on the Internet; http://www.wikipedia.org.

Giacomo Indiveri. The definition of neuromorphic and neuromorphic engineering as given by Wikipedia[11], or Giacomo Indiveri is:

> The term neuromorphic was coined by Carver Mead, in the late 1980s to describe Very Large Scale Integration (VLSI) systems containing electronic analog circuits that mimic neuro-biological architectures present in the nervous system. In recent times the term neuromorphic has been used to describe both analog, digital or mixed-mode analog/digital VLSI systems that implement models of neural systems (for perception, motor control, or sensory processing) as well as software algorithms.

> Neuromorphic Engineering is a new interdisciplinary discipline that takes inspiration from biology, physics, mathematics, computer science and engineering to design artificial neural systems, such as vision systems, head-eye systems, auditory processors, and autonomous robots, whose physical architecture and design principles are based on those of biological nervous systems.
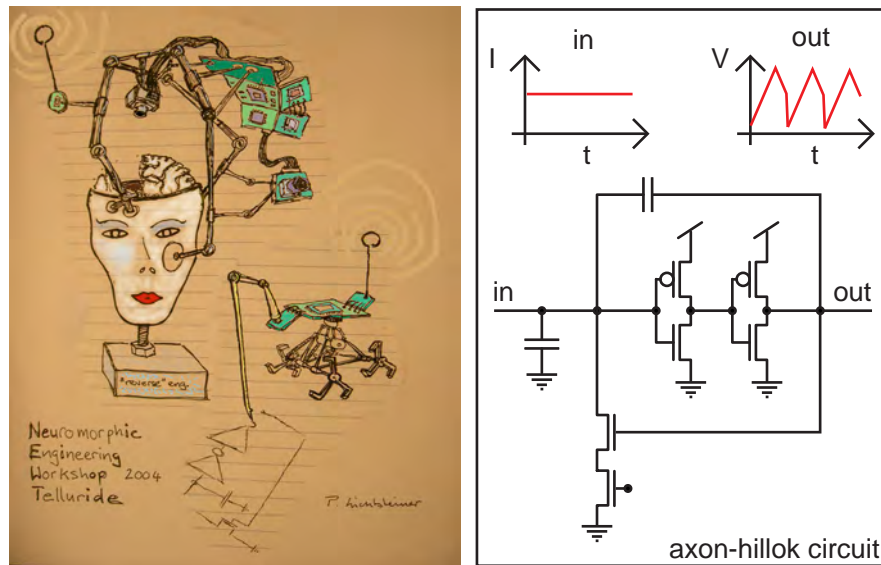
Figure 1.7: **A macabre vision of neuromorphic engineering depicting artificial neuromorphic devices trying to directly reverse-engineer a human brain. Of course, the first circuit that this devices synthesize will be the axon-hillock circuit, also sometimes called the integrate and fire circuit. The axon-hillock circuit, mimicking the axon-hillock (the spike generating site where the axon is attached to the soma) in a biological neuron, is depicted on the right in its most basic version.**

The term neuromorphic first appeared in a scientific publication in Carver Mead's paper *Neuromorphic Electronic Systems* [20] that was published in 1990.

---

[11]http://en.wikipedia.org/wiki/Neuromorphic

One year earlier, with the book on *Analog VLSI and Neural Systems*, Carver Mead laid down a clear foundation for the field of neuromorphic engineering [5].

The belief in the success of neuromorphic engineering was based upon two main arguments: Biological information processing is much more efficient and robust than the nowadays implemented digital information processing methods and biological information processing can be mimicked efficiently using analog VLSI. The first argument was based upon the interesting prediction that our brain is about 10 million times more energy efficient than the best possible digital technology [20]. The second argument is based upon analogies between neurons and electronic circuits, that people started to draw early on in the research of neurobiology and that is obvious because neurons use electrical signaling. Carver Mead found that the computational primitives in transistors can be efficiently used to mimic biological computational primitives if the transistors are used in the so called *subthreshold* domain [5].

An other argument that supports the neuromorphic approach are has become more interesting with the evolution of the microchip fabrication. The decreasing feature size in the microchip fabrication technology allows to implement increasing amounts of structures on a single chip, but the structures become more failure prone. A single faulty structure on a digital chip can make the chip useless. The combination of a higher count of structures with a lower reliability is obviously dangerous for the digital world. On the other hand, biological computation relies on huge amounts of failure prone structures to achieve good results. Neuromorphic engineering is therefore a potential way out of this reliability problem.

Neuromorphic engineering is often also justified as an approach of synthetic epistemology for understanding neural computation. Many neuromophic engineers have implemented theoretical models into neuromorphic chips or systems of neuromorphic chips with the aim to test theoretical neural models. One of these model systems, the neurologist's friend of Tobi Delbrück has become an appreciated tool. Nevertheless, neuromorphic engineering has not yet become a competitive alternative to computer modeling of neural systems, because most neuromorphic devices are not very user friendly, relatively small scale and the implemented structures seldom have enough flexibility such that they can be used conveniently for experiments.

The vision of Carver Mead was to build very large neuromorphic chips using up a whole wafer. This vision has not come true up to today but during the past twenty years, neuromorphic engineering has become an established research field with several institutions following this path all around the globe. Successful demonstrations of neuromorphic engineering principles include vision sensors (often called artificial retinas), silicon cochleas, silicon neurons, non-volatile memory and reconfigurable networks of silicon neurons.

1.4.1 Multichip Systems and AER

Recently different groups started efforts for relatively large multi-chip systems. These systems are based on chips that communicate with each other using the *Address Event Representation* (AER) protocol. The multi-chip system effort has the advantage that single component development runs in parallel and that the single components are less complicated and therefore much better realizable. Multi-chip systems can lead to highly flexible systems that can be built an

rebuilt in a modular fashion to meet specific needs. Furthermore, the combined efforts of a large group of people is has a larger impact and reduces redundant research by elevating communication among the different groups to a higher level.
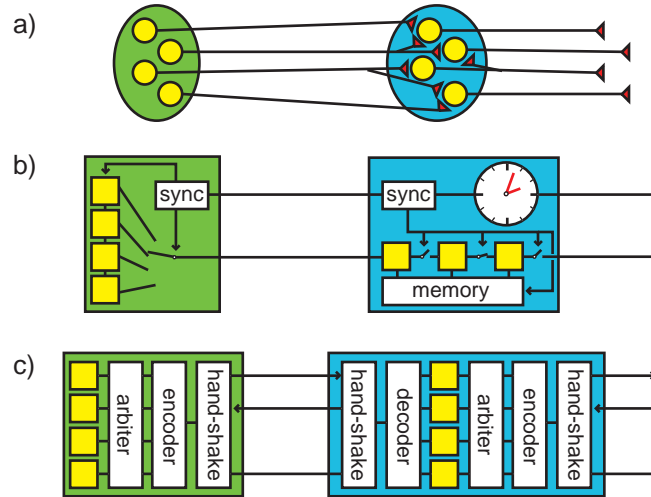

Figure 1.8: **Shows three different principles of communication from a visual sensor population in green to a processing population in blue. Part a) shows a schematic representation of the biological communication scheme, part b) a clocked sequential communication scheme and part c) an asynchronous multiplexed communication scheme.**

Early on in neuromorphic engineering, it became clear that a one of the major differences between brains and computers–the connectivity between processing units–was important for the emulation of neural processing. Directly implementing a neural connectivity pattern on chips was out of question because of space limitations and the basically 2 dimensional structure of chips, but on the other hand, using the standard sequential approach in electronics would right away diminish the potential power of neuromorpic systems. An asynchronous multiplexing strategy was adopted and a communication protocol called *Address Event Representation* that would enable multi-chip systems was developed. The original paper on AER arose from a group meeting in Carver Mead's lab at Caltech, where Dave Gillespie suggested an arbitrated communication scheme [21] for constructing a virtual bundle of axons based on asynchronous transmission of spikes. Massimo Sivilotti designed the initial version of the arbiter logic.

The neural communication relies on voltage pulses, so called action-potentials or spikes, that travel along an axon and elicit a release of neurotransmitters in synapses to other neurons. Every neuron communicates on its own axon and synapses. The interconnection of neurons is very abundant. The large amount of wiring in neural systems is accommodated in a 3 dimensional structure. The massive interconnections of neurons enables parallel and collective computation and is believed to be the main reason for the amazing computational power, robustness and efficiency of neural systems. Figure 1.8a) shows a schematic representation of the neural communication scheme. An important feature of

neural communication is that communication is driven by the occurrence of events.

Figure 1.8b) depicts the principle of an electronic camera or a machine vision system. A global clock generates a signal that allows a synchronized sequential read out of the population of sensor elements–the pixels. The processing happens also sequentially and can be governed by the same clock. The frequency of the clock output signals in the processing domain are nowadays as high as $3GHz$ enabling very fast sequential processing. In clocked systems the sequential readout and the following sequential processing happen in an arbitrary regular frequency that is independent of the occurrence of events.

The AER approach is depicted in Figure 1.8c). A pixel in the sensor population that wants to communicate the occurrence of an event, requests permission to take over a high-speed communication wire[12]. Upon permission the pixel communicates its own identy–its address–on this wire. A receiver is asked in a four phase handshaking cycle to read the address. Subsequently, the receiver routes a pulse to his processing unit that has the same address as the sending unit. This scheme establishes a virtual connection between sending and receiving units with a unity topology that is available upon request. By introducing an address mapping device between the two populations, an arbitrary connection topology can be achieved. AER, is a scalable emulation of the biological communication system that takes the limitations for direct wire connections in chips into account and makes use of the available speed in the communication channel.

AER can be considered as a real time communication system as long as the total activity of a population of sensory or computational units is not close to the bandwidth of the communication channel. The nature of AER leads to the occurrence of simultaneous and conflicting requests for the communication channel by different units. Two different tactics to deal with these conflicts emerged:

1. An arbitration scheme in which sending units have to compete for the access to the communication channel and

2. an ALOHA scheme in which conflicts are solved by throwing away all communication requests involved.

Each of these methods have their advantages (Eugenio Culurciello and Andreas Andreou recently presented a comparative study of AER communication schemes [22]) and decision on which to implement depends on the specific needs. The majority of the neuromorphic engineers have chosen the arbitration method because activity patterns that lead to many conflicts like synchronized activity of subpopulations of units are considered to be important especially when trying to implement collective computation schemes.

During the past few years, Kwabena Boahen has taken the lead in evolving the arbitrated AER that was started with the desing of Massimo Sivilotti. His efforts lead to a point-to-point connectivity scheme [23] with faster arbiters, non-greedy arbiters and peaked lately in a *A Burst-Mode Word-Serial Address-*

---

[12]Often instead of single communication wire, a bundle of shared high-speed wires, called a high-speed bus is used.

Event Link [24–26] that is made available for Tanner Tools[13] via *ChipGen*[14].

Future AER developments are likely to be combined with the highly evolved low voltage differential signaling (LVDS) that is widely used in traditional logic and provide for a very high-speed serial transmission. So far, the circuits that translate an address event into bit-string for serial communication and the LVDS circuits are to complicated and costly in area, that they are not used. With rising bandwidth requirements, certainly they need to be considered.

1.4.2   CAVIAR

The *Convolution AER Vision Architecture for Real-time* (CAVIAR)[15] project is combined effort of researchers of the four institutes:

- Instituto de Microelectrónica de Sevilla (IMSE), Consejo Superior de Investigaciones Científicas (CSIC), Spain.

- Arquitectura y Tecnología de Computadores, Universidad de Sevilla, Spain.

- AVLSI Group, Institute for Neuroinformatics (INI), Universithy of Zurich (ETH Zurich), Switzerland.

- Institutt for Informatikk, Universitetet i Oslo, Norway.

The two prime objectives of the CAVIAR project are the development of a general AER infrastructure for constructing neuromorphic multi-chip systems and the implementation of a perception-action demonstrator vision system that exploits this AER infrastructure. To meet this aim, the caviar project is divided into work packages for the development of the individual modules. The sensory front end of the CAVIAR system is a vision sensor–an artificial retina–that outputs irradiance changes in time in the AER-coded form. Several general purpose AER programmable-kernel convolution chips are used to extract features from the retina input. The next chip in the AER chain is a dimension-reduction competition chip that is used to select the most salient low-level feature extracted by the convolution chips. A learning chip that implements a spike-based learning rule on the spatio-temporal statistics of the AER inputs concludes the chain of AER chips. Besides the *Application Specific Integrated Circuits* (ASICs) built for the CAVIAR project, a considerable effort is put into the development of AER supporting structures like AER mappers, AER splitters and mergers and also AER-computer interfaces.

The CAVIAR project is surpassing all earlier efforts in building an AER multi-chip system in the number of people involved and the number of components developed. The vision sensor presented in this thesis is the serves as the front end input to the CAVIAR system and was developed to meet the CAVIAR project specifications. In section **??** we will present some results of the CAVIAR project.

---

[13]The Tanner Tools are CAD tools for chip-design produced and sold by Tanner Research, Inc, Pasadena, CA

[14]ChipGen, CAD tools for AER chips http://www.neuroengineering.upenn.edu/boahen/meth/tools.htm.

[15]CAVIAR is an EC 5th framework project of the Information Society Technology's (IST) program; IST-2001-34124. This CAVIAR project should not be confounded with the other IST program project called CAVIAR (*Context Aware Vision using Image-based Active Recognition*); IST-2001-37540.

1.5 State of the Art

The First AER Vision Sensor

The first AER vision sensor was built by Mahowald and Mead [27]. Mahowald's silicon retina was a demonstration of concept device that was unusable for any real world task. It was necessary to show it a very high contrast stimulus like a flashing LED to see any sensible response.

Adaptive Temporal and Spatial Filtering

Zaghloul and Boahen [28] took a major step by incorporating both sustained and transient types of cells with adaptive spatial and temporal filtering, meaning that the space and time constants vary according to the illumination level and spatio temporal contrast. This functionality-which of all devices built up to now comes closest to capturing key adaptive features of biological retinas-is achieved by the use of small-transistor log-domain circuits that are tightly coupled spatially by diffuser networks. However, the use of this circuit design style led to large mismatch: the pixel firing rates vary by a standard deviation of 1-2 decades and more than half the pixels do not spike at all for stimuli with 50% contrast. In addition the use of a passive phototransistor current-gain mechanism limits the dynamic range to about 3 decades and leads to a small bandwidth, particularly at low illumination. This chip also requires several low impedance biases very near the power rails that are difficult to efficiently generate. This chip was intended as a model of biology more than as a practical device. Boahen later argued vigorously [29] that heterogeneity (mismatch) can be viewed as being beneficial because it forces solutions to problems that biology faces. Nonetheless it is hard to see how neurons that do not fire at all or that occupy 100x the bus bandwidth of others are beneficial, and it does not answer how the event receivers should learn what the events mean, or if they are able to, how much this learning costs in terms of memory and computation. This chip represents a serious attempt at a new circuit design style that moves away from the classic style of *high-gain feedback with precise passive elements*. It is unfortunate that this design style has so many practical difficulties.

Spatial Contrast Extraction

Barbaro and Ruedi et al. took another big step with two devices [30,31]. They are closest to being dual in functionality to our temporal contrast vision sensor in that their outputs encode spatial rather than temporal contrast. Both devices (the first based on current mode and the second on voltage mode with greatly improved performance) relax the notion of a frame by transmitting events after a global reset, in the order of spatial contrast. Thus readout can be aborted early if limited processing time is available without losing information about high-contrast features. Each event is followed by another event that encodes gradient orientation. These chips are also the first to successfully implement steerable filters. The rate of change of the contrast threshold is varied dynamically after reset to control the bus occupation. *Non-maximum suppression* circuits can thin the detected edges to reduce the number of transmitted events. This device has very low 2% contrast mismatch and a large 6 decade dynamic range. The pixel also clamps the photodiode reverse voltage at a small value, potentially leading

to a substantial reduction in dark current. (Despite this mechanism, it is not explained why their photodiode dark current of 300 fA is so large.) They are presently in commercial development for automotive applications [32].

These devices are also notable for their non-arbitrated AER communication circuits that use collision detection to discard simultaneous events with an active hold-off mechanism to reduce such collisions [30,33]. The main limitation of this architecture is that does not reduce temporal redundancy, (i.e. do temporal derivatives) and is limited in temporal resolution to the frame rate. Power consumption at 300 mW is about 10x the other devices cited here.

Temporal Change Threshold Detection

In an ISSCC 2005 paper [34] that stemmed from earlier work by the group on temporal differencing [35], Mallik et al. reported a temporal change threshold detection imager, which modifies the traditional active-pixel-sensor (APS) CMOS pixel so that it can detect a quantized absolute change in illumination. This synchronous device can store the addresses of pixels that signal change in a FIFO, making a new type of synchronous AER sensor. It has the big advantage that it offers a normal APS mode with small NMOS-only pixels, but the disadvantages of limited 2.5 decade dynamic range and fixed illumination-change threshold, meaning that the single threshold for change is only useful when the scene illumination is very uniform. It is also frame based, so the events times are quantized to the limited global sample rate.

The arbitrated AER Digital Vision Sensor

Culurciello et al. [36–38] built several imaging devices that use AER to communicate the pixel intensity directly. They have the advantage of relatively small pixel size, but the big disadvantage that the bus bandwidth is allocated according to the local scene brightness. Because there is no reset mechanism and because the event interval directly encodes intensity, a pixel can take an arbitrary amount of time to emit an event. A single highlight in the picture can saturate the bus. This work is notable for the development of a new style of very fast non-arbitrated AER communication circuits.

The Time To First Spike Imager

Harris' group have built experimental devices called the Time To First Spike (TTFS) imager [39] and the Time-Based imager [40]. Each of these devices emits events after global frame reset according to the pixel illumination; bright pixels spike first followed by darker ones. The threshold can be dynamically varied to ensure that all pixels spike within a given frame time. This imager has the advantage that the pixels can have large dynamic range. The main disadvantages are large pixels and that uniform scenes produce massive collisions between events, making it hard for the device to encode uniformity. The uniformity characteristics for these devices are not specified.

A Foveated AER Vision Sensor

Finally, Hafliger's group [41] reported an experimental foveated AER vision sensor with center array of intensity-coding pixels based on Culurciello's and

a periphery of novel change-detecting pixels based on combining an adaptive photoreceptor [42] with a bump circuit [43]. This device is still at first silicon.

1.6 The AER Temporal Contrast Vision Sensor

1.6.1 The Goal of This Work

The goal of this project is to develop a vision sensor that responds purely to temporal changes with an AER output. The sensor serves as the front-end vision sensor to the CAVIAR system. The vision sensor shall surpass existing transient vision sensors in stability, sensitivity, uniformity of the response, dynamic range and bandwidth and reach a good enough quality so that the sensor can be used for real world applications.

This Sensor Fills a Gap

As discussed in section 1.3, biological vision is split into to pathways, the magnocellular pathway and the parvocellular pathway. The magnocellular pathway, that is not sensitive to color but very sensitive to temporal irradiance changes, is the primary pathway for object detection, object boundary detection and provides a basis for the perception of depth and motion. Most electronic vision sensors are imagers, responding to the absolute irradiance and lacking any temporal filtering. A few devices that asynchronously respond to temporal changes have been built so far [27,28,41], but all of them comprise also cells that respond to the magnitude of irradiance or to spatial contrast and in consequence do not provide fast, sensitive and low mismatch response to temporal changes. The sensor presented here is unique in this way.

High Compression and Meaningful Output

Being purely sensitive to temporal changes, our sensor makes optimal use of the asynchronous AER communication by only sending information when a change in a pixel is detected. The combination of the pure temporal change sensitivity and the AER communication achieves a high compression and high temporal resolution. Furthermore, the information provided by the output of our sensor allways means that something happened. In particular examples events are only produced by interesting objects, while the rest of the visual scene will never produce any response. This property can conveniently be used for observation of moving objects, like cars or people.

Freeing Vision From Frames

The time domain computation and the high resolution in time of our sensor enables a different approach to vision problems that is based on the temporal relation between incidents that produce events. Rigid objects for example can very conveniently segregated by their different relative motion. Tracking of objects becomes easier with our sensor because a moving object will elicit a continuous stream of events and in comparison to a frame based approach will not be lost between two snapshots. The output of our sensor is driven by the dynamics in a visual scene and bandwidth is most appropriately allocated to

the pixels communicating these dynamics. The output frequency of events is not arbitrarily given by a clock but is a direct result of the visual input; slow movements produce low frequency output while fast movement leads to high frequency output.

The Technical Goal

The technical goal of the work presented here was to develop a transient vision sensor that is sensitive to relative temporal irradiance changes and that responds to such changes with address events. This sensor has to discriminate between positive and negative irradiance changes. The sensor has to work reliably under a wide range of illumination conditions and its output has to be independent of these illumination conditions. The sensor has to be implemented in a medium resolution; at least 64×64 pixels.

1.6.2   The Challenge

A Good Engineering Trade-Off

The main challenge of this project is to design a pixel that works over a wide range of illuminations, consumes a relatively small amount of silicon area and power, works reliably in different environmental conditions, especially different temperatures, has a uniform response and a very high bandwidth. Almost all of these metrics oppose each other, i. e. a uniform response is usually achieved by low mismatch and that results in large transistors in contrast to the desire of keeping the area low, or high bandwidth usually requires high currents contrasting the wish for low power. The quality of the trade-off between these competing yearnings will ultimately determine the success of the pixel.

A High-Quality Prototype

Up to the present day, many neuromorphic chips are good enough for the demonstration of concept but do not catch serious interest outside the community of neuromorphic engineering. Although many design are revolutionary, the effort that went into the implementation of the design can not keep up with its revolutionary concept. We think that the success of neuromorphic engineering is closely bound to development and fabrication of high quality prototypes that can be used by people distant from the ones that built them. Therefore, we put a lot of effort in the evolution of our design onto a level where it is robust and self-contained enough, that it can be handed to other people. The challenges in making a high-quality prototype are manifold; the prototype has to be small, the design has to be robust enough to work well without tuning in different conditions, the device has to perform well without necessary adjustments and finally and most important, the prototype has to work very reliable.

1.6.3   The Achievements

The first steps of this work have been laid down by Jörg Kramer. He implemented a 48×48 AER transient sensor [44] that based its temporal derivative computation on rectified charge feedback [45] and responded to positive logarithmic irradiance changes with ON events and to negative logarithmic irradiance

changes with OFF events. Ensuing the tragic death of Jörg Kramer in 2002, we developed an improved version of his AER transient sensor with 32×32 pixels. This sensor was in use for the CAVIAR project and served as the input sensor of an AER 2-chip system for an orientation tuning experiment. Jörg Kramer's transient optical sensor, his ON/OFF transient sensor and our improved version of this sensor are discussed in detail in chapter 2.

The problem with these devices is mismatch in the feedback elements, making it difficult to set a low contrast threshold across a large array, potential self-sustained oscillations in the photoreceptor circuit and undesirable very sensitive low-impedance biases. Because we did not find any solution to this problem in Jörg Kramers design, we developed a novel pixel design that combines an active continuous-time logarithmic photosensor with a well-matched self-timed switched-capacitor amplifier.

The final implementation of this new pixel design resulted in a vision sensor chip that contains a 128128 array of pixels. The chip was built in a 0.35u 4M2P process yielding 40x40 um2 pixels with 9.4% fill-factor. The measured dynamic range of this chip is $120dB$, by far surpassing any comercially available imagers that reach intrascene dynamic ranges that are usually lower than $50dB$. The chip power consumption is $23mW$. The response latency varies slowly with illumination and is $15\mu s$ at $\geq 1klux$ pixel illumination, this response latency corresponds to the inter-frame time of a $66kHz$ frame rate imager. The mismatch in event contrast threshold corresponds to 2.1% scene contrast.

In chapter xxx we will describe the new pixel architecture and present three different implementations of the new pixel along with characterization results in chapters xxxx.

The successful implementation of the temporal contrast vision sensor allowed to use it in different applications of which two are described in section xxxx. Currently several people are successfully using our latest temporal contrast vision sensor for different vision applications, a short overview of these efforts is presented in section xxxx.

Chapter 2

The ON/OFF Transient Imager

2.1   Jörg Kramer

In September 2002, the article with the title *An Integrated Optical Transient Sensor* was published in the *IEEE transactions on circuits and systems* [45]. This was Jörg Kramer's last single authored publication. It appeared two months after Jörg Kramer died from heat exhaustion after climbing Telescope Peak from Badwater in Death Valley, California.

The article reports on a transient optical sensor as Jörg Kramer stated in the abstract of [45]:

> The implementation of a compact continuous-time optical transient sensor with commercial CMOS technology is presented. In its basic version, this sensor consists of a photodiode, five transistors and a capacitor. The proposed circuit produces several output signals in parallel. These include a sustained, logarithmically compressed measure of the incoming irradiance, half-wave rectified and thresholded contrast-encoding measures of positive and negative irradiance transients, and a signal that shows a combination of the sustained and the bidirectional transient response. The particular implementation reported in this work responds to abrupt irradiance changes with contrasts down to less than 1% for positive transients and 25% for negative transients. Circuit modifications leading to more symmetric contrast thresholds around 5% are also described. Due to their compactness these transient sensors are

> suitable for implementation in monolithic one- or two-dimensional imaging arrays. Such arrays may be used to sense local brightness changes of an image projected onto the circuit plane, which typically correspond to moving contours.

The optical transient sensor has been used as a front-end for various biologically-inspired image-processing circuits [46–48].

A biologically inspired vision sensor that responds to ON and OFF transients was part of the CAVIAR proposal issued in January 2002. The proposed basis for the approach was the optical transient sensor. The goals for the retina subproject was stated on the CAVIAR proposal page 5:

> **Sensing Retina Chip.** A medium-resolution retina chip will be provided that responds to local positive and negative irradiance transients in a two-dimensional input image. It perfomrs focal-plane brightness adaptation and data compression for efficient use of the sensing and communication bandwidths.

Jörg Kramer was the person responsible for the retina part of the CAVIAR project. Along with the proposal went the development of the first temporally differentiating retina that used AER communication. Jörg Kramer used the transient optical sensor pixel as front end for this sensor that he presented at the *IEEE International Symbosium on Circuits and Systems* in May 2002 [44] under the title: *An ON/OFF transient imager with event-driven, asynchronous readout.* The CAVIAR project got sponsored by the IST and was due to begin in June 2002. Ensuing the tragic events in the summer 2002, Tobi Delbrück took over Jörg Kramer's responsibilities in the CAVIAR project and in autumn 2002 I joined in as a PhD-student.

Jörg Kramer did carefully lay out a path for a successful implementation of the CAVIAR retina based on the optical transient sensor [45] and he already implemented a first version of the CAVIAR retina [44]. As an appreciation of the fundamental function Jörg Kramer's work had for our work presented here, we will dedicate the next sections to give a summary of the optical transient sensor and the ON/OFF transient imager.

2.2 The Optical transient sensor

Figure 2.1 shows the adaptive photoreceptor [49] on the left and the optical transient sensor [45] pixel on the right. The resemblance of this two pixel designs is obvious. The optical transient sensor is a combination of Tobi Delbrück's adaptive photoreceptor [49] with a rectifying and thresholding differentiating element [50].

The optical transient sensor comprises a photodiode D, a starved inverting amplifier built with M_n and M_pr, a rectifying differentiating element formed by M_{ON} and M_{OFF}, a capacitor C connected to the sources of M_{ON} and M_{OFF} and a feedback transitor M_{fb}.

The photocurrent I_{ph} arises from phototransduction in the photodiode D and linearly depends on the irradiance. Natural illumination conditions lead to a photocurrent that is small enough to keep M_{fb} in weak inversion (subthreshold
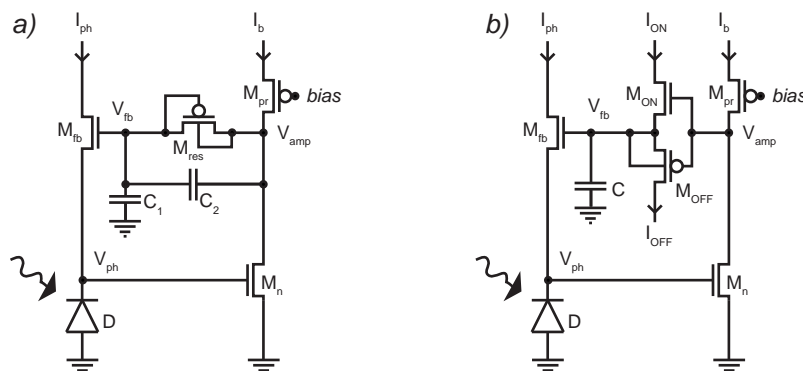
Figure 2.1: **Shows in a) Tobi Delbrück's adaptive photoreceptor and in b) Jörg Kramer's optical transient sensor pixel.**

domain). Because of the subthreshold operation of Mfb and because of the high-gain inverting feedback, that keeps the source potential V_{ph} almost constant, V_{fb} changes logarithmically with irradiance. The feedback loop significantly speeds up the circuit response with respect to a simple source-follower configuration XXXXrefToIntroPPSXXX, since the small photocurrent does not have to charge the relatively large photodiode capacitance [49]. Using the basic subthreshold domain transistor equation the relation between photocurrent and the gate of the feedback transistor can be written as:

$$I_{ph} = I_0 \cdot e^{(\kappa V_{fb} - V_{ph})/U_T} \iff V_{fb} = \frac{U_T}{\kappa} \ln\left(\frac{I_{ph}}{I_0}\right) + \frac{V_{ph}}{\kappa}, \qquad (2.1)$$

where I_0 is the current scaling parameter (also called the off-current), κ the subthreshold slope factor (in this case of M_{fb}), U_T denotes the thermal voltage kT/q given by the absolute temperature T, the Boltzmann constant k and the elementary charge q.

Equation 2.1 states that the output node V_{fb} is logarithmically dependent on the input irradiance. The difference at the output nodes V_{fb} of two adjacent optical transient pixels encodes a contrast rather than a brgihtness difference:

$$V_{fb1} - V_{fb2} = \frac{U_T}{\kappa}(\ln(I_{ph1}) - \ln(I_{ph2})) = \frac{U_T}{\kappa} \ln\left(\frac{I_{ph1}}{I_{ph2}}\right). \qquad (2.2)$$

Contrast encoding is generally more useful for vision than encoding of brightness, because most objects in our environment do not generate light but merely reflect it in a more or less diffuse manner. While the information about object surfaces is found in local image contrast, the global image brightness manly conveys information about the light source. Furthermore, a logarithmic representation enables potentially wide dynamic range due to the compression of the input irradiance range.

In steady state, the transistors M_{ON} and M_{OFF} are turned off because their shared gate voltage V_{amp} is adapted to their shared source voltage V_{fb}. An increase in irradiance leads a rise of V_{amp} relative to V_{fb}. The resulting voltage difference between gate and source of M_{ON} leads to a current I_{ON} charging C. A decrease in irradiance leads to a discharge of C through M_{OFF},

resulting in an OFF current I_{OFF}. The currents I_{ON} and I_{OFF} are proportional to the temporal derivative of V_{fb} and to the capacitance of C. Assuming that all parasitic currents and the delay through the inverting amplifier and the differentiating element can be neglected, we can easily derive the ON and OFF current as a function of the photocurrent. The relation between the ON (OFF) current I_{ON} (I_{OFF}) and the feedback node V_{fb} yields:

$$I_{ON\,(OFF)} = C \cdot \frac{d}{dt} V_{fb}. \tag{2.3}$$

Using equation 2.1 and equation 2.3 and assuming $\frac{d}{dt}V_{ph} \ll \frac{d}{dt}V_{fb}$ we obtain:

$$I_{ON\,(OFF)} = C \cdot \frac{U_T}{\kappa} \cdot \frac{d}{dt} \ln(I_{ph}) = C \cdot \frac{U_T}{\kappa} \cdot \frac{\frac{d}{dt} I_{ph}}{I_{ph}}, \tag{2.4}$$

Equation 2.4 points out that I_{ON} and I_{OFF} are proportional to the relative change in irradiance rather than to absolute change in irradiance. This property, that comes from the logarithmic current to voltage conversion, has a the benefit that the ON and OFF currents represent temporal contrast rather than temporal change of brightness.

Although the equations 2.1 to 2.4 describe some of the most important circuit characteristics in an intuitive way, they do not cover the real dynamics of the circuit, i. e. they do not explain the behavior of V_{amp}. Thorough calculations presented in Jörg Kramer's paper [45] lead to the following result for V_{fb}:

$$\begin{aligned} V_{fb}(t) &= V_{fb}(0) + \frac{U_T}{A_{ON}+1} \\ &\quad \cdot \ln\left(1 + \tau_{ON}^{-1} \int_0^t \left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{A_{amp}\kappa_{ON}} dt'\right). \end{aligned} \tag{2.5}$$

$$\tag{2.6}$$

A_{amp} is depends on the Early voltages V_{nE} and V_{pE} of M_n and M_{pr} and on the subthreshold slope factor κ_n of M_n

$$A_{amp} = -\frac{\partial V_{amp}}{\partial V_{fb}} = \frac{\kappa_n}{U_T}\left(V_{nE}^{-1} + V_{pE}^{-1}\right)^{-1}, \tag{2.7}$$

$$\tag{2.8}$$

The loop gain for ON transients A_{ON} is given by

$$A_{ON} = A_{amp}\kappa_{ON}\kappa_{fb}, \tag{2.9}$$

$$\tag{2.10}$$

where κ_{ON} and κ_{fb} are the subthreshold slope factors of M_{ON} and M_{fb}. The small-signal differentiator time constant for ON transients τ_{ON} is

$$\tau_{ON} = \frac{U_T C}{(A_{ON}+1) \cdot I_{ON}(0)}. \tag{2.11}$$

$$\tag{2.12}$$

Jörg Kramer gives the following relations for the ON transient current I_{ON} and the amplifier output voltage V_{amp} [45]

$$I_{ON}(t) = I_{ON}(0) \cdot \frac{\left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{A_{amp}\kappa_{ON}}}{1 + \tau_{ON}^{-1} \int_0^t \left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{A_{amp}\kappa_{ON}} dt'} \qquad (2.13)$$

$$V_{amp}(t) = V_{amp}(0) + A_{amp}U_T \left[\ln\left(\frac{I_{ph}(t)}{I_{ph}(0)}\right) - \frac{\kappa_{fb}}{A_{ON}+1} \right.$$
$$\left. \cdot \ln\left(1 + \tau_{ON}^{-1} \int_0^t \left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{A_{amp}\kappa_{ON}} dt'\right) \right] \quad . \qquad (2.14)$$

Analogous relation can be established for OFF transients [45]:

$$V_{fb}(t) = V_{fb}(0) - \frac{U_T}{(A_{OFF}+1)\kappa_{OFF}}$$
$$\ln\left(\tau_{OFF}^{-1} \int_0^t \left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{-A_{amp}\kappa_{OFF}} dt'\right) \qquad (2.15)$$

$$I_{OFF}(t) = I_{OFF}(0) \cdot \frac{\left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{-A_{amp}\kappa_{OFF}}}{1 + \tau_{OFF}^{-1} \int_0^t \left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{-A_{amp}\kappa_{OFF}} dt'} \qquad (2.16)$$

$$V_{amp}(t) = V_{amp}(0) + A_{amp}U_T \left[\ln\left(\frac{I_{ph}(t)}{I_{ph}(0)}\right) - \frac{\kappa_{fb}}{(A_{OFF}+1)\kappa_{OFF}} \right.$$
$$\left. \cdot \ln\left(1 + \tau_{OFF}^{-1} \int_0^t \left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{-A_{amp}\kappa_{OFF}} dt'\right) \right]. \qquad (2.17)$$

The equations 2.5 to 2.14 and 2.15 to 2.17 describe the optical transient sensor more precise than the equations 2.1 to 2.4, but they are derived without taking any parasitic capacitances (like the photodiode capacitance) or parasitic currents (like the leakage currents in M_{ON} and M_{OFF}) into account. Theses simplifications lead to a first order system, which does not describe potential stability problems [1] arising in this feedback system. However, the analysis is certainly worth the effort, if only to get an understanding of the behavior of V_{amp}.

If we consider only small transient signals $d/dt\, I_{ph}(t)$ and short time delays t, where

$$\tau_{ON} \gg \int_0^t \left(\frac{I_{ph}(t')}{I_{ph}(0)}\right)^{A_{amp}\kappa_{ON}} dt'$$
$$\tau_{OFF} \gg \int_0^t \left(\frac{I_{ph}(0)}{I_{ph}(t')}\right)^{A_{amp}\kappa_{OFF}} dt' \qquad (2.18)$$

---

[1]We encountered an oscillating behavior in our implementation of Jörg Kramer's pixel.

the effect of the feedback can be neglected and the equations 2.5 to 2.14 and 2.15 to 2.17 be greatly simplified [45]. In such circumstances, V_{fb} does not change, I_{ON} and I_{OFF} have a power low dependence on the relative photocurrent change and V_{amp} is a logarithmically dependent on the relative photocurrent change.

It is intriguing that the optical transient pixel computes 3 different useful functions of the input irradiance in parallel at 4 different terminals in continuous time using only 5 transistors, a photodiode and 1 capacitor. It does show how much can be achieved with a small number of transistors that are operated in the subthreshold domain in a most elegant way.

At the same time, the optical transient sensor is an example of a circuit that will not meet all the analytically elevated expectations, especially when implemented in an imaging array, because the output signals will be strongly affected by mismatch. The steady state voltage levels of the signals V_{fb} and V_{amp} will potentially vary much from pixel to pixel in an array. Furthermore, the readout of two voltage signals and two current signals while maintaining a low noise level in the pixels and keeping the transistor count low is not trivial. The amplifier output voltage V_{amp} enhances transients in a similar fashion as the adaptive photoreceptor circuit [49]. For this isolated function, however, the optical transient sensor is not competitive with the adaptive photoreceptor, in which the gain is controlled by a well matched capacitor feedback via C_1 (see Figure 2.1(a)). The drawbacks of this pixel do show up only in an imaging array and are of course not advertised in [45].

Because the ON and OFF currents I_{ON} and I_{OFF} are not affected by the steady state offsets, they represent the most promising output terminals of the optical transient pixel. It is, with this premises, not surprising that Jörg Kramer used these ON and OFF currents for the implementation of the transient optical sensor into the ON/OFF transient imager [44].

2.3 The ON/OFF Transient Imager

Using the ON and OFF currents of the optical transient sensor [45] in combination with analog-to-digital conversion circuits, Jörg Kramer built a focal plane vision sensor. The compact pixels of this sensor respond with to positive and negative irradiance transients on separate channels. This sensor is read-out using the AER protocol.

2.3.1 The Pixel

Figure 2.2 shows the pixel circuit of the ON/OFF transient imager. The currents I_{ON} and I_{OFF} computed in the front end of the pixel (Figure 2.1) are mirrored (by M_{mp1} and M_{mp2} for the ON currents and M_{mn1}, M_{mn2}, M_{mp3} and M_{mp4} for the OFF current) and fed into separate and independent analog to digital conversion structures for each of the two channels.

The mirrored versions of the currents I_{ON} and I_{OFF} are integrated on the capacitors C_{ON1} and C_{OFF1} respectively. A tunable, constant leakage current through M_{t1} and M_{t2} that is adjusted by the bias voltage *thres* provides a threshold for the magnitude of the signal currents and pulls the voltages V_{ON} and V_{OFF} to GND in the absence of a signal.
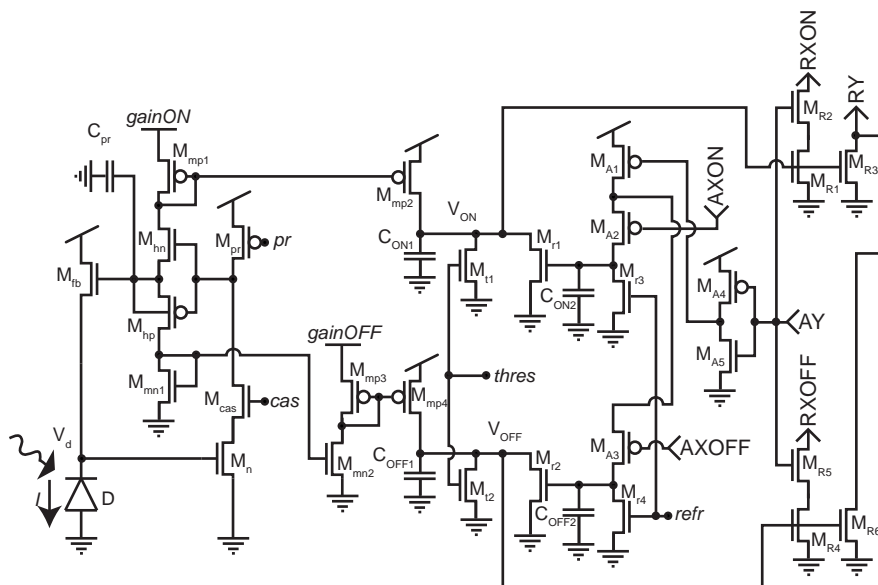
Figure 2.2: **Jörg Kramer's ON/OFF transient imager pixel.**

When V_{ON} (V_{OFF}) overcomes a certain threshold (given by a pull-up transistor in the peripheral AER circuits and discussed later in 2.3.2), a handshaking cycle is initiated that leads to the communication of an ON-event (OFF-event) and in return to the reset of V_{ON} (V_{OFF}) to GND via M_{r1} (M_{r2}). A prevailing signal current I_{ON} (I_{OFF}) will lead to another event and again to a reset. A higher threshold that V_{ON} (V_{OFF}) has to overcome to produce an event leads to a lower resolution of the analog to digital conversion in the pixel.

To limit the maximum output rate of a pixel, a so called refractory period that clamps the voltage V_{ON} (V_{OFF}) to GND for an adjustable time is implemented by the capacitor C_{ON2} (C_{OFF1}) and the transistor M_{r3} (M_{r4}). This refractory period is adjusted by the voltage on the node $refr$.

The nodes $gainON$ and $gainOFF$, which constitute the source potentials of the transistors M_{mp1} and M_{mp3} are separate power supplies, also called low impedance biases. They are used to amplify the currents I_{ON} and I_{OFF} according to the following equation (computed from the basic sub-trheshold equations and given only for gainON):

$$I'_{ON} = \frac{I'_0}{I_0} I_{ON} e^{(VDD-gainON)/U_T}. \qquad (2.19)$$

The potentially large amplification is necessary because the signals I_{ON} and I_{OFF} are very small. The cost of this so called *tilted mirror* is that the mirrored current $I_{ON'}$ is a significant current even when $I_{ON}=I_0 = 0$, which is the condition for no ON signal from the photoreceptor. Furthermore, the tilted mirror introduces significant mismatch. In a large array, an equal distribution of the powersupplies $gainON$ and $gainOFF$ requires low resistance wires.

The transistors M_{R1} to M_{R6} and M_{A1} to M_{A6} are needed for the AER communication and are discussed in the next section.

2.3.2 The Array and The AER

The ON/OFF transient imager uses AER to communicate events. Because we used the same AER communication principle in all our chips and because it appears here first, we will give a thorough discussion of the applied principle in the following paragraphs. In later chapters we will omit any further in depth AER discussions and refer back here.

The ON/OFF transient imager pixels are arrayed in a 2D matrix as shown in Figure 2.3a). The pixel-array is flanked on two sides by the AER communication infrastructure, which comprises two encoders, two communication logic blocks and two arbiters. The asynchronous communication to a receiver functions by three sequential four phase handshaking cycles:

1. *ROW HANDSHAKE.* The pixel pulls low the request RY (shown in Figure 2.3b)), which is shared among pixels in a row and is routed through the column arbiter. If the row, to which the requesting pixel belongs, is selected by the arbiter, the communication logic block sends an acknowledge signal AY back to the pixel. The signal AY is an active high signal[2] transmitted on a wire that is again shared among pixels in a row.

2. *COLUMN HANDSHAKE.* Upon receipt of the row acknowledge signal AY, the pixel triggers an active low column request signal that is routed along the column-wise shared wire $RXON$ for an ON event and along $RXOFF$ for an OFF event. The ensuing arbitration will consequently lead to a column acknowledge signal $AXON$ for an ON event and $AXOFF$ for an OFF event. The presence of both the acknowledge signal from the row and the column will reset the pixel and lead to the removal of both request signals.

3. *CHIP HANDSHAKE.* The grant of a column acknowledge signal leads to a chip request $REQUEST$ to receiver. The encoder uses the column acknowledge signal and the row acknowledge signal to program an Y-X-address that represents the location of the pixel that made an event. In a single sender single receiver configuration, this address is usually valid on a bus a certain set-time before the $REQUEST$ signal is activated. The receiver asserts the reading of the address by setting the $ACKNOWLEDGE$ signal active. Upon receiving the $ACKNOWLEDGE$ signal, the column and row acknowledge signals are removed and a new handshaking cycle can begin.

To guaranty the proper sequential handshaking and to solve conflicts between different pixels trying to send an event, the communication logic block and the arbiter are needed. The y-arbiter, the y-communication logic and the y-encoder are shown in Figure 2.4.

The column request wire RY is hooked up to the communication logic element. This element, also called the *arbLogic*, latches an active state of RY in the signal RY' until the $ACKNOWLEDGE$ signal is activated as shown in the truth table in Figure 2.5a). The signal RY' is hooked up the arbiter tree. If RY' is active and no other RYn' signal from a different row is active, the signal

---

[2]We call a signal an active high signal if the signal is in an active state pulled to the higher powersupply potential, which is usually called VDD. The active low signal is a signal that is pulled to GND when it is active.
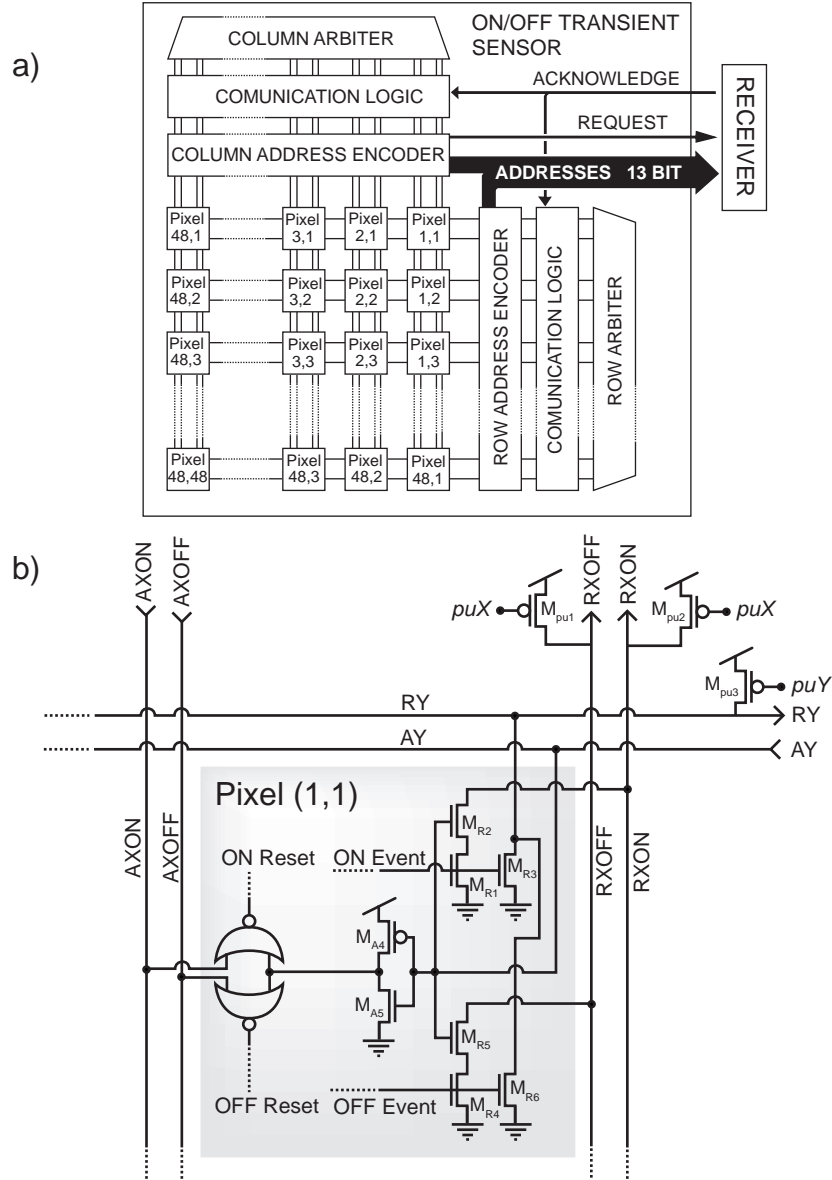
Figure 2.3: **The internal chip AER communication structure comprises the pixels-array, an encoder, a communication logic block and an arbiter for both the column and the row shown in a). The internal handshaking cycles use row-wise and column-wise shared request and acknowledge signals. The active low request signals RY, $RXON$ and $RXOFF$ are pulled high by the pull-up transistors M_{pu3}, M_{pu1} and M_{pu2} respectively, as shown in b). The acknowledge signals AY, $AXON$ and $AXOFF$ are driven by the handshake logic.**
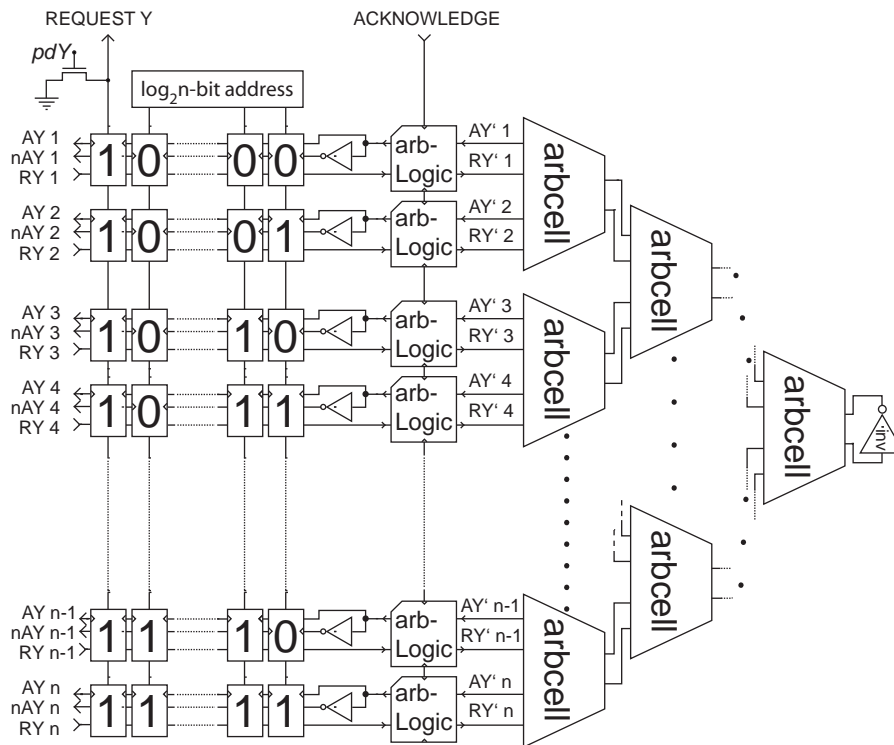
Figure 2.4: **The row AER communication infrastructure consists of the Arbiter, the Arbiter Logic and the Encoder.**

a)

ACK.	RY	RY'
0	0	p
0	1	1
1	0	0
1	1	1

b)

ACK.	AY'	AY
0	0	0
0	1	1
1	0	0
1	1	p

c)

0 = inactive
1 = active
p = previous state

Figure 2.5: **The communication logic works according to this truth table. a) shows the input from the row and b) the output to the row. A legend explaining the used terminology is depicted in c).**

AY' is activated by the arbiter after all previous RY' from other rows have been removed. The communication logic will activate the signal AY in response to the active state of AY' unless the $ACKNOWLEDGE$ signal is active, which leads to the preserving of the previous state of AY, as shown in the truth table in Figure 2.5b).

The rationale behind this scheme is the following: Latching RY' prevents bad states in the arbitration process because the latched signal RY' is stable. Furthermore because the RY' can not be removed until $ACKNOWLEDGE$, AY can not be removed until the chip gets an acknowledge $ACKNOWLEDGE$ signal from the receiver. By this restriction, it is made sure that the address on the bus, which is driven by AY is held valid until at least the activation of $ACKNOWLEDGE$. On the other hand, to prevent glitches, another AY signal can not be activated until the signal $ACKNOWLEDGE$ is again removed. Figure 2.6 shows one complete AER communication cycle with all signals involved in the AER communication.
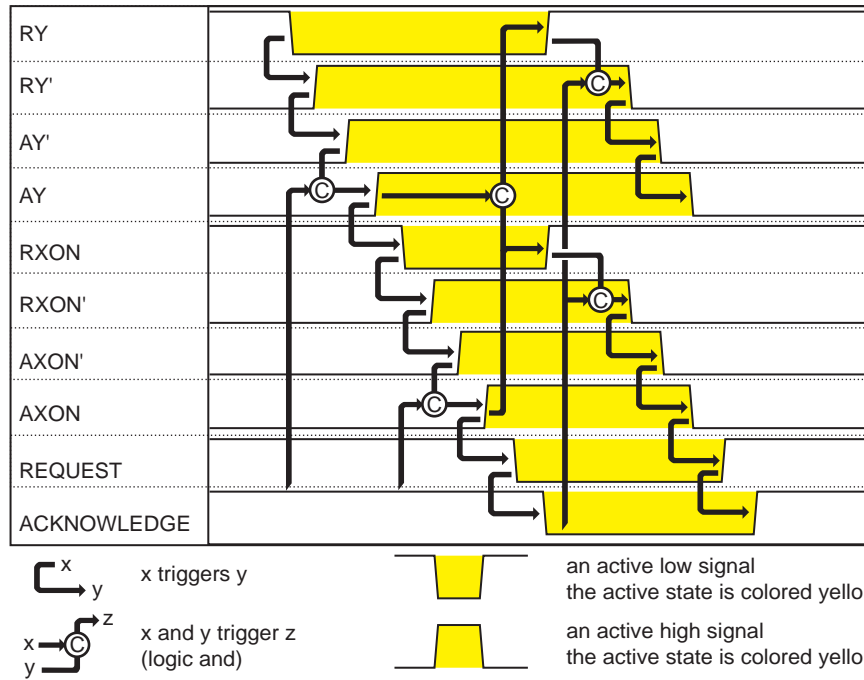

Figure 2.6: **Shows one complete AER cycle for an ON event.**

The applied AER communication principle is Kwabena Boahen's *point-to-point* scheme [23], which is an evolved version of the original scheme proposed in Carver Mead's lab [21] (see also section 1.4). For a detailed description of the circuits, please refer to appendix XXXX.

2.3.3 Experimental Results

Jörg Kramer implemented the ON/OFF transient imager in a $0.35\mu m$, CMOS, 2 poly, 4 metal[3] process. The pixels he designed have an area of $(32.8\mu m)^2$) with a fill factor of 9.2%[4]. The chip, which uses $3.3V$ powersupply, has a total size of $2.29mm \times 2.12mm$ [44].

Jörg Kramer characterized the ON/FF transient imager using LED's, an black and white paper stimuli. He measured a dynamic range of $66dB$, a maximum readout rate of $20MHz$ and a power dissipation ranging from $2.25mW$ in absence of any activity to $55.85mW$ for the maximum readout rate [44]. In [44] he showed figures of the imager response lacking any spontaneous activity and showing a reasonable symmetry between ON and OFF events.
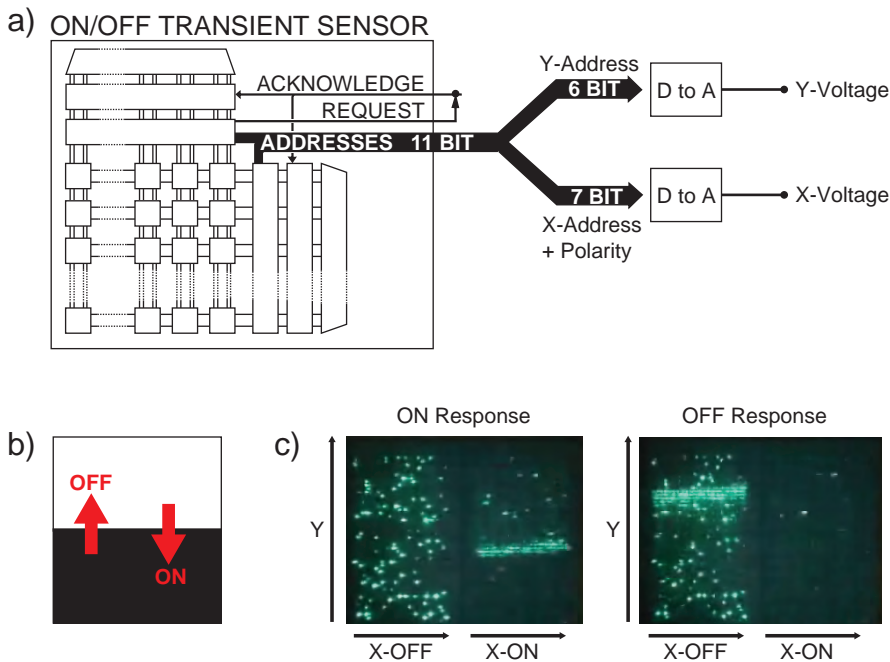

Figure 2.7: **Shows how the AER output of the transient ON/OFF sensor could conveniently be observed on an oscilloscope in real-time. Part a) shows the digital to analog conversion scheme, b) shows the standard stimulus used for experimentation with the sensor and c) shows photographs of the oscilloscope output.**

When Tobi Delbrück and I took over the quest of the development of a high quality AER vision sensor that responds to relative changes in illumination, we started to experiment with Jörg Kramer's device. Even though we spent considerable effort on understanding the different parameters of Jörg Kramer's sensor, we did not achieve a tuning of the sensor that would have let us reproduce Jörg Kramer's results.

---

[3]This is the way a the features of a certain chip fabrication process are described. Here it means that the minimum gate length in this process is $0.35\mu m$ and that the process is a standard CMOS process with 2 poly-silicon layers and 4 metal interconnect layers.

[4]The fill factor is the ratio between the photo-active area and the total area of a pixel

For the actual tuning of the ON/OFF transient imager, we used two *Digital to Analog Converters* (DACs) that translate the bit-pattern y-address and x-address plus the polarity (ON or OFF) into two analog voltages that we observed on an oscilloscope in the X versus Y mode. This digital to analog translation is shown in Figure 2.7a).

Figure 2.7c) shows the response of the ON/OFF transient imager to black and white paper stimulus, depicted in Figure 2.6b), moved in front of the sensor for our best effort tuning. The response clearly represented the input but we were not able to achieve at the same time a symetric response for ON and OFF. Furthermore, we could not find a parameter set that allowed to have low spontaneous activity paired with a reasonable sensitivity to illumination changes. Figure 2.7c) shows significant spontaneous activity in the OFF channel.

2.3.4   Conclusion

Our observations led us to the conclusions that Jörg Kramer's ON/OFF transient imager had some weaknesses that needed to be corrected by small changes in his design. We drew that conclusion from our experience while trying to tune Jörg Kramer's imager. The quick decision to redesign the ON/OFF transient was also motivated by the assumption that a full understanding of the circuits, the design and also the problems and limitations of this design could be gained rather by a creative effort than by a study of what Jörg Kramer left behind.

2.4   The Small CAVIAR Retina

Starting with Jörg Kramer's we designed an improved transient ON/OFF temporal vision sensor. That vision sensor was also intended as the front end vision sensor for the first CAVIAR system implementation, which was based on relatively small and cheap chips, and we therefore call it the *small CAVIAR retina*.

2.4.1   The Pixel

Figure 2.8 shows our pixel design. The places of the most significant changes are highlighted with colored patches. We did four major changes:

1. *Photodiode Capacitance.* The photoreceptor bandwidth dominates the temporal response of the pixel. For a temporal derivative operation, the bandwidth of the pixel limits the range of temporal frequencies over which the pixel can compute the temporal derivative. The photoreceptor bandwidth is constrained primarily by the photodiode capacitance and the feedback conductance. The bandwidth can be maximized by minimizing the photodiode capacitance. We used an n-well to p-substrate photodiode (underlaid with a light blue patch in Figure 2.8) instead of the transistor source/drain n-diffusion to p-substrate diode. This change decreases the time constant of the photodiode by a factor of about $10\times$. The drawback of using a n-well for the photodiode is that more pixel space is used, because the distance between n-wells that are needed for the implementation of p-FET transistors is relatively large photodiode n-well,
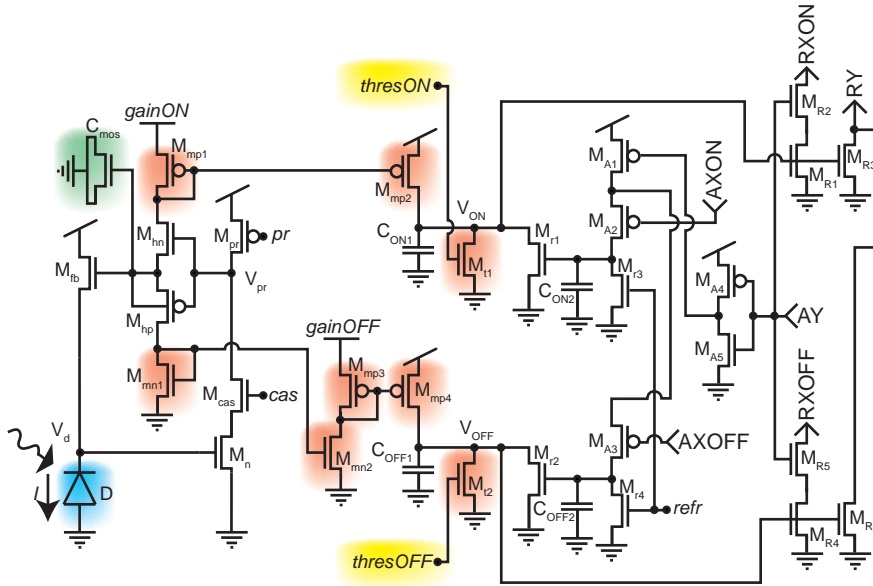
Figure 2.8: **Pixel-circuit design of our** 32×32 **transient vision sensor. The colored patches highlight places of important changes in respect to Jörg Kramer's design.**

2. *Asymmetry of Response.* Jörg Kramer's design, systematic mismatch between the ON and OFF channels made it impossible to simultaneously adjust the threshold parameter adequately for both channels (ON and OFF). Jörg Kramer's design used a common bias *thres* to set the threshold for transient detection that was common to the ON and OFF channel in the pixel (see Figure 2.2. To facilitate a symetric tuning of the ON and OFF thresholds, we have separated the threshold bias *thres* into *treshON* for the ON channel and *treshOFF* for the OFF channel as shown with a yellow highlight in Figure 2.8.

3. *Critical Transistor Sizing.* We have sized the critical transistors that result in generating the row request for the ON and OFF channels to minimize the effect of transistor mismatch given the constraint of limited pixel size. The critical transistors are displayed with a red highlight in Figure 2.8.

4. *Capacitor Sizing.* The pixel's photoreceptor storage capacitor, indicated in Figure 2.8 with a green colored patch, directly scales the ON and OFF transient output currents. This scaling is important because the ON and OFF leakage currents (the currents output during DC operating conditions) are scaled independently from the signal current. To make a larger signal-to-DC ratio, we made the capacitor as large as possible using a combination of poly1/poly2 and channel MOS capacitor capacitance.

2.4.2 Implementation of the Small Caviar Retina

To incorporate all the changes discussed above into a new pixel layout, we slightly enlarged the pixel size to $40\mu m \times 40\mu m$. Keeping the overall area of the chip the just below the minimal supported by the chip manufacturer and therefore keeping the cost minimal, necessitated a decrease of the number of pixels to be arrayed. We chose to assemble a 32×32 pixel array, which also conveniently filled up an encoding depth of 5 bit for the y-address, 5 bit for the x-address and 1 bit for the polarity.
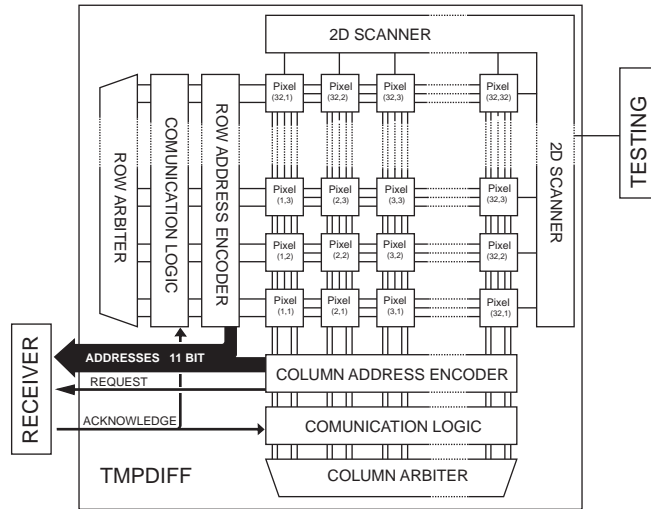

Figure 2.9: **The architecture of the small CAVIAR sensor. The core of this design is a 32×32 pixel array, flanked on two sides by the AER communication structure and on the other two sides by assemblies of shift registers that enable sequential access to pixel nodes for testing.**

Figure 2.8 depicts the architecture of the small CAVIAR retina. Two sides of the 32×32 pixel array are flanked by the AER communication infrastructure that we discussed in 2.3.2. Additionally to the AER output communication infrastructure, we implemented a scanner like structure built from two independent shift-registers. By appropriately clocking the shift registers, the nodes V_{pr}, V_{ON} and V_{OFF} of an arbitrary pixel can be connected to unity gain amplifiers that drive the signals onto output pins for testing purposes.

Figure 2.10 shows the pixel layout in a) and the chip layout in b). Notable is the large geometry of the 8 transistors that are sensitive to mismatch. To achieve as good matching as possible the transistors are placed in close proximity and have all the same orientation. The increased capacitance is place on the right side of the pixel and is implemented as a MOS capacitor. We carefully separated the analog and digital circuits to reduce coupling from the digital AER circuits into the sensitive analog circuits.

Figure 2.10b) depicts the placement of all the different modules in the pixel. Additionally to the scanner, the column and row AER circuits and the pixel-array, we implemented also a bias generator circuit on this chip. The bias generator circuit produces all the different pixel and chip biases according to a
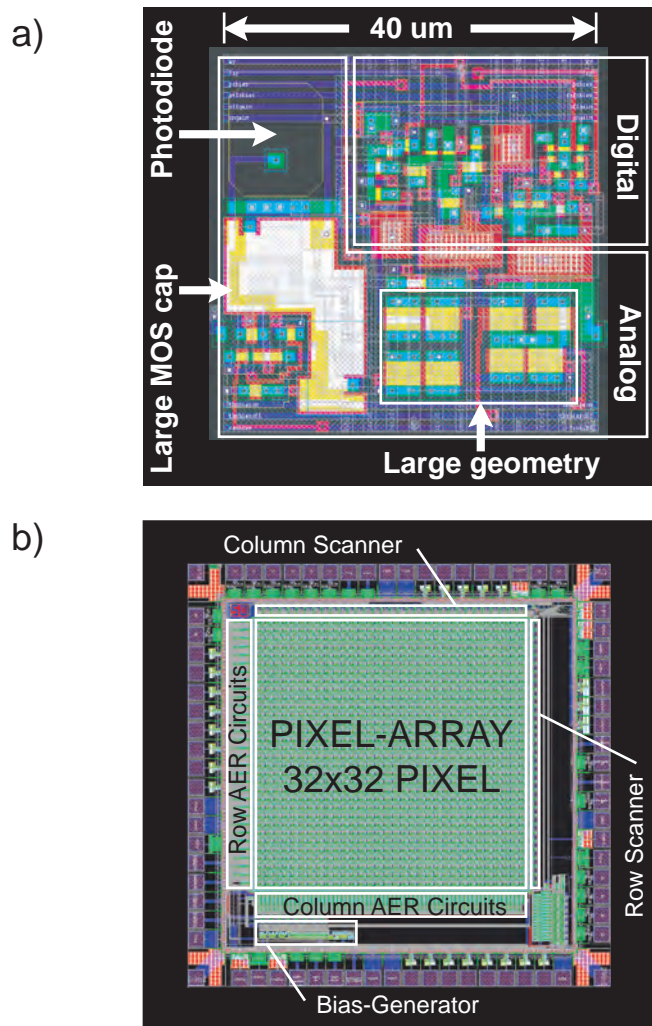
Figure 2.10: **The pixel layout of the small CAVIAR retina is depicted in a). The most important changes compared to Jörg Kramer's design are indicated. The layout of the whole chip is shown in b).**

hardwired program. Because it is very difficult to precisely predict the needed biases with circuit simulation, the pixel and chip biases can be overridden from the outside if not programmed precise enough or if tuning is needed. Furthermore, the flexibility of tuning the biases is very important for experimentation.
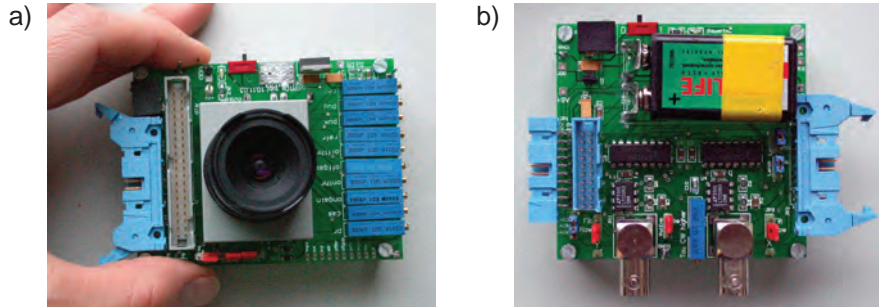

Figure 2.11: **On the left side of the vision sensor PCB, depicted in a), a forty pin and a twenty pin header complying with most AER standards are mounted. A battery-mount is placed on the back side of the PCB.**

The chip was fabricated in a $0.35\mu m$, 4 metal, 2 poly process. We built a custom *Printed Circuit Board* (PCB) that provides the tunable biases and connectors to interface the retina to receivers. This PCB is shown in Figure 2.11a). To make the device more mobile and independent of a lab power supply, we added a connector for battery-powered operation of the vision sensor. Figure 2.11b) shows the DAC PCB built by Tobi Delbrück that we used to display the AER events on the oscilloscope.

2.4.3 Experimental Results

Figure 2.12 shows raster plots of the response of the whole pixel array exposed to an LED modulated with a $8Hz$ square wave. The plots illustrate the clear separation of ON and OFF events to increasing and decreasing illumination changes. With the used bias settings the ON response shows higher variability than the OFF response, which could be due to cross talk. Although the majority of the pixels respond with one event per illumination change, some did not respond and others responded with several events.

Figure 2.13 illustrates the uniformity of the response of the retina to a stimulus of moving bars. We presented black bars on white paper on a rotating drum. The scene was projected onto the sensor plane with a $8mm$ $f/1.2$ lens. The bars were moving diagonally over the visual field of the vision sensor from bottom left to top right. We collected a total of 12'865'536 ON events and 12'288'000 OFF events. This small 4.7% difference in the total number of events, shows that the ON and the OFF channel can be tuned symmetrically with high precision. The coefficient of variation (standard deviation / mean) is 28% for the ON channel and 27% for the OFF channel. There are a few pixels with a very high spiking rate, we will call them hot pixels, but there no pixel that is totally silent. The gray scale images over the histograms show the two-dimensional distribution of the counted events. There is not obvious spatial pattern that can be observed
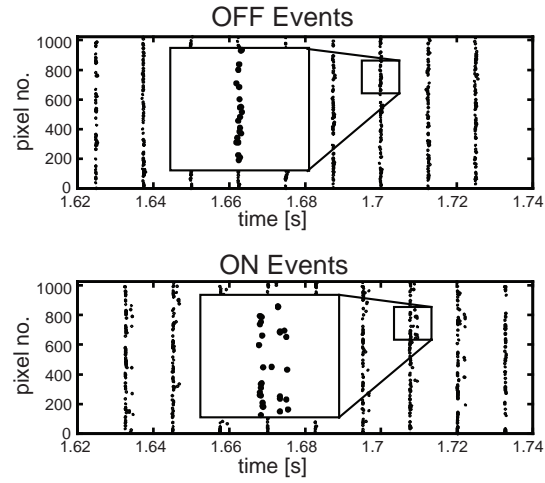
Figure 2.12: **Raster plots of the ON and OFF channels in response to a full chip modulated LED illumination. Each dot represents an event generated by a pixel, numbered on the y-axis. The LED was modulated with an** $8Hz$ **square wave with a contrast of 2:1. The LED directly illuminated the chip; no lens was used. The LED temporal contrast was chosen so that it nominally evoked one event per pixel per illumination change.**
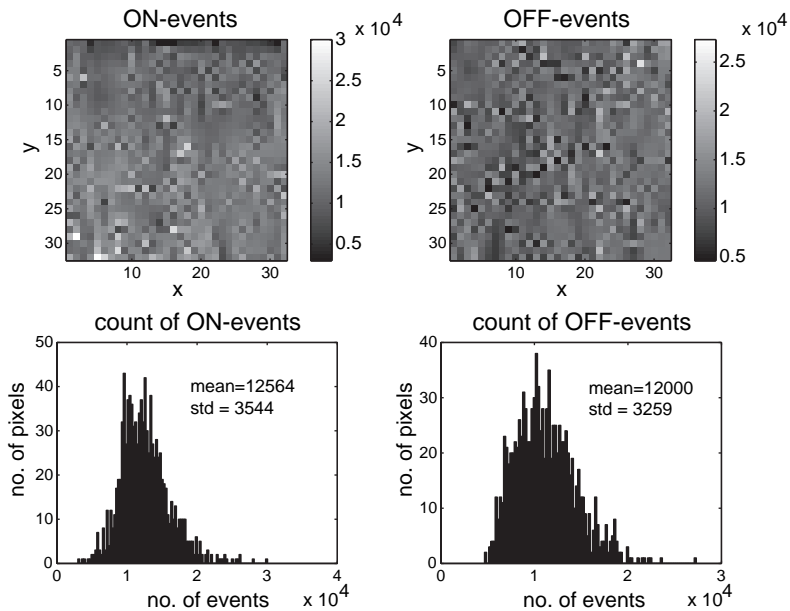

Figure 2.13: **Shows a spatial representation of mismatch in the gray scale images and histograms of the numbers of pixels that produced the same numbers of events of one type.**

but the grey scale images show the places of a few hot pixels. It is notable that the hot pixels do not appear in clusters.
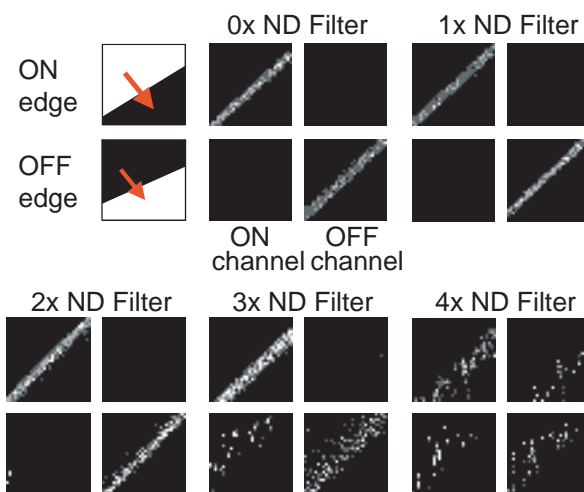

Figure 2.14: **Shows the ON/OFF responses of the small CAVIAR retina to a moving edge for different irradiance levels that where created by the placement of neutral density filters directly over the chip.**

Figure 2.14 shows the responses of the imager to a moving edge. The edge had a contrast of 10/1. To achieve a high irradiance variation between experiments we added a 1 decade attenuation neutral density filter every step. The numbers $0x$ to $4x$ indicate the number of 1 decade attenuation filters; $4x$ means an attenuation of 4 decades. Measured scene illuminance was $350lux$ and 4 decades attenuation lead to a scene illuminance of $0.035lux$.

The grayscale images show snapshots of spikes integrated over 50ms. The responses only begin to be affected by the scene illumination with the placement of the third neutral density filter, mainly in the OFF channel, but are still reliably detected even after the placment of the fourth filter.

At the time we published these results [51], we already doubted that this imager functions below moonlight scene illumination levels ($0.1lux$). And we stated that it was probable that the attenuation of the neutral density filters[5] was below one decade for the incandescent light we used to illuminate our scene. At a later time, when we experimented with the predecessor of this vision sensor, we made the effort of actually measuring the attenuation of the neutral density filters for different light sources and found that the attenuation for incandescent light sources is in the range of a factor of 5-7 instead of 10. For fluorescent light sources, we measured an attenuation of exactly 1 decade. We believe that nevertheless that the small CAVIAR retina has a dynamic range of about $100dB$ because we did not see any degradation of the response for much higher illumination levels than the $350lux$ we have been able to achieve in the lab.

---

[5]Kodak Wratten gelatine filters, No. 96

2.4.4 Conclusion

The ON/OFF transient photosensor developed by Jörg Kramer and evolved by us to the 32×32 small CAVIAR retina is a dense implementation that realizes significant computation and sensing in one feedback loop. We could achieve quite reasonable results for dynamic range (Figure 2.14), a nicely separated and symmetric response to positive and negative irradiance transients and very low spontaneous activity (Figure 2.13 and Figure 2.12). Because the small CAVIAR retina was performing well enough and implemented on a user friendly board, we could use it as an input device for the CAVIAR system and did also distribute it to our CAVIAR partners. We also used the small CAVIAR retina in a orientation tuning experiment done together with Elisabetta Chicca and Giacomo Indiveri (see section xxxx).

On the other hand, we could only achieve reasonable results with a considerable tuning effort. The small CAVIAR retina is very sensitive to changes of a some of the biasing parameters. Environmental changes like temperature changes and illumination condition changes necessitated adjustments. A contrast ration of 1.5 was the lowest irradiance change that the small CAVIAR retina could detect with careful tuning.

The photoreceptor feedback loop with the rectifying differentiating element makes up a third order system that has an inclination to oscillate for adverse bias settings. The small CAVIAR retina forced a continual trade off between sensitivity and stability that reduced its usefulness for experiments drastically.

Parallel to the development of the small CAVIAR retina, we worked out a different pixel design that we hoped would overcome the limitations of the small CAVIAR retina. Because initial results experiments with this new design were promising and because we did not find any satisfying way to get rid of the limitations of the small CAVIAR retina by improvements on the existing design, we shifted our efforts toward the implementation of the new design, which will be described in the next chapter.

Chapter 3

The Self Clocked Switch-Cap Pixel

3.1   A New Pixel Design

Jörg Kramer's ON/OFF transient imager [44] that is based on his beautiful and dense optical transient sensor [45] has some fundamental limitations. The most important limitation is that the ON and the OFF currents produced by the optical transient sensor are very small as shown by equation 2.4. For example, a constant increase in the logarithmic photocurrent, which corresponds to a doubling of the photocurrent every second, leads to an ON current that is $\approx 0.05V \times C_{pr}/s = 0.05pA$ for a reasonably large $1pF$ capacitance. These very small currents are fed through tilted mirrors in the ON/OFF transient imager pixel to obtain enough magnification such that the resulting signal can be thresholded. The mirrors, which are also not symmetric for the ON and the OFF channel, introduce a lot of mismatch at these low current levels, even when they are lain out very carefully. The small ON and OFF currents also put a direct limit to the sensitivity of the sensor because signal currents that are in the range of a few fA are lost in the transistor leakage currents unless very sophisticated methods are applied that are usually area-wise costly. Another important limitation of the design is that the photoreceptor circuit, although comprising in its most basic version [44] only 2 pFET transistors, 3 nFET transistors, a capacitance and a photodiode, is a complicated third order system that has the inclination to oscillate.

Because of these intrinsic limitations we started draw a lot of circuits in parallel to the design of the small CAVIAR retina. We decided that it would be worthwhile to implement our most promising idea in a test chip to try it out. Although the first trial was not totally successful, we were enough confident to switch our design efforts to this new design that is quite different to Jörg Kramer's design. Our hope was to overcome the fundamental limitations of Kramer's pixel in the longer term.

The new pixel design decouples the photo sensing from the differentiating and the rectifying thresholding. The separation makes the phototransduction fast and stable. We kept the logarithmic compression at the front end to achieve a large dynamic range. To achieve a low mismatch we applied a switch-cap principle that relies on precise capacitor matching rather than on relatively painful transistor matching. In the following paragraphs we will present this new pixel design.

3.2 The Principle of Operation

To get an understanding of the pixel, it is useful to first consider the abstracted model shown in Figure 3.1. The pixel consists of a photoreceptor front end that is capacitively coupled to an differencing amplifier. The photocurrent I produced by the photodiode is proportional to the pixel illumination. The photoreceptor responds logarithmically to intensity, thus it encodes relative illumination and therefore contrast. Due to its logarithmic response and the feeback configuration, the photoreceptor circuit has the desirable properties that it automatically controls individual pixel gain while at the same time responding quickly to changes in illumination. On the other hand it has the drawback that transistor threshold variation causes substantial DC mismatch between pixels, necessitating calibration when this output is used directly [52, 53].
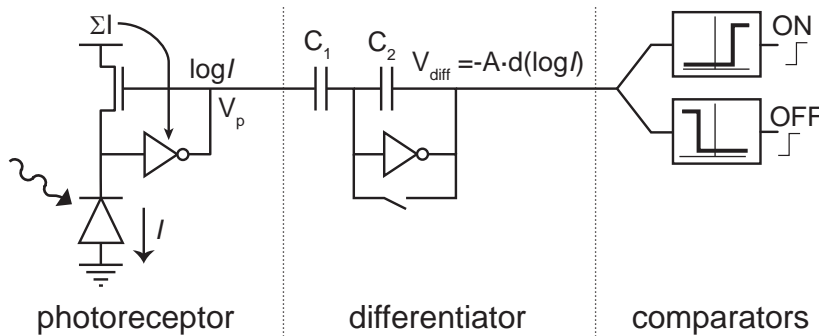

Figure 3.1: **Depicts the abstract pixel schematic. The inverters are symbols for single-ended inverting amplifiers.**

In our pixel the differencing circuit that follows removes all DC mismatch because it amplifies only the changes in the photoreceptor output. In other words, our pixel is sensitive to temporal contrast $C_{temporal}$ that we define with the following equation as:

$$C_{temporal} = \frac{1}{I(t)} \frac{dI(t)}{dt} = \frac{d(lnI(t))}{dt} \tag{3.1}$$
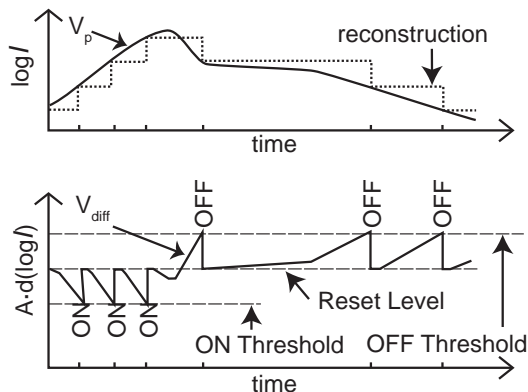
where I is the photocurrent.

Figure 3.2: **The principle of operation of the pixel.**

Figure 3.2 depicts the photoreceptor output voltage V_p and the capacitive feedback amplifier output voltage V_{diff} as a function of time in response to an arbitrary illumination waveform. The photoreceptor output voltage V_p is proportional to the logarithm of the light intensity at the photodiode. The voltage V_{diff} is given by

$$\Delta V_{diff} = \frac{C_1}{C_2} \cdot \Delta V_p \equiv A \Delta V_p \qquad (3.2)$$

where A is the closed-loop gain of the differencing circuit. If the voltage V_{diff} exceeds either the ON or the OFF threshold, the ON or the OFF comparator switches, and an ON or OFF event is generated. Ensuing the communication of this event, the inverting amplifier is balanced to a reset level by closing the reset switch. The reset switch is held closed for a brief, adjustable refractory period. The pixel circuit is quantizing voltage changes at the photoreceptor output in continuous time. Taking all the history into account, a reconstruction of the intensity waveform can be performed, as shown in Figure 3.2. The effect of inevitable comparator mismatch is greatly reduced by the gain A of the differencing circuit, in section 6.4.1 on page 79.

The differencing circuit is not a differentiator. Its output is not proportional to the instantaneous derivative of the log intensity but instead is proportional to the total amount of the change in relative intensity since the last reset. For smooth variations of the input waveform however, the time between quantization steps is inversely proportional to the temporal derivative of the illumination waveform, or in other words the quantization frequency represents the temporal derivative. In the next few paragraphs, we will consider in detail the operation of these component parts of the pixel circuit.

3.3 The Detailed Pixel Circuit

During the last 2 years we implement three different chips. During this time, the pixel circuit did change only slightly. To avoid redundancy we will discuss

here directly the most evolved circuits that have been implemented in our latest temporal contrast vision sensor that is presented in chapter xxxx.
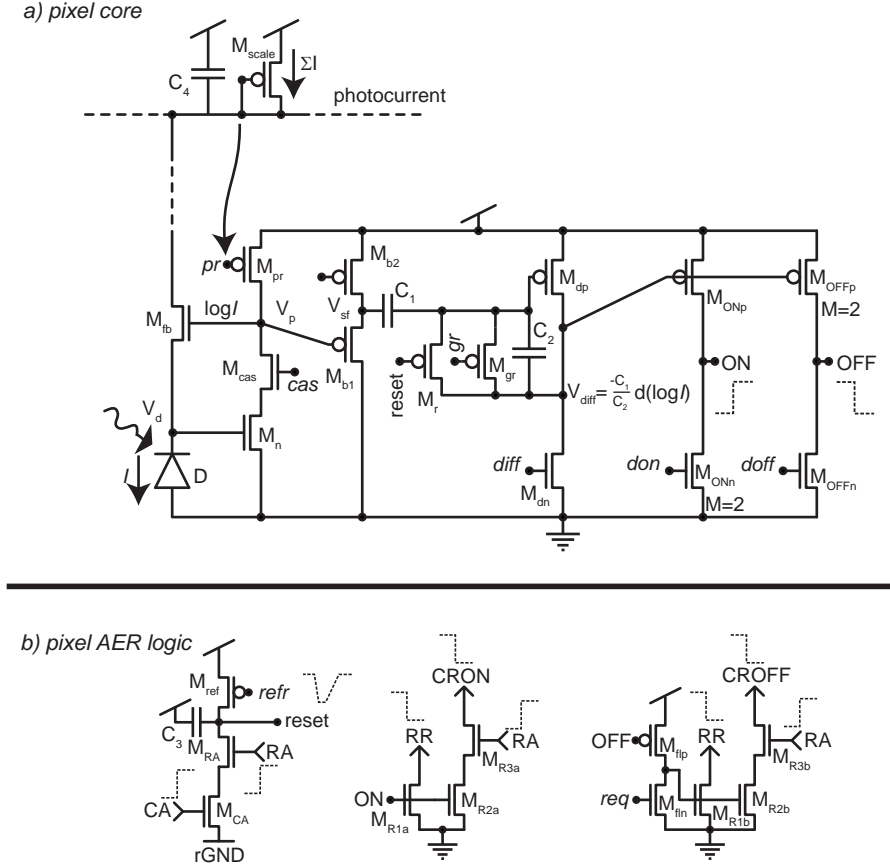

Figure 3.3: **Shows the detailed Pixel Circuit. The circuit is built using 27 transistors, 3 capacitors and one photodiode.**

Figure 3.3 depicts the detailed pixel circuit. The pixel circuit can be functionally divided into the pixel core (Figure 3.3a) applied to the analog computation and the AER circuits dealing with the pixel to periphery AER communication(Figure 3.3b).

The front end photoreceptor circuit comprises a photodiode D whose photocurrent is sourced by a saturated NMOS transistor M_{fb}. The gate of M_{fb} is connected to the output of an inverting amplifier that is built with the transistors M_{pr}, M_{cas} and M_n and whose input is connected to the photodiode. This well known transimpedance configuration (see e.g [42]) converts the photocurrent logarithmically into a voltage while it holds the photodiode clamped at a virtual ground. The bandwidth of the photoreceptor is extended by the factor of the loop gain in contrast to a passive logarithmic photoreceptor circuit. This extended bandwidth is valuable for detecting fast transients, especially in lower lighting conditions. In addition, this photoreceptor circuit includes the option of adaptive biasing, by which power consumption can be reduced and a constant

resonance of the photoreceptor can be maintained. Adaptive biasing is realized by using a fraction of the lowpass-filtered sum of the photocurrents of all pixels to directly generate the bias voltage for M_{pr} [54]. The implementation of this adaptive biasing that we used is shown in Figure 3.3 on top and consists of a series of parallel pFET transistors, here represented by M_{scale} and a capacitance $C4$. These circuits do not belong to the pixel but are shared by all pixels.

The photoreceptor output is buffered with a source follower to the voltage V_{sf} to isolate the sensitive photoreceptor from the rapid transients in the differencing circuit, and to drive the capacitive input of the differencing circuit. The floating node is the input to a capacitive-feedback inverting amplifier. The amplifier is balanced with a reset switch that shorts its input and output together, resulting in a reset voltage level. Using 3.10 and integrating 3.1 produces a direct relation between temporal contrast $C_{temporal}$ and the voltage V_{diff}

$$\begin{aligned}
\Delta V_{diff} &= -A \cdot \Delta V_{sf} = -A \cdot \kappa_{sf} \Delta V_p & (3.3) \\
&= -A \frac{U_T \kappa_{sf}}{\kappa_{fb}} ln \left(\frac{I(t+\Delta t)}{I(t)} \right) & (3.4) \\
&= -A \frac{U_T \kappa_{sf}}{\kappa_{fb}} \int_t^{t+\Delta t} C_{temporal}(y) dy & (3.5) \\
&\equiv -A \frac{U_T \kappa_{sf}}{\kappa_{fb}} \bar{C}_{temporal} \Delta t & (3.6)
\end{aligned}$$

where U_T is the thermal voltage and κ_z is the subthreshold slope factor of transistor M_z. The comparators (M_{ONn}, M_{ONp}, M_{OFFn}, M_{OFFp}) weigh the output of the inverting amplifier against global thresholds. These global thresholds are offset from the reset voltage to detect increasing and decreasing changes. An ON or OFF event is generated if the input of a comparator overcomes its threshold . The OFF comparator output is inverted by a starved inverter comprising the transistors M_{fln} and M_{flp}. The static bias $ioff$ reduces the power consumption and noise-producing dI/dt. Replacing the voltage V_{diff} in 3.3 by the comparator input thresholds and solving for the temporal contrast $C_{temporal}$ yields the minimal positive temporal contrast that triggers an ON event

$$\begin{aligned}
C_{threshold} &= (\Delta C_{temporal})_{min} = \Delta \log(I)_{min} \\
&= \frac{\kappa_{fb}}{A U_T \kappa_{sf} \kappa_{dp}} \left(\kappa_{dn}(\text{don} - \text{diff}) + U_T \log(2) \right) & (3.7)
\end{aligned}$$

and the minimal negative temporal contrast $C_{temporal}$ that elicits an OFF event

$$\begin{aligned}
C_{threshold} &= (\Delta C_{temporal})_{min} = \Delta \log(I)_{min} \\
&= \frac{\kappa_{fb}}{A U_T \kappa_{sf} \kappa_{dp}} \left(\kappa_{dn}(\text{doff} - \text{diff}) - U_T \log(2) \right) & (3.8)
\end{aligned}$$

where *don-diff* is the ON threshold and *doff-diff* is the OFF threshold. To simplify the equations we set the subthreshold factors of M_{ONn} and M_{OFFn} equal to M_{dn} and the same for the corresponding p-FETs. This simplification is justified by the fact that all the nFET/pFET transistors are of equal size and orientation and they are also place in close proximity. For large thresholds the voltage *don* can be high, such that the transistors M_{ONn} and M_{ONp} are operated in the above threshold domain. A derivation for the minimal positive temporal

contrast that triggers on ON event when M_{ONn} and M_{ONp} are operated in above threshold generates

$$
\begin{aligned}
C_{threshold} &= (\Delta C_{temporal})_{min} = \Delta \log(I)_{min} \\
&= \frac{\kappa_{fb}}{AU_T \kappa_{sf} \kappa dp} \left(\frac{1}{\kappa_{OFFp}} \sqrt{\frac{2\beta_{ONn}}{\beta_{OFFp}}} (\text{don} - V_{T,n}) \right. \\
&\quad \left. + V_{T,p} \kappa_{dn} \text{diff} - \log\left(\frac{I_{0,n}}{I_{0,p}}\right) U_T \right).
\end{aligned}
\tag{3.9}
$$

These equations take into account that M_{ONn} is 2 M_{dn} and M_{OFFp} is 2 M_{dp}. The minimal temporal contrast $C_{threshold}$ is dimensionless and is hereafter called contrast threshold, standing for the contrast sensitivity threshold of the pixel for a given comparator threshold. For smoothly-varying temporal contrasts, the rate of generated ON or OFF events can be approximated with

$$
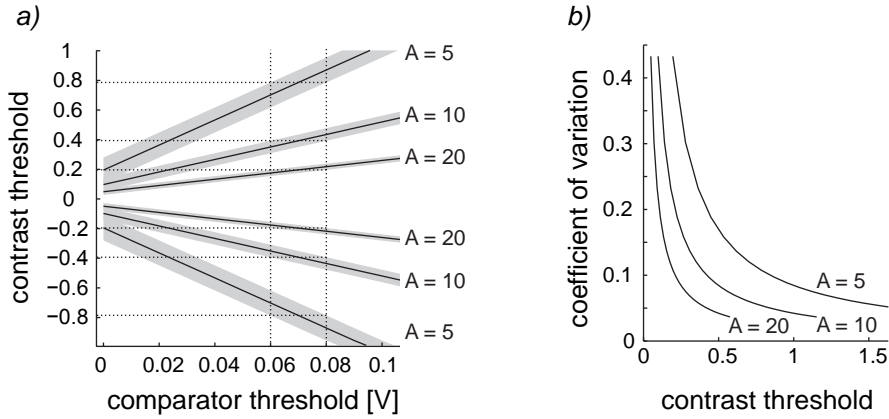f(t) = \text{EventRate}(t) \approx \frac{C_{temporal}(t)}{C_{threshold}}.
\tag{3.10}
$$

Figure 3.4: **Part a) shows the theoretical pixel contrast threshold as a function of the comparator threshold for 3 different capacitor ratios (5, 10 and 20). The gray area depicts an example 10 mV comparator mismatch. Part b) shows the predicted covariance as a function of contrast threshold for the 3 different capacitor ratios. The plots highlight the significance of the pixel gain.**

Equations 3.7 to 3.9 exemplify the benefit of this pixel design: The unavoidable comparator mismatch is scaled relative to the input with the differencing circuit gain A. This important principle is illustrated in Figure 3.4. The relations 3.7 to 3.9 are later characterized experimentally in Sec XXX.

The ON and OFF events are communicated to the periphery by the circuits in Figure 3.3b) that implement the 4-phase AE handshaking with the peripheral AE circuits. In the jargon of AER, we use *arbitrated word-parallel non-greedy* AER circuits. *Arbitrated* means that pixel events are queued and wait their

turn. *Word parallel* means that our x,y-address is communicated in parallel rather than multiplexed by row address followed by bursts of column addresses, and *non-greedy* means that the arbitration ensures that a row or column that is serviced by the arbiter is guaranteed not to be serviced again before all other rows and column that have registered requests have been serviced. Although the circuits used for the AER communication vary slightly compared to the circuits described insection 2.3.2, the AER communication scheme applied stays the same and is not repeated here.

Worthwhile to mention is that while row and column ON and OFF request signals (RR, $CRON$, $CROFF$) are generated individually, the acknowledge signals (RA, CA) are shared. They can be shared because the pixel makes either an ON or OFF event, never both simultaneously and because there exists only one reset in the pixel for both polarities. The column acknowledge signal is CA generated by a logic or function of the signals $CAON$ and $CAOFF$ at the border of the array before the CA enters the array. Using only one column acknowledge signal for both polarities saves a wire which is beneficial to save space and to reduce noise.

The reset switch M_r executes a reset of the pixel circuit by balancing the differentiator. Along with the reset an adjustable refractory period, during which the pixel cannot generate another event, is implemented by means of the starved NAND gate consisting of M_{ref}, M_{RA} and M_{CA} depicted in Figure 3.3b). This refractory period limits the maximum firing rate of individual pixels and can help prevent small groups of pixels from taking over all bus capacity under certain circumstances.

The transistor M_{gr}, which is only implemented in the 128×128 temporal contrast vision sensor, is an additional reset switch that can be set on an arbitrarily-selected set of columns, including the entire chip if desired. It operate by holding the selected pixels in reset, preventing them from accessing the bus so that the remaining pixels can use the full bus capacity.

In the first implementation of the self-clocked switch cap pixel, we experienced some problems with charge injection from the reset switch M_r onto the floating node. Charge injection is a phenomena that is well known in switch-cap design but is still very hard to quantitatively grasp. The usual ways of dealing with charge injection obstacles is to use a full transmission gate consisting of a pFET and a nFET that are switched in opposite directions. The two FETs have to be carefully sized so that the opposed charges injected by both of them precisely cancel out. This technique was no option in our situation because of the difficulties to create a precise counter signal in the limited space available in the pixel. We dealt with the charge injection by using a programmable low overhead switch drive at $rGND$.

Junction leakage in M_r and M_{gr} causes a low rate of ON events that can be eliminated by slightly turning on M_{gr}, at the cost of increased corner frequency for the differentiator.

3.4   The Small Signal Pixel Model

Because pixel dynamics are important, particularly with regard to how they change with illumination, we now consider a small signal model of the front end pixel circuit.

At low illumination the photoreceptor bandwidth is proportional to photocurrent [42, 49], because it is determined by the RC time constant formed by the photodiode parasitic capacitance and the resistance of the feedback transistor M_{fb} source, as seen from the photodiode. Hence at low illumination levels, pixel bandwidth will be limited by available photocurrent.

The conductance into the source of M_{fb} is increased by a factor of the loop gain around the loop formed by the amplifier M_n and M_{pr} because the photodiode is held at a virtual ground by the feedback loop. The cascode M_{cas} increases the bandwidth further by removing the effect of the parasitic capacitance between the drain and gate of M_n, that is called the Miller capacitance. These effects are modeled by the following idealized transfer function of the photoreceptor followed by the source follower [42]

$$H(s) = \frac{\frac{v_p}{U_T}}{\log(I_{DC} + i)} = \frac{1}{\frac{(\tau_{in}s+1)(\tau_{out}s+1)}{A_{pr}}} \frac{1}{\tau_{foll}s + 1} \qquad (3.11)$$

where v_p stands for the small signal photoreceptor output voltage, $I_{DC} + i$ denotes the DC photocurrent plus the small signal variation, τ_{in} is the time constant of the photodiode node V_d, τout is the time constant of the photoreceptor output node V_p, τ_{foll} is the time constant of the source follower output node V_{sf}, and A_{pr} is the DC open loop gain of the feed-forward amplifier formed by M_n, M_{cas}, and M_{pr}. By time constant, we mean the open loop C/g time constant of the node. Equation refeq:photoTransFunc characterizes a 2nd order resonant filter followed by a 1st order lowpass filter. For an infinitesimally fast feed-forward amplifier ($\tau_{out} = 0$) and source follower buffer ($\tau_{foll} = 0$), equation refeq:photoTransFunc simplifies to a lowpass filter, which has a time constant that is approximately equal to τ_{in}/A_{pr}.

Intuitively, the feedback raises the conductance looking into the source of M_{fb} by a factor of the loop gain. Because the photoreceptor is arranged in unity-gain configuration–the gain from gate to source of M_{fb} is about 1–the loop gain is approximately A_{pr}. Including in the computation the dominant remaining Miller capacitance C_m that exists between the gate and source of M_{fb} (not shown in Figure 3.3), we obtain the idealized minimum closed loop time constant of the photoreceptor

$$\tau_{min} = \frac{C_{eff}}{g} = \frac{C_d/A + C_m}{I_{DC}/U_T} \qquad (3.12)$$

where C_d is the parasitic photodiode capacitance, C_m is the Miller capacitance from the gate to source of M_{fb}, I_{DC} is the DC photocurrent, and A is the gain of the feedforward amplifier. In the case of a large A_{pr}, the time constant is dominated by the remaining Miller capacitance. The bandwidth, defined as the $3dB$ roll-off frequency, is given by $f_{3dB} = 1/2\pi\tau_{min}$.

In section XXX, we experimentally characterize pixel bandwidth and in section XXX we characterize pixel latency.

3.5 Précis and Hors d'Oeuvre

The pixel design shown in Figure 3.3 is the heart and principal achievement of this dissertation. It comprises 16 nFETs, 11 pFETs and 3 capacitors. The pixel

senses impinging irradiance with a photodiode and quantizes relative irradiance changes by communicating address events. The quantization happens by a novel self-clocked switch-cap circuit. The novel pixel circuit surpasses fundamental limitations of previous designs.

An event that is communicated from the self-clocked switch-cap pixel means that the logarithmic illumination changed by an accurate and predictable amount since the last time this pixel emitted an event. The quantization step, which is congruent to the pixel's resolution in relative illumination change, is adjustable by two independent thresholds for positive and negative illumination variations.

By carefully decoupling logarithmic phototransduction and analog to digital conversion stability problems are minimized and at the same time pixel bandwidth is extended to both the upper and the lower end. Due to the decoupling, the quantization computation is no longer corrupted by the overall lighting conditions, leading to a wide dynamic range and a constant response. Finally, the behavior of the decoupled structure is much better predictable allowing more reliable calculations and numerical simulations and ultimately lead to a much deeper understanding of the pixel. Of course, a better understanding promotes a better implementation of the pixel and allows to use the parametric space–given by 8 adjustable pixel biases–strategically[1].

In circuit design, theory is nice but the ultimate judgment of a design is made after it is implemented into a physical device. We sequentially implemented the novel pixel design in three chips:

1. *The Large CAVIAR Retina.* In the first implementation we arrayed 64×64 self-clocked switch-cap pixels. The resulting vision sensor was used as the front end input to the CAVIAR system and is described in chapter xxxx.

2. *The Double Line Sensor.* In an attempt to create a high speed capable device, we implemented a double line sensor that comprises two lines of each 64 pixels.

3. *The AER Temporal Contrast Vision Sensor.* The latest implementation and prime achievement of our efforts is a 128×128 pixel sensor that we now call the AER *Temporal Contrast Vision* (TCV) sensor.

In the following three chapters we will present these implementations.

---

[1]The parametric space given by biases often caused headaches, sore fingers and maybe even nightmares before a lot of twiddling and fiddling lead to a so called magic setup. From this point on, the presence of the magician was of prime importance for the operation of this magic setup, leading to a reluctance of normal people to even come close to such a magic device and sometimes cold fear when they were invited to use it.

Chapter 4

The Large CAVIAR Retina

4.1   The Implementation

The 64×64 pixels temporal contrast vision sensor was our first implementation of the novel self-clocked switch-cap pixel. Numerical simulations, calculations and results from similar test-pixels were encouraging but ultimately the only way to find out if a circuit design really works according to the expectations is to implement it.

The CAVIAR project structure set a time line for the implementation and fabrication for this chip and ultimately, this chip is the front end vision sensor of the CAVIAR project. Hence we call this second vision sensor chip for the CAVIAR project the large CAVIAR retina.

In the following sections we will describe the chip architecture, the pixel, the fabricated chip and the integration of the chip into a small self-contained vision sensor system on a PCB.

4.1.1   The Chip Architecture

Figure 4.1 depicts a schematic drawing of the large CAVIAR retina. The core of the chip comprises the pixel-array, the AER communication infrastructure and a bias generator circuit block. At the periphery of the pixel-array, we placed a standalone pixel with direct measurement access on different nodes among some other test-structures. The chip AER output is a 13 bit digital address that has the 6 bit x and y addresses and a polarity bit. The communication uses Kwabena Boahen's peripheral AER circuits [23] to losslessly communicate all events in a word parallel fashion. The arbitration scheme employed is based on so called greedy arbiters, which preferably serve the same input again. The advantage of this scheme is a considerable speed up compared to older non
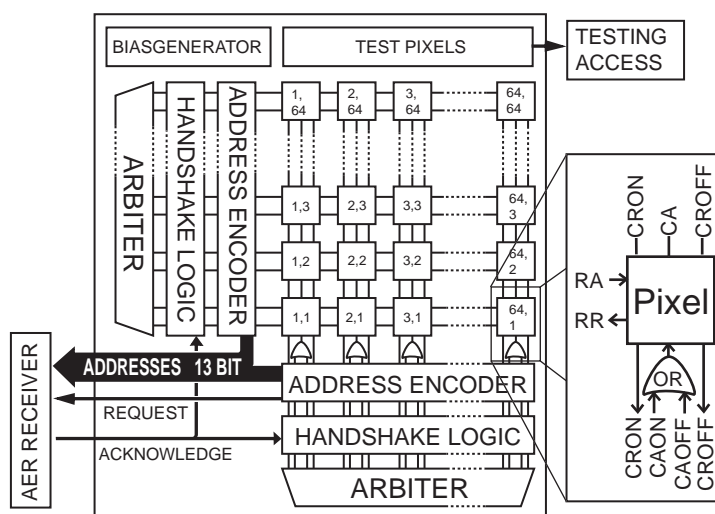
Figure 4.1: **Block diagram depicting the sensor architecture.**

greedy versions, the drawback is that rows an columns are served unequally when the communication channels start to saturate. For a thorough description of the AER handshaking cycles refer to section 2.3.2.

There are two different principles to reliably communicate address: The single-sender protocol that is intended for the direct connection of sender with a receiver and the multi-sender protocol applied in situations where the data bus is shared among more than two independent AER chips. In the single-sender situation, the data generally has to be valid for read before the request signal is activated and then it has to stay valid until the receiver sets the acknowledge signal. In multi-sender circumstances, only a sender that set up a communication link, meaning that it got an acknowledge from a receiver, is allowed to drive the bus, otherwise conflicting situations are likely to arise. This means that the data is set valid only after the reception of the acknowledge signal and is then kept valid until the acknowledge is removed.

In all our chips that implement the self-clocked switch-cap pixel both protocols are available and can be chosen by a jumper setting. This is made possible by tri-state logic in the pads, in which the output circuit can be disconnected from the rest of the circuit, putting the output in a high impedance state. By selecting the request signal for the enable of the tri-state output, a single-sender protocol is chosen and if the enable signal is connected to the acknowledge signal, a multi-sender communication is set up.

4.1.2   The Layout and the Silicon

In analog circuit design a careful layout is indispensable for proper circuit operation. In circumstances, in which a single element is arrayed, like in a vision sensor, the precision and care applied in the single element layout pays off even more. The size of a vision sensor pixel ultimately determines the size of a vision sensor chip of a certain resolution and therefore the prize of a sensor is often
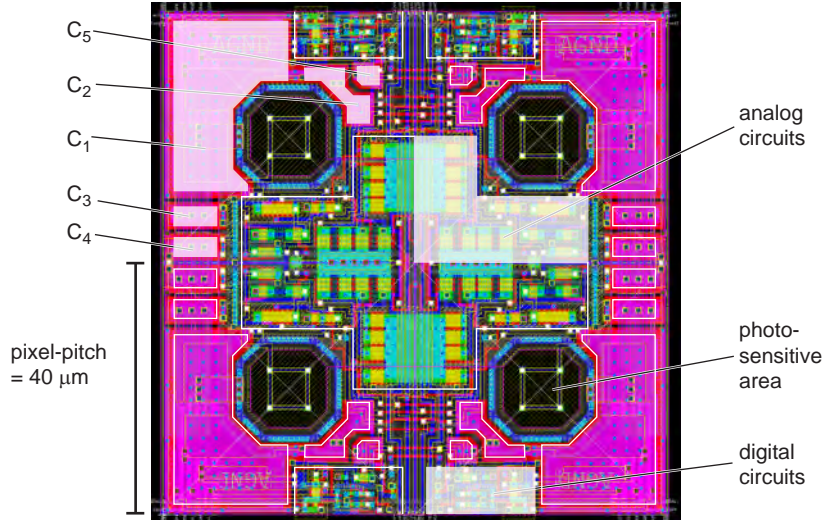
proportional to the size of the pixel.

Figure 4.2: **Layout of a block of 4 pixels.**

Figure 4.2 shows the layout of 4 pixels of the large CAVIAR retina. By arranging 4 neighboring pixels in a point symmetric fashion around their center, we gain the following advantages:

1. *Reduced Size.* Due to the mirroring, the biasing wires of two adjacent pixels are adjoined and can easily be shared. The sharing of the bias wires cuts in halve the number of wires needed, which results in a significant reduction of the pixel size.

2. *Reduced Noise.* In the chosen arrangement, analog parts of a pixel lie proximal to the analog parts of its neighbors and the very same is true for the digital parts and also capacitors. This effect allows an efficient shielding of analog circuits from noise produced by the digital circuits.

The nFET transistors M_{ONn}, M_{OFFn} and M_{dn} (please refer to the pixel schematic in Figure 3.3) are drawn in close proximity the same orientation and in a relatively large size to obtain the best possible matching, the same is true for the pFETs M_{ONp}, M_{OFFp} and M_{dp}. Most of the pixel area is used up by the capacitor C_1. The whole pixel array is covered by the topmost metal layer with openings only above the photo-sensing area to avoid any unfavorable carrier generation in the proximity of the transistors.

The chip was fabricated by *austriamicrosystems* AG[1] through the *Multi Project Wafer* (MPW) IC service provided by Europractice[2]. The silicon area used by large CAVIAR retina amounted to $3.3mm \times 3.4mm$. Figure 4.3 shows three photographs of the large CAVIAR retina chip taken through a microscope. The left image shows the whole array and the inlets each show a close up view of a few pixels. The lower inlet is a so called darkfield image that is made by
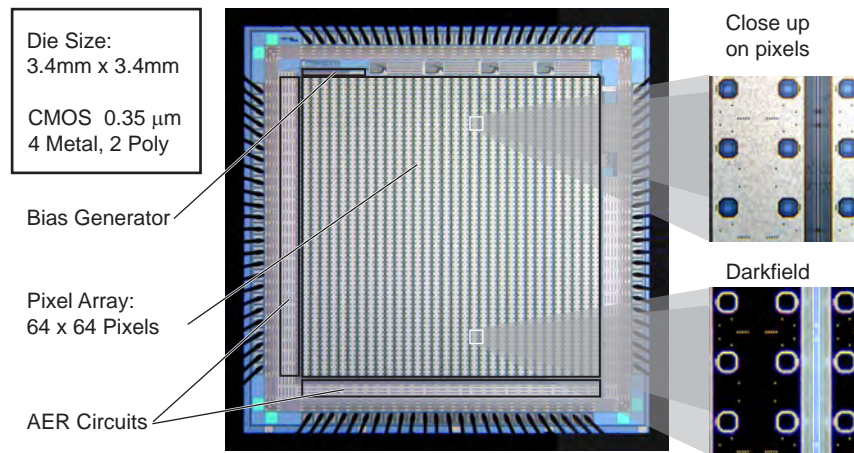
---

[1]http://www.austriamicrosystems.com
[2]http://www.europractice.imec.be

Figure 4.3: **Shows a die micro photograph of the** 64×64 **sensor with close up of the pixels. The lower close up of the pixel is taken with illumination from the side.**

illuminating the object under inspection from the rather than from the top, emphasizing the third dimension. Both inlets show the careful metal cover that is only open above the photodiode.
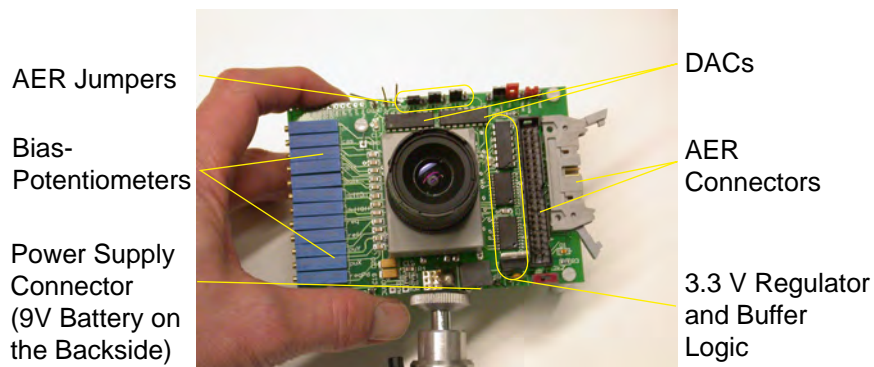
4.1.3   The PCB

Figure 4.4: **PCB.**

For a convenient use of the chip we built a camera system on a PCB. Figure 4.4 shows this system mounted on a tripod. The board contains a 3.3 V regulator that can be driven either by an external power supply or a 9 V battery that is mounted on the backside of the board and not visible on Figure 4.4. The AER output is available on a 40 pin and a 20 pin header making the system compatible to many AER receivers. A set of three jumpers on one side of the board allows to select between the single-sender and the multi-sender protocol,

to shortcut the request signal directly to the acknowledge signal and also to select a division for incoming 5 V acknowledge signals. The potentiometers on the board are used to adjust the biases. The board also is equipped with DACs that enable an on board translation of the digital AER addresses into analog x and y voltages to be used with an oscilloscope in the X versus Y mode the very same way shown in Figure 2.7.

We used this camera system to experiment with the large CAVIAR retina. In the next section we will describe a few of these experiments.

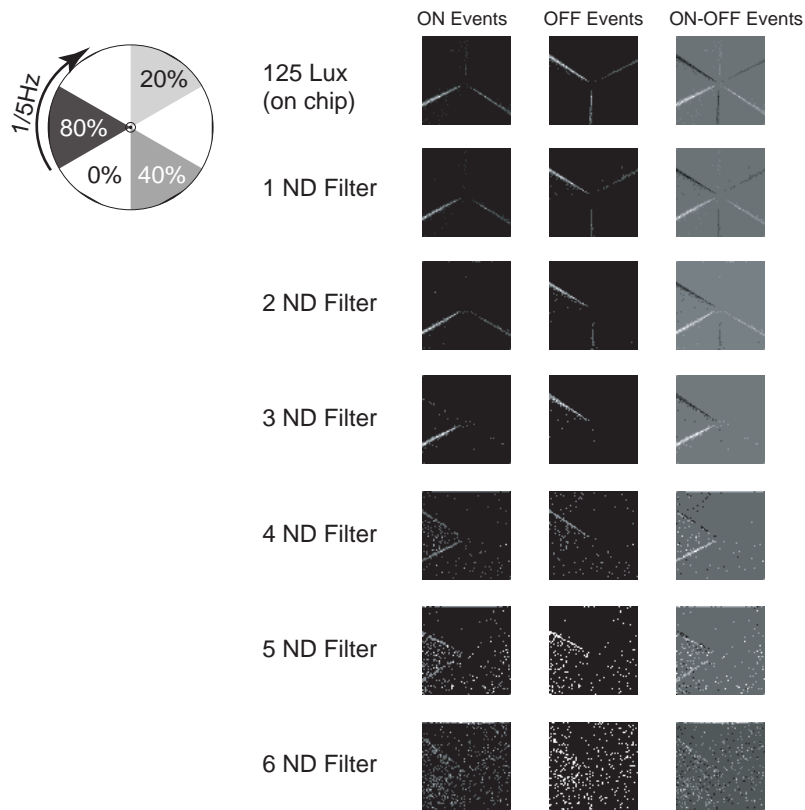## 4.2 The Experimental Results

Figure 4.5: **Shows gray-scale images of** $20ms$ **integrated AER retina output observing a rotating fan stimulus that was illuminated by a DC-powered halogen** $20W$ $12V$ **lamp. The disk was irradiated with approximately** $35W/m^2$**. The luminance of the white part was 1 knit=1000** cd/m^2**. The lens had aperture** $f/1.4$**. Full gray-scale is 6 events. For each row starting from the top, we added another decade of Kodak Wratten neutral-density filtering. This figure is adapted from [55].**

We characterized the large CAVIAR retina's responses and illumination requirements using the rotating fan-like stimulus shown at the upper left of Figure 4.5. We used this wedge stimulus to test usable contrast and speed response.

Features at the outer edge move faster, and the 6 wedge borders have different contrast. The three segments have a black density of 80%, 40% and 20% respectively. The lightest gray segment has a contrast of 0.8:1 relative to the white background. The large CAVIAR retina produces meaningful output over at least 5 decades of operating range and in response to scenes with only 20% contrast.
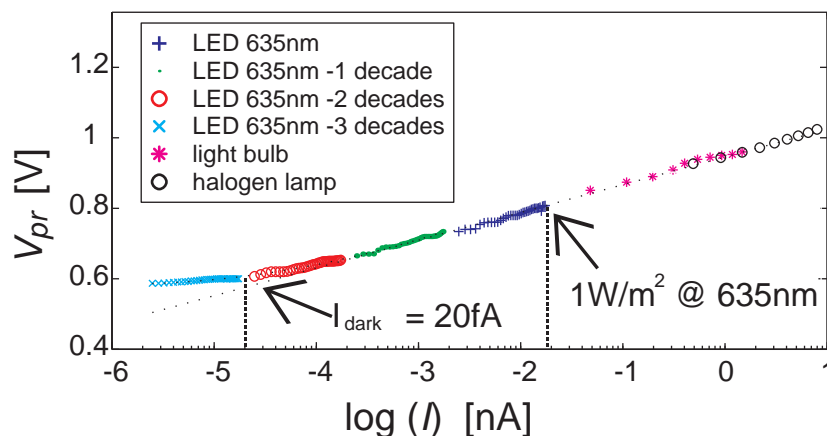

Figure 4.6: **The DC response of the front-end logarithmic photoreceptor circuit. The pixel photocurrent was computed directly from the sum of the entire array photocurrents to calibrate different illumination sources. The response is logarithmic over at least 5 decades.**

Although the fill factor (8.1%) of the retina pixels is small, the photodiode area ($130\mu u^2$) is large. The photodiode dark current can be inferred from Figure 4.6, which shows the DC response characteristics of the logarithmic photoreceptor. The measured dark current level is consistent with the vendor process specifications and is more less equaly split between area and perimeter leakage.

Events that we recorded in a public hallway are shown in Figure 4.7. The large CAVIAR retina generated about 200 events per $25ms$ time bin. This represents a data rate of 5-10% of an equivalent resolution gray-scale imager. This example shows that only objects that move relative to the retina produce output. This feature is potentially interesting for surveillance applications.

To explore high-speed imaging capabilities, we used the large CAVIAR retina to record motion of a drosophila fruit fly wing. The wing was attached to a tuning fork that was tuned to oscillate at the fly's natural wing beat frequency of about $230Hz$. The fruit fly wing glued to the metal rod is shown in Figure 4.8a). A sampled imager would require a frame rate of several times $230Hz$ to avoid substantial aliasing. Such high-speed imagers exist but produce huge amounts of data (see section 1.2.6). Figure 4.8b) shows the 96 events generated during a time slice of $310\mu s$. Figure 4.8c) displays the Y-addresses of the ON and OFF events triggered during two cycles of the wing beat. The red line is a running average of all the Y-addresses. This simple measure seems sufficient to reliable extract the wing beat frequency, phase, and amplitude.
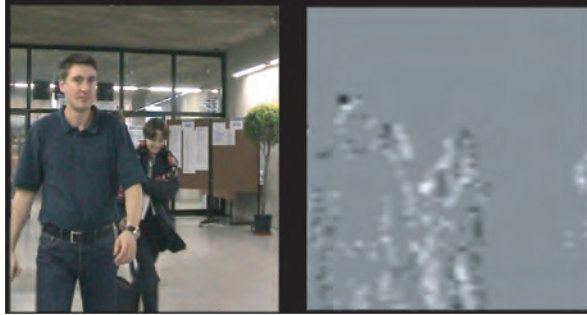
Figure 4.7: **Shows example natural data taken in a public hallway. The woman on the right is moving to the right and the man on the left is moving to the left. Both of them are darker than the background so the leading edges of each person tend to produce OFF spikes, while the trailing edges tend to produce ON spikes. The right gray-scale image shows the integrated spikes during 25ms. The left camcorder image was not aligned perfectly with the retina chip image, therefore, the head of a third person, who is walking behind the other two toward the left is only visible in the retina image. This figure is adapted from [55].**

4.3 Discussion

The large size CAVIAR retina, which is the first vision sensor that implements the self-clocked switch-cap pixel, responds with ON and OFF events to relative changes in irradiance. This asynchronous vision sensor integrates a 64×64 pixels that communicate with the AER protocol. The AER interface consists of a data bus with $X + Y + \text{Polarity} = 6 + 6 + 1 = 13$ bits of address, a request and an acknowledge signal. The request and the acknowledge signal implement a 4 phase handshake that can be operated in a single-sender or multi-sender mode. Table 4.1 lists all the specifications of the large CAVIAR retina.

The new pixel design proofed to be superior to the old one in almost all respects we care about. The large CAVIAR retina is operating over a dynamic

Pixel size	$40 \times 40 \ \mu m^2$
Pixel complexity	28 transistors (16 analog), 3 caps
Array size	64×64
Chip size	$3.3 \times 3.3 \ mm^2$
Chip supply current	$2mA$ at $3.3V$
Fabrication process	4M 2P $0.35\mu m$
Biasing requirements	12 generated gate biases
Fill factor	8.1% (PD area $130\mu m^2$)
Min. scene irradiance ($f/1.4$)	Approx. $1mW/m^2$ or $1lux$.
Photodiode dark current	(estimated) 20fA ($\approx 10nA/cm^2$)
Burst event rate	$\approx 4Mevent/sec$

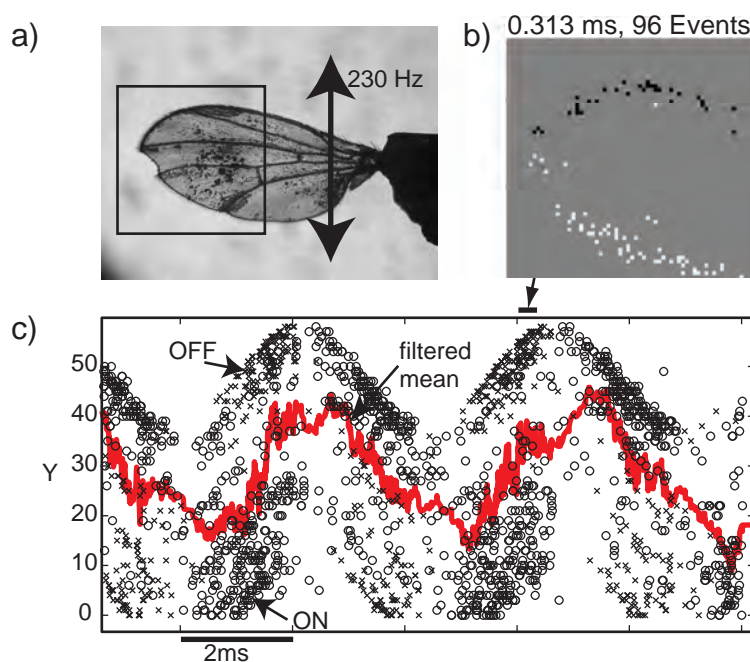Table 4.1: Specifications of the large CAVIAR retina

Figure 4.8: **Shows an experiment with an oscillating wing. Part a) shows the wing mounted on vibrating rod as viewed through a microscope. The oscillation frequency of** $230Hz$ **was measured independently using a microphone recording. The large CAVIAR retina was mounted on the microscope camera mount and imaged approximately the area indicated by the superimposed rectangle. Part b) shows a** $313\mu s$ **of events. On average there were about 50 events per time bin corresponding to a mean rate of 150k events/second. Part c) shows plots of the Y event addresses. Superimposed on the Y addresses is a running average that represents a measure of the vertical location of the wing. This figure is adapted from [55].**

range of about $120dB$ and its response is much more constant over this dynamic range than the response of the small CAVIAR retina. While the small CAVIAR retina could hardly resolve contrasts of 100%, the large CAVIAR retina reliably detects contrasts below 20%. The bandwidth of the self-clocked switch-cap design exceeds the bandwidth of the small CAVIAR retina dramatically on both ends. Unlike to the previous designs, slow movements are still reliably detected. The large CAVIAR retina has high-speed capabilities due to the increased bandwith on the higher end.

The large CAVAIR retina is one of the first AER imagers that can be used for real world applications. It has potential for high speed (low latency) manufacturing or inspection, and for low power surveillance, building automation, or visual prosthetic front ends. The large size retina is used by other people in combination with other AER chips and by a group that does real-time processing of the address event data with digital off-the-shelf components.

Chapter 5

The Double Line Sensor

5.1   The Implementation

Optical line sensors (mostly single line sensors) are traditionally used for high-speed imaging problems in industrial applications where the target objects move at controlled speed and under controlled lighting conditions perpendicular to the line of sensor elements. A typical application is quality inspection on a conveyor belt.

Conventional clocked line sensors produce huge amounts of data when employed in high-speed short-latency machine vision problems. For particular applications such as measurement tasks, object classification, object orientation monitoring etc. the bulk of the image data generated does not provide any information necessary to accomplish the function or to increase reliability or precision. One way to suppress redundancy is to use our temporally differentiating event-driven sensor architecture.

High-speed imaging typically depends on well controlled lighting conditions. Our event-driven switch-cap pixel detects contrast changes reliably over a wide dynamic range and is thus potentially suitable for use under difficult, variable lighting conditions i. e. road traffic monitoring or surveillance.

We implemented two parallel lines of the large CAVIAR pixel to test the capabilities of the event-driven switch-cap pixel in a double-line configuration. Th dual-line configuration provides additional functionality such as precision measurements of object velocities or trajectory angles by correlating data from the two parallel pixel lines.

5.1.1   The Chip Architecture

Figure 5.1 shows a block diagram of the sensor architecture, in which 128 pixels are arranged in two parallel lines of 64 pixels at a distance of $240 \mu m$. The pixels
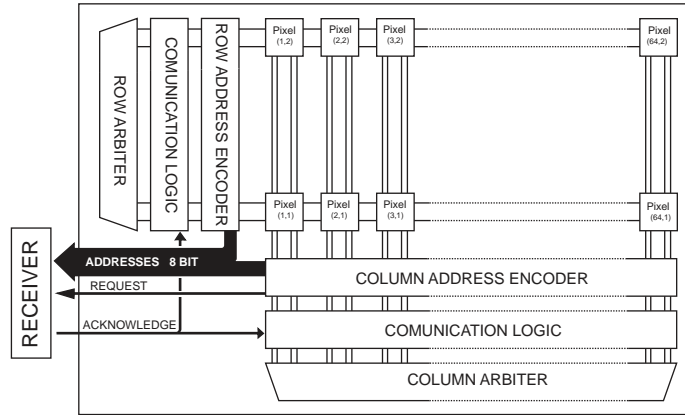
Figure 5.1: **Block diagram depicting the sensor architecture.**

communicate with the row and column arbiters via the pixel handshake signals request (RR, $CRON$, $CROFF$) and acknowledge (RA, $CAON$, $CAOFF$) shown in Figure 3.3. Row and column addresses are set in the respective encoders and communicated off chip on the 8-bit AER bus (6 bits column address, 1 bit row address, 1 ON/OFF polarity bit). The double line architecture is the same as the large CAVIAR retina architecture with a reduced resolution along one dimension. The AER communication follows the very same protocol implemented in the large CAVIAR retina.

5.1.2 The Layout
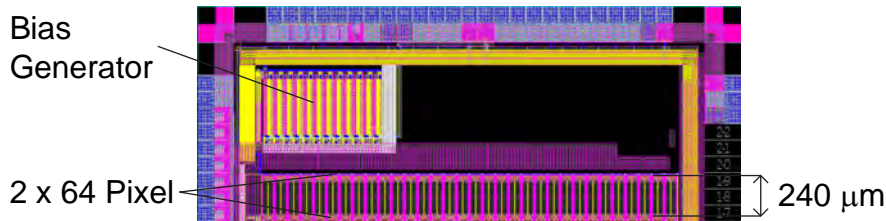
Figure 5.2: **Shows the dual-line sensor layout. The two lines of 64 pixels are spaced** $240\mu m$ **appart.**

The dual-line sensor was implemented on a chip together with test structures. Figure 5.1 shows the layout of the dual-line sensor. The test structures are not visible and are not discussed in this thesis. The pixel lines, each consisting of 64 large CAVIAR retina pixels, are space $240\mu m$ appart. The spacing of the two lines enhances the temporal separation of the individual lines and therefore the resolution for speed and orientation measurements. The bias generator, visible on top of the pixel lines is used for the whole chip and was the first

implementation of a programmable bias generator. The programmable bias generator allows bias adjustments made through a computer interface and is discussed in section XXXX.

5.2   The Experimental Results

We characterized the dual-line sensor using fast rotating paper stimuli and Kodak uniform density attenuation filters. The following sections discuss the time resolution, the dynamic range and an application example of the dual-line sensor.

5.2.1   Time Resolution

To measure the response of the line sensor to fast stimuli we presented a 50% gray circular spot on a white background to the sensor. The spot was drawn on a white paper disk in a distances of $9cm$ from the center. With an electromotor, we rotated the disk at variable speeds of up to more than 100 revolutions per second.
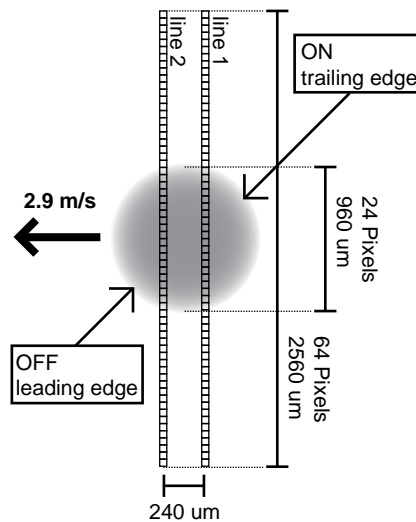
Figure 5.3: **Shows a gray (50% gray) spot stimulus projected onto the sensor focal plane. XXX**

Figure 5.3 shows the calculated speed of the stimulus image on the sensor focal plane with the disk rotating at 57 revolutions per second. The on-chip speed is a more practical metric than the actual stimulus speed, because the geometrical properties of the setup and parameters of the optical lens have not to be taken into account. The stimulus image spans over about 24 pixels, roughly one third of a line, and moves from right to left with an on-chip speed of $2.9m/s$.

Figure 5.4 shows the response of the line sensor to this stimulus at a scene illumination of approximately $1k$lux using a $f/1.4$ lens. The bandwidth of the receiver connected to line sensor limits the output to less than 10^6 events per

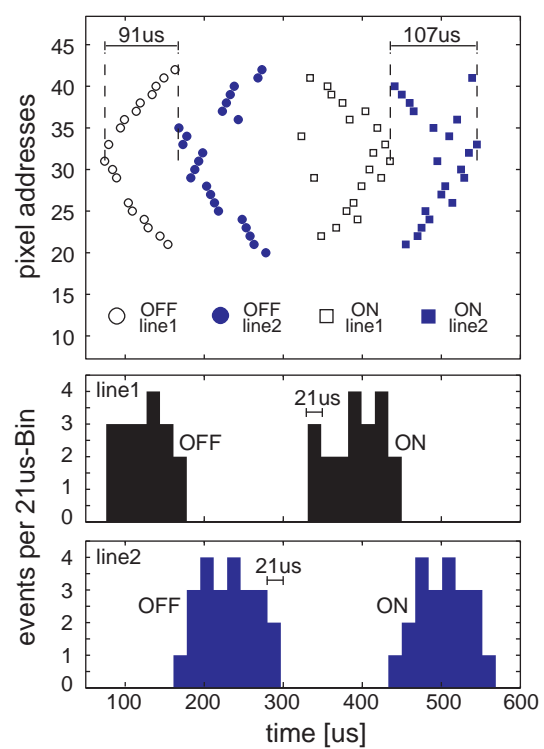Figure 5.4: **Plots the events generated in response to the described stimulus. The upper plot shows the individual events and the lower two plots show histograms of ON and OFF events in time for the first line and the second line. XXX**

second. The address events (color and shape coded for the different polarities (ON/OFF) and the two lines) are displayed in the upper plot. A marker represents an event that occurred at a certain time (x-axis) in a certain pixel (y-axis). The circles depict OFF events, the squares ON events. The data originating from the two different lines are distinguished by using empty and filled markers.

The OFF events track the leading edge of the circular spot and the ON events track its trailing edge. The first OFF event of line 2 appears $91\mu s$ after the first OFF event of line 1. Within this period, the leading edge shape is still temporally resolved, exhibiting a time resolution of around $10\mu s$. Effects of limited AE bus bandwidth appear as characteristic patterns of sequential event processing of certain OFF events of the second line and indicate that the sensor combined with the data receiver is operated near its limits. A faster receiver would extend the speed range until the intrinsic limit of the sensor is reached.

The lower panels of Figure 5.4 show the histograms of activity in time. By measuring the shift in time of the edges of the second line relative to the first line, the speed of the stimulus at the focal plane of the chip can be determined.
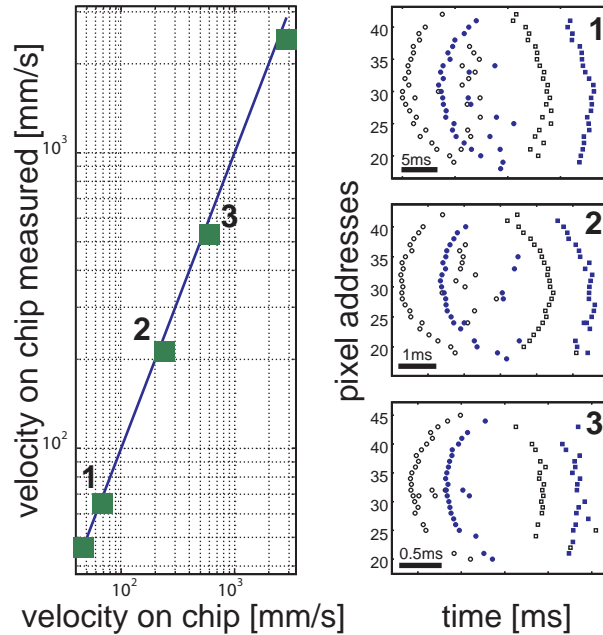
Figure 5.5: **Displays velocity measurements for various stimulus speeds plotted on a logarithmic scale and the corresponding events of the three labeled data points.**

Figure 5.5 shows a logarithmic plot of the speed measured by our dual-line sensor versus known stimulus speed. To quantify the speed we measured the time differences between the first OFF events of the first and second line as well as the time differences between the last ON events of both lines. The square data points represent measurements with the stimulus moving at different velocities from approximately $45mm/s$ to approximately $2.9m/s$ on the sensor plane. The data points are distributed along the line y = x. This result indicates that the velocity is reliably represented over a wide range of velocities.

Three data points are labeled and the three corresponding plots on the right of Figure 5.5 show that the event distribution looks very similar for speeds that are about two orders of magnitude apart. For very slow velocities, each pixel produces several events for a single stimulus edge. A simple correlation method applied to the address event data of the two lines can be used to precisely determine the delta time. This task can be performed in a straightforward way by a simple, inexpensive DSP.

5.2.2 Dynamic Range
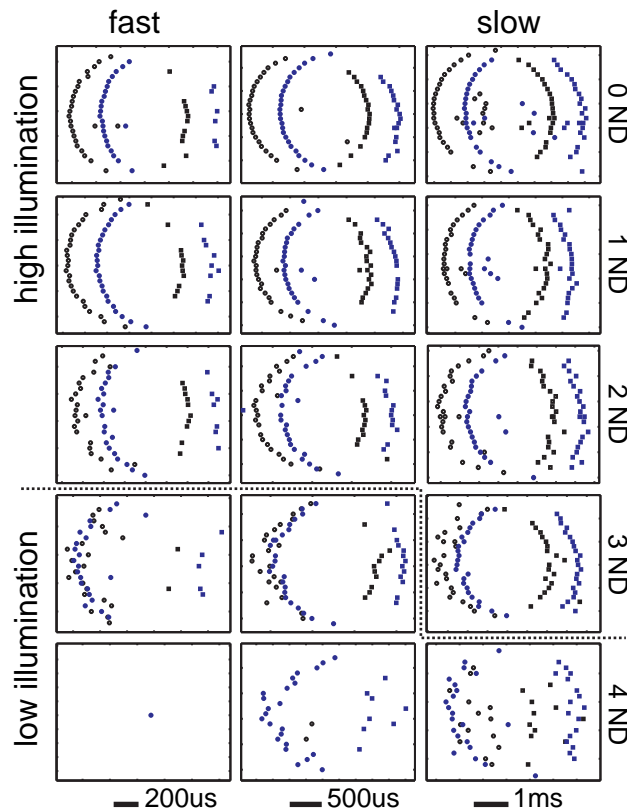
Figure 5.6: **Plots of the response of the dual-line sensor to the described dot stimulus with different illumination conditions and different speeds. The dashed line indicates the lower limit of the sensor performance.**

The dynamic range of the photoreceptors and pixel circuits, i. e. the region of scene illumination where the sensor produces useful data, spans from under 1lux to more than 10$k$lux , whereas the response of the sensor naturally is limited both by the speed of the stimulus and the lighting conditions.

Figure 5.5 shows the response of the sensor to the gray dot stimulus moving at different speeds and under various illumination conditions. The discrete levels of illumination were created by using a light source which generated 1$k$lux of scene illumination and applying uniform density filters of 1 to 4 decades of

attenuation. The dashed line indicates the performance limits of the sensor due to low lighting. At 1lux scene illumination, the dual-line sensor achieves a time resolution that is better than the equivalent time resolution of a conventional line sensor operating at 1000 frames per second (fps). At 10lux, the corresponding time resolution is higher than 10$k$fps.
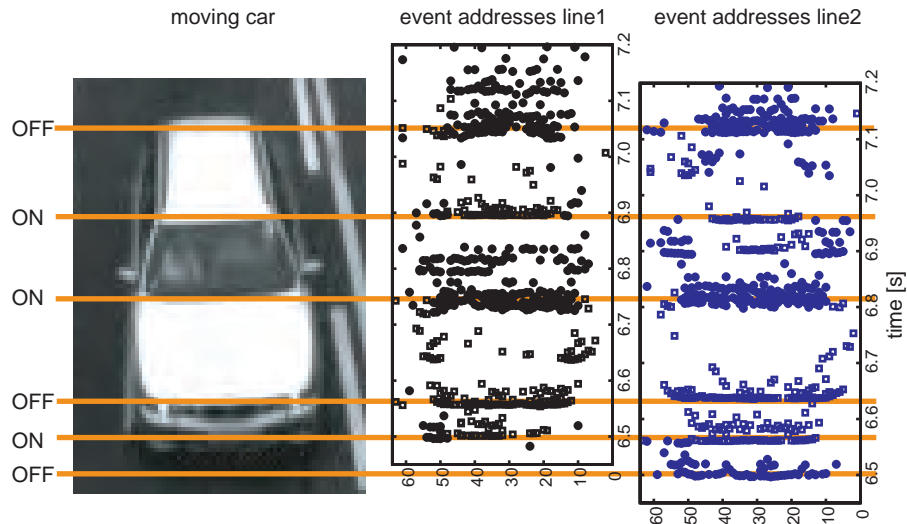
5.2.3   An Application Example

Figure 5.7: **Plots of the events in time in response to a car moving over the visual field of the lines. The plots are shifted relatively to each other and the picture of the car so that ON and OFF edges are aligned; the shift of the two lines can be used to geometrically calculate the speed of the observed car.**

Figure 5.7 shows the response of the dual-line sensor to a car in a traffic monitoring scenario. The shift of the second line in respect to the first line is 69ms, corresponding to a speed at the focal plane of 3.47mm/s. Again the delta time is best be established by using a correlation method. Knowing the relevant geometrical and optical parameters of the setup, the speed of the car on the road can be calculated from the focal plane speed in a straightforward way. The example car traveled at a speed of 56km/h.

5.3   Discussion

In this chapter we described a wide dynamic range dual-line optical sensor. The dual-line sensor implement two lines of 64 large CAVIAR retina pixels. The reduced resolution in one dimension yields lower output data rates and enables very high-speed real time vision in compact, low-cost and low-power imaging systems. Such systems can be employed in high-speed measurement applications. Due to the logarithmic response and the wide dynamic range of the self-clocked switch-cap pixel, the dual-line sensor can be applied to vision problems with uncontrolled and varying lighting conditions. The dual-line sensor is

Pixel size	$40 \times 40 \ \mu m^2$
Pixel complexity	28 transistors (16 analog), 3 caps
Array size	2×64
Fabrication process	4M 2P $0.35\mu m$
Biasing requirements	12 generated gate biases
Fill factor	8.1% (PD area $130\mu m^2$)
Min. scene irradiance ($f/1.4$)	Approx. $1mW/m^2$ or $1lux$.
Photodiode dark current	(estimated) 20fA ($\approx 10nA/cm^2$)
Sensor Latency	$< 100\mu s$
Burst event rate	$\approx 4Mevent/sec$

Table 5.1: Specifications of the dual-line sensor

therefore well suited for use in i. e. traffic monitoring applications as well as industrial applications where target object illumination cannot be controlled. The specifications of the dual-line sensor are summarized in table 5.1.

Chapter 6

The AER Temporal Contrast Vision Sensor

6.1 Preamble

The most important modifications that made our latest implementation of the self-clocked switch-cap pixel to one of the most successful AER sensors so far, are besides small alterations in the pixel design (which are discusse in section 3.3) changes we made in the bias generator circuit and the effort we spent on a compact camera system with a high-speed USB interface. Due to the relevance of the bias generator circuit, we will commence this chapter with a short description of the bias generator.

6.2 Programmable On-Chip Bias Generation

One long missing step in our chips was a reliable and programmable bias generation directly on the sensor chip. Although we started to implement pre-programmed bias generator circuits on our chips right with the small CAVIAR retina, we gained the experience that it is almost impossible to precisely predict the best (or even a good) set of biases even when using sophisticated simulation tools. Furthermore, an integrated programmable bias generator opens the possibility of adjusting the biases according to conditions, (i. e. daytime or nighttime, or when more bandwidth is desired) and dynamically under feedback control. This is analogous to the automatic gain control loop used in image sensors. For

example, the event thresholds or the refractory period can be increased if the output bus starts to saturate.

In comparison to the use of expensive high precision external DACs, an internal programmable bias generator has the following advantages:

1. *Price.* In the long run, it is much cheaper to use silicon area for the bias generation instead of using external DACs.

2. *Size.* An internal bias generator circuit uses a relatively small amount of chip area, that in most cases will not increase the package size in which the chip is placed. External DACs on the other hand use a significant amount of area that can surpass the size of the packaged chip, if a small package is used.

3. *Pins.* Our programmable bias generator uses 4 pins: data, clock, latch and reference resistor use each one pin. If external biasing is used, each bias has to be supplied via a separate pin.

4. *Temperatur.* The last but most important point is that our internal programmable bias generator makes the operation of the chip largely temperature insensitive. This can not easily be achieved with an external biasing scheme.

In order to take the vision sensor into the field it is useful to profit from the listed advantages. To build a manufacturable device that requires a minimum of external components, especially precision analog components, it is of prime importance to build a convenient internal bias generator scheme. Therefore we developed and integrated programmable bias generators [56] that are based on earlier work on pre-programmed biases [57]. Here we will describe the role of the bias generator in the AER temporal contrast vision sensor specifically.

The bias generator chip interface is a standard SPI serial interface. The fabricated circuit has 6 decades of overall range, but altogether a (possibly non-monotonic) resolution of better than 10^{-7} of the master current. The currents that are produced have PTAT-like characteristics, resulting in relatively constant gm behavior. 12 biases can be loaded over SPI in less than $1ms$, and they settle to new values in a few μs.

Each bias is a voltage that is generated in the bias generator, buffered and then connected to the core (which comprises the pixel array and AER communication infrastructure). Generally these bias voltages determine the level of current in a transistor but they are also used as cascode and voltage references. The pixel requires 8 bias voltages, one of them being a cascode bias (*cas* shown in Figure 3.3). In addition the pixel reset pulse level $rGND$ is programmable to allow for minimizing the reset switch charge injection. The peripheral AER circuits use 3 biases for the row and column pull-ups and the address encoder pull-down.

Each bias is programmable to a 24-bit fraction of the chip's master bias current I_m (Figure 6.1). The master bias current is generated by a Widlar bootstrapped mirror. Our typical setting uses a master bias current of about $40\mu A$. The master current is chosen by the value of an external resistor R_{master}. A $40\mu A$ master current is generated when a $10k$Ohm resistance is attached.

Each bias copies the master current and splits it using an independent current splitter. Bits loaded into a long shift register steer the current splitter
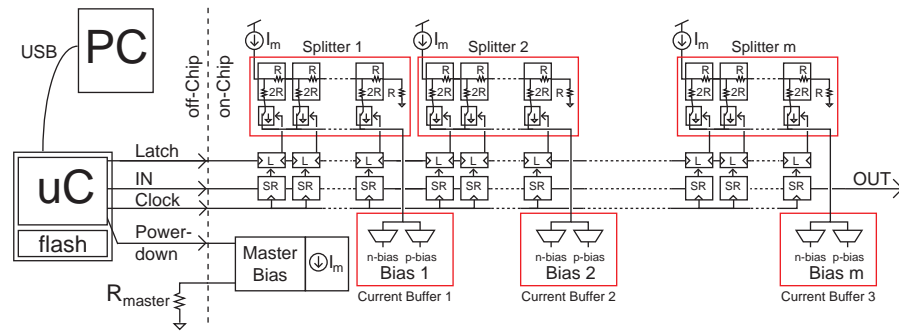
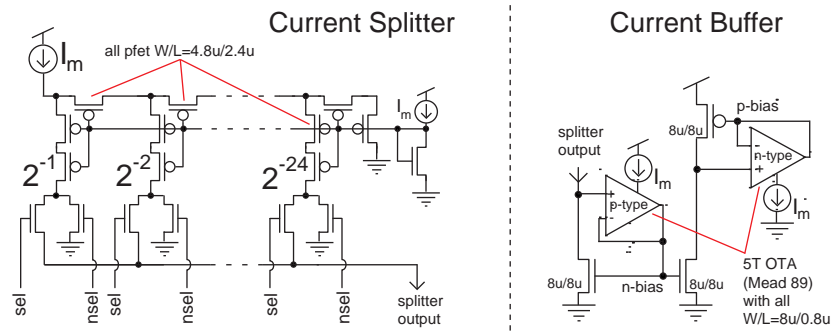Figure 6.1: **The biasgenerator schematic.**

Figure 6.2: **The splitter and the current buffer.**

output to an active mirror (Figure 6.2) that buffers the resulting n- or p-type bias voltage to the chip core. Cascode biases are created by using 2-diode connected transistors in series in the current buffer (CB). The unity-gain followers in the CB are biased with the master current.

The circuit layout intentionally used a p-type current splitter (as in [57] but different than i. e. [58]) despite the observation that p-FET matching is generally worse than n-FET matching, sometimes as much as 10x worse [59]. We optimized matching by using large unit transistors and ensuring equivalent metal coverage. The p-FETs were used in the current splitter because on vision sensors they can be protected from parasitic photocurrent effects [57].

6.3 The Implementation

6.3.1 The Chip Architecture
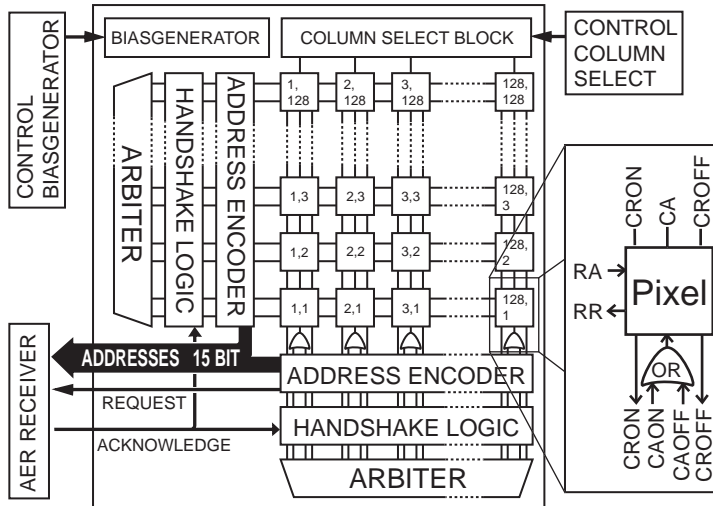
Figure 6.3: **Shows the chip architecture of the AER temporal contrast vision sensor.**

The pixels are embedded in a 128×128 array and handshake asynchronously with the peripheral circuits (Figure 6.3). Pixels have an x,y-address and in addition they communicate the type of event (ON or OFF). The chip output is a 15 bit digital address that has the 7 bit x and y addresses and a polarity bit. The communication uses the same AER principle as described in section 4.1.1, except that the arbiters are modified to work in a non greedy fashion [24–26]. For a thorough description of the AER handshaking cycles refer to section 2.3.2.

The bias generator circuit, described in the previous section, provides all necessary biases. It is connected with a *Serial Peripheral Interface* (SPI).

To reduce output data rate in case of limited post-processing power or to avoid bus arbiter saturation in high-speed imaging applications, the sensor is equipped with a configuration controller which allows the user to operate the sensor in one of four configurations:

1. Full array sensor.

2. Double-line sensor with $320\mu m$ line separation.

3. Quad-line sensor with 200, 400 and $800\mu m$ line separations.

4. Programmable. Any arbitrary number and combination of pixel-lines can be activated.

These configurations are hard-wired and are selected by tying the corresponding digital input pads to VDD and GND respectively. Choosing the programmable configuration enables a (SPI) connection. By loading the appropriate bits each of the 128 columns can be independently enabled or disabled. This feature can be used to configure the chip into a multi-line sensor of arbitrary line combination and separation. Potentially useful setups that reduce output data rate and unburden the arbiter could be to enable i. e. only every other line of the array, thus reducing vertical spatial resolution or reduce the active array to an arbitrary region-of-interest. The configuration can be dynamically changed while the sensor is in operation.

6.3.2   The Layout and the Silicon

The chip has been fabricated in a standard $0.35\mu m$ 4 metal 2 poly CMOS process. The process is a bulk silicon process rather than an epitaxially-grown surface process. Bulk silicon usually has larger dark current. This assertion is supported by the vendor process parameters for nwell-p-substrate junction leakage ($60aA/\mu m^2$ area, $0.27fA/\mu m$ edge, $6\ nA/cm^2$ for large area junction). These leakage currents are about $100\times$ those for a specialized imager process.
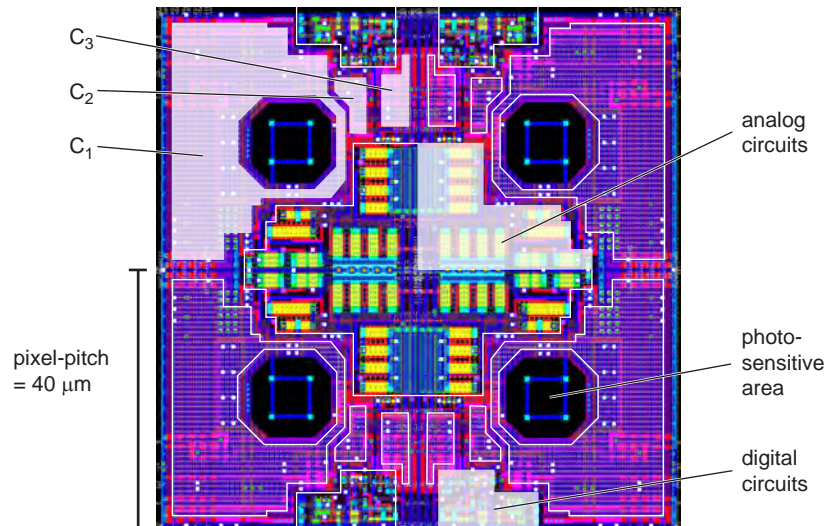

Figure 6.4: **Shows four pixels laid out in point symmetry. The three pixel capacitances, the pixel analog and digital transistors as well as the photodiode are indicated.**
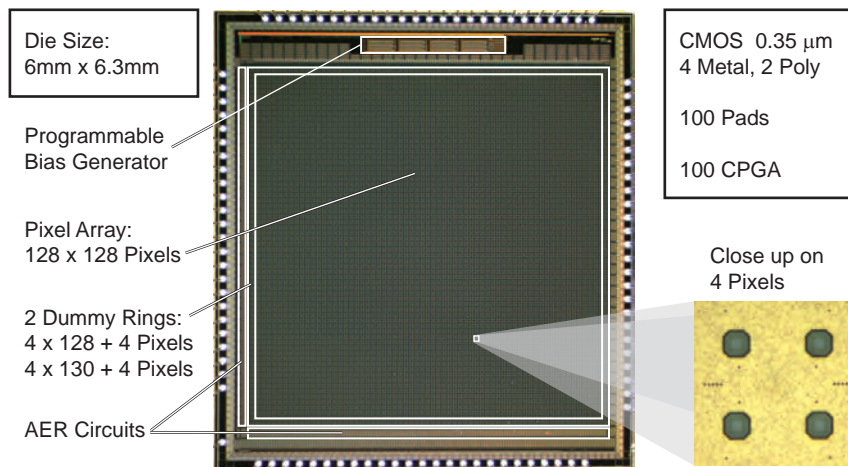
Die Size:
6mm x 6.3mm

Programmable
Bias Generator

Pixel Array:
128 x 128 Pixels

2 Dummy Rings:
4 x 128 + 4 Pixels
4 x 130 + 4 Pixels

AER Circuits

CMOS 0.35 μm
4 Metal, 2 Poly

100 Pads

100 CPGA

Close up on
4 Pixels

Figure 6.5: **Die photograph of the 128 x 128 Sensor.**

Figure 6.4 shows the layout of four pixels. Most of the pixel area is capacitance. Unit capacitors are not necessary because capacitor ratios for all pixels rather than absolute capacitor values are critical. Special features of this layout are the following:

1. The pixel array is padded by a ring of dummy quad pixels that are electronically held in reset.

2. The bias lines that run vertically through the array and the power lines that run horizontally are shared by neighboring pixels

3. The power supplies to the analog and digital parts of the pixel, indicated in Figure 6.4, are kept separate and no digital ground in the pixel is tied to the substrate

4. The only metal cuts in the pixel core are the photodiode openings

5. The analog transistors in the pixel that should match are drawn as similar, proximal and large as possible.

Figure 6.4 shows a die photograph of the AER temporal contrast vision sensor. Most of the chip area is pixel array. The peripheral AER circuits and bias generator occupy only about 5% of the area. The die size is $6mm \times 6.3mm$. The AER temporal contrast vision sensor has 100 pads of which about the halve is not necessary for normal chip operation and is used primeraly for characterization purposes.

6.3.3   The PCB

The AER temporal contrast vision sensor can be directly interfaced to other AER components that use the same word-parallel protocol or can be easily adapted to other AER protocols using simple commodity logic circuits like

a) Vision sensor system
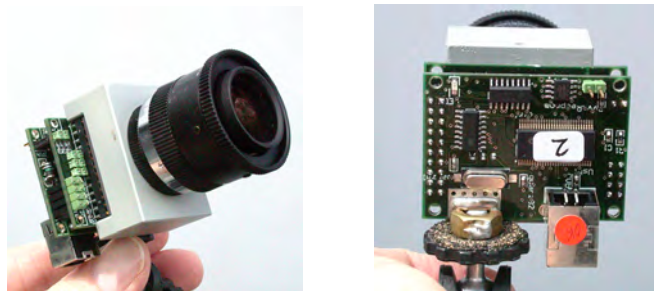
b) Vision sensor USB interface

Figure 6.6: **Shows the present implementation of the AER temporal contrast vision sensor camera system with USB2.0 interface. Part a) shows the vision sensor system and part b) is a schematic view of the USB hardware and software interface.**

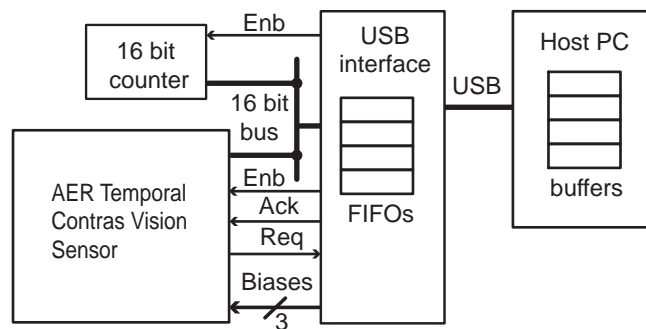CPLDs. For the separate use of the sensor we built a small camera system that uses only 7 active components: an AER temporal contrast vision sensor, a Cypress FX2LP USB2 interface chip, two 8 bit counters, a serial EEPROM, and a regulator. This compact camera system, built on two stacked PCBs is shown in Figure 6.6. The USB chip's integrated programmable chip interface, USB serial interface engine and FIFOs allow us to transmit events without any pauses that would corrupt the precise event timing (Fig. 6(b)). On the host side there is considerable complexity in hardware single-threaded platform like most PCs. Tobi Delbrückdeveloped a host architecture consisting of more than hundred JAVA classes for acquiring, rendering, and processing the non-uniformly distributed, asynchronous retina events in real-time on a PC.

The vision sensor sends AEs to the USB interface, which also captures timestamps from a free-running counter running at $100kHz$ on the same 16-bit bus. These timestamped events are buffered by the USB FIFOs to be sent to the Host PC. The PC also buffers the data in USB driver FIFOs, unwraps the 16 bit timestamps to 32 bit values, and offers this data to other threads for further processing. The same USB chip also uses a serial interface to control the vision sensor biases. Flash memory on the USB chip stores persistent bias values.

The camera system is directly powered by the USB connection. The compact and implementation in combination with the USB interface and power supply allows to use the AER temporal contrast vision sensor the same way any webcam can be used.

6.4 The Experimental Results

The metrics usually used for imager performance are quantities like resolution, fill factor, pixel size, imager size, die size, operating range (illumination), maximum read-out rate, supply voltage, and power dissipation (quiescent and maximum). All these parameters are relevant for the AER retina and all are readily determined by standard procedures.

In addition, because the effects of transistor mismatch have traditionally plagued neuromorphic aVLSI, we are particularly interested in the uniformity of the pixel responses. A spike should ideally have identical meaning to a receiver, regardless of its originating pixel.

6.4.1 Uniformity of Response

For standard CMOS image sensors the fixed-pattern noise (FPN) characterizes the deviation of the individual pixel response from the mean and therefore also the uniformity of response. For this vision sensor the equivalent measure is contrast threshold variation. The contrast threshold $C_{threshold}$ is the minimum relative illumination change needed to produce one event; it is the effective resolution in intensity changes. The $C_{threshold}$ is introduced in section 3.3 on page 3.3. In the following paragraphs the dominant sources of mismatch are discussed, the contrast threshold mismatch is introduced as a figure of merit that is in a first approximation independent of comparator threshold settings and finally the contrast threshold mismatch of the present sensor is characterized.

The dominant source of mismatch is expected to be found in the relative comparator mismatch because:

1. Device mismatch for transistors is in the order of 30% while for example capacitor mismatch is only in the order of 1

2. The comparators are simple two transistor devices without offset compensation.

3. The front end steady state mismatch is eliminated by differentiation.

4. Gain mismatch (kappa mismatch) in the front end is expected to be in the order of 1%.

For an analytical consideration of the mismatch only the mismatch in the comparators and in the reset level are taken into account. The event threshold mismatch is defined as follows:

$$\sigma_{contrast} = \text{1-sigma variation of } C_{threshold} \tag{6.1}$$

We can visualize the effect comparator mismatch has on the pixel mismatch and how the gain of the differencing circuit reduces mismatch as in Figure 6.7(a-c). Figure 6.7c) shows the mismatch in the reset and comparator thresholds separately as hypothetical probability density functions (PDFs). Figure 6.7b) shows the PDFs of the differences that define the actual event thresholds (the correlations between the resulting difference distributions introduced by the common reset are ignored here), and Figure 6.7a) shows these distributions referred back to the input log intensity through the gain of the differentiator, assuming no mismatch in the differentiator. (For simplicity a differentiator gain of A = 10 is assumed here.) It is graphically obvious that the PDFs are simply scaled by the gain, dividing mismatch by the gain. Setting different threshold levels results in a shift of the PDFs at the input, but will not change their shape. Ultimately the PDFs determine the maximum resolution that can be obtained with the sensor by lowering the thresholds.
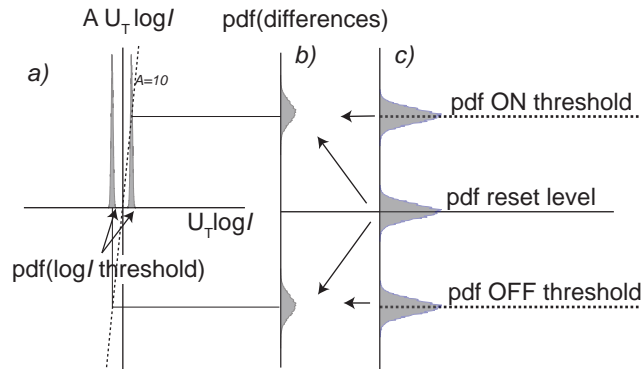

Figure 6.7: **Shows mismatch referred back to input. Part a) shows hypothetical pdfs (probability density functions) of input-referred mismatch using the pdfs of mismatch in the relative threshold between comparator reset and ON or Off level b), which arise from the individual offsets c).**

The PDFs at the input are directly given by the introduced $\sigma_{contrast}$. The $\sigma_{contrast}$ is independent from the threshold setting if we consider only the dominant mismatch sources. In the following paragraphs the measurement of the uniformity of the response is discussed.

The dynamic response property of the sensor requires a dynamic stimulation. A global intensity variation is not the ideal stimulus because it can saturate the bus. We therefore use a black bar that is moved through the visual field of the sensor. To avoid effects of the refractory period, we use a bar that has linear gradient edges and defocused the lens, creating a smooth gradient on the sensor plane. The stimulus was moved across the visual field of the vision sensor at a constant speed of 1pixel/10ms. The stimulus we used is shown in Figure 6.8a). The AER temporal contrast vision sensor responds with several ON and OFF events to the bar stimulus. The number of events N of one type–ON or OFF– that the sensor produces in response to the stimulus depends on the contrast threshold $C_{threshold}$ as follows:

$$N \pm \sigma_N = \frac{C_I}{C_{threshold} \pm \sigma_{contrast}} \approx \frac{C_I}{C_{contrast}} \left(1 \mp \frac{\sigma_{contrast}}{C_{contrast}} \right) \qquad (6.2)$$

where σ_N is the standard deviation in the number of events produced and C_I is the maximum logarithmic contrast given by

$$C_I = ln \left(\frac{I_{max}}{I_{min}} \right) \qquad (6.3)$$

with I_{max} corresponding to maximum and I_{min} to minimum luminous intensity reflected by the stimulus. The contrast threshold (which is independent for ON and OFF events) in turn depends on the comparator threshold settings according to equations 3.7 to 3.9 on page 50.

We conducted the experiment with the moving bar for 5 different threshold settings. Figure 6.8b)-c) show the responses of the pixels in a central column of the sensor for the two extremes of the threshold settings, reconstructed over time. For the reconstruction we computed the mean contrast threshold $C_{threshold}$ with the following relation:

$$C_{threshold} = \frac{C_I}{N}, \qquad (6.4)$$

which is obtained from equation 6.2. Figure 6.8b) shows a high threshold condition while Figure 6.8c) shows a more sensitive threshold configuration. The mean number of ON and OFF events for the least sensitive settings is 4.5 events and is 16.1 events for the most sensitive setting. The images show that there is a significant diversity in the initial condition of the pixels, emerging as the distribution times of the first event.

To quantify the pixel mismatch we counted events over a sequence of 40 stimulus presentations. Figure 6.9 shows a histogram of events per pixel for six different threshold settings. The x-axis is the number of events per stimulus presentation and the y-axis is the number of pixels that responded with an equal number of events. Each distribution has a Gaussian shape that is narrower for the less sensitive settings and broadens for more sensitive settings. The settings were chosen so that the number of ON events matched the number of OFF events.

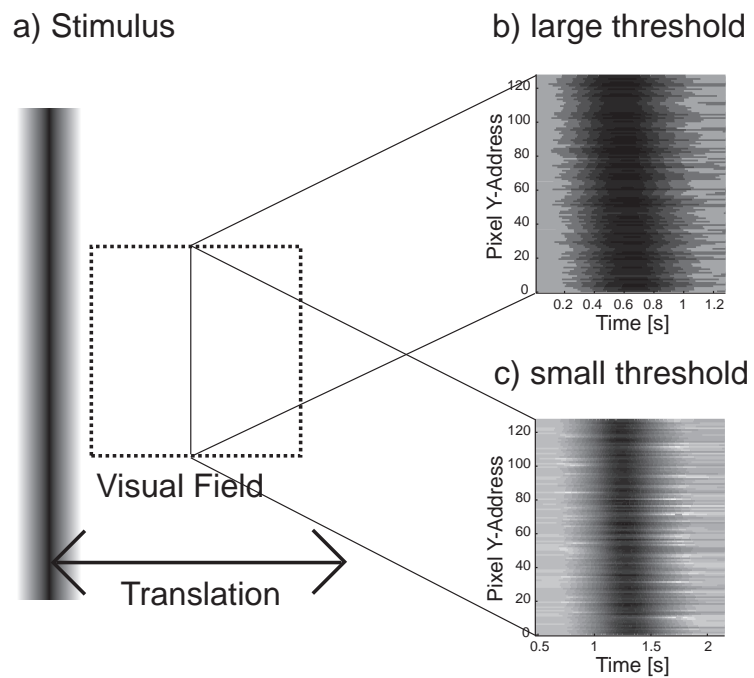Figure 6.8: **Shows setup and typical results for uniformity characterization. Part a) shows the stimulus. Parts b) and c) show the reconstruction of the responses of the center column of pixels. The upper gray scale image shows the response for a large threshold setting (fewer events) and the lower gray scale image shows the response for a lower threshold setting (more events). Note the different scales of b) and c).**
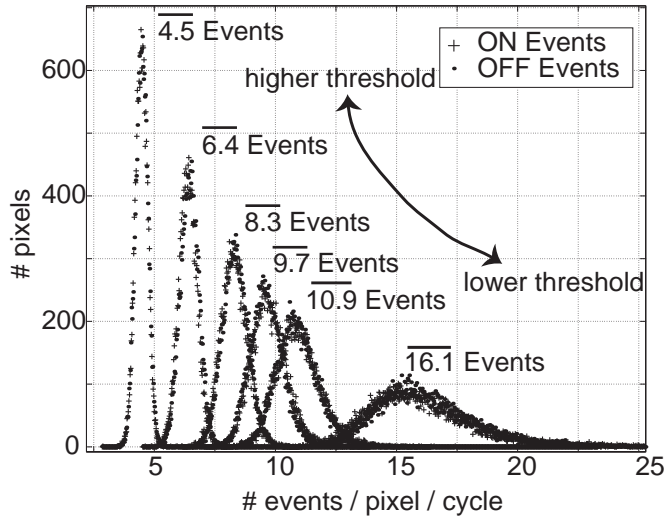
Figure 6.9: **Shows the distributions in the number of events recorded for 40 repetitions of the bar sweeping over the array. There are no dead pixels and only very few hot pixels–one pixel makes 4.6 times the mean ON events and 4.5 times the mean OFF events and one pixel makes 3.6 and 3.4 times the mean ON and OFF events respectively–the rest of the hot pixels lie within 2 times the mean. All the numbers are divided by 40 to show the number of events per sweep.**

We can compute the threshold mismatch from the width of these distributions combined with the known stimulus contrast of 15:1 using:

$$\sigma_{threshold} = \frac{\sigma_N}{N} \cdot C_{contrast}, \tag{6.5}$$

which is gained from equation 6.2. The results of computations made with equations 6.4 and 6.5 on the data shown in the histograms in Figure 6.9 are plotted in Figure 6.10 and Figure 6.11. In the following paragraphs we will discuss these results.

Figure 6.10 compares the plots the measured contrast threshold $C_{threshold}$ as a function of the comparator threshold settings. The error bars on the measurements show the $\sigma_{contrast}$ of the measured $C_{threshold}$. The solid lines show the analytical model as developed in section 3.3 on page 48. The dashed lines indicate the upper and lower deviation for an assumed $10mV$ relative mismatch in the comparators. The analytical prediction is close to the data in the case of OFF events. A close inspection shows that the slope is a bit flatter and that there is a slight offset. The prediction for the ON events is worse. This is partly due to the fact that the ON comparator was operated around threshold. Another observation is that the measured ON events indicate a lower contrast threshold than predicted while the OFF data indicates a higher contrast threshold than predicted. This could be explained by effect of the positive reset switch charge injected onto the floating node of the differencing circuit (see Figure 3.3 on page 49), that facilitates the emission of an ON event.

Figure 6.11a) shows the standard deviation $\sigma_{contrast}$ of the contrast threshold $C_{contrast}$ plotted versus the mean contrast threshold $C_{contrast}$. The solid
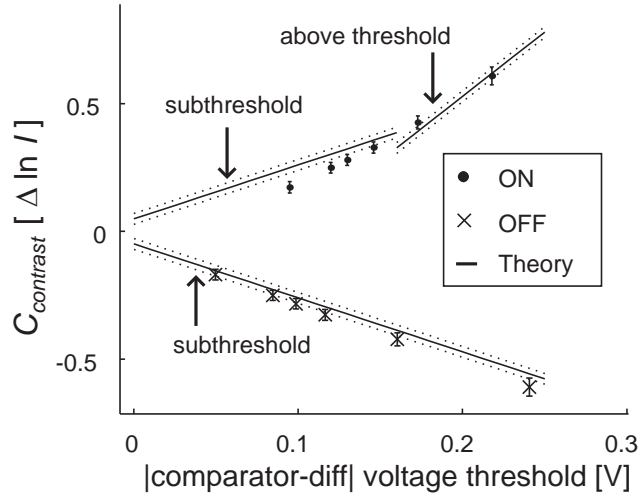
Figure 6.10: **Shows the measured $C_{threshold}$ as a function of the comparator threshold setting. These results are compared with the theory discussed in section 3.3 on page 48. The x-axis is the difference diff-don respectively diff-doff (see Figure 3.3 on page 49). The y-axis shows the theoretical curves and the measured $C_{threshold}$. 1-sigma threshold variation is shown as error bars. The dashed lines surrounding each theoretical curve are the computed variations in threshold assuming the relative comparator mismatch is $10mV$.**
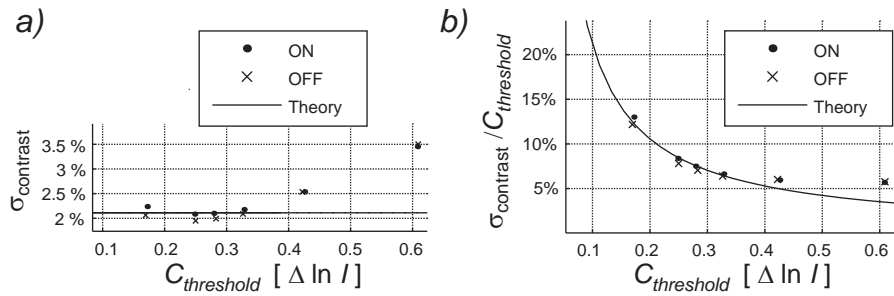
Figure 6.11: **Part a) shows the standard deviation of the measured response $\sigma_{contrast}$ plotted as a function of mean contrast threshold $C_{contrast}$. Part b) shows the relative standard deviation $\sigma_{contrast}/C_{contrast}$ (coefficient of variation) as a function of mean contrast threshold $C_{contrast}$.**

line shows the predicted mismatch assuming a $10mV$ relative mismatch in the comparator thresholds. For low $C_{contrast}$ (10% - 40% illumination change), the standard deviations lie within 2% - 2.5%. This is close to the theoretically predicted 2.1%. The mean $\sigma_{contrast}$ for all contrast thresholds is 2.3%. For higher contrast thresholds, $\sigma_{contrast}$ increases pointing to the fact that sources of mismatch in the front end are becoming dominant. Figure 6.11b) shows the resulting relative standard deviation $\sigma_{contrast}/C_{contrast}$ as a function of the measured contrast threshold $C_{contrast}$. The relative standard deviation has a $1/x$ behavior for low thresholds, where the theoretical prediction is very close to the measured data. A 10% log intensity threshold leads to 22% relative standard deviation. $\sigma_{contrast}/C_{contrast}$ saturates for higher thresholds to a minimum of about 6%.

In conclusion we can say that mismatch in the pixels is dominated by the mismatch in the comparators. This result meets the expectations and points out the benefit of the gain of the differencing circuit. The mismatch can be characterized by the contrast threshold variation $\sigma_{contrast}$ and it amounts to contrast of 2.1%.

6.4.2 Dynamic Range

Our vision sensor achieves a wide dynamic range due to the logarithmic compression in the front end photoreceptor circuit and the local (pixel intrinsic) event-based differencing operation (see Figure 3.3 on page 49. We define the dynamic range as the ratio of maximum to minimum lighting conditions at which events can still be reliably generated by high contrast stimuli.

To measure the intra-scene dynamic range we applied neutral density (ND) filters directly on the die[1], creating four quadrants of ND 0, ND 1, ND 2 and ND 3, where i. e. ND 3 means an attenuation of 3 decades. As a stimulus, we move a sinusoidal grating of contrast 15:1 diagonally across the visual field of the sensor. Figure 6.12a) shows the placement of the filters and Figure 6.12b) shows the stimulus and the direction of the stimulus movement. To explore a wide dynamic range, two different lighting conditions were applied. The choice fell on sunlight illumination and fluorescent illumination. The first light source was chosen because we do not have any stronger light source at hand in our lab and latter light source was chosen because fluorescent light has negligible irradiance in the infra red.

Figure 6.12c) show $80ms$ of integrated AER response to the stimulus under sunlight conditions using a $12mm$, $f/1.4$ lens. The luminance from the white part of the stimulus was $18knit$. With the same settings, the experiment was repeated applying a white paper luminance of $42nit$ under fluorescent lighting (Figure 6.12d)). For the sunlight case, there is little difference between the quadrant with no filter and the quadrant with ND 3 filter except a slightly ($1.5\times$) fewer total events, consistent with an increased gain at high photocurrents when the transistor M_{fb} (see Figure 3.3 on page Figure 49) operates in strong inversion. This shows that the response of the sensor is largely independent of the background illumination as long as the lighting is not too low.

The response of the vision sensor to the stimulus with white paper luminance of $42nit$ shows severe degradation of the response in the case of 3 decades of

---

[1]Kodak Wratten gelatine filters, No. 96
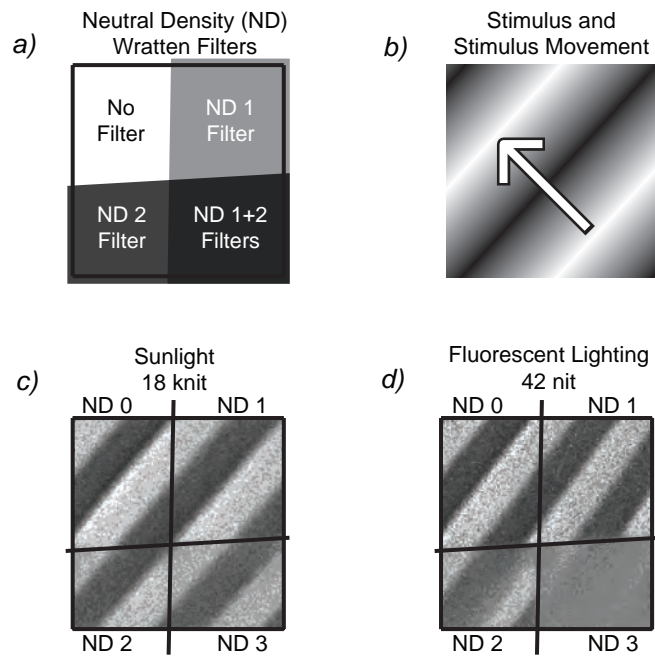
Figure 6.12: **Dynamic range measurements. Part a) shows the placement of the filters and b) the stimulus and the direction of the movement of the stimulus relative to the vision sensor. The gray scale images in c) and d) are generated using** $80ms$ **integrated output of the sensor. 3 ON (white) and 3 OFF (black) are the full scale. For the high intensity condition we used sunlight and for the lower illumination condition fluorescent scene illumination.**

attenuation. Only a fraction of the pixels produce one event per stimulus gradient. A response is still triggered reliably for a sharp black to white contrast (not shown). The response for 2 decades of attenuation shows reduced event numbers compared to response for ND 1 and ND 0, but a clear representation of the stimulus can still be identified. Thus we arrive at our dynamic range figure: The ratio from the white part of the stimulus seen under sunlight (18 knit) to the dark part of the stimulus under 2 decades attenuated fluorescent office lighting (42 nit/100/15 = 28 mnit) amounts to a dynamic range of about $643000 \approx 5.81$ decades $\approx 116dB$.

In order to more rigorously quantify the dynamic range of the sensor, we looked at the single-pixel dynamic range. We stimulate one randomly selected pixel near the center of the array using a modulated bright white LED. The LED is of the "blue + yellow phosphor"-type showing a spectrum between $425nm$ and $775nm$ with the main peak at $470nm$ and the second peak at $575nm$. The LED was modulated by a sinusoidal signal with variable DC offset and amplitude. The amplitude was adjusted to keep a constant contrast and to produce approximately one ON and one OFF event per stimulus cycle. The stimulus frequency is was to a few hertz.
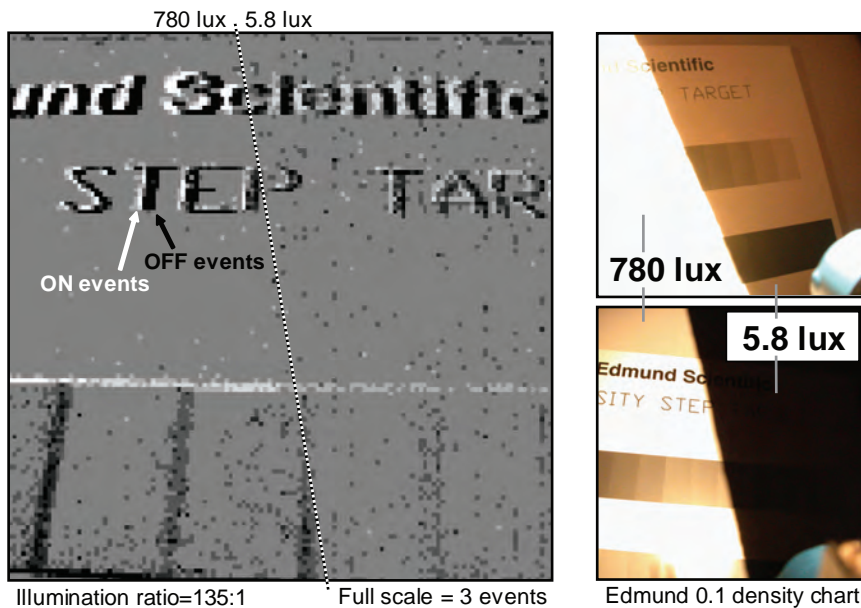
Figure 6.13: **Shows an example of a high dynamic range scene and the clearly superior intra-scene dynamic range of our sensor compared to a conventional digital camera.**

We got reliable response from the stimulated pixel in a range from $10mlux$ to $150klux$ at the sensor focal plane. This translates into 7.18 decades or $143dB$ of single-pixel dynamic range. The measurement has been repeated for a number of randomly selected pixels with similar results.

In conclusion, a dynamic range of more than 120dB can be safely claimed as the measurement on the single pixel resulted in a dynamic range of more than 140dB and because the upper limit in the experiment using the uniform

density filters could not be explored due to lack of a light source that produces stronger reflected luminance from our stimulus. The upper limit is interesting for applications that involve strong light sources under direct inspection.

This dynamic range surpasses the dynamic range of conventional cameras by almost a factor 1000. Figure 6.13 demonstrates the advantage of the contrast encoding of our AER temporal contrast vision sensor compared to the absolute illumination encoding of a conventional digital camera. The scene shown in Figure 6.13 has a measured illumination ratio of 135:1. The target that we moved relative to the sensor is a calibrated Edmund density chart with gray scale patches that differ 10% in reflectance relative to their neigboring patches.

6.4.3   Pixel Bandwidth

Here we show the characterization of the single pixel bandwidth to study the correctness of the theory developed in section 3.4 on page 52. In order to calculate the expected bandwidth, we need to know the photocurrent, which of course varies by many orders of magnitude depending on the illumination conditions. For simplicity we will choose a value of $1pA$, which is approximately 100 times the value of dark current we earlier measured for this process and photodiode shape [60] (see Figure 4.6 on page 60).
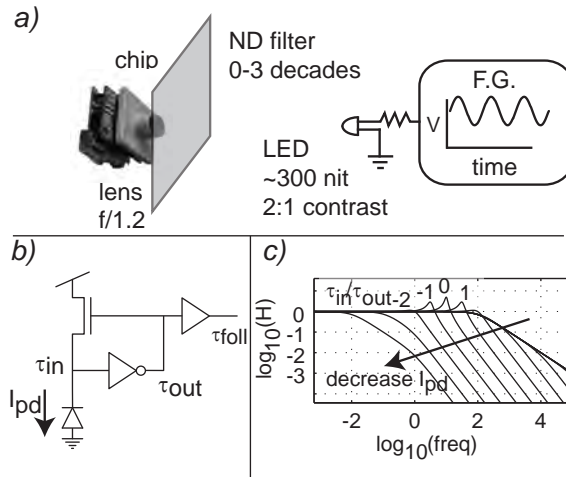
Figure 6.14: **Part a) shows the setup for the bandwidth measurements. The model circuit is shown in b) and c) shows the theoretical linear transfer functions based on equation 3.11 in which τ_{in} is varied over 8 decades while τ_{out} is held constant. These curves show the same response characteristics as the measured results shown in Figure 6.15.**

Taking estimated values for the photoreceptor closed loop gain $A \approx 100$, the photodiode capacitance $C_{pd} \approx 35fF$ and the miller capacitance $C_m \approx 0.3fF$, we can compute that at a photocurrent of $1pA$ and room temperature, the limiting bandwidth should be on the order of $25kHz$. Here C_{pd} is estimated from the process specifications for the n-Well to substrate capacitance ($0.08fF/\mu m^2$, $0.51fF/\mu m$) and the photodiode area and perimeter ($151\mu m^2$, $45\mu m$). It is interesting to note that the bandwidth is increased by using active feedback
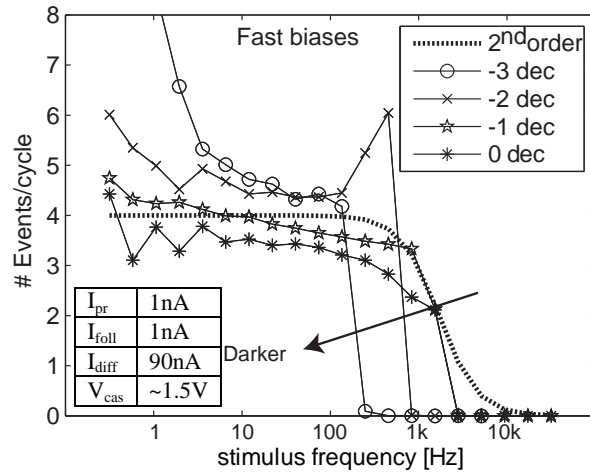
Figure 6.15: **Shows the measured responses to the modulated LED (see Figure 6.14). The curves are labeled with decade attenuation from bare LED with an unattenuated luminance of about** $300nit$ **using a** $6mm$ $f/1.2$ **lens. Data were collected from a single pixel over the duration of** $10s$**, then the number of events was divided by the stimulation frequency times the collection duration, leading to an average number of events per stimulation cycle.**

by a factor of about 15 compared to a passive logarithmic photoreceptor (from reducing the photodiode capacitance $35fF$ to an effective value of about $0.6fF$), and that the effective capacitance at the photodiode node consists of Miller capacitance and reduced photodiode capacitance in equal shares.

To study the validity of this theory, we measured the pixel bandwidth by imaging a LED as a stimulus to a few pixels near the center of the array Figure 6.14a)[2]. We then recorded the temporal "event transfer function" when varying the LED modulation frequency and recording the number of events from a single pixel collected over a fixed amount of time.

Figure 6.15 shows the average number of events (ON and OFF combined) generated per complete sinusoidal cycle of LED modulation for 4 illumination levels spaced by a decade. The thresholds for ON and OFF events were set to produce on average 5 events (2 ON and 3 OFF) per cycle at low frequency. We see from the data that at low illumination (-3 dec) the transfer function shows the characteristic of a first order lowpass filter. At higher illumination (-2 and -1 dec) there is some peaking that is due to resonance, either in the photoreceptor or the differentiator. The peaking indicates that at higher illumination levels the dominant pole is not at the photodiode node but somewhere else in the circuit. At the highest illumination level (0 dec), the measured single pixel bandwidth is about $3kHz$, and is clearly not determined by illumination but by some other dominant pole in the circuit.

Pixel bias currents for these measurements are shown in the table in Figure 6.15 and were adjusted to maximize pixel bandwidth while still maintaining

---

[2]We only stimulated a few pixels to limit the bus activity to ensure that the bandwidth was not being constrained by bus saturation.
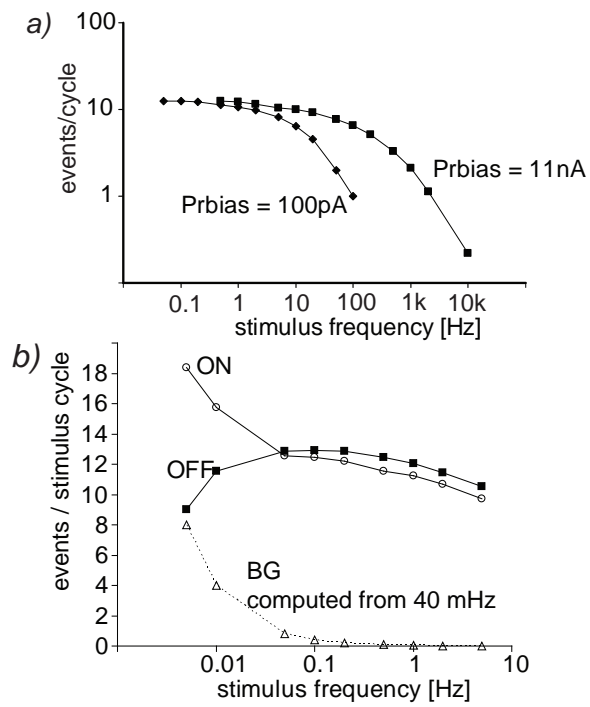
Figure 6.16: **Part a) shows the single pixel frequency response for two photoreceptor amplifier bias currents to demonstrate the control of bandwidth by the bias currents. Part b) shows events produced at very low stimulus frequencies. The BG curve is computed from measured** $40mHz$ **BG activity rate.**

normal operation of the vision sensor as a whole. Using these values, we can estimate the pole frequencies for nodes in the circuit. We will do these computations in the following paragraphs in order to understand if our understanding of bandwidth limits is correct.

The source follower drives the $500fF$ input capacitance of the differentiator and operates in weak inversion, so its corner frequency is computed as

$$f_{3dB} = \frac{g}{2\pi C} = \frac{I_{foll}}{2\pi C_1 U_T} = \frac{1nA}{2 \cdot \pi \cdot 500fF \cdot 25mV} = 13kHz. \qquad (6.6)$$

which is clearly not dominant.

The differentiator is an amplifier in feedback configuration with closed loop gain of about 20. Therefore its corner frequency is reduced by the closed loop gain. But its bias current is about 100 times higher than that of the source follower. Although it runs in moderate inversion, its corner frequency is unlikely to dominate given the previous computation.

Finally we come to the likely cause of the bandwidth limit, which is simply that the photoreceptor feedback amplifier does not obey the assumption taken for the validity of equation 3.12. This assumption was that τ_{out} is much smaller than τ_{in}/A in equation 3.11. We can examine this assumption again by computing τ_{out} from the known amplifier bias current I_{pr} and assumed gain A: the value of τ_{out} is given by

$$\tau_{out} = \frac{C_{out}}{g_{out}} = \frac{C_{out}A}{g_m} = \frac{C_{out}AU_T}{I_{pr}} \approx \frac{10fF \cdot 100 \cdot 25mV}{1nA} = 25\mu s \qquad (6.7)$$

equivalent to a bandwidth of about $6kHz$. Therefore we can understand the results in Figure 6.15 and Figure 6.16: At low illumination the bandwidth is dominated by the time constant of the photodiode node, while at high illumination, the bandwidth is dominated by the feedback amplifier. At the intermediate illumination level of -2 dec, the peaking is due to comparable τ_{out} and τ_{in}. In this condition, we expect from theory [42]

$$Q = \frac{sqrtA}{2} \qquad (6.8)$$

that the quality factor of the resonance should take the value $Q \approx 5$ for $A = 100$ assumed here. The resonance peak is reduced by the effect of pixel refractory periode. The larger the amplitude of the photoreceptor response is, the shorter the time between events becomes and as a result the effect of finite refractory period increases.

Figure 6.16a) shows that we can use the amplifier bias current to adjust the bandwidth. By increasing the current in M_{pr} by a factor of about 100 (from $100pA$ to $11nA$) the bandwidth expands by only about $50x$, probably because the amplifier starts to enter moderate-inversion operation.

At very low frequencies, the effect of the reset switch transistor junction leakage current, working in favor of ON and against OFF events, becomes significant. Figure 6.16b) shows events per stimulus cycle for very low stimulus frequencies between $10Hz$ and $5mHz$. The pixel under test was measured to have an average background ON-event rate of about $40mHz$ (one background event every 25 seconds) at room temperature. The BG curve shows how this background activity affects the measurement of very low frequencies. In the case

of complete absence of leakage, the usable frequency range appears to practically extend down to DC. Leakage limits the range for ON events to above $0.1Hz$ at room temperature.

6.4.4   Temporal Noise

Understanding how noise manifests itself in the vision sensor is important for practical and theoretical reasons. For applications under low illumination, where vision is photon starved, shot noise (photon counting noise) will likely dominate other noise sources. Theoretically, it is important to comprehend the noise mechanisms in this new type of vision sensor. A complete understanding of all noise sources in this vision has not been obtained yet and as we will discuss next, will not be easy to acquire. Nevertheless, we can make useful observations and reason intuitively about at least some of the phenomena.

Referring to the pixel circuit shown in Figure 3.3 on page 49, we expect multiple sources of noise:

1. shot noise in the photocurrent,

2. shot noise in the dark current,

3. shot noise in the feedback transistor M_{fb} source current,

4. amplification noise in the two amplifiers, the buffer, and the two comparators,

5. kTC reset noise in the differentiator,

6. noise in the charge injected from the reset transistor channel,

7. other noise sources such as power supply, substrate coupling, and capacitive coupling.

These noise sources have a variety of bandwidths and feed back into preceding parts of the circuit in a variety of ways. The picture as a whole looks very complicated. Despite this complexity, we hope that in analogy to many other systems, certain noise terms dominate and are not too difficult to comprehend.

Considering first the front end continuous-time logarithmic photoreceptor circuit, we know from [42, 49, 61, 62] that as long as the dominant pole is at the photodiode node, then the noise has a band-limited white spectrum with a first order lowpass power spectrum. The total noise voltage is also constant and independent of illumination as long as the illumination dominates the dark current. This total noise power is spread over a bandwidth that is determined by the effective integration time of the photoreceptor, as shown in section 3.4 on page 52. The magnitude of the total noise is given by [42, 62]

$$\Delta \bar{V}_{pr}{}^2 = \frac{kT}{C_{eff}} \tag{6.9}$$

where $\Delta \bar{V}_{pr}{}^2$ is the average mean square variation of the photoreceptor output. kT is the thermal energy and C_{eff} is the effective capacitance of the photodiode node taking into account the feedback speedup (section 6.4.3 on page 88).

This equation asserts two things:

1. that the feedback amplifier does not contribute any noise in the pass-band of the photoreceptor because its noise is spread over a much larger bandwidth, and

2. the photoreceptor effectively integrates over a constant number of photons, regardless of illumination level.

As the photocurrent becomes smaller, the noise becomes concentrated in a smaller passband and the noise per unit frequency becomes larger. Therefore, as illumination decreases, more noise will be concentrated in the passband of the pixel as a whole, and so there will be more jitter in the timing of the events that are generated.

Considering kTC noise in the differentiator reset, Fowler et al. [63] have recently reviewed methods for reset noise reduction that appear similar in style to refractory mechanism used in the pixel to reset the differentiator, where the reset switch is switched on quickly but off slowly, gradually decreasing the bandwidth of the feedback until the switch is completely off.

Fowler et al. showed in [63] that it is not possible to compute these effects using linear circuit models or models that assume thermal equilibrium, because the circuit is not in thermal equilibrium during reset. However, we can make an imprecise estimate of the magnitude of the reset noise assuming that it contributes a mean square charge variation of kTC to the floating differentiator node, where C is the total node capacitance which is almost equal to the input capacitance C_1. This fluctuation can be referred back the input through the differentiator input capacitor C_1, i. e., a charge fluctuation on $diffIn$ is equivalent to a fluctuation in the photoreceptor/buffer output voltage that produces the same voltage fluctuation on $diffIn$.

To characterize the temporal variability, we measured the single pixel responses to a sinusoidally-varying $10Hz$ LED stimulus (Figure 6.17). The events were recorded for about 30 seconds so there are 300 LED modulation cycles of data. The response consists of sets of about 4 ON and 4-5 OFF spikes that are plotted in the individual event raster plots a) and event histograms b). The events do not occur completely out of phase because the $\ln I$ variation is strongest near the minimum of the sinus cycle. The variations in the event times as measured by the full width at half maximum of the larger histogram peaks is about $1ms$ in this recording, but depends on the stimulus frequency, contrast and luminance as well as on the pixel contrast threshold $C_{contrast}$. In addition, the variability becomes larger for the trailing ON events (indicated with an arrow), which are just at threshold and therefore show the largest influence of noise. For these events, the variability is increased to about $3ms$.

In the next measurement with the same stimulus we used neutral density filters to attenuate the LED luminance over 5 decades. Figure 6.17c) summarizes these measurements. The number of events produced per stimulus cycle hardly changes over 4 decades but starts to decrease rapidly at 4 decades attenuation, suggesting that the effective contrast is being reduced by dark current. The variability in the number of events produced per cycle remains almost constant over luminance, suggesting that this variability arises mainly from other sources.

The temporal resolution or "sharpness" of the response decreases slightly with increasing luminance. Figure 6.17c) shows the jitter (black markers), measured as the full width at half maximum of the largest histogram peak in each plot. It seems evident that the jitter goes reciprocally with the magnitude of the
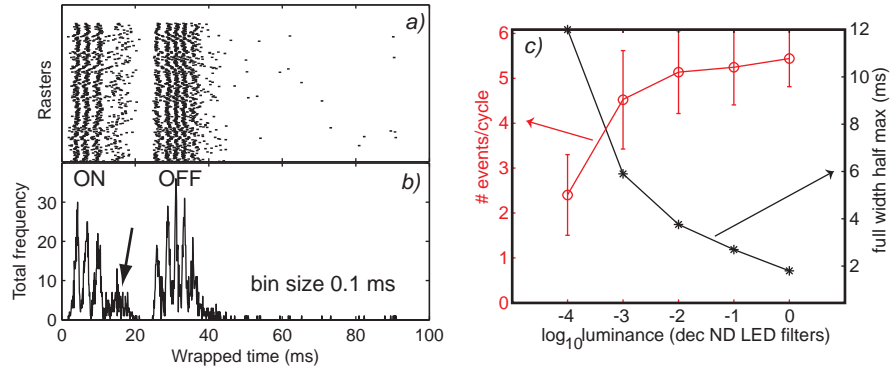
Figure 6.17: **Shows the event raster plots (in a)) and the event histograms (in b)) of the responses of a single pixel to a sinusoidal $10Hz$ LED stimulation with a full contrast of $72/21$. Part c) shows the summary of the measurements over 5 decades with the $10Hz$ sinusoidal stimulus. The total number of events per cycle is plotted using red markers and can be read out with the left y-axis. The full width half maximum variation measure is plotted with black markers with its corresponding y-axis on the right.**

response, but this could be related and not causal. We observed that the jitter in event timing depends on the absolute rate of change of logarithmic intensity much more than on the frequency of stimulation. When we reduced the LED frequency to $1Hz$, we observed a jitter in spike timing at -2 decade luminance of about $80ms$, representing, roughly, a doubled fractional jitter. The jitter is greatly reduced by using a square wave stimulus, as discussed in the next section.

In summary, the temporal noise mostly affects the precise timing of the events and not their production. Noise goes monotonically inverse with illumination, suggesting that the main source of noise is at the front end of the photoreceptor. This total shot noise is roughly constant with illumination because the integration time scales inverse with illumination, but as illumination decreases, more noise appears in the pixel passband for the event production. However, measurements suggest that for most operating conditions, the noise bandwidth is not dominated by the front end photoreceptor but by later parts of the pixel circuit. Thus, noise as expressed by event timing jitter is only a weak function of illumination.

6.4.5   Latency and Latency Jitter

Sensor latency is important for vision systems requiring quick answers, for instance, when the sensor output is used as part of a feedback controller. The latency of the sensor response (meaning the time it takes from the occurrence of the illumination change at the photodiode to the output of an address event) arises from contributions by the pixel, the asynchronous bus arbiter and the rest of the AER periphery. The total AER latency is estimated to be of the order of some hundred ns to a few μs. The variable latency from the pixel is, at least at low illumination, dominated by the photoreceptor and as such is a function of the photoreceptor bandwidth. As derived in section 3.4 on page 52, the pho-

toreceptor bandwidth at low illumination levels is proportional to illumination. However, this assumption is only true if the photoreceptor feedforward amplifier is much faster than the photodiode node, which is typically not true except at the lowest illumination levels. Therefore a basic prediction is that the latency should increase when we decrease the illumination level, and the increase should be proportional to reciprocal illumination at very low illumination that is still larger than the dark current level.
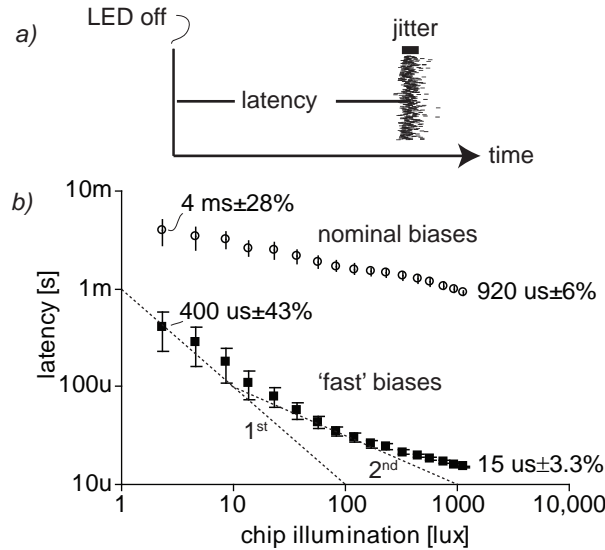
Figure 6.18: **Shows the sensor latency and the latency jitter (error bars) versus illumination in response to a 30% step increase of the single pixel illumination. Part a) shows the measurement of repeated single event responses to the step (jitter is shown by the error bars). Part b) shows the results with two different bias settings. The chip illumination means illumination of the sensor plane (the silicon) and not scene illumination.**

The latency was measured using a low-contrast (30%) periodic $10Hz$ step stimulus at variable DC luminance (Figure 6.18). The thresholds were set to produce exactly 1 event of each polarity per stimulus cycle. The overall latency is plotted versus stimulus chip illuminance for two sets of measurements, one at the nominal biases, the other at higher current levels for the M_{pr} photoreceptor and M_{b2} source follower biases (Figure 3.3 on page 49). The plots show the measured latency and the 1-sigma response jitter. The dashed lines show a reciprocal (1st) and reciprocal-square-root (2nd) relationship between latency and illumination.

The most interesting aspects of this data are the following:

1. The minimum latency is only $15\mu s$, representing an effective single pixel bandwidth of about $66kHz$.

2. The latency is a very soft function of photocurrent. It is reciprocal with illuminanceonly only at the very low illuminance.

3. At the nominal biases that we generally use, the latency is still only $4ms$ at the lowest illuminance of a few lux.

4. The jitter in the step response is a small fraction of the latency regardless of illuminance.

The latency only changes very slowly with illuminance for the nominal biases because other mechanisms besides photoreceptor bandwidth limit the latency. In summary, this vision sensor's low latency makes it attractive for real-time control systems.

6.4.6   AE Bus Bandwidth

The AE transmission cycle time has a lower bound of about $1.2\mu s$ using the integrated bias generator with an external $10k$Ohm resistor, which sets the master current to about 40μA. This choice is optimal for normal operation because it still allows setting acceptably small bias currents for situations when we want make the pixels slow. However, this cycle time limits the maximum event output rate with our USB2.0 connection to about $600keps$, forming the major limitation to high speed operation for this vision sensor.

As long as the event rate is below the bus saturation limit by a factor of about 0.8, the output events are accurately timed. Above this limit, the vision sensor enters a self-scanned serial mode where all active pixels are sequentially read out. This happens because the AE circuits are not greedy but serve a certain row and column only once before other waiting rows and columns are served. As a result all pixels that want to send events out get to send them out, but their exact timing accuracy is lost and reduced to the effective rate at which all active pixel can be read out. The transition between accurately-timed events and scanned operation is very sharp.

6.4.7   Temperature Dependence

The effect of temperature variation is important for real world operation, especially in the variable weather conditions experienced in the outdoor unconditioned environments in which this type of vision sensor is expected to be used. Here we will consider the effect of temperature first from the theory of pixel operation and then characterize it experimentally.

Over the range of $-20\,^{\circ}$C to $+60\,^{\circ}$C the absolute temperature increases by a factor of only 1.3, so effects that are linear in absolute temperature will have a small effect on the AER temporal contrast vision sensor operation. Effects that are exponential can have a huge effect over this range. We can expect that temperature will affect the operation in the following ways, listed approximately in order of decreasing importance.

1. Temperature will dramatically affect the photodiode dark current (which is a leakage current). The physics underlying this is that increased temperature increases the average carrier energy, both elevating the carrier concentration and the energy that carriers have to affect generation processes. Leakage current reduces the contrast of real illumination signals, reducing sensitivity at low illumination. Furthermore, leakage increases shot noise at low illumination, increasing the jitter in event timing. Dark current
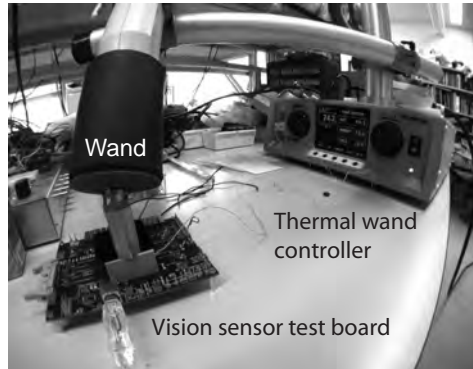
Figure 6.19: **Temperature Dependence Measurement Setup.**

is known from the literature to change exponentially with temperature, doubling about every $6-8\,°\text{K}$ [64]. If we take $7\,°\text{K}$ as typical, then we can expect a increase of about $50\times$ from room temperature to $+60\,°\text{C}$.

2. In a similar way temperature will also affect the junction leakage in the differentiator reset transistor M_r and the global reset transistor M_{gr} (Figure 3.3 on page 49). Hence we expect that the background (BG) ON activity doubles every $6-8\,°\text{K}$ increase in temperature. At room temperature we typically observe BG activity of about $40mHz/pixel$. At $+60\,°\text{C}$ this BG activity can increase by a factor of 50 to $2Hz$, resulting in a significant chip-level BG activity of about $33keps$.

3. Temperature will also affect the overall gain of the system through the initial phototransduction. This has a gain that is proportional to $U_T = kT/q$, so the gain should increase proportional to absolute temperature (PTAT), i. e. it should change by about 30% over the total range of temperature.

4. Temperature will also variably affect the bias currents. In general the integrated bias generator is a PTAT-like source. Specifically, the bias generator generates a constant g_m current for the master current. The most important effect of increased temperature on bias currents is that the minimum possible current that can be distributed by mirroring will increase [57]. At high temperature, the present implementation of bias generator will not be able to distribute a bias voltage that generates a very small current. Hence the longer time constants will decrease. A secondary effect of changing temperature is that the bias currents will scale differentially with temperature because they are all scaled copies of a single strong inversion level master current. While the large currents will not change much, the small currents will not change as much as they should, i. e., they will scale as less than PTAT. Therefore the g_m's of the low-biased amplifiers will tend to decrease with temperature.

To control device temperature we used a thermal wand[3] to control the package temperature. This bench-top device is very convenient because it can put a

---

[3]Temptronic ThemoStream, www.temptronic.com

device at a known temperature in a feedback loop using a thermocouple sensor and it only requires a source of compressed air. It has the drawback that going below freezing without condensation requires a source of dried air, which we did not have available. Therefore we only investigated temperature from about $+10\,^{\circ}\mathrm{C}$ to $+70\,^{\circ}\mathrm{C}$ in the following experiments.
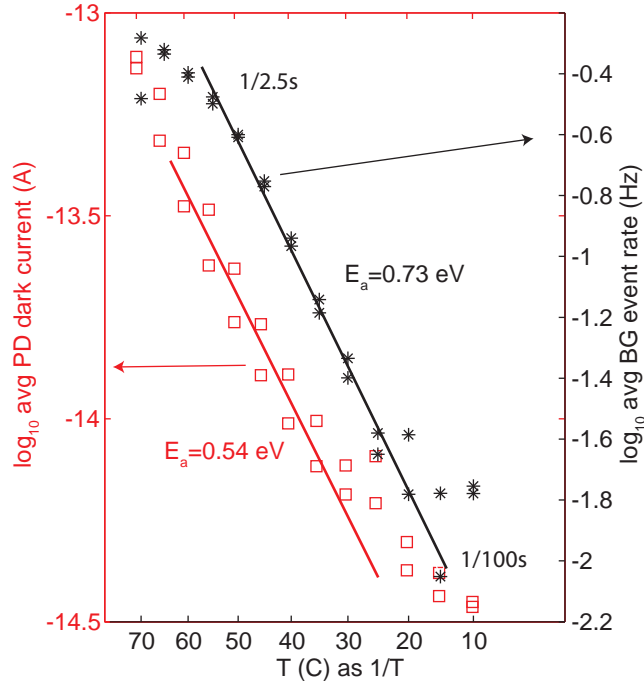
Figure 6.20: **Shows the logarithm of the BG activity and the logarithm of the photodiode dark current as a function of** $1/T$. **The fit indicates an underlying process with activation energies of about** $0.73eV$ **for the BG activity and** $0.54eV$ **for the photodiode dark current. The error bars show a standard error of measured event rate. The temperature was swept up and back down to ensure repeatability and soak-in. Higher points at lowest temperatures may be the result of condensation on the chip die or the package. The device was allowed to stabilize (soaked) for 120 seconds after reaching the set temperature. The two points at each temperature are the results of a sweep up and down in temperature.**

Figure 6.20 shows the steady-state background activity and the photodiode dark current as a function of temperature. The photodiode dark current is measured from the global node *photocurrent* (Figure 3.3 on page 49). Junction leakage is often fitted by the function

$$I_{leak} \propto \exp\left(\frac{-E_a}{kT}\right) \tag{6.10}$$

where I_{leak} is the leakage current, E_a is an activation energy, k is Boltzman's constant and T is absolute temperature. This activation energy is related to the energy barrier. Typically E_a has a value of ranging from $0.5 - 1eV$. Given

this range of E_a, around room temperature the current doubles about every $6 - 8\,°$K . Plotting I_{leak} on a log scale versus $1/T$ can give a hint as to whether the underlying process is diffusion or space-charge dominated. Specifically, from equation 6.10 we get

$$\ln I = const - \frac{E_a}{kT} \tag{6.11}$$

 so the slope of the logartihm of the background activity plotted versus $1/T$ where T is absolute temperature should yield E_a. The measured data (Figure 6.20) shows that this model is valid over a range extending from about $15 - 50\,°$C, and the slope of a linear fit indicates a value $E_a = 0.73eV$ for the BG activity and $0.54eV$ for the photodiode dark current.

This result is often 'linearized' around room temperature to indicate how much the effect changes per degree and how many degrees result in doubling. This comes from equation 6.11

$$\begin{aligned}
\frac{d\ln X}{dT} &= \frac{E_a}{kT^2} \\
&= \frac{q \cdot 0.73V}{1.38 \cdot 10^{23}J/\,°K \cdot (300\,°K)^2} = 9.4\% \text{ per } °K \\
&\rightarrow \quad 7.4\,°K \text{ to double} \tag{6.12}
\end{aligned}$$

 Above $50\,°$C the activity stops increasing so rapidly with temperature but this is a measurement artifact. At $60\,°$C pixels spontaneously spike every $2.5s$, which is still acceptably low for many tasks.

We now consider the dark current and what it implies about minimum scene illumination. The individual photodiode dark current at room temperature is about $4fA$, consistent with results from earlier silicon [60] (see also Figure 4.6 on page 60). Using the fact that for white light $1lux = 10^4\,$photons$/\mu m^2/s$ [65], a quantum efficiency of 50%, and a pixel area of $151\mu m^2$, we compute an equivalent pixel illumination of about $33mlux$.
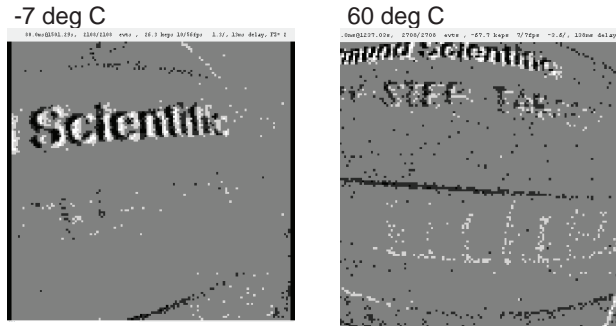

Figure 6.21: $80ms$ **snapshots of events collected at low and high temperature. Bias settings were left untouched. Full scale is 2 events.**

Using the geometric relation where L_{chip} is the chip illumination, L_{scene} is the scene illumination, f is the lens aperture ratio (assumed $f/1.2$, see appendix B.2 on page 115), and assuming scene reflectance of 20%, this pixel level

dark current is computed to be equivalent to a scene illumination of about 60 times higher illumination, or 2 lux. Thus the dark current at room temperature cuts the contrast in half at about 20 times 0.1 lux full moonlight level [65]. This indirect computation is consistent with our practical experience-the vision sensor is usable under street lighting but not under full moonlight. (Moonlight is also blue-ish with lower quantum efficiency.) Using a low leakage imager process would extend the dynamic range by a factor of 100. The dynamic range is cut in half approximately every $8\,°C$ increase in temperature.

Figure 6.21 shows snapshots of images of the Edmund density step chart (compare with Figure 6.13 on page 87) collected at low and high temperature under moderate indoor lighting (white surface luminance of about $15nit$), holding all bias current settings the same. There are no glaring differences between the images. At $60\,°C$ the background activity rate increases to about $20keps$. At low temperature it is possible to set lower thresholds without creating massive background activity.

We have used this sensor in a wide variety of conditions spanning a hot summer and a cold winter. There are no wildly obvious changes that render the device unusable. The most important effects-change in background activity and reduced dynamic range-will matter in some applications, but dark current will be greatly reduced by an optimized process.

6.4.8   Power Consumption

Measured current consumption of the vision sensor chip, running with $3.3V$ supply, using our typical bias generator settings is $7 \pm 0.4mA$, so typical chip-level power consumption is about $23mW$. This supply current is dominated by the analog core supply, which accounts for more than 90% of the power. The rest of the supply current ($\approx 300 - 600nA$) is supplied to the digital parts of the pixel, the AER periphery, and the pad frame. Power consumption depends weakly on activity, increasing about 10% with high activity.

About 80% of the power is burned by the integrated bias generator ($5.5mA$). This is only because of over design of the bias buffers. The remaining supply to the core ($1.5mA$) is divided into $16k$ pixels that each burn $93nA$. The programmed pixel supply is dominated by the differentiator and comparator supplies. When the differentiator is balanced it burns a programmed current $I_d = 34nA$. The ON and the OFF comparator burn each I_d. The total differentiator plus comparator current that we program is computed to be $136nA$ which is quite close to the current inferred from measurement.

In summary, substantial power reduction could be obtained by optimizing the bias generator. For example, if we could reduce the bias generator supply to $0.5mA$ then the total current consumption would be reduced to less than $3mA$. We can already realize substantial power reduction simply by running the entire device more slowly (reducing the master current I_m), but this reduces the maximum event rate. This reduction could be beneficial in situations where peak activity is expected to be rather low.
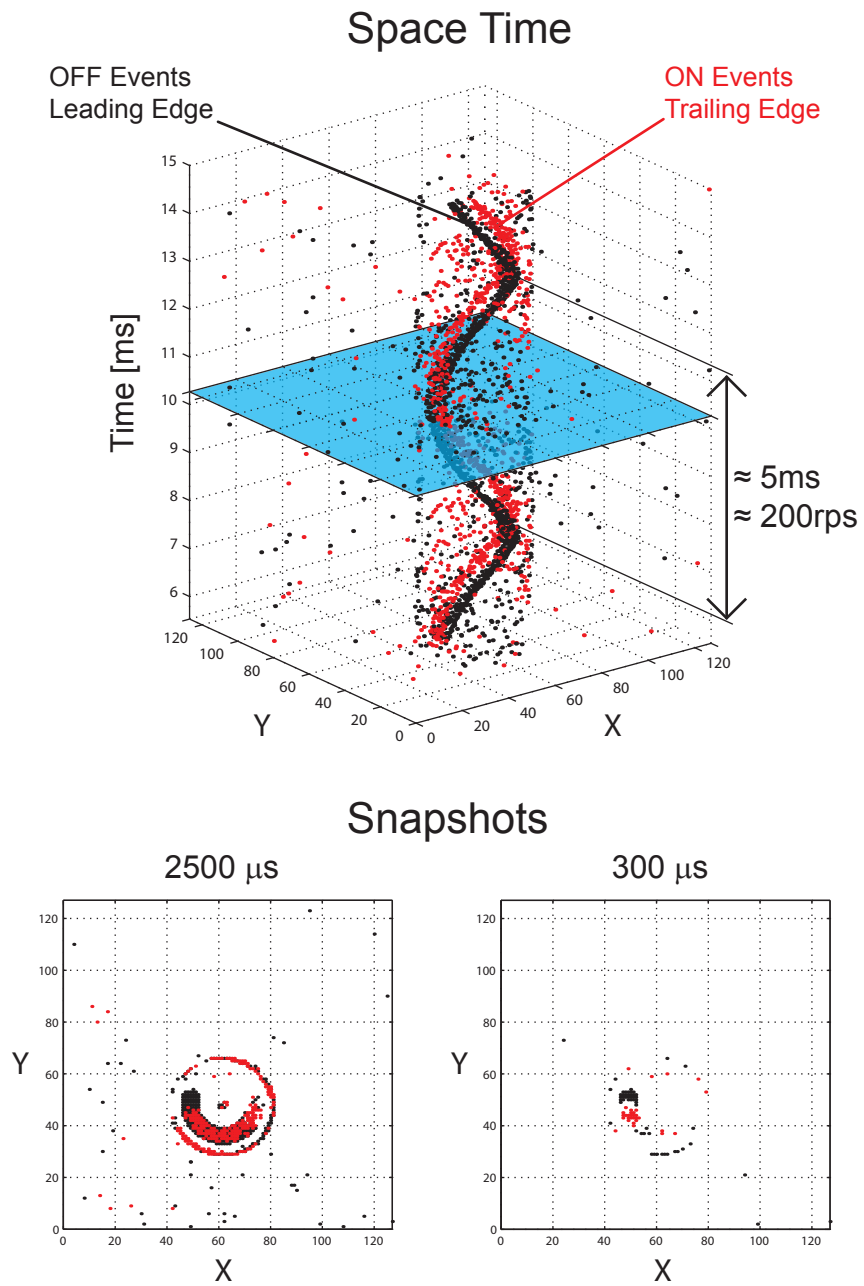
Figure 6.22: **Response of the vision sensor to a fast rotating disk with a black dot.**

Pixel size	$40 \times 40 \ \mu m^2$
Pixel complexity	28 transistors (16 analog), 3 caps
Array size	128×128
Chip size	$6 \times 6.3 \ mm^2$
Chip supply current	$7 \pm 0.4 mA$ at $3.3V$
Fabrication process	4M 2P $0.35 \mu m$
Biasing requirements	12 generated gate biases
Fill factor	8.1% (PD area $130 \mu m^2$)
Dynamic Range	$120 dB$
Photodiode dark current	(estimated) 20fA ($\approx 10 nA/cm^2$)
Burst event rate	$\approx 600 kevent/sec$

Table 6.1: Specifications of the AER temporal contrast vision sensor.

6.5   Example Data

6.6   Discussion

The AER temporal contrast vision sensor achieves 120 dB intra-scene dynamic range in an non-optimized (non-imager) process, making it usable under uncontrolled illumination conditions. The pixel contrast mismatch referred to the input (log intensity) is 2.1% and the best uniformity we ever achieved with this type of sensor. The latency of the response is a soft function of photocurrent and is typically around $1ms$ using nominal biases and indoor illumination. Fast biases and higher illumination allow latencies as short as $15\mu s$, which corresponds to the inter frame time of a $66kfps$ high-speed imager. Power consumption is $23mW$. The vision sensor also integrates a programmable bias generator that allows temperature-independent and process-independent operation and can be used for the dynamic control of operating parameters. Table **??** lists all the specifications of the temporal contrast vision sensor.

Applications areas for this vision sensor include high-speed low bandwidth imaging, wireless sensor networks, surveillance and traffic monitoring under uncontrolled lighting conditions, industrial vision for manufacturing or inspection, autonomous navigation systems (i. e. lane finding, flying vehicles), human interface devices (i. e. eye-trackers), and visual prosthetics.

Chapter 7

Conclusion

7.1 Retrospection

We start this chapter with a brief retrospection of the our work progress. Ensuing we will review some of the applications in which our sensors served as input device. Finally we will compare our latest vision sensor to other state-of-the-art vision sensors, outline the advantages of our sensor and give a short list of possible future improvements to our design.

7.1.1 Starting With Jörg Kramer's Retina
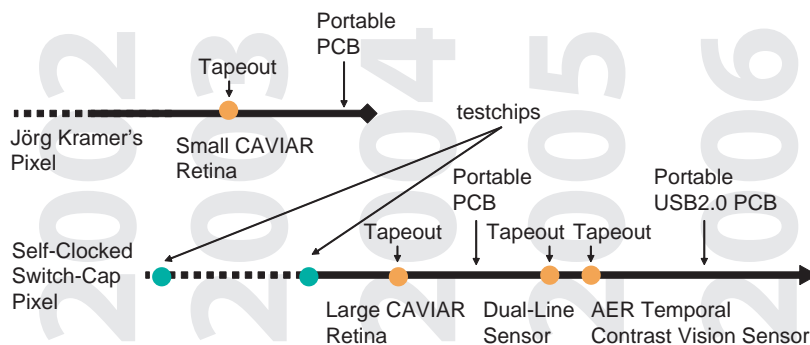

Figure 7.1: **Shows all the chips we taped out during this project.**

This project was started by Jörg Kramer. He presented the first successful implementation of an AER transient vision sensor in 2002 [44] that he based

103

on his integrated optical transient sensor [45]. We followed the path that Jörg Kramer has lain out and taped out the small CAVIAR retina chip in 2003. This chip enhanced Jörg Kramer's design in respect to the symmetry of the response between ON and OFF events, the dynamic range and the uniformity of the pixel responses at the prize of a larger pixel size. Our own implementation of Jörg Kramer's design did teach us some basic limitations of this design. In response to this realization we decided to put our confidence in a new design that resulted from different design trials we made on two test-chips. We implemented the new pixel, which we call the self-clocked switch-cap pixel in a 64×64 vision sensor. The tape-out of this chip was exactly one year after the tape-out of the small CAVIAR retina. This 64×64 sensor, which became the large CAVIAR retina, exceeded our expectations and the new pixel design became the design of our choice. The last two vision sensor chips we got manufactured were the dual-line sensor and the AER temporal contrast vision sensor. Both of these designs were sent for fabrication in spring 2005. Figure 7.1 shows the timetable of chip manufactoring.

For all of the four major chips (small CAVIAR retina, large CAVIAR retina, dual-line sensor and AER temporal contrast vision sensor) we designed one or more PCBs. The PCBs, although less laborious to design, are very important for a successful operation of the integrated circuits, especially if one desires to use his integrated circuits away from the lab bench. Figure 7.2 shows all the different PCBs and chips we developed for this project during the last 4 years.

The integrated circuits on the PCBs made camera systems that could be used for sensor characterization and also as input devices to AER receiver and transceiver chips, to AER to computer interfaces or, if equipped as a USB vision sensor, for computer based vision processing.

7.2   Applications

Ultimately the success of a sensor will be determined by its application. In the development and evolution of a novel sensor design, the application may not be the immediate concern, nevertheless it is always the final destination. In this chapter we will shortly discuss some of the applications.

7.2.1   The Orientation Selectivity Experiment

One type of applications of neuromorphic devices is to use them as a platform to study neuronal processing principles. Together with Elisabetta Chicca, Giacomo Indiveri and Rodney Douglas, we did use the small CAVIAR retina in an experiment to study orientation selectivity [66]. The following paragraphs describe this experiment. The following passages are adapted from our publication [66].

A regular and similar structure of neuronal connections within each cortical area suggests that the cortex may use a common core processing circuit, or *canonical microcircuit*, that can be tuned to perform specific tasks [67]. The canonical microcircuit, and its later extensions, emphasize the role of first order recurrent connections between cortical neurons. These recurrent connections between thresholded neurons support cooperative-competitive processing, in which networks of neurons participate collectively in the generation of an appropriate

AER Temporal Contrast Vision Sensor USB2.0 Camera Systems

AER Temporal Contrast Vision Sensor Test Boards

Dual-Line Sensor Board

Large CAVIAR Retina Test Board

All Chips

DAC Board

Large CAVIAR Retina Board

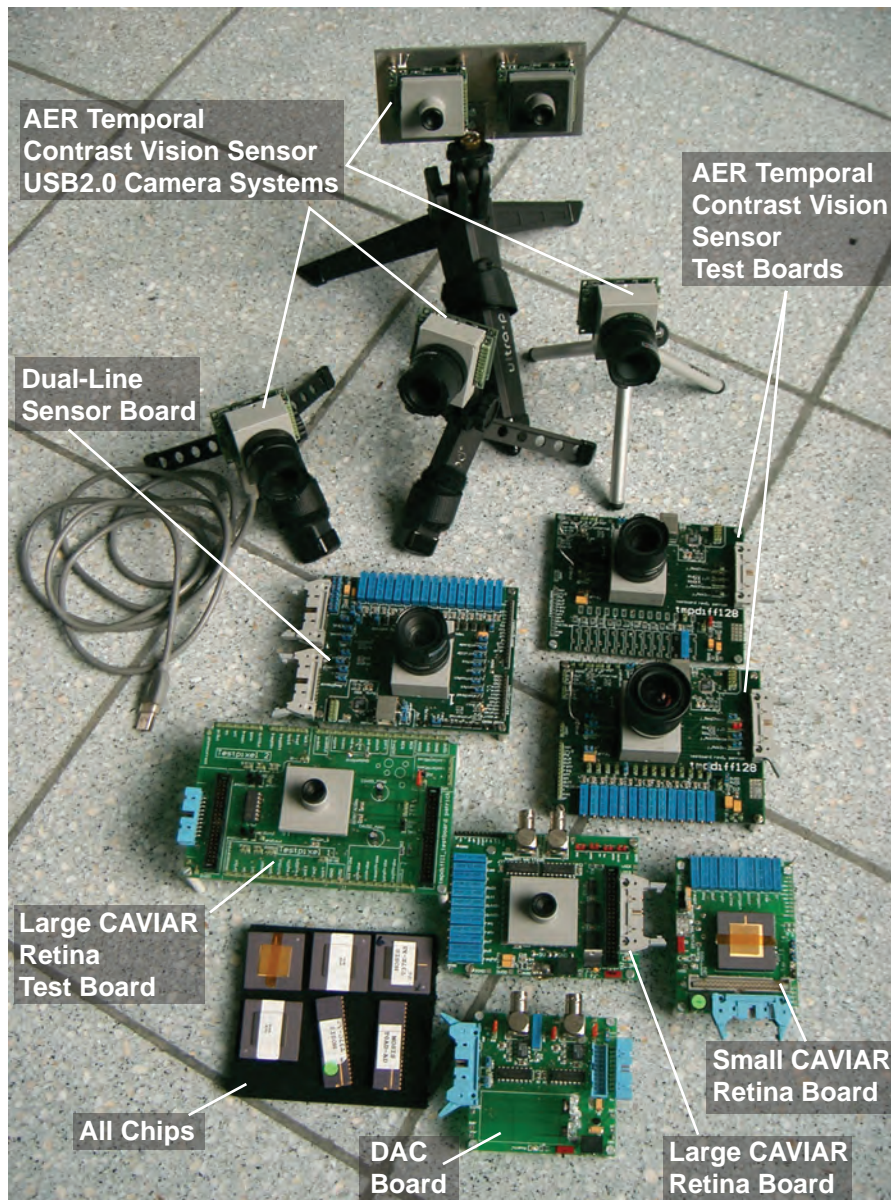**Small CAVIAR Retina Board**

Figure 7.2: **Shows all the PCBs we developed during this project.**

interpretation of their sensory input. The output of a given neuron depends dynamically on the activity of all neurons in the network. As a result, these networks are able to perform complicated non-linear operations, such as the winner-take-all function. The computational abilities arising from competition are especially useful for feature extraction and pattern classification problems.

Several theoretical recurrent models have been proposed to explain the origin of orientation selectivity in primary visual cortex and as a means to understand cortical circuitry and cortical computation (for a review see [68]).

To implement orientation selectivity we used a VLSI recurrent network of spiking neurons and interfaced it to the small CAVIAR retina. The pattern of connectivity described in these feed–back models is reflected in the local recurrent connections of our VLSI network of spiking neurons: similar orientation cooperate through lateral excitation and different orientations compete through global inhibition.

Several hardware models of orientation selectivity have been proposed in the past [48, 69–72]. Our approach differs from those of from all previous hardware models because it decouples the sensing stage from the computational stage, is based on collective dynamics of populations of neurons and is has different populations of neurons selective to different orientations.

The orientation selectivity system components

The system consists of two neuromorphic VLSI chips, a PCI-AER board [73] and supporting hardware (see Fig. 7.3). The neuromorphic chips are the small CAVIAR retina and a recurrent competitive network of integrate-and-fire neurons and dynamic synapses (IFWTA chip). The PCI-AER board is a custom device which provides a communication bridge between the AER and the PCI bus of a host computer. It supports a real-time routing, programmable connectivity, and monitoring and stimulation of address events. The supporting hardware comprises a custom Digital to Analog Converter (DAC) board [74] for setting the analog biases of the neuromorphic chips, an LCD screen for presenting visual stimuli, and a workstation for hosting and controlling the PCI-AER board, programming the DAC board and controlling the LCD screen.

The IFWTA chip

The architecture of the IFWTA chip is shown in Fig. 7.4. It is a two-dimensional array containing a row of 32 Integrate-and-Fire (I&F) neurons, each connected to a column of afferent synaptic circuits. Each column contains 14 AER excitatory synapses, 2 AER inhibitory synapses and 6 locally connected (hard-wired) synapses. When an address-event is received, the synapse with the corresponding row and column address is stimulated. If the address-events routed to the neuron integrate to the neuron's voltage threshold for spiking, then that neuron generates an address-event which is transmitted off-chip. The AER input synapses can be used to implement arbitrary network architectures, by (re)mapping address-events via the PCI-AER board.

Synapses with local hard-wired connectivity are used to realize a competitive soft winner-take-all (WTA) network with nearest neighbor and second nearest neighbor recurrent interactions (see Fig. 7.4 and Fig. 7.5): 31 neurons of the array send their spikes to 31 local excitatory synapses on the global inhibitory
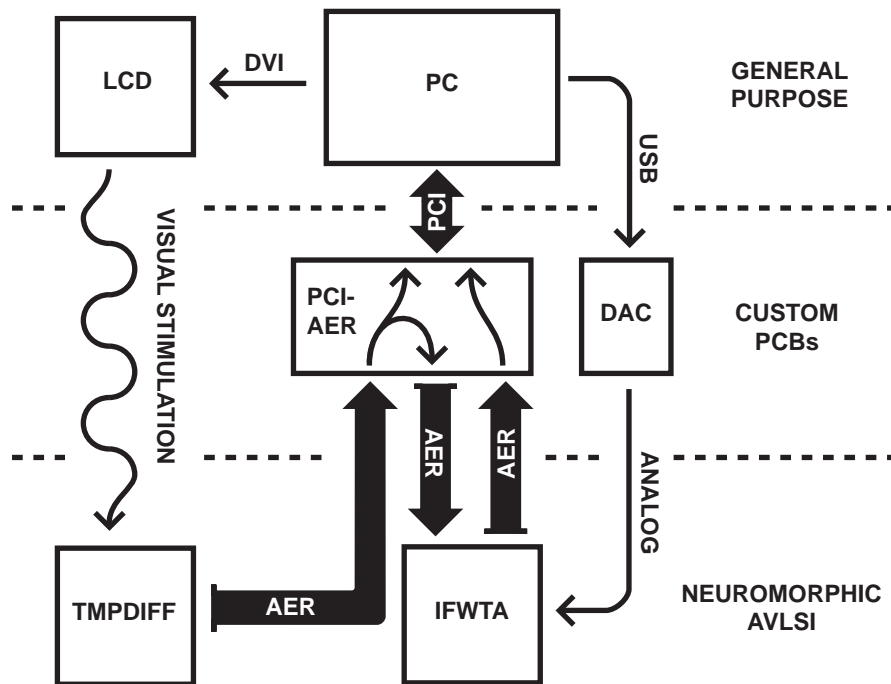
Figure 7.3: **AER vision system setup. The PCI-AER board routes output events of the TMPDIFF chip in response to visual stimuli to the IFWTA chip and monitors the activity of both chips. The PC controls the LCD screen for stimulus presentation, the PCI-AER board and the DAC board. This figure is adapted from [66].**
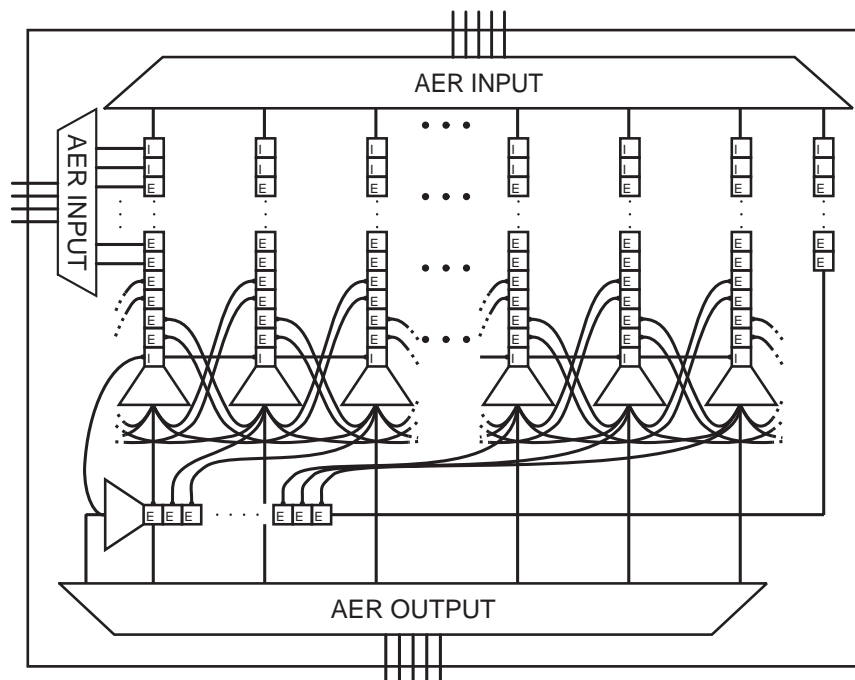
Figure 7.4: **Chip architecture. Squares represent excitatory (E) and inhibitory (I) synapses, trapezoids represent I&F neurons. The I&F neurons can transmit their spikes off-chip and/or to locally connected synapses. This figure is adapted from [66].**

neuron; the inhibitory neuron, in turn, stimulates the local inhibitory synapses of the 31 excitatory neurons; each excitatory neuron stimulates its first and second neighbors on both sides using two sets of locally connected synapses. The first and second neighbor connections of the neurons at the edges of the array are connected to pads. This allows us to leave the network open, or implement closed boundary conditions (to form a ring of neurons [75]), using off-chip jumpers. A detailed description of the IFWTA chip was presented in [76].
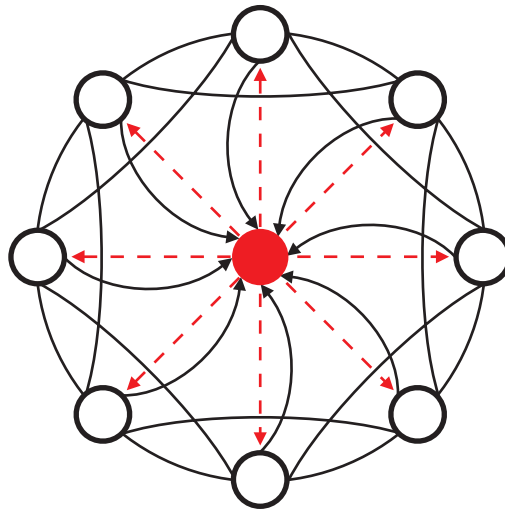
Figure 7.5: **Schematic representation of the connectivity pattern implemented by the internal hard-wired connections (closed boundary condition). Empty circles represent excitatory neurons and the filled circle represents the global inhibitory neuron. Solid/dashed lines represent excitatory/inhibitory connections. Connections with arrowheads are monodirectional, all the others are bidirectional. This figure is adapted from [66].**

Experimental Results

Orientation selectivity is achieved by appropriately mapping feed-forward connections from the small CAVIAR retina pixels to the IFWTA chip neurons (via the PCI-AER board), and by activating the local recurrent connections on the IFWTA chip. The feed-forward mapping is set so that each IFWTA neuron is excited by all the pixels of the TMPDIFF chip belonging to a central bar with a specific orientation. We implemented 31 sets of mapping tables that map 31 differently oriented bars onto the 31 excitatory neurons of the IFWTA chip.

In our experiments we displayed to the small CAVIAR chip flashing oriented white bars on a dark background using an LCD display. The activity of the retina chip was monitored by the PCI-AER board and transmitted (via the PCI-AER board mapping tables) to the IFWTA chip.

We repeated the same experiment for two different conditions, in terms of the local connectivity of the IFWTA chip. In the first condition the IFWTA chip was tuned to implement a purely feed-forward model. Subsequently, we ac-

tivated the recurrent connectivity to implement the feed-back model. Three sets
of local synapses were used: first neighbor excitatory to excitatory synapses to
simulate the mutually excitatory connections among cells with similar preferred
orientation, inhibitory and excitatory synapses connecting the global inhibitory
neuron to the excitatory neurons and vice versa to simulate the mutual inhibi-
tion among cells with different preferred orientation.

The recorded activity of the IFWTA neurons was used to compute the mean
firing rate of each neuron in response to the stimuli and tuning curves were
obtained by plotting these data for each neuron as a function of orientation (see
Fig. 7.6).

The TMPDIFF central pixels are mapped to all neurons, therefore each
IFWTA neuron is also receiving input events when its non-preferred orientation
is presented to the retina. The effect of this "base line" input is clearly visible in
the feed-forward model, where the activity of the IFWTA neurons simply reflect
the input from the retina. In this case, the frequencies in the tuning curves are
greater than zero for all orientation and a maximum is observed at the preferred
orientation. In the feed-back model the "base line" activity is suppressed and
the activity in response to the preferred orientation is amplified.

With this experiment we have been able to reproduce the orientation se-
lectivity found in biological vision systems. The succesful implementation of
such an a multi-chip system shows that neuromorphic AER chips, that were
not specifically designed for this experiment can be modularly added together
to implement a certain functionality. The success of such systems is made pos-
sible by the sophisticated supporting PCI-AER board [73]. On the other hand,
this type of application is in direct competition with numerical simulation on
conventional computing hardware.

7.2.2 CAVIAR

The recent efforts in multi-chip systems has lead to larger systems that are
flexible due to their modular structure. These systems can achieve significant
computation in real time. Our sensor serves as the front-end of CAVIAR, which
is such a multi-chip system. In the context of CAVIAR, our small CAVIAR
retina, the large CAVIAR retina and also the AER temporal contrast vision
sensor was successfully used. We did distribute a camera system comprising the
small CAVIAR retina and all necessary infrastructure to operate and interface
this retina to our partners at the beginning of 2004. The small CAVIAR retina
was used by our partners to interface it to their AER components. In 2005
the large CAVIAR retina did replace the small CAVIAR retina and was used
in a first successful assembly of the CAVIAR system. In the following sections
we will describe the application of the large CAVIAR retina in the assembled
CAVIAR system [?].

The CAVIAR Components

During the development of the large CAVIAR retina, the Institute of Neuroin-
formatics entered a loose partnership with the Information Technologies Divi-
sion of the Austrian Research Centers Seibersdorf (ARCS)[1]. ARCS, encouraged

---

[1] ARC Seibersdorf research GmbH, Tech Gate Vienna, Donau-City-Str.1, A-1220, Austria,
http://www.arcs.ac.at/
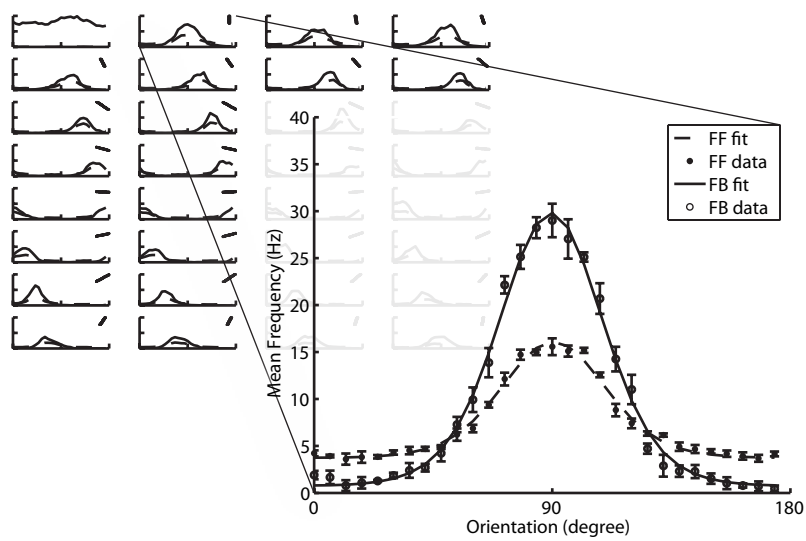
Figure 7.6: **Tuning curves for the feed-forward (dashed line) and the feed-back (solid line) model of orientation selectivity. The mean frequency (Hz) of each neuron is plotted as a function of stimulus orientation. The top left graph shows the activity of the inhibitory neuron, the other graphs show the activity of the excitatory neurons (a bar representing the retinal pixels mapped to the neuron, i.e. its preferred orientation, is shown in each plot). This figure is adapted from [66].**

by the results from our large CAVIAR retina, started a considerable effort in integrating our vision sensor into their *SMART-EYE* vision system[2].

With the development of the USB2.0 camera system for the AER temporal contrast vision sensor, a new interesting topic emerged; Real-time event based processing of AER events on a computer. Tobi Delbrück developed a sophisticated JAVA software architecture to capture, display and process address event packages delivered from the USB2.0 camera system in real time. Do to high-speed capabilities, the high compression efficiency and the asynchronous output of the AER temporal contrast vision sensor and a performance optimized software architecture, the camera-USB2.0-PC system can directly be applied to some high-speed vision applications.

In the coming sections, we will, in a very compact way, present the orientation tuning experiment, some CAVIAR results and an experiment performed with a fruit fly.

7.2.3 The Orientation Selectivity Experiment

The main achievement of this work is the implementation of a high-quality frame-free transient vision sensor that represents a concrete step towards solving vision problems in the event-based, data-driven, redundancy-reducing style of computation that underlies the power of biological vision. This sensor responds to relative changes in intensity. This response property discards most illuminant information, leaving precisely timed information about object and image motion. This information is useful for dynamic vision problems. The pixel design uses a novel combination of continuous and discrete time operation, where the timing is self-generated. The use of self-timed switched capacitor architecture leads to well-matched pixel response properties and fast, wide dynamic range operation. Table 1 shows the key performance metrics and compares them with other work. The vision sensor achieves 120 dB intra-scene dynamic range in an non-optimized (non-imager) process, making it usable under uncontrolled conditions. The pixel contrast mismatch referred to the input (log intensity) is 2.1In addition, we developed new techniques for characterizing this sensor, including metrics for matching, for pixel bandwidth, and for pixel latency. We characterized the sensor for these metrics over wide illumination range. The main areas that could benefit from improvement are as follows: The AER bus bandwidth limits high speed imaging for 'busy' scenes. The chip should be built in a low leakage imager process. The bias generator should be modified to reduce its power consumption and to allow for smaller bias currents in order to limit bandwidth when desired. An integrated means for measuring average scene brightness would be beneficial for automatic bias control. It remains to be seen how much lack of any DC response hinders application. Applications areas for this vision sensor include high-speed low bandwidth imaging, surveillance and traffic monitoring under uncontrolled lighting conditions, wireless sensor networks, industrial vision for manufacturing or inspection, autonomous navigation systems (e.g lane finding, flying vehicles), human interface devices (e.g. eye-trackers), and visual prosthetics. The processing of this vision sensor's output for vision is beyond the scope of this journal [43, 47, 48].

blabla

---

[2]http://www.smart-systems.at/downloads/highlights_06_ITN_tds.pdf

7.3 Impact

blabla

Appendix A

Abbreviations and Mathematical Conventions

Throughout the thesis, we use the following mathematical symbols:

CCD	kjl
ν	spike rate of Poisson frequency
V	Voltage or membrane potential
p	Probability distribution, transition probability in a Markov chain
P	Probability, with the special case:
$P(\mu, n)$	Poisson distribution
t,T	time or time interval
n, m, p, i	number or index of spikes
N, k	number or index of neurons
$\hat{d}$	difference of the center of two Gauss distributions with $\sigma = 1$

The following abbreviations are used:

Caviar	Convolution AER (Address-Event-Representation) Vision Architecture for Real-time, EC 5th framework project IST-2001-34124
VLSI	Very Large Scale Integration
aVLSI	analog Very Large Scale Integration
CCD	Charge-Coupled Device
BNC	Bayonet Niell-Concelman Connector
Matlab	mathematical program of Mathworks, Inc.

Appendix B

Optics

B.1 Optical Components

Cold Mirror Has a dielectric coating that is designed to reflect the visible region of the spectrum and transmit the infrared. Uses include specimen illumination, microscopy and any application where heat needs to be removed from the reflected beam.

B.2 Lens Systems

A vision sensor is usually preceded by a lens system. It is critical to join a vision sensor with optics that match the sensor well. The most important terms are explained in the following glossary.

B.2.1 Glossary of Terms

This glossary list the most important terms for lens systems. Most of this terms are used throughout the thesis. The glossary was assembled from

http://www.pentax.co.uk/cctv/glossary.html,
http://www.edmundoptics.com/techSupport/index.cfm,

Angle of View The angle formed by the two lines from the secondary principle point to the image sensor is called the angle of view (see figure B.1). It is computed the following way:

$$\theta = 2 \cdot \frac{1}{tan\frac{D}{2f}} \tag{B.1}$$

D : Diagonal of Sensor
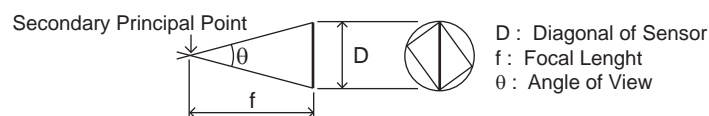f : Focal Lenght
θ : Angle of View

Figure B.1: **The Angle of View**

Aperture The opening of the lens, the size of which is controlled by the iris and is measured in F numbers. Generally, the lower the F number, the larger the aperture and consequently more light can pass through the lens. In some machine vision applications, soft aperture filters the normal (sharp) video signal to diffuse sharp edges or features that may cause unwanted harmonics during image processing.

Auto Iris An electronic circuit fitted to a lens that enables the iris to automatically adjust to control the amount of light falling upon the image device. There are two main types of auto iris - Video and DC Iris. A video iris lens requires a video signal input from the camera and converts it to a DC input through its built in amplifier circuit. A DC iris lens, often referred to as direct drive, does not have this circuitry and requires simply a DC input from the camera.

Back Focal Length (BFL) The distance between the last surface of a lens to its image plane.

C Mount and CS Mount The two standard mountings used for connecting lenses to cameras. The threads are identical on these mounts: 1 inch diameter x 32 threads per inch. C-Mount uses 17.52 mm flange back distance, CS-Mount uses 12.5 mm. CS-Mount cameras can be converted to C mount by using a 5mm spacer element. CS-Mount lenses cannot be used with C-Mount cameras.

Circle of Confusion The circle of confusion, (also known as disk of confusion, circle of indistinctness, blur circle, etc.), is an optical spot caused by a cone of light rays from a lens not coming to a perfect focus when imaging a point source. To calculate a camera's depth of field, the size of a circle of confusion that can be considered to be an acceptable focus needs to be known. The maximum acceptable diameter of such a circle of confusion is known as the maximum permissible circle of confusion, the circle of confusion diameter limit, or the circle of confusion criterion, but is often incorrectly called simply the circle of confusion.

Circle of Least Confusion The circle of least confusion is the smallest optical spot a lens can make, for example by picking a best focus position that makes a good compromise between the varying effective focal lengths of different lens zones due to spherical or other aberrations. Diffraction effects from wave optics and the finite aperture of a lens can be included in the circle of least confusion, or the term can be applied in pure ray (geometric) optics.

Depth of Field When an object is focused, it can be observed that an area in front of and behind the object is also in focus. The range in focus is called depth of field. Depth of field increases when: the focal length is shorter, the aperture (F number) is larger or the object distance is longer. When the depth of field is extended to infinity, the focusing distance is called hyper focal distance. The depth of field can be calculated by using the

following formulae (B.2-B.4):

$$H \quad = \quad \frac{F^2}{C \cdot f} \tag{B.2}$$

$$T_1 \quad = \quad \frac{B(H + f)}{H + B} \tag{B.3}$$

$$T_2 \quad = \quad \frac{B(H - f)}{H - B}. \tag{B.4}$$

F is the F number, T_1 is the near limit, T_2 is the far limit, H is the hyper focal distance, f is the focal length, B is the object distance measured from the sensor plane and C is the circle of confusion diameter limit.

F Number Refer to Aperture.

Field of View The relationship between the angle of view and the distance of the object from the lens as depicted in figure B.2. The field of view varies with the focal length of the lens as follows:

$$\frac{w}{W} = \frac{h}{H} = \frac{f}{L}. \tag{B.5}$$

W : width of object
H : height of object
w : width of sensor
h : height of sensor
f : focal lenght
L : object distance
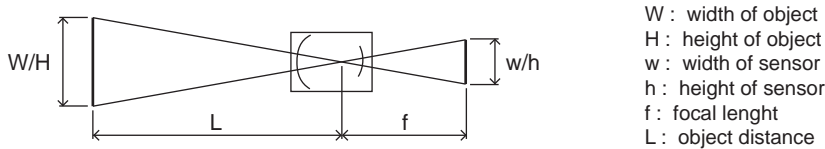
Figure B.2: **The Field of View.**

Format Lens systems are designed for certain formats of sensors. The most commonly used formats are:

- 1 inch, 12.7mm / 9.252mm [height / width]
- 2/3 inch, 8.8mm / 6.6mm [height / width]
- 1/2 inch, 6.4mm / 4.8mm [height / width]
- 1/3 inch, 4.8mm / 3.6mm [height / width]

Lens systems that are specified for larger formats can be used also on smaller formats. Usually the price of the lens system scales with the format size it is specified for.

Monofocal Lens A lens having only one focal length and therefore a fixed field of view depending on the format of the sensor.

Varifocal Lens A lens which has a variable focal length that can be set manually. Ideal for situations where the required view falls between the focal lengths of two monofocal lenses. Any adjustment in the focal length will require the lens to be re-focused.

Zoom Lens A lens which has a variable focal length that can be set manually or remotely. Unlike the varifocal lens, if the back focus is correctly set, the image should stay in focus as the focal length is altered.

Zoom Ratio The zoom ratio of a lens is simply the ratio between its maximum and minimum focal length, e.g. a 12 to 240mm lens has a zoom ratio of 20 times. Consequently the ratio between the maximum and minimum sizes of the images produced on the monitor will also be 20 times. However it will not give the exact size of the image as this is only apparent once the focal length is known.

B.3   Optical Units

B.3.1   Radiometric Units

The SI derived radiometric units are *energy, power, irradiance, radiant intensity* and *radiance.*

Power. Power, in radiometric applications often called radiant flux, is the derivative of energy with respect to time, dQ/dt, and the unit is the watt (W). Power is used for non-integrating detectors and continuous sources.

Energy. Energy is the integral over time of power, and is used for integrating detectors and pulsed sources. Irradiance, also called flux density, is measured in W/m^2.

Irradiance. Irradiance is power per unit area incident from all directions in a hemisphere onto a surface that coincides with the base of that hemisphere. A similar quantity is radiant exitance, which is power per unit area leaving a surface into a hemisphere whose base is that surface.

Radiant Intensity. Radiant intensity is measured in W/sr. Intensity is power per unit solid angle. The integral of radiant intensity over solid angle is power. Radiance is measured in $W/m^2/sr$.

Radiance. Radiance is power per unit projected area per unit solid angle. Radiance is the derivative of power with respect to solid angle and projected area, $dF/dwdA \cos q$ where q is the angle between the surface normal and the specified direction. The integral of radiance over area and solid angle is power.

Photon quantities are also common. They are related to the radiometric quantities by the relationship $Q_p = \hbar\omega = h\nu = hc/\lambda$ where Q_p is the energy of a photon at wavelength λ, h is Planck's constant, $\hbar = h/2\pi$, ν the frequency and c is the velocity of light (e.g. at a wavelength of 1 mm, there are approximately $5 \cdot 10^{18}$ photons per second in a watt and conversely, at 1 mm wavelength, 1 photon has an energy of $2 \cdot 10^{-19}$ joules).

B.3.2   Photometric Units

Photometry is a measurement method that attempts to quantify light intensity in a way to match the human perception. Therefore, photometry uses a weighting function in frequencies to account for the specific spectral response of the

human eye. To account for the two types of receptors in human eyes, photometry is divided up in scotopic photometry and photopic photometry. Scotopic photometry is applied to dim lighting conditions, in which the rod photoreceptors are dominant for vision and photopic photometry deals with bright conditions, in which the cone photoreceptors are dominant. The region between the photopic and the scotopic vision, in which both types of photoreceptors contribute to vision, is called the mesopic vision.

The basic units for photometry are *lumen (lm), lux (lx), candela (cd)* and *nit*.

Candela. The candela is one of the seven base units of the SI system and it is defined as follows (see [77]):

> The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540x10^{12}$ hertz and that has a radiant intensity in that direction of $1/683$ watt per steradian.

Because the candela can be defined in other base units of the SI system, it is not independent and could be removed as base unit. It remains a SI system base unit because of historical reasons. 1 candela is equal to 1 lumen per solid angle.

Lumen. The lumen is an SI derived unit and stands for luminous flux. The lumen is derived from the candela and is the luminous flux emitted into one unit solid angle (1 sr) by an isotropic point source having a luminous intensity of 1 candela. It is analogous to the unit of radiant flux (watt), differing only in the eye response weighting. If a light source is isotropic, the relationship between lumens and candelas is $1cd = 4\pi lm$ (an isotropic source having a luminous intensity of 1 candela emits 4π lumens into space, which is 4π steradians).

Illuminance. Illuminance, called lux, is another SI derived quantity which denotes luminous flux density. Lux is lumens per square metre, or lm/m^2. Typical values range from 100 000 lx for direct sunlight 1 lx for full moon.

Luminance. Luminance, called nit, is analogous to radiance, differentiating the lumen with respect to both area and direction. One nit is one cd/m^2. Luminance is not included on the official list of derived SI quantities but often used to characterize the "brightness" of flat emitting or reflecting surfaces. Typical laptop computer screens have between 100 and 250 nits.

The human eye spectral response is characterized experimentally. The relative effectiveness of spectral light to produce photoisomerziation in the cones is called the spectral luminous efficiency function for photopic vision, $V(\lambda)$. The relative effectiveness of spectral light to produce photoisomerziation in the rods is called the luminous efficiency function for scotopic vision, $V'(\lambda)$. These two empirical curves look very much like Gaussian functions. The Color Vision Lab at USCD provides on their web pages [78] data files of $V(\lambda)$ and $V'(\lambda)$.

B.3.3   Conversion of Units

The photometric units and the radiometric units can be converted into each other, as they represent the same quantities.

QUANTITY	RADIOMETRIC	PHOTOMETRIC
power	watt (W)	lumen (lm)
power per unit area	W/m^2	lux (lx)
power per unit solid angle	W/sr	candela (cd)
power per area per solid angle	$W/m^2/sr$	nit

The definition of the candela says that there are 683 lumens per watt at a frequency of 540 THz, which is a wavelength of 555 nm. This is the wavelength that corresponds to the maximum spectral responsivity of the human eye. The conversion from watts to lumens at any other wavelength involves the product of the power (watts) and the $V(\lambda)$ value at the wavelength of interest. In order to convert a source with non-monochromatic spectral distribution to a luminous quantity, the situation is decidedly more complex and the spectral nature of the source needs to be known. An even more complicated situation is at hand if a conversion from lumens to watt is is required for a non-monochromatic source.

B.4   Optical Measurement Methods

As stated above, the scale factor is 683 lumens per watt for yellow-green photons (540 nm). You get less lumens per watt away from this peak. For "white" light, defined as distributed like sunlight over energy over the visible range, scale is 250 lumens per watt, or 4 mW/lumen. Incandescent 60W bulbs put out around 600 lumens; they only make 10 lm/W because most watts make invisible infrared photons. Useful relations between lumens and watts (from Al Rose) are

$$lux \approx 4mW/m^2 \approx 10^4 \frac{photons}{\mu m^2 \cdot s} \; (white\ sunlight). \tag{B.6}$$

If a surface is a diffuse reflector, it looks equally bright at all angles; it emits a constant luminous intensity in all directions, a constant rate of photons per unit solid angle from your viewpoint. This means actual luminance (lumens per steradian per m^2 surface) drops off as $\cos\theta$ away from the surface perpendicular, but luminous intensity (lumen per steradian) is constant. Integrating total lumens emitted over a hemisphere from a m^2 illuminated by a lumen, tells you that you need factor of $1/\pi$ to get total lumens emitted to come out to the lumen you shined on. For a perfect diffuse reflector, the relation between luminance L and incident illuminance I (lm/m^2) is

$$L(nit) = I(lux)\frac{\cos\theta}{\pi}. \tag{B.7}$$

Note the $cos\theta$ goes away if you use projected surface area, luminance is then lumen per steradian per projected m^2. A perfect white reflector illuminated with 1 lux has 0.34 nit luminance; conversely, a surface with 1 nit luminance is equivalent to a perfect diffuser illuminated by 3.1 lux. Now we will relate illuminance of a scene to illuminance of the sensor chip. This is just geometry.

Suppose 1 *lux* illuminates 1 m^2 white diffuse reflector, imaged through aperture f. The 1 lumen reflected is spread into hemisphere of 2 π steradian; the lens intercepts a fraction of it and concentrates it into a smaller area on the sensor. The formula relating imager illuminance L_{chip} (*lux*) to scene illuminance L_{scene} (*lux*), for perfect (white) reflectance, small magnification (objects a lot more distant that focal length) and no lens absorption, is computed in Fig. 22:

This result is very useful for estimating the practical implication of a measured dark current level. The implication of this result is that chip illumination is much smaller than scene illumination: Scene reflectance (averaging to 18

Appendix C

aVLSI Basics

Appendix D

AER infrastructure

Bibliography

[1] J. Adkisson, J. Gambino, T. Hoague, M. Jaffe, J. Kyan, R. Leidy, D. Mc-Grath, R. Rassel, D. Sacket, and C.V. Stancampiano. Optimazition of Cu interconnect layers for $2.7\mu m$ pixel image sensor technology: Fabrication, modeling, and optical results. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[2] M. Murakami, M. Masuyama, S. Tanaka, M. Uchida, K. Fujiwara, M. Kojima, Y. Matsunaga, and S. Mayumi. $2.8\mu m$-pixel image sensor vMaicovicon^{TM}. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[3] N. Karasawa, D. Sugimoto, H. Ohki, K. Hatano, and H. Yoshida. A 2.2-um squarre pixel IT-CCD constructed from single-layer electrode. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[4] T. Delbrück. Library essentials, Analog VLSI and neural systems by Carver Mead, Addison Wesley, 1986. *The Neuromorphic Engineer*, 1(1):11, 2004. http://ine-web.org/research/newsletters/index.html.

[5] C.A. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, Reading, MA, 1989.

[6] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, and R. Douglas. *Analog VLSI:Circuits and Principles*. MIT Press, 2002.

[7] M. Riordan. The end of AT&T: Ma Bell may be gone, but its innovations are everywhere. *IEEE Spectrum*, 42(7):46–51, May 2005.

[8] E. R. Fossum. Active pixel sensors: Are CCDs dinosaurs? In M. M. Blouke, editor, *Charge-Coupled Devices and Solid State Optical Sensors III, Proceedings of the SPIE*, volume 1900, pages 2–14, 1993.

[9] E. R. Fossum. What to do with sub-diffraction-limit (SDL) pixels? – a proposal for a gigapixel digital film sensor (DFS). In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[10] B. Pain, T. Cunningham, S. Nikzad, M. Hoenk, T. Jones, B. Hancock, and C. Wrigley. A back-illuminated megapixel cmos image sensor. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[11] J. Bogaerts, B. Dierickx, P. De Moor, D. Sabuncuoglu Teczan, K. De Munck, and C. Van Hoof. High-end CMOS active pixel sensor for hyperspectral imaging. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[12] B. Benwell, N. O, G. Allen, J. Huras, and M. Ledgerwood. $12k$ $5\mu m$ linescan CCD sensor with $320mhz$ data rate. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[13] S. Kleinfelder, S. Lim, X. Liu, and A. El Gamal. A 10000 frames/s cmos digital pixel sensor. *IEEE Journal of Solid-State Circuits*, 36(12):2049–2059, December 2001.

[14] A. Joshi, D. Chiaverini, K. Jung, V. Douence, G. Wijeratne, G.M. Williams, and M. Loose. Scalable architecture for high-resolution video-rate CMOS imaging systems on chip. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[15] T. Honda, H. Yamamoto, H. Okamoto, and R. Nagayoshi. Development of 2/3-type 1-mega pixel progressive scan CCD for HDTV capable of high frame rate of $96fps$. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[16] D. Stoppa, V. Luigi, A. Simoni, L. Gonzo, M. Malfatti, and G. Pedretti. A 50-pixel CMOS sensor for TOF-based real time 3d imaging. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[17] M. Lehmann, T Oggier, B. Büttigen, Chr. Gimkiewicz, M. Schweizer, R. Kaufmann, F. Lustenberger, and N. Blanc. Smart pixels for future 3D-TOF sensors. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[18] R. W. Rodieck. *The First Steps in Seeing*. Sinauer Associates, Inc., Sunderland, Massachusetts, 1998.

[19] J. Zihl, D. Von Cramon, and N. Mai. Selective Disturbance of Movement Vision After Bilateral Brain Damage. *Brain*, 106(2):313–340.

[20] C. Mead. Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10):1629–36, October 1990.

[21] J. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti, and D. Gillespie. Silicon auditory processors as computer peripherals. *IEEE Transactions on Neural Networks*, 4:523–528, 1993.

[22] E. Culurciello and A. G. Andreou. A comparative study of access topologies for chip-level address-event communication channels. *IEEE Transactions on Neural Networks*, 14(5):1266–77, September 2003.

[23] K. A. Boahen. Point-to-point connectivity between neuromorphic chips using address-events. *IEEE Transactions on Circuits and Systems II*, 47(5):416–34, 2000.

[24] K. A. Boahen. A burst-mode word-serial address-event link – I: Transmitter design. *IEEE Circuits and Systems I*, 51(7):1269–80, 2004.

[25] K. A. Boahen. A burst-mode word-serial address-event link – II: Receiver design. *IEEE Circuits and Systems I*, 51(7):1281–91, 2004.

[26] K. A. Boahen. A burst-mode word-serial address-event link – III: Analysis and test results. *IEEE Circuits and Systems I*, 51(7):1292–300, 2004.

[27] M. Mahowald. *An Analog VLSI System for Stereoscopic Vision*. Kluwer, Boston, MA, 1994.

[28] K. A. Zaghloul and K. Boahen. Optic nerve signals in a neuromorphic chip: Parts 1&2. *IEEE Transactions on Biomed. Eng.*, 51:657–675, 2004.

[29] K. Boahen. Neuromorphic microchips. *Scientific American*, pages 56–63, May 2005.

[30] M. Barbaro, P.-Y. Burgi, A. Mortara, P. Nussbaum, and F. Heitger. A 100100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding. *IEEE Jour. of Solid-State Circuits*, 37(2):160–172, Feb. 2002.

[31] P.-F. Ruedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P.-Y. Burgi, S. Gyger, and P. Nussbaum. A 128 /spl times/ 128 pixel 120-db dynamic-range vision-sensor chip for image contrast and orientation extraction. *IEEE Jour. of Solid-State Circuits*, 38(12):2325–2333, Dec. 2003.

[32] E. Grenet, S. Gyger, P. Heim, F. Heitger, F. Kaess, P. Nussbaum, and P.-F. Ruedi. High dynamic range vision sensor for automotive applications. In *Proceedings of the SPIE*, volume 5663, pages 246–253, 2005.

[33] A. Mortara. A pulsed communication/computation framework for analog VLSI perceptive systems. In T. S. Lande, editor, *Neuromorphic Systems Engineering*, pages 217–228. Kluwer Academic, Norwell, MA, 1998.

[34] U. Mallik, M. Clapp, E. Choi, G. Cauwenberghs, and R. Etienne-Cummings. Temporal change threshold detection imager. In *2005 IEEE ISSCC Digest of Technical Papers*, pages 362–363. IEEE, 2005.

[35] V. Gruev and Etienne-Cummings, R. A pipelined temporal difference imager. *IEEE J. Solid State Circuits*, 39:538–543, 2004.

[36] E. Culurciello, R. Etienne-Cummings, and K. Boahen. Arbitrated address-event representation digital image sensor. *Electronics Letters*, 37(24):1443–1445, Nov 2001.

[37] E. Culurciello, Etienne-Cummings, R., and K. Boahen. A biomorphic digital image sensor. *IEEE J. Solid State Circuits*, 38:281–294, 2003.

[38] E. Culurciello and Etienne-Cummings, R. Second generation of high dynamic range, arbitrated digital imager. In *2004 IEEE International Symposium On Circuits and Systems*, volume 4, pages IV–828–831, Vancouver, May 2004. ISCAS '04: Vancouver, Canada.

[39] X. Qi, G. Xiaochuan, and J. Harris. A time-to-first-spike CMOS imager. In *Proc. IEEE Int. Symp. on Circuits and Systems*, volume 4, pages IV–824–827. IEEE, May 2004.

[40] Q. Luo and J. Harris. A time-based CMOS imager. In *Proc. IEEE Int. Symp. on Circuits and Systems*, volume 4, pages IV–840–843. IEEE, May 2004.

[41] M. Azadmer, J. P. Abrahamsen, and P. Häfliger. A foveated aer imager chip. In *IEEE International Symposium on Circuits and Systems, ISCAS 2005*. IEEE, May 2005.

[42] T. Delbrück and C.A. Mead. Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits. In C. Koch and H. Li, editors, *Vision Chips: Implementing vision algorithms with analog VLSI circuits*, pages 139–161. IEEE Computer Society Press, Los Alamitos, CA, 1995.

[43] T. Delbrück. "Bump" circuits for computing similarity and dissimilarity of analog voltages. In *Proc. Int. Joint Conf. Neural Networks*, pages I–475–479, July 1991.

[44] J. Kramer. An ON/OFF transient imager with event-driven, asynchronous readout. In *Proc. IEEE International Symposium on Circuits and Systems*, May 2002.

[45] J. Kramer. An integrated optical transient sensor. *IEEE Transactions on Circuits and Systems II*, 49(9):612–628, Sep 2002.

[46] G. Indiveri, A.M. Whatley, and J. Kramer. A reconfigurable neuromorphic VLSI multi-chip system applied to visual motion computation. In *Proceedings of the Seventh International Conference on Microelectronics for Neural, Fuzzy and Bio-inspired Systems; Microneuro'99*, pages 37–44, Los Alamitos, CA, April 1999. IEEE Computer Society.

[47] G. Indiveri, R. Mürer, and J. Kramer. Active vision using an analog VLSI model of selective attention. *IEEE Transactions on Circuits and Systems II*, 48(5):492–500, May 2001.

[48] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbruck, T. Burg, and R. Douglas. Orientation-selective aVLSI spiking neurons. *Neural Networks*, 14(6/7):629–643, 2001.

[49] T. Delbrück and C. A. Mead. Adaptive photoreceptor with wide dynamic range. In *1994 IEEE International Symposium On Circuits and Systems*, volume 4, pages 339–342, London, May 1994. ISCAS '94: London, England, 30 May–2 June.

[50] J. Kramer, R. Sarpeshkar, and C. Koch. Pulse-based analog VLSI velocity sensors. *IEEE Transactions on Circuits and Systems II*, 44(2):86–101, February 1997.

[51] P. Lichtsteiner, T. Delbruck, and J. Kramer. Improved ON/OFF temporaly differentiating address-event imager. In *11th IEEE International Conference on Electronics, Circuits and Systems.*, pages 211–214. IEEE, December 2004.

[52] M.Lose, K.Meier, and J.Schemmel. A self-calibrating single-chip cmos camera with logarithmic response. *IEEE Journal of Solid-State Circuits*, 36:586–596, 2001.

[53] S.Kavadias, B.Dierickx, D.Scheffer, A.Alaerts, D.Uwaerts, and J.Bogaerts. A logarithmic response cmos image sensor with on-chip calibration. *IEEE Journal of Solid-State Circuits*, 35:1146–1152, 2000.

[54] T. Delbruck and Oberhof D. Self-biasing low-power adaptive photoreceptor. In *2004 IEEE International Symposium On Circuits and Systems*, pages IV–844–847, 2004. ISCAS '04.

[55] P. Lichtsteiner and T. Delbrück. 64×64 event-driven logarithmic temporal derivative silicon retina. In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, Karuizawa, Nagano, Japan, 2005. IEEE.

[56] T. Delbrück and P. Lichtsteiner. Fully programmable bias current generator with 24 bit resolution per bias. In *2006 IEEE International Symposium On Circuits and Systems*, Kos, 2006. ISCAS '06: Kos, Greece.

[57] T. Delbrück and A. van Schaik. Bias current generator with wide dynamic range. *Jour. of Analog Integrated Circuits and Signal Processing*, 43:247–268, 2005.

[58] B. Linares-Barranco, T. Serrano-Gotarredona, and R. Serrano-Gotarredona. Compact low–power calibration mini–DACs for neural arrays with programmable weights. *IEEE Transactions on Neural Networks*, 14(5):1207–1216, September 2003.

[59] T. Hamamoto, S. Sugiura, and S. Sawada. On the retention time distribution of dynamic random access memory (DRAM). *IEEE Transactions on Electron Devices*, 45(6), 1998.

[60] P. Lichtsteiner and T. Delbrück. A 64×64 AER logarithmic temporal derivative silicon retina. In *Research in Microelectronics and Electronics, 2005 PhD*, pages 202–205. IEEE, 2005.

[61] R. Sarpeshkar, T. Delbrück, and C. A. Mead. White noise in MOS transistors and resistors. *IEEE Circuits and Devices Magazine*, 9(6):23–29, November 1993.

[62] P. Abshire and A.G. Andreou. Capacity and energy cost of information in biological and silicon photoreceptors. *Proceedings of the IEEE*, 89:1052–1064, 2001.

[63] B. Fowler, M. D. Godfrey, and S. Mims. Reset noise reduction in capacitive sensors. *IEEE Transactions on Circuits and Systems I*, page accepted, 2006.

[64] A. J. P. Theuwissen. *Solid-state imaging with charge-coupled devices.* Kluwer, Dordrecht, The Netherlands, 1995.

[65] A. Rose. *Vision: Human and Electronic.* Plenum Press, New York, 1973.

[66] E. Chicca, P. Lichtsteiner, T. Delbruck, G. Indiveri, and R. J. Douglas. Modeling orientation selectivity using a neuromorphic multi-chip system. In *Proceedings of IEEE International Symposium on Circuits and Systems*, 2006. (In Press).

[67] R. J. Douglas and K. A. C. Martin. Neural circuits of the neocortex. *Annual Review of Neuroscience*, 27:419–51, 2004.

[68] D. Ferster and K. D. Miller. Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.*, 23:441–71, 2000.

[69] T. Serrano-Gotarredona, A. G. Andreou, and B. Linares-Barranco. AER imager filtering architecture for vision–processing systems. *IEEE Transactions on Circuits and Systems – I*, 46:1064–71, 1999.

[70] P. Venier, A. Mortara, X. Arreguit, and E. A. Vittoz. An integrated cortical layer for orientation enhancement. *IEEE Journal of Solid–State Circuits*, 32(2):177–86, 1997.

[71] G. Cauwenberghs and J. Waskiewicz. Focal–plane analog VLSI cellular implementation of the boundary contour system. *IEEE Transactions on Circuits and Systems – I*, 46(2):1064–71, 1999.

[72] T. Y. W. Choi, P. A. Merolla, J. V. Arthur, K. A. Boahen, and B. E. Shi. Neuromorphic implementation of orientation hypercolumns. *IEEE Transactions on Circuits and Systems I*, 52(6):1049–60, 2005.

[73] V. Dante, P. Del Giudice, and A. M. Whatley. PCI-AER – hardware and software for interfacing to address-event based neuromorphic systems. *The Neuromorphic Engineer*, 2(1):5–6, 2005. http://ine-web.org/research/newsletters/index.html.

[74] M. Oster. Tuning aVLSI chips with a mouse click. *The Neuromorphic Engineer*, 2(1):9, 2005. http://ine-web.org/research/newsletters/index.html.

[75] R. Hahnloser, R. Sarpeshkar, M. Mahowald, R. J. Douglas, and S. Seung. Digital selection and analog amplification co-exist in an electronic circuit inspired by neocortex. *Nature*, 405(6789):947–951, 2000.

[76] E. Chicca, G. Indiveri, and R. J. Douglas. An event based VLSI network of integrate-and-fire neurons. In *Proceedings of IEEE International Symposium on Circuits and Systems*, pages V-357–V-360. IEEE, 2004.

[77] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gómez-Rodríguez, H. Kolle Riis, T. Delbrück, S. C. Liu, S. Zahnd, A. M. Whatley, R. J. Douglas, P. Häfliger, G. Jimenez-Moreno, A. Civit, T. Serrano-Gotarredona, A. Acosta-Jiménez, and B. Linares-Barranco. AER building blocks for multi-layer multi-chip neuromorphic vision systems. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, Dec 2005.

[78] Bureau International des Poids et Mesures. The international system of units. World Wide Web page, 2006. http://www.bipm.org/en/si/.

[79] Colour and Vision Research Laboratories. Luminous efficiency. World Wide Web page, 2006. http://cvision.ucsd.edu/lumindex.htm.

Acknowledgments