

# Analog and digital implementations of retinal processing for robot navigation systems

**Author(s):**
Moeys, Diederik Paul

**Publication Date:**
2016

**Permanent Link:**

https://doi.org/10.3929/ethz-a-010897825 →

DISS. ETH NO. 24047

# *ANALOG AND DIGITAL IMPLEMENTATIONS OF RETINAL PROCESSING FOR ROBOT NAVIGATION SYSTEMS*

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

Presented by

DIEDERIK PAUL MOEYS

MEng Electronic and Electrical Engineering, University College London (UCL)

Born on 31.12.1990

Citizen of The Netherlands/France

Accepted on the recommendation of:

Prof. Dr. Tobi Delbruck

Prof. Dr. Kevan A. C. Martin

PD. Dr. Shih-Chii Liu

Prof. Dr. Bernabé Linares-Barranco

2016

## Abstract

This thesis presents both analog and digital implementations of visual processing inspired by the biological retina. The aim is to use this processing for a tracking task to allow robotic navigation within a chase scenario. Although the full extent of retinal processing is not yet fully understood due to its enormous complexity, its interesting aspects are brought into the context of electronic engineering and computer science: in analog design, signal processing and machine learning.

First, the basic functioning of the biological retina is explained as well as its analogies with the Dynamic Vision Sensor (DVS) or "Silicon Retina", which is the sensor used for the task. This sensor is the fundamental hardware on which this entire doctorate is based on. Its activity-dependent spiking output is used to mimic the photoreceptors' response to light changes, at high speeds and with large dynamic range. The sensor's output data is then processed either in the jAER JAVA framework or on FPGA, where neural circuits of the Retinal Ganglion Cells (RGC) are mimicked in order to extract useful information. Personal work done in this area has proved that it is indeed possible to use an ensemble of Object Motion Sensitive (OMS) RGCs in order to detect object motion and track it, under the condition of little or no ego-motion. The current OMS algorithm fails for self-motion of the DVS as the entire and more complex cellular interactions of RGCs is not re-created and the entire processing relies on a single cell-type.

To overcome this problem, the approach of machine learning was adopted as a consequence. Specifically targeting a predator/prey scenario, a Convolutional Neural Network was setup in order to provide steering directions to the predator robot and allow it to move, in an arena, following its prey. A relatively small network was used and fed with DVS histograms created at an activity-dependent rate. The convolution kernels learned, i.e. the pattern that the network looks for in the scene, still resemble the receptive fields of the biological retina. However, the empirical knowledge surrounding machine learning algorithms does not yet allow to fully make a detailed comparison. Practical on-field tests were performed and surprisingly successful results were obtained.

Finally, to further improve the basic technology on which the algorithmic work relies on, another, more sensitive, silicon retina was designed. A more sensitive sensor means more details, and therefore information, picked up in a scene. This thesis presents its design and characterization according to newly set and more precise standards, defining Signal-to-Noise Ratios for example. The contrast detected can be down to 0.95% for negative logarithmic changes in light intensity with this sensor. The high gain comes at the cost of a reduced intra-scene dynamic range; therefore, an adaptation mechanism is also present in the sensor in order to match the scene's median illumination. Another interesting application for using this sensor, in the context of calcium imaging in neurons, to detect their activity, is also discussed and preliminary results are shown.


*Keywords: Neuromorphic Engineering, Dynamic Vision Sensor (DVS), Active Pixel Sensor (APS), Dynamic and Active Vision Sensor (DAVIS), Silicon Retina, Characterization, Low-power, Temporal Contrast, Automated Gain Control, FPGA Signal Processing, Object Motion Cell, Convolutional Neural Network, Robot, Photodiode, Color Sensor.*

## Sommario (Italian)

Questa tesi presenta implementazioni sia analogiche che digitali di elaborazione visiva ispirata alla retina umana. Lo scopo è utilizzare l'elaborazione della retina per fare tracking e permettere la navigazione robotica nel contesto di uno scenario predatore/preda. Nonostante la complessità della retina non sia ancora totalmente conosciuta, alcuni suoi aspetti interessanti sono estratti e messi nel contesto dell'ingegneria elettronica e della scienza dei computer: nel design analogico, nell'elaborazione di segnali e nel machine learning.

Per prima cosa, il funzionamento basilare della retina biologica è spiegato, insieme alle sue analogie con il Dynamic Vision Sensor (DVS) o "Retina di Silicio", che è il sensore utilizzato per risolvere il problema. Questo sensore è l'hardware fondamentale sul quale si basa questo intero dottorato. Il suo spiking-output che dipende dall'attività presente nella scena, è utilizzato per imitare la risposta dei fotorecettori ai cambiamenti di luminosità, a velocità altissime e con grande range dinamico. L'output del sensore è poi elaborato o dal framework jAER in JAVA o da FPGA, dove i circuiti neurali delle Cellule Retinali Ganglionali (CRG) sono imitati per estrarre informazioni utili. Lavoro compiuto in quest'area ha provato che è possibile infatti di utilizzare un insieme di CRG Sensibili al Movimento di Oggetti (SMO) per detettare oggetti in movimento e seguirli, in condizioni di sensore statico. L'attuale algoritmo SMO sbaglia però quando il sensore si muove visto che l'intero complesso di cellule della retina non è interamente riprodotto e l'intera elaborazione dei dati dipende da un solo tipo di cellula.

Per superare questo problema, è stato adottato l'approccio del machine learning come conseguenza. In specifico, interessandosi ad uno scenario predatore/preda, un Network Neurale Convoluzionale è stato creato per dare direzioni di manovra al robot predatore e permettergli di muoversi, in un arena, seguendo la sua preda. Per questo scopo, un network relativamente piccolo, che prende i dati del DVS sotto froma di istogramma con frame-rate variabile, è stato usato. I motivi di convoluzione che il network cerca nella scena, assomigliano ancora ai campi recettivi della retina biologica. Ciononostante, la conoscenza empirica che circonda gli algoritmi di machine learning non permette ancora di comparare dettagliatamente biologia e computer vision. Test pratici su campo sono stati svolti e risultati sorprendentemente positivi sono stati ottenuti.

Infine, per migliorare ancora la tecnologia base su cui gli algoritmi lavorano, un altra retina di Silicio più sensibile è stata sviluppata. Un sensore più sensibile significa più dettagli raccolti in una scena visiva (e quindi più informazione). Questa tesi ne presenta il design e la caratterizzazione secondo nuovi e più precisi standard, definendo anche il rapporto segnale-rumore per esempio. Il contrasto detettato può scendere fino allo 0.95% per cambiamenti logaritmici di intensità. L'alto guadagno del sensore però viene con il costo di un range dinamico intra-scenico ridotto; perciò, un meccanismo d'adattamento è anche presente per ovviare al problema e sincronizzare il sensore all'illuminazione mediana della scena. Un'altra interessante applicazione nel contesto della visualizzazione del calcio nei neuroni per detettarne l'attività è discussa in questa tesi, insieme a risultati preliminari. Sono anche sicuro che nessuno leggerà questa tesi quindi saluto te, Diederik del futuro.

**To family, friends, my love and wiener dogs.**

# Acknowledgements

I would first of all thank the reviewers of this thesis who will have to go through all these pages. Thank you Tobi, Kevan, Bernabè and Shih-Chii: much appreciated!

I would like to thank my supervisor Prof. Dr. Tobi Delbruck for supervising me during this entire PhD. I would also like to thank Dr. Sim Bamford for helping me with the layout of the new generation SDAVIS192 sensor (whose initial idea comes from Bernabè and Tobi). I would like to thank Dr. Shih-Chii Liu for supervising me for the design of the AQC cochlear circuit which extended into my PhD too.

I would like to thank the INI Sensors' meeting in general for interesting discussions and Daniel Neil for very valuable help in the context of neural network. He instructed me from scratch, seriously.

I would like to thank Dr. Federico Corradi and Luca Longinotti for teaching me that programming is not so bad and that I can even do it myself. Their help was really valuable in setting up everything for sensors' characterization and neural network testing. I would also like to thank Chenghan Li for helping me with the measurement setup and for the wise comments on the characterization definitions and design considerations.

I would like to thank Fabian Voigt from HIFO for supervising me and Gemma Taverni while testing the SDAVIS192 sensor with fluorescence imaging. I would like to thank the INI people that made the working environment nice and interdisciplinary.

I would like to thank Prof. Dr. Alejandro Linares-Barranco, Antonio Rios-Navarro and the entire ATC group in Seville for helping me with FPGA design and the good times at ISCAS. I would also like to thank the folks at the University of Göttingen (Prof. Dr. Tim Gollisch's lab) and at the University of Ulster (Prof. Dr. Martin McGinnity's lab) for helping me in the project and welcoming me at their institutions.

I would like to thank the people at IMEC too who helped me and Chenghan with the sensors' characterization, and the Telluride Neuromorphic Engineering Workshop 2015 participants for introducing me to new ideas (especially the IBM guys).

More than all, I would to thank my family, especially my mother for helping me with homework since age 5 and my father for helping me get into the best UK universities. I wouldn't even be here if it wasn't for them.

And finally, I really have to thank my goddess Elisa for putting up with me all these years and for making my life wonderful.

Oh, I forgot, I would definitely NOT thank the "reviewer 2" of every submission I have ever made, who made my life miserable at paper submission times.

This research was supported by the European Commission project VISUALISE (FP7-ICT-600954) and by the Samsung Project NPP. The Seebetter project funded the wafer run in which SDAVIS192 was designed.

# Table of contents

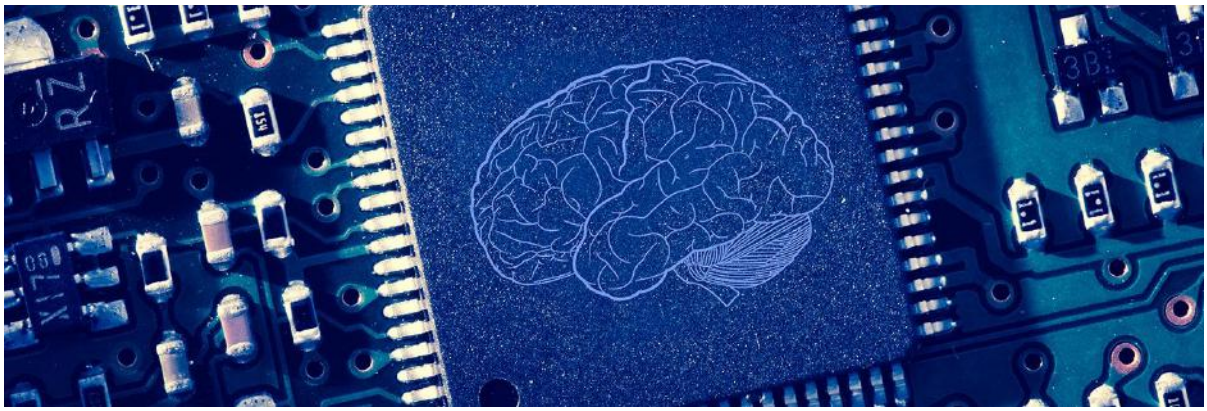

Image adapted from [1]

# Chapter 1

## 1. Introduction

This introductory chapter gives, at first, an overview of the thesis. It also aims to introduce the concept of neuromorphic engineering and its historical milestones. In particular, attention is given to the design and evolution of the Dynamic Vision Sensor (DVS) or silicon retina. The basic concepts of section 1 will then be built upon in chapter 4, where the DVS characterization methods and the newly developed SDAVIS192 are introduced.

## 1.1. Thesis overview

After the introduction presented in this chapter, the section "Retinal Ganglion Cell implementation: the Object Motion Cell" introduces the biology of the retina. It is also shown how, what we understand about retinal ganglion cells neural circuits can be implemented into software and digital design with the purpose of object detection and tracking, which could then be used in a robotic chase scenario. The chapter includes the pre-print copy of the conference paper [2], presented at ISCAS 2016. The subsequent chapter, "Convolutional Neural Network for steering a robot in a predator-prey scenario" looks at an alternative approach to solve the same task with the help of convolutional neural networks. Successful results are shown in the context of the predator/prey robotics scenario, where the predator robot receives steering directions depending on the tracked prey position. The chapter also includes a pre-print copy of the conference paper [3], presented at EBCCSP 2016. The chapter "Characterization and applications of a high-gain Dynamic and Active pixel Vision Sensor" looks at the design of the SDAVIS192. This sensor is an improved and more sensitive DAVIS, designed to detect lower contrast changes, which could encode important features in the visual field. The same chapter explains how to characterize Dynamic and Active Pixel Vision Sensors (DAVIS) used in the experiments, to evaluate their features and therefore their fitness to the task. Preliminary results for another application, in the field of fluorescent imaging, are also discussed. The chapter includes a pre-print copy of the journal paper [4]. The "Conclusions" chapter summarizes the investigations' results. Finally, "Appendix I" introduces the setup used in chapter 4 and "Appendix II" presents the design of photodiode test structures in order to estimate dark current of the technology used for the fabrication of SDAVIS192. Although the testing was never finished, this appendix is useful as general insights about the photodiode technology used in SDAVIS192 are given.

## 1.2. Neuromorphic Engineering

This thesis introduces work done in the field of Neuromorphic Engineering (NE). NE consists of trying to understand the underlying principles of the brain and reproducing them in hardware and software, in order to achieve fast and intelligent computations which diverge from standard approaches.

The brain, however, is immensely complex and a full understanding of even its most basic functions is hard to grasp (especially by engineers such as myself) and the knowledge surrounding it is based on hypotheses. In neuroscience, brain parts are studied in great detail, but their overall functional integration is still based on suppositions. Therefore, currently, NE tries to reproduce the only aspects of the brain that appear as smart solution to engineering problems. This results in picking interesting concepts and integrating them into the current technology.

NE was first conceived by Prof. Dr. Carver Mead [5] around the end of the 1980s. He had investigated the claims that the subthreshold characteristics of transistors were very similar to the characteristics of the ion channels of the biological neuron's membrane. As a matter of fact, just like the ionic conductance of the membrane (dependent on the number of ion channels open) is exponentially on the membrane's potential, the subthreshold diffusion current of the transistor is exponentially dependent on the potential barrier height. Fruitful discussions with his colleagues Max Delbruck and Paul Müller led him to believe that through the recreation of basic neural circuits and functions in Very Large Scale Integrated Circuits (VLSI) it is possible to create intelligent systems based on the real nervous system [6]. The nervous system is stochastic, massively parallel, power-efficient

(the entire brain burns about 20 W of power), asynchronous, analog and incredibly adaptive. This makes it different from the fast, deterministic, digital, clocked power-hungry and fixed computers. Constructing VLSI systems based on the brain's principles has the advantage of both sides: intelligent and low-power computations and the speed of electrical systems. Electrons' mobility is seven orders of magnitude larger than the ion's mobility while the brain's power consumption per "operation" is roughly seven orders of magnitude smaller than a computer's [7].

Mead started to teach the influential Caltech's "Physics of Computation" course, together with Richard Feynman and John Hopfield (other pioneers of the field) and to develop the basic neuromorphic circuits. These included, of course, neuron and synapses [8][9] but also silicon cochleas [10][11] and, in particular, silicon retinas, the main focus of this thesis. A detailed description of the biological retina can be found in section 2.1 of chapter 2.

## 1.3. Silicon Retinas and the Dynamic Vision Sensor

Although an initial work on a silicon retina appeared in 1970 [12], it was with Misha Mahowald that the silicon retina was truly developed [13][14][15][16]. Based on the logarithmic photoreceptor of Mead [17] the voltage-mode retina gave temporal contrast by producing an event whenever the photocurrent would exceed its low-pass filtered version by a certain amount. This was improved by Kramer in [18] by adding two polarities to the event resulting from the detection of the change. This was the precursor to the Dynamic Vision Sensor (DVS) of [19][20] and finally [21].

The DVS was the first stable asynchronous temporal contrast vision sensor. The DVS produces, asynchronously, a positive ON event (ON spikes) when a positive logarithmic change in light intensity happens at a specific pixel location, and a negative OFF event (OFF spikes) when the opposite happens. Every pixel is completely independent from each other and their output consists of an x and y event coordinate together with a request, handled via an Address Event Representation (AER) arbiter tree. The AER protocol is a four-phase handshake [22], within which an element raises its request (phase 1) waiting for the acknowledge signal of the next cascaded element. If the element acknowledges back (phase 2), the first element withdraws its request (phase 3) and so does the second element with its acknowledge signal (phase 4). The event coordinate is then timestamped by an off-chip Field Programmable Gate Array (FPGA) and sent off-board.

The simplified schematic of the DVS pixel is shown in Fig. 1A (taken from [21]). Light is detected through the logarithmic photoreceptor of [23] which gives an output voltage proportional to the natural logarithm of the photocurrent detected by the photodiode. A feedback with large gain ensures that whenever the photocurrent decreases the node voltage of the photodiode, more current is sourced by the nFET transistor which sources the current to the photodiode. This effectively clamps the photodiode node to a virtual ground. This logarithmic stage is what gives to the DVS its large Dynamic Range (DR). The logarithmic operation is in fact the compression that is required achieved large DR. The logarithm of a certain contrast change is approximately equal to the fractional increase from one level to the next giving this contrast, as equation (1.3.1) points out.

$$\ln(1+\varepsilon) \approx \varepsilon \quad for\ \varepsilon \ll 1 \tag{1.3.1}$$

The next stage of the DVS is the differencing amplification where only changes in the output of the photoreceptor (marked as Vp) are amplified. This low-mismatch switched-capacitor amplification stage magnifies changes in Vp and feeds them (through its output node Vdiff) to two comparators which generate ON or OFF events. Events are generated whenever Vdiff crosses either one of the comparators thresholds. The events translate into AER requests which are then acknowledged. The acknowledge signal then causes the reset of the differencing amplifier to be activated and Vdiff is reset to a value in between the two comparators' thresholds. Once the amplifier is reset, it can start to integrate again, on its sensitive node, a new change of Vp.

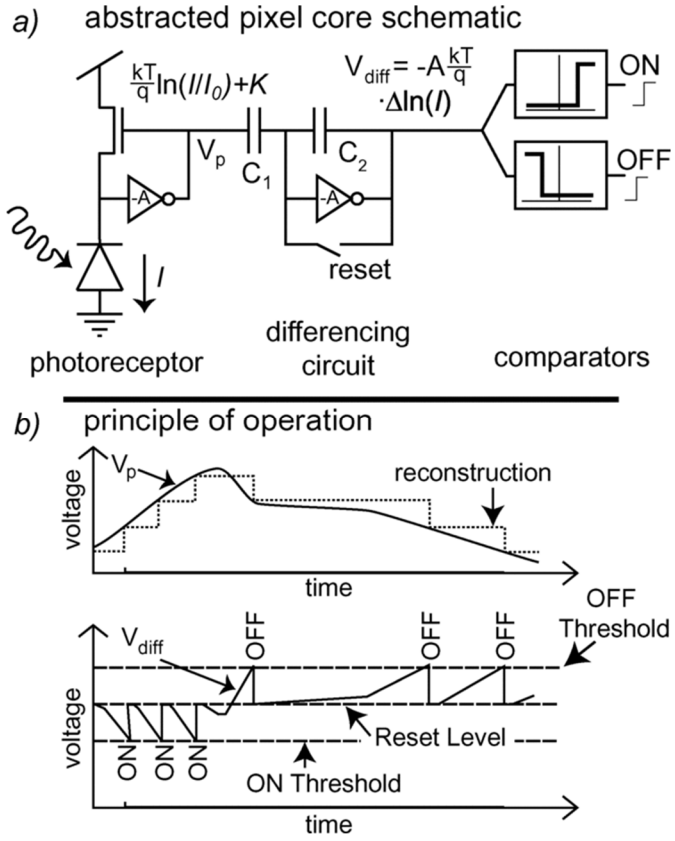


*Fig. 1 Top: abstracted schematic of the DVS. Bottom: ideal principle of operation of the DVS. Image taken from [21].*

The fact that every pixel is independent of each other and that only relevant information is sent out from the sensor through the AER protocol, makes the data rate of the chip activity-dependent and relatively low, compared to frame-based systems. A low data-rate means that data can be sent out faster and that micro-second resolution inter-spike intervals can be achieved, together with a high DR. In practice, if the x and y coordinates of ON and OFF events are displayed in a 2D grid, what is visible is the contour of the moving object or the parts of the scene which get brighter or darker (such as a light bulb turning on or off). See, for example, Fig. 7, which is the output of JAVA-based jAER software [24] used to display the output of the DVS sensor. This feature makes the DVS suitable for many applications, but the most obvious that come to mind are surveillance, odometry [25][26], optical flow estimation [27][28] and tracking [29][30][2][31].

The DVS of [21], has since evolved to the Dynamic and Active Pixel Vision Sensor (DAVIS) sensor of [32] which combines the DVS capabilities with the Active Pixel Sensor (APS) capabilities of a normal

frame-based camera (the ability to take an overall intensity readout, i.e. a frame). The fusion of the two methods comes with the area cost of only the extra APS transistors, since the two circuits share the same photodiode (the major cost factor). For a more detailed explanation of the DAVIS circuit please refer to chapter 0, where a new-generation more sensitive DAVIS is introduced. Since [32], no publication on the DAVIS sensor itself has been made apart from the CDAVIS [33]: a combination of 3 state-of-the-art color APS pixels for every monochromatic DAVIS pixel.

Previous work described in [34] had the focus of increasing the contrast sensitivity of the DVS sensor. A single pixel only was however fabricated and tested. Other designs of the DVS concept were developed in Seville. In particular [35][36] improved the sensitivity of DVS circuit with extra pre-amplifiers. Increased contrast detection was also achieved by Dr. Minhao Yang in [37]. The SDAVIS192 circuit presented in section 4 aims to achieve similar results.

Finally, various DVS-like sensors were also developed in parallel to the DVS. Examples are the Singapore's time-to-first spike sensors [38][39], which encode intensity with the quantity of spikes generated, and the DVS-inspired Asynchronous Time-Based Image Sensor (ATIS) of [40][41][42]. The ATIS outputs both a signed spiking output encoding for relative changes in intensity and the time it took from such event generation to the reference-voltage crossing of the separate photodiode dedicated to intensity readout. This way, the intensity readout is reconstructed from the time it took to integrate a certain voltage from the detected intensity. Although less bio-inspired, more complex processing element pixels have also been developed by Dudek [43][44], computing simple computer vision algorithms (erosion, hole filling, etc.. ) within the array of pixels.

Image adapted from [45]

# *Chapter 2*

## 2. Retinal Ganglion Cell implementation: the Object Motion Cell

The work of the following chapter focuses on the behavior of Retinal Ganglion Cells (RGC) and how what we understand about them can be mimicked in software and hardware to solve object detection and tracking tasks. The aim of tracking, if successful, is to then possibly steer a predator robot in the direction of the tracked prey robot. To understand the function and importance of RGCs, these are first introduced in their biological context. Then, a specific non-standard type RGC, the Object Motion Cell (OMC) is presented, along with its software and Field Programmable Gate Array (FPGA) implementations. Advantages and limitations of this approach are discussed.

## 2.1. Biological relevance of Retinal Ganglion Cells

In a simplified cross-section of the eye, as the one shown in Fig. 2, the basic components can be observed.

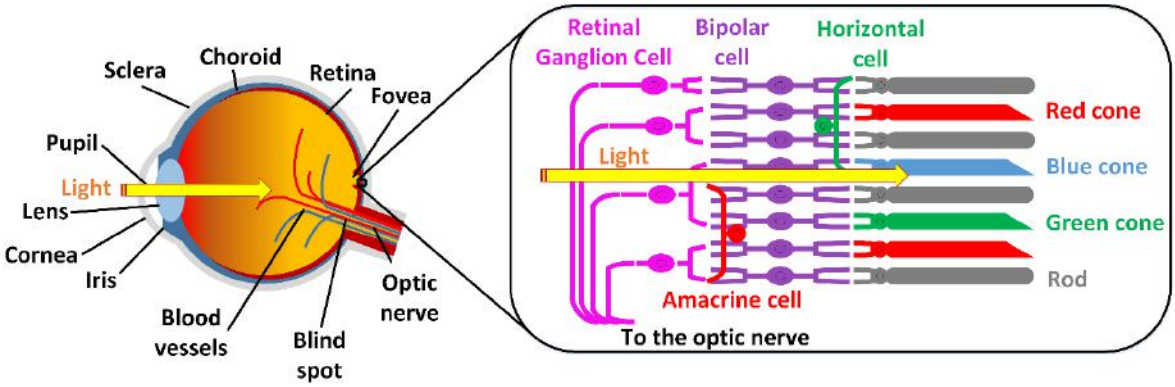


*Fig. 2 Cross-section schematic of the eye with enlargement of the retina.*

Light travels through the cornea (the transparent part of the front of the eye) in an amount decided by the aperture of the pupil, set by the muscles *dilator pupillae* in the iris. The light is focused by the lens through the vitreous humor (the transparent fluid which fills the eye) to the photosensitive retina, whose cross-section is enlarged in Fig. 2. The retina consists of over sixty different types of neural cells [46], covering with its 200 μm thickness [47] approximately 72% of the internal surface of the eyeball. The retina is supplied with nutrients and oxygen thanks to blood vessels lying inside the choroid, the connective vascular tissue of the eye which interposes between retina and sclera (the white of the eye).

The retina receives light from its back part, through a layer of neural cells, called Retinal Ganglion Cells (RGC). The light travels through until it reaches the layer of rods and cones (named after their shape). These are photosensitive cells which, unlike other neurons, hyperpolarize (get inhibited) when stimulated by light. It is important to note that cones are mostly sensitive to color and intensity changes. These photoreceptors are distinguished in rods and red/green/blue cones. Cones, which are about 7 million, are situated mainly inside the fovea, a small cavity in the retinal tissue which is the point of focus of the lens and which encodes the most of the visual information for daylight vision (photopic vision). 64% of cones are sensitive to red, 32% to green and 2% to blue. The rods, which are about 120 million in the human retina, are instead mainly distributed around the fovea (where they are absent). They are sensitive to peripheral motion and are suited for vision in the dark (scotopic vision) since their photosensitivity is centered around the 550 nm wavelength (although color discrimination at night is not possible since only one type of photoreceptor is active). Their response is however slower than that of the cones. Both photoreceptor types are absent at the blind spot, the place where the optic nerve exits the eye.

Indirect interaction between photoreceptors is set by horizontal cells, providing inhibitory feedback [48]. The function of these cells is thought to be local gain control: the horizontal cell network is connected by conductive gap junctions to sense the spatiotemporal average scene illumination to subtract it from the output of the photoreceptors it is connected to. This gives the photoreceptors a much wider operating range and enhances edge-perception.

These photoreceptors then synapse with bipolar cells, which act as intermediaries with the RGCs. There are about 12 types of bipolar cells [49] and each can extract a different information from the multiple photoreceptors it is connected to. There are for example ON, OFF and ON-OFF type bipolar cells. The first, ON, respond to an increase in firing rate for a light stimulus. The second, OFF, respond to a decrease in firing rate to a light stimulus, which conversely corresponds to an increase in firing rate for a "dark spot" stimulation. OFF bipolar cells are the most common type in the retina, perhaps as a result of evolution. Recognizing a dark object (a possible predator coming from above) could prove fundamental to survival. The third, ON-OFF, only respond with rectified activity for the transition of the stimulus. The bipolar cell type ON or OFF is dictated by the expression of the glutamate receptor, which effectively inverts the sign of the membrane potential change in the cell [50]. Other bipolar cells respond to particular wavelengths and act as band-pass filters. Bipolar cells then signal to the postsynaptic RGCs, which processes the combined information of various cells. It is important to note that the information signaled by the bipolar cells is rectified through a synaptic non-linearity before being integrated by the RGCs dendrites [51]. In the Object Motion Cell behavior modelling, which is described in the following chapters, the bipolar cells are considered the subunits of the retina. Bipolar cells connecting directly to RGCs contribute to excitation while inhibition is mediated by fast amacrine cells, which release GABA or glycine. RGCs gather inputs from various bipolar cells which compose their Receptive Field (RF). These can be overlapping and their size is about 1 mm. RGCs axons then proceed as the optic nerve through the optic chiasm to the superior colliculus in the midbrain and to the Lateral Geniculate Nucleus (LGN) of the thalamus. These centers of higher level of processing then project to V1, the visual cortex, which in turns maps to other regions of the cortex for more abstract visual interpretations.

The RGCs' architecture is important as it constitutes the first basic processing of the scene inside the retina. There exist 15-20 types of RGCs, some sensitive to object color, size, direction [52], speed of motion [53] and spatial frequency [54]. The first types of RGCs were first described in 1938 by the Nobel Prize winners Granit, Hartline and Wald [55]. The type of RGC strictly depends on what type of bipolar cells it contacts and on the spatial stimulation of its RF [56]. The RFs of RGCs are, as a matter of fact, weighted spatially giving rise to the ON/OFF center OFF/ON surround topology, whereby a light stimulus in the center of the RF causes excitation of the RGC and inhibition, if in the periphery of the RF. The center-surround topology can be modelled as a Difference Of Gaussians (DOG) [57]: a steep, narrow Gaussian function dominating in the center of the RGC's RF, antagonized by a wider, but smaller in amplitude, Gaussian function which takes over in the surround. The difference in firing rate between the ON/OFF surround and OFF/ON center encodes the contrast of the stimulus. Populations of RGCs, each tuned to detect a particular feature, also encode for scene spatial structure and stimulus strength in their absolute and relative spike timing [58]. Some RGCs are also tuned to very specific intermediate contrasts as their firing pattern maximally responds to them [59]: this allows to discriminate different light-intensity transitions. Direction-selective RGCs can even encode for stimulus direction of motion (if it matches their preferred direction [60]), velocity (in their steady-state spike pattern) and acceleration (in their transient responses) [61]. The RGC's temporal spike code has been successfully modelled with the interplay of fast and slow amacrine cells' inhibition [62]. The variety of information produced by the various populations of RGCs is then conveyed and processed at brain structures downstream the visual pathway.

Ganglion cells are tuned to a variety of object sizes depending on the size and morphology of their RFs. Smaller RFs tend to be more sensitive to fine-grain and high-contrast stimuli, for example. As regards colors, color-opponent RGCs have also been described [63]: with their center-surround morphology, color-opponent RGCs are represented by Red-Green and Blue-Yellow sensitive cells. An interesting behavior is the increase in size of RFs during dark adaptation, as rod-pathways start to substitute cone-pathways [64].

A particular type of RGC, the melanopsin ganglion cell, predicted in 1927 [65] and confirmed in 1998 [66], does not make use of cones and rods as photoreceptors, but is instead photo-sensitive itself thanks to a photo-pigment. The melanopsin, as a matter of fact, although its low sensitivity to light, allows this RGC to detect ambient light changes through its dendrites. Since these RGCs represent only 1% of the total RGCs and are about half a millimeter large [67], the spatial resolution is also low. Melanopsin RGCs are responsible for a number of reflexes to light, for example pupil constriction in response to light, which makes them the biological counterpart of auto-iris in electronic cameras. Finally, they contribute to regulate the circadian rhythm, signaling the light/dark phase of the day which sets the cycle to the Supra-Chiasmatic Nucleus (SCN) of the hypothalamus.

More recently new types of RGCs have been discovered through electrophysiological recordings on salamander and mouse retinas [47], [68]–[70]. The newly identified RGCs are sensitive to looming (approaching) motion of dark object [68] and to local object motion (while being inhibited by global motion) [14][15]. These two RGC types were the focus of the EU project VISUALISE [71]. This work has been carried on by the VISUALISE project partners of the University of Göttingen, under the supervision of Prof. Dr. Tim Gollisch. By projecting various natural scene stimuli, such as photographs and videos, on the animal's retina, lying onto a 16 x 16 Multi-Electrode Array (MEA), a multitude of extracellular recordings were obtained and analyzed. Thanks to the complex stimuli the approach and object motion RGCs were again observed and also new types of RGCs of yet unknown functionality were documented.

## 2.2. Göttingen VISUALISE Workshop

A VISUALISE workshop in Göttingen was organized in February 2014 and I joined it as a representative of the University of Zurich. For a week I worked closely with experienced researchers in the field of visual computation and retina's biology, assisting in experiments and participating in discussions. I observed the preparation of the salamander Axolotl (Ambystoma Mexicanum) retina. This consists of the adaptation to the dark of the salamander's eyes and of its immersion in ice to anesthetize it before operation. After an hour the salamander's retinas are carefully extracted and placed onto a special MEA to keep them alive during stimulation. The entire laboratory is kept completely in the dark to avoid external light stimuli. A picture of the setup is shown in Fig. 3.

The special MEA has a glass cylinder where the salamander's retina is placed. While contacting the electrodes the retina is immerged into a solution of nutrients, constantly being pumped into the cylinder, which keeps it alive for a few hours. The top projector then projects the stimuli onto the retina.

I have also recorded data with a DVS silicon retina, under the same experimental setup used for electrophysiological recordings of real salamander retinas, as seen in Fig. 3. The same stimuli

proposed to the salamander retina were directly projected and focused onto the pixel array of the DVS sensor, without the use of an on-board camera lens. The obtained data was used by the partners of the Italian Institute of Technology (IIT) of Rovereto, Italy, to compare natural and silicon retina's reactions to visual stimuli and to validate the programmed visual computation models. The acquired dataset is available under request.

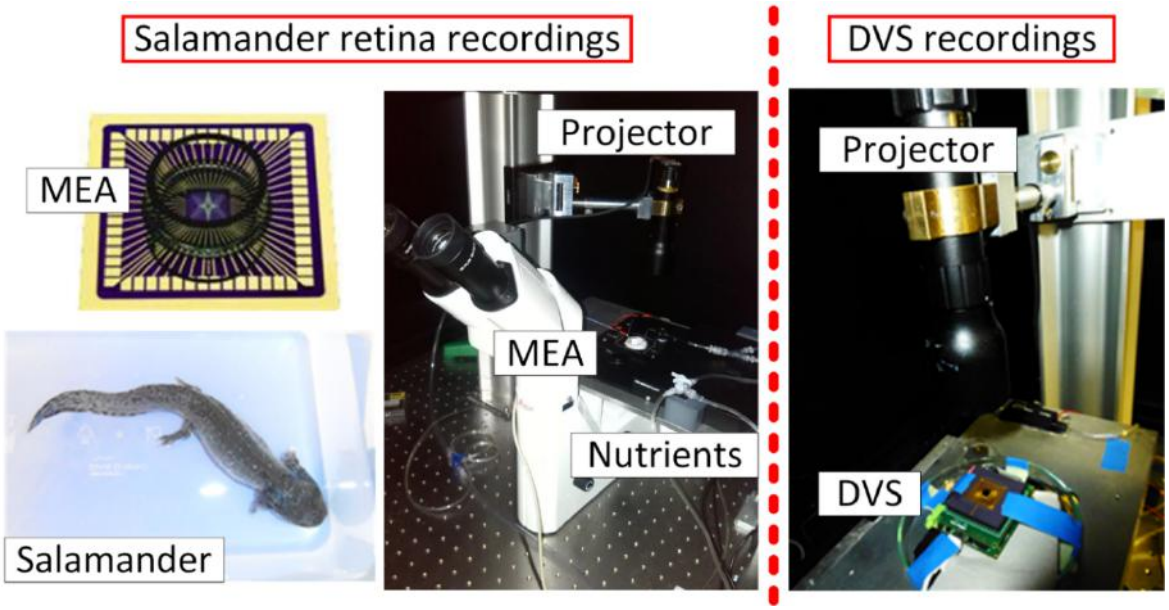


*Fig. 3 Left of dotted red line: retina electrophysiological recording setup. Right of dotted red line: DVS recordings setup.*

## 2.3. Object Motion Cell implementation in software and FPGA

The idea of mimicking RGCs in software and hardware is to use the fast early vision system feature extraction and processing to allow object detection and tracking even under demanding visual conditions. Such algorithm fits perfectly with the fast DVS sensor, which mimics the fast photoreceptors of the retina. An example of this work is the model of the Approach Cell (AC) RGC, discovered by Botond Roska's laboratory [68], started by Tobi Delbruck in 2013, continued by me in 2014, currently being implemented by Hongjie Liu and available at [72]. This cell model detects approaching dark objects by calculating the imbalance between ON and OFF activity of its subunits (bipolar cells). If motion is not approaching but only lateral, then ON and OFF activity cancel out. Such algorithm can be used in a robot which needs to avoid a predator or objects falling from above.

Continuing on this line of thought, the object motion sensitive RGC [69], [70] was the object of the following published work. This cell fires for local object motion in the center of its field of view (up to a radius of 250 μm of the total 0.5-1 mm RF) and gets inhibited by global synchronous motion in its periphery (the remaining field of view), for example the result of a saccade. This non-standard center-surround response RGC, just like other RGCs, gets its non-linear rectified inputs from at least 20 of its subunits: the bipolar cells (each with RF of 20-50 μm). Fast transient bipolar cells at the center of the RF excite the RGC while surround inhibition is mediated by fast starburst amacrine cells. Rectification comes from the synapses from bipolar cell to RGC. If a saccade happens, then all subunits are stimulated and their net synaptic input is summed up instantaneously. This results in the silencing of

the Object Motion Cell (OMC). If instead only a local motion happens at the center of the OMC, then excitation prevails and the cell fires.

This model was implemented first in jAER [24] at first, the software which processes the events which come from the DVS sensor. To create a neuro-inspired object tracker, an array of 16 x 16 bipolar cells subunits resulting in 15 x 15 OMCs was created. Their 2 x 2 excitation centers were tiled across the field of view of the DVS camera with unity stride. All OMCs share the same surround which is as wide as the entire field of view of the sensor. This particular architecture allows to cluster the spikes of each OMC, sensitive to object motion at the specific location of its excitation center. A cluster of spiking OMCs can then be assigned a tracker, with its geometrically computed center of mass. In a staring scenario (the sensor has no ego-motion), the algorithm works very well [73] and up to two trackers can be activated. In the case instead where the DVS also has ego-motion, the OMC is inhibited by the synchronous scene motion and the tracking stops. Even boosting the strength of the center excitation to still allow the OMC to react to strongly moving objects does not help as features from the background suddenly appearing in the field of view of the DVS get recognized as objects. The algorithm is just too simple for complex tasks. This is the main problem of the OMC algorithm which does not solve the object-background segmentation problem necessary for robot navigation. Nonetheless, since the tracker works well for a staring sensor, it was further implemented into FPGA hardware. Due to the restrictions in memory of the device and, even with large design simplifications, the number of OMCs implemented on FPGA is of only 9 cells, just enough to recognize simple object directions of movement.

Both software and hardware FPGA models have been published in the International Symposium on Circuits And Systems 2016, which took place 22-25 May in Montreal, Canada. The conference paper, titled “Retinal ganglion cell software and FPGA implementation for object detection and tracking” is a collaboration between the Robotic and Technology of Computers Laboratory of the University of Seville, Spain and the Sensor´s group at the Institute of Neuroinformatics, ETH Zürich and University of Zürich. The authors are: Diederik Paul Moeys, Antonio Rios-Navarro, Tobi Delbruck and Alejandro Linares-Barranco. The paper, reported in the following pages, was written during my two-week stay at the University of Seville in September, were I worked under the supervision of Prof. Dr. Alejandro-Barranco. Antonio Rios-Navarro was responsible for the Opal Kelly-based communication system which allowed parameters to be passed to the algorithm.

# *Retinal ganglion cell software and FPGA model implementation for object detection and tracking*

Diederik Paul Moeys[1], Tobi Delbruck[1], Antonio Rios-Navarro[2] and Alejandro Linares-Barranco[2]
[1]Institute of Neuroinformatics, University of Zürich and ETH Zürich, Switzerland
[2]Robotic and Technology of Computers Laboratory University of Seville, Spain

## Abstract

This paper describes the software and FPGA implementation of a Retinal Ganglion Cell model which detects moving objects. It is shown how this processing, in conjunction with a Dynamic Vision Sensor as its input, can be used to extrapolate information about object position. Software-wise, a system based on an array of these of RGCs has been developed in order to obtain up to two trackers. These can track objects in a scene, from a still observer, and get inhibited when saccadic camera motion happens. The entire processing takes on average 1000 ns/event. A simplified version of this mechanism, with a mean latency of 330 ns/event, at 50 MHz, has also been implemented in a Spartan6 FPGA.

# I. Introduction

To increase power efficiency, decrease data rate and latency, neuromorphic sensors have been developed over the last 30 years. The Dynamic Vision Sensor (DVS) [21] is an example of such category of devices which draws inspiration from the real functioning of the retina. This vision sensor outputs temporal contrast of logarithmic intensity, asynchronously, through Address Event Representation (AER). The AER protocol encodes the x-y address of where the change happened, to which a microsecond timestamp is added. The in-pixel processing imitates the inherent processing of the Retinal Ganglion Cells (RGC) in biological retinas with microsecond resolution, allowing tasks such as high-speed tracking [29]. Therefore, to investigate further the efficient processing of visual information of the brain, this work tries to mimic further the pre-processing intrinsic to the retina, computed by a specific type of RGC, described in [70]. The Object Motion Cell (OMC) detects local motion by getting excited by small moving objects and being inhibited by large synchronous global motions of the scene (saccades). This feature can be used to extract basic information about motion of a target object. A simulation of an analog imager based on the OMC was previously attempted in [74] but was centered around a different model and technology.

# II. The Circuit Model

## A. The observed neural mechanism

The mechanism of the OMC is possible due to its excitatory-inhibitory center-surround morphology. Although the Receptive Field (RF) of the cell is in practice more like a two-dimensional (2D) Laplacian function, it can be simplified to a 2D top-hat function, with positive weight in its center and negative on its outside. The basic algorithm on which this cell is based is summarized in Fig. 4 and is reported in the following steps. The RF of the cell is composed of subunits of similar sizes: these represent the single bipolar cells of the retina. The central subunits are excitatory while all the other subunits are inhibitory. Since the bipolar cells are not inhibitory themselves, their inhibition is theorized to be mediated by fast amacrine cells [70]. When a change in brightness is detected somewhere in the RF of the cell by a hyperpolarizing cone, the membrane potential of the bipolar cell connected to it is increased linearly. Then, a non-linear rectifying transformation is applied to it. While the subunits are integrating they also decay due to an ionic leakage, adapting to the present visual situation. The RGC contacting the bipolar cells of the exciting center and the inhibiting amacrine cells then integrates the net synaptic input (the difference between excitation and inhibition) on its own membrane. If this is higher than its response threshold, the cell fires. The cell works such that if there is a perfectly synchronous motion in the inhibitory surround, the center excitation is cancelled. Otherwise if the excitatory subunits are triggered and not compensated for, the cell spikes.

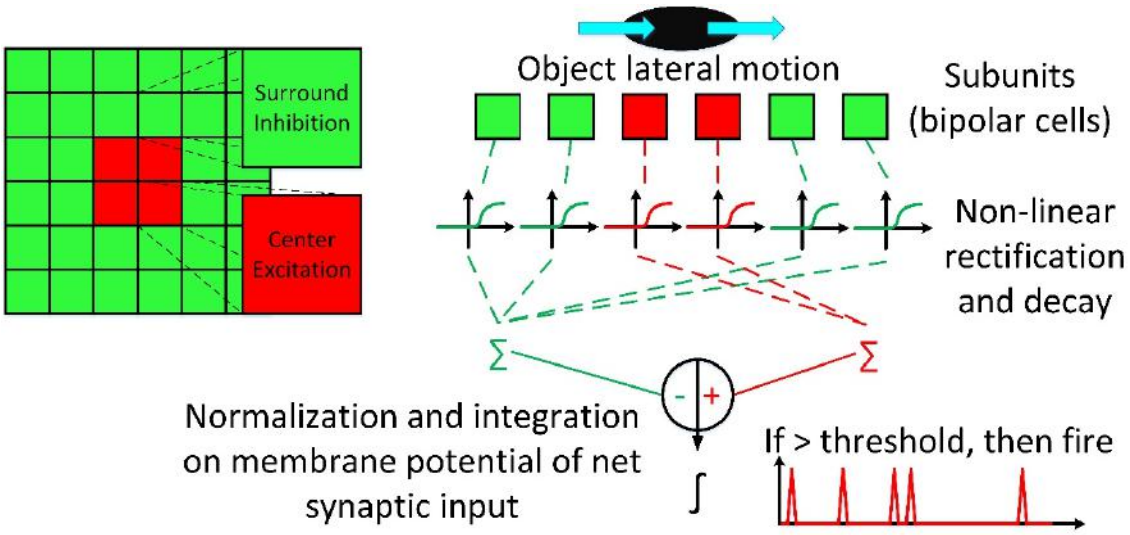


*Fig. 4 Object motion cell's simplified computation.*

## B. Single jAER OMC implementation

To mimic the behavior of a simple OMC, an algorithm has been developed in jAER, the software which processes the events of the DVS neuromorphic sensor [24]. To create the subunits of the OMC, these are set by subsampling the address of the incoming events, making it possible to scale the size of the single subunits by a power of two, to better fit the size of the object to be detected. Then all subunits are set to be inhibitory apart from the central four which are excitatory. When an event is received at a particular x, y coordinate within a subunit, its membrane potential is increased linearly by one unit and its non-linearity is calculated. This non-linear rectification can be set to be, in the developed model, of exponential type of any order or of exponential tangent type. In the first case, a clipping is artificially placed on the membrane potential of the single subunit, so that if too much activity is registered, a single subunit cannot constantly dominate over other weak subunits. In the second case, the natural saturation of the exponential tangent already inherently performs this operation. To model the ionic leakage which makes the subunit's membrane potential decay if no further activity is detected in its RF, an adjustable exponential decay with time constant $\tau_s$ is computed at every event timestamp received and applied to the subunit.

For the whole array of subunits, the total surround inhibition is computed by adding the non-linearized membrane potentials of each subunit and by normalizing. The same is done for the center excitation. The excitation can be scaled by a synaptic weight $\alpha$. This empirically ensures stability and allows the OMC to be adjusted to different visual scenes. The final membrane potential of the RGC is computed by integrating the net synaptic input and, if larger than the adjustable threshold, the cell fires, signaling object motion detection. The RGC's membrane potential is also decayed exponentially with time constant $\tau_n$. If the exponential tangent non-linearity is chosen, then the computation of the OMC can be modelled as:

$$\begin{cases} if\ \ V_m e^{-t/\tau_n} \leq V_{IF} \quad do\ not\ fire \\ \quad if\ \ V_m e^{-t/\tau_n} > V_{IF} \quad fire \end{cases} \tag{2.3.1}$$

Where $V_m$ is the integrated membrane potential of the OMC expressed in equation (2.3.2), $t$ is the timestamp time, $V_{IF}$ is the Integrate and Fire (IF) threshold to be overcome to spike.

$$V_m = \int_0^t \left( \alpha \frac{\sum_{i=1}^{4} tanh\left(V_{ex_i} e^{\frac{-t}{\tau_s}}\right)}{4} - \frac{\sum_{i=1}^{k-4} tanh\left(V_{in_i} e^{\frac{-t}{\tau_s}}\right)}{k-4} \right) dt \qquad (2.3.2)$$

where $\alpha$ is the weight of the excitation, $V_{ex_i}$ and $V_{\mathrm{in}_i}$ are the *i*th excitatory and inhibitory subunit membrane voltages respectively and $k$ is the total number of subunits. As an option, the total inhibition (or excitation) can be computed, not only by the membrane potential but by its difference to the neighboring subunits. This local normalization removes the problem of global dimming, which could trigger the RGC response.

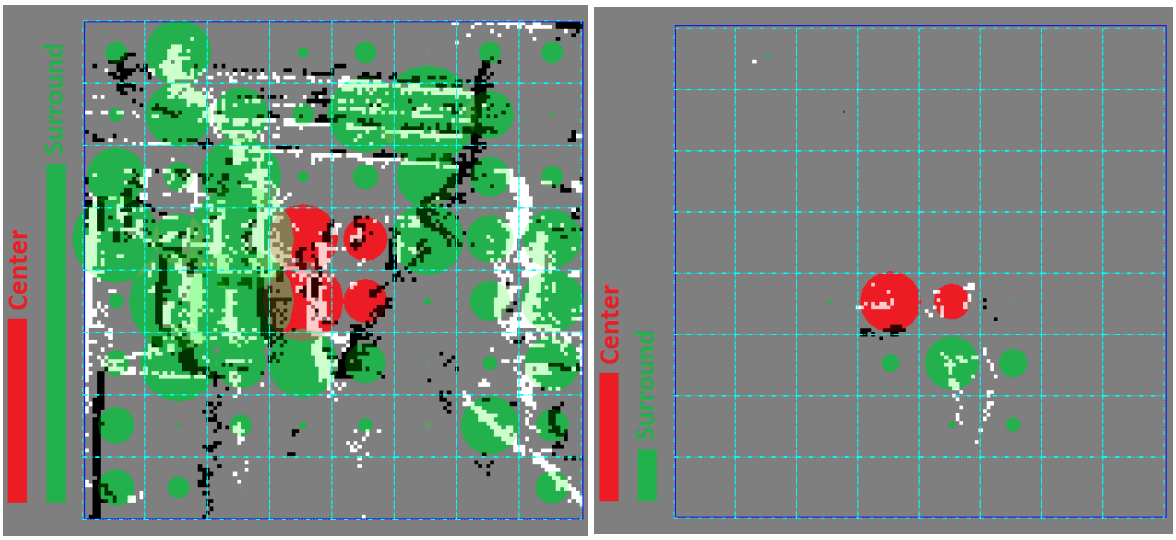


*Fig. 5 jAER output showing the DVS camera output events (ON in white, OFF in black). Left: saccade inhibiting the OMC. Right: single object movement exciting it. The green disks represent the activation of the inhibitive subunits and the red ones represent the excitation. The bars on the side compare the normalized activities. The positions of the subunits are highlighted by a grid.*

The simple OMC was initially tested in jAER with 8 x 8 subunits. The algorithm works for different natural visual stimuli such as the ones shown in Fig. 5. On the left image, the DVS camera output shows a saccade view of the office. On the right side, only a person moves in the scene. As can be seen, when the camera is moved in a saccade, the inhibition subunits are maximally active (green disks) and compensate for the excitatory input (red disks). If the local motion at the center of the RF is instead not compensated, then the cell fires. This can be seen in the graph of Fig. 6, which illustrates a plot of center excitation and surround inhibition with a second order non-linearity added to each subunit.

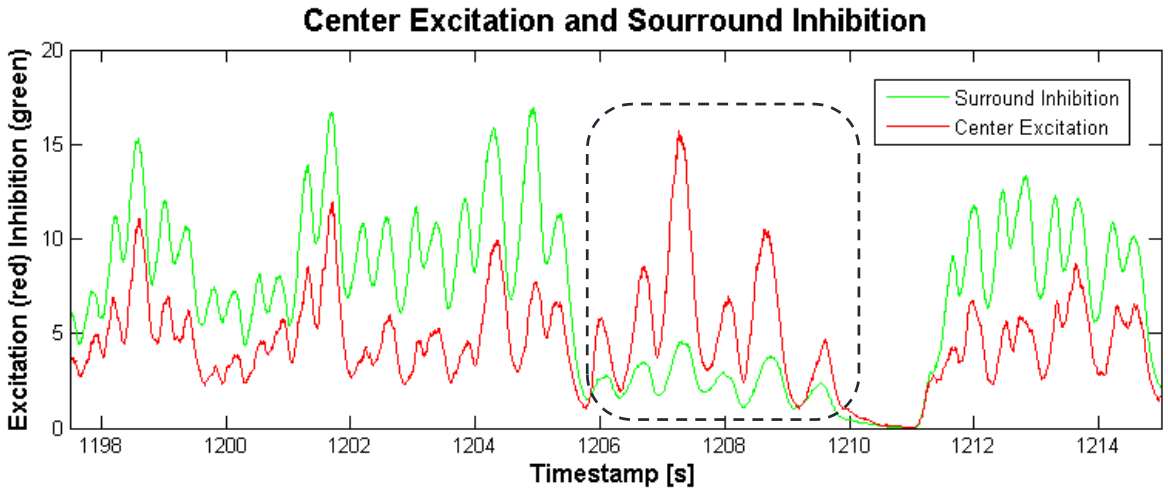


*Fig. 6 Plot of surround inhibition versus center excitation for the video of Fig. 2 in arbitrary units. The non-circled parts correspond to global motion (excitation smaller than inhibition). The circled part corresponds instead to local motion (excitation larger than inhibition).*

# III. Multiple OMC Tracker jAER Implementation

## A. Multiple jAER OMCs implementation

To make use of the OMC for tracking, it is important not to just detect the object's presence but also its direction. To achieve this goal, an array of 16 x 16 subunits was set up in jAER. By sliding the 2 x 2 excitation center of the OMC across all subunits with unity stride, 15 x 15 OMCs (this number is dictated by the size of the target) can be constructed. To speed up computation, all subunits are regarded as the inhibiting surround, including the central ones, so that the total inhibition can be computed just once for all OMCs with minimal error. This would mean that the term $k-4$ in the right hand side of equation (2.3.2) simplifies to $k$. The final result is that now all these overlapping cells respond to object motions at specific locations. To make the algorithm more robust, the average event rate is used. For the DVS128 sensor, if this is lower for example, than 500 ev/s (events/second), then the IF threshold of the RGC can be set to a high value. This is because for such low event rates the pixels producing a response constitute mostly random activity due to leakage in the reset transistor of the DVS pixel [21]. Increasing the threshold prevents the cell from firing for no relevant activity. An upper threshold can also be set if the activity is too high: in such case it is likely that the sensor either moves very close to a high contrast wall or that the target is too close and covers the entire field of view. In this case, tracking is not necessary and can be suppressed. Any event rate in between these boundaries can be associated with a moving object. The event rate numbers used in this design have been obtained empirically by placing a 128 x 128 DVS sensor on a moving robotic platform following another robot.

## B. Tracking

The single OMCs spiking in the presence of a moving object can be easily clustered, so as to obtain the position of the moving object by correlation. A tracking scheme was implemented to draw a containing box around the last 3 spiking cells close in time and space, and to find its center of mass by geometry. If some OMCs spike due to a second object moving in the scene, therefore beyond the

reach of the first tracker, then these outputs can be associated to a second tracker. Double tracking is shown in the left image of Fig. 7, for a still camera staring at the scene where two objects are moving independently. The trackers reset automatically when no OMC near them spikes within an adjustable amount of time, and reappear at another location where an object seems to be moving. This way the tracker can be reused once the first object stops moving and the reset event can trigger the memorization of its last known position.

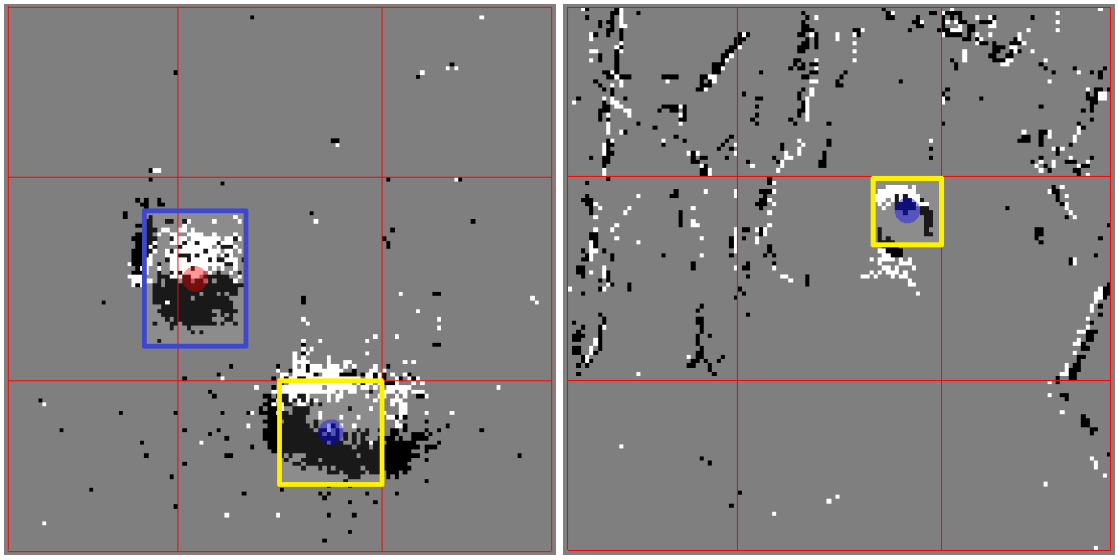

*Fig. 7 15 x 15 array of OMCs: on the left, two objects activate both the trackers (yellow and blue with respective blue and red center of mass), on the right the moving object can still be tracked while the camera is moving in the scene thanks to excitation weight adjustment. The position of the center of mass with respect to the 9 red quadrants can be used by a robot to plan its next action.*

By dividing the field of view in 9 quadrants and knowing the position of the center of mass of the tracker within one of them, it is possible for the robot carrying the sensor to plan its next move and follow, or just shoot, the target. The size of the target can also be roughly estimated with a very simplified inverse pinhole camera model in the x direction knowing the width of the target object and ignoring lenses' distortion. The numbers need to be heavily low-pass filtered in order to obtain a reliable measurement, but the order of magnitude of the result is at least consistent with the ground truth.

Since the algorithm works in such a way that the OMC gets inhibited in the case of global motion, the same happens when the observer is moving and tracking gets suppressed. To still allow the OMCs to fire and continue tracking even while the observer is moving at moderate speed, parameters need to adapt. An approach to solve this problem effectively is to increase the weight of excitation $\alpha$ by a fixed amount when a certain activity, indicating apparent scene motion, is detected. This way the firm movement of an object can still cause the OMCs to fire even though inhibition is stronger. This algorithm works however only if the apparent motion of the scene is slower than the one of the target object. Also, if corners or high-contrast features suddenly appear in the field of view, these might be detected as objects to be tracked. Only the temporal and spatial correlation of the OMCs spiking which are part of the tracker can guarantee that the correct object is still followed. This can be seen in the right image of Fig. 7.

# IV. FPGA implementation

## A. Hardware used and multiple OMC FPGA implementation

The FPGA design was approached to explore the possibilities of implementing complex cell types in logic to exploit parallelism. The OMC mechanism was prototyped in the Spartan6 XC6S1500FXT Xilinx FPGA board developed in [75], called AERNode board. The design of this platform allows multi-board communication with conventional parallel-handshake-AER chips, serial Low-Voltage Differential Signaling (LVDS) connections or robots with the adequate motor interfaces. A daughter board based on an OpalKelly module, called OKAERTool, is used for monitoring, or sequencing, and logging, or playing, events from and to the AERNode board. It is in fact able to sequence events from its on board DDR2 128MB SDRAM to the AERNode board and to monitor its output through USB2.0 in jAER. OKAERTool is fundamental for debugging the design implemented in FPGA.

Due to the limited number of resources of the FPGA available (gates and memory), the OMC implementation strategy was changed into a much more simplified one. Five OMCs' centers were fitted in the center of each quarter of the field of view and in its center to obtain the most basic directions of a moving object. Every OMC center consists of 2 x 2 subunits. The design consists of two levels: a Mother Cell (MC) which deals with the four-phase AER handshake protocol (request and acknowledge) with the outside neighboring blocks and five inner Daughter Cells (DC) which, in parallel, each calculate the excitation of a particular OMC. The MC calculates the inhibition (set to be the entire field of view, as in the jAER model) and feeds it along with the incoming request to the DCs if the input event falls within their excitation center. The MC also manages the global high-priority decay of the subunits by a counter. The DCs then propagate their request, in case of firing, to the following stage through the MC. Since the cells all work in parallel and store their firing in a one-hot coded output vector, the processing delay does not scale with the number of DCs active. The active low requests are anded so that the request to the next stage is active for at least one DC active.

The algorithm of each single DC follows the most basic jAER implementation, however, to reduce the use of resources variables are restricted to 16 bit values and every operation is simplified. Since divisions (for normalization of inhibition and excitation) are performed by even numbers, these are done by multiples of 2 by bit-shifting. The same bit-shift operation substitutes the non-linearity and a saturation is achieved with a comparator when a certain value is attained. The global decay is also achieved with bit-shifts and it is adjustable by the counter's limit. Finally, only one multiplication is present, the one for the integration of the single daughter OMCs' membrane potential. Three parameters can be set via Serial Peripheral Interface (SPI) through the OKAERTool. These are the IF threshold, the decay counter's limit and the excitation weight $\alpha$.

## B. System integration and hardware

The system of OMCs, enclosed by the MC, was integrated with a pre-existing system architecture. This was achieved by creating a separate, parallel processing branch through a splitter and a merging element. This can be seen in Fig. 8. The input request and parallel data which the OMC receives are the same that the cascade of filter elements receives (in this case the Hot Pixel Filter, which filters out addresses of pixels with high spike rate). In case of firing, the OMC sends its request further to a

merging arbiter along with its output data (the one-hot-coded firing DC's address). The arbiter decides randomly which branch will be serviced first and acknowledges it after reading its data. The acknowledge signals of both branches propagate back to a Muller C-element latch which combines them into a single one which is then sent back to the event source (the DVS or the event sequencer).

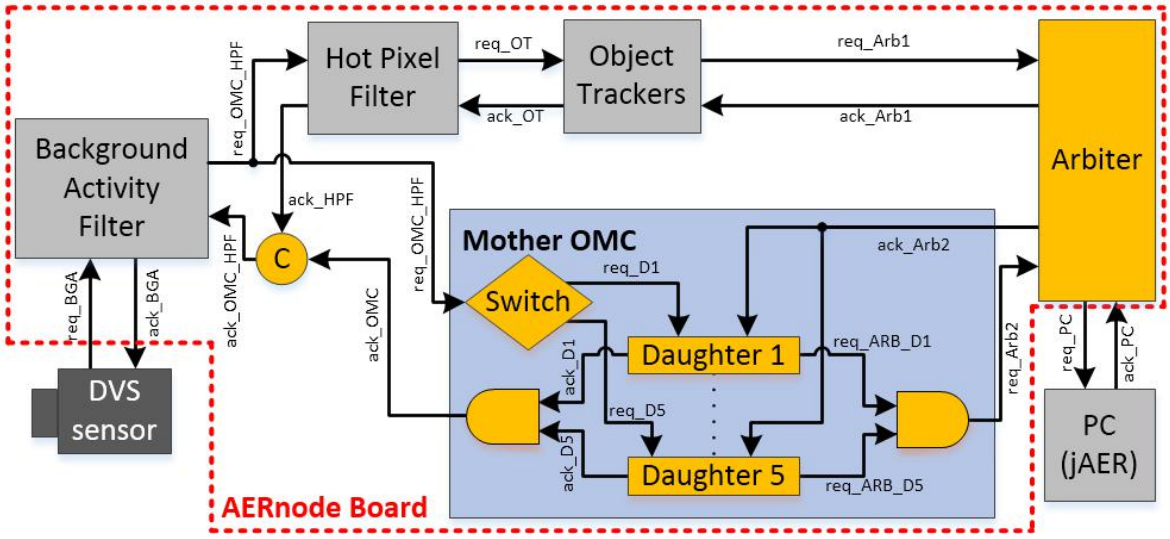


*Fig. 8 Integration of the OMC with the existing architecture of [30].*

A problem which was encountered during design was that if both branches would be requesting at the same time, the arbiter would be waiting for the request withdrawal of the branch requesting first, before servicing the second branch. The request withdrawal cannot happen until the acknowledge signal of the branch is propagated back to the event source. Since, due to the C-element, this cannot happen until both acknowledge signals are received, the system effectively deadlocks. To overcome this problem caused by the dependency of the merger from the splitter, a feature of collision detection was added to the arbiter: the latter can in fact now acknowledge the two branches one after the other even though the request of the first is not yet de-activated. The implementation of Fig. 8 for 5 DCs uses 4% of the slice registers, 11% of the slice Look-Up Tables (LUT) and occupies 16% of the available slices. For 9 DCs, the resource consumption changes to 5%, 12% and 19% respectively.

# V. Results

## A. Latency and power consumption comparison

To estimate the delay of the OMC tracker in both jAER and FPGA, the time taken for an input event to be processed was measured. For the jAER OMC tracker, the nanoTime() method of class System was used to measure the processing time for 3 different numbers of OMCs (though the difference between 5 and 9 cells is irrelevant). It should be noted, however, that this delay increases by 25-30% if the input events are not being read directly from the hard disk (as it is the case for the reported numbers) but if they are obtained in real-time from the DVS sensor. This is because events are processed at a higher speed (at the maximum of the system's capabilities) if read from a logged file

rather than if the data is obtained from the real world. For the FPGA performance, the numbers were obtained using the Xilinx ChipScope tool, since the processing delay was below the microsecond event timestamping resolution. The results are summarized in Table 1. The OMC in FPGA takes 22 and 11 clock cycles, depending if the incoming event falls in one of the DC's RF, to process an input event and complete the AER handshake with a 50 MHz clock. As regards the power consumption comparison, the FPGA has a factor of at least 100 of advantage over even the small embedded Intel Next Unit of Computing (NUC).

## VI. Conclusion

This paper offers two implementations of the OMC model for the purpose of object detection and tracking: one software-based and FPGA-based. A comparison study is presented between these two systems and highlights the power consumption and latency advantages of the FPGA. Due to its parallelism, the nanosecond latency does not scale with the number of OMCs implemented. This paper however does not yet present the next step: the recreation of the 15 x 15 OMCs tracker of jAER into FPGA. This is because the resources of the Spartan6 constrain for the moment the OMC array size. Knowing that the number of occupied slices for the 5 OMCs design is 16% and for the 9 OMCs design this is 19%, it can be approximately estimated that at most another 108 OMCs, each taking 0.75% of the resources, can be fitted into the current design (still 117 below the desired 225). The work to come will therefore focus on further reducing the size of the DCs and on the optimization in the integration with the existing architecture. At the moment, the output of the OMCs is already used to choose the location where to initialize the trackers of [30] and to validate their operation: the object tracker which is active can now only exist if its center of mass location falls within the RF of a DC firing.

| **Specification Table** | | | |
|---|---|---|---|
| | jAER (64-bit Intel NUC, 4 GB RAM, i5-4250U, 1.30 GHz) | jAER (64-bit PC, 16 GB RAM, i7-4770K, 3.50 GHz) | FPGA (Xilinx Spartan6, 50 MHz) |
| Latency of 5x and 9x OMCs | ~500 ns/ev at 0.2 Mev/s, at CPU load < 5% | ~250 ns/ev at 0.2 Mev/s , at CPU load < 2% | 220 or 440 ns/ev at any event rate |
| Latency of 15x15 OMCs | ~1000 ns/ev at 0.2 Mev/s , at CPU load < 5% | ~500 ns/ev at 0.2 Mev/s , at CPU load < 2% | NA |
| Power | 6.2 W static + 6.2 W for running jAER + 2.48 W for 5/9x OMCs or 3.72 W for 15x15 OMCs | A few hundreds of Watts | 0.775 W static + 0.05 W for 5/9x OMCs |

*Table 1 Specification Table of the FPGA OMC.*

## Acknowledgments

This research is supported by the European Commission project VISUALISE (FP7-ICT-600954) and by the Spanish grants (with support from the European Regional Development Fund) BIOSENSE (TEC2012-37868-C04-02) and the Andalusian Council Excellence grant MINERVA (P12-TIC-1300). We thank the Sensors group at INI and the RTC lab in Seville.

## 2.4. Live demonstration

A live demonstration paper [76] associated with the conference paper [2] of the previous chapter was also accepted at ISCAS 2016. The paper has the same four authors in the same order: Diederik Paul Moeys, Tobi Delbruck, Antonio Rios-Navarro and Alejandro Linares-Barranco. It is simply titled: "Live Demonstration: Retinal ganglion cell software and FPGA implementation for object detection and tracking". The live demonstration of the implemented systems was performed at ISCAS 2016 as seen in Fig. 9, with great interest from the conference participants who could freely interact with the two setup present: the software jAER-based model OMC tracker running on the Intel NUC PC and its simplified FPGA model running on the AERNode Spartan6 board. The cluster object tracker of [30] was also demonstrated: the clusters were initialized and validated by the OMC FPGA object detector.


*Fig. 9 Live demonstration setup at ISCAS 2016.*

A video demonstration of tracking applications at low and high-speed of the OMC is available at [73]. In the first segment, the ability of a single OMC to distinguish object motion from background motion thanks to inhibition (green disks) and excitation (red disks) subunits is presented. In the second segment, low-speed tracking is presented. The pink squares represent the OMCs' centers (tiled across the entire field of view) which are firing. The yellow and blue boxes with the blue and red dot respectively represent the extracted and tracked clusters with their center of mass. Finally, in the third segment, high-speed tracking of a spinning dot rotating at 1000 rpm is presented. The video is slowed down to highlight the microsecond resolution of the DVS.

## 2.5. More details on the FPGA implementation

The journal paper titled “Low Latency Event-based Processing for Dynamic Vision Sensors on FPGAs” was submitted to Transactions on Circuits And Systems I (TCAS I) [77]. The paper was rejected and is under new revision for a second submission to the same journal. The paper is a collaboration between the Robotic and Technology of Computers Laboratory of the University of Seville, Spain and the Sensor´s group at the Institute of Neuroinformatics, ETH Zürich and University of Zürich. The authors (among which I am third) are: Alejandro Linares-Barranco, Francisco Gomez-Rodriguez, Diederik Paul Moeys, Fernando Perez-Peña, Gabriel Jimenez-Moreno, Shih-Chii Liu and Tobi Delbruck. The paper combines the detailed description of four processing blocks, implemented in FPGA into the Spartan6 of the AERNode board used in [2] and previously in [30]. The first is the background activity filter: a processing step which only lets through address events which are correlated in space and time. This removes background noise events due to leakage in the reset switch of the DVS pixel [21]. The second is the mask filter: this processing step detects hot pixels with a high firing rate independent of the scene activity. These non-functioning pixels, if not detected and removed in the computation, can cause problems to the following processing steps by creating false positive detections and taking over computation. The paper then features the cluster object tracker of [30] which estimates position, velocity and center of mass of the tracked object. Finally, the object motion detector of the previous chapter [2] is also featured. The description of the Finite State Machine (FSM) of the OMC is given in all details, including the state diagrams of mother and daughter cell. More details on the latency of the filter are also given.

## 2.6. Conclusion

The study of the OMC has led to an interesting investigation in the functionality of RGCs in the early visual processing. The usefulness for tracking in a static scenario has proved that this bio-inspired approach has potential and should be further explored. However, the limitations of the algorithm while the camera is in self-motion, show that the OMC is not sufficient alone to solve entirely the problem of background-foreground separation necessary for robust object tracking. The algorithm has to be complemented by other methods, possibly other models of RGCs. The idea would then be to merge the information stream from the OMC into another, acting as a weight for the overall algorithm’s decision.

The OMC could be used in the future as a complement of the deep learning architecture introduced in chapter 3, by weighting the regions of the field of view containing object movement, when no ego-motion is present.

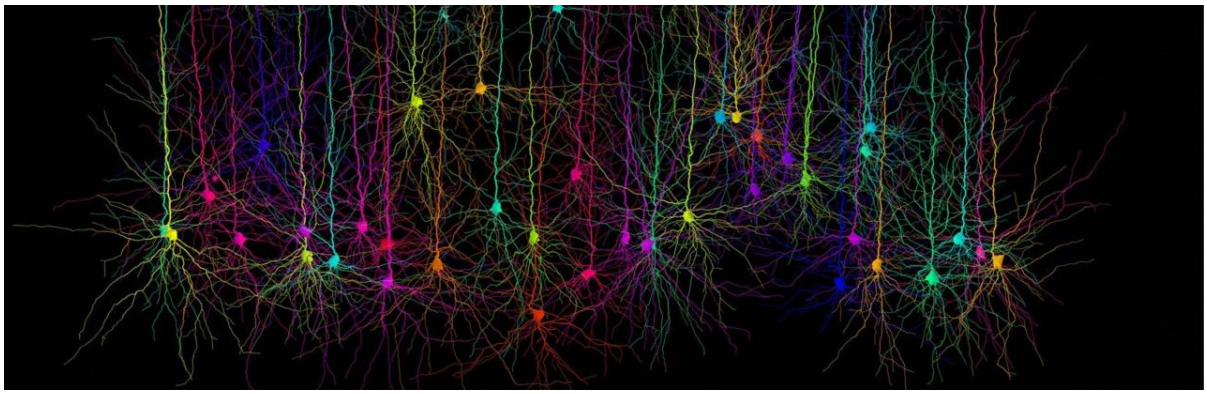

Image adapted from [78]

# *Chapter 3*

## 3. Convolutional Neural Network for steering a robot in a predator-prey scenario

Due to the limitations of the Object Motion Cell (OMC) algorithm in the context of tracking during ego-motion, a new investigation in the field of machine learning was proposed. The following chapter introduces briefly the basic concepts of Artificial Neural Networks (ANN), Deep Neural Networks (DNN) and in particular of Convolutional Neural Networks (CNN) and how they can be used in this context to solve the problem. Although the approach is different, the method does not stray away too much from the computation type of Retinal Ganglion Cells (RGC), at least conceptually. As it will be shown, the patterns that the network looks for in a scene are very similar to the basic Gabor orientation filters and other more complex receptive fields of retinal circuits. Overall the work focuses on the concrete predator/prey scenario, the aim of the entire work, in which a predator robot obtains steering directions from the positioning of the prey in its field of view. Successful trial runs are presented as well as ongoing improvements are presented.

## 3.1. Convolutional Neural Networks

### 3.1.1. Machine learning and computer vision

Machine learning is a sub-topic of computer science in which computers learn to solve a particular task by constructing an internal model of the task. While no specification is given about the model, the algorithm does not follow fixed instructions, but, instead, needs to learn from a large set of example inputs to make predictions about the output. The algorithm is therefore data-driven and it is evaluated on a test set by its percentage accuracy after being trained on a training set. If the desired outputs (labels) of the computer are given, then the learning is supervised. If the computer needs to infer labels depending on the features of the data instead, the learning is said to be unsupervised.

Machine learning is nowadays extremely popular in the field of computer vision [79] as the request for visual intelligence increases (face recognition, self-driving cars, quality inspection, to name a few). In particular, the center of focus is on image classification, where the task is to label an image with a class label depending on its content. To do so, computers need to learn both dimensionality reduction (extracting only the important features of the image) and expansion (extracting features that are not immediately visible from the combination of other features). Image classification needs to be robust and deal with image distortion, occlusions, different viewpoints, illumination, scale, colors and so on.

The complexity of the training algorithm (number of parameters) can be increased in order to increase its accuracy (or equivalently decrease its error) on the training set as well as the test set. However, beyond a certain point of complexity, the algorithm can start to overfit the data as the training error further decreases but the test error increases again. One of the hard tasks of machine learning is to achieve generalization and to avoid overfitting by providing an as large as possible training dataset. Data availability (with included labels) is therefore one of biggest concerns in machine learning and datasets of a million images are not uncommon.

### 3.1.2. Artificial Neural Networks and Deep Learning

In machine learning, Artificial Neural Networks (ANN) are networks of interconnected nodes (artificial neurons) vaguely similar to the neural circuits of the animal brain which can approximate linear or non-linear functions. The ANN can have various topologies (the arrangement and connections of the units), but mainly consist of an input layer, a single or more hidden layer and an output layer whose units each represent a label. The neuron-to-neuron connections are weighted and are subject to the activation functions of the neurons. The activation of a neuron is the function that the neuron applies to its weighted and summed inputs. This could be a sigmoid, a hyperbolic tangent or a Rectified Linear Unit (ReLU). Weights, which are the strength of the connections and resemble biological synapses, change during training as well as biases, which are just offset terms added before non-linearities at each stage of the network.

These parameters can be tuned by a variety of learning algorithms. An algorithm always has score function and a loss (or cost) function. The score is the function $f(x_i, W, b)$ which gives, for an input, the scores of the output classes [80]. For a linear classifier this is shown in equation (3.1.2.1):

$$f(x_i, W, b) = Wx_i + b \tag{3.1.2.1}$$

where $x_i$ is the input image, $W$ is a set of trained weights and $b$ is a bias vector, also trained.

The loss function is needed to minimize the error between computed and desired output for a certain input and it is more complex. It is then used to adjust the weights $W$ and biases $b$ to obtain a lower error or, equivalently, a higher accuracy. The loss function is the key for successful learning and many different and complex ways to solve this have been investigated [81]. Once the loss function is computed, backpropagation is applied to the ANN. This consists in changing the weights and biases in order to minimize the loss. This is done in various ways but the most common is Stochastic Gradient Descent (SGD), where the values of weights and biases are changed proportionally to the negative of the gradient of the loss function, expressed as a function of the network's parameters (in order to reach a minimum) [79]. The input to output gradient is computed stage by stage with the chain rule as the output of the network is composed of linear operations of functions of other differentiable functions. The method is said to be stochastic as the recursive optimization works on a system which is not precisely mathematically modelled, but only estimated with noisy observations (it is subject to random variables). Only the rough properties of the non-modelled system are extracted in order to attempt to find minima and maxima. Because each neuron is an independent unit in each layer, ANNs can be implemented in a parallel fashion for high-speed and can pair well with Graphical Processing Units (GPU).

Although conceived as early as 1943 [82], ANNs only became popular again because of the success of Deep Learning (DL) in the early 2000s in the field of image recognition (see section 3.1.3). Deep Neural Networks (DNN) are ANNs which have more than one hidden layer. The complexity of functions they can perform is therefore increased. Each layer extracts from the activation of the previous layer a different representation at a higher level of abstraction. These abstractions are not easily understood by humans as the network deepens but current work tries to investigate it [83]. DNNs lack a strong theoretical background regarding their methods and improvements as most of these come from trial and error and iterative processes.

### 3.1.3. Convolutional Neural Networks

Convolutional Neural Networks (CNN) are a subtype of DNN feedforward architecture particularly suited for 2D data (images) [84]. Feedforward signifies that the computation is unidirectional from input to output [85]. CNNs are inspired by the work of Hubel and Wiesel on the cat's visual cortex and by their discovery of the concept of receptive fields [86][52][51][50][49][48][47]. Starting from the first primitive Neocognitron [87] of 1980, CNNs first breakthrough was with Yann LeCun, who improved CNNs in 1989 [88] by successfully applying backpropagation for training, in the context of postal code digit recognition [89]. Since then, dramatic improvements to CNNs and their applications in visions have been made [90][91][92][83][93][94][95], all mainly based on GPU. CNNs are also currently being converted to the spike domain [96][97] in order to fully exploit neuromorphic hardware implementing on-chip convolution [98][99][100][101].

An example of a CNN can be seen in Fig. 10B. As the name suggests, CNNs involve layers of convolution after the first input layer. A patch of pixels forming a pattern (a kernel of *K x L* weights) is slid all across the *M x N* image with a certain stride in order to convolve with *(M-K+1) x (N-L+1)* pixels (assuming a stride of 1). If a patch of the image matches strongly the pattern of the kernel, the resulting output pixel (obtained by dot product) will have a high value. In the opposite case, the value will be low. The output image resulting from a convolution is called a feature map. With a single grey input image, in the first convolution layer, each kernel produces one feature map. If the kernels that the network learns are Gabor filters (orientation detectors with a certain spatial frequency), then the

resulting output image of the convolution process will be an image where that particular orientation will be highlighted. Every convolution layer performs therefore, through the kernel, a weighted sum, which is then subject to the pixel-wise activation of the neuron. Nowadays, the preferred activation is the aforementioned ReLU [92]. This function passes either zero if the pixel value is negative or the pixel value. Interleaving convolution layers are often max-pooling layers which also increase accuracy. In this layer of stride 2, only the highest value of every four pixels is selected to compose the output feature map, which is therefore down-sampled. The trick of down-sampling serves the purpose of applying the following convolution at a larger scale, in order to extract larger features at a higher abstraction level, since the image is already processed. The convolution behavior is similar to the function of simple cells in the visual cortex and max-pooling roughly approximates the behavior of complex cells [91]. If the second convolution layer has, for example, four output feature maps, then it will apply four different kernels to the four output feature maps of the previous layer. Each of the four feature maps that comes from a single feature map of the previous layer gets summed pixel by pixel. Therefore, although there are 16 kernels, only four output maps are produced and fed to the following layer. Finally, one or more linear classifiers (fully connected layers, also called perceptrons) expand the dimensionality of the feature maps for further levels of abstraction and map to the output layer. The output layer consists of a set of output neurons each corresponding to a label of the input. The strongest output is selected as the winner. Usually, the outputs predictions of the network are passed through a softmax (normalized exponential) function which acts as Winner-Take-All (WTA) and "squashes" the output values. This way, normalized and more intuitive class probabilities are obtained: the outputs range 0-1 and sum to 1 in total (as if they were probabilities). If the network guesses wrong, the example is used for backpropagation to minimize the cost function and change the weights and biases according to a certain learning rate (the step taken to change the weights).

A certain percentage of the weights can be randomly reset to zero during training in order to help highlight stronger features. This is called dropout [102][103] and it can significantly improve the network accuracy. This can also prevent overfitting or help to get out of a local minimum in training. Other techniques to prevent overfitting are called regularization techniques. One of the most common is weight decay [104], were weights are decayed to zero in order to keep weight values as small as possible. This prevents the weights to favor particular images and grow disproportionally.

Many training sets for various tasks are available to train classifiers such as CIFAR10 [105], Caltech101 [106], COIL20 [107] and the Street View House Numbers (SVHN) Dataset [108]. One of the most famous testbenches is the MNIST hand-written digit recognition dataset [109] (60000 training images and 10000 testing images) on which the current lowest error record was set by [110] to be 0.23%. Many competitions take place every year to assess networks in the task of image recognition. The most famous competition is ImageNet Large Scale Visual Recognition Competition (ILSVRC) [111], currently containing 14.2 million labeled images belonging to one or more of the 22k classes. The competition regularly brings out the new state of the art networks: AlexNet (2012) [92], ZFNet (2013) [83], GoogLeNet (2014) [93], VGGNet (2014) [94] and ResNet (2015) [95] and CUimage (2016).

In conclusion, CNNs are interesting because they learn to extract features from their input example training set which would otherwise need to be handcrafted. This results in their applicability in many scenarios such as: tracking [112][113], image- [114][94], face- [115][116] and location-recognition [117], image segmentation [118][119], self-driving cars [120][121] and so on.

## 3.2. Steering a robot with a DAVIS sensor

A small CNN with only 4 output classes (Left, Center, Right and Non-visible) was developed to process the spiking and frame data of a DAVIS sensor and to then give steering directions to a four-wheeled predator robot. The steering directions match the position of the second prey robot, human-controlled, roaming around in a robot arena at the University of Ulster. The work was one goal of the VISUALISE project [71]. The CNN, running in jAER in an i7 on-board PC, would communicate via UDP its decisions directly to the Robot Operating System (ROS) controlling the servo motors of the robot. The ROS controller was initially developed by Dr. Gautham Das of the University of Ulster and then by PhD student Emmett Kerr, Dr. Philip Vance and myself. The training of the CNN was performed offline in Caffe [122] and the CNN obtained was exported to XML file format to be then re-imported in jAER. This was the most tedious part of the work and would have not been possible without the help of Dr. Federico Corradi. The exported kernels of the convolutions had to be written out in XML from a Python script in order to be re-imported in Java, which would expect a format identical to the Matlab Deep Learning Toolbox. All of the languages give a different reordering with the *reshape* function which complicated things further. The entire work took a lot of debugging time. Finally, PhD student Daniel Neil managed to apply a deconvolution analysis using the Lasagne neural network software in order to visualize what the receptive fields and patterns the network look for in a scene (see Fig. 15 and Fig. 16, where guided-backpropagation is applied). A deeper analysis of the significance of the CNN's kernels is currently under investigation. A video of the final review of VISUALISE, where the work is demonstrated, is available at [123].

A conference paper was published on the work done in the Second International Conference on Event-Based Control, Communication and Signal Processing (EBCCSP) 2016, which took place 13-15 June in Krakow, Poland. The conference paper, titled "Steering a Predator Robot using a Mixed Frame/Event-Driven Convolutional Neural Network" [3] is a collaboration between the Intelligent Systems Research Centre, University of Ulster, Londonderry, Northern Ireland and the Sensor´s group at the Institute of Neuroinformatics, ETH Zürich and University of Zürich. The authors are: Diederik Paul Moeys, Federico Corradi, Emmett Kerr, Philip Vance, Gautham Das, Daniel Neil, Dermot Kerr and Tobi Delbruck. The paper, reported next, was written during my three stays at the University of Ulster.

FUN FACT: after the paper with this catchy title was uploaded on Arxiv [124], we were contacted by *Motherboard*, the technology online newspaper of VICE [125] for an interview. The newspaper produced quite a decent article regarding the work done [126] and so did CNBC shortly after [127]. Little did we know that the work would go immediately viral: many online newspapers copied and rephrased the material to fit their clickbait purpose, adding information which did not exist. The scientific work got turned to Terminator references and apocalyptic scenarios. [128] is my personal favorite: "Rise of the 'Terminators': Super-intelligent predator robot is taught to hunt down prey in chilling experiment" with sentences like "The test comes as experts warm AI could wipe out a tenth of the global population in five years." Some more examples here: [129], [130] and [131]. In any case, occupying the first 10 pages of Google just by typing "robot hunt" gave us some exposure as well as distrust for low-level online media!

# *Steering a Predator Robot using a Mixed Frame/Event-Driven Convolutional Neural Network*

Diederik Paul Moeys[1], Federico Corradi[1], Emmett Kerr[2], Philip Vance[2],
Gautham Das[2], Daniel Neil[1], Dermot Kerr[2], Tobi Delbruck[1]
[1]Institute of Neuroinformatics, University of Zürich and ETH Zürich, Switzerland
[2]Intelligent Systems Research Centre, University of Ulster, Londonderry, Northern Ireland

## Abstract

This paper describes the application of a Convolutional Neural Network (CNN) in the context of a predator/prey scenario. The CNN is trained and run on data from a Dynamic and Active Pixel Sensor (DAVIS) mounted on a Summit XL robot (the predator), which follows another one (the prey). The CNN is driven by both conventional image frames and dynamic vision sensor "frames" that consist of a constant number of DAVIS ON and OFF events. The network is thus "data driven" at a sample rate proportional to the scene activity, so the effective sample rate varies from 15 Hz to 240 Hz depending on the robot speeds. The network generates four outputs: steer right, left, center and non-visible. After off-line training on labeled data, the network is imported on the on-board Summit XL robot which runs jAER and receives steering directions in real time. Successful results on closed-loop trials, with accuracies up to 87% or 92% (depending on evaluation criteria) are reported. Although the proposed approach discards the precise DAVIS event timing, it offers the significant advantage of compatibility with conventional deep learning technology without giving up the advantage of data-driven computing.

# I. Introduction

The DAVIS is a neuromorphic camera which outputs static Active Pixel Sensor (APS) image frames concurrently with dynamic vision sensor (DVS) temporal contrast events [21][32]. DVS address-events (AEs) asynchronously signal changes of brightness. Deep neural network (DNN) technology use has become widespread and there are a large number of tools available for application of this technology. Our aim in this study was to develop the simplest-possible use of DNN technology as applied to DAVIS sensor output. We wanted to preserve something of the data-driven nature of the DAVIS DVS output together with maximum compatibility with existing DNN tools. The aim of this paper, which was a work in parallel with [132] and more functional alternative to [2], is to concretely explore an application in robot navigation.

# II. Setup and recordings

The purpose of the network is to steer the predator robot in the direction of the prey robot. The simplest possible implementation of the predator needs to recognize in which of the 3 vertical regions of the field of view the prey is. These are Left/Right/Center and Non-visible (LCRN). The *N* output signals that a search for the prey should be initiated. This approach is inspired by the seminal work in [133], [134] and most directly by the forest trail detection and following work of [135], but applied in an indoor scenario with an event-based vision sensor. This limited dimensionality of 4 outputs was determined as the minimal useful case; other networks that also output the prey distance or analog position in the field of view were ruled out by the limited availability of training data at different prey distances. The following two sections introduce the setup used to gather the 500'000 images dataset to train the Convolutional Neural Network (CNN) to perform the task.

## A. Robots

Fig. 10A shows the two Robotnik Summit XL mobile robots [136] used in the experiment and the arena. These 750 x 540 x 370 mm robots have four omnidirectional Mecanum wheels that allow skid-row kinematics thanks to small rollers on them. They weigh 40 kg and they can move up to 3 m/s (4 body lengths/second, or the equivalent of 60 km/h for a 4.5 m long car) in the 9.5 x 6.7 m arena. Our experiments were limited to maximum 1.5-2 m/s to prevent possible damaging crashes. The robots are fitted with an Inertial Measurement Unit (IMU) and a laser scanner to detect and avoid collisions. The robots are controlled through the Robot Operating System (ROS) framework [137], running on the embedded PC with Linux with Intel Core i7 processor. They communicate through WiFi 802.11n to allow access to their operating system. The DAVIS sensor is mounted on the predator robot through a simple mounting hole with lock/wing nut. The 2.6 mm wide angle lens provides a horizontal field of view of 81 degrees. Lighting was varied by turning the room lights on and off, but the Vicon tracking system, which caused flicker highlights on the floor, was turned off. The windows were not shaded and experiments were conducted under sunny and cloudy conditions. The floor is specular and has stripe patterns, and there are background objects above the walls of the arena. Fig. 10B shows the overall system architecture of the predator robot as described in later sections.

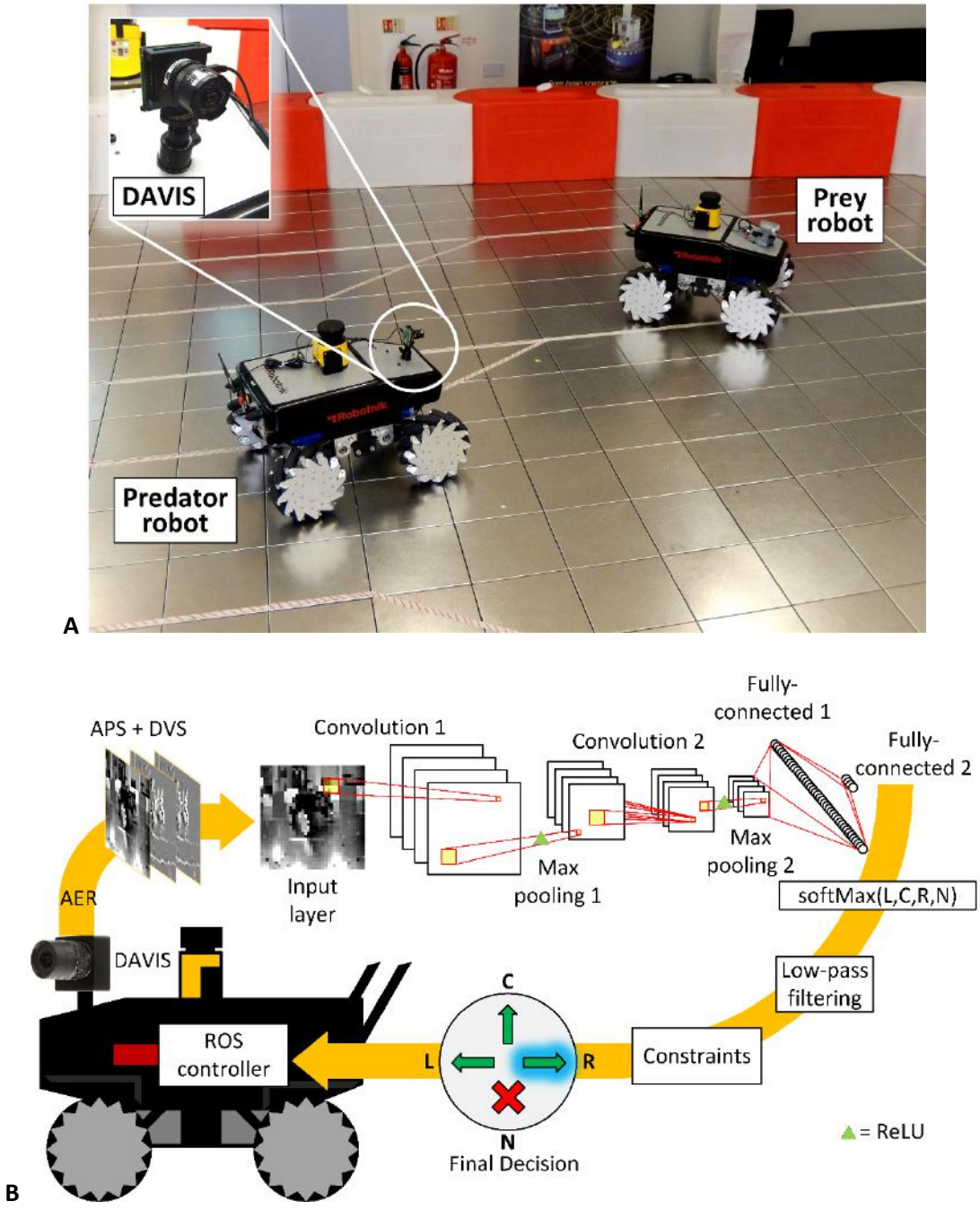


*Fig. 10 A: Summit XL predator (left) chasing the prey (right) in the robot arena of the University of Ulster, Londonderry. B: overall closed-loop system: the DAVIS sensor generates APS and DVS data which is alternately fed to the 4C5-R-2S-4C5-R-2S-40F-R-4F convolutional neural network. The results are filtered and the final decision is used in conjunction with the laser scanner output to control the Summit XL behavior in the ROS controller.*

## B. Recordings and preprocessing of the data

Twenty DAVIS recordings with a total duration of about 1.25 hour were obtained by driving the two robots in the robot arena of the University of Ulster in Londonderry, as seen in Fig. 10A. The predator robot, fitted with the recording DAVIS, followed the prey robot, initially by teleoperation

and later autonomously. The prey robot was teleoperated or controlled by a semi-random policy under laser range finder control. DVS and APS data was obtained under conditions to cover variations in lighting, relative position and distance between robots and speed. Various arrangements of background objects such as a black wheelchair and the interference of people walking in front of the camera were recorded.

The 240 x 180 APS frames were captured in global shutter-mode using auto-exposure (with typical exposure times of 5 ms) from the sensor at an average of 15 fps. A 36x36 CNN input image size was selected as the minimum size by which the robot can still be recognized by human eye and the easy divisibility of the pixel number in the three steering regions.

DVS data is integrated to 36x36 frames as 2D histograms obtained by integrating 5'000 ON and OFF events in 200 possible gray level values, i.e. starting from a pixel value of 0.5, each DVS ON event increases the gray value by 1/200 and each OFF event decreases it by -1/200. The subsampled pixel address (2D histogram bin) is computed by integer division of the event coordinates. Since the DVS frames are sparse, active DVS frame pixels accumulate about 50 events.

The APS frames are resized down to 36 x 36 with nearest-neighbor interpolation. We used nearest neighbor rather than more accurate methods for computational efficiency targeting embedded application on low power processors. Using jAER [24], the software that processes DAVIS data, both DVS and APS data are converted into uncompressed .AVI video format for later pre-processing with MATLAB.

In total, about 75'000 APS and 275'000 DVS frames were obtained. The data was inflated to 500'000 frames by creating falsely over- and under-exposed APS data by shifting the gray values of the frames by a fixed amount and by clipping the data out of range. This increases the amount of training data and the robustness of the network to new exposures. At the same time, this data augmentation balances out the APS/DVS frames ratio for training, bringing it up to 45% and 55% of the whole training set respectively. An example of the raw recording along with APS and DVS frames is shown in Fig. 11.

The ground-truth positions of the prey robot were obtained by manual labelling of the robot position in jAER by capturing the mouse pointer position on the screen during playback of the recordings (using the jAER filter TargetLabeler). Depending on which third of the visual field the mouse pointer falls within, the LCRN label is assigned to the frame. 11% of L, 18% of C, 15% of R and 56% of N compose the final training image dataset, which corresponds to the first 80% of each recording. The remaining 20% of each recording, with LCRN percentages differing by only up to 2%, is used as test set.

Frames are not shuffled randomly before the train and test set separation, because frames are consecutive in time, and so they are highly correlated. Therefore, presenting frames in the test set that are consecutive to those in the training set would result in false higher accuracy, since the network was trained on very similar images, but reduce its robustness to new inputs.

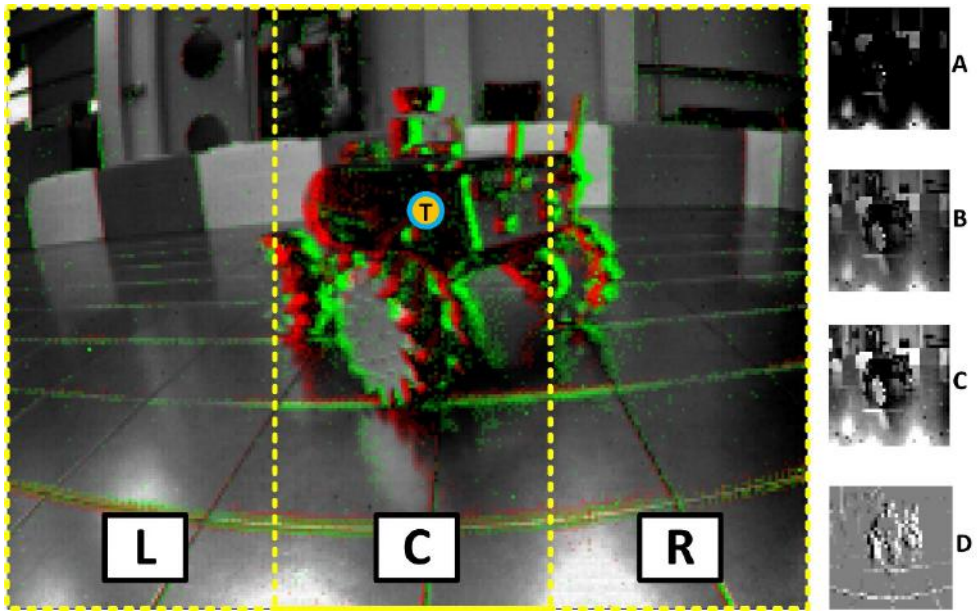


*Fig. 11 Example of raw recorded data with overlaying APS gray value data and DVS data (red are OFF events, green are ON events). The field of view is divided into the three regions and the target is labelled. A, B and C are the extracted 36 x 36 APS frames with falsely created exposures. D is a subsampled DVS histogram.*

## III. Network Topology, Input Normalization, Training, and Error Analysis

We used a CNN because of its proven performance with image recognition [91]. In initial feasibility studies, we first tried to train separate CNNs to process the APS and DVS frames. However, the limited amount of training data invariably resulted in overfitting the training data. To increase the amount of training data, we found that training a single CNN driven sequentially by both types of frames resulted in higher overall accuracy. This approach also simplifies the software architecture. As shown later, it results in CNN first-layer feature detectors that detect robot features in both the APS and DVS frames.

Correct input normalization of DVS and APS frames proved essential for training a network that would work reliably. APS frame input pixel values $F$ are rescaled to the new range $R_n$ 0 to 1 with (3.2.1) to obtain the normalized pixel values $F_n$:

$$F_n = R_n \frac{F - \min(F)}{(\max(F) - \min(F))} \tag{3.2.1}$$

Since DVS histograms are not real images, their histogram level which corresponds to zero events (gray in the image) is held at 0.5 (half of $R_n$), to avoid unwanted flickering, dependent on the ratio of ON and OFF events collected. Variation of such gray background would just complicate the recognition task. The DVS histograms are then clipped at three times their standard deviation σ computed around the mean μ of the originally acquired DVS histogram. This clipping removes outliers and allows the DVS histogram to cover the full $R_n$ keeping 99.7% of the information. The entire dataset extraction and generation from 1 hour of recordings through MATLAB takes about 90 minutes. Fig. 12 shows normalized LCRN examples for APS and DVS.

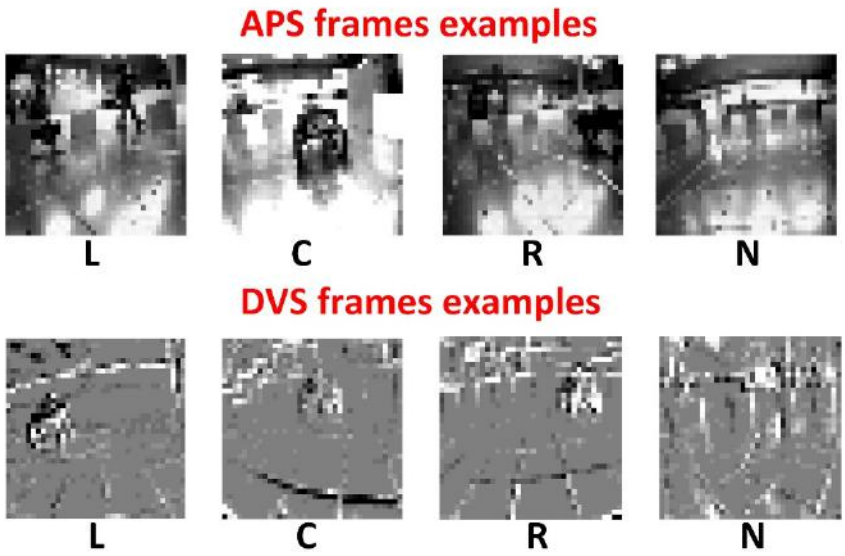


*Fig. 12 Normalized LCRN examples for APS and DVS at 36 x 36 resolution.*

To overcome the speed limitations of the MATLAB toolbox, previously used in [132], the CNN was trained with the Caffe framework [138]. The CNN consists of the following layers: a 36 x 36 input layer, a convolution layer with $N$ output feature maps with $n$ x $n$ kernels (denoted here as $NCn$), a max pooling layer with stride 2 (denoted as $2S$), another convolution layer with $M$output feature maps with $m$ x $m$ kernels ($MCm$), another max pooling layer with stride 2, a fully connected layers of $p$ neurons (noted as $pF$) and finally a 4-neuron output layer that is fully connected to the previous layer. The activation function at the end of the convolution and at the output of the fully-connected layers was tested with both sigmoid activation (noted as $S$) and Rectified Linear Unit (ReLU) activation (noted as $R$) types. The network was trained using a softmax loss function on the output layer. The size of the network was chosen through a manual optimization process aimed at the minimum number of features required to perform the task with acceptable accuracy. The performance of some of the various networks explored for the task is shown in Fig. 13. Running 100'000 training iterations required about 40 minutes of compute time on the largest CNN in Caffe running in an Ubuntu VM, using CPU-only mode. We used a VM for convenience, since this VM could easily be shared among the authors and moved to different PCs.

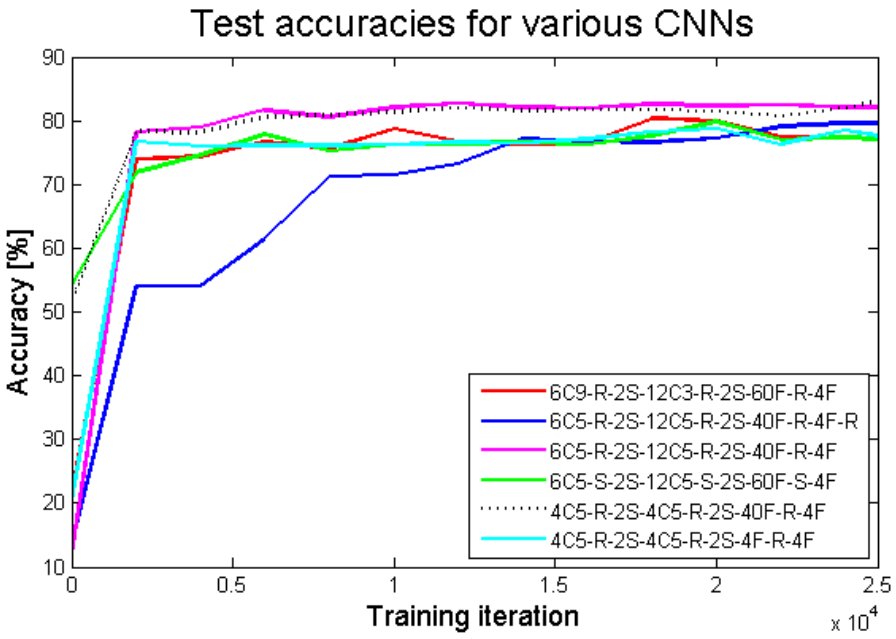


*Fig. 13 Testing set accuracies of various networks versus number of training iterations. The selected runtime network is plotted with the dotted line.*

The maximum achieved accuracy was 93% on this training dataset and 87% on the test dataset. Inaccuracy was due to a number of reasons, summarized in Fig. 14 and the rest of this section. The first consists of DVS histograms containing no information: even though no movement takes place in the scene (both robots pause for a moment), the DVS still integrates 5'000 uncorrelated events due to the leakage in the reset switch of the pixel [21]. The result can be seen in Fig. 14I. These frames are however relatively few since the reset switch leakage of the pixel causes ON events only every 10 seconds. With 240*180 pixels (43'200 in total), the time to reach 5'000 uncorrelated events is just over 1 second. This noise frame rate is well below the normal DVS frame-rate but still affects the over accuracy on the dataset. These background activity DVS frames could be easily filtered out by the jAER BackgroundActivityFilter, but during our experiments we did not use this filter.

Similar to no-movement in the scene, which makes the DVS effectively blind, DVS histograms can be integrated at a higher frequency if both the predator and the prey are still but another object moves, creating events. If a person passes by in this situation, for example, the prey is falsely predicted to be invisible (Fig. 14J). Other noise events can disrupt the proper computation on DVS histograms. The main one is parasitic capacitive coupling and non-optimal biasing. This coupling between the global electronic shutter and the DVS event generation mechanism causes a burst of DVS events on each frame [32] and creates events correlated with the sample rate of the APS, filling up the 5'000 events allowed in the DVS histogram, sometimes covering up the prey robot (especially if far away). This noise can be recognized by the scanning lines (Fig. 14O,P) which indicates strong coupling activity which saturates the event arbiter. In some images, when the robot is far away, this artifact overlaps the prey (Fig. 14O,P) and sometimes the prey is too small even for human detection (Fig. 14C,D,K). In other cases features of the background can take over and be recognized as the prey (Fig. 14G,H,L). An interesting example is a black wheelchair that was present in the corner of the robot arena. This looked in the downsampled image like the prey robot and the only way to recognize it was with correlation with previous frames where the prey position is known to be different. Since the network

learns without any frame-to-frame correlation, then this is a problem that remains unsolved for frames where the robot is far away and of very similar shape to the black wheelchair.

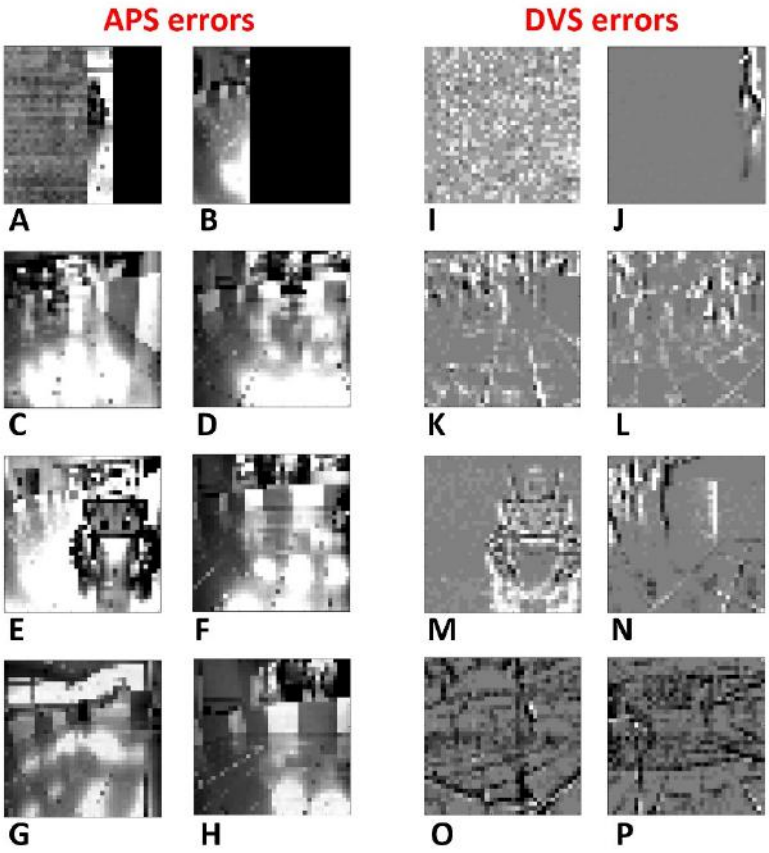


*Fig. 14 Examples (36 x 36 images) that cause wrong network predictions. A and B: corrupted and dropped APS frames. C and D: prey far away and overlapping with other similar shapes. E and F represent ambiguous APS frames: in E, they prey is very close, covering two thirds of the field of view, in F, only one wheel of the prey is present. G and H: false positives caused by a black wheelchair and a high-contrast poster. I: DVS histogram which integrated random background noise due to no movement present in the scene. J: DVS histogram where the prey and predator are still but a moving person triggered the DVS frame integration. This makes the prey effectively non-visible from the DVS output. K: prey too far away to have good resolution. L: wheelchair appearing as a robot to the network. M and N: same ambiguity of E and F. O and P: noise activity integrated in the DVS histogram due to coupling between APS and DVS in the DAVIS sensor.*

Furthermore another factor which deteriorates APS images is the occasional corruption and partial loss of APS frames (Fig. 14A,B). This problem in our experiments was due to incorrect setting of the USB buffer sizes. Since the host side buffers were set too small and APS data arrives in large bursts, some data is occasionally dropped, including frame start events, resulting in corrupted frames. When a black stripe covers the prey, it is impossible even for a human, without previous frame correlation, to know the current position of the robot. In our experiments this problem was handled by low-pass filtering the network predictions.

Finally, the main reason for errors consists in the ambiguity of the frame when the robot crosses boundary regions and the label oscillates around it. Other ambiguities examples are when only a wheel of the robot is visible at the extreme edge of the image or when the prey is close enough to cover two thirds of the field of view of the predator (Fig. 14E,F,M,N). Since the data is hand-labelled

using a subjective interpretation of the robot position in the image, the ambiguous frames are bound to deteriorate the overall accuracy. To prove the importance of this second factor, the accuracy of the test-set was recomputed to eliminate the most ambiguous frames in which the labelled target position is within 1 pixel (in the 36 x 36 image, corresponding to about 6 pixels in the 240 x 180 original image) of the four boundary region. This on average improves the accuracy by 3%. Increasing this error margin further improves the accuracy of the networks. The rest of the ambiguous images are the ones where the prey robot is very close to the predator and more than one LCRN region is covered by it. There are however still some frames for which the correct decisions are very obvious to human eye where the network fails for no apparent reason.

Test and train accuracies are close to each other (on average there is a 5% difference between the two) and indicate that overfitting is minimal. Overfitting was also minimized by using dropout (randomly setting to zero in each iteration a fraction of the weights to reduce weights' co-adaptation) of 20-30% of the first fully-connected layer. The gradient-based optimization method chosen was "Adam" [139], to deal with the large amount of training data.

The size of the network was chosen observing the effect of each of its parts. Regarding the convolution, increasing the size of the square kernel increases the network's accuracy. However, computing time and overfitting eventually increase too. A size of 5 x 5 was found to be optimal. The number of feature maps per convolution layer was reduced to the minimum necessary. We found that a surprisingly small number of features were needed. When we used more than about 4 features per layer, kernels started to repeat or ended up with near-zero weights. The optimum number of neurons in the first fully-connected layer was 40; providing more units did not increase accuracy but leaving out the first fully connected layer significantly reduced accuracy. A ReLU activation function provided similar accuracy as sigmoidal activation, but was much less likely to become stuck in a local minimum where the network would detect only the class most present in the training data. A lower accuracy is also observed for networks terminating with ReLU units, therefore this activation function was used at convolution outputs and the first fully-connected layer output. Since the network decision is always taken as the maximum activated output unit, this finding does not make sense at first, but was probably the result of back-propagating the unbounded ReLU outputs or not backpropogating error reduction from negatively-activated ReLU's.

After exploring many different combinations of parameters, we finally settled on a runtime $4C5\text{-}R\text{-}2S\text{-}4C5\text{-}R\text{-}2S\text{-}40F\text{-}R\text{-}4F$ architecture providing low computational cost and acceptable accuracy. Its kernels of the first and second convolution layer are shown in Fig. 15. The kernels in the first convolution layer extract the most basic features of this predator/prey context with spatially-confined filters. As a matter of fact, kernel 0 and 1 seem to highlight the high contrast of the wheels of the prey and kernel 2 and 3 the edges of the arena walls above and to the right of the robot. The kernels of the second convolution layer combine the first layer features, but their interpretation is more difficult.

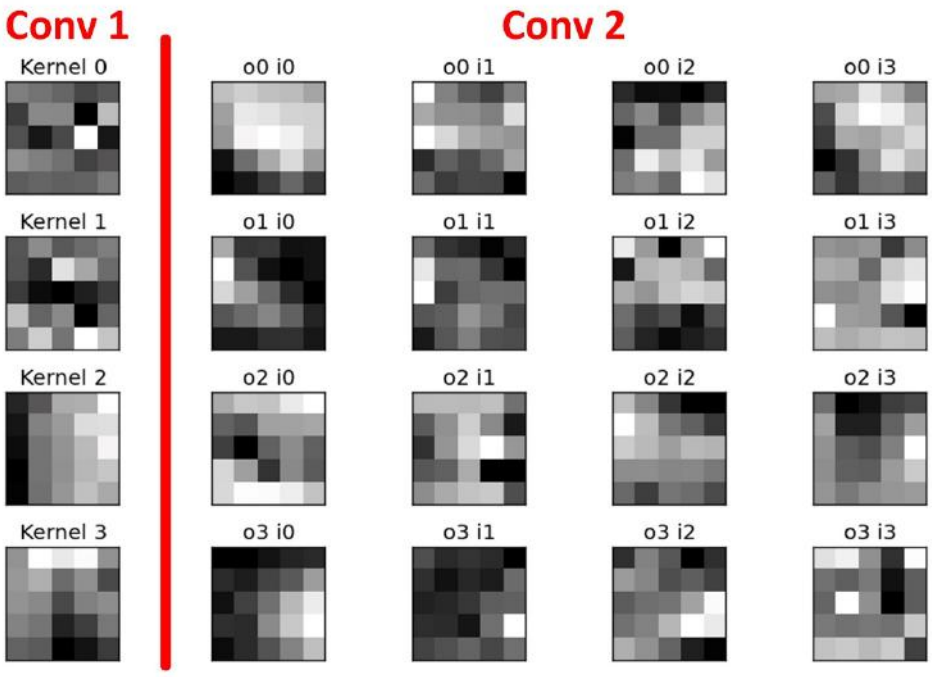


*Fig. 15 Left of red line: 4 kernels of the first convolution layer 4C5 for the chosen network (4C5-R-2S-4C5-R-2S-40F-R-4F). Right: 16 kernels of the second convolution layer 4C5. For example, o2 i1 is the kernel for output unit 2 of the second convolution layer that filters output feature 1 of the first subsampling layer.*

In Fig. 16, the guided-backpropagation method [140] of saliency visualization was employed to determine the parts of a particular input image that resulted in a strongly winning C activation. The guided backpropagation process, which is a mixture of backpropagation based on the input data and a deconvolution of the gradient, hides the influence of negative gradients which decrease the activation of the target neuron while highlighting regions that strongly affect the target neuron. The example shows that for this input, the robot wheel and dark body cause the strong activation, while other features such as the grid and highlights on the floor and the walls and background are ignored.

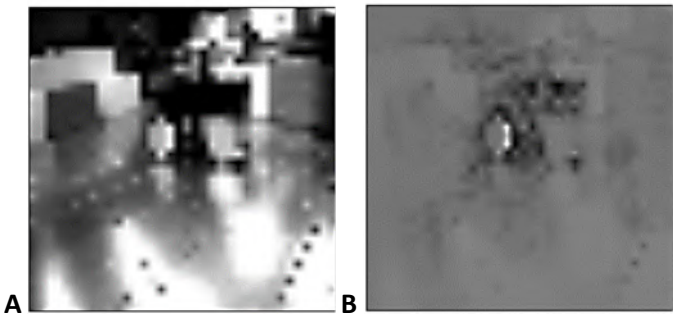


*Fig. 16 Saliency visualization using the guided-backpropagation method. A: the input image (36 x 36) presented to the network. B: Saliency overlay, with highly salient regions transparent to expose the input data that caused the active class label (robot in center).*

## IV. jAER Implementation

jAER is a software project for visualizing, recording, and soft real time processing of event based sensors [24]. A CNN runtime engine along with necessary utilities such as the target labeler were implemented in jAER as the Java package eu.visualize.ini.convnet. The pre-trained weights of the kernels and of the fully-connected layer are loaded from an XML file describing the network into jAER. A python script cnn_to_XML.py in the jAER scripts/python/caffe_utils folder reads the Caffe network architecture and weight files to produce the XML file. The CNN takes as input the downsampled and normalized 36 x 36 APS frames and the subsampled and normalized 36 x 36 DVS histograms of 5'000 events. The forward CNN pass requires between 1-2 ms compute time for the chosen network size on the robot PC, which has a Core i7 processor running Java JDK1.8. The final decisions of the CNN, along with every intermediate activation map, can be visualized during runtime as shown in Fig. 18. When playing back labeled data in jAER, the error rates for both APS and DVS frames can be computed online and compared with the offline results of the training.

## V. Robot Communication

jAER runs concurrently with ROS on the robot. jAER sends its steering decisions to ROS using User Datagram Protocol (UDP). The two integers are sent as two bytes. The first is a sequence number 0-255 to check for lost datagrams and the second is the robot steering direction encoded as L=0, C=1, R=2, N=3. Communicating through UDP allows jAER to run independently as a server on the same computer and not as a ROS slave node. This way its processing can be monitored and controlled without modification of jAER's architecture. The UDP messages are sent at each novel decision. The computations run in jAER at a jAER processing rate of 240 Hz, which is near the fastest that DVS frames are generated. Thus, even though jAER processes the DAVIS data in packets, most packets result in no decision and few packets produce more than one decision. Processing at a lower rate in jAER, with larger packets, would result in multiple decisions per packet, all sent in a very short time, which would not make sense. Fig. 17 shows the measured interval distribution for commands sent to ROS for the trial run #7. The median communicated decision rate was 91 Hz. The APS frame rate was about 15 Hz, and therefore the other 76 Hz rate of decisions was caused by DVS frames. With a DVS frame size of 5'000 events, an average rate of DVS events of about 76 Hz*5'000 = 380'000 DVS events per second (eps) rate is consistent with our observations of the average DVS event rate during this recording. For the fastest trial run describe later, where the robot ran at 2 m/s, the average DVS event rate was about 600 keps, resulting in average DVS frame rate of about 120 Hz, although peak rates often reached the jAER processing cycle rate of 240 Hz.

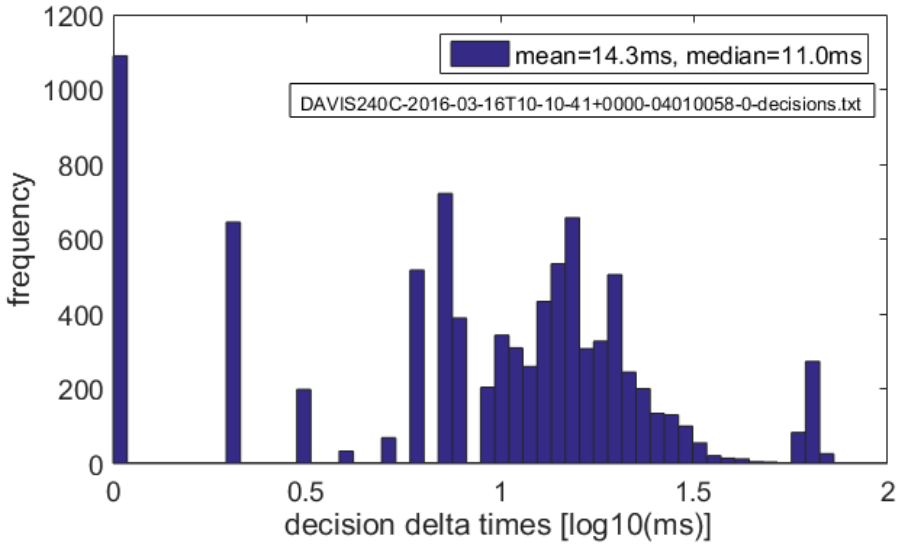


*Fig. 17 Decision output interval distribution.*

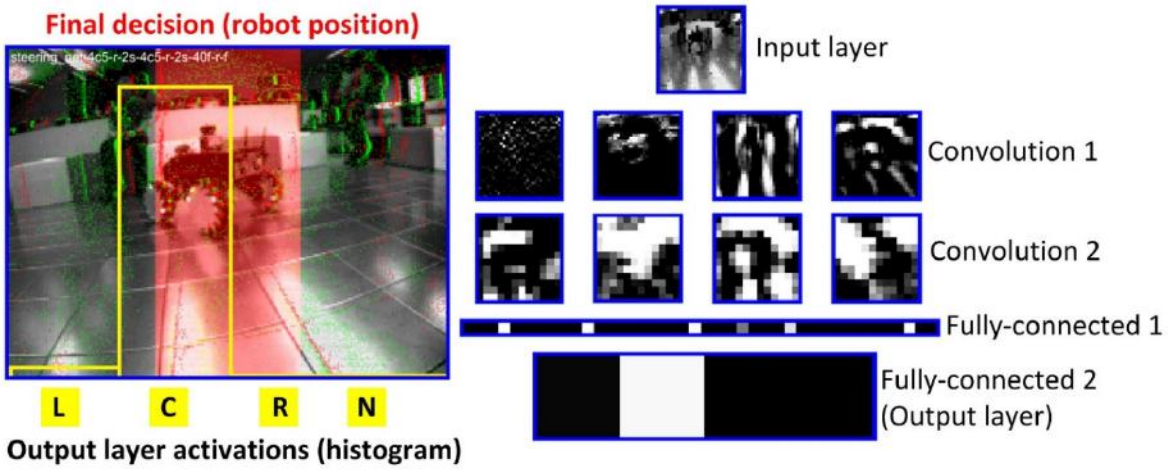


*Fig. 18 JAER visual output of the jAER CNN's processing (4C5-R-2S-4C5-R-2S-40F-R-4F). On the right-hand-side: the activations maps of input, convolution and fully-connected layers respectively. On the left-hand-side: final network decision: the four analog outputs of the four-neuron of the second fully-connected output layer are represented as an histogram of four bars in yellow. The red rectangle shows the final result of the softmax operation highlighting the section of the field of view where the prey is in. In case of non-visibility, no red rectangle is displayed. See video [141].*

## VI. Behavior and Decision filtering

The behavior of the predator robot allows several heuristic filtering options to reduce noise in the CNN decision output. Low-pass filtering is possible and an advantage of real-time, since frames are consecutive and strongly correlated.

The low pass filtering maintains an analog LCRN state for each possible LCRN CNN decision output. The output (winner) of the low pass is the LCRN state which has the largest value. These LCRN states

are bounded from 0 to 1. A parameter $\alpha$ specifies a step size. For each decision, the corresponding state is increased by $\alpha$ and all other states are decreased by $\alpha$. For our trial runs when low-pass filtering was used, we set $\alpha$=0.25. For example, if the current winner was L with state value 1 and state C was at 0.25, it would require 2 consecutive C decisions for C to be the new winner, because L would follow the sequence 0.75, 0.5, and C would follow the sequence 0.5, 0.75. Thus, after 2 decisions, C would become the new winner.

When the ROS controller receives L or R decisions it unconditionally turns left or right respectively at an angular velocity of $\pi/3$ rad/s and with the maximum allowed linear velocity (chase mode). The maximum linear velocity suitable for the arena used in these experiments is about 1.5 m/s. Trials at 2 m/s showed that safety distances in the presence of an obstacle need to be increased to avoid crashes, effectively reducing the useful area of the arena by more than a meter per side. At higher speeds it is also difficult for the teleoperated prey to evade the predator. Finally, the linear speed of the predator robot is also regulated by an underlying model of potential fields, an obstacle-avoidance algorithm for path planning. According to this algorithm a vector field is established over the area visible by the laser scanner and obstacles have a repulsive force dependent on distance that reduce the linear velocity of the robot. The aim of the robot behavior is forcing the prey to be in its center C (which corresponds to the central 27° of the FOV of the predator with the chosen lens) and accelerating towards it until the minimum safety distance is detected by the laser scanner in the center 40° FOV. Upon initialization, if the predator does not see the prey, or if it loses it for more than 5 seconds, the predator goes into search behavior (wander mode) and moves around randomly. If instead the prey becomes non-visible after being on its left or right it spins in the direction it last saw the prey (either left or right respectively) at an angular speed of 1.5 rad/s and zero linear speed (rotate mode). This behavior motivates the first constraint to the CNN output: CNN decisions that indicate that the prey is again visible on the opposite side of the field of view from which it disappeared can be discarded. Two more logical constraints can be applied. The first one is that the prey cannot switch instantly from center to non-visible. The final constraint is that the prey cannot pass from left to right without passing the center. This logical information can be used as post-processing of the CNN's output to increase its accuracy, at least in the autonomous scenario (some the data in the training set were generated by teleoperation of the predator).

The robot proceeds forward to the detected prey until the laser scanner detects imminent collision and stops. If the robot dashed forwards while detecting the prey in its center and a collision was detected, the prey is considered captured (prey caught mode). After this, the predator robot spins in the correct direction to center the prey and the prey moves away, controlled by the user or by the automatic navigation protocol. The chase starts again after a few seconds. All information about the state of the robot behavior (chase, wander, rotate and prey caught modes) is sent back to jAER through UDP, for display and recording.

## VII. Closed Loop Results

Once a CNN network with acceptable accuracy was trained, the robot control loop was closed by letting the predator robot run following the computed steering commands. While the prey was manually driven by an operator at the University of Ulster, the predator was being driven by decision

outputs. Initially we controlled and monitored the jAER CNN interface live from Zürich through a TeamViewer connection. The last 8 trial runs were conducted on site.

A video of one chase sequence (trial run #8) is available at [123], and a video of the CNN activity during part of this sequence is available at [141]. Video [123] shows the robot arena in which the predator drives at 1.5 m/s when moving forwards. Low-pass filtering with $\alpha$=0.4 and heuristic decision filtering, as described in the previous section, were used to smooth the decision output. The synchronized jAER running the CNN with both APS and DVS frames is also shown. The ground truth prey locations were labeled and online accuracy statistics are shown in the video. These statistics are based on the unfiltered raw CNN decision output, before low pass and heuristic filtering. Video [141] shows a view like Fig. 18 of the network activity during part of the same run. The topographic arrangement and responses of the convolutional layers becomes more obvious in this video.

For each of the eight trial runs, the ground truth was labelled offline. Speed was varied from 0.5 to 2 m/s and decision constraints were not applied in two trial runs. In three trial runs the lighting conditions were also altered (certain lights were turned off) to check for robustness, although no difference was noticed due to the auto exposure control of the APS frame capture of the DAVIS and the automatic local gain control of the DVS pixels. The decisions of the network were recorded and timestamped, to estimate the accuracy, which is shown in Fig. 19 for the first seven trial runs. Accuracy is plotted against the number of pixels $p$ in the boundary overlap, that is to say, the number of pixels that constitute the margin within which the decision is can still be considered correct if it is either one of the two neighboring LCRN regions forming the particular overlap and the ground truth falls within this same margin. If, for example, the ground truth labeled location is outside the C region but to the right or left within a distance *p* of the edge, then a C decision is still considered correct. And similarly, if the ground truth location is within p of the outer edges (indicating that the robot is only partially visible), then an N decision is still considered correct. Any L, C, or R decision when the ground truth label is N is a false positive and is always considered incorrect. These criteria are illustrated in Fig. 20.

It can be observed that increasing the overlap regions where the target location is ambiguous increases the accuracy. The various accuracies start from different levels since the number of ambiguous frames changes in every run. If the prey robot moves slower, then it is more likely to be mostly covering more than one third of the field of view generating ambiguous frames. Or it could be around the central region and oscillating in position around its boundaries. These cases quantitatively lead to a higher error rate as compared to the hand-labelled ground truth but qualitatively they are irrelevant as the robot will barely move if the decisions oscillate around a boundary. From Fig. 19, it can be seen that the use of constraints improves accuracy by about 6-7%. Interestingly, it can also be noted that when the maximum speed of the robot is increased, the accuracy is also increased. This is probably due to the fact that since the predator is faster, the prey has less time to escape. Therefore, the prey cannot move out from the predator's center C into the ambiguous steering regions that decrease accuracy. The achieved accuracies are comparable to [135], although the context of the robot arena and the application (the chase scenario) are probably of lower complexity than natural forest trails.

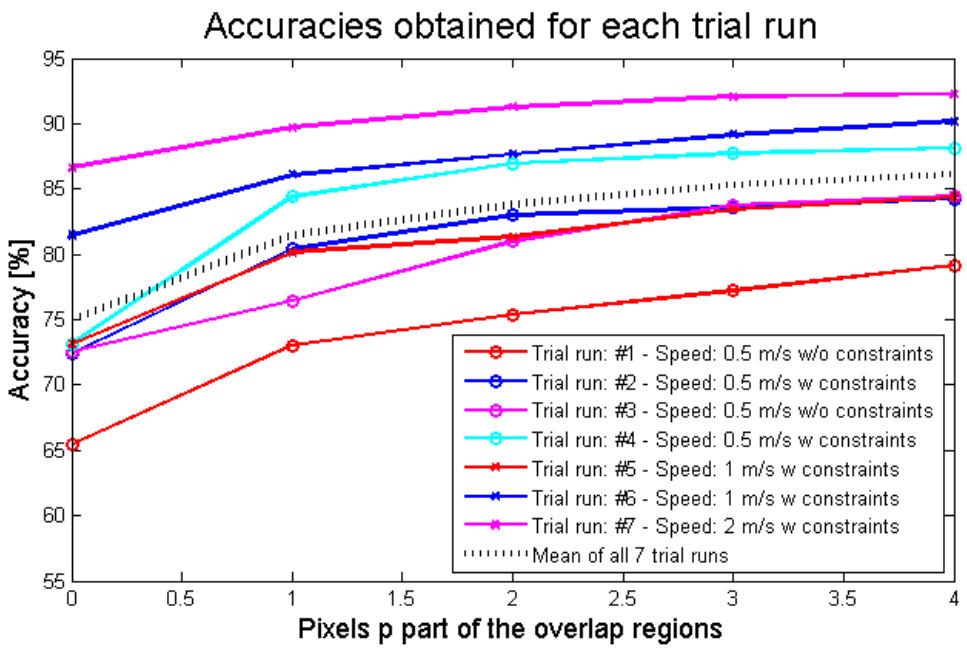


*Fig. 19 Accuracies of the seven trial runs versus number of pixels p part of the overlap regions of the 36 x 36 image.*

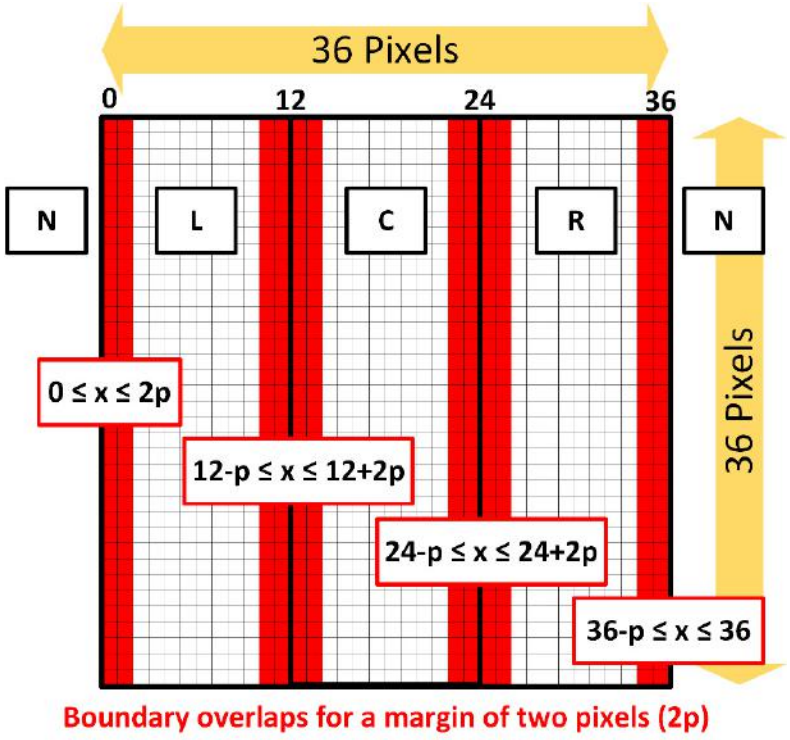


*Fig. 20 Example 36 x 36 frame and overlap regions (shown in red) of width multiple of p, within which the decision is still considered correct if the ground truth labeled target location is within either one of the two neighboring LCRN regions forming the particular overlap (and the ground truth labelling falls within it).*

The DVS frame error rate was slightly higher than that of the APS frames. For example, in trial run 8 (not included in Fig. 19), the overall raw error rate before decision filtering was about 13%; for APS

the error rate was about 9% and for DVS the error rate was about 14%. If we decreased the DVS frame to only 1'000 events, then the DVS error rate rose to 21% but the DVS frame rate rose to about 500 Hz. Thus there is a tradeoff between DVS frame size (in events) and decision accuracy. This tradeoff is to be expected, because integrating fewer DVS events results in a higher sample rate but a more quantized image, also with statistics on which the CNN was not trained.

## VIII. Conclusion

This paper proposed a method for combining the well-developed frame-based field of convolutional neural architectures with the data-driven processing of neuromorphic engineering using a simple approach aimed to be maximally compatible with existing training tools. With a processing rate proportional to the scene activity, encoded in the number of events, it is possible to reduce the amount of computing power whenever it is not needed. Although using DVS frames throws away the precise temporal information contained in the events, it still provides the advantages of the local temporal contrast response of the DVS pixels, which extracts features robustly in wide dynamic range lighting conditions. These conditions are a problem for conventional APS frames, e.g. as observed in forest scenes in [135].

In contrast to the training based on a simulated road in [134], learning user steering commands in [133], and learning trail versus non-trail using three cameras mounted on a human's head while they walk along trails in [135], our training was based on hand labeled prey target locations. This was feasible because there was only one target and we had only to label its 2D location. This labeling could be done almost in real time during playback, so as a fraction of the entire training process it represented only a tiny fraction of the effort.

The developed chase system ran robustly even with accuracies of around 80%. Our main finding and biggest surprise was the small size of CNN required for solving this problem. More surprisingly, it reliably detects the absence of the prey with about the same accuracy as it detects the presence and location of the prey. That means that this tiny CNN must detect the absence of a conjunction of features characterizing the prey in a rather complex (but static) background scene, which is a much more difficult task. The selected runtime network has only about 10'000 parameters. If we define each necessary multiply or add as one operation, then the forward pass requires only about 350'000 operations. (The current jAER CNN implementation has a much higher operation count because of matrix indexing computations.) This low operation count puts the computational cost well-within the range that could be serviced by small embedded application processors in a more optimized implementation. Although we did not explore it in our experiments, the computational cost could have been substantially reduced by adaptively controlling the DAVIS APS frame capture. For example, APS frames might be triggered only when the DVS event rate is low, or when the DVS output layer produces an ambiguous analog decision.

In future work, the size of the prey robot could also be taken into account to determine the distance from the predator robot. The only reason this was not done in the current project was lack of time and initial training data. The apparent height of the prey robot in the recording could also be used to infer the distance of the prey, however, this was not possible with the available recordings. This height constancy is due to the camera mounting, which points the recorded scene's vanishing point to the

horizon line, so that the position of the prey does not vary significantly in height with prey distance. The more general problem of identifying the walls of the arena and the robot's direction and distance from them could also be studied. These specific problems for the robot arena could be considered as prototypes for vehicle and path detection in autonomous driving (i.e. on roadways or factories), but with the advantage of smaller datasets and networks that are much easier to study.

The runtime jAER CNN implementation and Caffe conversion script are open-source. The database of recordings, trial runs and Caffe datasets are available on request.

## Acknowledgments

This research is supported by the European Commission project VISUALISE (FP7-ICT-600954), SeeBetter (FP7-ICT-270324) and Samsung. We would like to thank the Sensors group at INI Zürich, Luca Longinotti from iniLabs GmbH, and the Intelligent Systems Research Centre of the University of Ulster.

## 3.3. Conclusion

The simple 4-class steering network introduced in this work, managed surprisingly well in the task of giving steering directions to predator robot. The work showed that combined a DVS sensor with variable frame-rate output can be compatible with the state-of-the-art technology of CNNs, while still maintaining the efficient aspects of neuromorphic sensors. One of the next steps will be of course to run the CNN on a smaller less power-hungry platform, in order to prolong battery life.

The main limitation in training the CNN was the accuracy of the dataset, which should also be improved in the future by data re-labelling. However, it is possible that the large size of the dataset makes up for its imprecision and ambiguity.

The interesting aspect of the work was that the problem and network size were simplified to the minimum computation possible, still obtaining significant results. Current work, however, aims at giving an analog angle and distance of the relative position of the prey robot to ensure a smoother guidance of the robot, which appeared jerky in some trial runs. Distance can be extracted from the size of the prey robot in the field of view of the predator as this varies across the recordings (unlike vertical position).

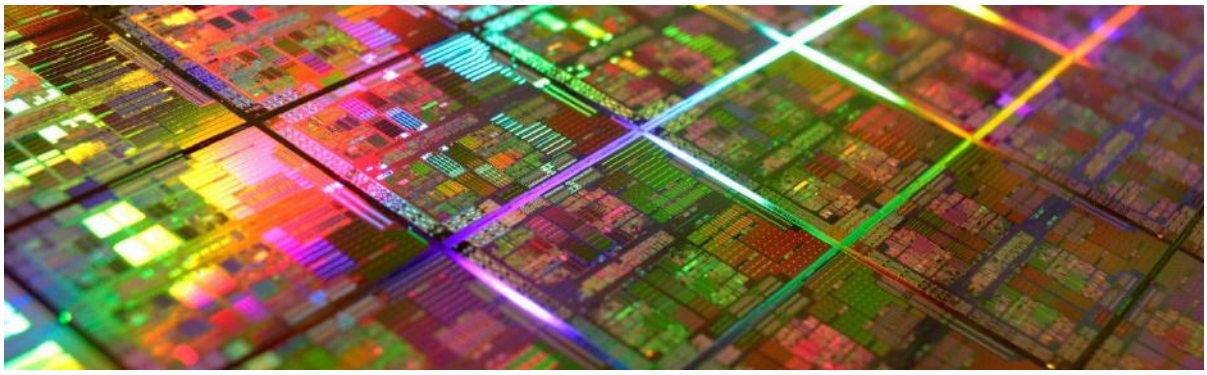

Image adapted from [142]

# Chapter 4

## 4. Characterization and applications of a high-gain Dynamic and Active pixel Vision Sensor

Taking advantage of the SEEBETTER project funding for a full wafer run with CMOS Image Sensor (CIS) 180 nm technology [143], a new version of the Dynamic and Active pixel Vision Sensor (DAVIS) was designed. The aim of this silicon retina is to improve upon the previous generation by adding increased contrast sensitivity capabilities to the pixel in order to improve the amount of details perceived, which could contain precious information. This feature can make the sensor suitable for further applications where it is necessary to either detect small changes in light intensity, such as neuroscientific calcium imaging, or capture many details of a scene even with lower optical array resolutions. The sensor was manufactured and characterized following the rigorous methods introduced in this chapter. The overall results show better contrast sensitivity with good SNR indeed, and highlight the problem, and reason, for APS to DVS coupling which does not enable their simultaneous use. All measurements performed resulting in the plots of this section are recorded and available on the iniLabs Network-Attached Storage (NAS) server. Bias settings are stored as well as the scripts used to generate the plots. Initial results for an application in neural imaging are also

discussed. The work was only possible thanks to the entire Sensor groups of INI and in particular to Dr. Sim Bamford (help with layout), Vicente Villanueva (PCB design), Luca Longinotti (software support), Dr. Federico Corradi and Chenghan Li (help in characterization). Also, most importantly, Prof. Dr. Bernabè Linares-Barranco and Prof. Dr. Tobi Delbruck came up with the idea of this sensor design.

In this chapter also, objective characterization techniques for Dynamic and Active pixel Vision Sensor are presented as they are needed for the evaluation of the improved SDAVIS192 sensor. The aim is to apply measurement standards to the simple Active Pixel Sensor (APS) intensity readout and to set standards to evaluate the main features of Dynamic Vision Sensors. Only knowing quantitatively, the features of the sensor allows to estimate its fitness to the task, no matter which one. The idea of this work comes from the fact that most designers of DVS sensors evaluate their sensor performance according to different and sometimes, non-transparent and subjective criteria. This damages researchers whose numbers are not as good because obtained with different conditions and triggers credibility issues from researchers in the same field and in particular in state-of-the-art technology designers. The aim of this work, which sets conditions and measurement procedures, is therefore to conform the various designers to a common standard, to which they can contribute and improve upon. This work is a collaboration between Dr. Federico Corradi, Chenghan Li and myself and would have not been possible without it.

## 4.1. The Sensitive Dynamic and Active Pixel Vision sensor

The interest in bio-inspired architectures since the late eighties, has led to the development of neuromorphic systems. The manifesto of these systems consists in observing and following the brain's underlying mechanisms that allow fast and low-power computations. Following this approach, neuromorphic sensors have been developed, in particular, the Dynamic Vision Sensor (DVS) [21], which goes beyond the frame-based approach of all normal digital cameras. Each pixel is a single processing unit on its own capable of telling if there has been a positive or negative logarithmic temporal contrast. ON and OFF event coordinates, of where the positive or negative temporal change respectively took place, are thereby generated and communicated off chip through the Address Event Representation (AER) protocol. Through this preprocessing, only useful information is extracted from the scene and sent out for further computations. This allows a lower data-rate strictly dependent on the activity in the visual field, which in turn means a lower power consumption and higher speeds.

To ensure compatibility with normal frame-based cameras the Dynamic and Active Pixel Vision Sensor (DAVIS) was developed in [32], combining DVS and Active Pixel Sensor (APS) readout. A variant of this, consisting of a single DAVIS pixel surrounded by three state-of the art pinned photodiode APS pixels was then also designed in [144]. Other methods of combining light intensity readout with DVS functionality include the ATIS sensor of [40] and the CeleX of [38].

More recently, to satisfy the need for more sensitive image sensors the work of Barranco et al. [36] achieved a sensor with 0.9% contrast sensitivity, adding more gain to their own variant of the DVS pixel design. Inspired by this idea, to approach application areas such as fluorescent imaging microscopy and aerodynamics, the silicon retina SDAVIS192 was designed. This chip was developed using the 180 nm Towerjazz CIS process. In the Sensors' group of the Institute of Neuroinformatics, already a similar intent was pursued first by Prof. Dr. Delbruck [34] (with only a single test pixel

fabricated) and by Dr. Minhao Yang, although with a much more complex high-gain Asynchronous Delta Modulation DVS (ADMDVS) [145].

SDAVIS192 was developed and manufactured taking advantage of funding from the SEEBETTER project [143] for a shared full-wafer run on Towerjazz 180 nm CIS technology. The access to this technology allowed to have two versions of the same sensor: one with an RGBW Color Filter Array (CFA) and a monochrome one. All characterization results which will be shown in this chapter regard the monochrome SDAVIS192, unless otherwise specified.

The design, characterization results and applications of SDAVIS192 were submitted for publication to Transaction on Biomedical Circuits and Systems in March 2017 [4]. The journal paper, titled “A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications” is a collaboration between the Brain Research Institute, University of Zürich, Switzerland and the Sensor´s group at the Institute of Neuroinformatics, ETH Zürich and University of Zürich. The authors are: Diederik Paul Moeys, Federico Corradi, Chenghan Li, Simeon Bamford, Luca Longinotti, Fabian F. Voigt, Stewart Berry, Gemma Taverni, Fritjof Helmchen and Tobi Delbruck. Fabian F. Voigt, Stewart Berry, Gemma Taverni and Fritjof Helmchen were instrumental for the neural imaging application.

# *A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications*

Diederik Paul Moeys[1], Federico Corradi[2], Chenghan Li[1], Simeon Bamford[2], Luca Longinotti[2], Fabian F. Voigt[3], Stewart Berry[3], Gemma Taverni[1], Fritjof Helmchen[3], Tobi Delbruck[1]
[1]Institute of Neuroinformatics, University of Zürich and ETH Zürich, Switzerland,
[2]iniLabs GmbH, Zürich, Switzerland, [3]Brain Research Institute, University of Zürich, Switzerland

## Abstract

Applications requiring detection of small visual contrast require high sensitivity. Event cameras can provide higher dynamic range and reduce data rate and latency, but most existing event cameras have limited sensitivity. This paper presents the results of a 180 nm Towerjazz CIS process vision sensor called SDAVIS192. It outputs temporal contrast Dynamic Vision Sensor (DVS) events and conventional Active Pixel Sensor (APS) frames. The SDAVIS192 improves on previous DAVIS sensors with higher sensitivity for temporal contrast. The temporal contrast thresholds can be set down to 1% for negative changes in logarithmic intensity (OFF events) and down to 3.5% for positive changes (ON events). The achievement is possible through the adoption of an in-pixel preamplification stage. This preamplifier reduces the effective intrascene Dynamic Range (DR) of the sensor (70 dB for OFF and 50 dB for ON), but an automated operating region control allows up to at least 110 dB DR for OFF events. A second contribution of this paper is the development of characterization methodology for measuring DVS event detection thresholds by incorporating a measure of Signal to Noise Ratio (SNR). At average SNR of 30 dB, the DVS temporal contrast threshold Fixed Pattern Noise (FPN) is measured to be 0.3%-0.8% temporal contrast. Results comparing monochrome and RGBW color filter array DVS events are presented. The higher sensitivity of SDAVIS192 make this sensor potentially useful for calcium imaging, as shown in a recording from cultured neurons expressing calcium sensitive green fluorescent protein GCaMP6f.

# I. Introduction

The interest in bio-inspired architectures has led to the development of event-based neuromorphic systems to achieve faster and lower power sensing and computing by using the brain's sparse, asynchronous digital communication architecture [146][147][12]. Neuromorphic sensors have been developed, in particular, the Dynamic Vision Sensor (DVS) [21]. Each DVS pixel reports if there has been a positive or negative brightness (log intensity) change since the last event. ON and OFF event coordinates are thus asynchronously generated and communicated off chip through the Address Event Representation (AER) protocol [12][22]. This sparse information about scene brightness changes allows lower data-rate and latency, which in turn means lower power consumption and higher speeds.

To enable compatibility with conventional frame-based cameras the Dynamic and Active Pixel Vision Sensor (DAVIS) was developed in [32], combining DVS and Active Pixel Sensor (APS) readout. A color DAVIS, consisting of single DAVIS pixels surrounded by three 5-transistor global-shutter pinned photodiode APS pixels was also built in [144]. Another method of combining light intensity readout with DVS is the ATIS [40]; it offers high Dynamic Range (DR) intensity but large pixel size.

To satisfy the need for more sensitive dynamic vision sensors to serve application areas such as bolometry, fluorescence microscopy and fluid dynamics, pixel variants were designed to achieve higher contrast sensitivity by increasing photoreceptor front-end gain. These areas require sensitivity to contrasts below 10%. Microbolometers require sensitivities of around 1% to be sensitive to 1°K temperature changes, because the temperature coefficients of the micromachined devices are roughly around 1.5%/°K [148].

Of particular interest within fluorescence microscopy is the recording of functional signals in living cells and tissues, for example by imaging calcium or voltage changes in neurons using fluorescent indicators. The newest genetically encoded calcium indicators have changes of fluorescence over the baseline ΔF/F of 20% for a single action potential in cultured neurons and rise times of 50-150 ms [149].

Sensing low contrast is difficult because noise in the signal can obscure the small relative variation of the signal and transistor mismatch makes it difficult to set low thresholds. The fundamental shot noise variance in a collected signal of $N$ photoelectrons is $\sigma^2 = N$, and reliably sensing a fractional deviation $C = \Delta N/N$ (single pixel contrast) requires a signal that is a multiple of the noise. If we take $m$ standard deviations ($m\sigma$) as our required signal to noise ratio (SNR), then we can compute the required $N$ [150]:

$$\mathrm{C} = \frac{\Delta N}{N} > m\sigma = \frac{m}{\sqrt{N}}$$
$$N > \frac{m^2}{C^2} \qquad (4.1.1)$$

E.g. if C=0.01 and $m$=3, then $N$=9x104. Thus a 3-sigma detection of a 1% contrast requires collecting 90k $e^-$, which exceeds the typical 10k e- Full Well Capacity (FWC) of conventional CMOS image sensors by a factor of ten [9].

In the previously mentioned application areas, the collected light is often the result of a secondary process such as fluorescence, where the primary exciting light is blocked by a filter, and the emission is a small fraction of the excitation, typically between $10^{-5}$-$10^{-6}$ [151]. The level of excitation illumination is limited by the small molecular cross sections for absorption [152] and by the low concentrations of fluorophores commonly used. The above calculation also ignores other noise sources such as 1/f noise in the readout. This discussion assumes single pixel reliable detection of given contrast, but area can trade off for contrast, as explained in [150]. Detecting a smaller contrast

trades off linearly with increasing the linear dimension of the feature to be detected.

Past DVS developments aimed towards higher sensitivity include a bolometer DVS [153] that used two switched capacitor amplifiers in series to achieve gain with modest capacitor ratios; the color-change DVS in [34] that used a sample and hold with capacitive amplification in the front-end stage; the sensitive DVS in [36] that used a stacked diode voltage amplifier; and the delta-modulator DVS in [145] that used a programmable-gain operational amplifier. We adopted the practical approach pioneered in [36]. Although it sacrifices intrascene DR, it leads to a simple and compact design. However, since the circuit in [36] consumes the photocurrent, it is incompatible with the DAVIS sensor since the photocurrent is also needed for integration by the APS circuit. To overcome this problem, the SDAVIS192 was designed and fabricated. SDAVIS192 simplifies the gain stage of [36] and moves the operating region control outside the IC, where it can be controlled more flexibly by a designed control policy, rather than a fixed analog feedback loop.

One of the main additional contributions of this paper is to improve measurement of neuromorphic vision sensor specifications towards establishment of standards. Specifically, this paper reports our methodology for measuring the temporal contrast threshold and suggests a measure of signal to noise ratio (SNR) for the temporal contrast. We use this method to measure the temporal contrast threshold of SDAVIS192. The rest of this paper is organized as follows: Sec. II describes the design of SDAVIS192. Sec. III reviews characterization methods in the literature and presents setup and protocols used for the characterization of SDAVIS192 and the measurement results. Finally, Sec. V shows an application of SDAVIS192 in the field of neural imaging.

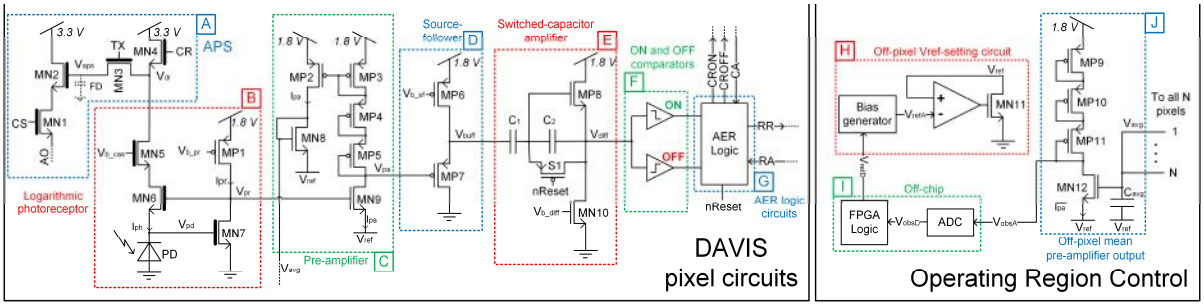


*Fig. 21 SDAVIS192 pixel, thick gate lines denote 3.3V transistors. A: APS, B: logarithmic photoreceptor, C: pre-amplifier stage, D: p-type source-follower, E: difference amplifier, F: ON and OFF comparators, G: in-pixel AER logic circuits, H: off-pixel circuit for setting $V_{ref}$ and bias generator, I: off-chip FPGA state machine and ADC, J: off-pixel mean pre-amplifier output circuit.*

## II. Design

This section describes the design of SDAVIS192, including the operating point control and IC layout.

### A. Concept and preamplifier design

The schematic of the SDAVIS192 pixel is shown in Fig. 21. The pixel improves on the original DAVIS [32], which in turns improves the original DVS pixel (to detect logarithmic brightness changes over time) of [21] by adding the well-known APS part to obtain signal intensity frames (box A in Fig. 21). The APS readout shares the same photodiode of the DVS circuit and it is isolated from it by the cascode transistor MN5. The APS readout is reset when MN4 is activated through the Column Reset (CR) signal. The reset charge, and alternatively, the illumination-dependent integrated charge, are integrated on the Floating Diffusion (FD) node Vaps when the transmission gate MN3 has been connected (integration starts) and disconnected (integration end), depending on the signal TX. Once the pixel's column is selected through the Column Select (CS) signal through the transistor MN1, the integrated

FD value is read through the source-follower pair composed by MN8 and a column-parallel source follower bias transistor (not shown).

The DVS circuitry of SDAVIS192 consists of the logarithmic photoreceptor with output $V_{pr}$ (box B) which compresses the input photo-signal, a preamplification stage with output $V_{pa}$ (box C) to boost the signal, a p-type source-follower buffer to avoid signal coupling with output $V_{buff}$ (box D), a switched capacitor amplifier with output $V_{diff}$ (box E), and two comparators to generate ON and OFF events (box F). Then AER communication circuits (box G) do a four-phase handshake by generating the pixels' Row Request (RR), Column Request OFF (CROFF) and Column Request ON (CRON) signals and receive the Row Acknowledge (RA) and Column Acknowledge (CA) signals.

The main improvement over previous DAVIS sensors is the introduction of a preamplification stage (box C), adapted from [32] to the DAVIS pixel circuit. It consists of a low-mismatch subthreshold voltage amplifier. While in [36] the amplifier is composed by a load of 4 series diode-connected nFET transistors driven by a pFET transistor, the proposed implementation uses only 3 diode-connected pFETs (transistors MP3-5 driven by the nFET MN9) to match the 1.8 V supply. The preamplifier of [36] was coupled to the photoreceptor through a source-driven active current-mirror [154] which is incompatible with the DAVIS APS readout. The SDAVIS192 unifies the preamplifier design of [36] with the front-end photoreceptor of [21][32] to achieve a DAVIS pixel with higher front-end gain. The preamplifier, whose input is $V_{pr}$ and whose output is $V_{pa}$, works on the principle of a common source amplifier with diode-connected load, with subthreshold transistors, as explained in [154]. A pFET active-load would have far too much and uncontrolled gain, together with an additional bias line. A very short pFET to get low enough gain would have a very process-dependent gain dependent on the Early voltage. The minimum size diode-connected transistors occupy less than 5% of the total pixel area. The gain $A_1$ of the preamplifier can be derived by equating the current in MP3-5 and MN9 [36]:

$$A_1 = -\frac{\kappa_n}{\kappa_p}\left(1 + \frac{1}{\kappa_p} + \frac{1}{{\kappa_p}^2}\right) \tag{4.1.2}$$

where $\kappa_p$ and $\kappa_n$ are the back gate coefficients of p and n FETs, respectively. For $\kappa_p$=0.7 and $\kappa_n$=0.8, $A_1$=-5.1. The relationship between the change in logarithmic intensity $\Delta\ln(I)$ and $V_{diff}$ is given by:

$$\Delta V_{diff} = -A_2 \Delta V_{buff} = -A_2 \kappa_p \Delta V_{pa} = -A_1 A_2 \kappa_p \Delta V_{pr} = \frac{-U_T A_1 A_2 \kappa_p}{\kappa_n} \Delta\ln(I) \tag{4.1.3}$$

where $A_2$ is the gain of the difference amplifier corresponding to the capacitive ratio $C_1/C_2$, and $U_T$ is the thermal voltage.

## B. Operating Region Control

The additional gain introduced by A2 reduces the DR of the pixel and therefore leads to possible clipping. In high illumination conditions, $V_{pa}$ saturates at $V_{ref}$ plus the saturation voltage of MN9 (4UT). For low illuminations, $V_{pa}$ will begin to saturate at the saturation voltage of MP3 plus the two drain-to-source voltages of MP4 and MP5, which are larger because of bulk effect. Fig. 22 shows the range of $V_{pa}$:

$$V_{ref} + 4U_T < V_{pa} < V_{dd} - 4U_T\left(1 + \frac{1}{\kappa_p} + \frac{1}{{\kappa_p}^2}\right) \tag{4.1.4}$$

For $V_{ref}$=0.2 V and $\kappa_p$=0.7, the possible swing is 1.053 V.

The operating points of the preamplifiers can be globally adjusted through the $V_{ref}$ common source.

$V_{ref}$ control is based on [36], which faced the same saturation problem with the preamplifier. In [36], the operating point was adapted by changing the gate voltage of the preamplifier rather than the source, since the DVS circuit is coupled to the photoreceptor by a source-driven active current mirror. The analog adaptation circuit continuously adapted the reference gate voltage depending on the off-pixel recreated average photocurrent. A digital control approach is amenable to software design, and was chosen to replace the analog feedback of [36] that is prone to oscillations (personal communication from Linares-Barranco and previous experience with analog feedback [17]). This concept is illustrated in Fig. 23. A replica-biasing circuit external to the pixel was therefore designed to set the source reference bias voltage $V_{ref}$ (hence the operating point) of the preamplifier of every pixel depending on the average scene illuminance.

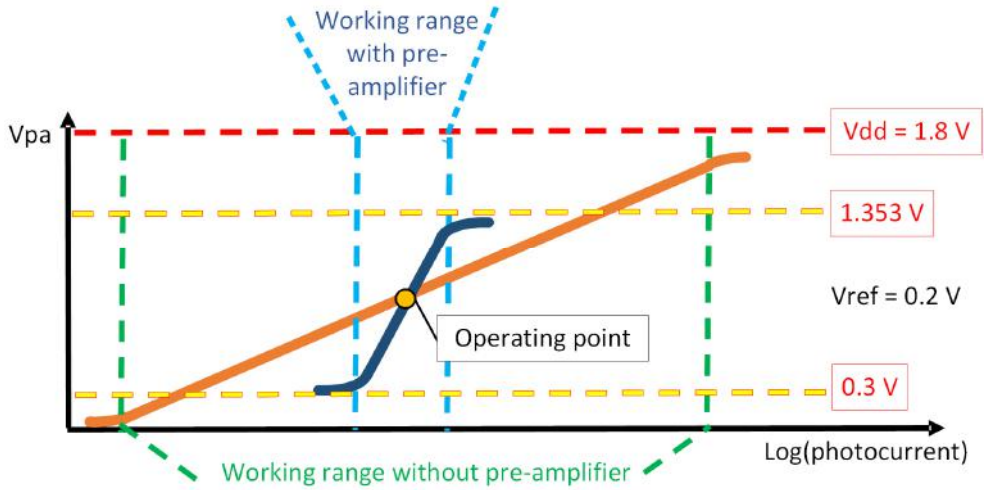


*Fig. 22 Dynamic range reduction with increasing gain (conceptual plot).*

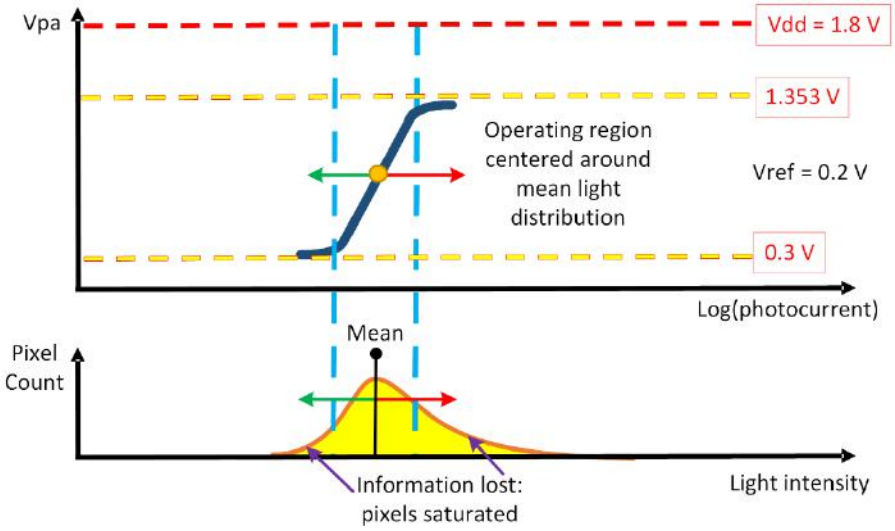


*Fig. 23 Operating region shift calculated from the mean intensity distribution.*

The aim of the adaptation replica biasing circuit is to check if the preamplifier's outputs are saturating at either one of their limits, and if they do so, to digitally control $V_{ref}$ to bring back the signals to their correct operating range. As seen in Fig. 21, the adaptation circuit (box H, I, J) makes a copy of the preamplifier's current, by mirroring it with the pFET MP2 and then averaging it onto the

diode-connected nFET MN8 in parallel with the shared 8.8 pF $C_{avg}$ pMOS capacitor (box J) to smooth this value ($V_{avg}$) with a time constant dependent on the average preamplifier current $\overline{I_{pa}}$ of the N pixels. $V_{avg}$ is then fed into the preamplifier replica (MP9-11 and MN12) to re-create the mean pixel condition. $V_{avg}$ is given by:

$$V_{avg} = \frac{U_T}{\kappa_n}\left(\frac{V_{ref}}{U_T} + \ln\left(\frac{\overline{I_{pa}}}{I_0}\right)\right) = \frac{U_T}{\kappa_n}\left(\frac{V_{ref}}{U_T} + \ln\left(\frac{\sum_{n:1}^{N} I_{pa_n}}{I_0 N}\right)\right) = \frac{1}{\kappa_n}\left(V_{ref} + U_T \ln\left(\left(\sum_{n:1}^{N} e^{\ln(I_n) - V_{ref}/U_T}\right)/N\right)\right) = \frac{U_T}{\kappa_n}\ln\left(\sum_{n:1}^{N} I_n/N\right) \quad (4.1.5)$$

where $I_0$ is the subthreshold off current, $V_{pa_n}$ is the n'th preamplifier output voltage and $I_n$ is the n'th photocurrent. $V_{avg}$ is then roughly proportional to the logarithm of the photocurrent. $V_{obsA}$ is then the average of the array's $V_{pa}$.

$V_{obsA}$ is connected to an off-chip 10-bit Analog-to-Digital Converter (ADC), whose output $V_{obsD}$ connects to the FPGA. The latter, according to a Look-Up Table, chooses the appropriate action to bring back the preamplifier in its correct operating range. This in practice means sending out the desired $V_{refD}$, which is then in turn converted into an analog value $V_{refA}$ by the on-chip coarse-fine bias generator of [155] (also used to provide all other biases). The ADC sample rate is adjustable up to 10 MHz. A row-parallel distributed low-dropout voltage regulator drives $V_{refA}$ to the array $V_{ref}$, by sinking as much current as needed through MN11 [155].

## C. Layout

The pixel size is 18.5 x 18.5 μm$^2$ and the transistor count is 47. Digital and analog signals were carefully separated at the pixel level in order to avoid the crosstalk between APS and DVS that also affected [32]. Where possible, metal layers connected to the analog supply rails isolate digital and analog signals. A weak ground connection turned out to be a mistake causing strong APS-DVS coupling due to ground fluctuations at every reset of the APS array, creating flashes of DVS events. This makes DVS and APS unusable simultaneously.

The photodiode has a 21.2% Fill Factor (FF), and is surrounded by a shallow active guard ring. A deep p-implant under the n-well prevents the n-well from stealing photo-generated carriers from the photodiode. A Buried Photodiode (BPD) is used. In a BPD, the junction is buried beneath the silicon surface by a shallow p-type implant that reduces SiO2 interface trap-states and therefore the dark current of the photodiode. An ohmic n-type contact penetrates the p-type implant. Salicidation was blocked on the source and drain of the APS structure to further reduce dark current. The reset transistor S1 of the DVS is also non-salicided. The redesign of the pixel allowed the resizing of C1, to increase the gain of the difference amplifier from 20 to 30 with respect to [32]. The overall added gain in the amplifier's passband is a factor of 6.

The array size of sensitive pixels is 188 x 192, which is less than the effective size of the array (208 x 192) because 20 columns of test pixels are present on the left-hand-side. Fig. 24 shows the layout of the SDAVIS192 chip on the Towerjazz 180 nm 6M1P technology. In the chip microphotograph, only the optical array is visible; in the RGBW split (see next paragraph) the rest of the chip is covered with

a red+blue optical filter, to shield parasitic photodiodes from visible light.

A wafer split of the dedicated mask run included an RGBW (Red, Green, Blue, White) Color Filter Array (CFA) to add DVS color sensitivity as reported in [156]. Also, an SDAVIS192 test pixel is connected to 3 analog multiplexers to observe most internal nodes. APS readout uses the same internal 10-bit column-parallel ADC of [144].

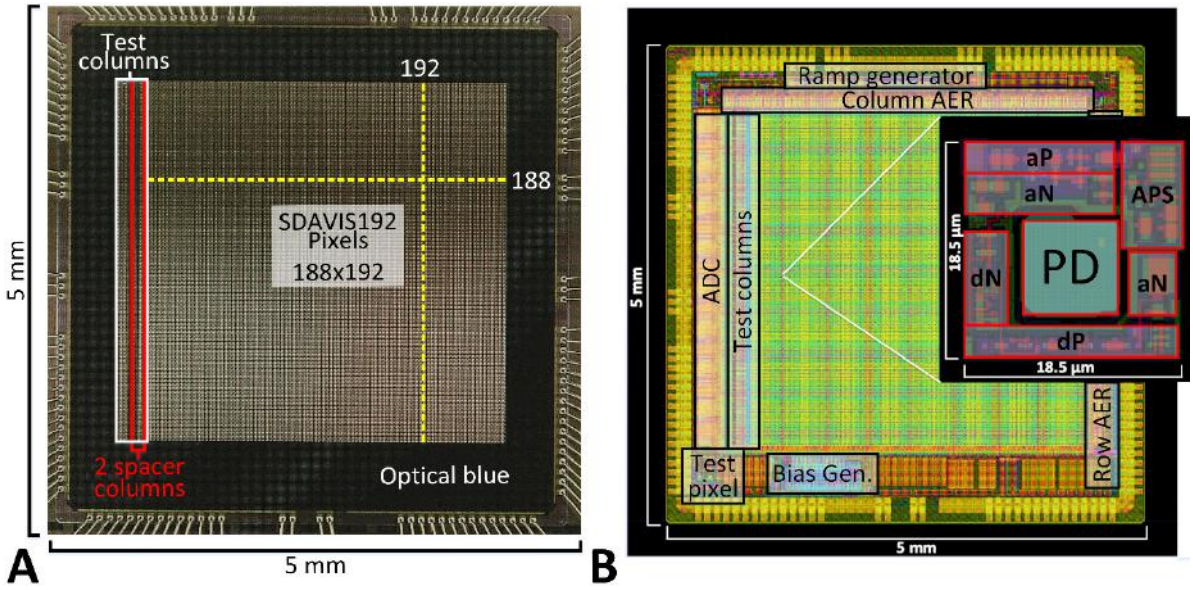


*Fig. 24 A: SDAVIS192 microphotograph. B: SDAVIS192 IC layouts (showing active, poly and M1). In single pixel: PD denotes the photodiode and aP/aN and dN/dP denote analog and digital n or pFETs respectively.*

## III. Characterization

Unless otherwise specified, all characterization results refer to the monochrome SDAVIS192 chip.

### A. Measurement setup

To perform the measurements in which a uniform light source sweep is required, the setup of Fig. 25 was used. This setup is a improvement on the set-ups used for testing [32],[145],[157]. An operational amplifier in feedback configuration was used to change the intensity of the light linearly with a function generator. This circuit forces the current through the high-brightness LED source to follow Ohm's law as the voltage on the 1 Ω resistor changes. DC and AC components can be applied to the light whose illuminance at the integrating sphere output port can linearly span 10 to 3500 lux. Neutral Density Filters (NDF) were used to attenuate this range by various factors. To stimulate only a small part of the pixel array (to not saturate the AER bandwidth), the integrating sphere is removed. An interposing surface with a hole of diameter 0.5 mm only lets a small fraction of the fiber's output reach the sensor, which is mounted with a lens (1/2" 4.5 mm f/1.4).

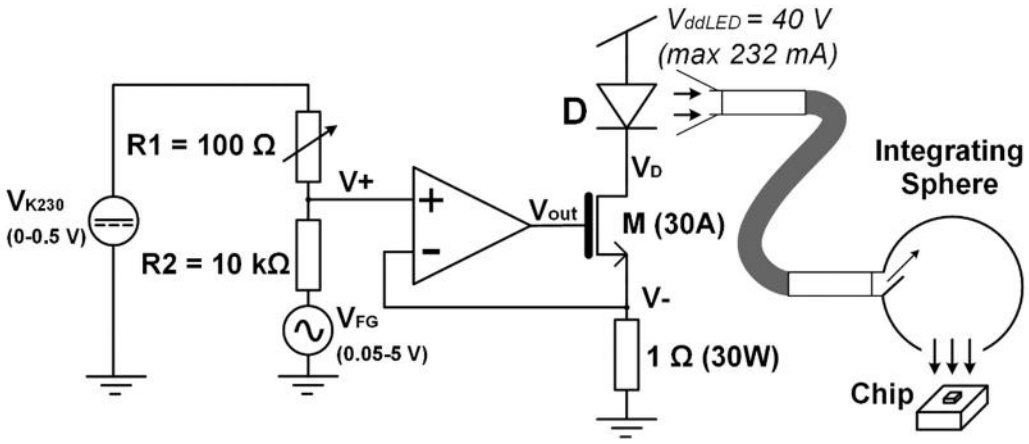


*Fig. 25 Testing setup for uniform light intensity sweep.*

## B. Test Pixel and Preamplifier Characterization

The transient response of the preamplifier was obtained from the test pixel that instruments pixel internal nodes. Fig. 26 illustrates the circuit response to a 5 Hz, 1.67 contrast, 1 klux light oscillation with the setup of Fig. 25. The photoreceptor output $V_{pr}$ responds with a 23 mV peak-to-peak oscillation. $V_{pa}$ was not observable, but $V_{buff}$ corresponds to $V_{pa}$ with a negative offset of roughly a threshold voltage. $V_{pa}$'s inferred value matches the observed behavior of $V_{obsA}$ (observed by exposing the entire array at different illuminations with various bias settings). For a different $V_{ref}$ bias, the transistor MN9 of Fig. 21 is in a different operating region and the transistor's $\kappa_n$ changes. The $V_{buff}$ to $V_{pr}$ voltage gain ranges from 0.5 to 3.15 depending on $V_{ref}$. In the region 200-400 mV the gain is about constant.

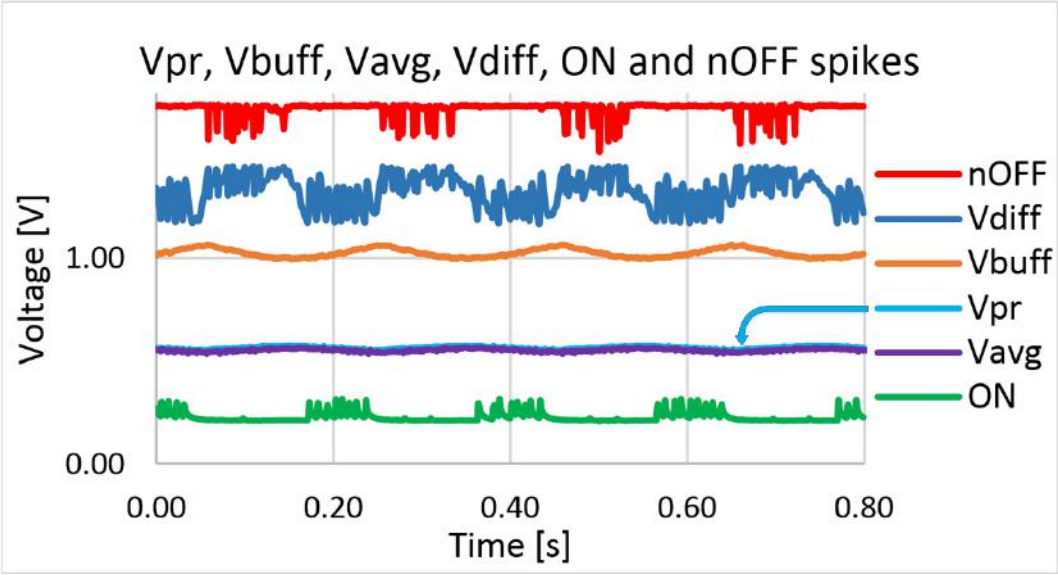


*Fig. 26 Test pixel oscilloscope traces.*

The gain is lower than the estimated preamplifier gain because of the source-follower's gain $\kappa_p$ lower than unity. A small ripple on every waveform is due to back-coupling of the ON and OFF events generated in the comparators' stage, though it is insufficient to generate further events. In Fig. 26, $V_{avg}$, which represents the average $V_{pr}$ value of the array, is identical to $V_{pr}$ (all pixels see the

same scene). $V_{diff}$, the output of the differential amplifier, amplifies changes in $V_{pa}$ by about 30 times. ON and active-low nOFF events appear in the corresponding rising and falling parts of the light cycle although their full amplitudes are not visible due to the low timescale resolution set on the oscilloscope.

## C. DVS Contrast Threshold, SNR and FPN

Contrast sensitivity characterization estimates the sensor's temporal contrast threshold, which is defined as the minimum percentage temporal contrast necessary to trigger either an ON or OFF DVS event. In the ideal case, for a sinusoidal input, the sensor produces exactly $s_{ON}$ events when the light intensity increases and exactly $s_{OFF}$ events when it decreases and no events should be generated for no temporal contrast. In reality, $n_{ON}$ noise events can appear during the negative part of the cycle, where only OFF events should be present and vice versa for $n_{OFF}$ events, as Fig. 27 shows. More generally, the expected cyclic pattern of events, which could ideally be used to reconstruct the signal, can show jitter and corruption. This can be due to oscillations and coupling in the pixel but also to a temporal noise activity. This activity can be due to the leak in the reset switch [21], or, especially in lower lighting, it can be due to temporal noise at the photodiode exceeding the small thresholds and triggering events.

In order to achieve a higher contrast sensitivity, the ON and OFF thresholds of the pixel must be lowered to the minimum possible. Due to mismatch in the pixels, some end up having their $V_{diff}$ passing threshold at reset making those pixels fire all the time (so-called 'hot pixels'). If contrast thresholds are computed without taking into account this type of fixed-pattern noise, it becomes hard to outperform the high contrast sensitivities of the state-of-the-art.

The single test pixel of [34] achieved a 50X reduction in contrast threshold from 15% by greatly increasing photoreceptor gain, but only by also running at very low bandwidth; additionally the continuous-time photoreceptor adaptation caused additional events that are difficult to interpret. Lichtsteiner et al. [21] quantified threshold variation by counting events produced by the passage of a stimulus of large 15:1 contrast. This method focused on measuring fixed pattern variation, while temporal variation was addressed by a separate experiment quantifying latency jitter. Posch et al. [153] quantified contrast threshold in terms of response probability for a given step change stimulus, but did not characterize variation between pixels whether temporal or fixed-pattern and thereby gave no information about the usability of the sensor as sensitivity is increased. Serrano- Lenero-Bardallo et al. [35] and Gottaredona and Linares-Barranco [36] followed the method of Lichtsteiner et al. [21] but used a TFT monitor stimulus with a large contrast stimulus of 400%. Unlike Lichtsteiner et al. [21], Gottaredona and Linares-Barranco [36] did not establish the asymptote at which contrast threshold mismatch becomes independent of contrast threshold; rather, they justified a 1.5% contrast threshold at 0.9% fixed-pattern noise by including an image which shows a large edge is visible to a human. They also stated that pixels could self-oscillate for very low thresholds (since, due to comparator mismatch, some thresholds will always be crossed [35]), but did not quantify the problem. Brandli et al. [32] and Yang et al. [145] used a modulated LED stimulus and integrating sphere to improve homogeneity of stimulus, (as proposed in [157]). Yang et al. [145] rightly pointed out that in order for the sensor to be performant, with low contrast threshold, the pixel response needs to be appropriately band limited and the illumination must be sufficiently high to reduce temporal noise. However, the 1% contrast threshold of the sensor was justified by placing an arbitrary cut-off coefficient of variation of contrast threshold of 35%.

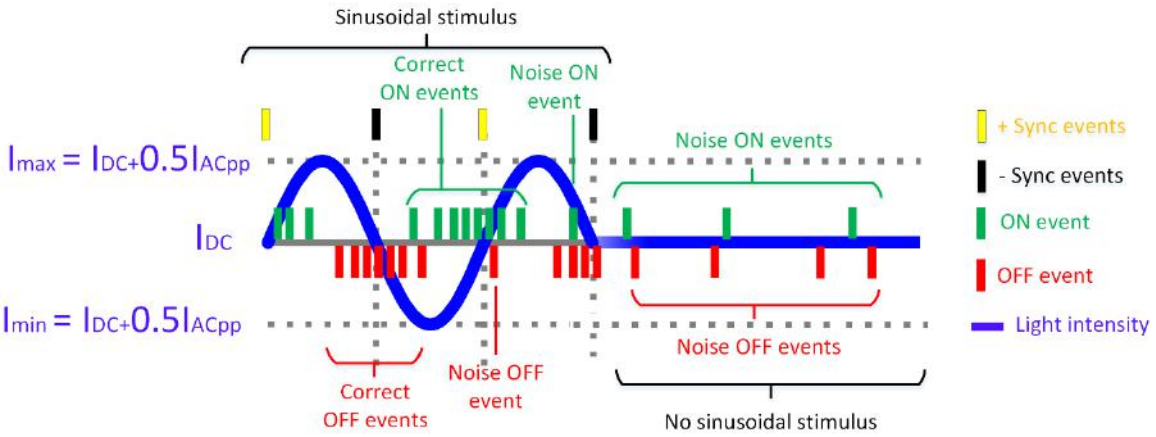


*Fig. 27 Typical noisy DVS activity for sinusoidal and no sinusoidal input.*

The method proposed here estimates noise activity and the quality of the signal through a Signal-to-Noise Ratio (SNR). By so doing we characterize the usability of the sensor at various contrast thresholds and in varying conditions and bias levels. Counting all events of the same type in the correct part of the cycle (ON in upswing and OFF in downswing) and dividing by the number of periods, gives the number of events generated per cycle by the sensor, $s_{ON}$ and $s_{OFF}$. In all measurements, it was assured that no coupling events would appear in the wrong part of the cycle (ON events in the downswing and OFF events in the upswing). Recording the sensor's output for no contrast present after the AC stimulation allows estimation of the noise activity present in the time corresponding to a cycle in the AC stimulation: $n_{ON}$ and $n_{OFF}$. These may not correspond to the actual noise during AC stimulation, since all pixels are being stimulated and signal overtakes noise (also thanks to the refractory period). However, the influence of noise is still present in the cycle and does affect sensitivity and must be taken into account. $n_{ON}$ and $n_{OFF}$ constitute the noise floor and without prior knowledge of the signal, they would not allow understanding whether recorded events constitute signal or noise. As observed from measurements, $n_{ON}$ and $n_{OFF}$ do not increase over time but rather depend only on the specific bias settings and illumination. For their measurement, we wait two seconds after the end of the AC stimulation, to let the pixels recover from possible refractory period effects. Subtracting the $n_{ON}$ and $n_{OFF}$ from the $s_{ON}$ and $s_{OFF}$, gives the actual signals $s_{ON} - n_{ON}$ and $s_{OFF} - n_{OFF}$. Knowing these and the contrast of the light intensity gives the sensitivity of the sensor.

If a sinewave of DC light intensity $I_{DC}$ is supplied with an AC peak to peak oscillation equal to $I_{ACpp}$, the contrast in the upward phase of the cycle can be defined as $TCON_{ON}$, which is the maximum value of light intensity $I_{max}$ over the minimum $I_{min}$, assuming a monotonic change:

$$TCON_{ON} = \frac{I_{max}}{I_{min}} = \left(\frac{I_{DC} + 0.5I_{ACpp}}{I_{DC} - 0.5I_{ACpp}}\right) \tag{4.1.6}$$

In the case of the contrast in the downward phase of the cycle ($TCON_{OFF}$), the ratio is reversed. To convert from a fractional increase $\varepsilon$ (for $\varepsilon \ll 1$) to a percentage contrast threshold [21] for $s_{ON} - n_{ON}$ or $s_{OFF} - n_{OFF}$ events recorded (noise subtraction), the logarithmic threshold of the DVS can be approximated as:

$$\theta_{ON}[\%] = \frac{\ln(TCON_{ON})}{s_{ON} - n_{ON}} = \frac{\ln(1+\varepsilon)}{s_{ON} - n_{ON}} \approx \frac{\varepsilon}{s_{ON} - n_{ON}} \tag{4.1.7}$$

$$\theta_{OFF}[\%] = \frac{\ln(TCON_{OFF})}{s_{OFF} - n_{OFF}} \quad (4.1.8)$$

This approximation is valid for $\varepsilon$<0.2 (as it is the case for SDAVIS192) for an error within 3%. $\theta_{ON}$ and $\theta_{OFF}$ are computed from the median event count per cycle minus the median noise ($s_{ON} - n_{ON}$ and $s_{OFF} - n_{OFF}$) of all pixels of the array. The median count is used to filter out outlier hot-pixels. This effect was observed in measurements when the mean contrast threshold would be 2% and the median 30%, highlighting the masking effect of hot pixels.

With the setup of Fig. 25, for a particular homogenous illumination, a 1.67 $TCON_{ON}$ contrast is applied with a 1 Hz sinusoid and events are recorded from the sensor with the APS turned off. The frequency is low enough to not saturate the output bus, and at the same time higher than the mean firing of the background leak activity (typically 0.1 Hz). Thresholds are set initially once to a good sensitivity (and the sensor produces correlated activity) and then bias, contrast and background illumination sweeps are performed. The half-cycles of the sinewave in which to count events are timestamped through sync events gathered by the FPGA, from the VFG function generator. At the end of 10 cycles, an extra 10 s are recorded with DC input count $n_{ON}$ and $n_{OFF}$. For every sensitivity level, a Signal-to-Noise Ratio (SNR) is computed as:

$$SNR_{ON}[dB] = 20\log_{10}\left(\frac{s_{ON} - n_{ON}}{n_{ON}}\right) \quad (4.1.9)$$

$$SNR_{OFF}[dB] = 20\log_{10}\left(\frac{s_{OFF} - n_{OFF}}{n_{OFF}}\right) \quad (4.1.10)$$

The SNR ratios were taken as signal minus noise over noise, because the minimum $s_{ON}$ or $s_{OFF}$ which can be recorded during AC stimulation is $n_{ON}$ or $n_{OFF}$ (just pure thermal or leakage noise). If SNR was taken as $s_{ON}/n_{ON}$ or $s_{OFF}/n_{OFF}$, then in the worst case scenario, the SNRs would just be 0 dB for $s_{ON} = n_{ON}$ or $s_{OFF} = n_{OFF}$. With our SNR method, SNRs fall below zero dB for $s_{ON} < 2n_{ON}$ or $s_{OFF} < 2n_{OFF}$. Finally, for $s_{ON} = n_{ON}$ or $s_{OFF} = n_{OFF}$, SNR is equal to minus infinity dB (no input-dependent signal).

To test how far the sensitivity of SDAVIS192 can be pushed while still maintaining reliability, ON and OFF threshold biases were swept with a sensor illumination of 1 klux in various combinations. Fig. 28A plots the resulting $SNR_{ON}$ and $SNR_{OFF}$ versus $\theta_{ON}$ and $\theta_{OFF}$. The lower the contrast threshold, the more the signal with respect to the temporal noise (and therefore the higher the SNR). If contrast threshold is lowered further, however, the pixels start to self-oscillate and the SNR abruptly drops to minus infinity dB ($s_{ON} = n_{ON}$ or $s_{OFF} = n_{OFF}$). Minimum contrast thresholds were established by the lowest contrast threshold achievable with SNR>0. These are at 1 klux for SDAVIS192 are $\theta_{OFF}$=0.95% with $SNR_{OFF}$=28.5 dB and $\theta_{ON}$=3.45% with $SNR_{ON}$=18 dB. Approaching the limit of contrast sensitivity, more OFF than ON events are generated and therefore OFF has a lower contrast threshold. This is the case because the two different bias current references of the comparators determine their different speeds, which limit the encoding favoring one type of event [145]. For prior DAVIS sensors it is the opposite, due to the SDAVIS192 inverting preamplifier. In the case of DAVIS240C [32], minimum contrast thresholds coincide with the ones of Fig. 29 at 1 lux: $\theta_{ON}$=27.6% with $SNR_{ON}$=19 dB, $\theta_{OFF}$=20.5% with infinite $SNR_{OFF}$ (no $n_{OFF}$).

Fixed Pattern Noise (FPN) can be estimated from the pixels' mismatch (spread) in firing rates, as done by Lichtsteiner et al. [21]. Plotting an event histogram, however, shows that the distribution of firing rates is not normal and that there are a certain amount of kurtosis and skew. Standard deviation (normally used to compute FPN) is a summary statistic of a more complex underlying distribution. The

Cumulative Distribution Function was used to provide lower and upper percentiles of number of events per cycle, corresponding to one standard deviation in a normal distribution (15.8%-84% percentiles divided by 2). These can then be used to obtain $\mathrm{FPN_{ON}}$ and $\mathrm{FPN_{OFF}}$ expressed as percentage contrast [21]. For the same set of biases of Fig. 28A, $\mathrm{FPN_{ON}}$ and $\mathrm{FPN_{OFF}}$ are plotted in Fig. 28B as a function of contrast threshold. The higher the contrast threshold, the higher the FPN: [21] attributed this trend to amplifier gain mismatch. For the sets of biases to achieve the lowest $\theta_{\mathrm{ON}}$ and $\theta_{\mathrm{OFF}}$, $\mathrm{FPN_{ON}}$=0.2-1% contrast and $\mathrm{FPN_{OFF}}$=0.8-0.9% contrast for SDAVIS192. In comparison, for DAVIS240C, $\mathrm{FPN_{ON}}$=7.5-15% contrast and $\mathrm{FPN_{OFF}}$=5.5-8.5% contrast.

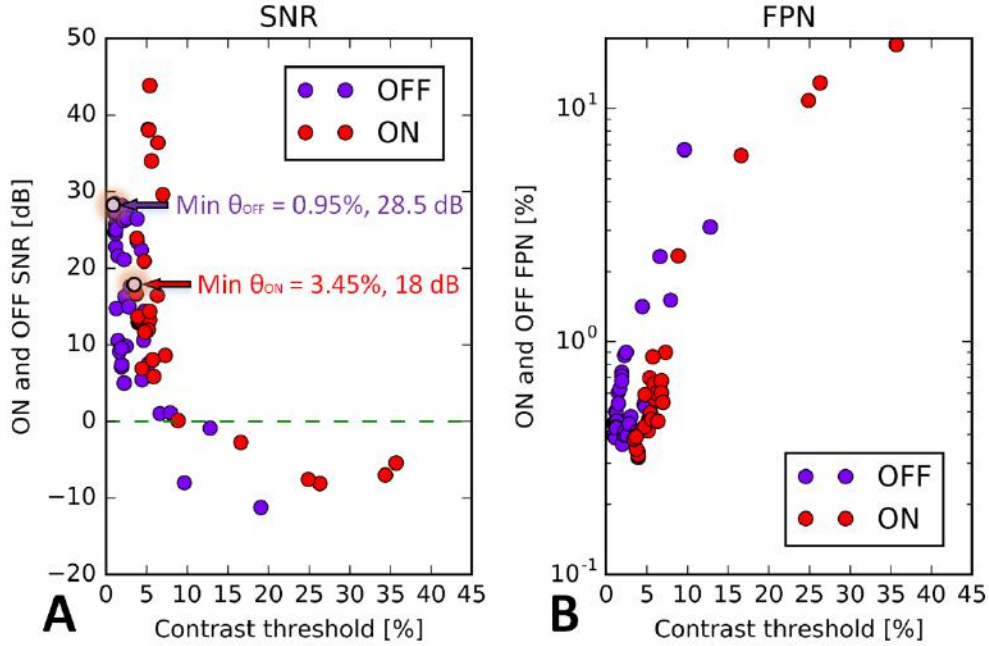


*Fig. 28 SDAVIS192 $SNR_{OFF}$ and $SNR_{ON}$ (A) and percentage $FPN_{OFF}$ and $FPN_{ON}$ (B) vs the respective contrast thresholds at 1 klux, with IR cut filter.*

As sensitivity increases, temporal noise becomes an important factor since the number of "hot-pixels" firing increases. Due to their low threshold, these pixels lock themselves in a state of high firing of one type. If a pixel is considered 'hot' if it fires more than 200 times per cycle, the number of ON hot-pixels for SDAVIS192 is almost zero, while OFF hot-pixels reaches 0.5% for $\theta_{\mathrm{OFF}}$<1.2%.

## D. DVS Dynamic Range

In [21], [32], [36], [35] and [145] DR was obtained by observing the sensor under extremely low and high light conditions by 'visually recognizing' a correlation in the events which would match the stimulus. However, this is not case of this work because, although the sensor still shows some correlation with the input signal (a 1 Hz flashing can be recognized even beyond the given DR of SDAVIS192), noise is dominant and SNR<0. If the noise is larger than the signal, then the sensor is difficult to use in practical applications. Prior work also made no distinction between ON and OFF events.

Here the illumination was varied in the range 0.01 to 2 klux. An IR-cut filter was used to ensure correct NDF attenuation. Fig. 29 illustrates contrast thresholds versus illuminance level measured at the sensor for a single set of OFF and ON biases, for both DAVIS240C and SDAVIS192, which work best at most illuminations. It therefore does not mean that the sensor cannot work outside the range

shown in Fig. 29, however, the sensor exhibits its larger DR in this range. In the case of DAVIS240C, for the selected set of biases giving SNR>0, the DR is 60 dB for ON and OFF. For SDAVIS192, the global operating range bias was changed only when necessary: the same set of biases, which would still give SNR>0 and contrast thresholds lower than 100% was kept whenever possible (this change is marked with the green vertical line occurring at 0.1 lux). The intra-scene DR of SDAVIS192 (for the same set of biases) is at least 50 dB for ON and at least 70 dB for OFF. "At least" means that the DR could be higher, but the uniform light intensity of the setup of Fig. 25 cannot reach 10 klux to verify this claim. The relatively small DR of SDAVIS192 comes from the fact that the sensor has high gain. The DR of DAVIS240C is however similar, although the smaller gain, because of the smaller $s_{ON}$ and $s_{OFF}$ with respect to $n_{ON}$ and $n_{OFF}$ does not guarantee SNR>0 on a larger illumination range. Thus SDAVIS192 has more points in the graph of Fig. 29 where SNR>0 and contrast threshold is less than 100%. The parabolic shape of Fig. 29 shows that the event count decreases as the sensor reaches either ends of its intra-scene DR. Thanks to the global $V_{ref}$ bias sensitive to mean illumination, the total operating range of the DVS sensor spans at least 110 dB for OFF for the same set of biases. For the same ON bias, no extension of the DR can be made and this remains at least 50 dB (similarly to the 54 and 60 dB of [35] and [36]).

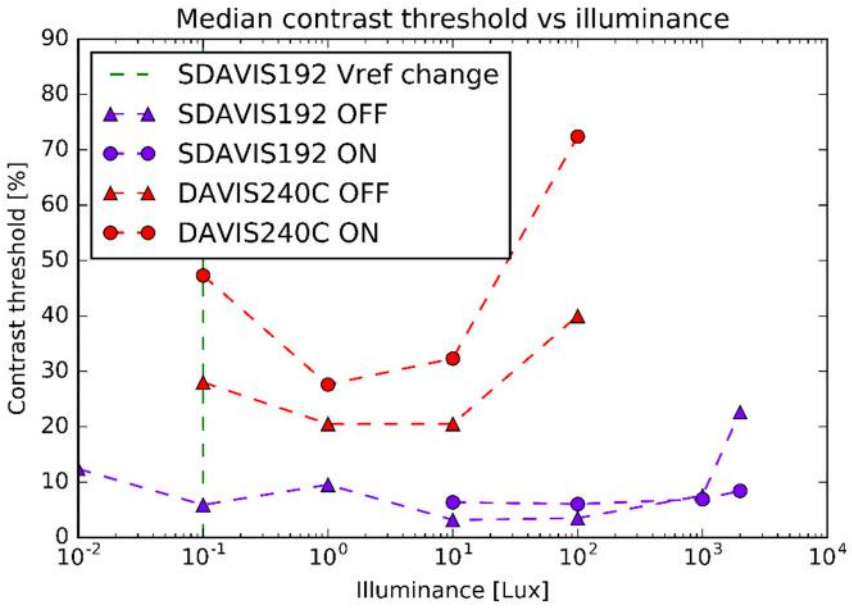


*Fig. 29 SDAVIS192 and DAVIS240C ON and OFF contrast thresholds vs illumination level. No point at a particular illumination means SNR<0.*

The ON pathway has a lower DR than the OFF pathway from two phenomena that take place at low and high intensities. At low light levels, the photodiode dark current decreases the signal contrast in the photocurrent. At the same time, junction leakage in the S1 reset transistor continually produces a fictitious OFF temporal contrast. The combination of these two effects means that at low intensity, the sensor favors making OFF events. The APS dark current (estimated at 16k $e^-$/s for SDAVIS192, in section III.F) reduces the photocurrent contrast. At 0.01 lux the estimated photon contrast with respect to the estimated dark current is less than 2%. However, across the entire array, pixels that are near threshold respond synchronously, resulting in a median contrast sensitivity of 10%. At high intensities, parasitic photocurrent in the S1 reset transistor source-drain junction generates an increasing number of OFF events, again producing a fictitious OFF temporal contrast. This parasitic photocurrent problem was already known from [21] and [32] and therefore particular care was taken

to shield the reset transistor with metal and an active guard ring. However, the shielding is not perfectly effective. SDAVIS192 functions at 3.5 klux, while DAVIS240C, stops working at 2 klux as parasitic photocurrents in the insufficiently shielded bias generator deadlock the chip for this set of biases. (DAVIS240C has no optical shield on peripheral circuits.)

## E. DVS Pixel Bandwidth

The Frequency Response (FR) of the single pixel was first obtained by Lichtsteiner et al. in [21] together with the second-order filter behavior of the pixel. The pixel FR can be measured using the same setup as for contrast threshold, but with only a few pixels being stimulated, and using a sinewave as the modulating signal. We use a lens to image a pinhole. Knowing the diameter of the pinhole and the circular field of view of the Tektronix J17 luminance meter, we can measure the pinhole luminance, which could be set to a maximum of 30 cd/m$^2$. Geometrical optics is then used to compute the sensor illuminance. By sweeping the 1.67 $\mathrm{TCON_{ON}}$ contrast sinewave frequency and measuring the number of events produced per half cycle, with noise subtracted, it is possible to obtain a pixel FR. Since high-bandwidth biases are not the purpose of SDAVIS192, but rather high-contrast sensitivity biases with good SNR, only these are presented. Fig. 30A illustrates ON and OFF FRs with a contrast of $\mathrm{TCON_{ON}}$=1.67. $\theta_{\mathrm{OFF}}$ and $\theta_{\mathrm{ON}}$ are both about 5%. With these biases, the 3dB cutoff is about 20Hz and the absolute cut-off is about 100 Hz. The preamplifier, weakly biased to achieve higher sensitivity, is the main reason for this FR that hardly changes with intensity. The roll-off is approximately that of a second-order lowpass filter and no ringing was observed.

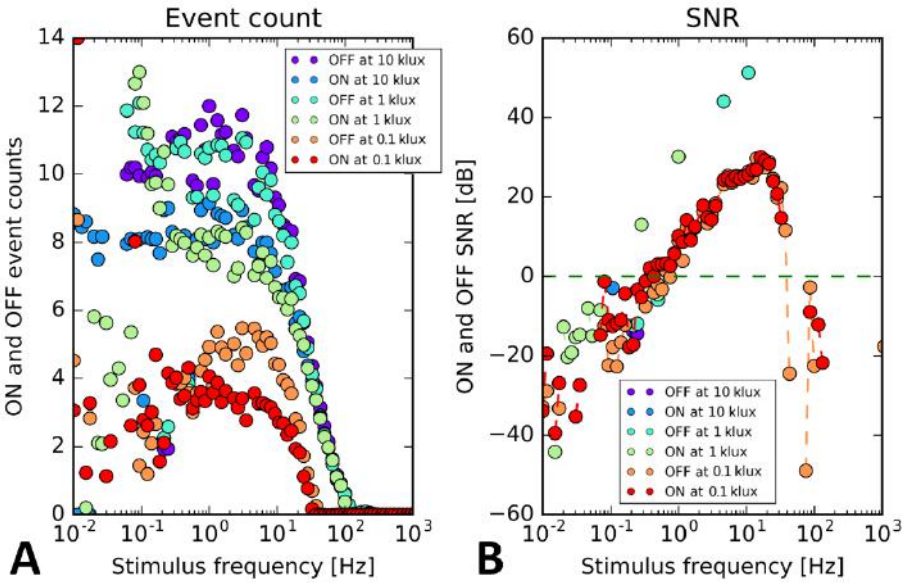


*Fig. 30 A: Pixel bandwidth of SDAVIS192 as a function of frequency (with contrast of 1.67) for bias settings giving good contrast sensitivity. 10 events per cycle corresponds to a 5% temporal contrast threshold. B: Corresponding ON and OFF SNRs (SNRs at 1 and 0.1 klux are not visible since infinite). Illuminance values are at the sensor focal plane.*

Noise in the measurement is considerable below 0.1 Hz since recordings of 10 cycles of activity can take minutes. During these long recordings any sudden source of noise can be can be falsely detected as signal. The corner frequency of the pixel is thus uncertain and can only be defined by looking at the specific SNR. In [21], no lower cut off was measured. Below the corner frequency, $s_{\mathrm{ON}} = n_{\mathrm{ON}}$ and $s_{\mathrm{OFF}} = n_{\mathrm{OFF}}$ . In this work, $s_{\mathrm{ON}} - n_{\mathrm{ON}}$ approaches zero around DC as background noise events take over and hide the signal. As illuminance decreases, dark current contrast reduction reduces the event

count, but the lower cut-off frequency remains very similar. As can be seen in Fig. 30B, SNRs plummet as soon as the event count decreases and the fixed background noise takes over. For these biases, noise is almost non-existent at 10 and 1 klux chip illuminations (very large SNRs). At 10 lux, noise becomes visible and SNRs appear in the range from -40 to 30 dB. In the work of [35], [36] and [145], unrealistic illuminations of up to 50 klux were used to estimate FR.

## F. APS Characterization and Dark Current

APS was characterized using the Photon Transfer Curve (PTC) measurement procedure of the EMVA1228 standards [158]. All results of APS characterization are reported in Table 2. The 3.39% APS FPN of SDAVIS192 is about three times the 0.97% of DAVIS240C because the latter uses an external ADC rather than the on-chip column parallel ADC of SDAVIS192. DAVIS240C, has an 18k $e^-$/s APS dark current (0.858 nA/cm2) at 25°C. The SDAVIS192 APS dark current should ideally be one third of such value for similar transistor and photodiode sizing, since it uses BPD instead of Surface Photodiode (SPD), but is only 21% smaller: 16k $e^-$/s (2.6 fA or 0.748 nA/cm2). This high leakage is close to vanilla n-well photodiode performance (1 nA/cm2). This values are also worse than the 4 fA measured by Lichtsteiner et al. [21] on the larger nwell photodiodes manufactured with a non-CIS process. The reasons for this problem might be large transistor junction leakage in the APS readout pathway. Read noise $\sigma_{\mathrm{read}}$ (sensitivity) are less than 1 Digital Number (DN, 1.26 mV for SDAVIS192 and 1.47 mV for DAVIS240C) and are therefore limited by ADC quantization: 61 $e^-$ and 57 $e^-$ respectively.

## G. QE Characterization

Measurements of absolute Quantum Efficiency were performed at the laboratory of Interuniversity Micro Electronics Center (IMEC), Belgium. The setup consisted of a Xenon-based white light source whose light passes through a monochromator, which, through a series of mirrors and shutters, only lets through a specific wavelength. PTC spectral measurements were performed at each wavelength. Comparing the photo-generated current to the current measured from a Newport 918 series reference photodiode with Newport 1936 power meter, the results shown in Fig. 30A were obtained for the monochrome sensors. The QEs of DAVIS240C and SDAVIS192 can be compared fairly since their pixels have the same size and similar FF. The FF determines the maximum QE for each sensor if no microlens is present (as for DAVIS240C). Both sensors are front illuminated with nearly the same metal stacking. As Fig. 30 shows, SDAVIS192 has an overall higher QE spectrum. The thick microlenses (of size 15x15 µm2 and thickness 3 µm with additional spacer), appear to be effective, since the peak QE of 24% is higher than the 21.2% FF. The diffuse incident light in the QE measurement probably underestimates the actual QE using a lens, since in that case the light would likely be better focused onto the photodiode. In comparison, the QE of the DAVIS240C (FF of 22%) peaks at 7% at 450 nm, with a magnitude 3.4X smaller than that of the SDAVIS192.

Part of the increase in QE is also due to the use of a BPD for SDAVIS192 with respect to a SPD in DAVIS240C. The BPD almost doubles the junction area and reduces the chance that photo-generated carriers are captured in trap-states at the interface with the SiO2. It can be seen that the deeper BPD is slightly more sensitive to longer wavelengths while the SPD is slightly more sensitive to shorter wavelengths that could be absorbed in the BPD surface cover implant. Both photodiodes have a sharp cut-off around 350 nm because UV light is reflected by the N-BK7 package cover glass. The repeatable complementary fluctuation at 850 nm is not explained. For longer wavelengths, the QE decreases due to the shallow photodiodes and small reverse bias voltage.

Overall, the thick microlens combined with BPD appears to approximately triple the QE for this diffuse incident illumination. The QE of a color split of the SDAVIS192, with a CFA on top of the pixels with RGBW pattern, was also characterized in [156] and the QE spectrum of the single RGBW channels is shown in Fig. 30B.

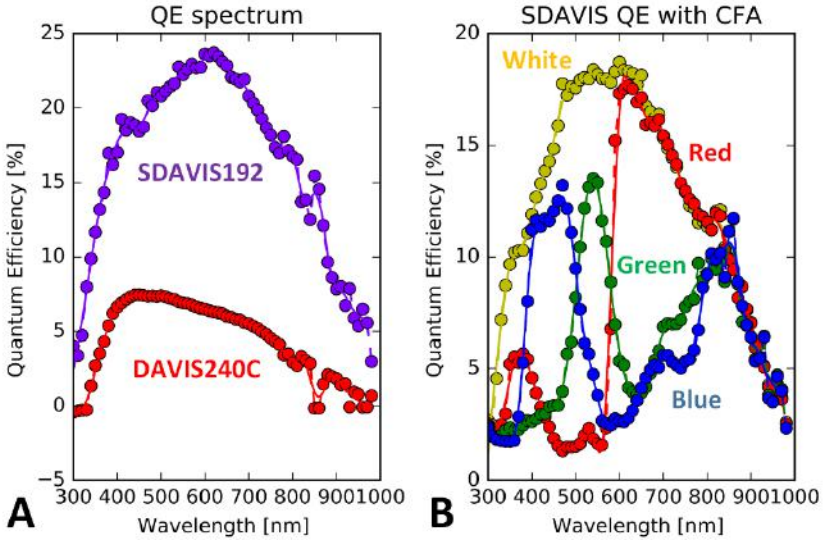


*Fig. 31 A: External Quantum Efficiency (QE) spectrum of Active Pixel Sensors of SDAVIS192 and DAVIS240C . B: Measured external QE spectrum of the RGBW channels of SDAVIS192. Results obtained at IMEC.*

## H. Sample Monochrome and RGBW Outputs

Fig. 32 compares 30 ms DVS time slices from SDAVIS192 with DAVIS240C (video available at [159]). Fig. 32A shows the setup for the moving hand and Edmund density step chart. The density step chart has 10 steps of 0.1 density. The contrast of each step is 100.1=1.26. The DVS gray scale is 10 events, i.e. from gray, 10 events ON or OFF makes the image full white or black. SDAVIS192 produces on average 10X more events revealing more gray scale and detail and better distinguishes the hand from its shadow.

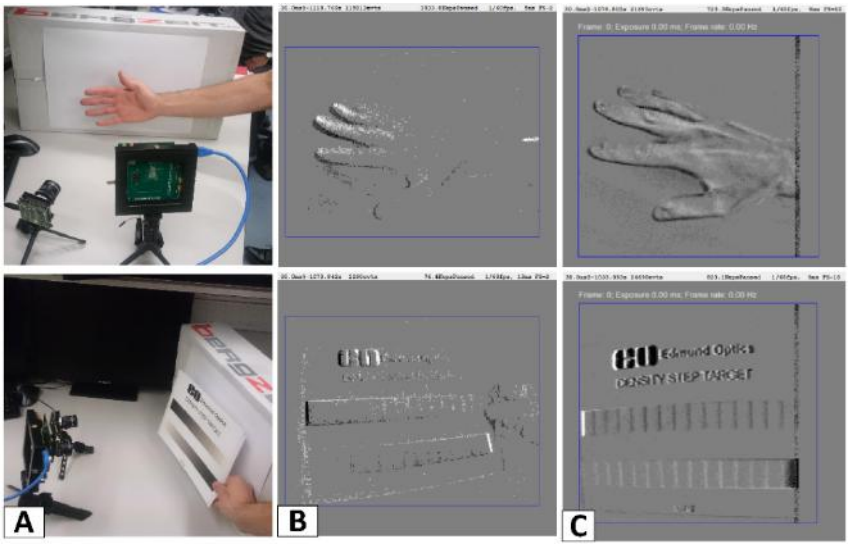


*Fig. 32 A: Moving stimuli: hand and Edmund Optics density step chart; B: DAVIS240C DVS raw output with Full Scale (FS) of 1; C: SDAVIS192 DVS raw output with full scale of 10 events. Both B and C show 30 ms DVS time slices. ON events are represented in white and OFF events in black.*

| | This work: **SDAVIS192** | Brandli et al. [32], DAVIS240C, measured with current setup | Serrano-Gotarredona et al. [36] | Yang et al. [145] |
|---|---|---|---|---|
| *Functionality* | Asynchronous temporal contrast + APS | Asynchronous temporal contrast + APS | Asynchronous temporal contrast | Asynchronous temporal contrast |
| *CMOS technology* | 180 nm 1P6M MIM CIS | 180 nm 1P6M MIM CIS | 350 nm 2P4M OPTO | 180 nm 1P6M |
| *Chip size mm²* | 5x5 | 5x5 | 4.9x4.9 | 3.2x1.6 |
| *Array size* | 192x188 | 240x180 | 128x128 | 60x30 |
| *Pixel size µm²* | 18.5x18.5 | 18.5x18.5 | 30x31 | 31.2x31.2 |
| *Fill factor* | 21.2% | 22% | 10.5% | 10.3% |
| *Peak absolute QE* | 24% | 7% | N.A. | N.A. |
| *Pixel complexity* | 47 transistors, 2 MIM caps, 1 MOS cap, 1 BPD, 1 microlens | 48 transistors, 2 MIM caps, 1 MOS cap, 1 SPD | N.A. | N.A. |
| *DVS interface* | 8-bit word-serial AER | 8-bit word-serial AER | Word-parallel AER | 6-bit word-serial AER |
| *Supply voltage* | 3.3/1.8 V | 3.3/1.8 V | 3.3 V | 1.8 V |
| *Power consumption* | DVS only (best): 15 mW (100k eps), 20 mW (2M eps) @ 100 lux. Normally 60 mW. APS only: 50 mW (1 fps), 60 mW (30 fps) @ 100 lux | Minimum (unconfirmed) at 4 mW (low activity), 15 mW (high activity) | 2.6 mW (1k eps)<br>4 mW (100k eps)<br>95 mW (20M eps) | 0.72 mW (10k eps) |
| *DVS dynamic range* | ON: at least 50 dB intrascene. OFF: at least 70 dB intrascene and at least 110 dB overall. With SNRs > 0 | ON: 60 dB intrascene. OFF: 60 dB intrascene. With SNRs > 0 | 60 dB intrascene, 120 dB overall[a] | 130 dB[a], obtained by waiving a hand in front of the sensor and arbitrarily recognizing it by eye |
| *DVS minimum contrast sensitivity* | ON: 3.45% (28.5 dB SNR)[b].<br>OFF: 0.95% (18 dB SNR)[b] | ON: 27.6% (19 dB SNR)[c].<br>OFF: 20.5% (Inf dB SNR)[c] | 1.5%[a] | 1% for 35% coefficient of variation[a] |
| *Max bandwidth* | 54 Meps (self-ack)<br>11 Meps (FPGA), 50 fps | 54 Meps (self-ack)<br>11 Meps (FPGA), 50 fps | 20 Meps | 10 Meps |
| *DVS optimized minimum latency* | ON: 35 µs @ 30 kcd/m²<br>OFF: 10 µs @ 30 kcd/m² | ON: 15 µs @ 30 kcd/m²<br>OFF: 20 µs @ 30 kcd/m² | 3.2 µs @ 2 klux[a] | N.A. |
| *DVS normal latency* | 0.1-2 ms[b] | 0.1-2 ms[b] | Up to 6 ms @ 0.2 lux | N.A. |
| *DVS FPN* | ON: 0.2-1% contrast<br>OFF: 0.8-0.9% contrast | ON: 7.5-15% contrast<br>OFF: 5.5-8.5% contrast | 0.9% contrast[a] | 35% coefficient of variation at 1% contrast[a] |
| *DVS bandwidth* | ON, OFF: 0.1-100 Hz[d] | ON, OFF: 0.1-1k Hz[d] | N.A. | N.A. |
| *APS dynamic range* | 56.12 dB | 52.57 dB | - | - |
| *APS FPN* | 3.39% | 0.97% | - | - |
| *APS DSNU* | 25.79 $e^-$ | 21.38 $e^-$ | - | |
| *APS dark signal* | 16k $e^-$/s (0.748 nA/cm²) | 18k $e^-$/s (0.858 nA/cm²) | - | - |
| *APS readout noise* | 61 $e^-$ | 57 $e^-$ | - | - |
| *APS conv. gain* | 22 µV/$e^-$ | 24 µV/$e^-$ | - | - |

*Table 2 Specification table. [a]ON and OFF not specified separately. [b]At 1 klux. [c]At 1 lux. [d]At 10 klux chip illumination (30 cd/m² source).*

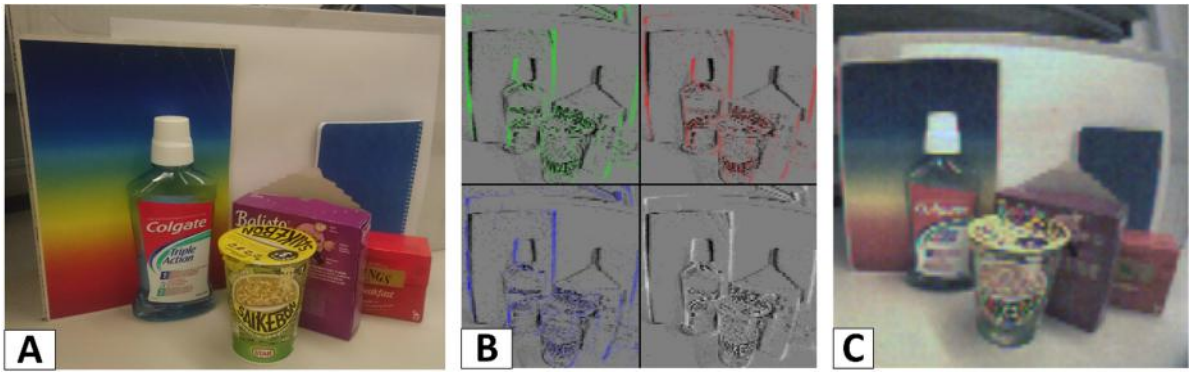


*Fig. 33 A: ground truth (cellphone camera); B: RGBW events separate by color (RGBW ON events are represented in RGBW respectively and OFF events in black); C: interpolated RGBW APS frame. Test pixels are ignored.*

Fig. 33 shows results from the RGBW split of SDAVIS192 that can provide alternating RGBW frames and RGBW DVS events [156]. Fig. 33A shows the ground truth scene. Fig. 33B shows the separate RGBW DVS channels. Fig. 33C shows a linearly interpolated APS frame.

Table 2 compares the design and measured specifications of SDAVIS192 with previous work.

## IV. Application in Neural Imaging

The high sensitivity of SDAVIS192 makes it potentially suitable for imaging of neuronal activity using calcium-sensitive florescence. Current imaging systems use expensive frame-based CCD or scientific CMOS cameras with 75 dB or more of DR and peak QEs up to 70%, but only about 15 fps frame rate when recording at full resolution. The output of a minute of recording can fill over 40 GB of storage. A DVS sensor can potentially reduce redundancy and increase time resolution because it responds only to fluorescence changes which streamline recordings from large neuron populations.

As a feasibility test of this capability and the first reported test of a DVS sensor for fluorescent imaging, the SDAVIS192 sensor was connected to the output of a microscope (a Zeiss Axioscope FS with a 40x NA 0.8 water dipping objective) with a 50-50 splitter for comparison with a Hamamatsu Orca V2 sCMOS. Hippocampal organotypic slice cultures were prepared from P6 mice(29) (all animal procedures were carried out according to the guidelines of the Center for Laboratory Animals of the University of Zurich and were approved by the Cantonal Veterinary Office) and virally transduced to express the “fast” calcium indicator GCaMP6f [149]. After 3 weeks in culture, slices were transferred to an upright microscope and during the experiment were perfused with aCSF containing bicuculline (50 µM), to block GABAA receptors and enhance neuronal excitation. Spontaneous action potentials result in calcium influx, causing a conformational change of the genetically encoded fluorescent protein, changing its emission properties. The green fluorescence is separated from the 470 nm excitation light by an emission filter (525/39 nm). Fig. 34 shows the parallel recording of several neurons with the Orca camera and of the SDAVIS192 DVS output at a time point when the highlighted neuron is becoming brighter. The 100 ms of DVS activity clearly highlights the active cell and its dendrites compared with the relatively silent background.

The frame rate of the Orca camera was limited to 10 Hz by the 100 ms exposure needed to image at the low intensity. Quantitative analysis of APS frames from SDAVIS192 show that the brightest parts of the scene produce a photocurrent of about 67k e-/s, equivalent to about 0.1 lux, and only about 4

times the dark current.

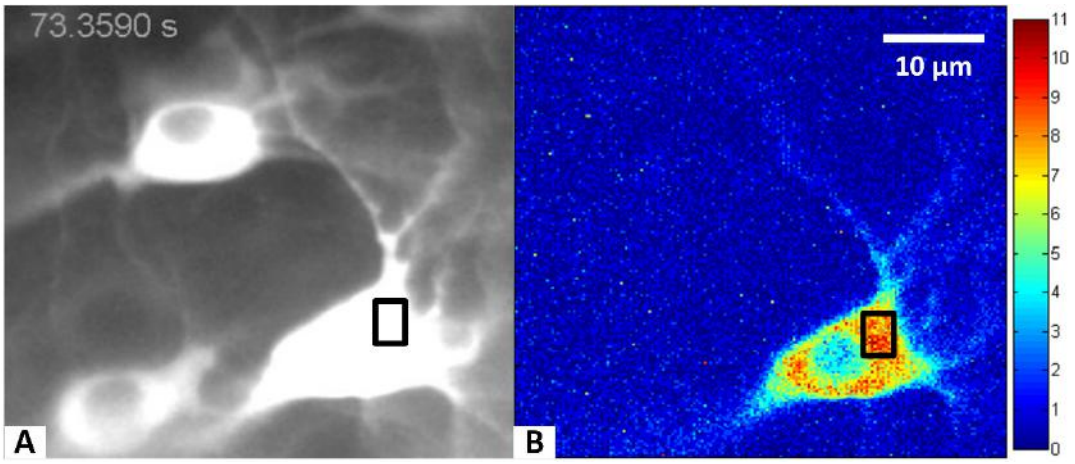


*Fig. 34 A: ORCA CCD camera frame compared to (B) the corresponding 100 ms DVS time slice of SDAVIS192. The regions of interest are indicated. Recordings: 40x-GCamp6f-fullLED-15-26-crop (APS) and 152640x (DVS).*

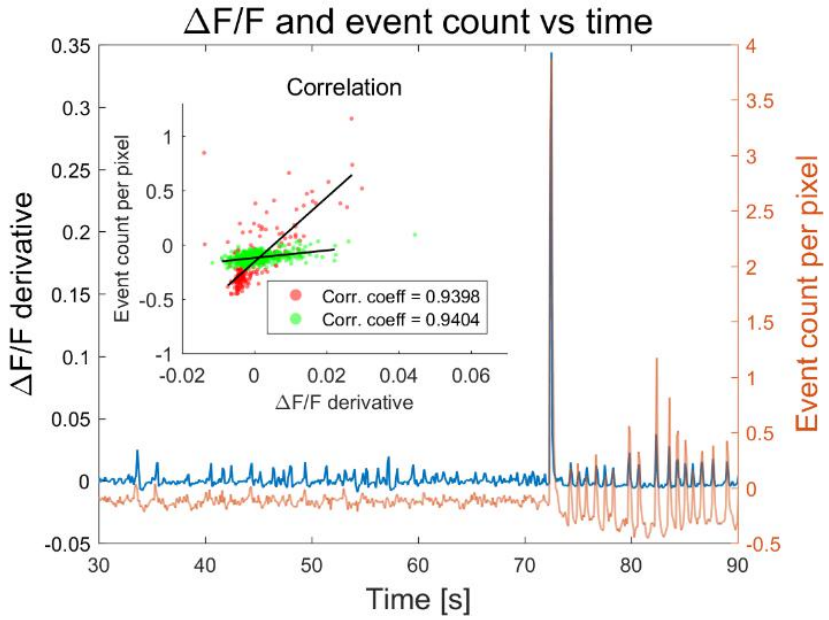


*Fig. 35 Synchronized $\Delta F/F$ of the region of interest of the Orca camera frame aligned in time with the DVS binned activity corresponding to the same area. The green scatter plot corresponds to the correlation before the bicuculline's effect (up to 72 s) and the red refers to the correlation after this.*

Fig. 35 compares the 100 ms bins of DVS activity of the selected 15x10 pixel Region Of Interest (ROI) of Fig. 34 with the $\Delta F/F$ of the same region over 60 s of activity before and after the bicuculline effect. $\Delta F/F$ is the normalized derivative of grey scale ROI as detected by the Orca camera over time. It is the percentage change from the previous point in time of the raw ROI average grey value. As can be seen, the DVS activity is synchronized in time with the fast transient activity, although slow dynamics are not visible in this simple binning analysis. The inset of Fig. 35 shows that the two signals correlate with a coefficient of 0.94. The dark background is out of the intrascene DR of SDAVIS192, which results in the pixels exposed to it to be noisy and sometimes self-oscillate. $V_{ref}$ in this dim illumination is non-ideal, but the signal is still correlated (green scatter plot of Fig. 35), although with smaller contrast

sensitivity. $V_{ref}$ was not adapted during the recording so as to avoid producing noise events due to the switching of the bias setting and to leave the sensor in an operating range working well with the brightness of the neurons.

When bicuculline takes effect at 72s, the sensor begins to work in its normal operating range (red scatter plot of Fig. 35). As the average illumination of the ROI increases, so does the sensitivity of the sensor. For example, at 50 s, the contrast threshold is 13.3% on average (ON and OFF together). After the large 35% increase in illumination at 72 s, at which the contrast threshold is 7.56%, the average contrast threshold from such point onward becomes 3.45% as the background illumination is higher and matches $V_{ref}$. The DVS sensitivity is still less than optimal as the ON and OFF contrast thresholds were slightly increased to reduce the quantity of events due to noise in low illuminations. An FFT spectral analysis of the DVS activity at bin width from 10ms to 200ms shows that the fluorescence signal is limited to about 5Hz bandwidth, confirming that in this experiment the 10Hz Orca frame rate was sufficient. The limited signal bandwidth is due to the slow kinetics of this calcium indicator.

## V. Conclusion

The design of the SDAVIS192 and its characterization show its lower contrast threshold (down to 0.95% for OFF and 3.45% for OFF) and at the same time, higher average DVS SNR of 30 dB. This reduces the intrascene DR to 50 and 70 dB for ON and OFF, but the digital adaptation mechanism inspired by [36] extends it up to 110 dB. Though the sensitivity of the sensor is similar to that reported in [36] and [145], the methods described in this work introduce a quantitative estimate of the sensor's performance through SNR, to assess the boundaries of its operation capabilities (minimum contrast threshold, latency and maximum DR). The 1:1 comparison with DAVIS240C shows how the numbers first reported in [32] change depending on the criterion selected and on the desired signal quality (SNR>0). Overall, all specifications were also obtained for ON and OFF separately.

While this work also reports standard APS characterization results, it presents the first detailed DVS quantum efficiency characterization, that was introduced in [156]. SDAVIS192 is also the first DAVIS sensor to produce, along with the color APS frames first introduced in [145], color RGBW DVS events (as used for image reconstruction in [156]).

A real calcium imaging experiment, as opposed to the simulation of [145], is shown in this work. The preliminary results are promising (0.94 correlation and detection of all intensity transients), but future work will improve the quality of the SDAVIS192 sensitivity in low light conditions, which is critical in this field. The aim is to move to in-vivo imaging with Voltage Sensitive Dyes (VSD) to resolve single action potentials. VSDs have latencies below 0.2 ms, but also $\Delta F/F$ below 2% [160]. Attempts at in-vitro color fluorescence microscopy, where different VSDs operate at different color bands, will also be made with the color-sensitive version of SDAVIS192.

## Acknowledgments

This research was supported by the European Commission project VISUALISE (FP7-ICT-600954), SeeBetter (FP7-ICT-270324), the Swiss National Science Foundation (31003A_149858; F.H.), the US NIH BRAIN Initiative (1U01NS090475-01, F.H.) and Samsung Electronics Corporation. We would like to thank the Sensors group at INI Zürich, in particular Vicente Villanueva for the PCB designs. We would also like to thank Bernabè Linares-Barranco from the National Microelectronics Center, Seville,

Spain, for useful discussions and Vasyl Motsnyi, from IMEC research institute, Leuven, Belgium, for his help in QE measurements. Furthermore, we would like to thank Dubravka Göckeritz-Dujmovic for preparing slice cultures and Urs Gerber.

## 4.2. More details on the Active Pixel Sensor characterization

The APS part of the DAVIS sensor produces intensity readouts at every pixel (frame). The Active Pixel Sensor (APS) characterization was overlooked in the previous section as it considered standard for all frame-based cameras. The following subsections analyze the characterization methods to obtain the main specifications of an APS sensor. The standards the APS characterization are based on are the European Machine Vision Association (EMVA) standards 1288 [161].

### 4.2.1. Photon Transfer Curve theory

The Photon Transfer Curve (PTC) is an important and standard measurement for APS cameras as the conversion gain of the sensor and various other specifications can be estimated from it [162]. This is performed with the uniform intensity sweep measurement setup described in the Appendix I. The PTC is normally obtained by sweeping the uniform light intensity and by plotting the spatio-temporal variance (noise) of the recorded frames' gray value versus their spatio-temporal mean value. The slope of this curve, at a particular point can be used to estimate the conversion gain of the sensor, as will be explained in the following sections. Frames need to be obtained with global shutter readout so that all pixels integrate in the same instant.

In order to start with the measurement, the sensitivity curve must be obtained first. This is the curve of the spatio-temporal mean gray value versus light intensity. The light intensity can be swept to obtain the curve but since a carefully controlled light source would be needed, this is not possible with the resources available at INI. As a matter of fact, since the LED of the setup is first of all not monochrome but white, its spectral characteristics will change as a function of bias (control voltage). To overcome this issue, the PTC can also be performed by sweeping the exposure duration of the sensor, to equivalently change the light integrated on the sensor's photodiodes. The results should be identical to sweeping the light intensity. In order to do so, the light intensity of the setup, described in more detail in Appendix I, is set, from the integrating sphere, to 1 klux (by setting $V_+$ accordingly, as in equation 4, with only its DC component $V_{+DC}$ active) and the exposure duration of the sensor is swept from 1 to 100 µs. At every exposure, a fixed number of frames (about 300) is recorded. It is important to note that the recorded frames are already the difference between integrated and reset ADC values, as Differential Double Sampling (DDS) is performed. The difference between DDS and Correlated Double Sampling (CDS) is illustrated below in Fig. 36 and consists in the timing at which the two samples are taken.

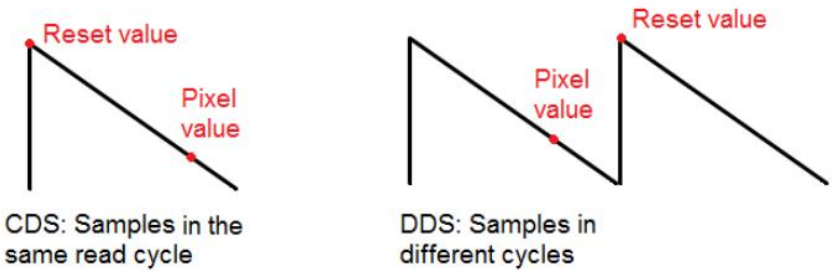


*Fig. 36 Comparison of Correlated Double Sampling (CDS) and Differential Double Sampling (DDS) techniques.*

Offset Fixed Pattern Noise (FPN), called Dark Signal Non-Uniformity (DSNU) and explained in section 4.2.7, is the fixed variation among all pixels' gray values in the dark. Together with reset voltage temporal (thermal) noise, it can be removed by CDS [163]. This sampling technique involves sampling the pixel's value and the reset voltage value (which changes in time) in the same read cycle. With DDS instead, only offset FPN can be removed, since the pixel's reset value is obtained in another read cycle (which will be different). The reset value of the pixel (taken an interval of time before with a different array-scan) is also output along with the value of the pixel. Subtraction and hence DDS is then performed off-chip.

The spatio-temporal mean of every pixel is first computed in time (across the 300 frames) and then in space (for a large patch of the frame, perhaps 100 x 100 pixels). The mean value is expressed in Digital Numbers (DN) which represent the gray value. An example of expected APS sensitivity curve is shown in Fig. 37.

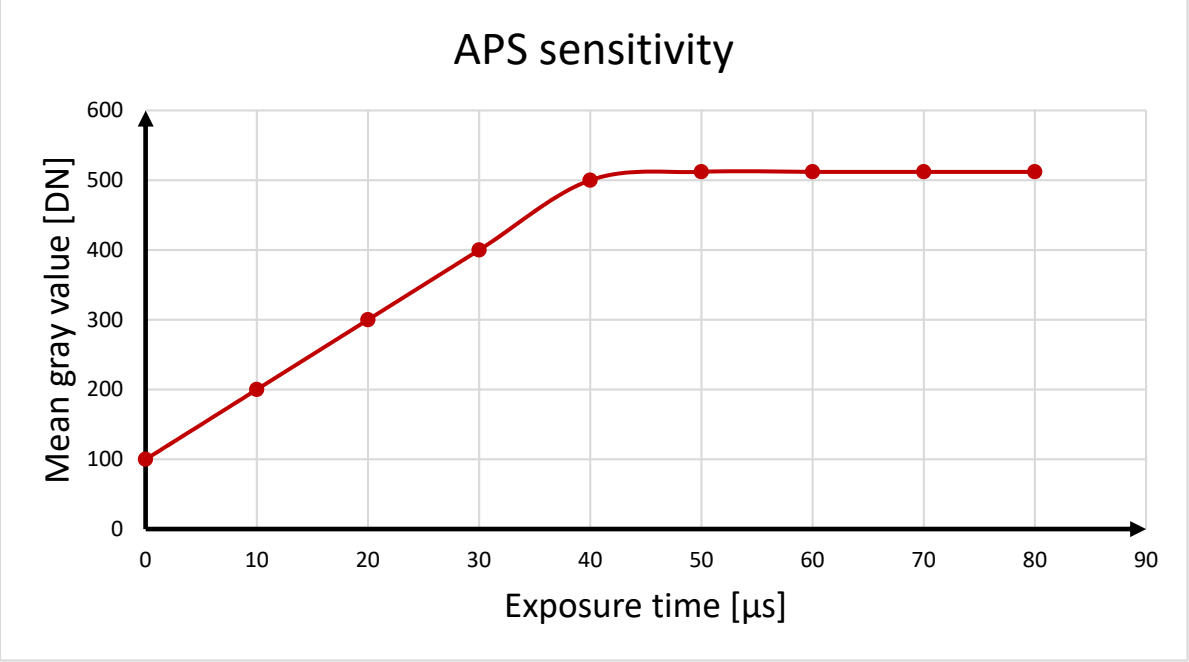


*Fig. 37 Example of sensitivity curve for the APS.*

As can be seen in Fig. 37, the curve starts linearly at a certain mean gray value. The starting point of the curve depends on the illumination chosen as even the shortest exposure duration (1 µs) is

sufficient for photocurrent to appear. For low illuminations, this offset will be close to zero. It will not be exactly zero as noise sources such as dark current contribute to this value. As more light falls onto the sensor, the mean gray value increases linearly. The DN number saturates when the charge collected from the photodiode fills the well entirely. This stage is called Full-Well (FW) and from here on, no more charge can be collected. Ideally, the lower ADC reference should be set in order to correspond with this level so as to fully use the ADC range and minimize quantization noise. The slope of the sensitivity plot is proportional to the photo-current and therefore on the illumination level.

For each of the points collected in the sensitivity plot (which according to literature should be at least six [162]) the variance can be computed in order to obtain the PTC. The PTC is based on the fact that since the sensor can be considered a black box with light as an input and DN as an output and the only noise introduced at the input is the shot noise due to the particle nature of light, the difference between output noise and input noise is the noise of the sensor itself [164]. The measured signal $S_{tot}$ is equal to the sum of signal offset $S_{off}$ with the signal created by the converted photons as equation (4.2.1.1) shows [162]:

$$S_{tot} = GN_o + S_{off} \tag{4.2.1.1}$$

where $G$ is the conversion gain of the sensor and $N_o$ represents the photo-generated electrons. The resulting temporal noise is then the sum of variance due to read and shot noises multiplied by $G^2$:

$${\sigma_{tot}}^2 = G^2({\sigma_{read}}^2 + {\sigma_{shot}}^2) \tag{4.2.1.2}$$

Shot noise, ${\sigma_{shot}}^2$, is due to the random variation of light and is expressed by equation (4.2.1.3):

$${\sigma_{shot}}^2 = 2e\bar{I}B \tag{4.2.1.3}$$

where $e$ is the electron charge, $\bar{I}$ is the average photo-current and $B$ is the bandwidth considered. Shot noise variance ${\sigma_{shot}}^2$ is therefore directly proportional to the signal and the standard deviation $\sigma_{shot}$ (the root mean square (rms) value of shot noise) is proportional to the square root of the signal. Plotted on a log-log scale against mean signal (mean number of incident photons on a pixel), the standard deviation $\sigma_{shot}$ appears as a straight line with slope 0.5. Shot noise can be modelled as a Poisson process, which means the mean (its deriving signal) is equal to the variance (noise). This gives (4.2.1.4):

$${\sigma_{shot}}^2 = N_o = \frac{(S_{tot} - S_{off})}{G} \tag{4.2.1.4}$$

And in turn, a total temporal noise of:

$${\sigma_{tot}}^2 = G^2{\sigma_{read}}^2 + G(S_{tot} - S_{off}) \tag{4.2.1.5}$$

The Signal to Noise Ratio equation can also be therefore derived in (4.2.1.6):

$$\frac{S_{tot}}{\sigma_{tot}} = \frac{(S_{tot} - S_{off})}{\sqrt{G^2{\sigma_{read}}^2 + G(S_{tot} - S_{off})}} \tag{4.2.1.6}$$

Since shot noise can therefore very well be modelled and other sources of noise can be individuated and recognized. An example of PTC plotting standard deviation vs mean signal can be seen in Fig. 38, where CDS or DDS is present. The total signal $S_{tot}$ is represented on the x-axis but $S_{off}$ is not subtracted. The variance $\sigma_{tot}$ is represented by the y-axis.

As can be seen Fig. 38, three regions can be identified. The first and leftmost region is the read noise dominated region: it is due to temporal noise in the readout circuitry ${\sigma_{read}}^2$ and it forms the noise floor of the sensor (it can be found by the interception of the PTC curve with the y-axis). For no incoming photons it is the only noise source visible, lower-limiting the PTC with a slope of zero. The second region is the one where shot noise ${\sigma_{shot}}^2$ takes over, with a slope of 0.5. If $S_{off}$ would be subtracted from $S_{tot}$, the conversion gain $G$ could be obtained by the x-axis intersection of the line fitted in the 0.5 slope region of the graph of Fig. 38. This estimate could however be imprecise, since the slope is generally smaller than 0.5 and the logarithmic plot is very compressive, making it hard to choose the right region to fit (the ideally shot-noise limited region). Finally, the last region is the FW region, where variance drops dramatically due saturation of the pixel output.

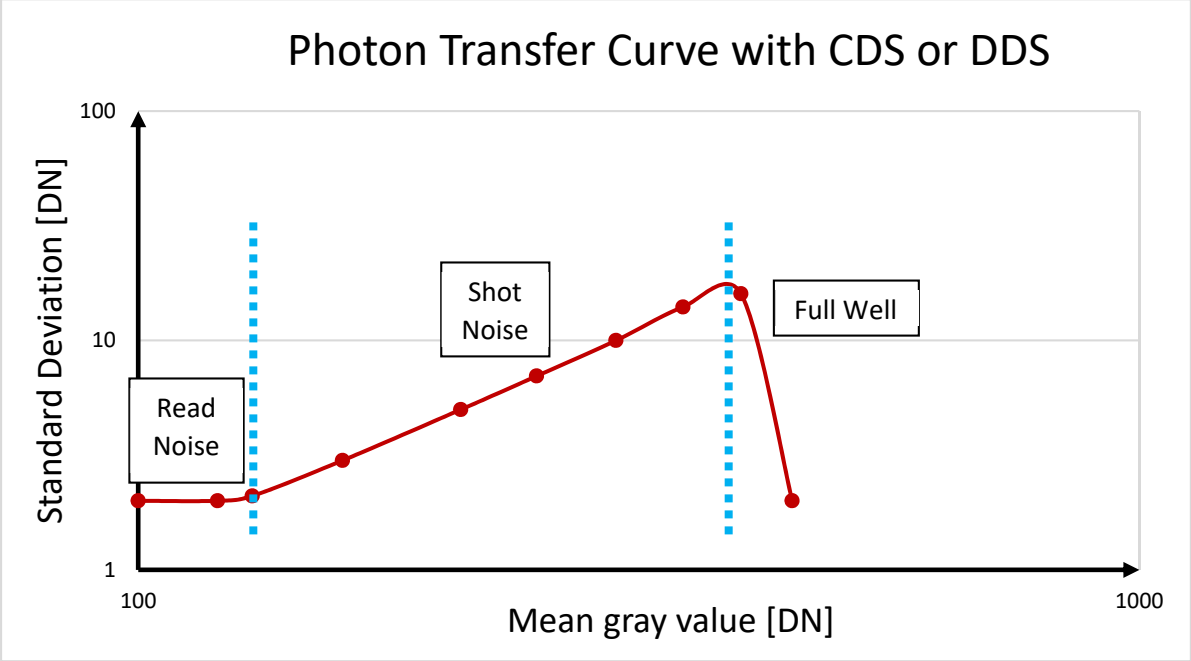


*Fig. 38 Example of photon transfer curve in log scale representing standard deviation vs mean signal.*

If the PTC is plotted as temporal variance in $DN^2$ vs mean signal on a linear scale, in the shot-noise limited part of the curve, the conversion gain $G$ can be obtained from the right-hand-side of equation (4.2.1.5), since ${\sigma_{read}}^2$ (the y-axis intercept of the curve) is negligible.

As said before, with the current setup, 300 frames are collected for each exposure level and at an illumination of 1 klux. These can be subdivided in square regions for separate PTC processing, if they are generated by different pixel designs. From these, the temporal variance ${\sigma_{tot}}^2$ can be extracted with equation (4.2.1.7), based on the EMVA standard 1288 [161]. This method is in theory more robust than [161] (since it uses not just two frames but 300), although no significant difference is observed when comparing to the strict 1288 standards for PTC.

$$ {\sigma_{tot}}^2 = \frac{1}{MNK}\sum_{x=1}^{M}\sum_{y=1}^{N}\sum_{z=1}^{K}\left(F(x,y,z) - \frac{1}{K}\sum_{z=1}^{K}F(x,y,z)\right)^2 \qquad (4.2.1.7) $$

where $F(x, y, z)$ is the pixel $x, y$ of the frame $z$, if there are $K$ $M$x$N$ frames. The second term in the subtraction is simply the temporal mean which is subtracted from the current pixel value $F(x, y, z)$. The spatial averaging is computed at last by the x and y summations on the left-hand-side of equation (4.2.1.7).

### 4.2.2. Conversion gain theory

The conversion gain $G$, expressed in microVolts per electron, is the gain of the sensor from detected electron to voltage created (also expressed in DN/$e^-$). Theoretically, this can be estimated with the simple equation for charge $Q$ to voltage conversion:

$$V = Q/C \tag{4.2.2.1}$$

where $V$ is the voltage on the APS node and $C$ is the conversion capacitance of such node (marked FD in Fig. 21). According to Dr. Brandli, for most of the new DAVIS sensors, this should be 18 fF with respect to ground. However, after Chenghan Li's accurate simulation of the pixel's behavior, since the n-type MOS capacitor is close to Vdd, the effective capacitance is reduced to less than 8 fF. This is because the MOS channel is not yet fully inverted and the oxide capacitance $C_{ox}$ is not the only capacitance present. The additional depletion capacitance $C_{dep}$ of the channel in subthreshold in series lowers the total capacitance value, therefore giving more conversion gain. For a single electron of charge $e$ (replacing $Q$ in equation (4.2.2.1), the conversion gain can be estimated in the range 20-25 μV/$e^-$.

As mentioned, the conversion gain can be extracted from the shot-noise limited part of the PTC. This is because increasing the mean input signal $S_{tot}$ to the sensor (with increasing exposure time) results in an output $GS_{tot}$. The variance ${\sigma_{tot}}^2$ on the other side increases as $G^2 S_{tot}$ [164] since for a Poisson process the mean signal is equal to signal variance. Therefore the ratio of the two quantities (and the slope of the PTC) gives the conversion gain $G$ in DN/$e^-$. To convert this value into microVolts per electron, the number must be multiplied by the ADC range divided by the number levels it covers (for example 1024). The new generation of DAVIS chips allows to choose between external and internal 10-bit (1024 levels) ADC and this can be done through the switch "useInternalADC". The external ADC THS1030 [165] has the reference voltages set by resistive dividers, giving an overall ADC range of 0.648 V for all DAVISFX3 chips. The internal 10-bit ADC has configurable ADC references set to give a good APS image. At the moment these are set to give an ADC range of 0.596 V for DAVIS346B and 1.29 V for SDAVIS192. DAVIS240C and DAVIS240B, both based on DAVISFX2, have an external ADC, set to an ADC range of 1.501 V. The aforementioned parameters are needed to convert $G$ in Volts per electron as equation (4.2.2.2) shows.

$$G[\mathrm{V}/e^-] = G[DN/e^-]\frac{ADC_{range}}{N_{levels}} \tag{4.2.2.2}$$

where $ADC_{range}$ is the ADC range and $N_{levels}$ is the number of levels of the ADC (1024 for 10 bits).

The slope should be obtained in the shot-noise limited region of the PTC. To check, $\sigma_{shot}$ can be plotted in a log-log plot against the mean gray value in DN. Where the slope is 0.5, the PTC slope can be computed. If the slope is lower than 0.5, then other noise sources are present and can influence the conversion gain. As a matter of fact any source of variance could falsely result in a conversion gain $G$ higher than in reality (as was observed for the external ADCs, perhaps for their different sample and hold readout).

### 4.2.3. Sensitivity results

The measurement of Photon Transfer Curve (PTC) was performed for the monochrome SDAVIS192 and all of the Towerjazz sensors with the setup illustrated in Fig. 25. The sensitivity plot for the white light source and at 1000 lux, is shown in Fig. 39 for a patch of 100 x 100 pixels in comparison with the results of the DAVIS240C sensor. DAVIS240C is identical to the sensor of [32], but it was manufactured with the same 180 nm Towerjazz process, for a fairer comparison, and it does not have the malfunctioning test pixels mentioned in [32]. Every point of the curve is the spatio-temporal average of 300 frames at the same exposure.

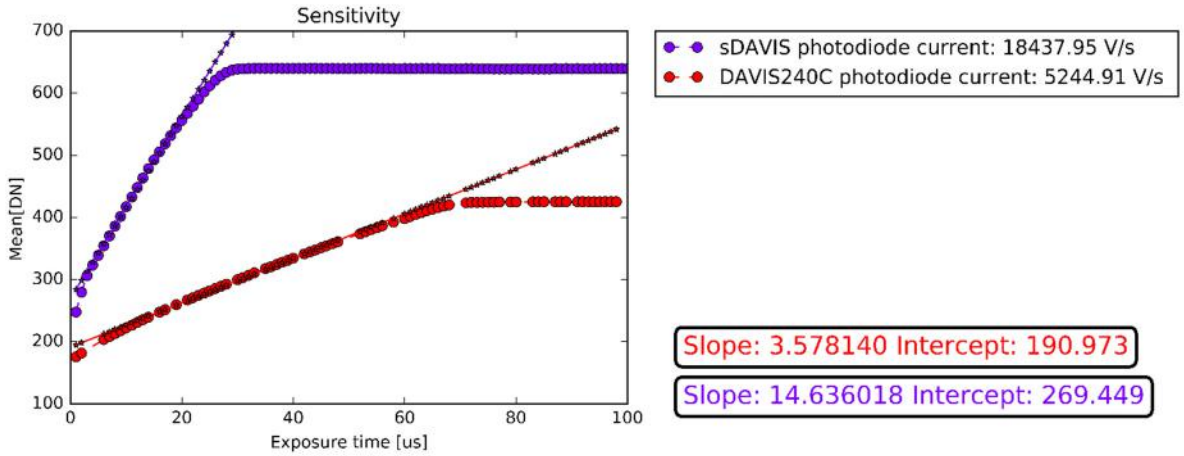


*Fig. 39 Sensitivity plot of SDAVIS192 Active Pixel Sensor (violet) and DAVIS240 pixels (red) at 1 klux illumination.*

As can be seen from Fig. 39, the SDAVIS192 saturates at about 640 DN ($DN_{max}$), corresponding to a Full Well Capacitance (FWC) of 36'648 $e^-$ (using the conversion estimated from simulation of 22 $\mu V/e^-$). In comparison, the reference 240C design saturates at 425 DN ($DN_{max}$), with a FWC of 25'957 $e^-$ (for a simulated conversion gain of 24 $\mu V/e^-$) A single DN corresponds to 1.26 mV (1.29 V/1024 ADC levels) for SDAVIS192 and to 1.47 mV (1.501 V/1024 ADC levels) for DAVIS240C. The SDAVIS192 seems to be four times more sensitive to light level than DAVIS240C, as the slope of its sensitivity curve is four times higher for the same illumination. The low reference voltage corresponding to 1024 DN is not ideally placed where the saturation of the sensor occurs and this in turn reduces the possible Dynamic Range (DR) of the sensor. The choice of reference voltages was however not optimized to achieve just the 4 extra dB of DR that could in theory further be obtained, to reach the theoretical limit of 60.2 dB with such ADC. The reason for this is that the APS part of the DAVIS circuit is not the main feature of the DAVIS sensor and a 60 dB DR is anyway well below the state-of-the-art for APS imaging.

### 4.2.4. Photon Transfer Curve and conversion gain results

The extracted linear PTC curves of variance ${\sigma_{tot}}^2$ versus spatio-temporal mean signal level are shown in Fig. 40 for DAVIS40C and SDAVIS192 obtained with the setup illustrated in chapter 6 at 1 klux. As can be seen the variance of DAVIS240C is about half that one of the SDAVIS192. This is to be expected since the DAVIS240C uses an external state-of-the-art ADC with lower mismatch than the custom-made single-slope ADC of the SDAVIS192 has. Fitting a straight line in the rising edge of the plot, the respective conversion gains $G$ can be obtained. As equation (4.2.1.5) states, $G$ is the slope of the line fitted in the shot-noise limited part of the curve. For DAVIS240C this is 0.021 DN/$e^-$ and for

SDAVIS192 this is instead 0.023 DN/$e^-$. These can be converted in μV/$e^-$ through equation (4.2.2.2) and correspond to respectively 31.29 μV/$e^-$ and 29.56 μV/$e^-$. These numbers do not match the simulated 24 μV/$e^-$ and 22 μV/$e^-$, this is because measuring conversion gain is extremely difficult, as any oscillation in the light and power fluctuation in the chip can give a conversion gain higher than possible. To obtain reasonable numbers, the simulated values of conversion gain are used to convert numbers into electrons in the next sections.

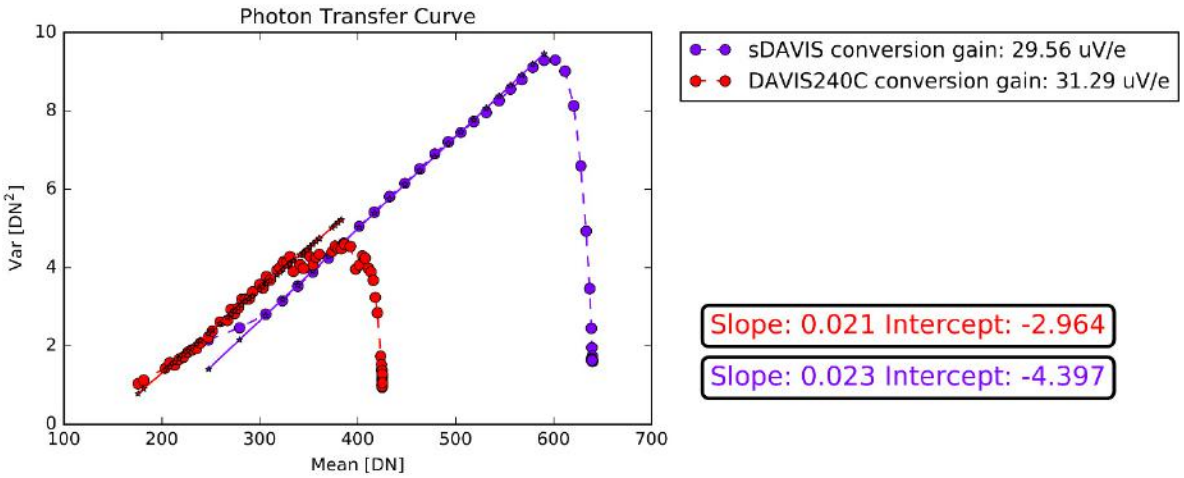


*Fig. 40 Linear plot of Photon Transfer Curve (variance vs spatio-temporal mean) of SDAVIS192 Active Pixel Sensor (violet) and DAVIS240C pixels (red) at 1 klux.*

The PTC plot of Fig. 40 (representing variance vs mean) can be re-drawn in a log-log plot and by substituting the standard deviation to the variance and by subtracting the initial value of the PTC spatio-temporal mean to the x-axis, as shown in Fig. 41. Fitting a straight line in the last part of the curve, before saturation occurs, allows to see whether the PTC measurement is reliable: if the slope is not 0.5, then an error in the conversion gain estimate is present. For DAVIS240C and SDAVIS192 these are 0.34 and 0.378 respectively, confirming that an important error occurs in the estimation of the conversion gain. In this case, for the previously extracted values, the error is roughly 37% larger numbers. From Fig. 41, the conversion gain can be again estimated from the x-axis intersection of the straight line, though this measure is even more imprecise. The resulting conversion gains of DAVIS240C and SDAVIS192 are 18 μV/$e^-$ and 20.56 μV/$e^-$ respectively and seem to better match the simulated values, although this shouldn't be the case.

Furthermore, it was noticed that measurements of conversion gain $G$ were heavily dependent on illuminance level, highlighting that increasing noise ${\sigma_{tot}}^2$ was recorded with increasing light intensity. Since the noise of the light source grows with light intensity, this gives a $G$ much larger than in reality (and simulation), and increasing with illumination level. $G$ can however also vary with signal as it is dependent on capacitance and this can vary a little with signal. Another possible partial explanation for the higher conversion gain, devised by Chenghan Li and which can be found in his thesis, is the additional temporal noise of charge injection whenever the transfer gate of the APS is turned off. The explanation is only partial since the APS transfer gate transistor MN3 in Fig. 21 is very small and cannot therefore contribute too much charge injection.

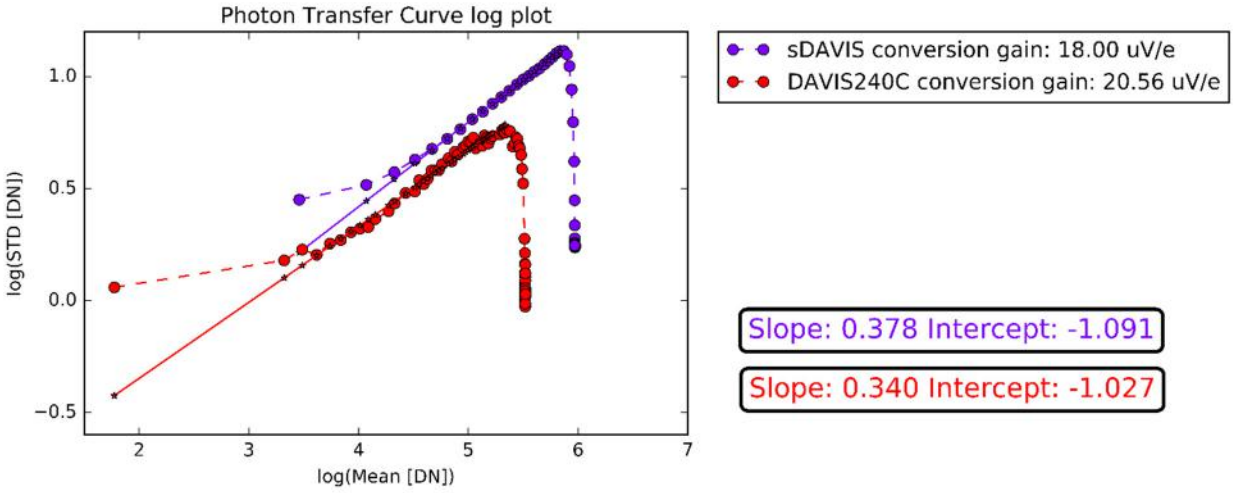


*Fig. 41 Log-log plot of Photon Transfer Curve (standard deviation vs spatio-temporal mean with initial offset subtracted) of SDAVIS192 Active Pixel Sensor (violet) and DAVIS240 pixels (red) at 1 klux.*

### 4.2.5. Read noise theory

The read noise is the noise level at zero illumination $\sigma_{read}$. Multiplying the DN of $\sigma_{read}$ (the rms standard deviation, not the variance) by the conversion gain $G$, gives the noise floor in electrons. This is in turns gives the sensitivity of the detector: the minimum size the signal has to have to have to be recognized as such.

### 4.2.6. Dynamic Range theory

Knowing the read noise allows to estimate the Dynamic Range (DR). The read noise is in fact the smallest detectable signal and, in DN, it corresponds to the temporal noise for no exposure ($\sigma_{read}$). The maximum detectable signal is simply the highest value in DN the signal can reach in the sensitivity plot ($DN_{max}$). The DR in dB can be computed as equation (4.2.6.1).

$$DR = 20\log\left(\frac{DN_{max}}{\sigma_{read}}\right) \tag{4.2.6.1}$$

For a 1024-level ADC and noise level less than 1 DN (which by quantization needs to be approximated to 1 DN), the DR is theoretically as high as 60.2 dB.

### 4.2.7. Fixed Pattern Noise theory

The Fixed Pattern Noise (FPN) is the mismatch among pixel sensitivities. It can be calculated for every exposure length as the average standard deviation of the temporal mean of the pixels with respect to the spatio-temporal mean of the signal as equation (4.2.7.1) shows:

$$FPN = \sqrt{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\left(\frac{1}{K}\sum_{z=1}^{K}F(x,y,z) - \frac{1}{MNK}\sum_{x=1}^{M}\sum_{y=1}^{N}\sum_{z=1}^{K}F(x,y,z)\right)^2} \tag{4.2.7.1}$$

where $x$ and $y$ are the coordinate of the pixel $M$x$N$ pixel array and $z$ is the selected sample of all the $K$samples taken.

FPN is expected to rise when pixels approach full well. This is because while some pixels will saturate first at some DN level, while others will keep increasing in level until their mismatched Full Well

Capacity (FWC) is reached. This will make their relative value even more different and FPN will increase.

Total FPN can also be calculated along the x and y axis of the sensor ($FPN_{tx}$ and $FPN_{ty}$ respectively) to observe the total contributions along each axis due to pixel and ADC mismatch. As shown in (4.2.7.2) and (4.2.7.3), this is achieved by calculating the spatio-temporal variance for every column or row with respect to its spatio-temporal mean and by averaging among all such columns or rows.

$$FPN_{tx} = \frac{1}{M}\sum_{x=1}^{M}\sqrt{\frac{1}{MN}\sum_{y=1}^{N}\sum_{x=1}^{M}\left(\frac{1}{K}\sum_{z=1}^{K}F(x,y,z) - \frac{1}{MK}\sum_{x=1}^{M}\sum_{z=1}^{K}F(x,y,z)\right)^2} \quad (4.2.7.2)$$

$$FPN_{ty} = \frac{1}{N}\sum_{y=1}^{N}\sqrt{\frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\left(\frac{1}{K}\sum_{z=1}^{K}F(x,y,z) - \frac{1}{NK}\sum_{y=1}^{N}\sum_{z=1}^{K}F(x,y,z)\right)^2} \quad (4.2.7.3)$$

To observe just the importance of ADC column mismatch, which strongly affects one dimension, the pixel to pixel mismatch can be averaged out by re-computing $FPN_{tx}$ and $FPN_{ty}$ as $FPN_{ax}$ and $FPN_{ay}$ as shown in (4.2.7.4) and (4.2.7.5). These are obtained by calculating the variance of the average column/row temporal mean with respect to the total array spatio-temporal mean and by averaging spatially along such column/row axis.

$$FPN_{ax} = \sqrt{\frac{1}{M}\sum_{x=1}^{M}\left(\frac{1}{NK}\sum_{y=1}^{N}\sum_{z=1}^{K}F(x,y,z) - \frac{1}{MNK}\sum_{x=1}^{M}\sum_{y=1}^{N}\sum_{z=1}^{K}F(x,y,z)\right)^2} \quad (4.2.7.4)$$

$$FPN_{ay} = \sqrt{\frac{1}{N}\sum_{y=1}^{N}\left(\frac{1}{MK}\sum_{x=1}^{M}\sum_{z=1}^{K}F(x,y,z) - \frac{1}{MNK}\sum_{x=1}^{M}\sum_{y=1}^{N}\sum_{z=1}^{K}F(x,y,z)\right)^2} \quad (4.2.7.5)$$

$FPN_{ax}$ and $FPN_{ay}$ are then smaller than $FPN_{tx}$ and $FPN_{ty}$ and only encode for pure row and column mismatch. The axis which does not have the ADC mismatch ($FPN_{ax}$ or $FPN_{ay}$), will remain similar to $FPN_{tx}$ or $FPN_{ty}$. However, the axis which is subject to ADC mismatch, should exhibit a larger and roughly constant FPN. The visual result of ADC mismatch is evident horizontal or vertical stripes in the frame.

FPN can be separated into two contributors: Dark Signal Non-Uniformity (DSNU), which is the random offset (dependent on temperature) of each pixel in the dark, and the Photo Response Non-Uniformity (PRNU) which is the random variance in pixel gain as a function of light. DSNU can be almost completely eliminated by frame subtraction as it is done in CDS: subtracting the signal read from the reset read. In DDS this is not the case completely as the dark signal has temporal noise which does not get cancelled. As for PRNU, this remains and is visible. According to the EMVA standard 1288 [161], DSNU can be computed as the FPN in the dark at the exposure $e_{50}$ which corresponds to the

exposure at which the signal reaches 50% of its saturation level in the normal PTC curve (in the light). As for PRNU in percentage this can be computed as:

$$PRNU = \frac{\sqrt{{\sigma_{tot\_50}}^2 - {\sigma_{dark}}^2}}{S_{tot\_50} - S_{dark}} \quad (4.2.7.6)$$

where ${\sigma_{tot\_50}}^2$ is the spatio-temporal variance at $e_{50}$, ${\sigma_{dark\_50}}^2$ is the spatio-temporal variance in the dark at $e_{50}$, $S_{tot\_50}$ and $S_{dark}$ are the spatio-temporal means of the signal in the light and dark respectively at $e_{50}$. If ${\sigma_{dark}}^2$ is small enough (thanks to DDS or CDS) and $S_{dark}$ is zero, PRNU can be approximated by $\sigma_{tot\_50}$ divided by $S_{tot\_50}$.

### 4.2.8. Fixed Pattern Noise results

FPN can be obtained from equation (4.2.7.1) and be plotted for both sensors in Fig. 42 (expressed in DN) and Fig. 43 (expressed in percentage). In Fig. 42 it can be seen that FPN is lower for low illumination levels as most of the DSNU is eliminated by dark frame subtraction. It then very slowly rises for with signal level due to PRNU until it abruptly jumps to a higher value once full well occurs, as described in 4.2.7. In Fig. 43 instead, the FPN in percentage highlights the effect of PRNU (gain mismatch) which depends on illumination.

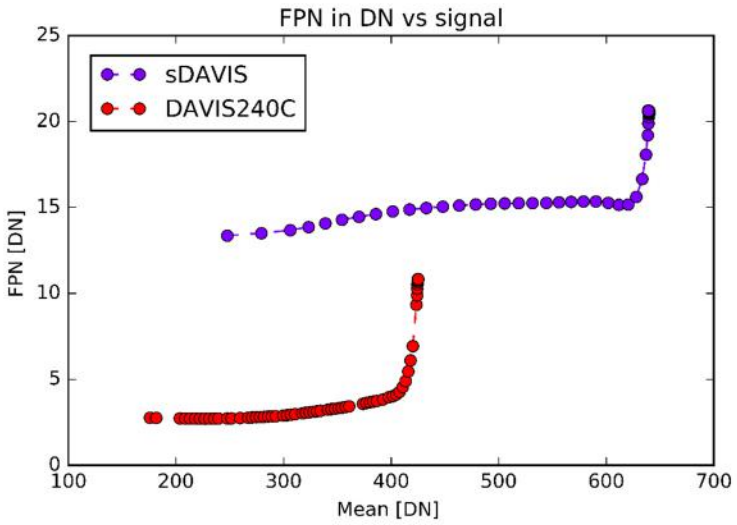


*Fig. 42 Fixed Pattern Noise expressed in DN vs spatio-temporal mean of SDAVIS192 Active Pixel Sensor (violet) and DAVIS240C pixels (red) at 1 klux.*

FPN at the exposure $e_{50}$, which corresponds to 50% of the ADC saturation value is generally quoted as the total FPN of the sensor. For DAVIS240C this is 2.91 DN (0.97%) at 300.5 DN of spatio-temporal mean. For SDAVIS192 this is instead a bit more than 3 times worse: 15.05 DN (3.39%) at 443.8 DN of spatio-temporal mean.

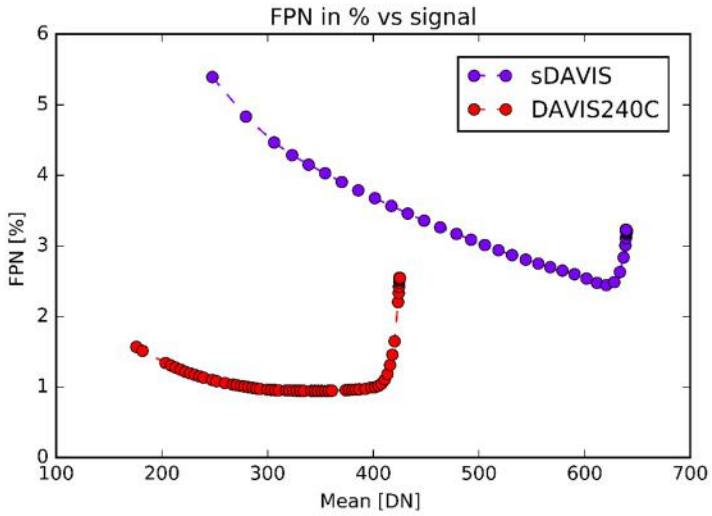


*Fig. 43 Fixed Pattern Noise expressed in percentage vs spatio-temporal mean of SDAVIS192 Active Pixel Sensor (violet) and DAVIS240C pixels (red) at 1 klux.*

As stated in section 4.2.7, FPN can be divided into its x and y components $FPN_{tx}$ and $FPN_{ty}$ which account for pixel plus ADC column/row mismatch. The plots in DN and percentage value for these FPNs for both sensors, obtained from equations (4.2.7.2) and (4.2.7.3) are shown in Fig. 44 and Fig. 45. No large visible difference appears between such plots and Fig. 42 and Fig. 43, as the ADC column mismatch is masked by the pixel's mismatch. At $e_{50}$, $FPN_{tx}$ and $FPN_{ty}$ are 2.5 DN and 2.76 DN respectively (0.83% and 0.92%) for DAVIS240C and 14.8 DN and 14 DN respectively (3.32% and 3.16%) for SDAVIS192.

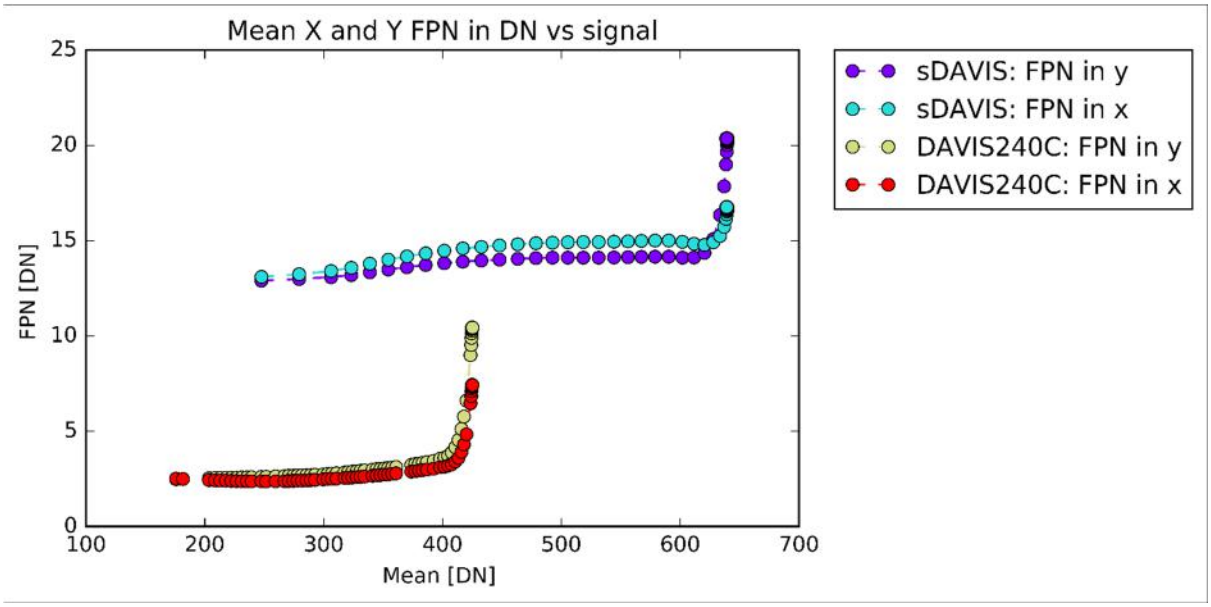


*Fig. 44 X and Y Fixed Pattern Noises $FPN_{tx}$ and $FPN_{ty}$, from equations (4.2.7.2) and (4.2.7.3), expressed in DN vs spatio-temporal mean of SDAVIS192 Active Pixel Sensor (light blue and violet) and DAVIS240C pixels (red and yellow) at 1 klux.*

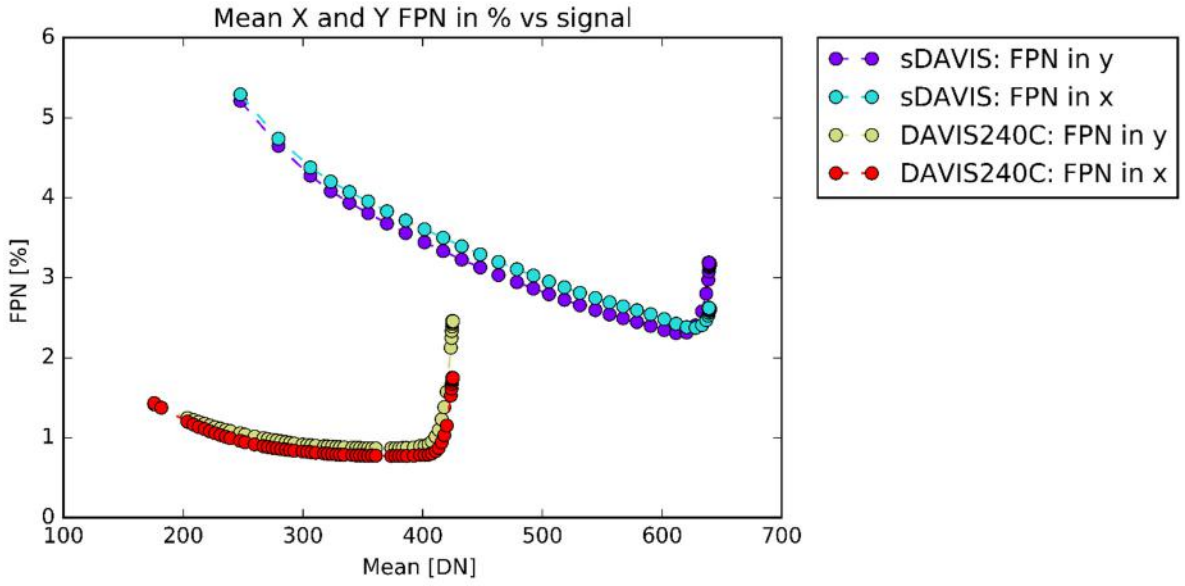


*Fig. 45 X and Y Fixed Pattern Noises $FPN_{tx}$ and $FPN_{ty}$ from equations (4.2.7.2) and (4.2.7.3), expressed in percentage vs spatio-temporal mean of SDAVIS192 Active Pixel Sensor (light blue and violet) and DAVIS240C pixels (red and yellow) at 1 klux.*

To truly observe the effect of ADC column mismatch, $FPN_{ax}$ and $FPN_{ay}$ from equations (4.2.7.4) and (4.2.7.5) must be observed (reported in Fig. 46 and Fig. 47).

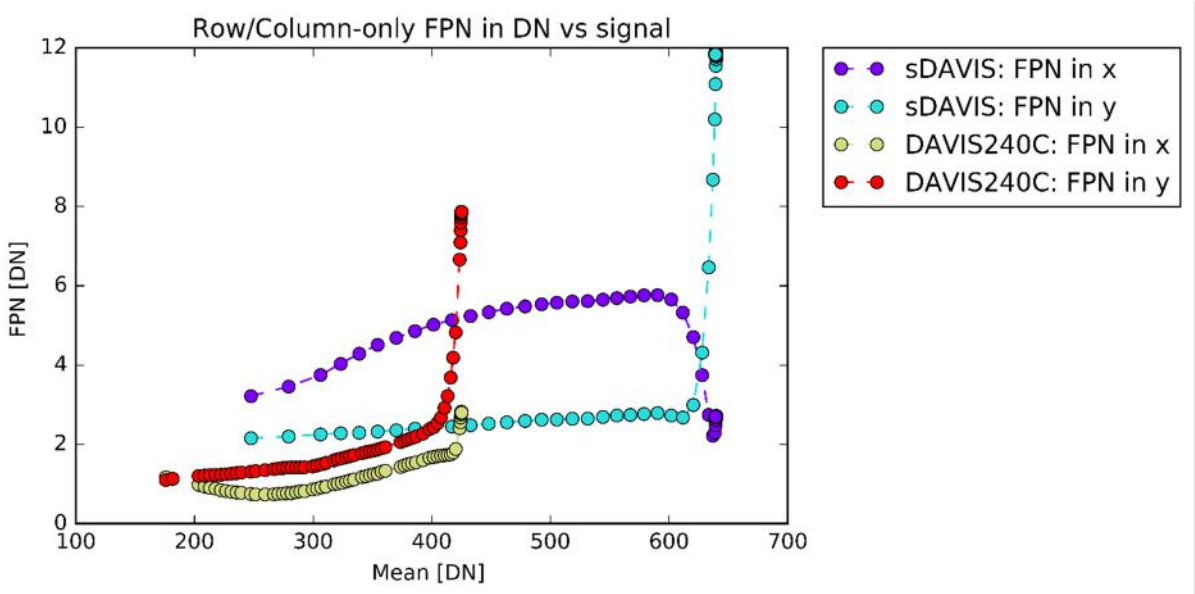


*Fig. 46 X and Y Fixed Pattern Noises $FPN_{ax}$ and $FPN_{ay}$ from equations (4.2.7.4) and (4.2.7.5), expressed in DN vs spatio-temporal mean of SDAVIS192 Active Pixel Sensor (violet and light blue) and DAVIS240C pixels (yellow and red) at 1 klux.*

In these FPNs (Fig. 46 and Fig. 47), the pixel-to-pixel x and y mismatch has been averaged off. This means that the FPN recorded is due to column and row mismatch such as the one caused by the ADC column mismatch. As a matter of fact for SDAVIS192, while $FPN_{ay}$ looks just like FPN for SDAVIS192, $FPN_{ax}$ highlights a roughly constant FPN in DN (and therefore decreasing in percentage) due to ADC

mismatch. The reason for the curved shape of $FPN_{ax}$ might be explained with the superposition of PRNU. Since for SDAVIS192 mismatch appears on the x-axis, the ADC is column-parallel. This effect is reduced for DAVIS240C as the ADC is external and state-of-the-art. At $e_{50}$, $FPN_{ax}$ and $FPN_{ay}$ are 0.89 DN and 1.47 DN respectively (0.29% and 0.49%) for DAVIS240C and 5.3 DN and 2.5 DN respectively (1.20% and 0.57%) for SDAVIS192

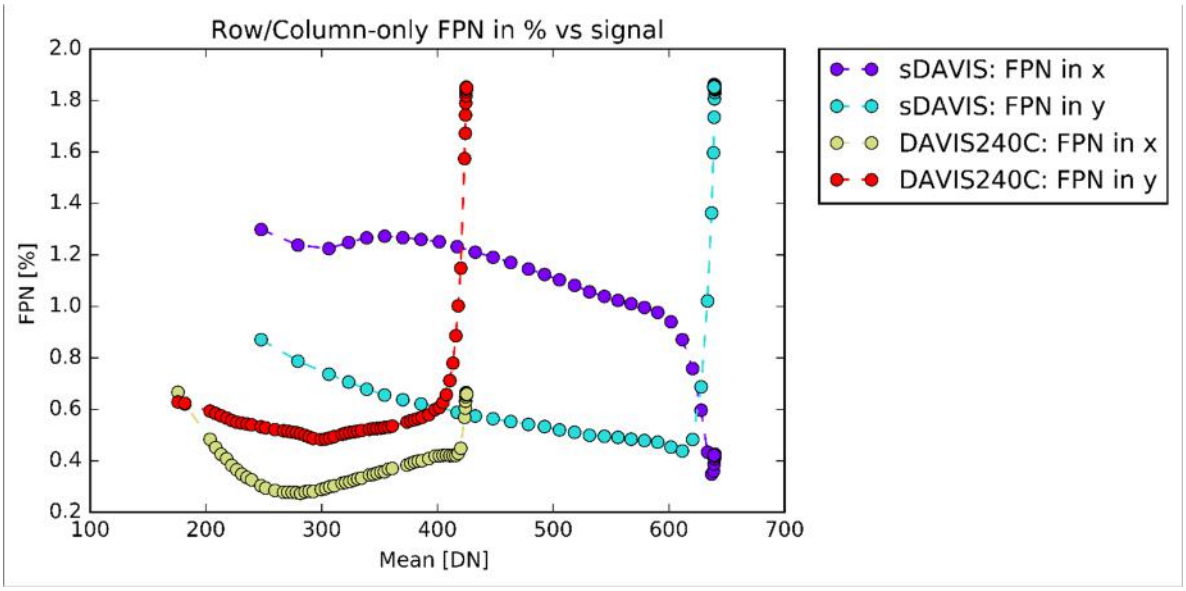


*Fig. 47 X and Y Fixed Pattern Noises $FPN_{ax}$ and $FPN_{ay}$ from equations (4.2.7.2) and (4.2.7.3), expressed in percentage vs spatio-temporal mean of SDAVIS192 Active Pixel Sensor (violet and light blue) and DAVIS240C pixels (yellow and red) at 1 klux.*

### 4.2.9. Dark current

Dark current is the contribution of the leakage current of the photodiode in dark, as described in 7.1, and of all the transistor junctions' leakage of the readout circuit positioned on top of the photodiode. Dark current can be estimated by performing PTC measurements in the dark for very long exposures. The sensitivity curve can be obtained, starting from the zero value (if the ADC high reference voltage is set correctly) and increasing with slope $dDN/dt$ directly equal to the dark current expressed in DN/s. This is because the dark current $I_D$ is the current in a capacitor where the voltage $V$ discharges linearly in time from the capacitor $C$.

$$I_D[V/s] = \frac{ADC_{range}}{N_{levels}} \frac{dDN}{dt} \tag{4.2.8.1}$$

Using the conversion gain $G$, obtained from the PTC in the light, allows to convert $I_D$ in electrons without assuming a value for $C$. As a matter of fact, to convert this current is $e^-$/s, $I_D$ needs to be divided by $G$:

$$I_D[e^-/s] = \frac{ADC_{range}}{N_{levels} G} \frac{dDN}{dt} \tag{4.2.8.2}$$

and multiplied by the electron charge $e$ to obtain $I_D$ in A/s:

$$I_D[A/s] = eI_D[e^-/s] \tag{4.2.8.3}$$

### 4.2.10. Dark current, read noise, Dynamic Range and DSNU results

The rest of the specifications can only be obtained with the PTC measurements in the dark. If the same procedure is followed with the sensors completely covered, the sensitivity curves of Fig. 48 are obtained. This time the slope is equal to the dark current $I_D$ in DN/s. For DAVIS240C, after the conversion, this corresponds to 18'357 $e^-$/s (or equivalently a dark droop rate of 0.44 V/s). The number, for SDAVIS192, should theoretically be one third of such value for similar transistor and photodiode sizing, but buried instead of surface photodiode, as mentioned in 7.2.1 and 7.2.2. The value of $I_D$ is instead only about 21% smaller: 15'993 $e^-$/s (dark droop rate of 0.35 V/s). Assuming all of the photodiode area is contributing to $I_D$, DAVIS240C and SDAVIS192 leak respectively at 0.858 nA/cm$^2$ and 0.748 nA/cm$^2$. This result is unexpected but reflects reality: a one second exposure is sufficient to almost saturate the sensor in the dark. The same is observed in all of the other Towerjazz waver sensors DAVIS128, DAVIS346 and DAVIS640. DAVIS240C does not saturate as fast, although $I_D$ is very similar: the reason for this is that, as can be seen in the sensitivity plot in the dark of Fig. 48, the lower reference voltage of the external ADC clips the dark current until about 400 ms, effectively giving a dark image. This high leakage is on average 16 times what is considered good (50 pA/cm$^2$) and close to vanilla n-well photodiode performance (1 nA/cm$^2$). The reasons for this problem might be three:

1. Wrong layout of the photodiode: a mask was forgotten or wrongly placed during the layout of the pixel. All chips use the same buried photodiode layers because they take one model as an example (the error might have been duplicated over and over);
2. The leakage of the transistor junctions is too large: too many transistors stacked on top of the photodiode leak considerably into this sensitive node. The leakage current is larger than the dark current. Unfortunately, I could not find the process specifications regarding junction leakage;
3. The numbers for dark current present in the datasheet refer to the photodiode operated under a different condition, i.e. with greater reverse bias. Although dark current should grow exponentially with reverse bias, perhaps the photodiode was designed for low dark current in this specific condition.

This problem completely spoils the performance in the dark of SDAVIS192 and sensors with same APS pixel design.

From Fig. 48, since the sensor's spatio-temporal mean starts from 0.3 DN for DAVIS240C and from 0.05 DN for SDAVIS192, the read noise $\sigma_{read}$ (and therefore the sensitivity) is limited by ADC quantization, i.e. 1 DN (61.08 $e^-$ and 57.26 $e^-$ respectively, using the simulated values of conversion gain). Using equation (4.2.6.1), and by substituting the respective $DN_{max}$, the Dynamic Range (DR) of both sensors is 52.57 dB and 56.12 dB respectively. As stated in section 4.2.6, both are 7.63 dB and 4.08 dB below the theoretical maximum (which could be possibly achieved by changing the ADC's low reference to match the APS' circuit maximum voltage swing).

By observing the FPN in the dark in Fig. 49, the DSNU, described in section 4.2.7, can be inferred. This corresponds 21.38 $e^-$ and 25.79 $e^-$ (using the simulated conversion gains of 24 and 22 μV/$e^-$) for DAVIS240C and SDAVIS192 respectively at the spatio-temporal mean level corresponding to $e_{50}$ in the dark. Since this value is very small, even using equation (4.2.7.6), the PRNU turns out just to be the

FPN at $e_{50}$ in the light (0.97% and 3.39%, or 2.91 DN and 15.05 DN for DAVIS240C and SDAVIS192 respectively).

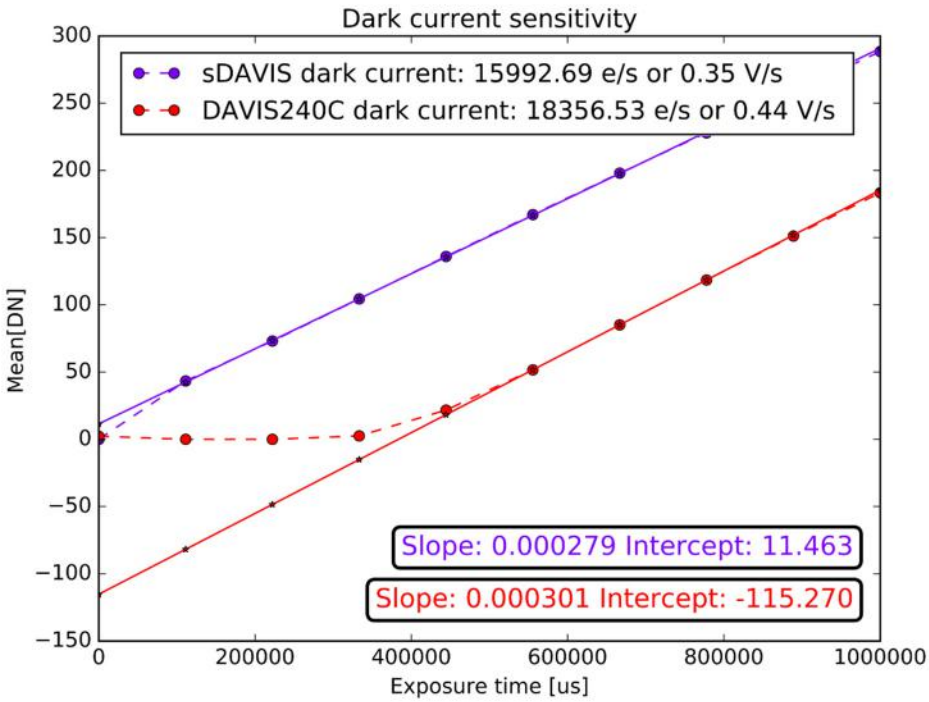


*Fig. 48 Sensitivity plot of SDAVIS192 Active Pixel Sensor (violet) and DAVIS240 pixels (red) in the dark. The value in electrons per second is computed with the simulated conversion gains.*

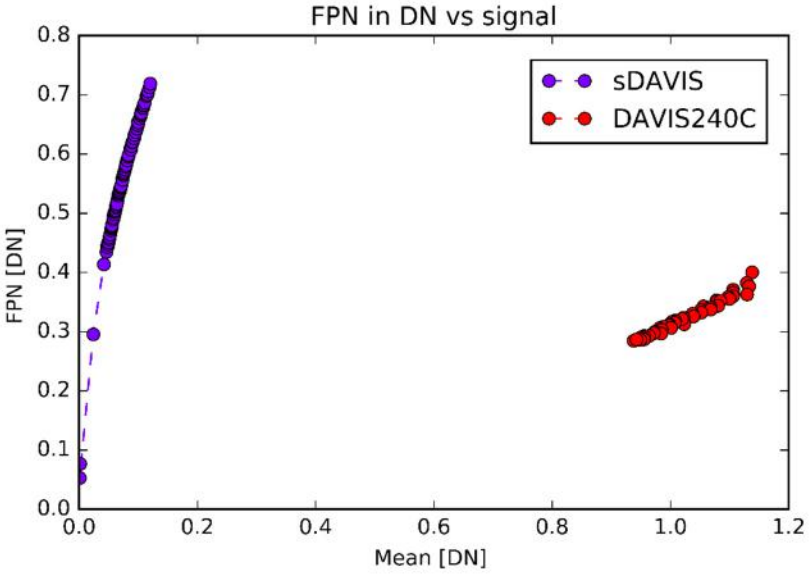


*Fig. 49 Fixed Pattern Noise in the dark, expressed in DN vs spatio-temporal mean of SDAVIS192 Active Pixel Sensor (violet) and DAVIS240C pixels (red).*

### 4.2.11. External Quantum Efficiency theory and setup

External Quantum Efficiency (EQE) is the ratio of photo-generated electrons $e_{gen}$ and number of total incident photons $N_{ph}$ shining on the total pixel area.

$$EQE = \frac{e_{gen}}{N_{ph}} \tag{4.2.10.1}$$

EQE can be at most unity and is wavelength-dependent. It is to be distinguished, however, from Internal Quantum Efficiency (IQE) where $N_{ph}$ becomes $\alpha N_{ph}$, the number of absorbed photons.

$$IQE = \frac{e_{gen}}{\alpha N_{ph}} \tag{4.2.10.2}$$

where $\alpha$ is smaller than unity and therefore IQE>EQE. IQE removes the importance of reflections on the cover glass and on various material interfaces but it is not a good way to estimate the real performance of the sensor.

EQE can be obtained through the repetition of the PTC measurements for different monochromatic wavelengths whose photon flux (photons per unit area) is known. A reference photodiode, whose EQE is known, is therefore needed for the measurement. For each wavelength, the photodiode current $I_{pd}$, obtained from the slope of the sensitivity plot in the light is computed. Then, the dark current $I_D$ is also computed as described in 4.2.9. The difference between these two currents is the actual photo-generated $I_{ph}$ as shown in (4.2.10.3).

$$I_{ph} = I_{pd} - I_D \tag{4.2.10.3}$$

Once $I_{ph}$ is converted into $e^-$/s, with the computed conversion gain $G$, the photo-generated current per unit area can be obtained since the pixels area $A_{pix}$ is known (18.5 x 18.5 µm$^2$ for most of the sensors):

$$I_{ph}[e^-/sm^2] = \frac{I_{ph}[e^-/s]}{A_{pix}} \tag{4.2.10.4}$$

Then since the EQE of the reference photodiode is known, its irradiance $E_D$ (expressed in W/m$^2$), corresponding to $I_D$, and $E_{pd}$, corresponding to $I_{pd}$, are also known. Both $I_D$ and $I_{ph}$ are measured twice close together in time in order average out temperature and temporal variations of the reference photodiode readout. The obtained irradiance $E_{ph}$ corresponding to the $I_{ph}$ of the reference photodiode can be converted into photon flux $P$. Knowing that the wavelength $\lambda$ of the light the energy of a single photon $E(\lambda)$ can be estimated from (4.2.10.5) and the number of incident photons per second per unit area (photon flux $P$) can be obtained by dividing the irradiance $E_{ph}$ by $E(\lambda)$ as in (4.2.10.6).

$$E(\lambda) = \frac{hc}{\lambda} \tag{4.2.10.5}$$

where $c$ is the speed of light and $h$ is Planck's constant.

$$P[N_{ph}/sm^2] = \frac{E_{ph}}{E(\lambda)} \tag{4.2.10.6}$$

EQE can then simply be computed as the ratio of (4.2.10.4) and (4.2.10.6), resulting in (4.2.10.1). EQE can then be plotted as a function of wavelength.

To obtain reliable measurements of absolute Quantum Efficiency, these were instead performed at the laboratory of IMEC [166] in Leuven, Belgium, in June 2016, by Chenghan Li and myself and with the help of the local researcher Vasyl Motsnyi. Dr. Federico Corradi also helped in writing the script for controlling the measurements.

The setup used is illustrated in Fig. 50. The Newport white light source used is Xenon-based. Its light passes through a Newport monochromator, which, through a series of mirrors and shutters, only lets through a specific wavelength. The minimum wavelength-step resolution of the monochromator is 10 nm and its selected wavelength is controlled externally from a PC running Labview (not shown in Fig. 50). The sensor's board is connected through a special plate to the output of the integrating sphere so that homogeneous light falls onto the sensing array (identically to the setup illustrated in Appendix I). The output of the sensor is monitored and recorded by cAER, a C-based version of jAER [167]. cAER also varies the exposure of the sensor and sends commands to Labview about wavelength sweep and shutter opening or closing.

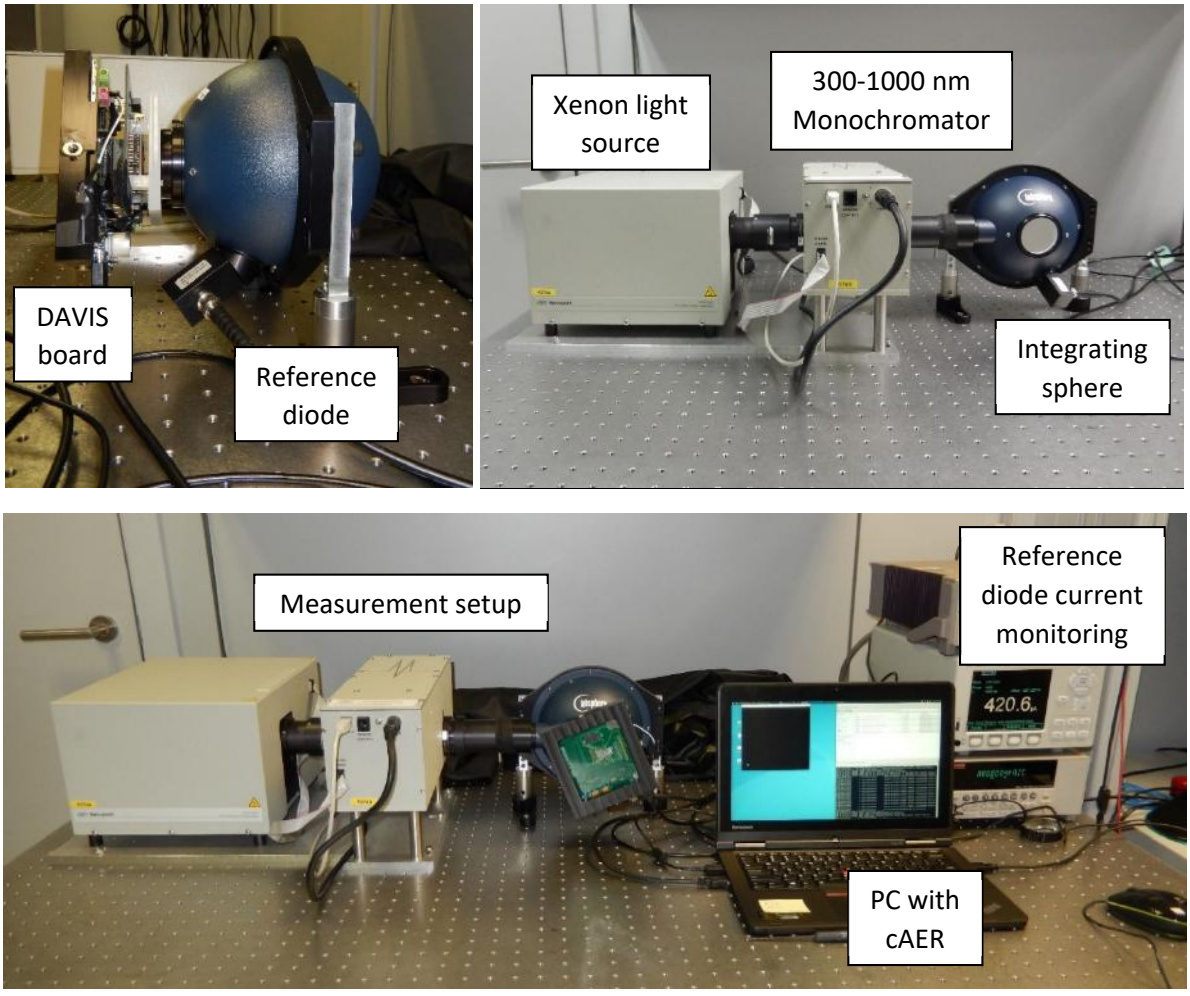


*Fig. 50 Setup, by Vasyl Motsnyi, of light source, monochromator and integrating sphere facing the board containing the chip under test at IMEC, Leuven, Belgium.*

The results for the sensitivity were obtained for every DAVIS sensor available at each swept wavelength of the light source, so as to compute the final QE of the sensor (External Quantum

Efficiency). The only difference at every wavelength is the slope of the sensitivity plot, proportional to the photo-generated current $I_{ph}$, as absorption of Silicon is dependent on wavelength and junction depth. Comparing $i_{ph}$ with the photon flux detected by a reference photodiode, the results described and shown in Fig. 31 were obtained to show the comparison between the reference design of [32] and SDAVIS192.

## 4.3. Notes on power consumption

Power consumption of the vision sensor only, excluding therefore the large power consumption of the mother board with FPGA and daughter board supporting it, can be estimated for separately for both the analog 1.8 V and 3.3 V (A1.8 and A3.3) and digital 1.8 V and 3.3 V (D1.8 and D3.3) power supplies. This can be achieved by putting the Keithley 236 source-measure unit in series with each supply separately (through its terminals TRIAX1+ and BNC) in V-I mode with 0 V. This allows to measure a current consumption $I$ of up to 100 mA and therefore, by knowing the voltage of the supply $V$ (either 1.8 V or 3.3 V), the total power consumption $P$ through equation (4.3.1).

$$P = VI \tag{4.3.1}$$

The power consumption can be further divided into static and dynamic, depending on if APS and DVS are enabled or not. DVS power consumption depends on activity and APS on frame-rate. Both depend on the signal quality chosen, given by the bias settings, and on the scene illumination. For these reasons, power consumption must be estimated under a variety of settings and conditions. Recordings of the sensor output must be obtained, along with the used bias settings, to verify power consumption claims. Finally, the low-power bias settings, used to obtain the lowest power consumption, should allow the vision sensor to still give a mildly decent and humanly recognizable image. This is of course subject to personal judgment, and this is the reason why recordings should be obtained for future critical analyses.

Power consumption measurements were performed for SDAVIS192. Analyses were performed for either APS or DVS active. The reason for this is that strong coupling exists between APS and DVS. The capture of a single frame creates a flash of DVS events. Although DVS events coinciding with a frame can be removed in jAER, this creates a burst in power when both event- and frame-capture are active. The ideal situation would be to alternate DVS and APS. An example for this could be to first calibrate the sensor with APS and DVS turned off, and then start recording with DVS and no APS for small temporal changes (calcium imaging).

The measurements presented in Fig. 51 and Fig. 52 show the power consumption of SDAVIS192 tuned to have very good APS performance.

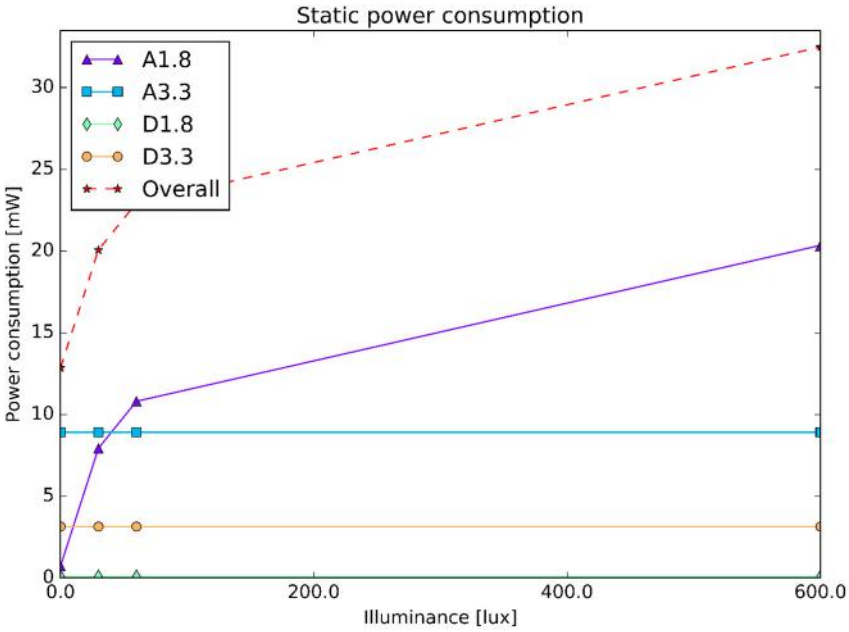


*Fig. 51 Static power consumption (APS and DVS turned off) of SDAVIS192, tuned to have good APS performance, as a function of illuminance.*

Fig. 51 shows the static power consumption (APS and DVS are turned off) of the analog and digital 1.8 V and 3.3 V power supplies (A1.8, A3.3 and D1.8, D3.3) as a function of illumination. As can be seen, A1.8 has a strong dependency on illumination which brings the static power consumption from 12.86 mW to 32.48 mW. This non-negligible 19.62 mW increase in power consumption from 0.1 lux to 600 lux of ambient illumination points out the need to state the illumination conditions of the power consumption measurement. The higher the current in the photodiodes or parasitic nodes, the higher the current consumption.

Measuring D1.8 proved to be very hard at all settings. The reason is that although its sub-milliWatt power consumption, D1.8 requires a high current at initialization. This current happens to be larger than 100 mA, the compliance of the Keithley 236 source-measure unit connected in series with the power supply in V-I mode and 0 V to measure current. The Keithley therefore limits the current at start-up and 9/10 times does not allow the sensor's bias generator to initialize and the sensor to function correctly. Repeated attempts allowed to sporadically start the sensor correctly and measure D1.8. Anytime a new configuration chain would be sent to the chip the same problem would happen (for example changing a bias or enabling/disabling APS or DVS).

Fig. 52 shows the total power consumption for the same biases of all power supplies as a function of illuminance and frame rate. It can be seen that an increase in frame rate of 30 times does influence power consumption visibly as more frames are sampled and transmitted off-chip. A3.3 is also supposed to increase with light intensity (more current drawn by the photodiode) but the fact that the supply does not increase visibly in Fig. 52 probably signifies that the increase in power consumption is minimal over the baseline for this bias setting. For other bias settings A3.3 was noted increase slightly with illuminance but just 1.5 mW at most for a change in two decades of illumination.

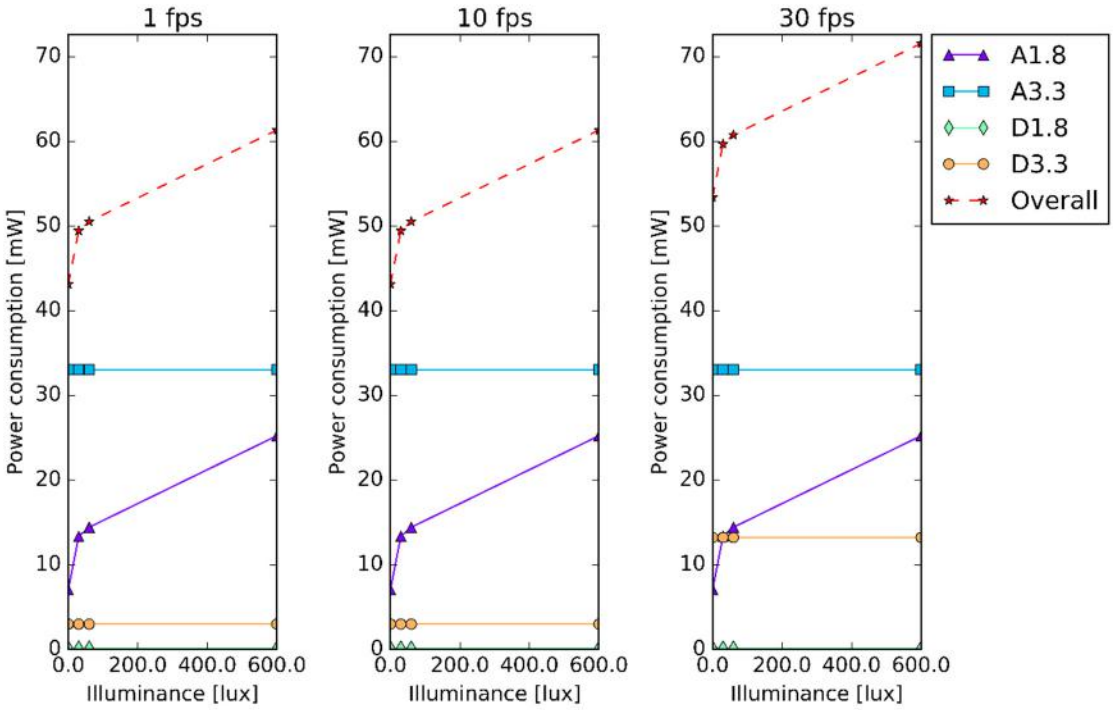


*Fig. 52 Total power consumption of SDAVIS192, tuned to have good APS performance, as a function of illuminance and frame rate.*

The overall power consumption, at these bias settings for good APS, therefore ranges from 43.12 mW to 71.53 mW. Illumination and frame-rate play an important role in this almost 30 mW increase. The power consumption of APS is higher than the one of DAVIS240C stated in [32] for two reasons. The first is that in [32] the APS was barely working in order to minimize biases and power consumption. The APS sensor would, however, never be used with such settings. The second reason is that SDAVIS192 has an internal single-slope ADC which was designed by Dr. Raphael Berner and implemented by Chenghan Li, Dr. Sim Bamford and Prof. Dr. Tobi Delbruck. This significantly increases the chip power consumption but reduces the overall board power consumption. For the DAVIS240C board, the external THS1030 10-bit high-speed ADC used dissipates at least 100 mW of power [165].

The same measurements were performed for the DVS enabled and the APS turned off for two bias settings. The first bias setting, just like in [32], is optimized for power consumption and the DVS is working with minimum performance (this is illustrated in Fig. 53 and Fig. 54). The second bias setting used is the one that is normally used for the high-contrast operation (this is illustrated in Fig. 55 and Fig. 56). As regards the first bias setting for DVS of Fig. 53, static power consumption is again dependent on illumination as more current drawn by the photodiode increases the power consumption of A3.3. A1.8 also peaks but at medium illuminations. The few numbers of point in this curve, however, do not allow to explain this behavior fully. The large increase in A3.3 of about 10 mW, here and in the following plots, for 208*192-192*4 = 39’168 pixels in total, means a power consumption of about 0.255 µW per pixel. For the 3.3 V supply, this signifies an increase in current of 77.3 nA per pixel. This number seems a little too large and might indicate that another source of

current consumption is light dependent, perhaps some parasitic photodiode in the 3.3 V circuitry whose behavior depends on bias settings. This was not present, as a matter of fact, in Fig. 51.

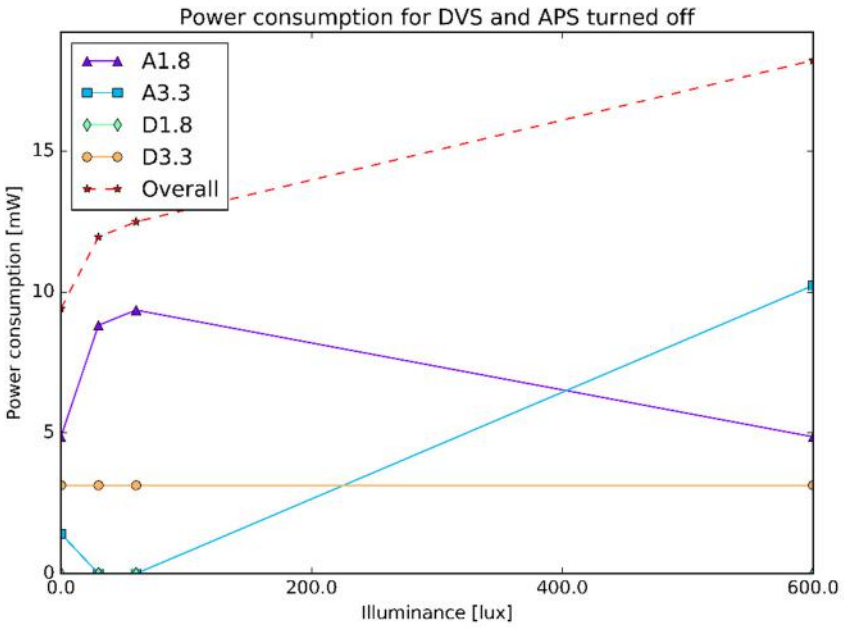


*Fig. 53 Static power consumption (APS and DVS turned off) of SDAVIS192, tuned to have minimum DVS power consumption, as a function of illuminance.*

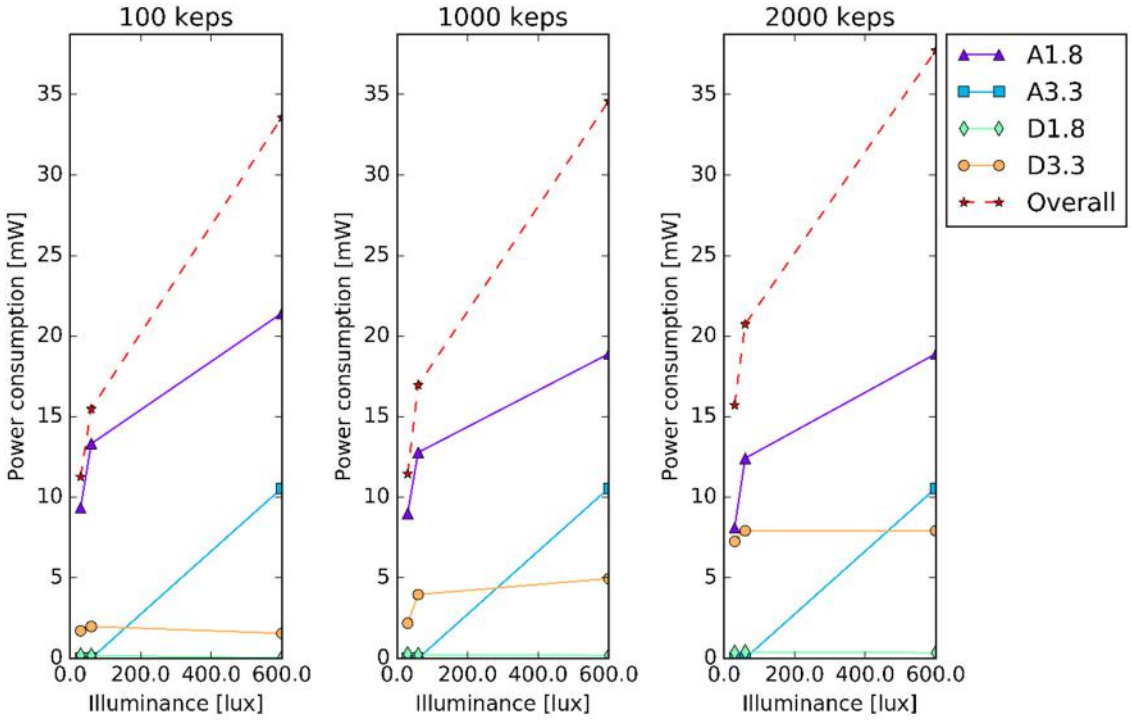


*Fig. 54 Total power consumption of SDAVIS192, tuned to have minimal DVS power consumption, as a function of illuminance and DVS activity.*

As can be seen in Fig. 54, all power consumptions seem to increase with illumination but D1.8. The increase in D1.8 is just too small though to be noticed as only being about 1% of the 70 μW unstable baseline. As more data is sent over to the FPGA through 3.3 V lines, A3.3 also increases in power consumption with increasing activity. All other supplies are approximately constant with respect to activity. Depending on DVS activity and illuminance, the chip's total power consumption, for this specific bias settings for minimum power consumption, ranges from 11.27 mW to 37.72 mW.

An interesting fact to note is that DVS is fully functional even by disconnecting A3.3 and D3.3. This should theoretically not be the case as A3.3 is needed for the DVS front-end and D3.3 for communicating to the FPGA. Furthermore, both 3.3 V supplies are necessary to the pads. The probable cause of this are diode junctions in the pads, which turn on whenever any of the 3.3 V supplies drop below 1.8 V. This prevents them from dropping any further. If the front-end is connected to 1.8 V, DVS can still work fine but APS does not have enough headroom to function. As regards D3.3, as long as the logical 1 is above 3.3/2 = 1.65 V, then the communication to the FPGA works as the digital 1.8 V supply suffices. If 3.3 V supplies would be completely disconnected, sacrificing APS for the sake of power consumption, the latter would drop between 9.56 mW and 19.24 mW for these specific set of biases. This could mean up to a factor of two in power saving. This power consumption is still higher than the 5 mW to 14 mW stated in [32] for low and high activity respectively. The analysis of [32], however, lacks details, bias settings and illumination conditions. Furthermore, SDAVIS192 is bound to have a higher DVS power consumption because of the strong shifted-source bias driving the pre-amplifier stage and provided by strongly biased OTAs present on every row of the array (192 in total).

If the DVS performance is improved for maximum contrast sensitivity (with APS turned off), then the graphs of Fig. 55 and Fig. 56 are obtained. A3.3 static power consumption seems independent of light in this case as probably the background consumption is already consistent.

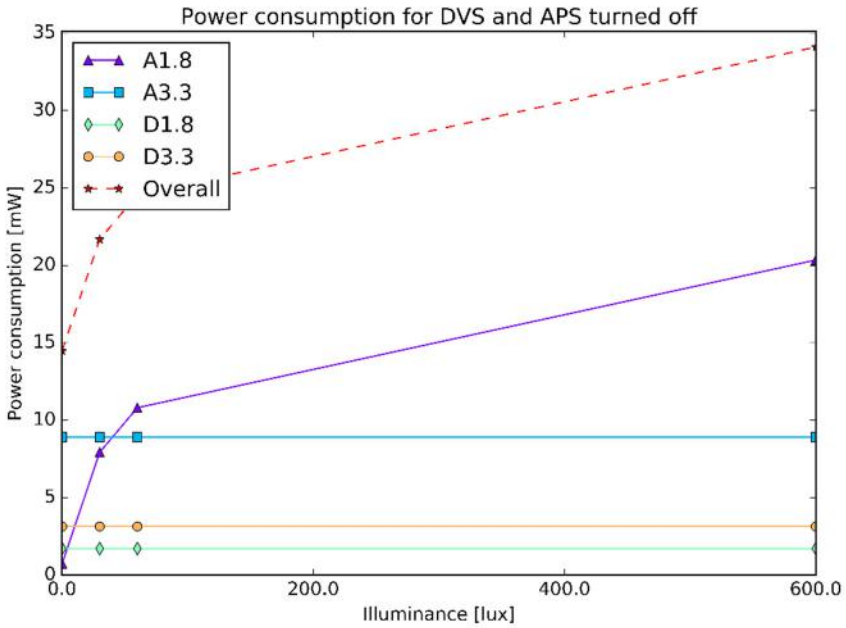


*Fig. 55 Static power consumption (APS and DVS turned off) of SDAVIS192, tuned to have the best DVS performance, as a function of illuminance.*

As regards the total power consumption, it can be seen in Fig. 56 that changes as a function of activity and illuminance are also modest for all supplies but A1.8, where an increase of 5 mW is recorded over 2 decades of illumination.

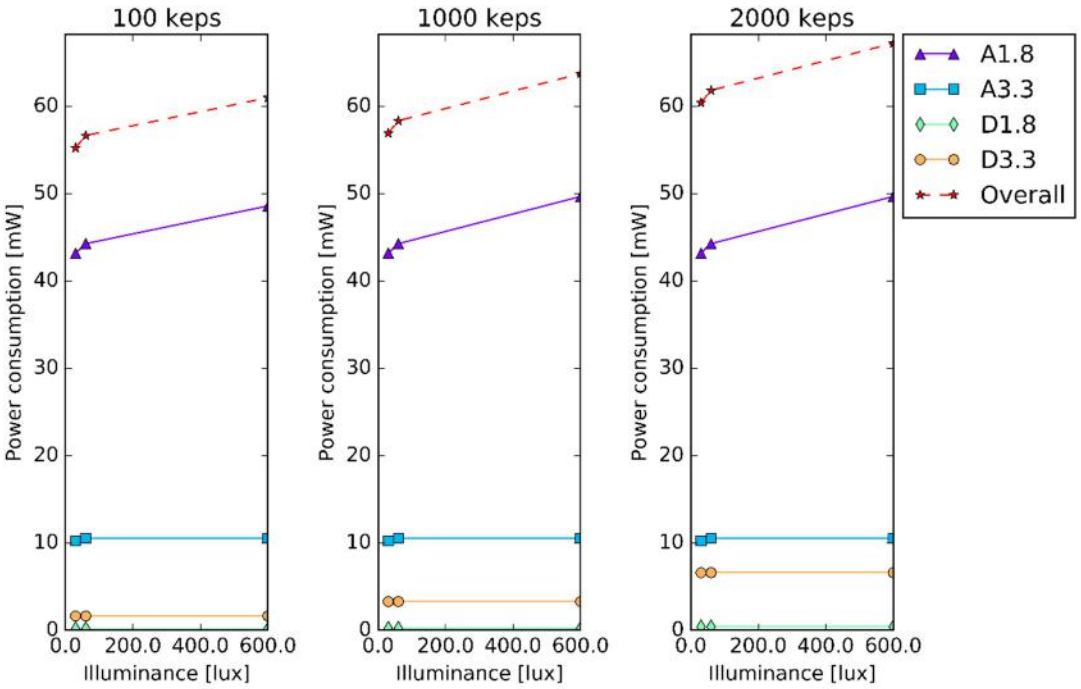


*Fig. 56 Total power consumption of SDAVIS192, tuned to have the best DVS performance, as a function of illuminance and DVS activity.*

For great DVS performance and APS turned off, the power consumptions range 55.59 mW to 67.24 mW, which is two to five times more than with respect to the previous bias settings for low power. If A3.3 and D3.3 are disconnected, this range drops from 43.38 mW to 50.08 mW.

For all above power consumption measurements recordings were obtained to show the quality of the sensor's output for such bias settings (also saved) and are available on request.

## 4.4. Automated Operating Region Control

### 4.4.1. Steady-state analysis

To estimate the operating range of the pre-amplifier, the mean pre-amplifier output VobsA (referring to Fig. 21) was observed under different illumination and Vref conditions. First of all, however, having access to both the chosen VrefA (the input to the shifted-source bias OTA) and Vref (the actual shifted-source bias voltage value), the linearity of the shifted-source OTA was investigated. This can be observed in Fig. 57.

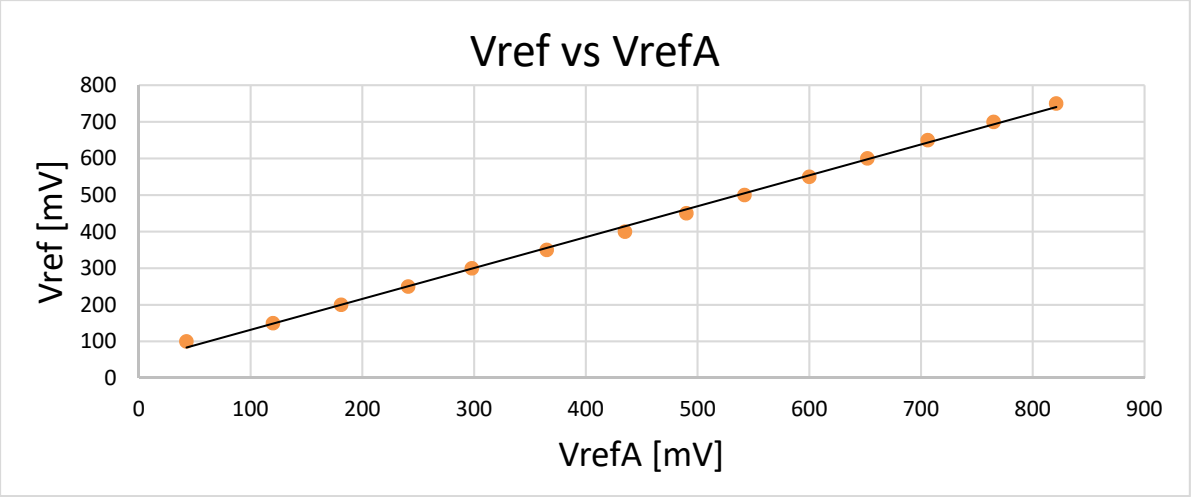


*Fig. 57 Shifted-source OTA transfer characteristics.*

As can be seen, the OTA behaves reliably above 100 mV when the transistor MN11 of Fig. 21 reaches saturation. The bias generator current set through jAER or cAER, resulting in the voltage VrefA, was also noted down. As a matter of fact, knowing this current allows to write the look-up table in the FPGA that will set the pre-amplifier shifted-source bias in the correct range.

To characterize the mean pre-amplifier output, the set-up of Fig. 58 was put in place. This consists of a 9.34 klux, 150 cd (22.34 lumen), 20-25°, 460-470 nm, 3.2-3.5 V super-bright white 5 mm LED powered through a 220 Ω resistor by a 5V supply, illuminating the optical surface of the chip (the photodiode array) through a 5 mm LED-holder pinhole sealed in black tape. The current through the LED is about 20 mA. All 4 important nodes are observed with the oscilloscope: Vref, VrefA, VobsA and Vavg. A set of Neutral Density Filters (NDF) was then slid between the LED-holder and the chip, to attenuate the known illuminance and therefore change the brightness of the LED. The available NDF used attenuated the light by $10^{0.5}$ (factor of 0.32), $10^{1}$ (factor of 10), $10^{2}$ (factor of 100), and $10^{3}$ (factor of 1000).

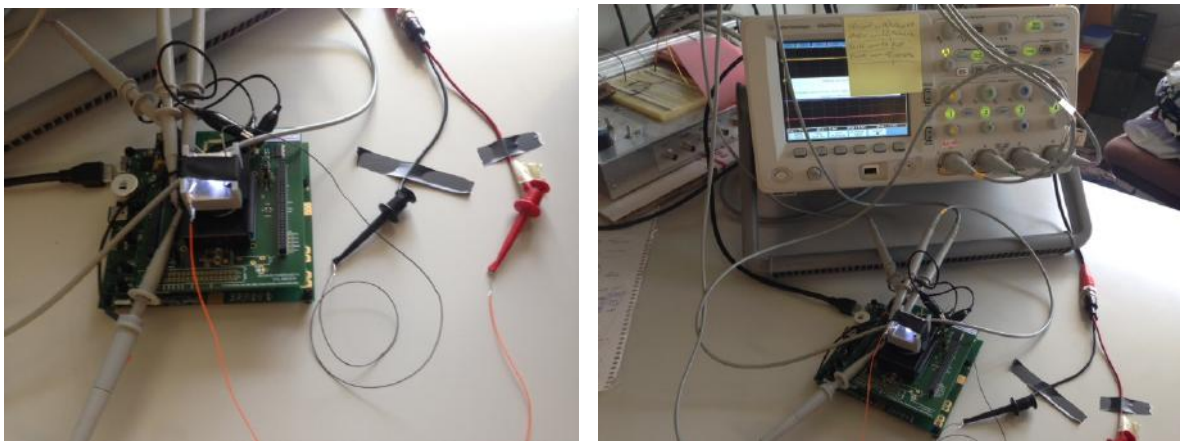

*Fig. 58 Set-up to characterize the pre-amplifier of SDAVIS192.*

The first node that was studied was Vavg. Referring to Fig. 21, Vavg is the gate of the mean pre-amplifier circuit. It is also the drain and gate voltage of the in-pixel diode-connected transistor MN8

and it represents the mean Vpr, the gate voltage of MN9. While sweeping Vref, Vavg was observed and the result is reported in Fig. 59.

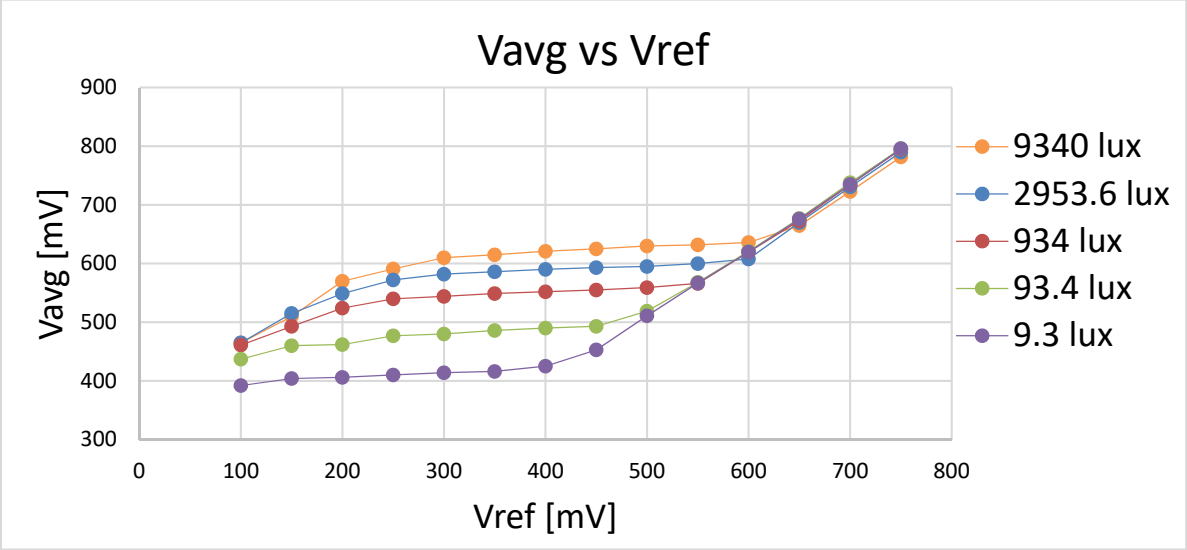


*Fig. 59 Vavg changing with respect to Vref, for 5 different NDF attenuations.*

Three regions in the graph of Fig. 59 can be identified. In the central region, where the slope of the curves is zero, Vref has no effect on Vavg. As a matter of fact, as Vref rises, the gate-to-source voltage (Vgs) of MN9 (Fig. 21) is reduced (as Vpr is only set by the lighting conditions and is therefore fixed), Ipa, the pre-amplifier's and mean pre-amplifier's current is also reduced and the Vgs of MN8 (which is the difference between Vavg and Vref) also drops. For Vref increasing by the same amount for MN8 too, Vavg stays fixed and the Vref change is compensated. When Vref rises up to Vpr (the current value of Vavg), the third region on the RHS of the graph begins. At this point the polarity of the transistor MN8 is reversed, as the gate (and drain) voltage drops below the source of the transistor. From this point onwards, Vavg just follows Vref exactly and transistor MN9 turns off. As for the LHS of the graph, when Vref drops below 250 mV, Vavg decreases. This is because the transistor MN8 gets out of saturation and the drain-to-source voltage (Vds) dependence of the current in the mean pre-amplifier circuit starts to show. At this point the mirroring operation begins to fail, as the different Vds of MN9 and MN8, due to their different drain connections, do not allow a proper current copying.

This can be better observed by plotting Vavg as a function of the Vgs of MN8 (which is the same as the Vgs of MN12), as done in Fig. 60. No matter the increase in Vgs of MN8, Vavg stays constant as Vref compensates (by decreasing) for the increased Vgs and therefore current Ipa. When the Vgs becomes zero (as Vref reaches Vpr and therefore Vavg), Vavg starts to be risen by the transistor MN8 flipped in polarity and now follows Vref and no longer Vpr (MN9 turns heavily off). The end of the curve reflects the decrease in Vavg due to the entrance in the triode region of MN8.

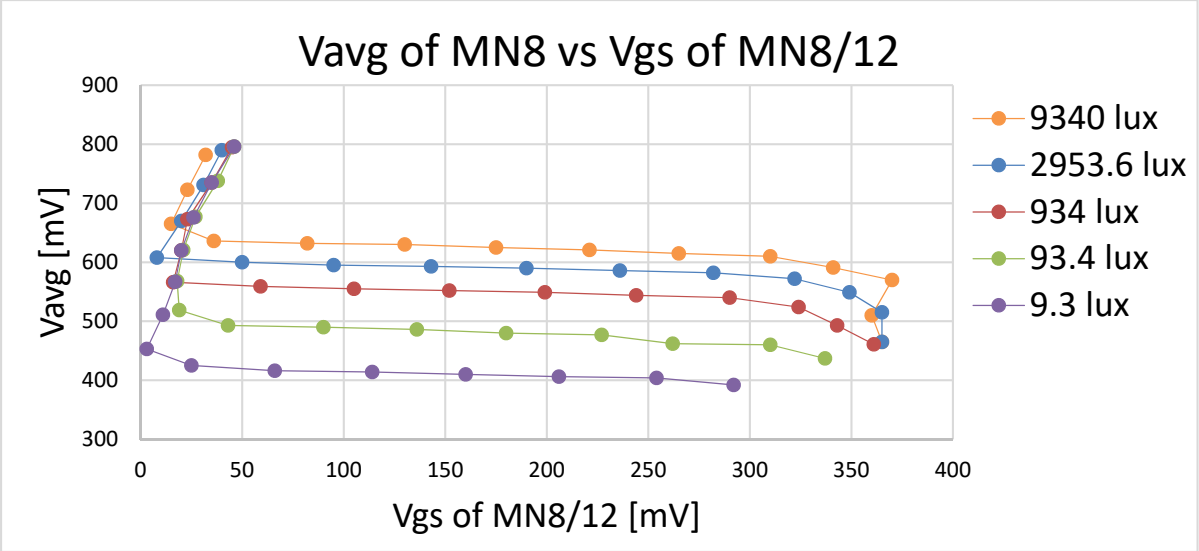


*Fig. 60 Vavg changing with respect to Vgs of MN8 and MN12, for 5 different NDF attenuations.*

The rise in Vref also affects the driving transistor of the pre-amplifier: MN12 (shown in Fig. 21). The nFET exits from the saturation region due to the drop in Vds of MN12 (which is the difference of VobsA and Vref), observed in Fig. 61 and Fig. 62 for Vref lower than 100-400 mV (depending on the illumination) and Vgs of MN12 higher than 300 mV respectively.

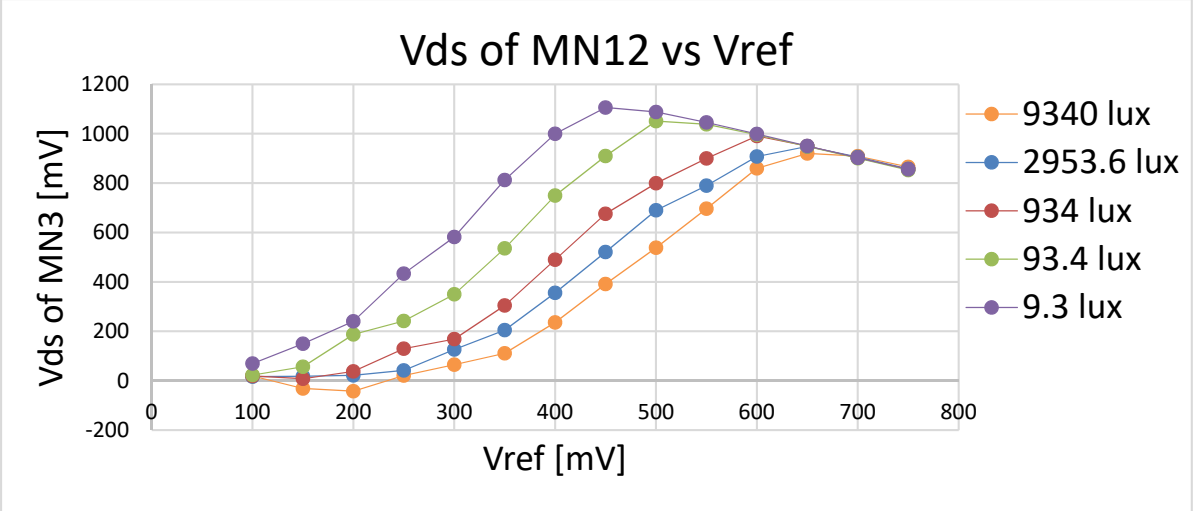


*Fig. 61 Vds of MN12 changing with respect to Vref, for 5 different NDF attenuations.*

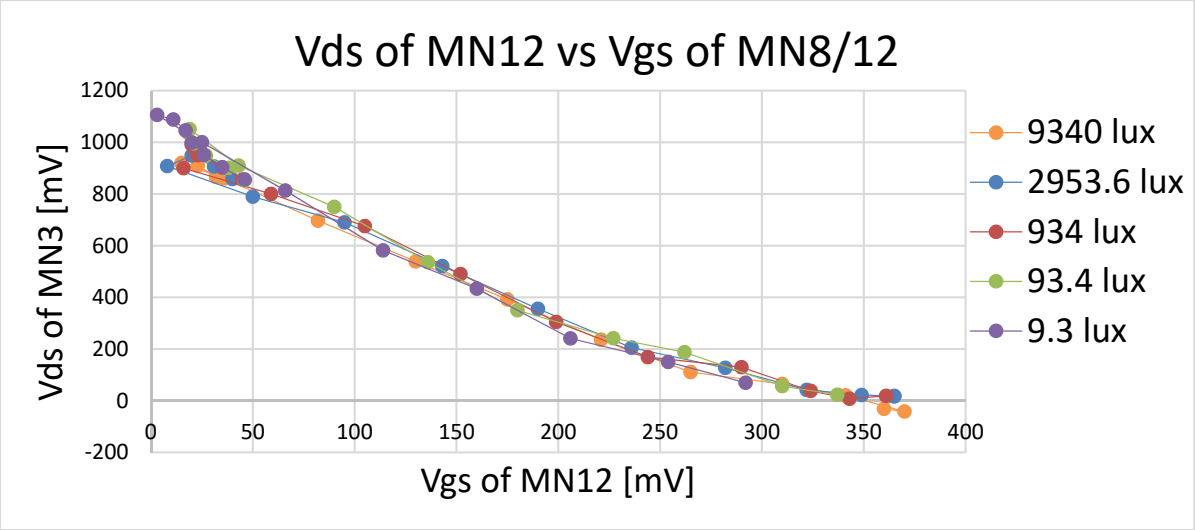


*Fig. 62 Vds of MN12 changing with respect to Vgs of MN8/12, for 5 different NDF attenuations.*

The effect of light on all of the previous curves can be explained as follows: referring to Fig. 21, for a higher light intensity (no NDF is the highest in this case) Vpd (Fig. 21) is discharged more, Vpr is then increased along with Vavg (which is an approximate copy of it). Increasing light intensity changes the range in which Vavg is following Vpr correctly and therefore the range in which the circuit is operating correctly.

As regards VobsA, the mean pre-amplifier's output, this also has a dependence on light, as it directly depend on Vavg. This can be observed in Fig. 63 and Fig. 64, where VobsA saturates at around 100 mV on the lower side (for low Vref, when MN12 has a large enough Vgs and small enough Vds to get out of saturation) and at 1.6 V on the higher side. It is important to note that the chip's measured A1.8 Vdd is 1.73 V, as provided by the supporting Printed Circuit Board (PCB), this therefore affects the presented measurements and results in a mismatch with the predictions.

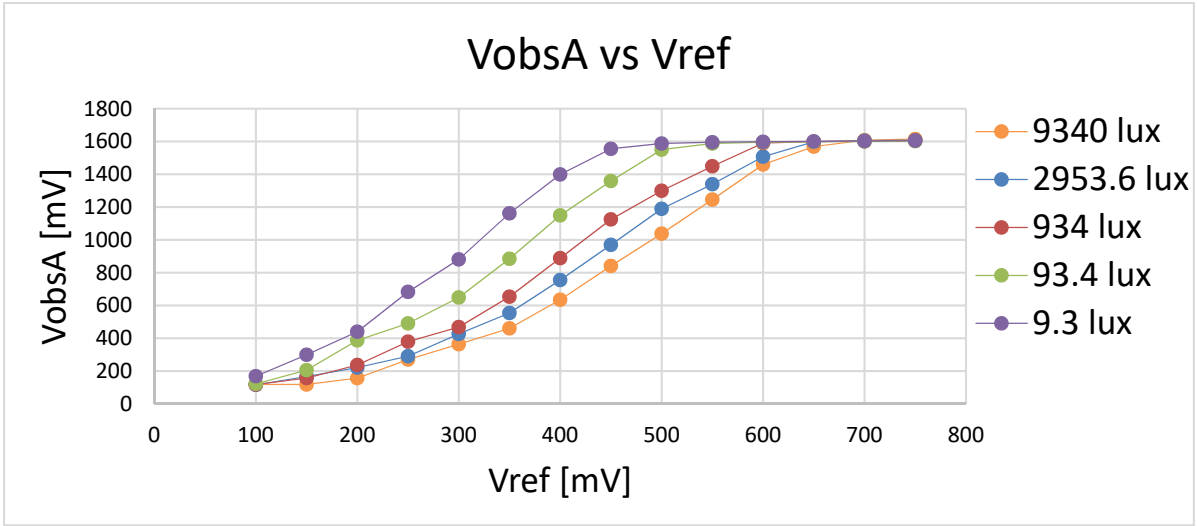


*Fig. 63 VobsA changing with respect to Vref, for 5 different NDF attenuations.*

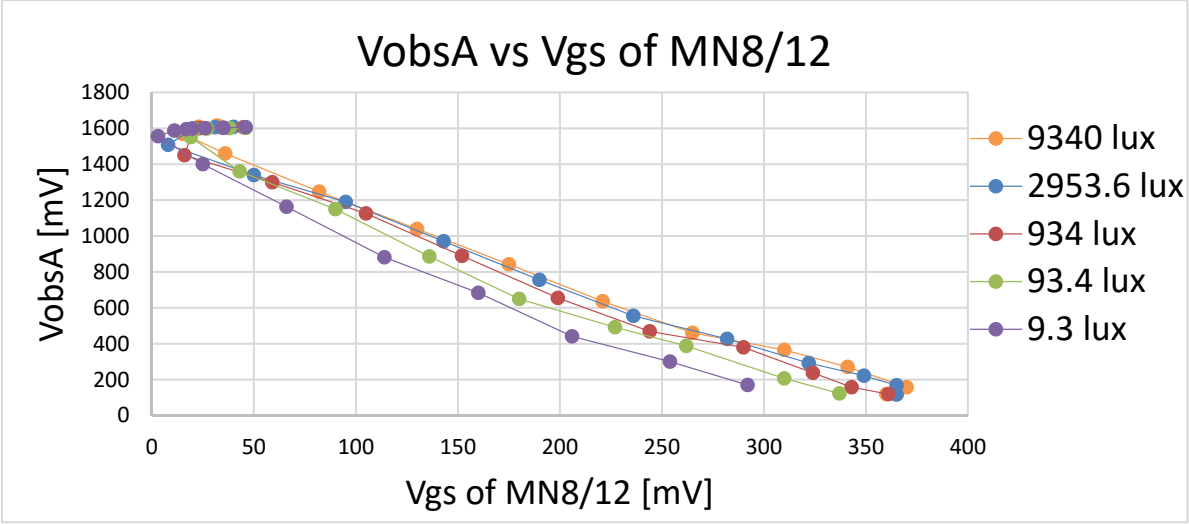


*Fig. 64 VobsA changing with respect to Vgs of MN8/12, for 5 different NDF attenuations.*

Being VobsA the output of MN12 (which has negative gain) whose input is Vavg, its response to light is inverted. A negative swing in Vavg corresponds to a positive swing and vice versa. Increasing light makes the operating region of the amplifier shift: the saturation on the high side occurs for higher values of Vref and on the low side for higher values of Vref. The opposite happens for lower light conditions. This also means that it is possible to keep the amplifier in range (without saturation) by shifting Vref depending on the illumination (and therefore current value of VobsA). Finally, Fig. 65 illustrates the direct dependence of VobsA on light. As can be seen, for a factor of 1000 of illumination change, VobsA can span 800 mV of voltage values within its operating range. For a wrongly-matching Vref, the pre-amplifier can easily saturate.

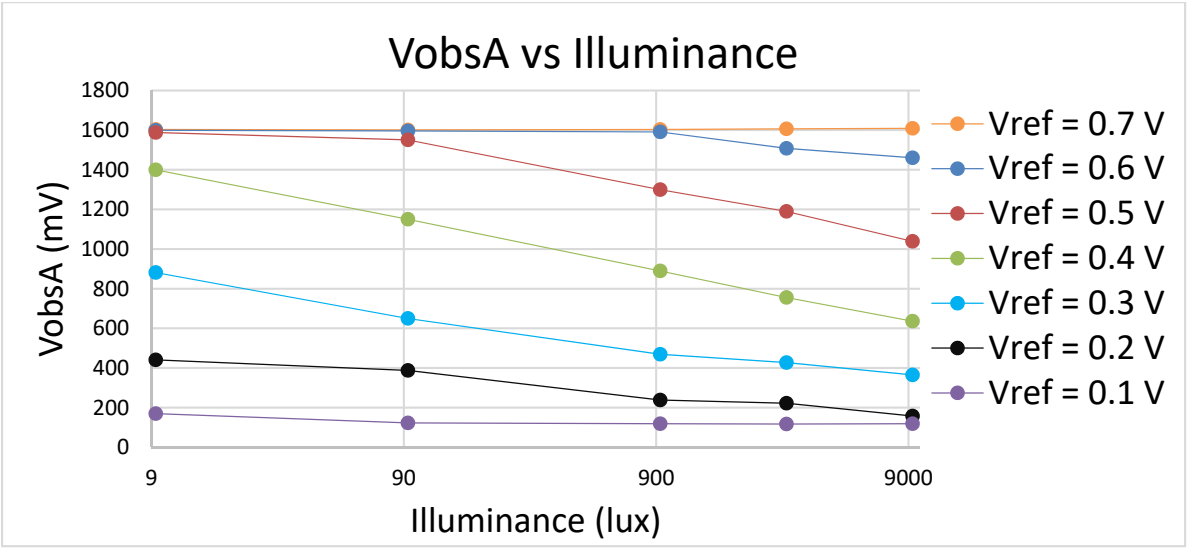


*Fig. 65 VobsA changing with respect to 5 different NDF attenuations, for different Vref.*

### 4.4.2. Test pixel transient analysis

The time response of the pre-amplifier was obtained thanks to the test pixel present at the bottom left of the pixel array. The test pixel provides access points to all SDAVIS192 internal nodes, buffered

through analog multiplexers AMUX0-2, which are biased by "LocalBufBn" and enabled by the configuration bit "UseAOut". The test pixel itself is enabled by setting configuration bit "ResetTestPixel" false. Fig. 66 illustrates the circuit response to a 5 Hz, 50% contrast, 1 klux light oscillation with the uniform intensity measurements setup of chapter 6. As can be seen, the photodiode node Vpd's response is below the resolution of the oscilloscope. The photoreceptor output instead responds with a 22.7 mV peak-to-peak oscillation, due to the high transconductance of MN7 (refer to Fig. 21).

Vpa, the direct output of the pre-amplifier was not observable as it was not possible to connect to this node due to layout reasons. Vbuff, the output of the following buffer stage (Fig. 21), corresponds to Vpa with a negative offset of roughly a threshold voltage. Fig. 66 shows the 7 possible outputs Vbuff for a logarithmic sweep in bias of Vref (obtained by sweeping the coarse current of Vref from 1 to 7 and by keeping the fine value constant to 20). Vbuff behaves as pointed out in section 4.4.1 for VobsA, just with a constant offset for the different biases, as its circuit, and therefore behavior, are the same of VobsA. For each Vref bias, the transistor MN9 of Fig. 21 is in a different operating region and the transistor's $\kappa$ changes. This gives a Vbuff to Vpr voltage gain ranging from 0.5 to 3.15 (slightly below expectation due to the source follower's gain less than unity) and a saturation of Vbuff for these bias settings and illumination on the low-end of 400 mV and on the high-end of about 1.66 V. A small ripple on every waveform is due to back-coupling of the ON and OFF spikes generated in the comparators' stage, though it is insufficient to generate further spikes in a positive feedback fashion.

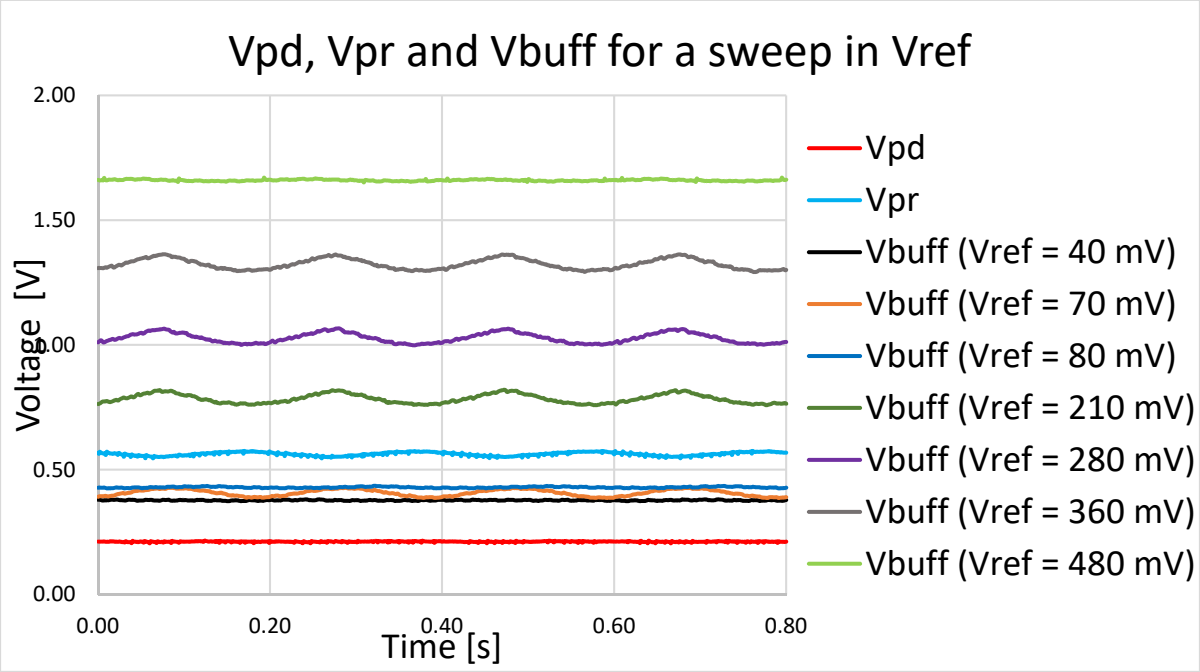


*Fig. 66 Oscilloscope trace of photodiode voltage Vpd, output of photoreceptor Vpr and all possible pre-amplifier buffered outputs Vbuff for a logarithmic sweep in pre-amplifier source bias Vref.*

The Vbuff to Vpr AC gain plot as a function of illuminance is visible in Fig. 67. In the region 200-400 mV the gain is about constant. In an average scene, in fact, the sensor works well with a 300 mV Vref. The bump at 0.07 V is not explained.

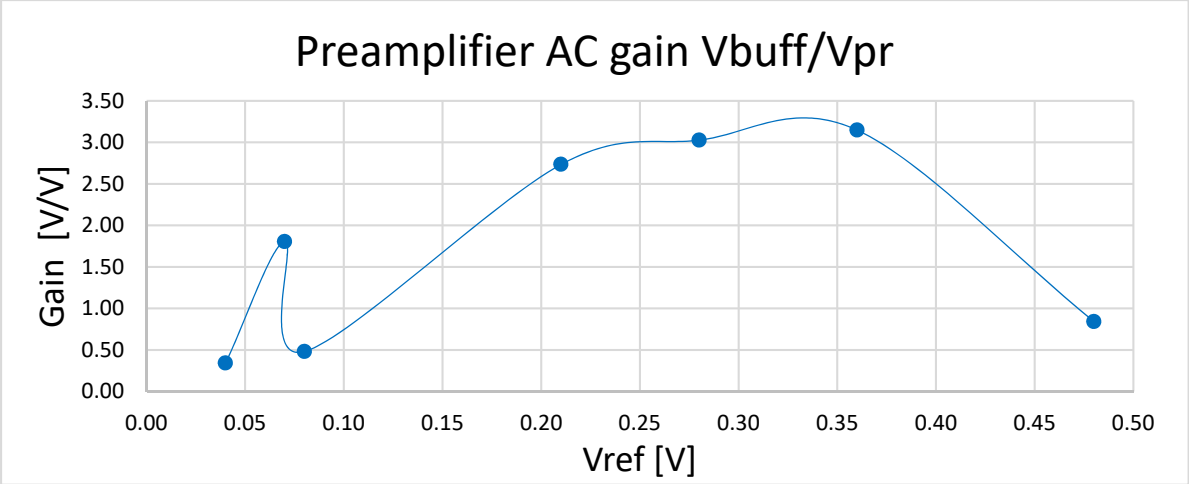


*Fig. 67 Plot of Vbuff/Vpr AC gain as a function of Vref bias.*

### 4.4.3. Finite State Machine

Having obtained all available characterizations of the pre-amplifier, the FPGA Finite State Machine (FSM) which can change Vref depending on the value of VobsA can be defined. This simple FSM is illustrated in Fig. 68. On Reset, the FSM is in *Idle* where the Vref change flag is set to be inactive. If enabled, the FSM proceeds to 4 consecutive identical stages called *ADCsample1-4* which sum the four samples of VobsD. These are obtained at a selectable time distance of up to a few seconds, set by a counter. The counter overflow *COVF* signal notifies the FSM to proceed to the next stage.

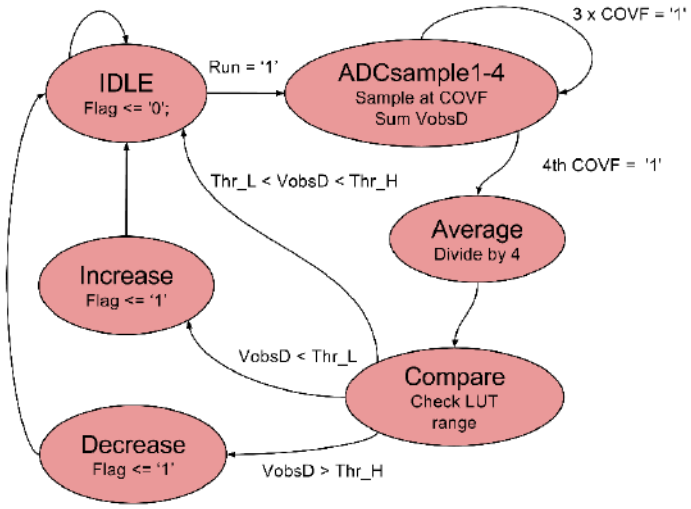


*Fig. 68 Finite State Machine (FSM) of SDAVIS192 operating region control.*

Once all four samples are summed and stored the average is computed in the state *Average*. The average VobsD value is compared with the Look-Up Table (LUT) which contains the high and low thresholds *Thr_H* and *Thr_L* which set the acceptance criterion of VobsD. If the criterion is met, the FSM returns to *Idle*, otherwise the FSM takes the appropriate response to decrease or increase the Vref bias if the VobsD value exceeds any of the two thresholds. This is done in the states *Increase* and *Decrease* where a flag signaling a bias change is set. This way it is possible to know which events are generated by noise in the transitions from one source bias level to another and to filter them out.

From experience, in an analog automated feedback and gain control [168], it is difficult to observe what state the system is in, and activity and adaptation might affect each other leading to unstable states or self-crosstalk.

## 4.5. APS and DVS coupling problem

As mentioned, DVS does not function properly while APS is enabled. All DVS pixels flash together saturating the bandwidth. To investigate this behavior, a constant light source of 1 klux was held on the array and an exposure of 1 ms was set in Global Shutter mode (GS). Using to the test pixel to observe all internal nodes, huge 100 ns transient oscillations of up to 0.8-1 V were observed on all internal nodes of the circuit (Vpd, Vpr and Vpa). The transients were also observed in the analog power supplies A1.8 and A3.3 as illustrated in Fig. 69. Since the photodiode has no connection to any Vdd but only to the grounded substrate, the transients are present in the ground of the chip and board and couple into all DVS transistors and the general power supply. The huge swings caused in Vpd (compared to the miniscule true signal) generate plenty of events, breaking down the DVS functionality. Using external A1.8 and A3.3 only reduces oscillations amplitudes to two thirds of the original value. To solve the problem, stronger voltage regulators must be put in place as well as very large bypass capacitors. The current bypass capacitors present on the PCB and inside the chip itself are just not sufficient.

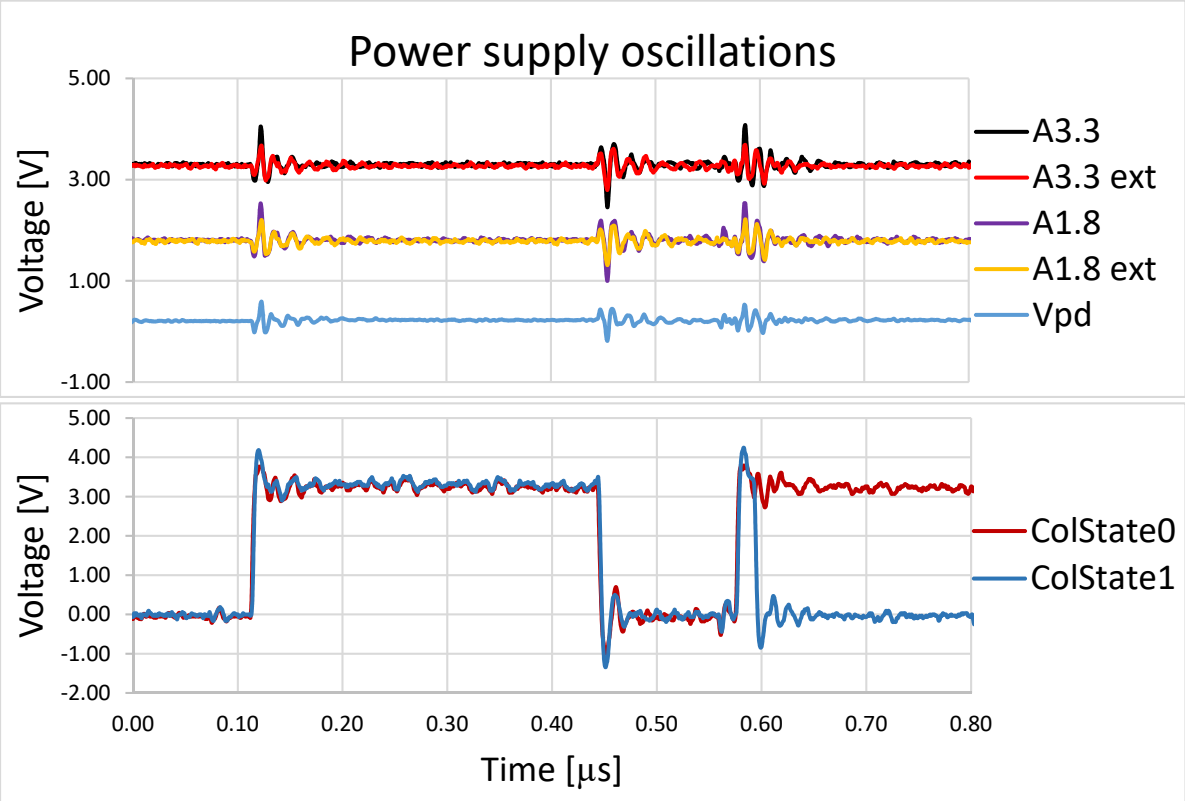


*Fig. 69 Oscilloscope traces of power supply and ground oscillations due to coupling with the column reset signal ColRes in Global Shutter mode.*

To understand where the coupling is coming from, it is sufficient to take a look at the APS signals. The APS Differential Double Sampling (DDS) which involves the two frame reads (reset and signal) is governed by an on-chip state machine. This bases its decisions according to two external signals coming from the FPGA: "ColState0" and "ColState1". When both of these signals are 0, the FSM is in its *Idle* state and frame exposure can take place. When the signal combination is 01 or 10, either the reset or the signal read respectively takes place. For the combination 11, the "ColRes" pixel reset signal (marked CR in Fig. 21) is activated and the photodiode is reset. Whenever ColRes switches active or inactive due to the ColStat0/ColState1 combinations, ground coupling appears, as Fig. 69 shows. Reset of the pixel happens column-wise in Rolling Shutter (once per column within a reset read) and concurrently for all pixels in Global Shutter during the reset read plus one time for each column right before its reset read (to improve signal quality). Unfortunately, there is no direct access to ColRes, even in the test pixel. Its behavior is only deducible from ColState0 and ColState1. It is possible that ColRes, a 3.3 V signal, strongly couples into analog ground GndA since it is runs on metal 4 along the pixel and it perfectly overlaps GndA on metal 2 all the way. The total pixel array capacitance between ColRes and GndA is 192 x 192 x $2.65 \times 10^{-16}$ F = 9.77 pF. The number was obtained through parasitic extraction at design stage and it was thought to be acceptable. The reason for this the particular arrangement in which GndA completely covers ColRes is due to the careful layout of the pixel in which the idea was to shield the digital signal from all other sensitive nets with a ground cover. This could however be the source of the problem. Another possible explanation is that ColRes simply takes too much power in resetting the array both in Global Shutter and in Rolling Shutter modes and this makes the overall power drop.

## 4.6. Application in neural imaging

To target an application for SDAVIS192, current work is being done to use it in the context of neural imaging. Dr. Yang's paper [37] made a similar attempt with a recording of calcium imaging playing on a computer screen, but with therefore an unrealistic scenario. A frame of the video played is shown in Fig. 70.

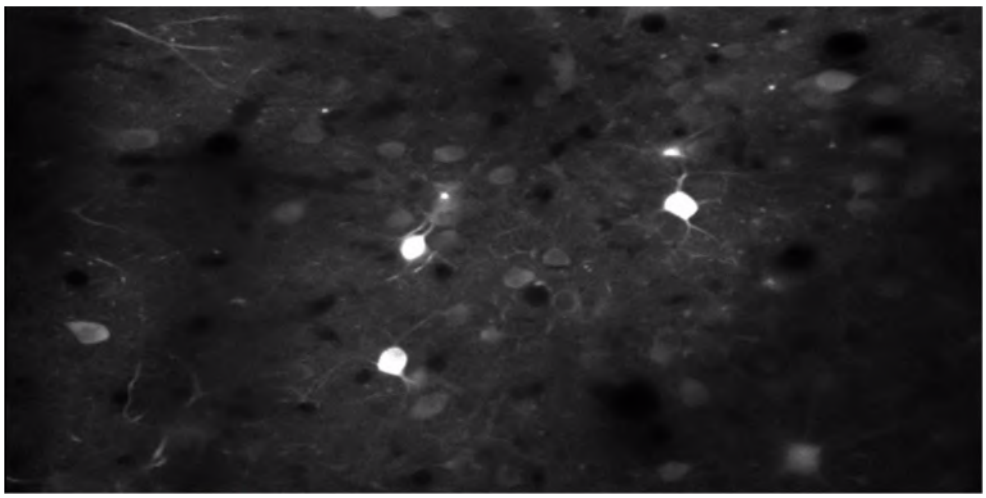

*Fig. 70 GCaMP (a genetically encoded calcium indicator) blinking dependently on activity, observed with 2-photon microscopy (photo from Andreas Keller, INI).*

Together with the help of the PhD student Gemma Taverni, we initially worked under the supervision of Dr. Ariel Gilad (a postdoc of Prof. Dr. Fritjof Helmchen) at University of Zürich Brain Research Institute (HIFO). The aim was to try to substitute the SDAVIS192 sensor to a conventional camera setup for in-vivo two-photon imaging. The setup used for the imaging is shown in Fig. 71. It consists of a double-objective lens system focusing a blue excitation light source (LED) onto the thinned skull of a genetically engineered mouse. Through the thinned skull, it is possible to see the brain and its blood vessels. The mouse is held down by a screw and housed in a toilet paper roll. The mouse has a water feeder and a whisker stimulator on the side with which various tasks can be performed. Depending on the action, different pathways become active: reward (water given) or fear (white noise burst). With the particular setup, layer 2 and 3 were being activated. As neural activity takes place, the calcium released by synaptic activity activates the gene-encoded calcium fluorescent indicator. If the indicator captures, within a short time, two photons, it emits green light. A mirror sends up to the sensor the green fluorescent light, filtered by a very narrow-bandwidth green filter.

The problem with the setup, which made the measurements impossible, was the inability to focus on the brain surface of the mouse. The setup was kept slightly out of focus for HIFO's experiments, but this would kill the contrast needed to detect anything with SDAVIS192. We were not allowed to touch the focusing of the double objective (various people were using the same setup) and to get too close to the mouse head (because it would become nervous and agitated). For this reason we then switched to the more basic and safe setup of Fig. 72 to observe moving fluorescent beads. This work was a collaboration with PhD student Fabian Voigt, expert in optics, of the Prof. Dr. Urs Gerber group.

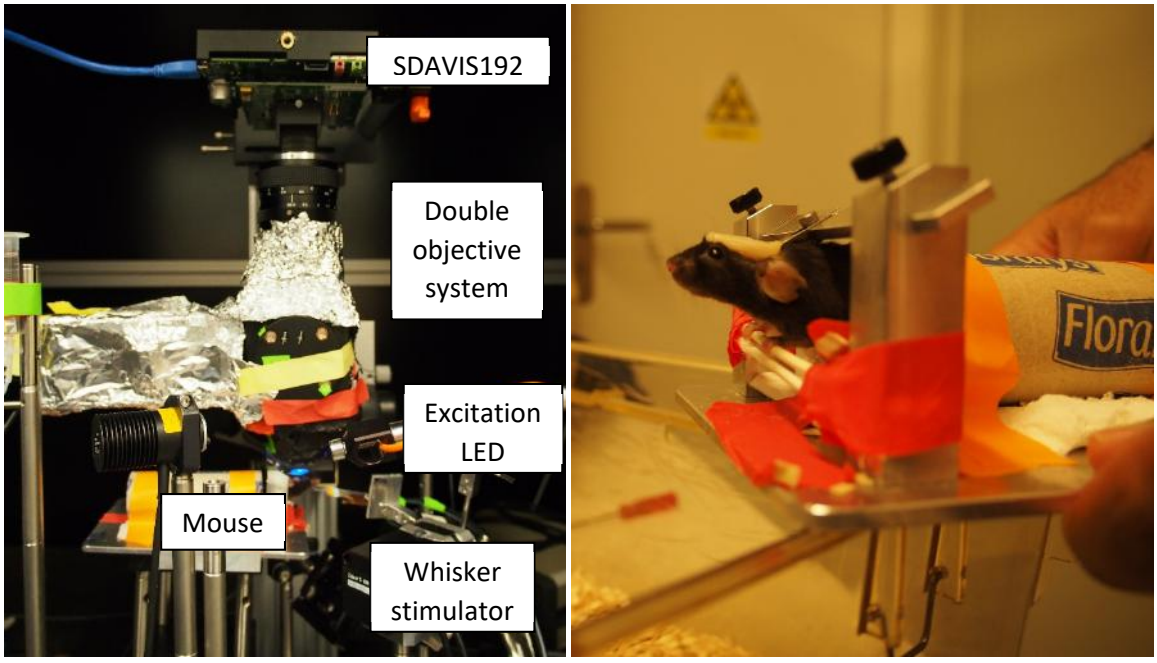


*Fig. 71 Setup for in-vivo calcium imaging.*

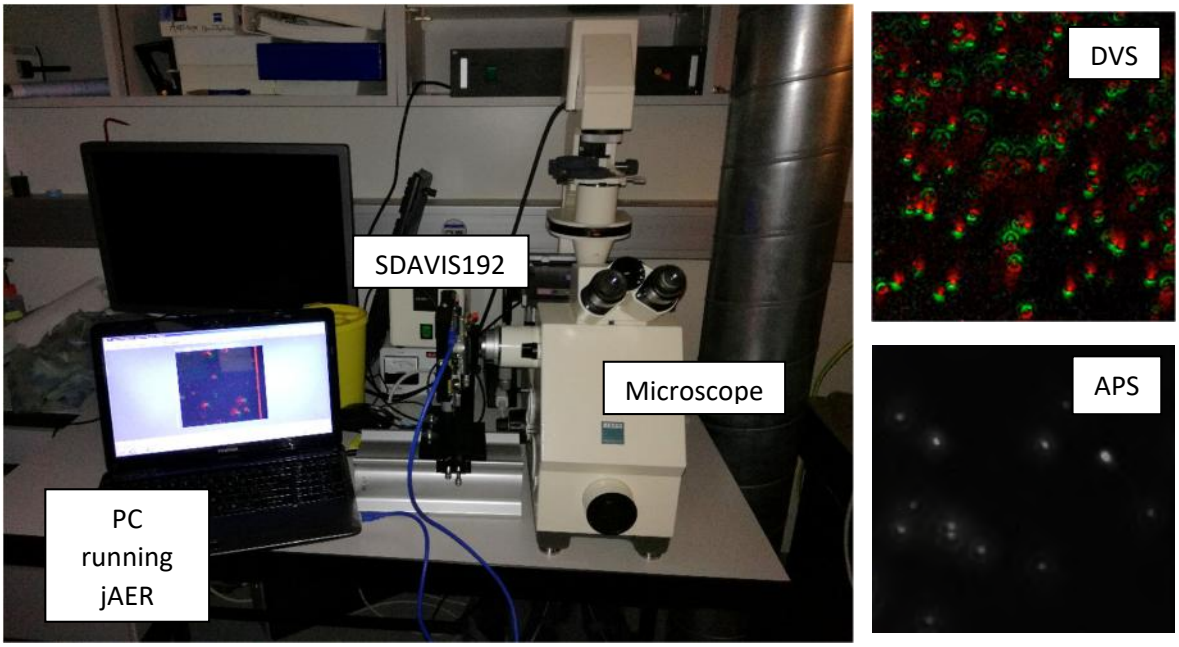


*Fig. 72 Left: Setup for fluorescent beads observation. Right: DVS and APS outputs of SDAVIS192.*

The setup of Fig. 72 consists of a microscope whose output port can be switched between camera sensor and visual output. The SDAVIS192 replaces the setup's camera and the live output can be seen in jAER and in the right side of Fig. 72. 1 µm wide fluorescent beads are moving because of Brownian noise in a water drop of solution 1:10. They move on the surface of the water, therefore some are at the bottom of the water drop and some are at the top. Depending on which one we focus, diffraction patterns (rings) appear for the unfocused beads. The beads have high contrast but they are visible with SDAVIS192 even with a low optical output power: 3 nW (when the excitation blue LED is tuned to the minimum output power possible). Since the experiment shows that it is possible to reduce the contrast further, the next step is to work with brain slices in-vitro, to prove the total feasibility of the experiment, before going back to in-vivo imaging with a dedicated setup. This work is currently being done.

## 4.7. Description of test columns not yet analyzed

As mentioned, at the side of the optical array, various test columns are present (although they have not yet been tested). The entire optical array of pixels then consists, from left to right:

1. The DasisPixelHP pixel (the pixel version including a high-pass filter to reduce uncorrelated activity, not described in this work due to possible patenting issues) in an array of 6 x 192 (The first 4 columns have no salicidation on the APS circuitry and on the reset transistor, while the remaining 2 columns do not have any), followed by a spacing of 2 empty pixels to allow re-wiring. All pixels have micro lenses;
2. The apsDVSPosFB pixel (the current reference design: the improvement of [32] with positive feedback as a new feature only) in an array of 6 x 192 (the first 4 columns, as opposed to the last 2, have no salicidation on APS circuitry and reset transistor and the central 2 columns have no micro lenses), followed by a spacing of 2 empty pixels to allow re-wiring;

3. The DavisPixelSenseNoBuff pixel (the SDAVIS192 sensitive pixel with the pre-amplifier but with no unity gain buffer and no salicidation on APS circuitry and on the reset transistor) in an array of 4 x 192 (half of which with salicidation and half without);
4. The DavisPixelSense pixel (the SDAVIS192 sensitive pixel with the pre-amplifier and with the buffer and no salicidation on APS circuitry and on the reset transistor) in an array of 188 x 192.

All photodiodes are non-salicided.

Image adapted from [169]

# Chapter 5

## 5. Conclusions and discussion

This doctorate work, focused on mimicking the retina's intelligent visual processing for the purpose of robotic navigation through tracking, has delved through various fields. This allowed tackling the problems from many angles, gaining knowledge in each area.

Investigations on the possibility to use the processing of the Object Motion sensitive Cell (OMC) in conjunction with the Dynamic Vision Sensor (DVS) have shown that it is indeed possible to use the bio-inspired algorithm to detect object (the target prey's) motion. With a staring sensor it is also possible to cluster the last detected object positions and to track the overall movement of such object. It has however been understood that the OMC algorithm alone is not sufficient to solve the foreground-background separation problem and that this does not allow the tracking to continue when the sensor moves. The reason for which the algorithm inspired by the processing of the Retinal Ganglion Cell (RGC) fails compared with biology, is the huge complexity and interaction of the various RGC population types missing in the developed model. For future improvements, other types of RGCs and, possibly, higher centers of processing of the brain, should be included in the algorithm. Merging and weighting of these streams of information should allow a better understanding of the scene, without letting the algorithm be fooled by ego-motion. A work in this direction, though not biologically-inspired, is being conducted together with Prof. Dr. Alejandro Linares-Barranco, trying to merge and validate the object trackers of [30] with output stream of the OMC.

The output of the OMC could also be used to weight the regions of a frame when no ego-motion is present, in order to highlight in a DVS histogram where object movement is present. The DVS histogram could then be processed by the deep neural network introduced in chapter 3.

As for the continuation of this work, a master student of Dr. Matthew Tata, Scott Stone, from the University of Lethbridge, Alberta, Canada, has decided to study the application of such algorithm in the medical context. Specifically, in the context of humans suffering from emispatial neglect. These patients, ignore half of their visual stream, commonly as the consequence of a stroke. Just walking around without the perception of approaching and moving objects in some parts of their field of view can prove dangerous to them, therefore an algorithm which can detect these can be potentially very useful. By transducing the visual signal, through sensory substitution, into a sound, for example, collisions can be avoided and the attention of the patient can be heightened.

In an attempt to pursue more efficient tracking, for the purpose of robotic navigation, the approach to the task was shifted from a purely biological one to a more conventional machine learning one. A Dynamic and Active Pixel Vision Sensor (DAVIS) combining DVS and APS capabilities was mounted on a robot in a predator/prey scenario in the robotic arena of the University of Ulster, Northern Ireland. The sensor would produce both APS frames and DVS histograms at a variable frame-rate. These would then be fed into a JAVA-based implementation of a Convolutional Neural Network (CNN), trained on locating the prey robot in one of the three regions of its field of view. The result of the CNN would finally signal the robot controller, after an appropriate post-processing, the steering direction. The predator robot could therefore navigate in the arena following the prey robot. Successful trial runs showed impressive results.

The main limitation of this approach, however, is the ambiguity of the hand-labeled dataset but the results of the work are nonetheless outstanding. These initial results are currently being improved with the addition of distance in the robot control. This is obtained through the size of the prey robot in the visual field (S-M-XL) as this varies much more of the height of the robot in the recordings (the camera always looks at the horizon). A distribution of the prey's size is visible in Fig. 73.

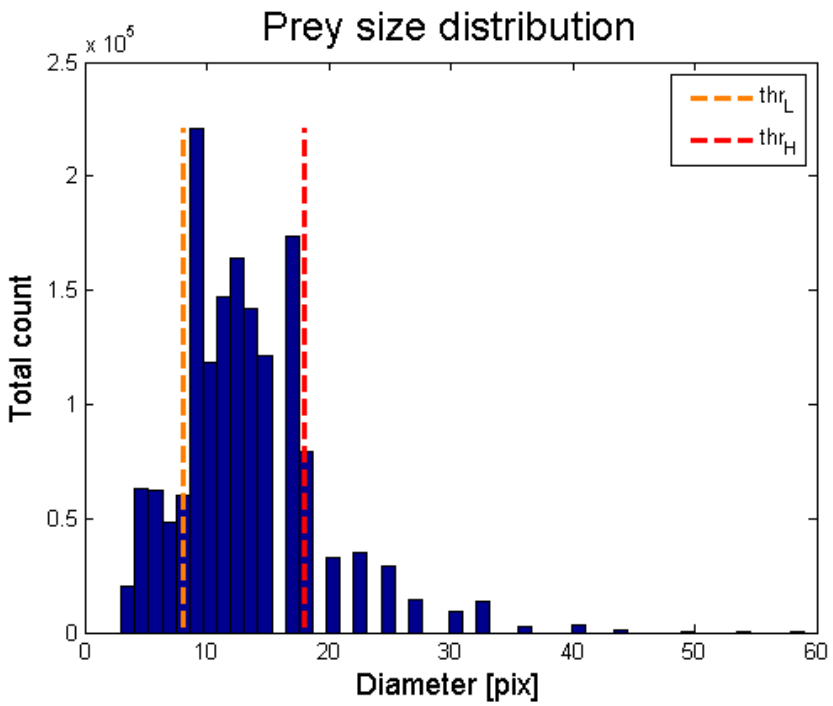


*Fig. 73 Prey size distribution with respect to a 36 x 36 input illustrating the separation in size. Sometimes the prey size can be larger than the resolution, if the prey robot is close enough.*

While this introduces further ambiguity, the intention is to extract an analog control signal (a relative position vector $p$) from the now 10 softmaxed outputs (all combinations of LCR and SMXL plus N), which goes beyond the correct or wrong output of the CNN. Output units containing R, C and L can be averaged together and summed vectorially, as Fig. 74 shows, to compute the angle of the vector $p$. The magnitude of the vector can be instead computed from the averaged SMXL units. Low-pass filtering and further constraints can then be applied again to correct the output as the network sees a time-correlated input (with respect to the shuffled video frames it sees in training). The resulting system can be seen in Fig. 74.

The results of the new net will be compared to human performance. Preliminary results in this area on a few test subjects already show that for the previous LCRN network of about 80% accuracy exceeds by roughly 20% human performance. A deeper deconvolution analysis is being deployed by PhD student Daniel Neil to find out more in detail what features in the scene most affect its decisions. Finally, the Inertial Measurement Unit (IMU) on the sensor's Printed Circuit Board (PCB) will be used to turn off APS when not needed.

Overall, although this work makes use of the current state-of-the-art CNN technology, it does so by integrating it with a neuromorphic sensor. The sensor is used to generate histogram DVS frames, but at a variable frame rate dependent on activity. The CNN will eventually be replaced by a Spiking Neural Network (SNN) once its processing time decreases to a level comparable to the stat-of-the-art. An important point of this work is also to provide a practical application of deep neural networks, which too often, are only tested on standard image recognition datasets.

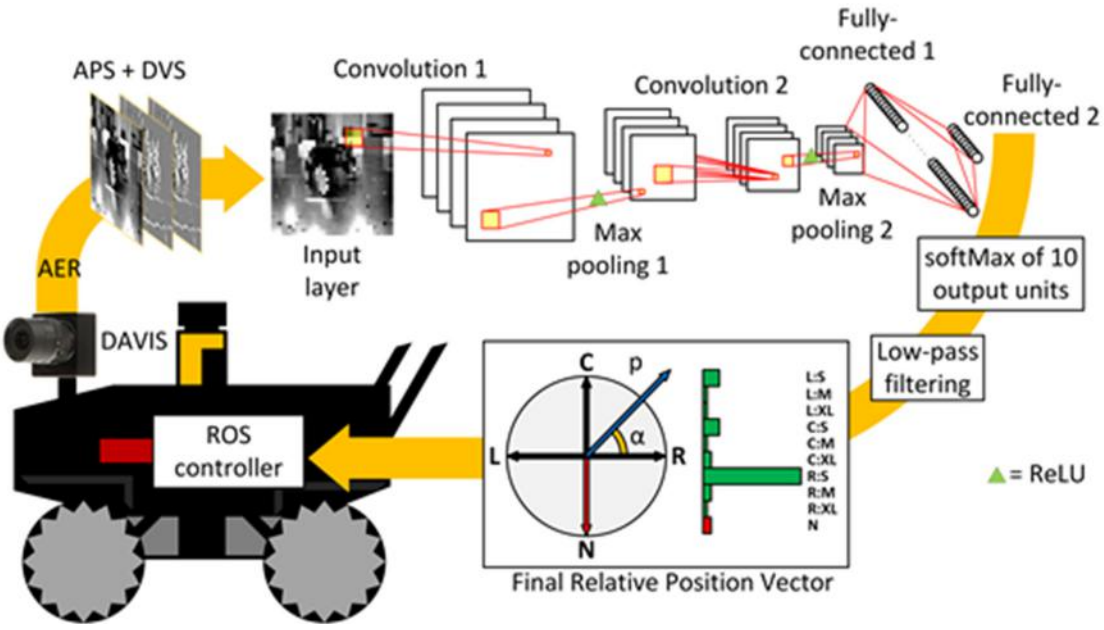


*Fig. 74 Complete system of steering control of the predator. A final relative angle and distance are computed from the position vector p as a function of the CNN's softmaxed outputs. Image adapted from [40].*

In order to improve the amount of details perceived by a DAVIS sensor, which could contain precious information regarding the tracking task, the high-contrast sensitivity silicon retina SDAVIS192 was developed. Integrating the preamplifier first described in [36], the sensor adds more gain to the DVS stream, resulting in a contrast sensitivity down to 0.95% with an SNR of 28.5 dB for OFF events and 3.45% with an SNR of 18 dB for ON events. Since more gain means a smaller intra-scene dynamic range (at least 50 dB for ON events and at least 70 dB for OFF events), a mean-lighting condition adaptation circuit moves the operating point of the pixel circuit, extending the operating range up to at least 110 dB for OFF events. This however means that information outside the intra-scene DR is lost. The sensor's visual output matches its characterization results, giving a detailed, almost tridimensional, image. The only design flaw, pinpointed through a test pixel to be the coupling between analog ground and APS column reset, is the extreme cross-talk between APS and DVS, which does not allow the two functions simultaneously.

Interesting preliminary results, in the field of imaging, have also shown a possible application of SDAVIS192. Current work is being done to use the sensor to detect weak calcium flashes of neurons expressing green fluorescent protein. The sensor, connected to a microscope, can effectively detect fluorescent beads moving in a solution and the next step will be to observe live in-vitro slices. Hopefully, the sensor will finally be used in-vivo, in case of positive results.

The PhD student Gemma Taverni, under the supervision of IniLabs' CTO Dr. Simeon Bamford, will continue to work on neural imaging and develop SDAVIS192 further in the coming years. Also, taking inspiration from this sensor, the master student Susanna Gobbi has developed, under Dr. Qiao Ning's supervision and mine, during the Neuromorphic Engineering II class, a two-stage amplifier SDAVIS192 sensor. In this case the pre-amplifier is substituted by an amplification stage similar to the one of the differential amplifier. This increases gain and complexity and area. However, the latter was not a

concern as the main application of this sensor is the detection of signal from four optical fibers gathering light from a mouse brain perfused with green fluorescent protein.

Further interesting work is currently being undertaken on the attempt to discern information based on the color of the RGBW DVS events. The color filter array present on the SDAVIS192 might in fact be useful to observe fluorescent dies of different colors in the same slice. A sneak-peek of the current work simple color reconstruction shows promising results, as shown in Fig. 75 [156].

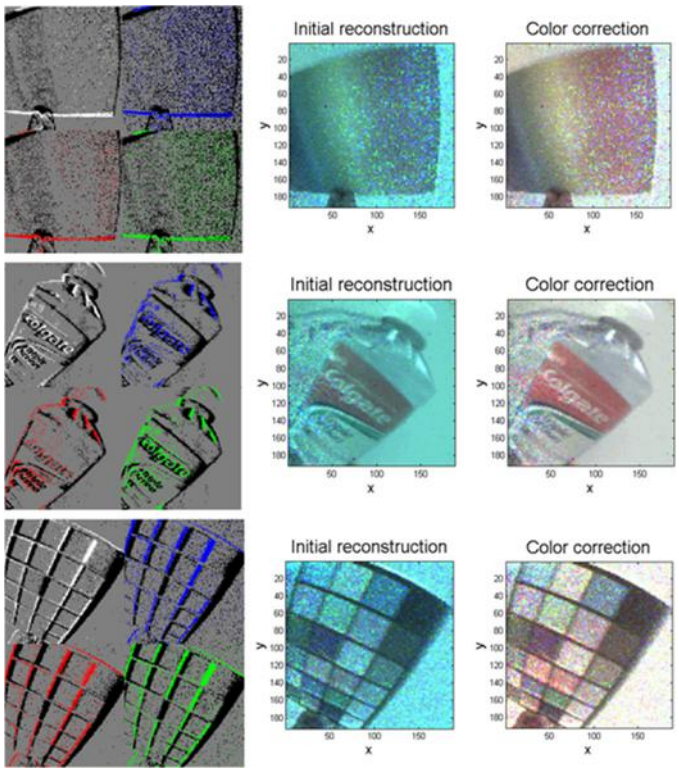


*Fig. 75 Left column: color RGBW DVS events from SDAVIS192 for three moving stimuli: a rainbow pattern, a mouthwash bottle and a Macbeth color chart. Events represented in RGBW represent color ON events of RGBW type respectively. OFF events of all type are represented in black. Central column: raw reconstruction from the 2 ms DVS bins before color correction (right column). Videos of the simple reconstruction are available at [170].*

In an effort to quantify the characteristics of the SDAVIS192 sensor, more accurate sensor characterization protocols and a more stable setup were developed. The aim of the work was to find a common testbench with which to test all spike-based sensors in order to finally be able to compare them fairly. An accurate characterization of all available sensors is about to start using these protocols. The entire open-source code is still currently being polished and will be subject to changes (the inclusion of a reconstruction evaluation), but it is already available in the jAER repository [24]. It is finally time to share the source code such that others can improve upon it and finally have more comparable and reproducible results.

# Bibliography


[1] "IBM's unveils the brain-inspired TrueNorth cognitive computer | Digital Trends." [Online]. Available: http://www.digitaltrends.com/computing/ibms-unveils-the-brain-inspired-truenorth-cognitive-computer/. [Accessed: 09-Aug-2016].

[2] D. P. Moeys, A. Linares-Barranco, T. Delbrück, and A. Rios-Navarro, "Retinal Ganglion Cell Software and FPGA Implementation for Object Detection and Tracking," in *ISCAS 2016*, Montreal, 2016, p. accepted.

[3] D. P. Moeys *et al.*, "Steering a predator robot using a mixed frame/event-driven convolutional neural network," in *2016 Second International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, 2016, pp. 1–8.

[4] D. P. Moeys *et al.*, "A Sensitive Dynamic and Active Pixel Vision Sensor for Color or Neural Imaging Applications," *IEEE Trans Biomed. Circuits Syst I*, submitted 2017.

[5] S. Furber, "Large-scale neuromorphic computing systems," *J. Neural Eng.*, vol. 13, no. 5, p. 51001, 2016.

[6] C. A. Mead, "Neuromorphic Electronic Systems," *Proc IEEE*, vol. 78, no. 10, pp. 1629–36, 1990.

[7] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbruck, and R. Douglas, *Analog VLSI: Circuits and Principles*. Cambridge, Mass: A Bradford Book, 2002.

[8] G. Indiveri, "Neuromorphic bistable VLSI synapses with spike-timing dependent plasticity," in *Advances in Neural Inform. Process. Syst. 15 (NIPS)*, Cambridge, MA: MIT Press, 2003, pp. 1115–1122.

[9] G. Indiveri *et al.*, "Neuromorphic silicon neuron circuits," *Neuromorphic Eng.*, vol. 5, p. 73, 2011.

[10] R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "A low-power wide-dynamic-range analog VLSI cochlea," *Analog Integr Circuits Signal Process*, vol. 16, pp. 245–274, 1998.

[11] V. Chan, S.-C. Liu, and A. van Schaik, "AER EAR: A matched silicon cochlea pair with address event representation interface," *IEEE Trans Circuits Syst Spec. Issue Smart Sens.*, vol. 54, no. 1, pp. 48–59, 2006.

[12] S.-C. Liu, T. Delbruck, G. Indiveri, A. Whatley, and R. Douglas, Eds., *Event-Based Neuromorphic Systems*. John Wiley and Sons Ltd., UK, 2015.

[13] C. A. Mead and M. A. Mahowald, "A silicon model of early visual processing," *Neural Netw.*, vol. 1, no. 1, pp. 91–97, 1988.

[14] M. A. Mahowald, "Silicon retina with adaptive photoreceptors," in *SPIE/SPSE Symposium on Electronic Science and Technology: From Neurons to Chips*, 1991, vol. 1473, pp. 52–58.

[15] M. Mahowald, "The Silicon Retina," in *An Analog VLSI System for Stereoscopic Vision*, Springer US, 1994, pp. 4–65.

[16] M. Mahowald, "VLSI Analogs of Neuronal Visual Processing: A Synthesis of Form and Function," PhD Thesis, California Institute of Technology, Pasadena, CA, 1992.

[17] C. Mead, "A sensitive electronic photoreceptor," in *1985 Chapel Hill Conference on VLSI*, 1985, pp. 463–471.

[18] J. Kramer, "An on/off transient imager with event-driven, asynchronous read-out," in *IEEE International Symposium on Circuits and Systems, 2002. ISCAS 2002*, 2002, vol. 2, p. II-165-II-168 vol.2.

[19] 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*, 2004, pp. 211–214.

[20] P. Lichtsteiner, "An AER Temporal Contrast Vision Sensor," Ph. D. Thesis, ETH Zurich, Zurich, Switzerland, 2006.

[21] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128 x 128 120 dB 15 µs Latency Asynchronous Temporal Contrast Vision Sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[22] K. A. Boahen, "Communicating Neuronal Ensembles between Neuromorphic Chips," in *Neuromorphic Systems Engineering*, vol. 447, T. S. Lande, Ed. Springer US, 1998, pp. 229–259.

[23] T. Delbruck and C. Mead, "An electronic photoreceptor sensitive to small changes in intensity," in *Advances in Neural Inform. Process. Syst. 1 (NIPS)*, vol. 1, D. Touretzky, Ed. San Francisco, CA, USA: Morgan Kaufman, 1989, pp. 720–727.

[24] "jAER Open Source Project," *jAER Open Source Project*, 23-Mar-2007. [Online]. Available: http://jaerproject.org. [Accessed: 23-May-2016].

[25] A. Censi and D. Scaramuzza, "Low-Latency Event-Based Visual Odometry," in *IEEE International Conference on Robotics and Automation (ICRA 2014)*, Hong Kong, China, 2014.

[26] D. Weikersdorfer, D. B. Adrian, D. Cremers, and J. Conradt, "Event-based 3D SLAM with a depth-augmented dynamic vision sensor," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 359–364.

[27] T. Brosch, S. Tschechne, and H. Neumann, "On event-based optical flow detection," *Neuromorphic Eng.*, vol. 9, p. 137, 2015.

[28] P. Bardow, A. J. Davison, and S. Leutenegger, "Simultaneous Optical Flow and Intensity Estimation From an Event Camera," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 884–892.

[29] T. Delbruck and M. Lang, "Robotic Goalie with 3ms Reaction Time at 4% CPU Load Using Event-Based Dynamic Vision Sensor," *Front. Neurosci.*, vol. 7, p. 223, Nov. 2013.

[30] A. Linares-Barranco, F. Gomez-Rodriguez, V. Villanueva, L. Longinotti, and T. Delbruck, "A USB3.0 FPGA event-based filtering and tracking framework for dynamic vision sensors," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 2417–2420.

[31] "RectangularClusterTracker (jAER javadoc)," 2007. [Online]. Available: http://jaer.sourceforge.net/javadoc/net/sf/jaer/eventprocessing/tracking/RectangularClusterTracker.html. [Accessed: 23-Oct-2015].

[32] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "A 240x180 130 dB 3 us Latency Global Shutter Spatiotemporal Vision Sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, Oct. 2014.

[33] C. Li *et al.*, "An RGBW Color VGA Rolling and Global Shutter Dynamic and Active-Pixel Vision Sensor," in *2015 International Image Sensor Workshop (IISW 2015)*, Vaals, Netherlands, 2015, pp. 393–396.

[34] T. Delbruck and R. Berner, "Temporal contrast AER pixel with 0.3%-contrast event threshold," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 2442–2445.

[35] J. A. Lenero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco, "A 3.6 µs Latency Asynchronous Frame-Free Event-Driven Dynamic-Vision-Sensor," *IEEE J. Solid-State Circuits*, vol. 46, no. 6, pp. 1443–1455, Jun. 2011.

[36] T. Serrano-Gotarredona and B. Linares-Barranco, "A 128 x 128 1.5% Contrast Sensitivity 0.9% FPN 3 µs Latency 4 mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers," *IEEE J. Solid-State Circuits*, vol. 48, no. 3, pp. 827–838, Mar. 2013.

[37] M. Yang, S. C. Liu, and T. Delbruck, "A Dynamic Vision Sensor With 1% Temporal Contrast Sensitivity and In-Pixel Asynchronous Delta Modulator for Event Encoding," *IEEE J. Solid-State Circuits*, vol. 50, no. 9, pp. 2149–2160, Sep. 2015.

[38] S. Chen, W. Tang, X. Zhang, and E. Culurciello, "A 64 64 Pixels UWB Wireless Temporal-Difference Digital Image Sensor," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 20, no. 12, pp. 2232–2240, Dec. 2012.

[39] S. Chen and A. Bermak, "Arbitrated Time-to-First Spike CMOS Image Sensor With On-Chip Histogram Equalization," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 15, no. 3, pp. 346–357, Mar. 2007.

[40] C. Posch, D. Matolin, and R. Wohlgenannt, "An asynchronous time-based image sensor," in *IEEE International Symposium on Circuits and Systems, 2008. ISCAS 2008*, 2008, pp. 2130–2133.

[41] D. Matolin, C. Posch, and R. Wohlgenannt, "True correlated double sampling and comparator design for time-based image sensors," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2009, pp. 1269–1272.

[42] C. Posch, D. Matolin, and R. Wohlgenannt, "High-DR frame-free PWM imaging with asynchronous AER intensity encoding and focal-plane temporal redundancy suppression," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 2430–2433.

[43] P. Dudek and P. J. Hicks, "A general-purpose CMOS vision chip with a processor-per-pixel SIMD array," in *Solid-State Circuits Conference, 2001. ESSCIRC 2001. Proceedings of the 27th European*, 2001, pp. 213–216.

[44] P. Dudek, "Implementation of SIMD vision chip with 128 times;128 array of analogue processing elements," in *IEEE International Symposium on Circuits and Systems, 2005. ISCAS 2005*, 2005, p. 5806–5809 Vol. 6.

[45] "Special Snowflake: 4 Things (Other Than Your Fingerprints) That Make You 1 In 7.2 Billion." [Online]. Available: http://www.medicaldaily.com/special-snowflake-4-things-other-your-fingerprints-make-you-1-72-billion-327710. [Accessed: 09-Aug-2016].

[46] R. H. Masland, "The Neuronal Organization of the Retina," *Neuron*, vol. 76, no. 2, pp. 266–280, Oct. 2012.

[47] T. Gollisch and M. Meister, "Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina," *Neuron*, vol. 65, no. 2, pp. 150–164, Jan. 2010.

[48] L. Peichl and J. González-Soriano, "Morphological types of horizontal cell in rodent retinae: a comparison of rat, mouse, gerbil, and guinea pig," *Vis. Neurosci.*, vol. 11, no. 3, pp. 501–517, Jun. 1994.

[49] H. Wässle, C. Puller, F. Müller, and S. Haverkamp, "Cone contacts, mosaics, and territories of bipolar cells in the mouse retina," *J. Neurosci. Off. J. Soc. Neurosci.*, vol. 29, no. 1, pp. 106–117, Jan. 2009.

[50] C. W. Morgans *et al.*, "TRPM1 is required for the depolarizing light response in retinal ON-bipolar cells," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 45, pp. 19174–19178, Nov. 2009.

[51] T. Gollisch and A. V. M. Herz, "The iso-response method: measuring neuronal stimulus integration with closed-loop experiments," *Front. Neural Circuits*, vol. 6, Dec. 2012.

[52] F. S. Werblin, "The retinal hypercircuit: a repeating synaptic interactive motif underlying visual function," *J. Physiol.*, vol. 589, no. Pt 15, pp. 3691–3702, Aug. 2011.

[53] R. Nelson, "Ganglion Cell Physiology by Ralph Nelson – Webvision." [Online]. Available: http://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/ganglion-cell-physiology/. [Accessed: 12-Apr-2016].

[54] C. Enroth-Cugell and J. G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat," *J. Physiol.*, vol. 187, no. 3, pp. 517–552, Dec. 1966.

[55] H. K. Hartline, "Visual receptors and retinal interaction," *Science*, vol. 164, no. 3877, pp. 270–278, Apr. 1969.

[56] S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina," *J. Neurophysiol.*, vol. 16, no. 1, pp. 37–68, Jan. 1953.

[57] R. W. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli," *Vision Res.*, vol. 5, no. 11, pp. 583–601, Dec. 1965.

[58] T. Gollisch and M. Meister, “Rapid Neural Coding in the Retina with Relative Spike Latencies,” *Science*, vol. 319, no. 5866, pp. 1108–1111, Feb. 2008.
[59] M. Greschner, A. Thiel, J. Kretzberg, and J. Ammermüller, “Complex spike-event pattern of transient ON-OFF retinal ganglion cells,” *J. Neurophysiol.*, vol. 96, no. 6, pp. 2845–2856, Dec. 2006.
[60] H. B. Barlow and R. M. Hill, “Selective sensitivity to direction of movement in ganglion cells of the rabbit retina,” *Science*, vol. 139, no. 3553, pp. 412–414, Feb. 1963.
[61] A. Thiel, M. Greschner, C. W. Eurich, J. Ammermüller, and J. Kretzberg, “Contribution of individual retinal ganglion cell responses to velocity and acceleration encoding,” *J. Neurophysiol.*, vol. 98, no. 4, pp. 2285–2296, Oct. 2007.
[62] A. Thiel, M. Greschner, and J. Ammermüller, “The temporal structure of transient ON/OFF ganglion cell responses and its relation to intra-retinal processing,” *J. Comput. Neurosci.*, vol. 21, no. 2, pp. 131–151, Oct. 2006.
[63] H. G. Wagner, E. F. Macnichol, and M. L. Wolbarsht, “The Response Properties of Single Ganglion Cells in the Goldfish Retina,” *J. Gen. Physiol.*, vol. 43, no. 6, pp. 45–62, Jul. 1960.
[64] L. Peichl and H. Wässle, “The structural correlate of the receptive field centre of alpha ganglion cells in the cat retina.,” *J. Physiol.*, vol. 341, pp. 309–324, Aug. 1983.
[65] C. E. Keeler, E. Sutcliffe, and E. L. Chaffee, “Normal and ‘Rodless’ Retinae of the House Mouse with Respect to the Electromotive Force Generated through Stimulation by Light,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 14, no. 6, pp. 477–484, Jun. 1928.
[66] I. Provencio, G. Jiang, W. J. De Grip, W. P. Hayes, and M. D. Rollag, “Melanopsin: An opsin in melanophores, brain, and eye,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 1, pp. 340–345, Jan. 1998.
[67] S. Hattar, H. W. Liao, M. Takao, D. M. Berson, and K. W. Yau, “Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity,” *Science*, vol. 295, no. 5557, pp. 1065–1070, Feb. 2002.
[68] T. A. Münch, R. A. Da Silveira, S. Siegert, T. J. Viney, G. B. Awatramani, and B. Roska, “Approach sensitivity in the retina processed by a multifunctional neural circuit,” *Nat. Neurosci.*, vol. 12, no. 10, pp. 1308–1316, 2009.
[69] B. P. Ölveczky, S. A. Baccus, and M. Meister, “Segregation of object and background motion in the retina,” *Nature*, vol. 423, no. 6938, pp. 401–408, May 2003.
[70] S. A. Baccus, B. P. Ölveczky, M. Manu, and M. Meister, “A Retinal Circuit That Computes Object Motion,” *J. Neurosci.*, vol. 28, no. 27, pp. 6807–6817, Jul. 2008.
[71] “Visualise EU project.” [Online]. Available: http://www.visualise-project.eu/. [Accessed: 11-Aug-2015].
[72] “jAER / Wiki / filt.eu.seebetter.ini.retinamodel.ApproachCell.” [Online]. Available: https://sourceforge.net/p/jaer/wiki/filt.eu.seebetter.ini.retinamodel.ApproachCell/. [Accessed: 15-Apr-2016].
[73] D. P. Moeys, T. Delbrück, and A. Linares-Barranco, “Object Motion sensitive Cell (OMC).” [Online]. Available: https://youtu.be/9bbgITgpJG0. [Accessed: 15-Apr-2016].
[74] K.-C. Tseng and A. C. Parker, “A neuromorphic circuit that computes differential motion,” in *2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2012, pp. 89–92.
[75] T. Iakymchuk *et al.*, “An AER handshake-less modular infrastructure PCB with x8 2.5Gbps LVDS serial links,” in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 1556–1559.
[76] D. P. Moeys, T. Delbrück, A. Rios-Navarro, and A. Linares-Barranco, “Live demonstration: Retinal ganglion cell software and FPGA implementation for object detection and tracking,” in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 1445–1445.

[77] A. Linares-Barranco *et al.*, "Low Latency Event-based Processing for Dynamic Vision Sensors on FPGAs," *IEEE Trans. Circuits Syst. I*, p. in progress.
[78] Editor, "Computer simulation of pyramidal neurons in the cerebral cortex," *Inside the Brain*, 17-Jul-2012. .
[79] "CS231n Convolutional Neural Networks for Visual Recognition." [Online]. Available: http://cs231n.github.io/. [Accessed: 29-Sep-2016].
[80] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, 2012.
[81] Y. Anzai, *Pattern Recognition and Machine Learning*. Elsevier, 2012.
[82] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133.
[83] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *ArXiv13112901 Cs*, Nov. 2013.
[84] "Convolutional Neural Networks (LeNet) — DeepLearning 0.1 documentation." [Online]. Available: http://deeplearning.net/tutorial/lenet.html. [Accessed: 01-Oct-2016].
[85] "An Introduction to Convolutional Neural Networks - Teach." [Online]. Available: http://scarlet.stanford.edu/teach/index.php/An_Introduction_to_Convolutional_Neural_Networks. [Accessed: 02-Oct-2016].
[86] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J Physiol*, vol. 160, no. 1, pp. 106–154, 1962.
[87] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202.
[88] Y. LeCun *et al.*, "Handwritten Digit Recognition with a Back-Propagation Network," in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, Ed. Morgan-Kaufmann, 1990, pp. 396–404.
[89] Y. LeCun *et al.*, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput*, vol. 1, no. 4, pp. 541–551, Dec. 1989.
[90] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, pp. 2278–2324.
[91] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 253–256.
[92] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
[93] C. Szegedy *et al.*, "Going Deeper with Convolutions," *ArXiv14094842 Cs*, Sep. 2014.
[94] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ArXiv14091556 Cs*, Sep. 2014.
[95] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *ArXiv151203385 Cs*, Dec. 2015.
[96] C. Farabet *et al.*, "Comparison between Frame-Constrained Fix-Pixel-Value and Frame-Free Spiking-Dynamic-Pixel ConvNets for Visual Processing," *Front. Neurosci.*, vol. 6, Apr. 2012.
[97] J. . Perez-Carrasco, B. Acha, C. Serrano, L. Camunas-Mesa, T. Serrano-Gotarredona, and B. Linares-Barranco, "Fast Vision Through Frameless Event-Based Sensing and Convolutional Processing: Application to Texture Recognition," *IEEE Trans. Neural Netw.*, vol. 21, no. 4, pp. 609–620, Apr. 2010.
[98] A. Linares-Barranco *et al.*, "On the AER convolution processors for FPGA," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 4237–4240.

[99] L. Camunas-Mesa, A. Acosta-Jimenez, C. Zamarrefio-Ramos, T. Serrano-Gotarredona, and B. Linares-Barranco, "A 32 x 32 Pixel Convolution Processor Chip for Address Event Vision Sensors With 155 ns Event Latency and 20 Meps Throughput," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 58, no. 4, pp. 777–790, Apr. 2011.

[100] L. Camunas-Mesa, C. Zamarreno-Ramos, A. Linares-Barranco, A. J. Acosta-Jimenez, T. Serrano-Gotarredona, and B. Linares-Barranco, "An Event-Driven Multi-Kernel Convolution Processor Module for Event-Driven Vision Sensors," *IEEE J. Solid-State Circuits*, vol. 47, no. 2, pp. 504–517, Feb. 2012.

[101] J. Jin, V. Gokhale, A. Dundar, B. Krishnamurthy, B. Martini, and E. Culurciello, "An efficient implementation of deep convolutional neural networks on a mobile coprocessor," in *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2014, pp. 133–136.

[102] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[103] P. Baldi and P. J. Sadowski, "Understanding Dropout," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 2814–2822.

[104] M. A. Nielsen, "Neural Networks and Deep Learning," 2015.

[105] "CIFAR-10 and CIFAR-100 datasets." [Online]. Available: http://www.cs.utoronto.ca/~kriz/cifar.html. [Accessed: 02-Oct-2016].

[106] "Caltech101." [Online]. Available: http://www.vision.caltech.edu/Image_Datasets/Caltech101/. [Accessed: 02-Oct-2016].

[107] "CAVE | Software: COIL-20: Columbia Object Image Library." [Online]. Available: http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php. [Accessed: 02-Oct-2016].

[108] "The Street View House Numbers (SVHN) Dataset." [Online]. Available: http://ufldl.stanford.edu/housenumbers/. [Accessed: 02-Oct-2016].

[109] "MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges." [Online]. Available: http://yann.lecun.com/exdb/mnist/. [Accessed: 02-Oct-2016].

[110] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column Deep Neural Networks for Image Classification," *ArXiv12022745 Cs*, Feb. 2012.

[111] "ImageNet Large Scale Visual Recognition Competition (ILSVRC)." [Online]. Available: http://www.image-net.org/challenges/LSVRC/. [Accessed: 02-Oct-2016].

[112] J. Fan, W. Xu, Y. Wu, and Y. Gong, "Human Tracking Using Convolutional Neural Networks," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1610–1623, Oct. 2010.

[113] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust Visual Tracking via Convolutional Networks," *ArXiv150104505 Cs*, Jan. 2015.

[114] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, "Deep Image: Scaling up Image Recognition," *ArXiv150102876 Cs*, Jan. 2015.

[115] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5325–5334.

[116] S. U. Rehman, S. Tu, Y. Huang, and Z. Yang, "Face recognition: A novel un-supervised convolutional neural network method," in *2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)*, 2016, pp. 139–144.

[117] T. Weyand, I. Kostrikov, and J. Philbin, "PlaNet - Photo Geolocation with Convolutional Neural Networks," *ArXiv160205314 Cs*, Feb. 2016.

[118] M. Lai, "Deep Learning for Medical Image Segmentation," *ArXiv150502000 Cs*, May 2015.

[119] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *ArXiv160506211 Cs*, May 2016.
[120] M. Bojarski *et al.*, “End to End Learning for Self-Driving Cars,” *ArXiv160407316 Cs*, Apr. 2016.
[121] “DeepDriving.” [Online]. Available: http://deepdriving.cs.princeton.edu/. [Accessed: 02-Oct-2016].
[122] “Caffe | Deep Learning Framework.” [Online]. Available: http://caffe.berkeleyvision.org/. [Accessed: 01-Feb-2016].
[123] D. P. Moeys, E. Kerr, and T. Delbruck, “VISUALISE Project Predator-Prey Robot Synchronized Arena + jAER Video 2016 03 23 063047 - YouTube,” 23-Mar-2016. [Online]. Available: https://youtu.be/fL3YCIPxuhM. [Accessed: 23-Mar-2016].
[124] D. P. Moeys *et al.*, “Steering a Predator Robot using a Mixed Frame/Event-Driven Convolutional Neural Network,” *ArXiv160609433 Cs*, Jun. 2016.
[125] “Motherboard | Home,” *Motherboard*. [Online]. Available: http://motherboard.vice.com/de. [Accessed: 27-Sep-2016].
[126] “Scientists Taught a Robot to Hunt Prey,” *Motherboard*. [Online]. Available: http://motherboard.vice.com/read/scientists-taught-a-robot-to-hunt-prey-deep-learning-visualise. [Accessed: 27-Sep-2016].
[127] R. Ferris, “Scientists are teaching robots how to hunt,” *CNBC*, 06-Jul-2016. [Online]. Available: http://www.cnbc.com/2016/07/06/scientists-are-teaching-robots-how-to-hunt.html. [Accessed: 27-Sep-2016].
[128] B. M. R. and R. O. for MailOnline, “The end is nigh! A killer robot has been taught how to hunt predators,” *Mail Online*, 06-Jul-2016. [Online]. Available: http://www.dailymail.co.uk/sciencetech/article-3676651/Killer-robots-taught-hunt-target-prey-scientists.html. [Accessed: 27-Sep-2016].
[129] “Scientists developing new AI to let robots ‘hunt’ their prey like a predator,” *Blastr*, 06-Jul-2016. [Online]. Available: http://www.blastr.com/2016-7-6/scientists-developing-new-ai-let-robots-hunt-their-prey-predator. [Accessed: 27-Sep-2016].
[130] “Scientists are teaching robots how to hunt down prey,” *Engadget*. [Online]. Available: https://www.engadget.com/2016/07/05/robots-hunt-prey/. [Accessed: 27-Sep-2016].
[131] “Scientists build ‘predator’ robot that teaches itself to hunt,” *FastCo News*. [Online]. Available: https://news.fastcompany.com/scientists-build-predator-robot-that-teaches-itself-to-hunt-4012894. [Accessed: 27-Sep-2016].
[132] H. Liu, D. P. Moeys, G. Das, D. Neil, S. C. Liu, and T. Delbrück, “Combined frame- and event-based detection and tracking,” in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 2511–2514.
[133] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp, “Off-Road Obstacle Avoidance through End-to-End Learning,” in *Advances in Neural Information Processing Systems*, 2005, pp. 739–746.
[134] D. A. Pomerleau, “ALVINN: An Autonomous Land Vehicle in a Neural Network,” Jan. 1989.
[135] null, “A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots,” *IEEE Robot. Autom. Lett.*, vol. PP, no. 99, pp. 1–1, 2015.
[136] “Mobile Robot SUMMIT XL,” *Robotnik*. [Online]. Available: http://www.robotnik.eu/mobile-robots/summit-xl/. [Accessed: 08-Jan-2016].
[137] “ROS.org | Powering the world’s robots.” [Online]. Available: http://www.ros.org/. [Accessed: 02-Feb-2016].
[138] Y. Jia *et al.*, “Caffe: Convolutional Architecture for Fast Feature Embedding,” *ArXiv14085093 Cs*, Jun. 2014.
[139] D. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *3rd Int. Conf. Learn. Represent. San Diego 2015*, Dec. 2014.

[140] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for Simplicity: The All Convolutional Net,” *ArXiv14126806 Cs*, Dec. 2014.

[141] D. P. Moeys, E. Kerr, and T. Delbruck, “VISUALISE Project Predator Prey CNN activity 2016 03 24 070136 - YouTube,” 23-Mar-2016. [Online]. Available: https://youtu.be/lPF3Youpmqk. [Accessed: 24-Mar-2016].

[142] “Silicon Wafer wallpaper_other_health questions,pictures,fotos.” [Online]. Available: http://www.tianxinqi.com/news/Silicon-Wafer10lcbmxbsd. [Accessed: 09-Aug-2016].

[143] “www.seebetter.eu,” 2009. [Online]. Available: http://www.seebetter.eu/. [Accessed: 15-Mar-2015].

[144] C. Li *et al.*, “Design of an RGBW color VGA rolling and global shutter dynamic and active-pixel vision sensor,” in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 718–721.

[145] M. Yang, S. C. Liu, and T. Delbruck, “A Dynamic Vision Sensor With 1% Temporal Contrast Sensitivity and In-Pixel Asynchronous Delta Modulator for Event Encoding,” *IEEE J. Solid-State Circuits*, vol. 50, no. 9, pp. 2149–2160, Sep. 2015.

[146] C. A. Mead, *Analog VLSI and Neural Systems*, 1st edition. Reading, MA: Addison Wesley Publishing Company, 1989.

[147] S. C. Liu and T. Delbruck, “Neuromorphic sensory systems,” *Curr. Opin. Neurobiol.*, vol. 20, no. 3, pp. 288–295, Jun. 2010.

[148] D. Matolin, C. Posch, R. Wohlgenannt, and T. Maier, “A 64 x 64 pixel temporal contrast microbolometer infrared sensor,” in *IEEE International Symposium on Circuits and Systems, 2008. ISCAS 2008*, 2008, pp. 1644–1647.

[149] T.-W. Chen *et al.*, “Ultra-sensitive fluorescent proteins for imaging neuronal activity,” *Nature*, vol. 499, no. 7458, pp. 295–300, Jul. 2013.

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

[151] *Principles of Fluorescence Spectroscopy | Joseph R. Lakowicz | Springer*. .

[152] L. Kastrup and S. W. Hell, “Absolute Optical Cross Section of Individual Fluorescent Molecules,” *Angew. Chem. Int. Ed.*, vol. 43, no. 48, pp. 6646–6649, Dec. 2004.

[153] C. Posch, D. Matolin, and R. Wohlgenannt, “A two-stage capacitive-feedback differencing amplifier for temporal contrast IR sensors,” *Analog Integr. Circuits Signal Process.*, vol. 64, no. 1, pp. 45–54, Jul. 2010.

[154] T. Serrano-Gotarredona, B. Linares-Barranco, and A. G. Andreou, “Very wide range tunable CMOS/bipolar current mirrors with voltage clamped input,” *IEEE Trans. Circuits Syst. Fundam. Theory Appl.*, vol. 46, no. 11, pp. 1398–1407, Nov. 1999.

[155] T. Delbrück, R. Berner, P. Lichtsteiner, and C. Dualibe, “32-bit Configurable bias current generator with sub-off-current capability,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 1647–1650.

[156] D. P. Moeys *et al.*, “Color Temporal Contrast Sensitivity in Dynamic Vision Sensors,” presented at the 2017 IEEE Intl. Symp. on Circuits and Systems (ISCAS), Baltimore, MD, USA, 2017.

[157] R. Berner, P. Lichtsteiner, and T. Delbruck, “Self-timed vertacolor dichromatic vision sensor for low power pattern detection,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2008, pp. 1032–1035.

[158] “EMVA1228 Standard for Characterization of Image Sensors and Cameras Release 3.0.” European Machine Vision Association, 2010.

[159] Diederik Paul Moeys, “SDAVIS and DAVIS240C comparison.” [Online]. Available: https://www.youtube.com/watch?v=QsYeSguI0b8. [Accessed: 27-Nov-2016].

[160] Y. Gong *et al.*, “High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor,” *Science*, vol. 350, no. 6266, pp. 1361–1366, Dec. 2015.

[161] B. Jähne, “EMVA 1288 Standard for Machine Vision,” *Opt. Photonik*, vol. 5, no. 1, pp. 53–54, Apr. 2010.
[162] “How To Measure ‘Photon Transfer Curve’ (1) ? « Harvest Imaging Blog.” .
[163] A. El Gamal and H. Eltoukhy, “CMOS Image Sensors, an introduction to the technology, design, and performance limits, presenting recent developements and future directions,” *IEEE Circuits Devices Mag.*, Jun. 2015.
[164] David Gardner, Summit Imaging, “Characterizing Digital Cameras with the Photon Transfer Curve.” .
[165] “THS1030 | High Speed ADC (>10MSPS) | Analog to Digital Converter | Description & parametrics.” [Online]. Available: http://www.ti.com/product/THS1030. [Accessed: 08-Feb-2016].
[166] “imec.” [Online]. Available: http://www2.imec.be/be_en/home.html. [Accessed: 26-Jun-2016].
[167] L. Longinotti, “cAER: A framework for event-based processing on embedded systems,” Bachelor Thesis, University of Zurich, Zurich, Switzerland, 2014.
[168] D. P. Moeys, T. Delbruck, and S.-C. Liu, “Current-mode automated quality control cochlear resonator for bird identity tagging,” in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1734–1737.
[169] “tech_general.jpg (1954×983).” [Online]. Available: https://cassandrajohn.files.wordpress.com/2014/12/tech_general.jpg. [Accessed: 09-Aug-2016].
[170] Diederik Paul Moeys, “Real color RGBW DVS event reconstruction from white frame.” [Online]. Available: https://www.youtube.com/watch?v=xTsR7i-ho1U&feature=youtu.be. [Accessed: 31-Oct-2016].
[171] “LED light source / white / portable / for endoscopes - 50 000 lux - Fiberscope.net by MEDIT.” [Online]. Available: http://www.directindustry.com/prod/fiberscopenet-medit/product-40228-398276.html. [Accessed: 28-Mar-2017].
[172] R. Berner, P. Lichtsteiner, and T. Delbruck, “Self-timed vertacolor dichromatic vision sensor for low power face detection,” presented at the ISCAS 2008, 2008, pp. 1032–1035.
[173] “Simple black picture,” *Picky Wallpapers*. [Online]. Available: http://www.pickywallpapers.com/miscellaneous/clean-style/simple-black-picture/. [Accessed: 09-Aug-2016].
[174] B. M. Onat, X. Jiang, and M. Itzler, “A systematic approach to dark current reduction in InGaAs-based photodiode arrays for shortwave infrared imaging,” in *Proceedings of IEEE-Photonics Society Conference, Antalya, Turkey, TuM1*, 2009, vol. 231.
[175] N. V. Loukianova *et al.*, “Leakage current modeling of test structures for characterization of dark current in CMOS image sensors,” *IEEE Trans. Electron Devices*, vol. 50, no. 1, pp. 77–83, Jan. 2003.
[176] B. Van Zeghbroeck, “Principles of Semiconductor Devices.” [Online]. Available: http://ecee.colorado.edu/~bart/book/. [Accessed: 11-Feb-2015].
[177] A. Yadav, *Solid State Devices and Circuits*. Firewall Media, 2008.
[178] “TowerJazz: CMOS Image Sensor (CIS) Technology.” [Online]. Available: http://www.towerjazz.com/cmos-image-sensor.html. [Accessed: 11-Feb-2015].
[179] S. C. Liu, J. Kramer, G. Indiveri, T. Delbruck, and R. Douglas, *Analog VLSI: Circuits and Principles*. MIT Press, 2002.
[180] S.-W. Han, S.-J. Kim, and E. Yoon, “A low dark current CMOS image sensor pixel with a photodiode structure enclosed by p-well,” *J. Semicond. Technol. Sci.*, vol. 5, no. 2, pp. 102–106, 2005.

[181] S.-G. Lee, J.-W. Park, J.-M. Kim, S.-H. Bae, and Y.-J. Lee, *Image Artifacts by Charge Pocket in Floating Diffusion Region on CMOS Image Sensors*. INTECH Open Access Publisher, 2011.

[182] J. F. Cox, *Fundamentals of Linear Electronics: Integrated and Discrete*. Cengage Learning, 2001.

[183] "The art of Analog Layout (Semiconductor Fabrication)." [Online]. Available: http://wenku.baidu.com/view/fade9e717fd5360cba1adb85. [Accessed: 11-Feb-2015].

[184] L. J. Chen, *Silicide Technology for Integrated Circuits (Processing)*. The Institution of Engineering and Technology, 2009.

[185] D.-U. Kang *et al.*, "Comparative study of various pixel photodiodes for digital radiography: Junction structure, corner shape and noble window opening," *J. Korean Phys. Soc.*, vol. 60, no. 9, pp. 1413–1418, May 2012.

[186] I. Shcherback, A. Belenky, and O. Yadid-Pecht, "Empirical dark current modeling for complementary metal oxide semiconductor active pixel sensor," *Opt. Eng.*, vol. 41, no. 6, pp. 1216–1219, 2002.

[187] S. Taylor, "Low Noise NMOS Imager Fabrication with a Study on the Effect of Photodiode shape and dark current." Grand Valley State University.

[188] Hamamatsu, "Photodiode technical Information." .

[189] "Thorlabs.com - Tutorials." [Online]. Available: http://www.thorlabs.de/tutorials.cfm?tabID=31760. [Accessed: 11-Feb-2015].

[190] C. Hu, "MOSFET Technology Scaling, Leakage Current, and Other Topics." EECS 130 Integrated Circuit Devices, UC Berkeley, 2006.

[191] H. Liu, C. Brandli, C.-H. Li, S.-C. Liu, and T. Delbruck, "Design of a Spatiotemporal Correlation Filter for Event-based Sensors." IEEE, ISCAS, 2015.

[192] R. Berner, T. Delbruck, A. C. B. US, and A. L. B. US, "High-speed USB2. 0 AER interfaces," Diploma thesis, Dept. Elect. Eng., University of Zürich, ETH Zürich, and Universidad de Sevilla, Zurich, Switzerland, 2006.

Image adapted from [171]

# Chapter 6: APPENDIX I

## 6. APPENDIX I: Uniform light intensity sweeps setup

### 6.1. Circuit topology

To perform the measurements, for both DVS and APS, in which a uniform light source sweep is required, the setup of Fig. 76 was used. This proposed setup is a re-work of Dr. Raphael Berner's original set-up for testing the DollBrain chip [172]. The purpose of this setup is to provide a light source which is homogeneous over the whole sensor array, and to change the intensity of this light linearly with a known and accurate voltage source.

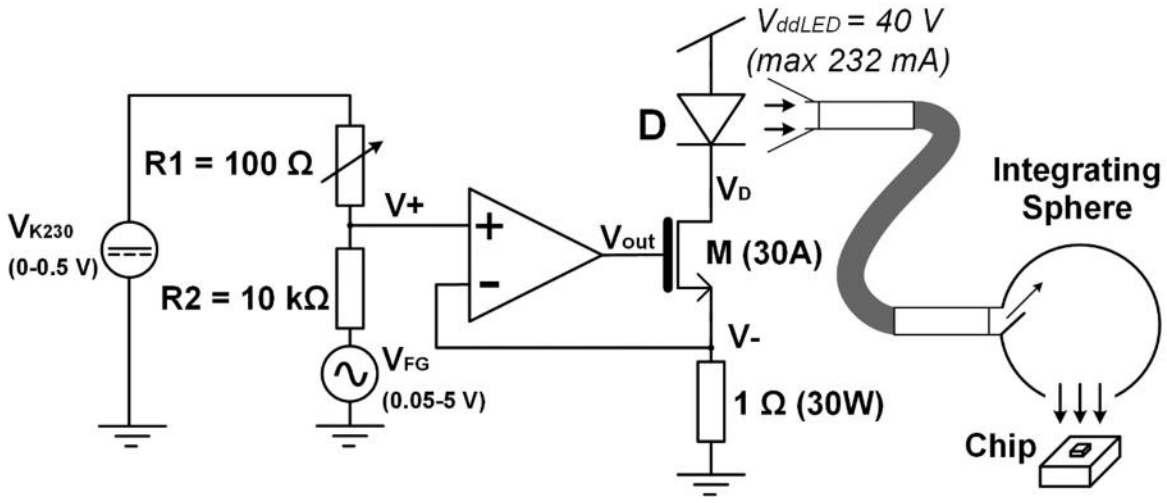


*Fig. 76 Testing setup for uniform light intensity sweep.*

The current through the Light Emitting Diode (LED) D, which is linearly proportional to light (at least in a restricted range) is set by a 1 Ω resistor and by the voltage across it $V_{-}$. This is possible thanks to the clamping action of the operational amplifier used in a feedback arrangement with a power n-type MOSFET M. Since in the negative feedback configuration $V_{+}$ and $V_{-}$ are forced to be equal, whichever voltage is set at $V_{+}$ will appear at $V_{-}$. According to Ohm's law, a voltage over a fixed resistance will produce a linearly dependent current. In this case, since the resistance is only 1 Ω, the current produced will be equal to the voltage $V_{-}$ set (and therefore $V_{+}$). The output of the operational amplifier, connected to the gate of M, will settle at a voltage corresponding to the gate-to-source voltage of M ($V_{out} - V_{-}$) necessary to provide the current set by the 1 Ω resistor. The LED's current will therefore be linearly set by the value of $V_{+}$. In order for this to work, the supply voltage of the LED, $V_{ddLED}$, needs to be at least as big as the maximum $V_{+}$ (and therefore current) which is intended to be used, $V_{+MAX}$, plus the saturation voltage of M, $V_{Msat}$, plus the forward voltage of D, $V_{Dfv}$. The voltage supply of the LED, $V_{ddLED}$, should then be:

$$V_{ddLED} = V_{+MAX} + V_{Dfv} + V_{Msat} \tag{6.1.1}$$

According to the datasheet of the selected components (the STMicroelectronics STP36NF06 MOSFET M and the Cree Inc. MHDGWT-0000-000N0HM230GCT LED lamp D) $V_{ddLED}$ was chosen to be 40 V (provided by the two supplies in series of the Thurlby Thandar Instruments (TTI) 2x 32 V 3 A). This is assuming a $V_{+MAX}$ of only 1 V for a current intended use of up to 1 A (for more current the supply can be easily adjusted). The value is quite large compared to the previous setup since the LED effectively consists of an array of LEDs in series which all have a forward voltage of about 3 V. The light of this chosen LED is warm white at 3000 K, although this values is defined at a particular current value only. Although the current of the used LED can only go up to less than 250 mA, to allow the use, if wanted, of much more power consuming LEDs, the circuit was designed to withstand up to a 30 A LED. The 1 Ω resistor has in fact a power rating of 30 W and the power MOSFET M can provide up to 30 A. The reason why this resistor is of only 1 Ω is that if it were larger, to reach 30 A the input voltage $V_{+}$ would have to be 30 V at least, larger than what can be provided my most devices. Also, to guarantee 30 A, the voltage supply of the operational amplifier needs to be at least 5 V to allow its output node $V_{out}$ to reach a gate-to-source voltage value for M large enough for this current to be

sunk. The LED is placed onto a heatsink which can withstand up to 47 W of power. This value is largely sufficient and heat-paste ensures that the heat can be efficiently transferred from the LED's back to this multi-winged heatsink. The power MOSFET M and the 30 W resistor both have for the moment a small 1 W heatsink made for TO-220 packages. These should be replaced for larger current uses. Probably the two packages could also be screwed onto the large heatsink of the LED.

The light emitted by the LED is then coupled into a thick optical fiber (the one from the Mitutoyo Lamplink light source) with a small metal cylinder in which the fiber is fitted. The cylinder is attached to the LED with hot glue and ensures that almost no light can escape sideways. The integrating sphere, from Edmund Optics (6'' general purpose integrating sphere system), then receives the fiber as an input. Inside the sphere, the light does not travel directly to the exit, situated at 90° from the entry point, but it is instead deviated by a small white metal panel. This makes the light travel a long way by being reflected against the white internal surface of the sphere, before finally appearing more homogeneous onto the image sensor's pixel array facing the exit point of the sphere without any lenses. It was noted that, while using Dr. Berner's setup during APS characterization, strange oblique patterns would appear in the image, therefore interfering with the correct measurement procedure. This means that even using the integrating sphere does not prevent heavy light interference patterns to appear at the output. The source of this problem was individuated by a quick consultation with Heptagon Oy's optical engineer Elisa Parola, who pointed out that the optical fiber was not behaving correctly. This was due to the fact that fiber's tips were wobbly, probably because of misuse or falls of the device. The tips are the most delicate part of the fiber and if any crack is present in them, then strong internal reflections can appear at the output, as observed. Another important effect is that most of the light will be reflected at the source, therefore limiting the operating light intensity range of the setup. This is the reason why Dr. Berner noted up to 2.2 klux in his PhD thesis about his setup and Dr. Minhao Yang, who had to perform similar measurements years later, had to redesign the circuit himself to increase the power disproportionally to get the same light intensities. Substituting the fiber allowed to use as little as 232 mA to achieve up to 4 klux compared to Dr. Yang's few tens of Amperes. Moreover, due to this higher power consumption, Dr. Yang's setup was particularly unstable and had to be switched off regularly to avoid overheating as power dissipation devices were not sized correctly. This in turn changed the light's color temperature and illumination conditions of the setup quite unpredictably, probably affecting the measurements.

As regards the input to the circuit, the connections and resistor values were modified from the previous designs of Dr. Berner and Dr. Yang. In both of their design the DC signal (setting the DC value of $V_+$) was set through the Keithley 230 programmable voltage source (K230) too, but the AC voltage source, provided by the HP function generator 33120a, was connected across R1. This bad arrangement led to a grounding of the K230's positive terminal and of $V_+$ during the operation of the function generator. This is due to the fact that one of the two terminals of the function generator is always connected to ground, the same ground to which the K230 connects to through the wall plug with its negative terminal. No other solution was found than to place the function generator between R2 and ground. The result of this configuration is two sources being divided by the same voltage divider in two different ways. Using the principle of superposition which states that every voltage source can be considered at a time by short-circuiting the others present, it can be seen that $V_{K230}$ is divided by the reistive divider set by R1 and R2 to give $V_{+DC}$, the DC component of $V_+$:

$$V_{+DC} = V_{K230}\frac{R2}{R1+R2} \tag{6.1.2}$$

In the case of the function generator instead, which constitutes the AC component of $V_+$, called $V_{+AC}$, the resistive divider is inverted and the result is:

$$V_{+AC} = V_{FG}\frac{R1}{R1+R2} \tag{6.1.3}$$

The two components, DC and DC, add to form $V_+$:

$$V_+ = V_{+AC} + V_{+DC} \tag{6.1.4}$$

Since the minimum AC voltage that the function generator can provide is 50 mV, too big to achieve low voltages at $V_-$ and therefore low light intensities, the resistive divider must be several times smaller than 1. In this case R2 was chosen to be 10 kΩ and R1 100 Ω. This turns out in a resistive divider of 0.99 (~1) for DC of 0.0099 (~0.01, one hundredth) for AC. Since $V_{K230}$ can be precise down to 1 mV, these values suit perfectly. However, to allow further flexibility, especially to recreate high luminosity conditions, the value of R1 can be tuned with a trimmer up to 1 MΩ.

## 6.2. Printed Circuit Board

The circuit was first built onto a small prototyping board and was later turned into a dual layer Printed Circuit Board (PCB) manufactured at Bastli ETH (the small laboratory in the AMIV student association of ETH). The schematic, made with EAGLE PCB 7.5.0 Light, shown in Fig. 77, is identical to Fig. 76 in terms of the basic circuit topology but adds some basic features. The first is an input switch to connect the power to the transconductance amplifier, the DC voltage component from the Keithley 230 and the AC voltage component from the Function Generator. The second is an LED to see whether the operational amplifier is powered up. The large 3000 K LED is not present on board, as it is attached to a large off-board heatsink.

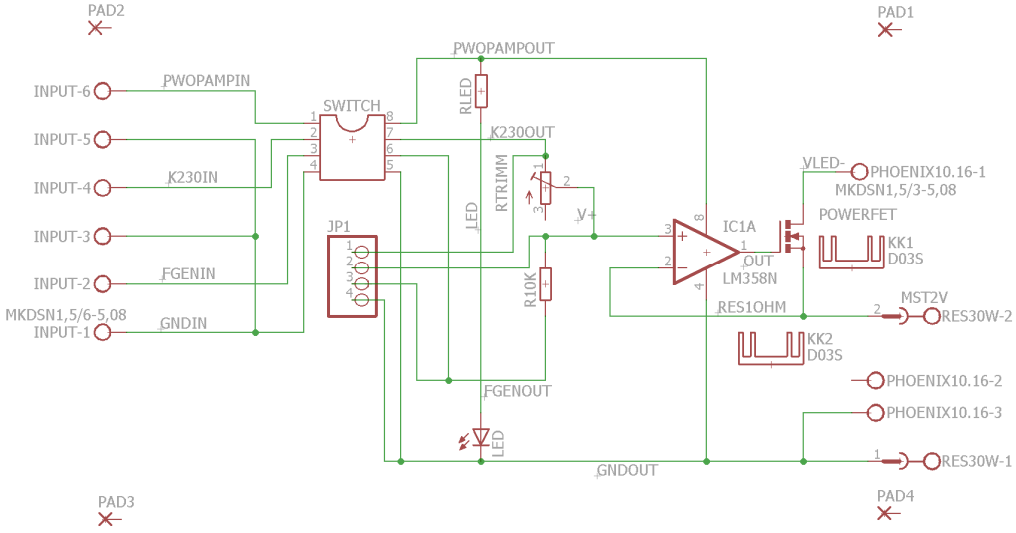


*Fig. 77 EAGLE PCB schematic (the 1 Ω resistor is replaced with a connector with similar footprint).*

All signals can be connected to the board through Phoenix connectors which can withstand up to 30 A. There are four header test points to connect to 1) $V_{K230}$, 2) $V_+$, 3) $V_{FG}$ and 4) ground. As for the

layout, shown in Fig. 78, no particular arrangement is considered other than the width of the lines where 30 A could flow and the space needed by the heatsinks of the power MOSFET and the 30 W 1 Ω resistor. The PCB was manufactured with a LPKF ProtoMat S63 at Bastli, the electronics laboratory of the AMIV, the electronics department (ITET) student association.

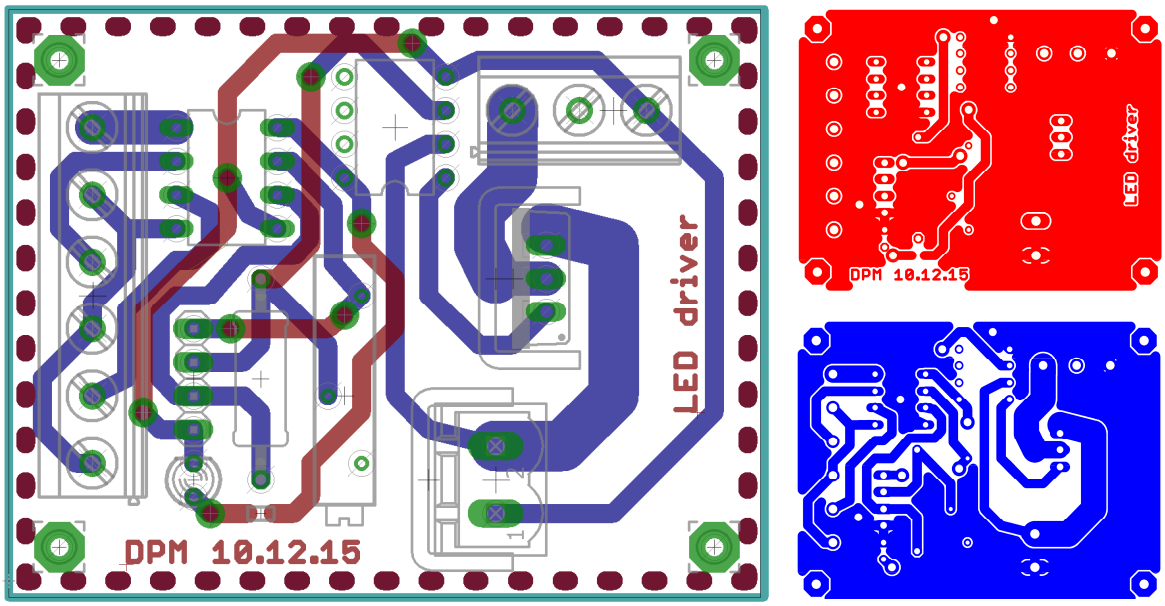


*Fig. 78 Left: EAGLE PCB layout. Top right: top side. Bottom right: bottom side. Top metal is indicated in red, bottom in blue, pads in green and milling layer (PCB outline) in light blue.*

A picture of the final setup is shown in Fig. 79.

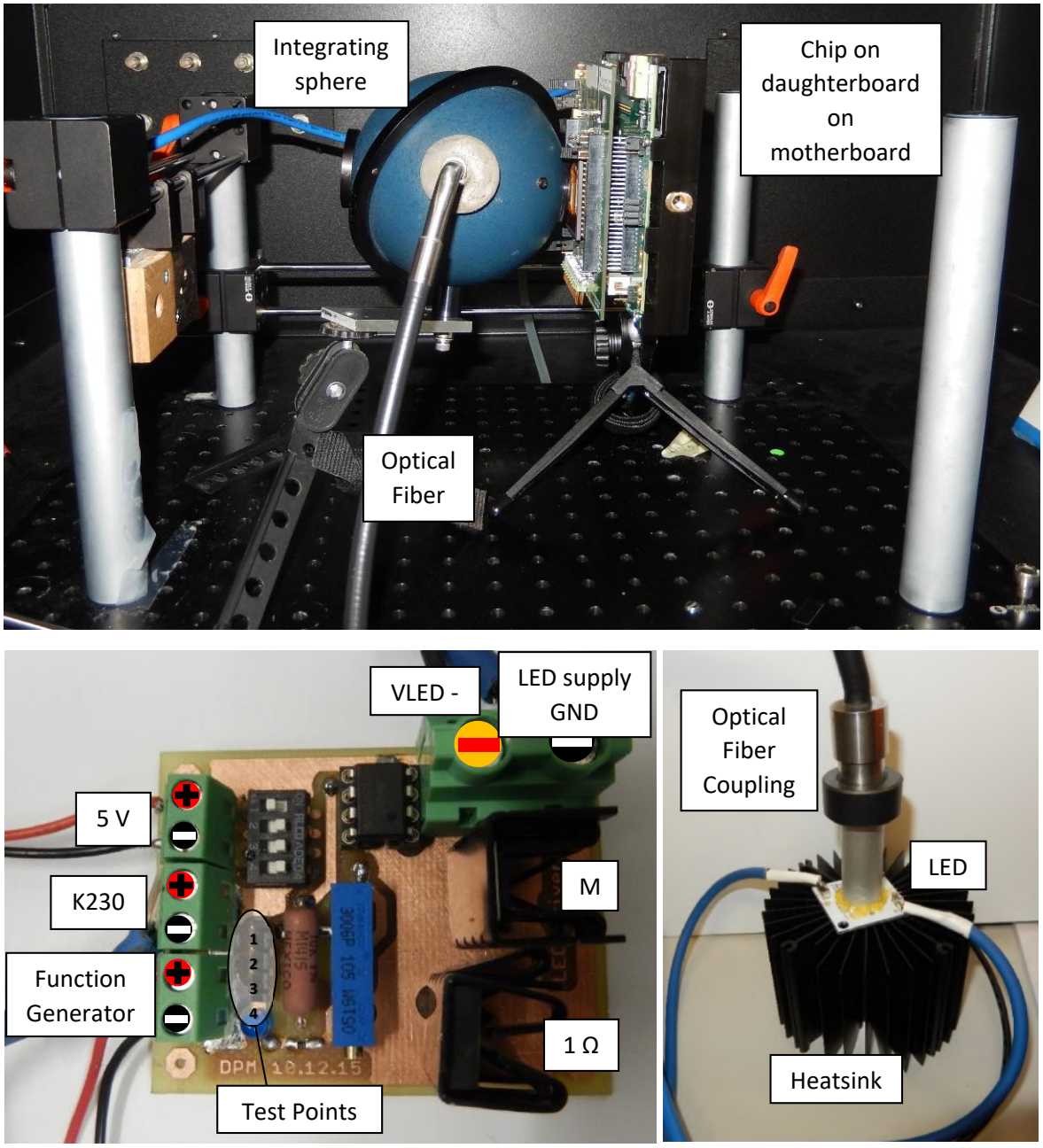


*Fig. 79 Top: Integrating sphere projecting light on the chip. Bottom left: PCB of the LED driver (all switches should be on the ON position, to the right, but currently the top switch is OFF, to the left, to turn off the blue LED to take a picture of the PCB). Bottom right: LED on heatsink with optical fiber coupler.*

## 6.3. LED characterization

The Cree Inc. MHDGWT-0000-000N0HM230GCT LED was tested with the setup of Fig. 76 and the photometer, set to white illuminance detection (head with white half-sphere diffuser). The photometer was positioned right in front of the integrating sphere output in a dark environment (dark box in dark room). $V_{FG}$ was set to zero and only the DC component $V_{K230}$ was swept from zero to the value at which the current of the LED would saturate. This saturation value happens to be 232 mA for the chosen $V_{ddLED}$ of 40 V.

For R1 = 100 Ω and R2 = 10 kΩ, therefore with a resistive divider ratio of 0.99, $V_{K230}$ was swept and the current across the LED was measured. This allowed to estimate, by averaging over the multiple data points obtained, the real value of the 1 Ω resistor in series with the LED. This turns out to be 1.104 Ω and the slight increase is due to the added milliOhm resistance of the cables connecting the LED to the PCB and the on resistance of the power MOSFET M. For every current value, the irradiance in mW/m$^2$ was measured as well as the illuminance in lux. The ratio of illuminance over irradiance for this particular LED, was also found to be 485.987 changing by ±2 over trial. This allowed to estimate either illuminance or irradiance. Measurements of irradiance were also performed with the resistive divider ratio of the DC component set to be 50 and 100 after tuning R1 to be 100 kΩ and 500 kΩ respectively (measured with the ohmmeter). These measurements were performed apart in time to allow to check for consistency in LED irradiance. The results are reported in Fig. 80 and Fig. 81.

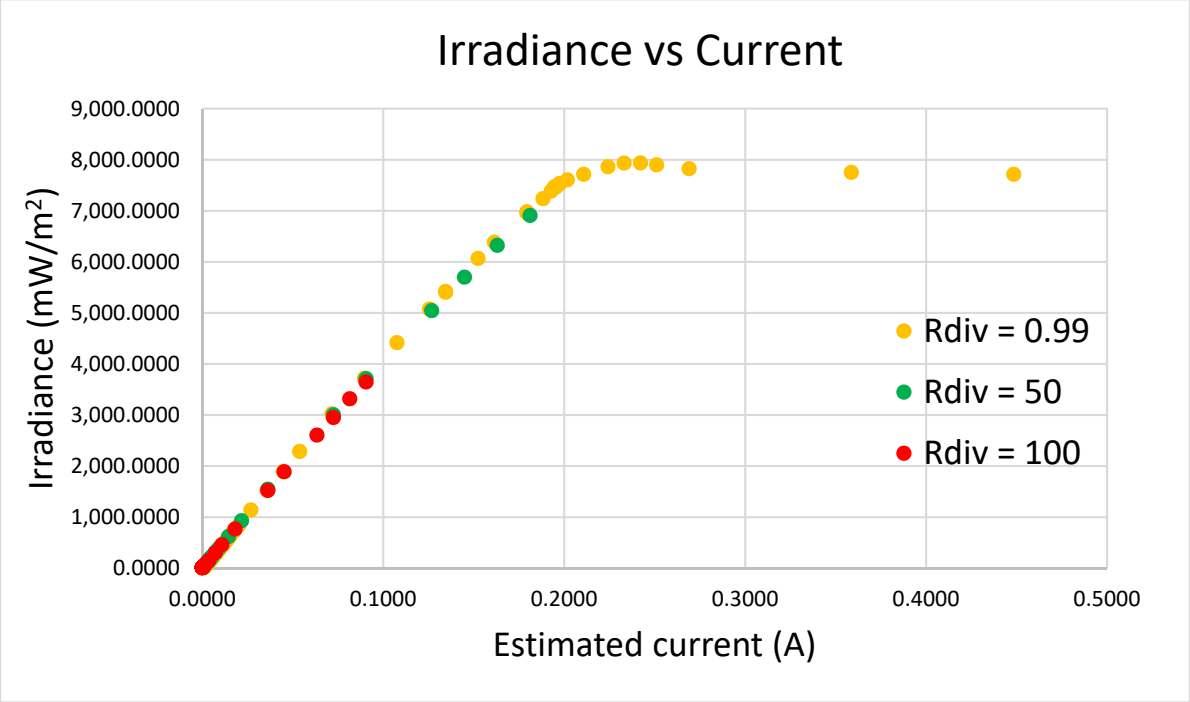


*Fig. 80 Irradiance vs current.*

As can be seen in all 3 cases, the measured irradiances versus estimated current (with known LED resistor) are consistent with each other. Three regions of operations can be distinguished. The first is the pre-threshold region (not-visible on a linear scale) where the irradiance is about 10 mW/m$^2$ until the current in the LED is about 50 µA. The linear region of the LED constitutes the second region and the one of interest. This extends from 50 µA to 200 mA with a consistent slope of about 40 W/m$^2$/A and intercept of about 15 mW/m$^2$ for the irradiance chart of Fig. 80 and a slope of about 19 klux/A and intercept of about 6.14 lux for the illuminance plot of Fig. 81. Finally, around 200 mA the saturation region occurs. The curve in the plot bends and reaches 232 mA for the given $V_{ddLED}$ of 40 V. As can be seen in Fig. 80 and Fig. 81, no points for the resistive divider ratios of 50 and 100 reach saturation, this is because very large $V_{K230}$ values are needed to do so.

In order to reach millilux values of illuminance, Neutral Density Filters (NDF) from Edmund Optics are used to attenuate the light coming out of the integrating sphere. NDFs of attenuation factors ranging from $10^{0.2}$ to $10^{3}$ were used in the experiments.

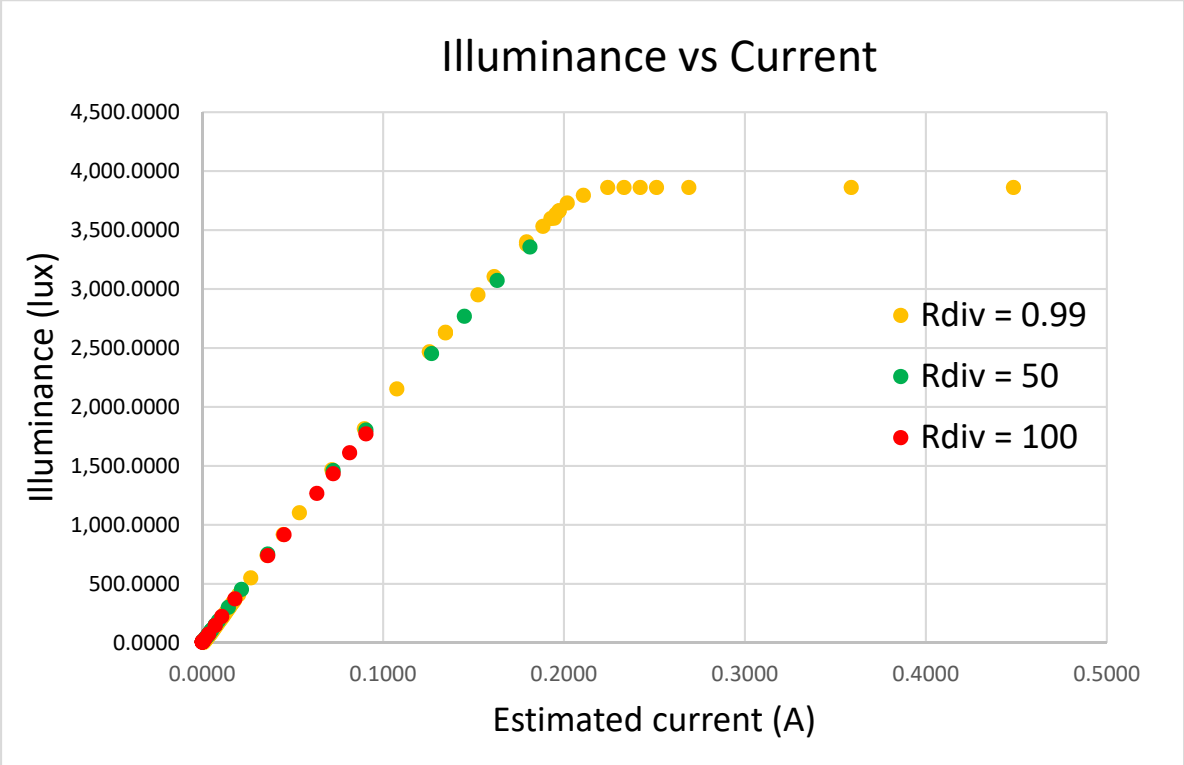


*Fig. 81 LED illuminance as a function of its current.*

## 6.4. Latency measurements setup

The estimate the sensor's latency it is necessary to stimulate only a small patch of the sensor in order not to saturate the sensor's bandwidth (by saturating the AER bus). To do so, a setup with a 5 mm LED is used to stimulate the sensor. The sensor is positioned at a 20 cm distance from the LED and it is focused on such distance with a lens. The LED can be flashed by modulating its power supply $V_{LED}$ from 0 to 5V which in turns modulates the current across it, reduced by a resistor of value 850 Ω. A circuit board was implement with 5 different LEDs of color white, yellow, red, green and blue, each current limited to a few milliAmperes by a 850 Ω resistor for a 5 V supply. A schematic of the circuit can be seen in Fig. 82. Each LED can be turned on with a switch independently. The setup also has a slider on which Neutral Density Filters can be mounted and interposed to reduce light intensity. These also tend to collimate the light and reduce the spread of the LED beam. A picture of the setup can be seen in Fig. 83 where hooks for a metal slider (to which NDF filters are taped to) are also present.

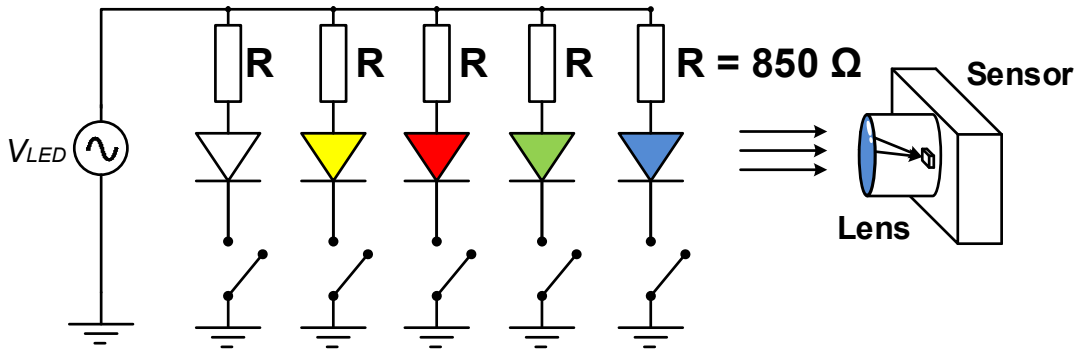


*Fig. 82 Schematic of the circuit board for latency measurements.*

To ensure that the latency measurements are not affected by the switching speed of the LEDs, the time constant $\tau_{LED}$ of the circuit can be calculated as:

$$\tau_{LED} = RC \tag{6.1.5}$$

where $R$ is the resistance limiting the LED current (850 Ω) and $C$ is the parasitic capacitance of the LED junction (the depletion capacitance) which can be found on the datasheet of the component (35 pF). The calculated $\tau_{LED}$ is 29.75 ns and is therefore negligible compared to the microsecond resolution of the DVS timestamping. The problem with this setup is the non-uniformity of the brightness of the LED (stronger in its center), but this can be fixed in by selecting the area of interest.

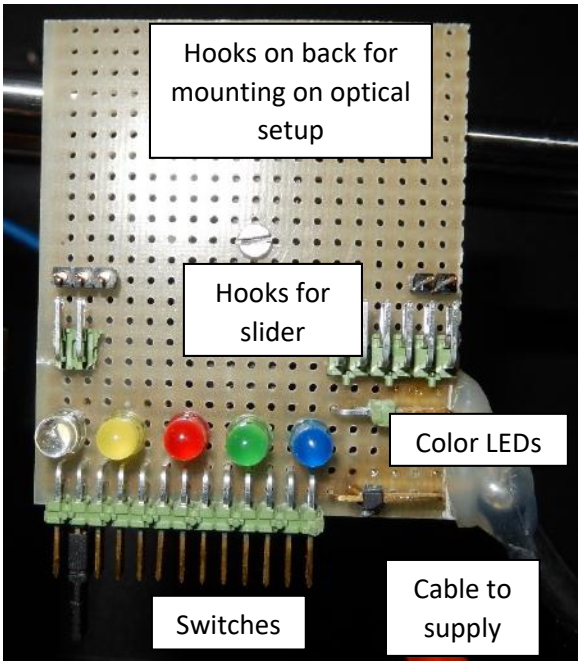


*Fig. 83 Picture of the circuit board for latency measurements.*

As an alternative to this setup, the optical fiber output of Fig. 76 can be pointed at the sensor directly from a distance without the integrating sphere. The setup however needs NDF filter of attenuation of at least $10^3$ to avoid the large spread of the beam.

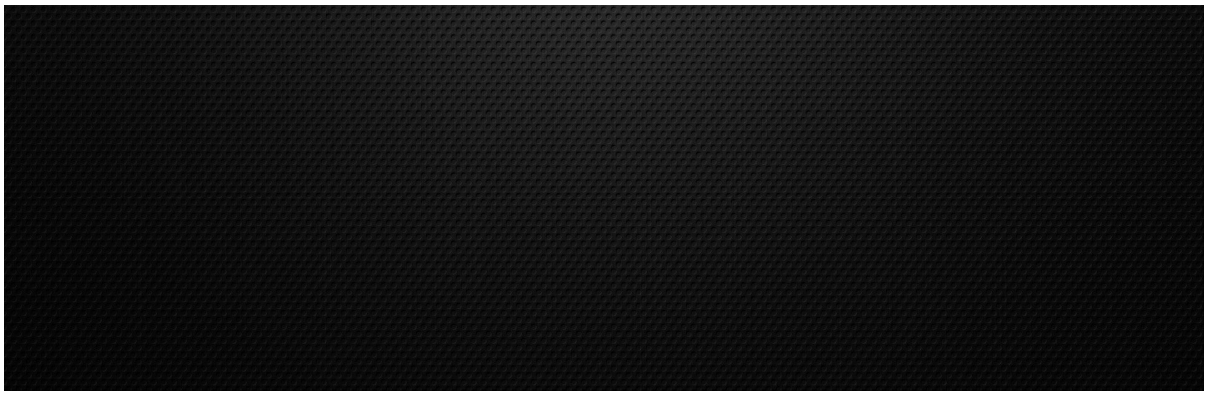

Image adapted from [173]

# Chapter 7: APPENDIX II

## 7. APPENDIX II: Dark current study attempt in photodiodes

The purpose of the work described in this appendix, was to estimate the dark current, and therefore the quality, of the different possible types and shapes of photodiodes employed in our wafer-run with Towerjazz Israel, in the CMOS Image Process at 180 nm. The initial measurements were not successful, but this could either be because of design flaws (of the chip hosting the photodiode array and a layer error in the surface photodiodes) or because of the primitive setup used for measurement. A new measurement attempt should be made soon to evaluate again the possibility to measure the structures, as soon as time allows it. However, the design of the structures and their layout can still be used in the future. The theory included in this section is also relevant for chapter 4.

## 7.1. Origin of dark current

Dark current is a current that flows in the photodiode even in the absence of light. This current is negligible in the forward-biased photodiode (photovoltaic mode, the one used to generate power) where diffusion current dominates. However, it is particularly evident in the reverse-biased photodiode (photoconductive mode, used for detection), where drift current is dominant.

Dark current is, first of all, partly caused by the random diffusion current of majority carriers which have enough velocity to cross the depletion region. Secondly, dark current is also composed by the drift current that randomly occurs when there are slight variations in the balance in the thermal generation-recombination of carriers within the depletion region [174][175]. It is also caused by background radiations, and especially by defects in the semiconductor's lattice [176]. These defects create trap states below the conduction band that allow valence-to-conduction-band transitions (and therefore current) more easily. Finally, a small part of the dark current occurs because of tunneling, if the reverse bias is strong enough to allow a band-to-band transition from the valence band of the p-substrate to the conduction band of the n-implant forming the photodiode.

Dark current is also dependent on doping concentrations and on the semiconductor's doping profile. It is one of the main limitations on a photodiode's performance as it reduces the sensitivity of the detector in low light conditions by masking the actual signal. It is usually measured in electrons/seconds ($e^-$/s) and it is because of its importance that it was thoroughly investigated in this experiment.

The dark current $I_D$ also contributes to an increase fixed pattern noise (FPN) and temporal noise. The latter is due to thermal and shot noise (the noise resulting from the discrete nature of current and its Poissonian distribution). These two added noises are reported below in (7.1.1) and (7.1.2):

$$i_{sD} = \sqrt{2qI_DB} \tag{7.1.1}$$

where $i_{sD}$ is the shot noise root mean squared (rms) current, $q$ is the electron charge constant, $I_D$ is the dark current and $B$ is the measurement bandwidth (the bandwidth of the noise).

$$i_{sT} = \sqrt{\frac{4k_BTB}{R}} \tag{7.1.2}$$

where $i_{sT}$ is the thermal noise rms current, $k_B$ is the Boltzmann's constant, $T$ is the temperature and R is the resistance across which this current flows (which also includes the internal resistance to the photodiode itself).

Dark current also defines the breakdown of the diode: the reverse breakdown voltage is as a matter of fact defined as the reverse-bias voltage for which the dark current is 10 μA [177].

## 7.2. Factors investigated

Various photodiode test structures were devised to observe the effects of the type of photodiode (buried or surface), of the shape and of the size, and of the presence of Salicide on the photodiode. The effect of leakage at the reset transistor's source to bulk junction was also investigated together with the effect of temperature.

### 7.2.1. Surface Photodiode

The first type of photodiode examined is the surface photodiode (SPD). This type of photodiode consists of n-doped implant (the same shallow active implant that would be used for an nFET, not an n-Well) into the p-doped substrate. Since the p-substrate is grounded by default, the photodiode is reverse-biased as long as its n terminal has a voltage positive or larger than 0 V. The sketch of an SPD is shown in Fig. 84:

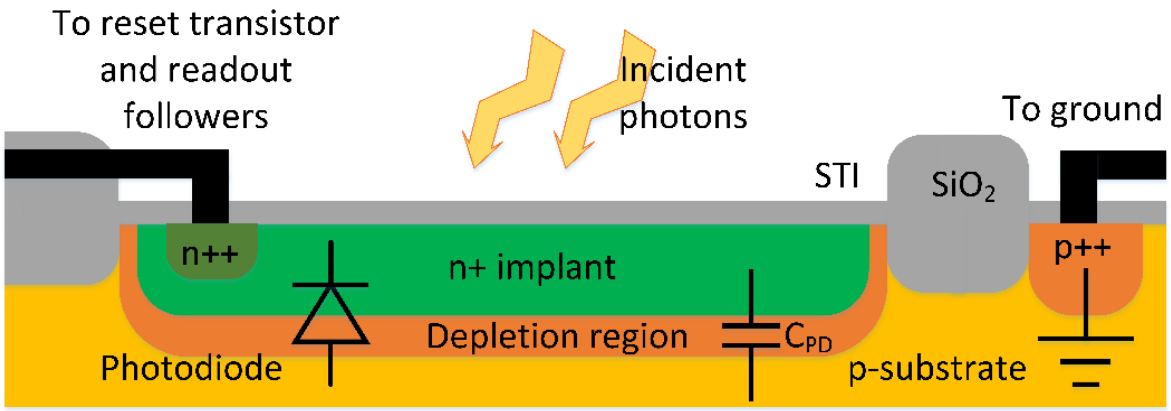


*Fig. 84 Surface photodiode cross-section.*

As can be seen there are contacts through the thin silicon dioxide layer that reaches the metal interconnect. The silicon dioxide is transparent and does not affect the quantum efficiency of the photodiode too noticeably. This type of photodiode is optimized for visible light as the junctions' depth is about 0.6 µm (with corresponding depletion region about 1 µm thick for no applied bias) and corresponds to such wavelength.

The reason why this diode, practically the simplest of its kind, was replaced in some of our chip designs by a buried photodiode is the worst dark current performance (1000 $e^-$/s, without salicidation, rather than 300 $e^-$/s at room temperature for a 10 µm x 10 µm photodiode [178]). This is because dark current is dependent on lattice imperfections. These are particularly abundant at interfaces of materials with different lattice constants (and hence structures) and in our case, at the interface between the junction of doped silicon and silicon dioxide [179]. This is particularly evident in the proximity of the shallow trench isolation (STI) where silicon dioxide was inserted after an etching in the silicon. Certain photodiodes' designs even propose an increase in distance between STI and n+ implant, with the result of a reduction in dark current [180].

### 7.2.2. Buried Photodiode

The buried photodiode (BPD), as the name implies, consists of an n-type implant buried just beneath the surface (the silicon dioxide interface). This is obtained by a first deep ion implantation of n-type and a by a second shallower p-type ion implantation cover up the top part of the n-implant (to bury the first implant). This in turn gives a lower dark current [181], as mentioned in the previous section, because of the silicon-silicon interface (although oppositely doped), the longer distance with STI and the larger pn junction surface (both above and below the implant) to capture the incident photons. Since the pn junction is deeper, longer light wavelengths can also reach it: the photodiode

is therefore not only optimized for visible light but for near infrared (NIR) light too. An example of BPD is shown in Fig. 85:

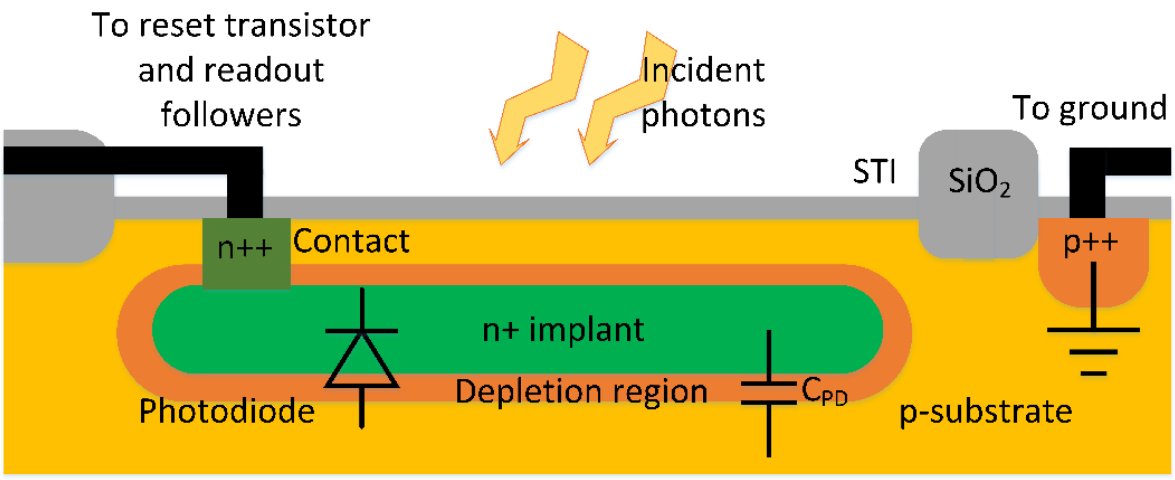


*Fig. 85 Buried photodiode cross-section.*

### 7.2.3. Salicidation and non-salicide block

Salicidation is a process step in the fabrication of semiconductor devices that is used to allow the formation of low-resistance electrical contacts between the semiconductors' active silicon diffusions, polysilicon gate and the metal interconnect. The resistance obtained is, as a matter of fact, even lower than the one that could be obtained by heavily doping the silicon. To quantify, a salicided contact has about 5 Ω of resistance, while a non-salicided contact has a resistance varying between 100 and 500 Ω according to the manufacturer. The doping of polysilicon, which becomes "clad poly", is also common for high-speed CMOS applications [182][183].

Salicidation involves the annealing of transistors' active regions with a thin sandwich of deposited metal: usually at direct contact with platinum (to enable a good electrical connection), subsequently covered with a refractory barrier metal such as tungsten (to provide resistance to heat and to electromigration wear). Finally, a last layer of copper-doped aluminum thin film is also usually included for lower and more stable resistance with the above metal interconnect.

The annealing reaction is made possible by heating the silicon wafer. The final result is a transition metal silicide, and since the metal only anneals with active regions only (and not with the surrounding areas covered in silicon dioxide $SiO_2$), the process is defined as self-aligned and hence the term Self-Aligned siLICIDATION: salicidation.

Since salicidation is an opaque layer, this process can be avoided on top of photodiodes thanks to a non-salicide (NS) mask. This results therefore in an increase in quantum efficiency of the photodiode, as more photons are capable to reach the pn junction and hence generate current.

NS also serves the purpose of reducing dark current [184] as the silicide's surface roughness and imperfections can increase it. It is also employed on sensitive nodes to reduce the effect of unwanted (dark) current. An example, in our case, is the reset transistor in the differencing amplifier of the DVS [21]. The introduction of NS does however introduce a larger resistance and therefore reduce the

device bandwidth (by increasing the overall RC time constant). NS must be then only present on nodes non critical to speed.

The effect of NS was investigated through the comparison of the dark currents of SPD and BPD with and without this mask.

### 7.2.4. Shapes

Since dark current is proportional to imperfections in the lattice, the measure of area and perimeter directly affect its magnitude [180]. More area involves more (imperfect) interface, but also means more fill factor (the parameter that defines the percentage of area of the pixel occupied by the photodiode), which in turns means more signal-to-dark-current ratio. More perimeter for the same area instead, for long photodiodes for example, means more interface at no extra gain in fill factor and therefore more dark current.

To study the effect of the perimeter-to-area ratio different photodiodes, for each BPD, SPD and salicidation combination, different structures were devised. First of all, a photodiode with minimum size, the minimum allowed by the Design Rule Check (DRC) of Cadence, was set up. Then, photodiodes multiples of this minimum length were also designed to cover up to four times the size and therefore, up to four time the perimeter-to-area ratio (once, twice, thrice and four times respectively). Since these photodiodes are really small, so is their individual dark current. For this reason, multiple same-size photodiodes are shorted together. More on the physical circuit design can be read in the section 7.3.

Finally, dark current is also proportional to the sharpness of the photodiodes' corners [185][186][187] (an accumulation of defects in the surface). To show this, for all photodiode types, a square, of 90° corners, and an octagon, of 135° corners, both of arbitrary shape but with the same perimeter-to-area ratio were devised. The octagon has also a slightly smaller area: if the dark current is still smaller with this type of photodiode this would show that rather than trying to maximize fill factor by adapting the shape of the photodiode to cover all the available space, it would be more convenient to aim at rounding the photodiode at the expense of its overall area instead.

### 7.2.5. Temperature

Drift current (and therefore dark current) is very temperature-dependent and doubles on average every for every temperature increase of 8-10°C [177][181][188][189]. This number is however a very rough estimate as it depends on other process-dependent factors, such as substrate thickness and resistivity, and electrical functions, namely at what reverse-bias voltage the photodiode is kept. This is due to the fact that electrons in the valence band become excited and are able to jump to the conduction band, effectively increasing the current and therefore shot and thermal noise.

### 7.2.6. Transistor leakage

Transistor leakage defines a current that flows even when the device is turned off and its junctions are reverse biased [190]. In an integrating photodetector this is a problem for the reset transistor (as it will be seen later in Fig. 88) on top of the photodiode. Its purpose is to act as an ideal switch and short Vdd to the diode in the reset phase, and to be an open circuit in the integration phase, when the signal is being collected. The problem is however that if the transistor leaks, during its off phase, then the integrated signal can be corrupted.

In our case the leak of interest is not the drain to source leak but the source to bulk (substrate). This leak can effectively decrease the photodiode's voltage, giving a false signal reading. To investigate such effect, a signal readout without a photodiode but just the reset transistor was set in place, to test the magnitude of such leak. To compare, another identical circuit with just the reset transistor was set with the difference of non-salicidation of such transistor. Since salicidation introduces more leakage current (due to surface defects), the expected leakage should be inferior to that of the salicided MOSFET.

### 7.2.7. Relative Quantum Efficiency

Although their minimal area, the test structures examined are still photodiodes and their quantum efficiency (QE) should be evaluated. The quantum efficiency is defined as the ratio between incident photons and electrons converted into current. It is therefore always less or equal (in the ideal case) to unity and varies as a function of wavelength. The latter is because of the wavelength-dependent absorption of the material. The measurements for QE will however be greatly affected by the fact that the contact to the photodiode itself covers great part of the photodiode's surface, especially in the case of minimum size, therefore reducing the light falling onto it. A rough proportion of photo-current through the different sizes should still be observed in any case.

## 7.3. Overall circuit

Photodiode test structures are organized in four large arrays at the bottom left corner of the AER Correlation filter chip [191]. The structures were kept as isolated as possible from the rest of the chip, apart from power ring connections and multiplexed analogue readouts. The four arrays cover the four combination of SPD, BPD and salicide block. Each of these array is composed by 336 cells, each including the following test photodiodes fitted together:

1. Two minimum-size photodiodes allowed by the process;
2. Two photodiodes with height equal to the minimum size and width equal to the double of it;
3. Two photodiodes with height equal to the minimum size and width equal to the triple of it;
4. Two photodiodes with height equal to the minimum size and width equal to the quadruple of it;
5. One Square photodiode (whose size is defined by the space available only);
6. One Octagonal photodiode (whose size is also only defined by the space available).

All of the above photodiodes are then enclosed by a substrate-ground contact. An example of four butted cells is shown in Fig. 86 for (BPD) and Fig. 87 (SPD). When arrayed together, photodiodes of the same type are shorted together, so as to increase their effective area, perimeter and, otherwise too small, dark current. Their connections are made by minimum width metal 1 and metal 2, however they occur both vertically and horizontally at every cell, therefore reducing the overall resistance of the paths.

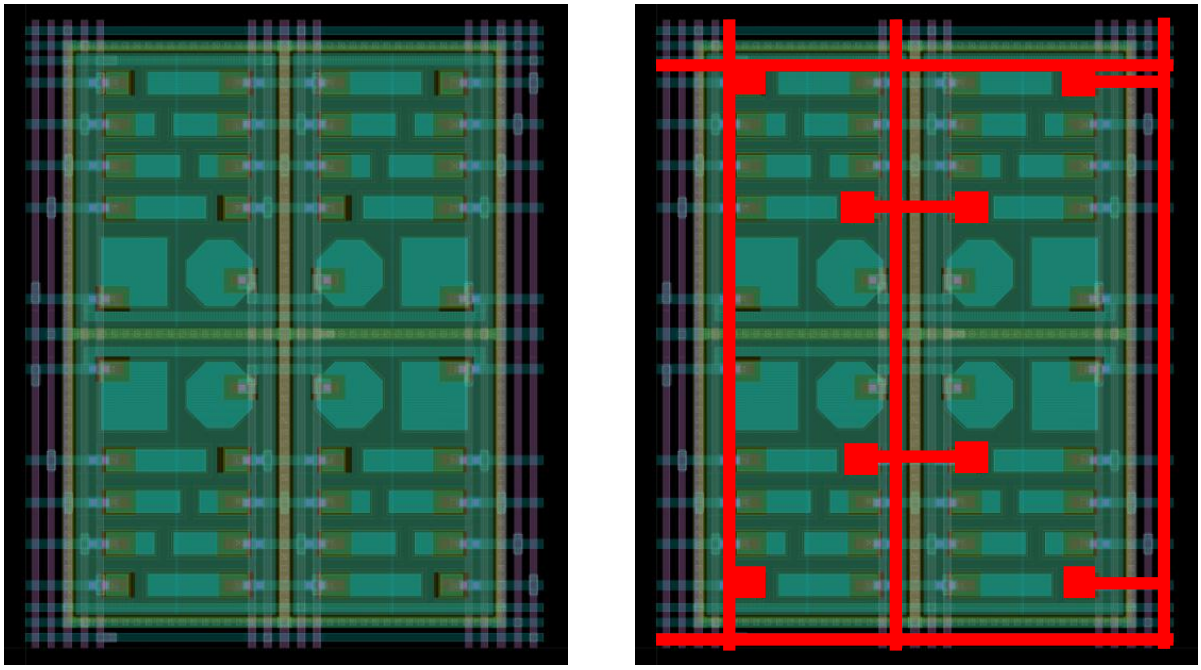

*Fig. 86 Left: Buried photodiode (BPD) cell of 2 by 2 subunits containing all shapes of photodiodes. Right: highlighted connections of the unit size BPD.*

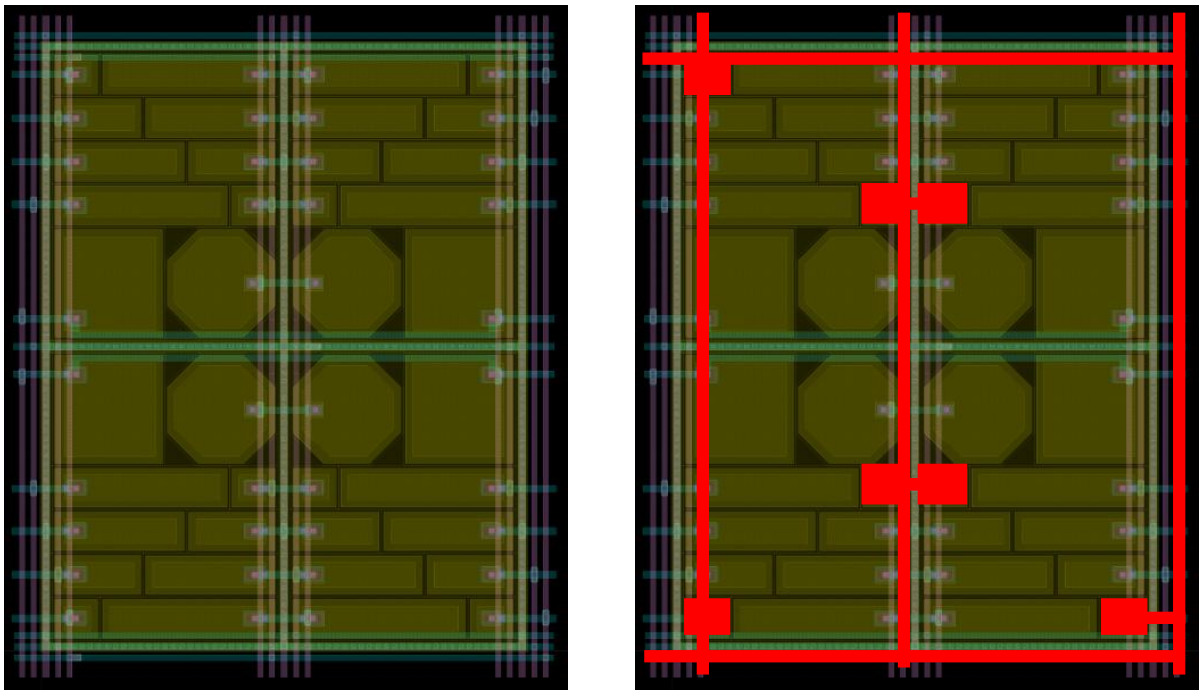

*Fig. 87 Left: Surface photodiode (SPD) cell of 2 by 2 subunits containing all shapes of photodiodes. Right: highlighted connections of the unit size SPD.*

Table 3 summarizes the test structures characteristics.

| Photodiode size | 1x1 unit | 1x2 unit | 1x3 unit | 1x4 unit | Square | Octagonal |
|---|---|---|---|---|---|---|
| SPD in parallel | 672 | 672 | 672 | 672 | 336 | 336 |
| SPD size (single) W/L or side length | 2.45 µm /2.20 µm | 4.89 µm / 2.20 µm | 7.34 µm / 2.20 µm | 9.78 µm / 2.20 µm | 6.12 µm | 2.55 µm |
| SPD area (single) | 5.37 | 10.73 | 16.10 | 21.47 | 37.45 | 31.40 |
| SPD area (total) | 3606.47 | 7212.95 | 10819.42 | 14425.89 | 12584.68 | 10549.34 |
| SPD perimeter (single) | 9.28 | 14.17 | 19.06 | 23.95 | 24.48 | 20.40 |
| SPD perimeter (total) | 6236.16 | 9522.24 | 12808.32 | 16094.40 | 8225.28 | 6854.40 |
| SPD perimeter/area | 1.73 | 1.32 | 1.18 | 1.12 | 0.65 | 0.65 |
| BPD in parallel | 672 | 672 | 672 | 672 | 336 | 336 |
| BPD size (single) W/L or side length | 1.07 µm / 1.07 µm | 2.14 µm / 1.07 µm | 3.21 µm / 1.07 µm | 4.28 µm / 1.07 µm | 2.94 µm | 1.21 µm |
| BPD area (single) | 1.14 | 2.29 | 3.43 | 4.58 | 8.61 | 7.07 |
| BPD area (total) | 769.37 | 1538.75 | 2308.12 | 3077.49 | 2894.38 | 2375.28 |
| BPD perimeter (single) | 4.28 | 6.42 | 8.56 | 10.70 | 11.74 | 9.68 |
| BPD perimeter (total) | 2876.16 | 4314.24 | 5752.32 | 7190.40 | 3944.64 | 3252.48 |
| BPD perimeter/area | 3.74 | 2.80 | 2.49 | 2.34 | 1.36 | 1.37 |

*Table 3 Test structures information. "Unit" defines the minimum size of the photodiode allowed by the process.*

Each type of photodiode of each of the four arrays (and therefore of each SPD, BPD and salicidation combination) is then connected to the n-type thick-gate reset transistor $M_{rst}$ (as can be seen in Fig. 88) which can effectively short the 3.3 V analog power rail to the photodiode. This transistor is very short but also very narrow (W = 0.42 µm and L = 0.35 µm). The latter means that is in not possible to test for the effect of salicidation on speed. This reset transistor cannot in fact supply a reset current to the entire array instantaneously and can therefore heavily influence the speed measurements. The reset signal itself is provided externally via a low-resistance pad.

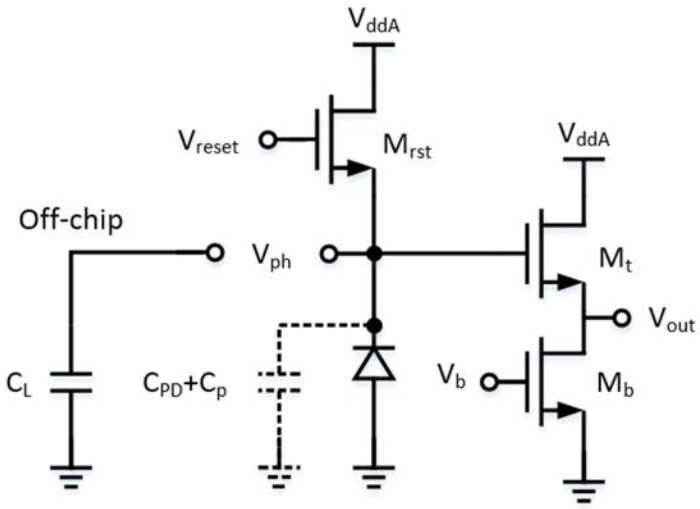


*Fig. 88 Read out circuit for each type of test photodiode. The off-chip capacitance is only connected to the four photodiodes whose node $V_{ph}$ is available via a low-resistance pad.*

A dedicated n-type thick-gate transistor source follower readout pair composed by transistors $M_t$ (W = 6.15 µm and L = 0.35 µm) and $M_b$ (W = 2 µm and L = 2 µm), is then used to read the integrated photodiode voltage value. This voltage then feeds into one of the chip's four analog multiplexers. The source-followers' bias voltage is provided externally directly via a low-resistance pad.

There are in total then 24 source followers' readouts for the test photodiodes (one for each BPD, SPD and salicidation combination) plus two source followers readouts that connect to two isolated reset transistors' sources. This was done to allow the readout of the transistors' leakage and to check for any improvement with the salicidation of $M_{rst}$. The overall circuit block diagram is illustrated in Fig. 89 and its layout counterpart in Fig. 90.

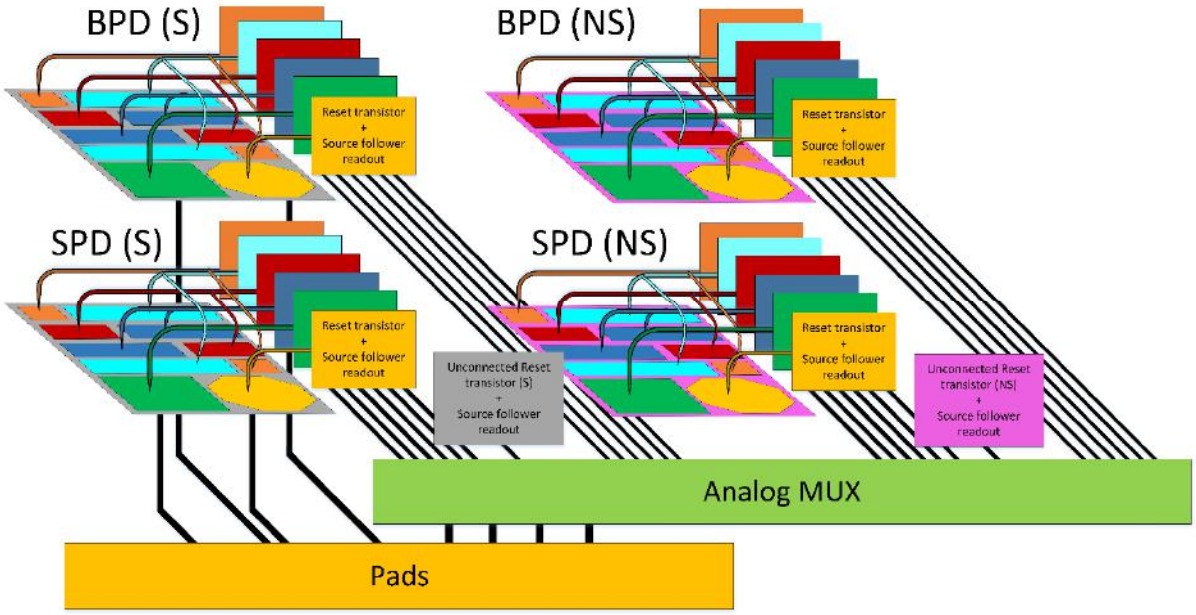


*Fig. 89 Overall circuit showing four cells of the four arrays (BPD and SPD with salicidation (S) and BPD and SPD non-salicided (NS)) and their 24 outputs going to the analog multiplexer (whose four outputs then go to pads). Two multiplexer inputs also come from two (NS and S) unconnected reset transistors. Four salicided BPD and SPD diodes are also connected directly to pad.*

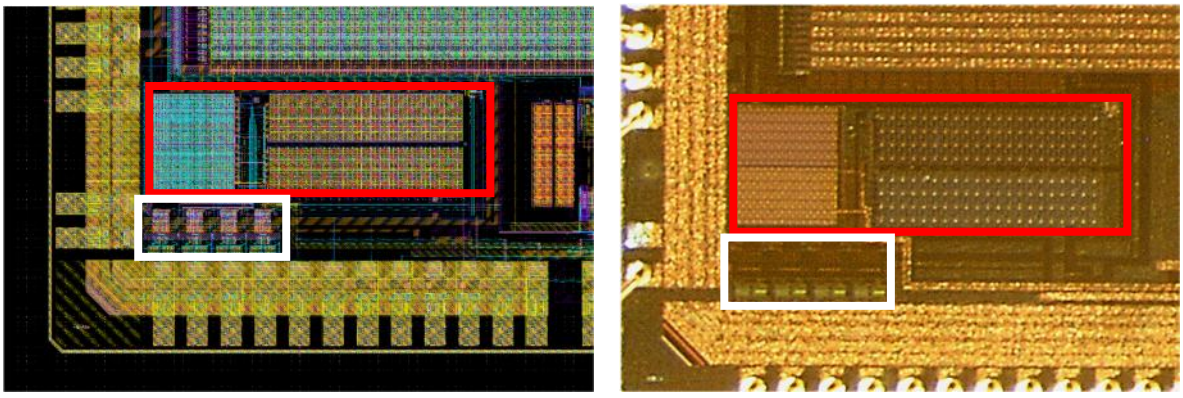

*Fig. 90 Overall circuit layout in Cadence (left) and photographed layout (right), at the bottom-left of the AER Correlation chip array. The photodiode array is highlighted in red and the analog multiplexers are highlighted in white.*

The chip configuration chain allows to input bits to switch between all of the multiplexers' inputs. This can be achieved through MATLAB code sent via USB to the USBtoMini Board developed by R. Berner [192].

## 7.4. Measurement setup

The dark current must be directly measured to estimate its magnitude. It is not possible to obtain its value from the source-follower readout. This is because, for the circuit shown in Fig. 88 the dark current $I_D$ is dependent mainly on the photodiode's large capacitance $C_{PD}$ as shown in (7.4.1):

$$I_D = C_{tot}\frac{dV_{ph}}{dt} = (C_p + C_{PD})\frac{dV_{ph}}{dt} \approx C_{PD}\frac{dV_{ph}}{dt} \tag{7.4.1}$$

where $C_{tot}$ is the total capacitance at the node $V_{ph}$ that includes the parasitic capacitance of the readout source follower transistor $C_p$ and the depletion capacitance of the photodiode $C_{PD}$.

As a matter of fact $C_{PD}$ is dependent on the voltage across it ($V_{ph}$, which determines the depletion region thickness), following (7.4.2) [176], which in turn is dependent on the dark current.

$$C_{PD} = \sqrt{\frac{q\varepsilon_S}{2(\varphi - V_{ph})}\frac{N_a N_d}{N_a + N_d}} \tag{7.4.2}$$

where $q$ is the electron charge, $\varepsilon_S$ is the absolute permittivity of the material, $\varphi$ is the diode's built-in potential barrier, $N_a$ is the acceptor's density level in the p-type side of the diode and $N_d$ is the donor's density level in the n-type side.

Furthermore, $C_{PD}$ is also dependent on area and perimeter as the electric field at the edges of the depletion region is fringing and therefore not homogeneous like in the center of the depletion region. On average $C_{PD}$ can therefore also be described by equation (7.4.3):

$$(C_{PD})_{avg} = C_{const} + C_A A + C_P P \tag{7.4.3}$$

where $(C_{PD})_{avg}$ is the average photodiode capacitance, $C_{const}$ is a constant capacitance (varying on the design and in the order of a few fF/µm), $C_A$ is the capacitance per unit area (this should be 0.47 fF/µm$^2$ for the BPD, according to the manufacturer), $C_P$ is the capacitance per unit perimeter (this should be 0.085 fF/µm for the BPD, according to the manufacturer), and $A$ and $P$ are area and perimeter of the photodiode respectively.

To avoid this dependence, the photodiode nodes $V_{ph}$ of 4 photodiodes (the octagonal and square BPDs and the octagonal and square SPD, all with salicidation) were brought to low-resistance pads, so as to connect them to a known capacitance $C_L$, large enough to dominate $C_{PD}$. Equation (7.4.4) then becomes:

$$I_D = (C_p + C_{PD} + C_L)\frac{dV_{ph}}{dt} \approx C_L\frac{dV_{ph}}{dt} \tag{7.4.4}$$

This way, the dark current $I_D$ can be fully evaluated as the voltage derivative can be measured and $C_L$ is known. Alternatively, for a very precise measurement of the current, $V_{ph}$ can be fixed by the

Keithley 236 source-measure unit. The device can in turn measure the current (equal to $I_D$) that it provides to keep $V_{ph}$ fixed. With the direct access to $V_{ph}$, the source-follower readout characteristics can also be evaluated. As regards all other photodiodes without direct access to their $V_{ph}$, only their relative relationships can be measured as no dedicated low-resistance pad was available at design time.

## 7.5. Problems and time constraints: measurement failure

The testing of the photodiodes test structures was performed on the hand-made chip-holder of Fig. 91 as the mother and daughter boards to host the AER Correlation filter of [191] were not yet available. Only the useful pins were connected to a pot-box.

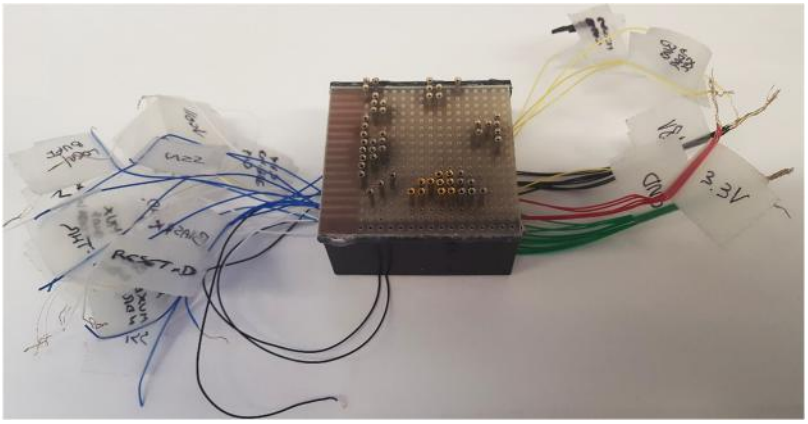


*Fig. 91 Initial testing platform for the photodiode test structures.*

The photodiode currents measured with the Keithley 236 source-measure unit were in the order of hundreds of microAmperes, no matter under which lighting conditions. Even the reset transistor could not reset the photodiode's voltage. Something was clearly wrong. This could be caused by the almost total absence of deep n-well contacts to Vdd in the AER Correlation filter. This would make unwanted diode junctions turn on far away from the Vdd contact area (two contacts are even 3 mm away on the 5 x 5 mm chip). Another reason for the large leakage could be that the biases of the AER Correlation filter are not set and the chip could be in a high-current state. With this very basic setup the biases could not be set. This could be done now with a mother and daughterboard for the AER Correlation filter. Unfortunately, for a question of time, this has not yet been done.

Further inspection of the photodiode arrays revealed that there was a mistake in the layout for surface photodiodes. I have misinterpreted the layer PD3 as a contact layer: it is instead the layer making the photodiode implant n-type. However this does not explain the large leakage current observed (it should just make a very small diode since PD3 is still there but just in minimum quantity).

With this knowledge and documentation, I hope a student, or myself, will continue this work soon.

## Curriculum Vitae

# DIEDERIK PAUL MOEYS

### Contact and personal information


| | |
|---|---|
| *Address:* | Affolternstrasse 103, CH-8050 Zürich |
| *Cell Phone:* | +41 76 736 6670 |
| *Email:* | diederikmoeys@live.com |
| *Date of Birth:* | 31.12.1990 |
| *Place of Birth:* | Annecy (France) |
| *Nationality:* | Dutch and French |


### Skills

*Expertise:* **Mixed-signal integrated circuit design**: focus on vision and auditory sensors (5 years).
Digital and software design of architectures for object **tracking**.
Machine learning: **Convolutional Neural Networks**.

*Computer skills:* **Cadence**, **TSPICE**, **ADS** (RF design), **CST** (Microwave Suite), **MultiSim**, **DipTrace** and **EAGLE** PCB software; **JAVA**, **MATLAB**, **Python**, **Caffe** (Deep Learning Framework), **VHDL**;
**ECDL** (European Computer Driving License).

*Languages:* **English**: full proficiency. **French**: mother tongue; **Italian**: mother tongue; **German**: Fluent (with basic understanding of Züritüütsch).

### Education and Awards

*Awards:*

- **2016 Institute of Engineering and Technology** (IET) **Postgraduate Scholarship** for excellence in research in the field of electrical, electronic or manufacturing engineering.
- **2013 Goldsmid Medal** for being the **top graduating student of the entire Engineering Faculty** of UCL.
- **2013 IET Prize** for being the **top graduating student of the entire Electronic Engineering department**.
- **2013 Dean's Top Graduate List for Excellence.**
- **2010 Goldsmid prize** for the best examination performance of the whole course.

*University Postgrad. 2013-2016:* **PhD, ETH Zürich, Switzerland**;
**Electronic Engineering**, Sensors' group of the Institute of Neuroinformatics (INI);
**International PhD Program in Neuroscience** (ZNZ). Defense on the 13/12/2016.

Chip-designer of a **high-contrast neuromorphic silicon retina** (Dynamic Vision Sensor, DVS) on a 180 nm wafer run, collaborating with **iniLabs Ltd**, under the EU project SeeBetter. Responsible for **dicing and bonding** of the wafers. Accurate sensors' characterization, setting measurement protocols for the field. Absolute Quantum Efficiency measurements performed at **IMEC** research institute, Belgium. Application of the sensor to **calcium imaging** of neural activity.

**FPGA** and JAVA-based bio-inspired processing of DVS data for the purpose of object detection and **tracking**, (VISUALISE project) at the RTC lab of the University of Seville and at the University of Göttingen.

Application of **machine learning (Convolutional Neural Networks)** to DVS data to **steer a predator robot** towards a prey robot, (VISUALISE and Neuromorphic Processor Project of **Samsung**). Successful trial runs performed during various visits to the University of Ulster. Wide media coverage including VICE and CNBC.

Testing of the **cochlear circuit** designed during the exchange at ETH Zürich.

Staff and participant at the Telluride **Neuromorphic Cognition Engineering Workshop 2015**. Worked closely **with IBM designers of the TrueNorth** chip to attempt using DVS data with a Neural Network.

**Supervisor of a master student's project** regarding the design of a high-gain DVS sensor.

*University Undergrad. 2009-2013:*

**Master of Engineering in Electronic and Electrical Engineering** (International Programme), **University College London (UCL), UK**
**First Class Honours MEng** with highest average across all years of **89/100**.
Average of third year **Erasmus Exchange at ETH Zürich**: **5.65/6**.

**Master project**: **hardware design of Quadrocopter** (power distribution and management system, safety circuits) and **design of an RF ambient radiation power-harvester (Rectenna)**. Strong team-leading abilities since being the group leader of 6 other different projects.

Projects undertaken at ETH Zürich regarding design and layout of a **Current-Mode Cochlear Resonator with Automated Q-Factor Control Loop** and of an **Active Pixel Sensor Camera Chip**.

*High-school 2004-2009:*

**Liceo scientifico Edith Stein**, Gavirate, Italy; *Esame di Stato Italiano* 2009: **97/100**.
**Oxford Advanced Study Program** 2008 in Oxford with results: Physics: A+ Mathematics: A.

## Employment History

*9/2013-9/2016:*

**Teaching Assistant of the ETH Neuromorphic Engineering course I and II**: introducing students from a wide range of backgrounds to analog circuit design, from basic semiconductor physics to the full chip design of the course's project. Responsible for hands-on laboratory sessions. Ability to explain complex concepts clearly and effectively with great feedback from students.

**Volunteer for presentations and demonstrations** to students visiting INI. Invited speaker at Liceo scientifico Gavirate and UCL to present research and talk to students about studying abroad. Volunteer at ETH Open Days and Scientifica 2016 (public dissemination in German).

*7-8/2015:* Staff at the Telluride **Neuromorphic Cognition Engineering Workshop 2015**.

*11/2014:* Staff at **IEEE Swiss Workshop on Memristive Devices and Neuromorphic Applications 2014**.

*6-9/2011:* **Designer** and tester of an analog data-acquisition PCB for benchmarking purposes, at **Whirlpool Corporation**, Cassinetta di Biandronno, Italy.

*6-12/2010:* **Mentor for students** in UCL: introducing students to university life and helping them with every subject, preparing them for tests and exams.

*3/2006-9/2009:* **Salesman** in ice cream shop and helper with deliveries to restaurants (dealing with customers and managing the shop in the owner's absence).

## Selected Publications

*Accepted:*

- "*Color Temporal Contrast Sensitivity in Dynamic Vision Sensors*", Diederik Paul Moeys et al., 2017 IEEE International Symposium on Circuits and Systems (ISCAS);
- "*Steering a Predator Robot using a Mixed Frame/Event-Driven Convolutional Neural Network*", Diederik Paul Moeys et al., 2016 IEEE Second International Conference on Event-Based Control, Communication and Signal Processing (EBCCSP);
- "*Retinal ganglion cell software and FPGA model implementation for object detection and tracking*", Diederik Paul Moeys et al., 2016 IEEE International Symposium on Circuits and Systems (ISCAS);
- "*Combined frame- and event-based detection and tracking"*, Hongjie Liu, Diederik Paul Moeys et al., 2016 IEEE International Symposium on Circuits and Systems (ISCAS).
- "*Current-mode automated quality control cochlear resonator for bird identity tagging*", Diederik Paul Moeys et al., 2015 IEEE International Symposium on Circuits and Systems (ISCAS).

## Interests

*Sports and interets:* Kayaking, Bouldering, Via Ferrata, Free-running, Snowboarding, Skateboarding, Black belt in Karate (2004), Hockey. Travelling, Cultivating carnivorous plants, Drawing.

## References

*ETH PhD Supervisor:* Prof. Dr. Tobi Delbruck, Head of Sensors Group at Institute of NeuroInformatics (INI), ETH Zürich.
Email: tobi@ini.uzh.ch, Telephone: +41 (0)44 635 30 38

*UCL Tutor:* Prof. Dr. Izzat Darwazeh, Head of the Communications and Information Systems Group, UCL London.
E-mail: i.darwazeh@ee.ucl.ac.uk, Telephone: +44 (0)20 7679 3945