VNIVERSITAT ̇ Q̈ ɪᴅᴇ VALÈNCIA

Development and Implementation of a Selective Change-Driven Vision Sensor for High Speed Movement Analysis

Submitted in partial fulfillment of the requirements for the degree of doctor of philosophy.

Doctoral program:

Tecnologías de la Información, Comunicaciones y Matemática Computacional

Presented by:

Pedro Diego Zuccarello

Directed by:

Dr. Fernando Pardo Carpio

and

Dr. José Antonio Boluda Grau

Valencia, February 2013

Resumen

Motivación y objetivos

Los sensores de imágenes tradicionales trabajan bajo unos principios muy simples y conocidos: el nivel de iluminación del entorno es muestreado y transmitido a intervalos de tiempo regulares; y todos los píxeles de la matriz, sin excepción, son transmitidos secuencialmente y en orden. Esto es así aunque no se hayan producido cambios en la escena bajo observación. Esto implica que una gran parte de la información que se genera y transmite puede ser considerada como redundante. En muchos casos esta estrategia es la más adecuada. Algunos ejemplos de ello son los escáneres, los sistemas de captura de imágenes para diagnóstico médico o los sistemas de video para entretenimiento. Todas estas aplicaciones necesitan la mayor cantidad de información posible sobre el entorno, aunque este no cambie o muestre variaciones muy pequeñas en intervalos de tiempo largos. Para otro tipo de aplicaciones, como los sistemas de visión artificial o las redes de sensores inalámbricas, la gran cantidad de información redundante que genera y transmite un sensor tradicional de imágenes puede convertirse en una limitación para la implementación de sistemas en muchos entornos reales.

Muchos sistemas de visión biológicos trabajan de manera completamente distinta a los sensores de captura de imágenes tradicionales. Una de sus principales características es que las celdas sensibles (el equivalente de los píxeles en tecnología de silicio) reaccionan de manera independiente y asíncrona a los cambios de iluminación.

Tomando como punto de partida los trabajos de C. Mead y M. Mahowald [1, 2] realizados a finales de los años 80, las últimas dos décadas han presenciado avances muy significativos en el diseño de sensores de visión, todos estos fundamentalmente orientados a transmitir y procesar solo la información considerada importante o relevante dentro de la escena bajo análisis. La mayor parte de estos diseños han tomado, en mayor o menor medida, el funcionamiento del sistema biológico de visión como base de sus desarrollos. El objetivo de muchos de los trabajos realizados en este área es imitar de la mejor manera posible, y mediante las más avanzadas tecnologías de silicio, el comportamiento de los sistemas biológicos en sus facetas visual, auditiva

y cognitiva (ver [2, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20] entre muchos otros). Otros trabajos han seguido otra filosofía, tomando la biología como fuente de inspiración, pero no como un objetivo en sí mismo (ver [5, 21, 22, 23, 24, 25, 26] entre muchos otros).

La estrategia de visión selectiva guiada por cambios (SCD por sus siglas en inglés) publicada en [27, 28, 29, 30], pertenece a este último grupo. Orientada a la detección y análisis de objetos moviéndose a alta velocidad, la estrategia SCD asume que solo un parte de la imagen muestra cambios entre dos *frames* consecutivos, mientras que la mayor parte de los píxeles permanecen igual. Esta hipótesis cobra especial sentido cuando se capturan *frames* a alta velocidad. Teniendo en cuenta que muchos de los píxeles de una determinada imagen no han cambiado respecto de sus valores en las imágenes anteriores de la secuencia, los algoritmos de procesado pueden utilizar la información ya almacenada para realizar sus cálculos. Es decir, que esta información redundante podría no transmitirse. Se podría incluso considerar que los píxeles de la matriz que muestran cambios pequeños, tendrán poco impacto en los resultados de los algoritmos. En la estrategia SCD estas hipótesis son trabajadas de forma tal que se consigue reducir sustancialmente la cantidad de información transmitida por el sensor, y por lo tanto la cantidad de información procesada fuera del mismo.

En la estrategia SCD ya no se trabaja con imágenes de forma estática, sino que la información es transportada y transmitida en la forma de un flujo de píxeles. Estos píxeles son seleccionados de forma tal que contengan solo la información con cambios temporales relevantes dentro de la escena bajo análisis. Bajo estas nuevas condiciones, sería necesario el rediseño de muchos de los algoritmos de visión tradicionales, ya que estos trabajan en base a una secuencia de imágenes estáticas transmitidas a intrevalos de tiempo regulares. El paradigma de procesado por flujo de datos (*data-flow processing*) [31] parace ajustarse de manera más adecuada a esta nueva forma de trabajo.

El principal objetivo de esta tesis es el diseño, la implementación y el testeo del primer sensor VLSI basado en la estrategia SCD de captura de información visual. En los trabajos previos a esta tesis, las ventajas de la Visión SCD fueron demostradas mediante simulaciones [27, 28, 29, 30, 32, 33]. La implementación en silicio propuesta como trabajo de esta tesis es el paso necesario para comprobar y cuantificar en entornos reales los beneficios que la estrategia SCD supondría para la implementación en hardware de muchos sistemas de visión artificial. Mas específicamente, los

principales objetivos de esta tesis son:

- Diseñar el primer chip VLSI de visión artificial siguiendo la estrategia SCD de captura y transmisión de información visual.

- Implementar un sistema completo de visión artificial basado en este nuevo sensor SCD. Para cumplir con este objetivo será necesario el diseño de una PCB, un firmware y un software de PC para la adquisición de datos.

- Habiéndose cumplido los objetivos anteriores, el último objetivo es mostrar, mediante la experimentación adecuada, como un sistema SCD es una alternativa válida a los sistemas de visión artificial tradicionales basados en secuencias de imágenes estáticas. De esta manera, se pretende cuantificar y analizar, las ventajas de este nuevo tipo de sistemas propuesto.

Metodología

Desde el punto de vista metodológico, esta tesis ha seguido la metodología *top-down full-custom* de diseño microelectrónico. Todo diseño microelectrónico se construye a partir de una serie de especificaciones. Tomando como base dichas especificaciones, se pueden identificar las principales limitaciones de los circuitos involucrados en el diseño. Esto permite proponer un diagrama en bloques de la implementación. Para el diseño de cada bloque en particular, se pueden buscar alternativas en la bibliografía pertinente. En algunos casos, las respuestas a los problemas de diseño pueden surgir como propuestas originales dentro de la investigación en curso.

En el caso particular de esta tesis, el énfasis se ha puesto en la captura de *frames* a alta velocidad (en el orden de 1-2 kfps). Otro de los puntos importantes a tener en cuenta es el diseño del bloque en el cual se seleccionan los píxeles que más han cambiado desde la última vez que fueron transmitidos. Los circuitos que más se adaptan a esta estrategia de comparación son analógicos [34, 35, 36, 37, 38, 39, 40]. Es este el motivo por el cual se ha escogido para el diseño del sensor una tecnología CMOS con comprobadas prestaciones de señal mixta analógico/digitales.

Una vez fabricado el chip, se procedió a seguir un plan de testeo con su correspondiente protocolo de validación. Debido a que el chip es un prototipo que no sigue ningún estándar particular, la placa de pruebas, el

protocolo de comunicación y el código del firmware y del software fueron
diseñados enteramente a medida.

Conclusiones

De acuerdo a lo expuesto en el primer capítulo de esta tesis, los objetivos
principales de la misma era: diseñar y fabricar el primer sensor siguiendo la
estrategia SCD de captura y transmisión de información visual, y mostrar,
mediante experimentación adecuada con el sensor fabricado, como un sis-
tema de visión artificial puede mejorar sus prestaciones cuando los princi-
pios SCD son utilizados en el diseño de su hardware y de sus algoritmos
de procesado. Todos estos objetivos fueron alcanzados. Se diseñó y fabricó
el primer sensor de visión SCD. Una vez fabricado, el mismo se empotró
dentro de un sistema portátil basado en un microcontrolador de 32 bits.
Este sistema sirvió para implementar algoritmos de seguimiento de obje-
tos moviéndose a alta velocidad. También sirvió para medir las principales
características del sensor.

Dentro del esquema del circuito, el bloque más importante, y el que
implicó un mayor desafío, es el bloque encargado de seleccionar los píxeles
que más han cambiado dentro de la matriz. Este tipo de comparaciones
globales de valores analógicos no son simples de realizar. En el caso de
esta tesis, para esta tarea se utlizó un circuito *winner-takes-all* (WTA). En
una primer etapa analógica el WTA selecciona un conjunto de valores que
muestran cambios de iluminación relevantes. Una señal de inhibición digital
se propaga a lo largo de toda la matriz para que solo uno de estos valores
sea transmitido, de esta manera evitando colisiones en la transmisión de
los píxeles. El circuito digital de inhibición ha sido publicado como una de
las contribuciones originales de esta tesis [77].

En esta primera implementación del sensor la resolución es todavía baja.
Esto es debido a que se trata de un primer prototipo. Se deja como trabajo
futuro la implementación y fabricación de una matriz de píxeles de mayores
dimensiones.

El proceso de caracterización del sensor muestra algunos problemas re-
lativos al ruido y al rango dinámico del mismo. El ruido es bastante mayor
de lo esperado. El rango dinámico es bajo, por lo que el sistema desarro-
llado puede utilizarse solo bajo determinadas condiciones de iluminación
favorables. El consumo de potencia por pixel es más elevado de lo que se
reporta en la bibliografía para otros sensores similares de reciente desarro-

llo. El transitorio de la señal del nivel de gris es largo y debería reducirse en futuras versiones del sensor. Todas estas cuestiones deben ser tenidas en cuenta como mejoras y trabajo futuro para futuras versiones del sensor SCD.

Los diseños VLSI en tecnología CMOS como el llevado a cabo en esta tesis son, habitualmente, difíciles de desarrollar. Los chips que llegan a obtener resultados lo suficientemente buenos como para llegar al mercado son el resultado de varias iteraciones de diseño y de la fabricación de varios prototipos. Esta tesis puede considerarse como la primer iteración de este proceso, y es evidente que todavía hay trabajo por hacer para llegar a los estándares de mercado. Aún así, bajo condiciones de iluminación adecuadas, se han realizado experimentos de seguimiento de objetos moviéndose a alta velocidad transmitiendo muy pocos píxeles del sensor hacia el sistema de procesado. Estos resultados, detallados en la sección 4.3 muestran como el hardware (sensor junto con el sistema de procesado) desarrollado en esta tesis es capaz de seguir objetos a muy alta velocidad utilizando tan solo el ancho de banda que utilizaría un sensor tradicional de baja velocidad de 25 fps, pero con la resolución temporal y las prestaciones de un hardware de captura de imágenes de alta velocidad trabajando a una tasa de 2000 fps. Esto demuestra y cuantifica como el ancho de banda y las prestaciones de un sistema de visión artificial pueden obtener una mejora sustancial al incorporar los principios SCD de diseño de hardware y procesado de la información. Esta experimentación constituye uno de los principales logros de esta tesis.

Contribuciones

- En los trabajos previos a la realización de esta tesis, las ventajas de la estrategia de visión SCD fueron presentadas solo en entornos simulados. Una de las mayores contribuciones de esta tesis es el diseño e implementación del primer sensor VLSI siguiendo los principios de sensado SCD. Este logro abre la puerta para el desarrollo e implementación de sistemas de visión artificial en los cuales es posible obtener altas velocidades de captura y procesado, pero utilizando hardware estándar de bajo coste.

- De todos los bloques circuitales diseñados e implementados, el que implicó el mayor desafío es el encargado de seleccionar el pixel que más

ha cambiado dentro de la matriz. La bibliografía muestra que los circuitos WTA pueden ser una solución cuando se desea averiguar el mayor de los valores entre un grupo de tensiones o corrientes analógicas. Cuando el número de celdas involucradas en la competición es muy grande, existe la posibilidad de que se seleccione más de un ganador como resultado de la misma. Una de las contribuciones originales de esta tesis es la propuesta de un subcircuito digital para lidiar con la situación de que se haya seleccionado más de un ganador en este tipo de etapas analógicas. El circuito finalmente implementado y fabricado consiste en una primer etapa analógica, en la que se podrían observar muchos ganadores, seguida de una etapa digital en la cual una señal de inhibición se propaga a través de la matriz para que solo uno de los píxeles seleccionados como ganadores en la etapa analógica pueda acceder al circuito de lectura exterior del pixel.

- Otro logro, que vale la pena mencionar, relacionado con la etapa WTA es que, al momento de escribir esta memoria, no hemos encontrado ningún reporte de la fabricación y testeo de este tipo particular de celda WTA analógica. A pesar de que el diseño de este amplificador WTA en particular fue propuesto den 1998 por Sekerkiran *et al.*, esta es la primera vez que se fabrica. En el experimento de seguimiento de objetos descrito en la sección 4.3 se muestra como el bloque de decisión basado en esta celda es capaz de encontrar el grupo de píxeles que más han cambiado dentro de la matriz.

- La implementación de un sistema de visión artificial SCD compacto, basado en el sensor VLSI diseñado durante esta tesis, y la experimentación llevada a cabo con él, pueden también considerarse uno de los mayores logros alcanzados durante esta tesis. A pesar de algunos problemas relacionados con el ruido y el rango dinámico del sensor (ver las secciones 4.2.5, 4.2.2 y 4.2.9 para una discusión detallada sobre estos temas) la experimentación descrita en la sección 4.3 muestra como, un sistema SCD sencillo basado en el sensor desarrollado en esta tesis es capaz de realizar el seguimiento de objetos moviéndose a alta velocidad con muy bajos requisitos de ancho de banda, pero con altas prestaciones en cuanto a resolución temporal y latencia en el procesado de la información.

Abstract

An artificial vision system is basically composed of a sensor, usually in VLSI CMOS or CCD technology, and a processing stage. Nowadays, in the vast majority of real-world implementations, the sensing part of the system is a traditional frame-based imager. These types of image sensors work under some very well known principles: the illumination level of the surrounding environment is sampled and transmitted at regular time intervals, even if no new relevant information is produced in the scene under analysis. A traditional frame-based image sensor is usually not able to evaluate if the information coming from a certain pixel is relevant or irrelevant. Since they do not perform any kind of analysis of the information being captured, the illumination level of all the pixels in the sensing matrix must be transmitted to be analyzed and processed at the processing stage. Many times, a huge amount of redundant non-relevant information is transmitted. The consequences of this are that valuable resources such as bandwidth and processing power are wasted. Furthermore, depending on the particular context and hardware configuration, the processing hardware may not even be able to cope with all the generated data.

Many of these problems can be overcome with the design of new sensing and readout strategies focused on the selection of relevant changing information. Over the last decade many relevant improvements have been achieved in this direction. Taking the biological vision system as a general guide and inspiration, an increasing number of very-large scale of integration (VLSI) vision sensors have been, and are being designed where the sparcity, asynchrony and event-driven generation of the information coming from the visual field is taken into account.

It is within this framework that Selective Change-Driven Vision (SCD) emerges as an innovative and original proposal. SCD Vision relies on the idea that a pixel showing a large change in intensity is an indicator of fast movements, and object edges around it. An SCD sensor is frame-based in the sense that successive frames are captured at a very high rate, but pixel readout is performed in an entirely different manner. The pixels are read out in order of relevance. The larger the change in illumination, the more relevant the pixel is considered to be. Not all the pixels in the sensing matrix need to be transmitted. As the pixels showing relevant changing

information are transmitted first, a small subset of pixels might be read
out, these being the ones conveying the most important information of the
scene under analysis.

In this thesis, the first VLSI CMOS vision sensor following SCD prin-
ciples is presented. A 32x32 pixel matrix was implemented and fabricated
in 0.35 μm 4-metal 2-poly silicon technology.

The most challenging part of this microelectronic design was the deci-
sion block, where the pixels undergoing the largest changes in the sensing
matrix are selected. This problem was solved by means of a winner-takes-all
(WTA) circuit. A large WTA network together with a proposal for single
winner selection was designed, implemented and its behaviour character-
ized.

The designed sensor was embedded into a small, but powerful artificial
vision system based on a 32-bit microcontroller. This system was used
to implement tracking algorithms as well as to characterize the main basic
features of the sensor. The experimentation carried out in this thesis shows
how a simple SCD system based on our SCD sensor is able to track fast
moving objects with just the bandwidth requirements of a low speed 25 fps
standard camera, but with the time resolution and performance of a high-
speed camera working at 2000 fps. This clearly demonstrates that band-
width and processing requirements are substantially reduced when SCD
hardware is used.

Agradecimientos y dedicatorias

En primer lugar, agradezco a Fernando Pardo y a José Antonio Boluda, no sólo por haber confiado en mí para realizar el trabajo que ahora presento y haberme permitido descubrir el fascinante mundo del diseño microelectrónico, sino por su activa participación en el testeo del sensor que aquí se describe y presenta. Sin su ayuda, esta tesis seguramente no hubiese llegado hasta donde llegó. Mis más sinceros agradecimientos a Paco Vegara por haberme ayudado una y otra vez con todo el cacharraje necesario para los experimentos. Su colaboración ha sido también fundamental en este trabajo.

A Tobi Delbrück, Shih-Chi Liu, Raphael Berner y Alex de la Plaza, gracias por su ayuda y sobre todo por su paciencia. Mi ignorancia en estos temas al principio de esta tesis era tan evidente como completa. Muchas gracias a Paco Serra-Graells y a Lluìs Terès, no sólo por confiar en mí para las muchas responsabilidades que implica llevar adelante un proyecto en el IMB-CNM, sino por haberme permitido ceder tiempo de mi nuevo trabajo para terminar esta tesis. A Wladimiro Díaz, gracias por su paciencia y ayuda en todo lo referido al sistema operativo Linux.

Quiero también agradecer a mis compañeros de trabajo, Guillermo, Katerine, Adrián y Alicia, que han hecho de nuestro lugar de trabajo un sitio ameno y lleno de compañerismo.

Si tengo que resumir los últimos cinco años y medio de mi vida con una sola palabra, la que, sin dudas, más cabalmente se ajusta, es la palabra 'aprendizaje'. Si de una enumeración se trata, esta incompleta lista puede servir de orientación: he aprendido la difícil, artesanal e interesante tarea de diseñar chips analógicos; he aprendido a diseñar circuitos impresos; he aprendido a programar el microcontrolador PIC; he aprendido las bases, y varios de los vericuetos, del exageradamente amado por unos y exageradamente odiado por otros, sistema operativo linux; he aprendido, y todavía sigo, día a día, aprendiendo, a ser padre; he aprendido, y todavía sigo, día a día, aprendiendo, a ser padre de una hija mujer; he aprendido, a través de enfrentar fantasmas ajenos, a enfrentar muchos de los fantasmas propios con los cuales todos convivimos. Estuve en el Instituto de Neuroinformática de Zürich, lugar que me fascinó desde el primer día por la capacidad y la experiencia de la personas, punteras casi todas en sus repectivas disciplinas,

que allí trabajan. Pero la fascinación más grande provino, tal vez, del hecho de la honestidad y completa humildad con la que algunas de las personas más brillantes que he conocido, y seguramente conoceré, reconocían su más completa ignorancia. Esto ha, definitivamente, dejado en mí una marca. Mi forma de trabajar y analizar la realidad que me rodea es ahora distinta. Aprendí lo que es el Acontecimiento, y me aferré a él con pasión; aprendí que la luna podría, si quisiera, tomar un helado; que la letra H es amiga de las vocales; y descubrí, que en la esquina de mi propia casa, hay un bosque, muy oscuro, y lleno de lobos. Aprendí sobre la alegría del llegar y sobre la dignidad, la entereza, la tristeza, y la aceptación del irse.

Agradezco a Max y a Sandra la confianza y una amistad que lleva ya varios años, y espero lleve muchos más. A Matu, a quien sigo considerando una de las personas con más espíritu científico de todas las que conozco, y a quien sigo queriendo como a un hermano; y a Laura, Valentín y David, su maravillosa familia, con quienes tan buenos momentos hemos compartido a lo largo de estos años. No puedo dejar de mencionar en estas páginas a Chema, quien es ya más porteño que yo mismo; a Majo, a quien conozco desde hace más años de los que puedo recordar, y a Malena y Aitana. Agradezco a Aníbal, por su amistad, por sus sabias y célebres frases, y por confiar en mi persona mucho más de lo que yo lo haré nunca. No puedo dejar de agradecer a Carlos; una parte de este trabajo, sin lugar a dudas, le pertenece. Quiero también agradecer a Esther, Miguel Ángel, Reme, Pablo y María, personas por las cuales siento muchísimo respeto y cariño, por haberme tratado desde el primer día como uno más de su familia.

Dedico un muy especial agradecimiento a Juan Antonio, Isabel, Gemma, Belén, Alberto, Nuria, Guillem y Alemnesh, por haberme ayudado a distraerme de la tesis en los momentos en los cuales nada salía bien y por haberme acompañado en los momentos en los cuales todo empezaba a cerrar, pero sobre todo, por tratar a mi familia como a la suya propia. Sus permanentes demostraciones de afecto forman parte de las cosas que nunca olvidaré.

Dedico esta tesis a mis padres, a quienes profeso el más profundo cariño, y a los cuales, a través de mi propia paternidad, comprendo cada día más. También dedico este trabajo a mi hermano, Santiago, cuyo sentido común ha sido siempre muy superior al mío; y a su familia, Malala, Renata y Genaro, por quienes siento un afecto infinito, y a quienes lamento no poder tener más cerca.

Finalmente, dedico no sólo este trabajo, sino el día a día de mi vida

entera a Esther, y a su sonrisa que todo lo ilumina, a Leo, mi pequeño sabio, y a Violeta, cuya contagiosa alegría hace que muchas de las cosas que me rodean cobren sentido. No puedo más que agradecerles por su apoyo, sus permanentes demostraciones de afecto y sobre todo, por su paciencia. Una gran parte del tiempo dedicado a esta tesis era, definitivamente, suyo. Espero poder compensarlos de ahora en adelante.

Contents

List of Figures

List of Tables

Abreviations and achronyms

AER address event representation

ADC analog-to-digital converter

B blue

BDJ burried double junction

CCD charge-coupled devices

CDS Correlated double sampling

CMOS complementary metal-oxide-semiconductor

CTIA capacitive transimpedance amplifier

DR dynamic range

DVS Dynamic Vision Sensor

FBK Bruno Kessler Foundation

FPGA field-programmable gate array

FPN fixed-pattern noise

GB green-blue

I2C Inter-Integrated Circuit

INI Institute of Neuroinformatics

I/O input/output

MIM metal-insulator-metal

OTA operational transconductance amplifier

PC personal computer

PCB printed circuit board

PIP polySi-insulator-polySi

PWM pulse width modulation

RG red-green

RGB red, green and blue

SCD Selective Change-Driven

sDVS sensitive Dynamic Vision Sensor

TCDS true correlated-double-sampling

USB Universal Serial Bus

UWB Ultra-Wide-Band

VLSI very large scale of integration

WSN Wireless Sensor Network

WSNs Wireless Sensor Networks

WTA Winner-takes-all

Introduction 1

Introduction

1

1.1 Motivation and objectives

Traditional frame-based image sensors, frequently called imagers, work under some very well known principles: the illumination level of the surrounding environment is sampled and transmitted at regular time intervals, even if no new information is produced. While this is perfectly suited for many real-world applications such as scanning systems, medical imaging or home video systems, that need accurate, precise and complete information concerning the lighting conditions of the surrounding environment, for other applications, such as artificial vision systems or Wireless Sensor Networks (WSNs), the huge amount of redundant information to be processed and transmitted becomes a crucial limiting factor. Perfected over millions of years of evolution, biological vision systems work on a completely different basis. One of the key features of these systems, is that each sensor cell (pixel) reacts independently and asynchronously to illumination changes. Starting in the late 1980's with Mead and Mahowald's work [1, 2], the last two decades have witnessed significant efforts towards overcoming the limitations of frame-based vision sensors and algorithms - namely by incorporating biological neuromorphic principles into very large scale of

integration (VLSI) hardware design. Work such as [2] and [3] pioneered in this field, and together with, [4], [5] and [6] show successful implementations of neuromorphic vision systems. Results in this field clearly show that bandwidth and processing power needs are substantially reduced when neuromorphic principles are applied.

The goal of many of the recent developments in this research field is to mimic, as closely as possible, the behaviour of the biological visual, auditory and cognitive systems by means of advanced silicon technologies (see [2, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20] amongst many others). Others, have followed a different approach by taking biology as an inspiration but not as an objective (see [5, 21, 22, 23, 24, 25, 26] amongst others).

SCD Vision presented and applied in [27, 28, 29, 30], belongs to this latter group, where biology principles serve as a general guide for VLSI implementation and algorithm design. Its working principle is oriented to fast movement detection and analysis under resource limited conditions. SCD Vision assumes that only part of the image changes between two consecutive snapshots and that the rest of the image remains unchanged. This assumption seems to be especially reasonable for very fast image rates. It, therefore, follows that any video processing algorithm does not need those parts of the image that have not changed at all, because a new result, for a new image, can be obtained just using the past results along with the changes introduced in the new image. Moreover, in all probability, those pixels that have changed little have almost no impact on the result of the processing, so it is possible that they are not needed, at least to some extent, depending on the application. SCD Vision implements these arguments in order to reduce the amount of data to be acquired, transmitted, and therefore processed, thus reducing the hardware required to process video and/or increasing the system's performance (usually both) [27].

As the only transmitted pixels are the ones that show new information catalogued as relevant by a certain criteria, the concept of image, conceived as a complete and static representation of the environment, doesn't seem to be valid any more. Instead, the information is conveyed by a continuous flow of pixels pinpointing relevant parts of the scene. This new, and seemingly much more interesting and useful approach, would require the redesign of many of the traditional computer vision algorithms. Under this new scenario, the data-flow processing paradigm ([31]) would be much more

adequate than traditional image-based analysis of sequences.

The main objective of this thesis is to design, implement and test the first VLSI chip following SCD Vision principles. Before the beginning of this research work, the advantages of SCD Vision had only been presented in simulated environments ([27, 28, 29, 30, 32, 33]). A silicon implementation is the necessary and logical step to prove and quantify how much SCD Vision can improve the performance of an artificial vision system in real situations where resources such as bandwidth and processing power are a substantial and important limiting factor. More specifically, the objectives of this research work are:

- To design the first vision chip following SCD principles.

- To implement a complete SCD hardware based on the designed chip. The development of a customized printed circuit board (PCB), firmware, and personal computer (PC) software for data acquisition are needed to fulfil this objective.

- Having fulfilled all the requirements of the previous items, to show, by proper experimentation under real-world scenarios, how an SCD Vision system based on the developed SCD hardware and sensor, is a valid alternative to traditional image-based artificial vision systems. The benefits of this newly implemented system will be properly quantified and analysed.

These objectives pave the way for future developments of SCD dataflow based algorithms and more specific hardware for applications such as WSNs or very fast 3D scanning systems.

1.2   Methodology

From the methodological point of view, the work carried out during this thesis followed a top-down full-custom microelectronic design approach. Any VLSI design must be constructed from a series of specifications. Based on these specifications, the bottlenecks of the real circuit can be identified, and a block diagram of the circuit can be proposed. Proposals for the design of every block can be found in the literature, or, in some cases, can

emerge as original contributions during the development of the thesis in the context of the research work.

In the particular case of this thesis, the emphasis is on achieving high frame rates, in the order of 1-2 kfps, and low power consumption. Another important problem to be solved is the comparison block, where the decision whether a pixel should be delivered or not is taken. The circuits that best suit the decision block requirements are analogue [34, 35, 36, 37, 38, 39, 40]. This is the reason why a complementary metal-oxide-semiconductor (CMOS) silicon technology with mixed-signal analogue/digital design options was chosen.

Once the chip was fabricated, a testing plan and validation protocol were followed. Due to the highly customized nature of the chip, a native original testing board, code and protocol were the chosen option.

1.3   Organization of the thesis

In order to successfully achieve the goals proposed in Section 1.1, the first step is to carry out an extensive review of the current state-of-the-art in vision circuit design. The outcome of this bibliographical research is reflected in Chapter 2. The large number of vision sensors developed in the last decade shows that this is a very active research field. Although the variety of vision sensing strategies is very wide, they all have the common purpose of reducing the large amount of redundant data to be transmitted and processed.

A complete section of Chapter 2 is dedicated to a detailed explanation of the SCD principles applied to artificial vision systems. The main differences compared to other vision sensing strategies are analysed, and its main advantages in fundamental aspects such as bandwidth occupancy and processing power requirements are discussed.

In Chapter 3 an in depth description of the proposed microelectronic VLSI circuit is given. The transistor-level description and the simulated performance of all the building blocks of this novel pixel design are described. The main characteristics of the pixel layout together with a set of microphotographs of the fabricated silicon die are also shown in this chapter.

A description of the experimental set up together with a detailed explanation of the experiments carried out and the results obtained are presented in Chapter 4. In this chapter, the sensor is characterized by its main basic features showing the advantages and drawbacks of this first VLSI SCD silicon implementation. The results of tracking experiments are also presented. These, are clearly oriented to show the advantages of SCD Vision over traditional frame-based artificial vision systems.

Finally, in Chapter 5 the conclusions and future work lines are discussed. New ideas are presented for improving the performance of the SCD pixel. The contributions of this thesis to this research field are also detailed in this chapter.

State of the art 2

State of the art

<div style="text-align: right">

2

</div>

Es polivalente en sus aplicaciones; sirve para enmendar a los presos, pero también para curar a los enfermos, para instruir a los escolares, para guardar a los locos, vigilar a los obreros, hacer trabajar a los mendigos y a los ociosos. Es un tipo de implantación de los cuerpos en el espacio, de distribución de los individuos unos en relación con los otros, de organización jerárquica, de disposición de los centros y de los canales de poder...

<div style="text-align: right">

M. Foucault

</div>

2.1 General Overview

From an architectural point of view, an artificial vision system is basically composed of a sensor, usually in VLSI CMOS or charge-coupled devices (CCD) technology, and a processing stage. Nowadays, in the vast majority

of real-world implementations the sensing part of the system is a frame-based imager. As explained in Chapter 1, in the context of artificial vision these types of sensors seem to be very inefficient and very unlike natural visual systems. Even though a new paradigm for vision sensor design was proposed back in 1992 with Mahowald's Ph.D. work [1, 2], in which sparsity, asynchrony and event-driven generation of data were taken into account, most of the efforts in this direction have been concentrated over the last ten years. This is because it was not until the mid or late 1990's, when CMOS technology was mature enough compared to the CCD devices, that CMOS pixel design began to show reasonable and acceptable results [41].

The vision sensors developed in this field in the last decade can be classified in many ways by considering technical aspects, i.e., its synchronous or asynchronous characteristics, the type of implemented photocircuit, whether they deliver grey level values or just spiking events, if they sense and compute temporal contrast, temporal difference, spatial contrast or spatial difference, or many other technical features. Another interesting possibility is to take into account whether the sensors were conceived trying to mimic the biological system (biomimetic design), or they just consider biology as a general inspiration but not as an objective. Biomimetic research aims to open a completely new and cutting-edge way of, not only sensing, but also processing visual information. As asynchrony is one of the most relevant aspects of this trend, completely new processing hardware and algorithmic design is needed under this paradigm. On the other hand, many other research works intend to improve the performance of traditional artificial vision systems in bandwidth and processing power requirements with new sensing strategies developing new VLSI sensors that fit into traditional available processing hardware. As this particular aspect of the vision chips developed in the last decade conditions many of its technical features, this is the classification chosen in this chapter to review some of the most relevant developments in this field.

2.1.1   Biomimetic event-driven vision sensors designs

One of the first relevant works in this field published at the beginning of the 2000's is [16], presented by Culurciello *et al.* in 2003. In this biomorphic arbitrated image sensor the photocurrents are integrated up to a fixed voltage threshold. A spike is generated whenever the integration signal crosses the threshold value; therefore, the magnitude of the photocurrent is inversely

proportional to the interspike interval. Events (spikes) are asynchronously transmitted out of the chip by an arbitration circuit. The original current-feedback event generator used in the photocircuit combines very little power consumption with high slew-rate response. With this strategy, a high dynamic range of 120 dB is achieved when no lower bound condition is placed for the output frequency per pixel. This value changes to 48.9 dB if the pixel update frequency is bounded to 30 Hz. Experimental results showed that the sensor was, at that time, similar to state-of-the-art frame-based CMOS imagers, in terms of speed and power consumption, although further work was needed to obtain similar results in image quality aspects.

In 2004 [12, 13], Zaghloul and Boahen reported an interesting VLSI model of the biological system by incorporating transient and sustained responses, these being characteristics of human vision. This retina performs spatial and temporal filtering adapting to illumination level. However, the performance of this design is severely limited by the FPN problem and its very low dynamic range.

In 2007 [42], a 32x32 pixel-matrix was designed where spatial contrast information was delivered as an address event representation (AER) data-stream. The spatial contrast is computed taking into account not only the pixel's photocurrent, but also information coming from neighbouring pixels as well. The most important contribution of this paper is that the local average of the photocurrents from the neighbouring pixels is obtained with very low mismatch via a diffuser network and an original calibration circuit presented by this same group in [43]. The reported 6.6% mismatch value is a remarkable achievement considering the very low level of the photocurrents usually present in this kind of sensors. As the spatial contrast is computed as the ratio of the pixel's illumination value and the average illumination of the neighbouring pixels, the DC value of this magnitude is not zero, so pixels continuously output spikes, even in the absence of new relevant information. This is the main drawback of this design. An alternative design proposed in 2010 [44] overcomes this problem by computing not the ratio, but the difference between the pixel's illumination and the average of its neighbours. This difference is normalized with respect to the ambient illumination level, so the result is ambient-light independent. This chip can also be operated in time-to-first-spike mode.

Amongst the variety of vision chips published in the last decade, what seems to be a milestone in this field is the Dynamic Vision Sensor (DVS)

[4] developed at the Institute of Neuroinformatics (INI) [1], Zürich. The clear intention of this work, presented in 2008, is to mimic, as closely as possible, the sensing stage of the visual biological system. Presented as one of the most relevant outcomes of the CAVIAR project [45], this temporal-contrast vision sensor completely abandons the notion of frames. Each pixel works in a completely independent and asynchronous manner quantizing local relative intensity changes. Every time a pixel detects a certain fixed contrast change, a spike (event) is generated. In order to avoid collisions, a special subcircuit [18, 19, 20] is used to arbitrate the events coming from different pixels. The output of the sensor is an asynchronous stream of pixel addresses configuring an AER of the visual field. This sensor exhibits a 2.1% mismatch in relative intensity event threshold, a dynamic range of 120 dB, a minimum event latency of 15 μs and 23 mW of chip power consumption. Although the vast majority of the systems for artificial vision processing are still frame-based, the DVS is being successfully used in many applications nowadays (see [46, 47, 48, 49]).

In 2010, a new version of the DVS pixel, called sensitive Dynamic Vision Sensor (sDVS), was presented in [50]. In this case, events are produced with only a 0.3% contrast in illumination, which represents a marked improvement (about 50 times better) compared to the original DVS design. The main difference compared to the DVS, is that another amplifying stage is introduced at the photodiode node taking advantage of the Miller effect of a PMOS transistor. However, this new design has many problems, such as mismatching, or the gain-bandwidth trade-off of the first amplifying stage, that must be resolved before a pixel matrix can be fully integrated.

In 2011, Leñero-Bardallo *et al.* published a sensor with the same working principles as the DVS but reporting latency as low as 3.6 μs [51]. Experiments show that objects and particles rotating at speeds as fast as 10 k revolutions-per-second can be successfully detected and tracked. This improvement is achieved at the expense of increased power consumption, a slightly increased FPN and a reduced intrascene dynamic range of 54 dB (although the overall dynamic range is close to 100 dB).

Recently, in 2011 [52], Posch *et al.* presented a 304x240-pixel AER vision sensor with a remarkable pixel design that outputs not only asynchronous events related to temporal contrast, but also the grey level value of the pixels associated with these asynchronous events. In other words, the

---

[1]http://www.ini.ch

grey level value is asynchronously delivered only for the pixels that changed. The event generation subcircuit is very similar to [4], while the grey level is encoded as a pulse width modulation (PWM) signal. A novel time-domain true correlated-double-sampling (TCDS) technique for obtaining the grey level value [53] yields array FPN smaller than 0.25%.

Other clear examples of biomimetic vision chips can be found in [6, 7, 54, 55].

Efforts are also being made to incorporate the event-based principles to transduce not only intensity, but also colour information. Following the same research and circuit design line of [4], in 2011 [56], Berner *et al.* achieve dichromatic colour sensitivity in standard 180 nm CMOS technology by means of a properly biased burried double junction (BDJ) to discriminate between incident photons with blue and red spectral characteristics. As the absorption length of incident photons in silicon increases monotonically with wavelength, the shallow junction of the BDJ is more sensitive to bluer light, while the deeper junction shows higher sensitivity in the red band of the spectrum. This same pixel also responds to log monochrome intensity changes. Two different signal pathways are implemented to amplify and integrate colour and intensity information, and different binary events are output responding to changes in wavelength or intensity of incident light. Despite several shortcomings, minutely detailed by the authors in the characterization of the pixel, the presented results show that the pixel responds to light wavelength changes as small as 15 nm, and to relative intensity changes as small as 10%. This promising design seems to point the way for future real-world performance implementations of colour event-based sensors. Also in 2011, another interesting work in this field was presented in [57] by Leñero-Bardallo *et al.* In this case, three buried junctions are fabricated in a standard 90 nm CMOS process to obtain three stacked photodiodes for separating colour information in three different channels: red-green (RG), green-blue (GB) and blue (B). In each channel, an integrate and fire neuron outputs spiking events; the frequency of these events is proportional to the inverse photocurrent of the corresponding p-n buried junction. A simple algorithm is used to convert pseudo-colour information provided by the sensor into red, green and blue (RGB) representation to be shown on a standard computer screen. Experimentation with the fabricated 22x22 pixel array shows satisfactory preliminary results, although much more work, also in this case, needs to be done to enable real-world

applications.

2.1.2   Other event-driven vision sensing strategies

In the late 1990's, Aizawa *et al.* presented a CMOS vision sensor with compression capabilities on the image plane [58]. The sensor is based on the so called conditional replenishment algorithm. The sensor has a memory where the last replenished value of every pixel is stored. Current pixel values are compared to those of the last replenished frame. The pixel-matrix is sequentially scanned, and only the values and addresses of the pixels for which the magnitude of this difference is greater than a certain threshold are transmitted. Once a pixel is transmitted, its corresponding value in the frame memory is replenished with the current pixel value. In this work a first 32x32 pixel-matrix prototype is presented. Each pixel includes a memory for the last replenished value and a comparator for the thresholding operation. A second 32x32 pixel-matrix prototype is presented in [59] where the thresholding and comparison functionalities are shared column-wise, in this way the pixel size can be reduced. Despite this new architecture, fill factor is only 1.9%. In both papers, the results show that images are very noisy and any of the silicon prototypes could be used to properly evaluate the advantages of the proposed algorithm.

In 2003, a very interesting chip with image contrast and orientation extraction was reported [5]. Each pixel integrates its photocurrent until a fixed global threshold-voltage is reached. At this moment, a comparison process between neighbouring pixels is fired to compute the contrast magnitude and direction (gradient). As data is transmitted out of the chip in decreasing order according to the contrast magnitude value, the chip first delivers relevant information and can be considered an event-based sensor, although this latter comparison and readout process requires a global chip-level synchronization similar to frame-based imagers. This sensor exhibits a remarkable dynamic range of 120 dB, detecting contrasts as low as 2% with an error on the orientation value of $\pm$ 3°. This sensor is being successfully used in the automotive industry [60].

The Smart Optical Sensors and Interfaces group [2] at the Bruno Kessler Foundation (FBK) [3] developed several vision sensors clearly oriented to

---

[2]http://soi.fbk.eu/en/home
[3]http://www.fbk.eu

WSNs applications. Prior research of this group shows interesting advances in frame-based smart vision sensors by incorporating on-chip image processing tasks (see [61] published in 2005 or [62] published in 2007). In 2009 [22] a novel pixel design for spatial-contrast extraction and binarization with event-based characteristics was presented. This 128x64-pixel-matrix sensor has a frame memory buffer so that the current spatial contrast can be compared against an arbitrary pattern stored in memory. This feature is fundamental for the built-in image processing capabilities of the chip. The sensor can be operated in two different modes: active or idle. In active mode, the matrix is scanned for pixels showing positive or negative differences with respect to the memorized spatial-contrast value. Only when a difference, positive or negative, is found the address of the pixel is latched and transmitted out of the chip. It takes about 150 μs to complete the scanning and readout process along the whole matrix. In idle mode, the matrix is scanned and the pixels showing positive or negative differences are counted without delivering any address out of the chip. At the end of the process, the final count is transmitted and used by any external processing as an estimation of the level of activity in the scene under analysis. Depending on the content of the frame buffer, different image processing tasks can be performed: contrast extraction, motion estimation or background subtraction. The measured power consumption is in the order of 100 μW when operated with and integration time of 10 ms and 25% of activity in the pixel matrix. Under proper lighting conditions, the integration time can be set down to 100 μs, which added to the readout time of 150 μs means that the sensor can be operated at a very high frame rate of 4 kfps. In 2011 [63], this sensor was thoroughly studied and characterized as a Wireless Sensor Network (WSN) node, showing that its very low power consumption and adaptability to scene activity make it a highly suitable alternative for this type of applications. This work was complemented, also in 2011, with [64], where a new version of the sensor optimized for a hierarchical energy management strategy is presented and tested.

Another interesting sensor suitable for WSNs applications is described in [25] (2010). The chip operates in three different modes: intensity, temporal difference and spatial contrast. In the first mode, the sensor acts as a traditional imager with Correlated double sampling (CDS) readout, while in temporal difference mode the intensity of the current frame is compared to the one immediately previous to it. An event is produced if the result of this comparison is greater than a certain threshold. In the third mode,

using a Winner-takes-all (WTA) circuit, every pixel compares its own intensity value against the intensity value of four of its neighbours, and this information is used to find contours and edges in the scene. As claimed by the authors, the main contribution of this work seems to be the compact implementation of all these three functionalities in a pixel with just 11 transistors. The application of WTA circuits for the inter-pixel intensity comparison task seems to be the key point of this contribution.

In [24] (2012), Chen *et al.* implement a very simple strategy for movement detection. When a new frame is buffered, the pixel matrix is scanned sequentially, and only the pixels that have changed above a certain threshold from one frame to the next one are transmitted, while the information in the remaining pixels is lost. The main advantage of this chip is that incorporates an Ultra-Wide-Band (UWB) transmitter. The experiments presented in [24] indicate a power consumption of 2.4 mW (0.9 mW for the pixel matrix and 1.5 mW for the transmitter) when the pixel matrix runs at 160 fps and the UWB transmitter works at 1.3 Mbps.

2.2   Selective Change-Driven Vision

Like all the sensors and artificial vision strategies described in the previous section, the aim of SCD vision is to reduce the amount of visual data to be transmitted and processed while keeping an accurate representation of the relevant information of the visual field. In SCD vision, events are collected as they are produced and selectively chosen in order of relevance according to the particular addressed problem. Although the SCD concept can be extended to many vision problems, in previous research [27, 28, 29, 32, 33, 65] the focus is on movement detection and analysis. In most movement algorithms, the most important event is the change in illumination level, since it may indicate a change in the scene being processed [66]. It seems that pixels that have undergone larger changes in illumination will offer richer information than others [27, 28] so this is the selective function used in previous SCD work and also applied in this thesis.

The SCD sensor developed in this thesis implements the principles mentioned above in the following way: frames are sequentially and synchronously captured at a very fast rate (1-2 kfps), but pixel readout is performed in a completely new manner. Pixels are now transmitted out

of the chip in descending order based on their illumination level difference with respect to the last transmitted value. When a pixel's address and illumination information are transmitted, the illumination value is stored in a memory within the chip so the next forthcoming value can be compared against it. As the ordering criteria takes into account the temporal difference with respect to the last read-out value, and not with respect to the value captured in the preceding frame, if a change in a pixel is not immediately processed because it is less relevant at that moment, there is no problem since it will be processed later when there are less important changes to process. Other state-of-the-art sensors, such as [24, 25], memorize events only from one frame to the next one; this means that if an event is not instantly processed that information is lost and its impact on the result is not calculated. The SCD update mechanism of the memorized grey level value is similar to the conditional replenishment algorithm proposed by Aizawa *et al.* in 1997 [58, 59]. The main differences are the thresholding operation and the readout sequence criteria: in [58, 59] the pixel matrix is sequentially scanned, while in SCD strategy pixels are ordered by relevance before being read out. According to [67], this SCD VLSI implementation can be classified as a frame-based time-difference sensor.

Although frame-based, this new sensing stage can be thought of as a source of information transmitting a continuous pixel flow, delivering the pixel's position and grey level. The pixel rate can be dynamically adapted to the available processing hardware capabilities and bandwidth resources. If the pixel rate is high, most of the pixels of the sensor are read-out even if they have changed little, but this sensor becomes really interesting when the pixel rate is reduced, even by several orders of magnitude, as it is still possible to obtain good results [68, 69]. It is also possible to keep the same pixel rate as a frame-based camera with traditional sequential readout scheme, but in this case the SCD offers another advantage and this is the possibility of processing information as though it were coming from a high speed camera. In fact, the SCD sensor is internally working at equivalent rates of 1-2 kfps, while the bandwidth requirement can be easily tailored to the available hardware capabilities, whatever they may be. In any case, the sensor is always delivering the pixels that will be of the most help in order to obtain the most accurate results for that bandwidth.

Another important advantage of SCD vision processing is that results of an algorithm are calculated immediately after every pixel. This means

that the result latency (time from scene event to algorithm response) can be very short (microseconds) compared to the latency of image processing which usually takes one frame acquisition time (milliseconds). This latency is very important in closed-loop control vision systems. Another advantage is that results are updated for every pixel, yielding a result flow with a much higher temporal resolution than that provided by a standard image processing system [27, 29, 69]. These aspects are very important in ultra-high speed movement tracking, or closed-loop vision systems.

At a first glance, it seems that keeping the results updated for every pixel can increase the workload, but this is not so for most cases since the same calculations, in a different order and fashion, are performed for every pixel. The only difference is that in the traditional case, we perform the calculation for every pixel in the image after the image has been taken (in a pair of nested loops), and in the other, the calculation is performed on the pixel as it is read-out, no matter in which order (see [28, 70] for an explanation about speeding-up strategies in the processing stages of an SCD systems). This holds for most first-stage algorithms, such as filtering, convolutions, basic object detection, and others. In higher level stages, it is possible to have some extra workload, which must be taken into account, but this can be reduced by smart algorithm tailoring [65].

From the hardware processing point of view, the change in the algorithm implementation does not imply a change in the processing architecture. Nevertheless, SCD vision works on events more than on a static flow of information. It is change-driven based, meaning that a change in illumination may drive the processing to perform. Data flow architectures can take advantage of this data-driven processing. In data flow architectures, it is the incoming data which fires the instruction execution, so there is no program counter but rather a graph of interdependent instructions that are executed when their required data arrive. These architectures are inherently parallel and can be implemented in reconfigurable hardware such as a field-programmable gate array (FPGA) [33].

The main difference between SCD Vision and the biomimetic approach is that in the former it is possible to obtain a constant synchronous pixel flow that can be easily adjusted to standard processing hardware capabilities. In biomimetic artificial vision systems, there is a non-continuous flow of spikes indicating events. In this way, there are time slots with no information to be processed, and other time slots where the pixel flow is

so high that the load could collapse any standard processing hardware. Asynchronous biomimetic vision chips require the design of custom specific asynchronous VLSI processing hardware [10, 11, 14, 45, 71, 72]. It seems that approaches like SCD Vision are, at present, much more flexible and much easier to translate into industrial applications.

Complete SCD vision systems, considering the sensing strategy described above followed by a processing stage, have been successfully simulated for optical flow computation, motion detection and tracking algorithms [27, 28, 29, 32, 65]. Real implementations with measured results have been recently published using the VLSI SCD sensor and processing hardware described in the following chapters of this document [68, 69, 70]. These results can be considered as direct and successful outcomes of this thesis.

2.3 Conclusions

All the sensors described in the preceding sections of this chapter have the common intention of reducing the huge amount of redundant non-relevant visual information produced by traditional frame-based imagers. Biological vision systems, perfected over millions of years of evolution, efficiently overcome this problem with cellular mechanisms that signal only relevant changing information in the visual field. Therefore, the biological system seems to be the perfect source of inspiration for the design of artificial vision systems. Most of the research and development efforts in this field have been concentrated over the last decade achieving relevant and important results. In many of the leading advances in this area, the focus is on mimicking as closely as possible the main features of the biological system (see Subsection 2.1.1). Other relevant works do not take biology as an objective, but as general guide for their developments by incorporating some of its principles to improve traditional sensing and processing strategies (see Subsection 2.1.2). Amongst this last group, SCD Vision is an interesting and original proposal where a frame-based temporal-contrast smart sensing strategy is used to signal relevant changes in the scene under analysis. This novel vision strategy represents a trade-off between biomimetic frameless designs and traditional frame-based sequential image-based systems. SCD Vision takes advantage of some of the data-reduction and smart-sensing

principles of the former ones while presenting a synchronous pixel-flow that can be easily handled by any standard processing hardware. In this chapter, the principles of SCD vision have been explained, showing how an artificial vision system working under this basis, can reduce its bandwidth and processing power requirements while still maintaining accurate results.

SCD sensor VLSI implementation 3

SCD sensor VLSI implementation 3

The basics of SCD Vision have been explained, described and analysed
in Chapter 1 and Section 2.2. This chapter is dedicated to an in-depth
description of the hardware implementation of the principles followed in
this thesis.

There are two approaches to producing a camera according to SCD
Vision principles. One is to make a custom-integrated circuit sensor, and
the other is to construct one from a high-speed camera (>1000 fps), adding
the necessary processing hardware to deliver a customizable number of
pixels from frame to frame, starting with the pixel that has undergone
the largest change since the last time it was read out in descending order.
One of the main objectives of SCD Vision is to work on resource-limited
systems (especially limited in size, power, or processing capabilities) and
also embedded systems, so this second approach hardly fits in with this
objective, because of the power, space, and performance requirement just
for the camera itself. The custom sensor seems more appropriate for this

objective, but it is more complex, initially expensive, and it is very difficult to produce a high-end product at the first attempt, especially in such an unexplored area. Nevertheless, a 32x32 SCD CMOS vision sensor has been designed and tested. In this chapter an in-depth description of this first silicon implementation of SCD principles is given. The sensor has been designed and fabricated using 4-metal 2-poly 0.35 μm austriamicrosystems' fabrication process.

3.1 Working principles

The SCD vision sensor developed in this thesis follows a temporal contrast strategy: successive snapshots of the environment are taken at a fast pace and pixels are selectively transmitted out of the chip in descending order per their illumination level difference with respect to their last transmitted value. When a pixel's address and illumination information are transmitted, the illumination value is stored in an internal pixel memory so that the next forthcoming value can be compared against it. Following this strategy, pixel readout frequency can be adapted to bandwidth availability. In the worst case scenario, one single pixel per captured frame is transmitted, this being the one with the largest illumination change in the entire scene.

For the silicon implementation of such a strategy, every pixel would require two memory elements: one for storing the last read out illumination level, and another memory element for the present one. The difference between the present and the previous illumination levels is physically represented as the output current of a wide-input-range OTA [73]. Another subcircuit is needed to select the maximum amongst all these currents. WTA circuits are a widely used solution in situations where a single winner has to be selected from a group of current or voltage competing values (see [74, 75, 76, 14] amongst many others). If we picture a WTA circuit as a black box, we can see that a set of currents is taken as input, and a set of voltages is produced as output. Every output voltage is associated with a single input current. Ideally, the output voltage associated to the highest input current should be well differentiated from the rest. According to the current polarities in our light sensing circuit, the winning pixel should output a voltage close to 0 V while the losers' output should be close to Vdd.

Given the always-limited discrimination ability of the WTA circuit, a large number of pixels increases the probability of having a certain number of cells competing with equal credentials. Therefore, multiple winners can be observed at the output nodes of the WTA amplifiers. To cope with this scenario, as one of the contributions of this thesis, a digital logic based on the propagation of vertical and horizontal inhibition signals has been proposed in [77]. Following a predefined path across the pixel matrix, inhibition signals are propagated so when the first winning output is found, the following outputs in the propagation path are forced to be close to Vdd. In favour of saving silicon area at the pixel level, only horizontal inhibition signals are used in this particular work; but the general idea of the inhibiting mechanism remains the same.

As soon as a single winner is selected, the output of the WTA digital logic activates the row and column address codifiers. In this way, the address of the pixel together with the captured grey level can be read out.

A block diagram of the circuit can be seen in Fig. 3.1. The rest of this chapter is dedicated to a detailed description of the different circuit blocks.
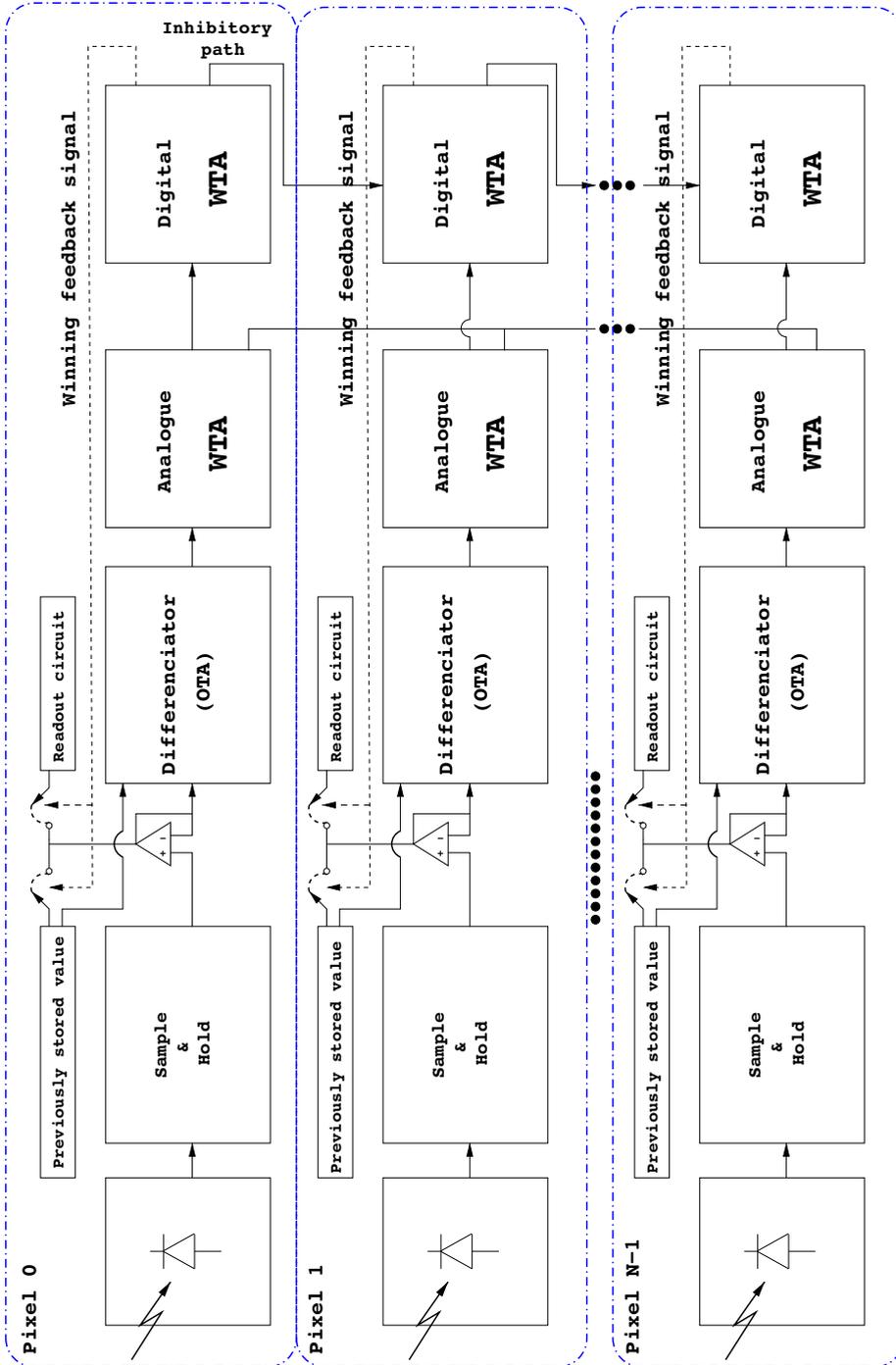
Figure 3.1. Block diagram of the pixel circuit. Address codifiers have been omitted. A detailed diagram of the codifiers can be seen in Fig. 3.15.

3.2 Circuit description

3.2.1 Light sensing and readout subcircuit

The light sensing subcircuit, pictured with a photodiode in Fig. 3.1 and detailed in Fig. 3.2 and Table 3.1, is based on a linear photocircuit [78] [79]. Clock CKphd signals the light sampling period. Transistor MPRST is a minimum size transistor acting as a PMOS switch. At the beginning of every CKphd period, switch MPRST is closed for approximately 1 μs, so the parasitic capacitance of the photodiode is charged to Vreset fixed voltage. During the rest of the sampling period, CKphd = Vdd, meaning that the Vphd node is floating. Under these conditions, incident light produces a photocurrent that forces Cparasitic to linearly discharge down to a certain voltage. The higher the photocurrent, the lower the voltage at the Vphd node.

About 1 μs before the end of CKphd period, sample and hold transistor MPSH and switch MPFLRSW close for about 700 ns, so that Cpresent is charged to a shifted version of the voltage at Cparasitic. Necessary current is sourced by a source follower buffer (transistors MPFLR1-2). In order to reduce non-linearity and to maintain the gain of the source follower stage as close to unity as possible, source and bulk of MPFLR2 are tied to the same potential.

Transistors MPUBF1-6 form a unity gain buffer that holds a copy of Vcap1 voltage at Vpresent node, and at the same time provides enough current to charge and discharge the Cprevious capacitor when the pixel wins the competition at the WTA stage and the voltage at the Vprevious node needs to be updated (see Subsection 3.2.3).

Voltage at node Vcap1 changes only at the beginning of the sampling period. If Vcap1 voltage is static and the MPSWPREV switch is open, only a small amount of current is needed to hold the voltage of the Vpresent node. This is why MPUBF5 sources a small amount of static current, and MPUBF6 is active only when WTAfeedback or CKsh signals are active.

Transistors MPROUT and MPROUT2 form the source follower readout subcircuit. Signal
WTAfeedback is fed back from the output of the digital logic of the WTA subcircuit handling MPROUTSW, MPUBSW1 and MPSWPREV switches in such a way that only the Vpresent voltage of the pixel showing the largest illu-

mination variation is read out at `Vreadout` node, and the voltage at its `Cprevious` capacitor is updated. Transistor `MPROUT` and node `Vreadout` are common to the whole pixel matrix.

Saturation region of `MPUBF5` and `MPUBF6` imposes:

$$\texttt{Vpresent} < \texttt{Vbufflow} + |\mathtt{V_{T\_UBF5}}| - \mathtt{V_{SG\_UBF1}}, \tag{3.1}$$

where $|\mathtt{V_{T\_UBF5}}| = |\mathtt{V_{T\_UBF6}}|$ is the absolute value of the threshold voltage of current sources `MPUBF5` and `MPUBF6` respectively, and $\mathtt{V_{SG\_UBF1}} = \mathtt{V_{SG\_UBF2}}$ is the source-gate voltage of input transistors `MPUBF1` and `MPUBF2` respectively.
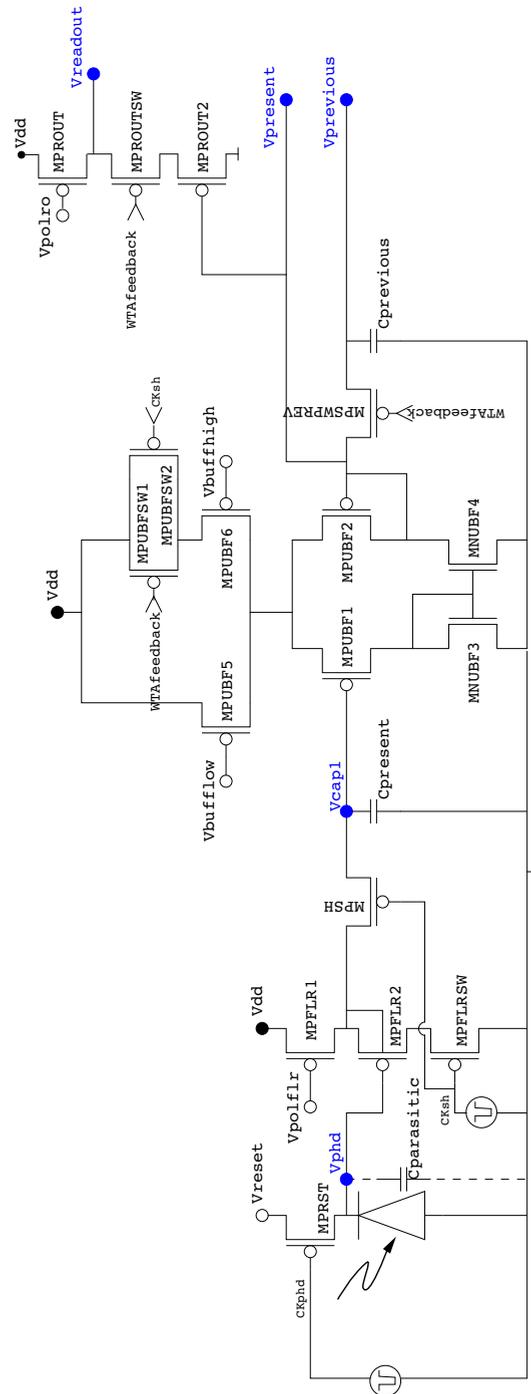
Figure 3.2. Schematic of the photocircuit.

Transistor	W/L [μm/μm]
MPRST, MPFLRSW, MPSH, MPSWPREV, MPUBFSW1, MPUBFSW2, MPROUTSW	0.4/0.35
MPFLR1, MPFLR2	0.7/0.7
MPUBF1, MPUB2	2/1.2
MNUBF3, MNUBF4	0.7/0.7
MPUBF5, MPUBF6	3/1
MPROUT	5/10
MPROUT2	4/0.7

Table 3.1. Transistor sizes for the photocircuit and readout subcircuits according to the schematic of Fig. 3.2.

On the other hand, source follower buffer at node Vphd constraints Vpresent to be Vpresent $>$ V$_{\text{SG\_FLR2}}$.

According to the simulations performed during this research work, the threshold voltage of PMOS transistors is V$_{\text{T\_PMOS}} \cong -0.9$ V for transistors with their source connected to ground, and V$_{\text{T\_PMOS}} \cong -0.7$ V when the source is tied to Vdd. Therefore, Vpresent can be considered to be approximately in the range 1.5 V $<$ Vpresent $<$ 2.5 V. The simulations depicted in Fig. 3.3 corroborate this analysis showing that, for the proposed biasing values, Vpresent is in the range 1.4 V $<$ Vpresent $<$ 2.6 V. If Vreset $>$ 2 V, MPUBF5 and MPUBF6 enter triode region.

Fadeouts and spikes in Vpresent are produced by changes in Vprevious at WTAfeedback negative transitions. WTAfeedback is, in this simulation, represented as a periodical pulse with T$_{\text{high}}$ = 1.98 ms and T$_{\text{low}}$ = 20 μs. In normal operation of the sensor, WTAfeedback is usually high, going low only when the pixel is signalled to be read out.

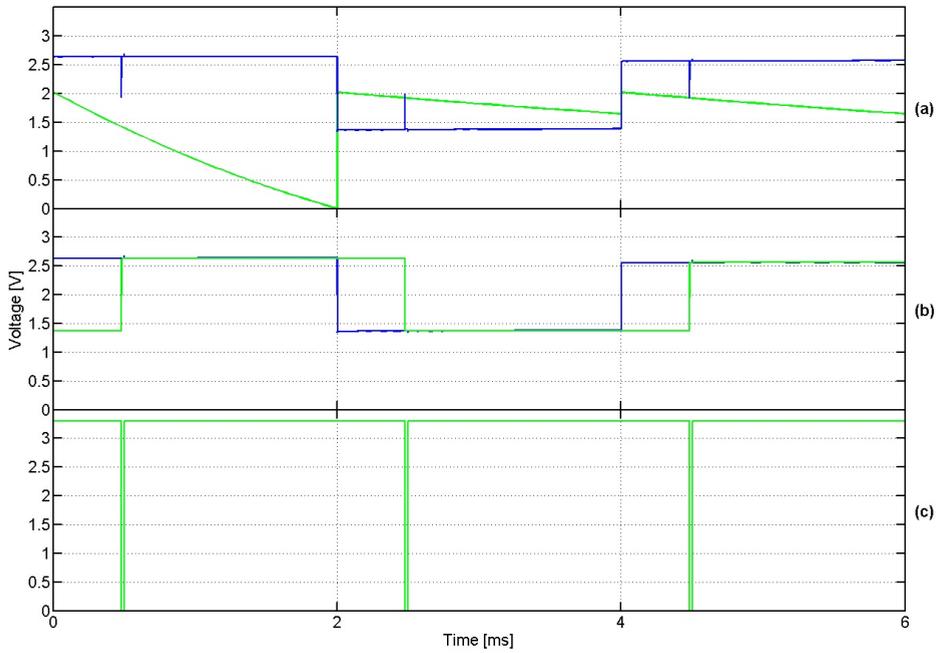**Figure 3.3.** Simulations of the photocircuit operation. Voltage signals at nodes (a)`Vphd` (green) and `Vpresent` (blue), (b) `Vpresent` (blue) and `Vprevious` (green), and (c) `WTAfeedback`. Integration time is 2 ms, low level pulse width of `WTAfeedback` is 20 μs, `Vreset`=2 V, `Vflr`=2.3 V, `Vbufflow`=1.8 V and `Vbuffhigh`=1.6 V.

3.2.2   The wide-input-range operational transconductance amplifier

The goal of this stage is to obtain an output current proportional to the absolute difference between voltages `Vpresent` and `Vprevious`. Since an OTA outputs a current proportional to the voltage difference between its input terminals, this kind of circuit could be used to accomplish this goal. The absolute value of this current could, afterwards, be obtained by implementing a full rectifier circuit.

According to the analysis performed in Section 3.2.1 `Vpresent` and `Vprevious` approximately range from 1.4 V to 2.6 V, so these are the input voltage values that must be used as design constraints. Another important constraint is that, for the considered input range, the output current must vary monotonically.

As the interest is focused on the absolute difference of the input voltages, the output of the OTA-rectifier stage should meet:

$$\texttt{Irect}_{(\texttt{Vpresent-Vprevious})} = \texttt{Irect}_{(\texttt{Vprevious-Vpresent})}, \qquad (3.2)$$

where `Irect`$_{(\texttt{Vpresent-Vprevious})}$ and `Irect`$_{(\texttt{Vprevious-Vpresent})}$ are the rectified output currents for (`Vpresent-Vprevious`) and (`Vprevious-Vpresent`) inputs respectively.

For analysis and design purposes, it is more meaningful to consider differential and common mode input voltages:

$$\texttt{Vd} = \texttt{Vpresent} - \texttt{Vprevious}, \qquad (3.3\text{a})$$

$$\texttt{Vc} = \frac{\texttt{Vpresent} + \texttt{Vprevious}}{2}, \qquad (3.3\text{b})$$

where `Vd` is the differential input voltage and `Vc` is the input common mode voltage. Fig. 3.4 depicts the range of values for both sets of input variables.

A differential amplifier, like the one depicted in Fig. 3.5 could act as an OTA taking `Iota` as output current. The problem is that if the length/width ratio of input transistors are $\texttt{W}_1/\texttt{L}_1 \geq 1$ and $\texttt{W}_2/\texttt{L}_2 \geq 1$, and, as in the austriamicrosystems technology used in this thesis, `Vdd`=3.3 V, the output current saturates long before the differential input voltage meets the desired range. On the other hand, if $\texttt{W}_1/\texttt{L}_1 \ll 1$ and $\texttt{W}_2/\texttt{L}_2 \ll 1$, then current source `MNSOURCE` would easily enter triode region for many of the values in (`Vc` $\pm$ `Vd/2`) range.
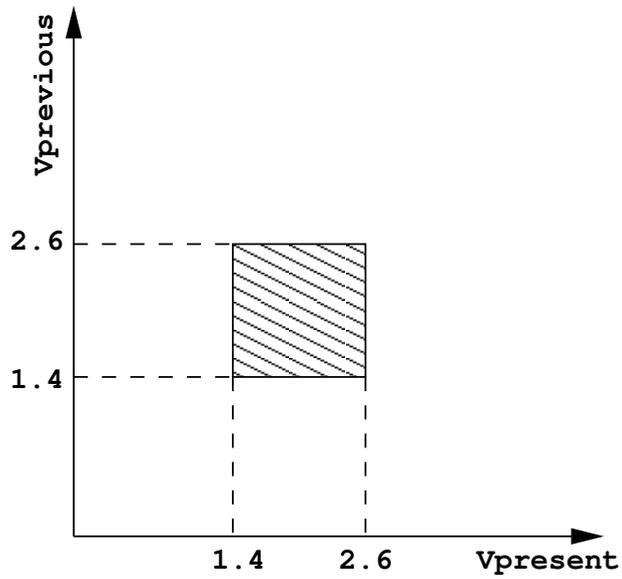
(a)
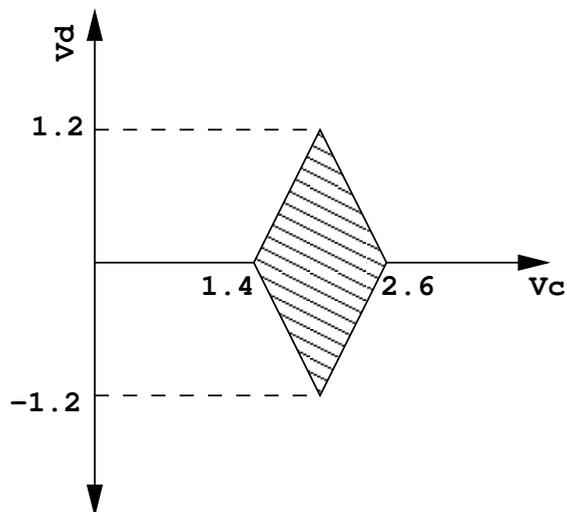
(b)

Figure 3.4. Streaked areas represent the OTA's input voltage range expressed in (a) `Vpresent-Vprevious`, and in (b) common mode (`Vc`)-differential mode (`Vd`) values.

Figure 3.5. Narrow input range OTA. W_1/L_1 and W_2/L_2 are the width/length ratio for transistors MN1 and MN2 respectively.
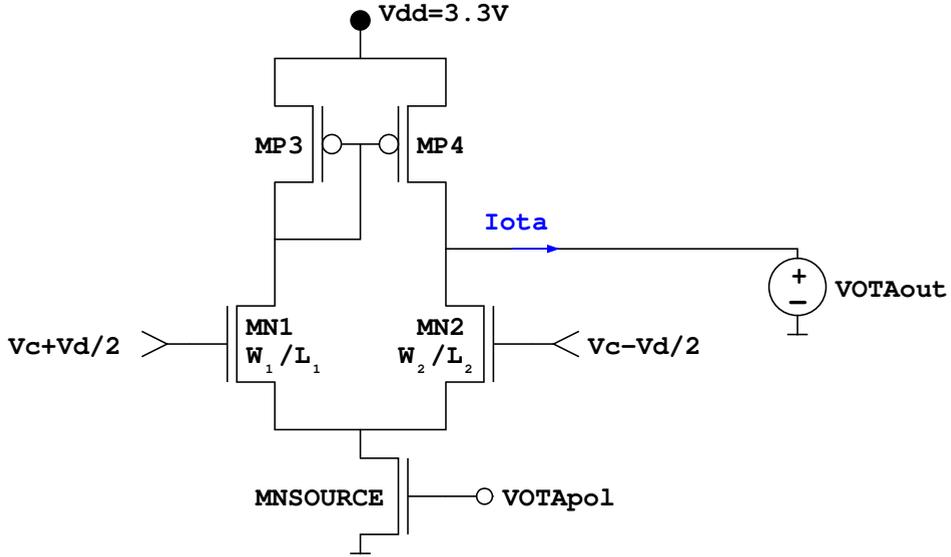
The solution followed in this thesis is to use two differential pairs with asymmetrical transistor sizes. In this circuit, which was first proposed by Nedungadi *et al.* [73], and it is labelled as the OTA block in Fig. 3.6, the output linear range is extended for a wider range of voltage input values. Transistor sizes for the present implementation of the circuit are detailed in Table 3.2. Transistors MNOTA1 and MNOTA4 together with current source MNSOURCE1 form one input differential pair, while the other one is formed by transistors MNOTA2-3 and current source MNSOURCE2.

In order to give some insight into the working principles of this circuit, the simulations in Fig. 3.7 show the variation of the drain current of the OTA's input transistors versus Vd for several values of Vc covering the input voltage range depicted in Fig. 3.4. The asymmetry on the W/L ratio between MNOTA1 and MNOTA4 causes the curve of the drain current of these two transistors to be shifted to positive Vd values, meaning that when Vd$= 0$, MNOTA4 is driving more current than MNOTA1. The equilibrium point at which both transistors drive the same amount of current is for Vd$= V_{eq14}$, with $V_{eq14} > 0$. Similarly, differential transistors MNOTA2 and MNOTA3 drive the same current when Vd$= V_{eq23}$, with $V_{eq23} < 0$.

Figure 3.6. Wide input range OTA and rectifier subcircuits.

Transistor	W/L [μm/μm]
MNOTA1, MNOTA2	1/10
MNOTA3, MNOTA4	2/2
MNSOURCE1, MNSOURCE2	4/2
MPOTA3, MPOTA4	4/2
MNRECT1	1.6/1
MPRECT1	0.8/1
MNRECTin	1/1
MPRECTin	2/1
MPRECT2, MPRECT3	1/0.7
MNRECT2, MNRECT3	2/0.7
MNRECT4	10/10
MNRECT5	4/10

Table 3.2. Transistor sizes for the OTA and current rectifier subcircuits according to the schematic of Fig. 3.6.
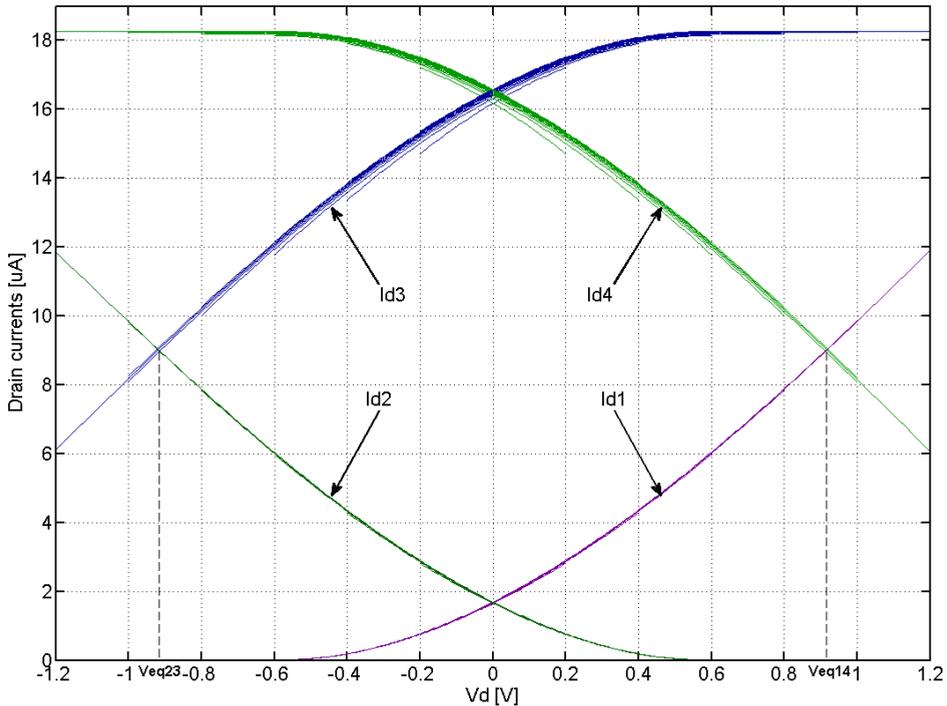

Figure 3.7. Drain current for transistors `MNOTA1-4`. For every transistor, a total of 11 curves are plotted for different `Vc` values ranging from 1.4 V to 2.6 V with a 0.1 V step. Simulations were performed with `VOTApol`=0.9 V.

PMOS transistors `MPOTA3-4` copy `Id1` to the OTA's output node so that,

$$\text{IOTAout} = \text{Id1} - \text{Id2}. \tag{3.4}$$

A graphical representation of `IOTAout` is plotted in Fig. 3.8 showing how the output current increases monotonically with increasing `Vd` values in the desired input range. The simulations also show that the stage is insensitive to common mode variations in the considered range.
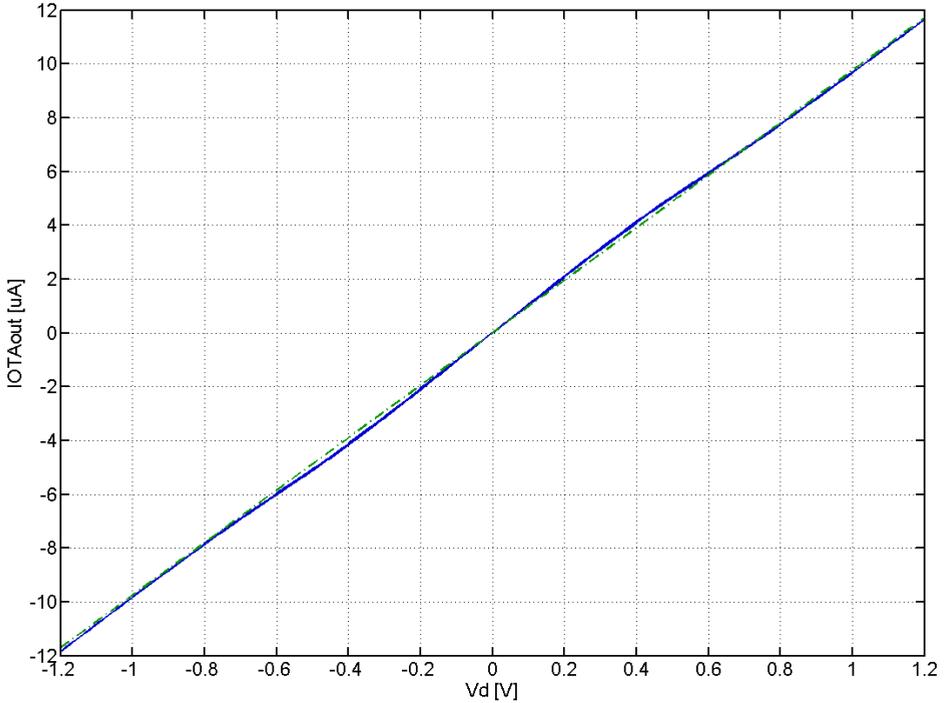
Figure 3.8. OTA's output current `IOTAout` (continuous blue curves). A total of 11 curves are plotted for different `Vc` values ranging from 1.4 V to 2.6 V with a 0.1 V step. Simulations were performed with `VOTApol`=0.9 V. Dashed-dotted green curve is a straight line with slope 12/1.23 μA/V.

Positive values of `IOTAout` force `MPRECT1-MNRECT1` inverter output to polarize transistor `MPRECTin` and to cut-off `MNRECTin`. Therefore $\text{Id}_{\text{Pin}} = \text{IOTAout}$ and $\text{Id}_{\text{Nin}} = 0$. Similarly, when `IOTAout` < 0, `MNRECTin` is in saturation and `MPRECTin` is cut-off, so $\text{Id}_{\text{Pin}} = 0$ and $\text{Id}_{\text{Nin}} = \text{IOTAout}$. As

Id_{Nin} is mirrored by `MPRECT2-3` transistors, in both cases, `IRECTmirror` is positive [80].
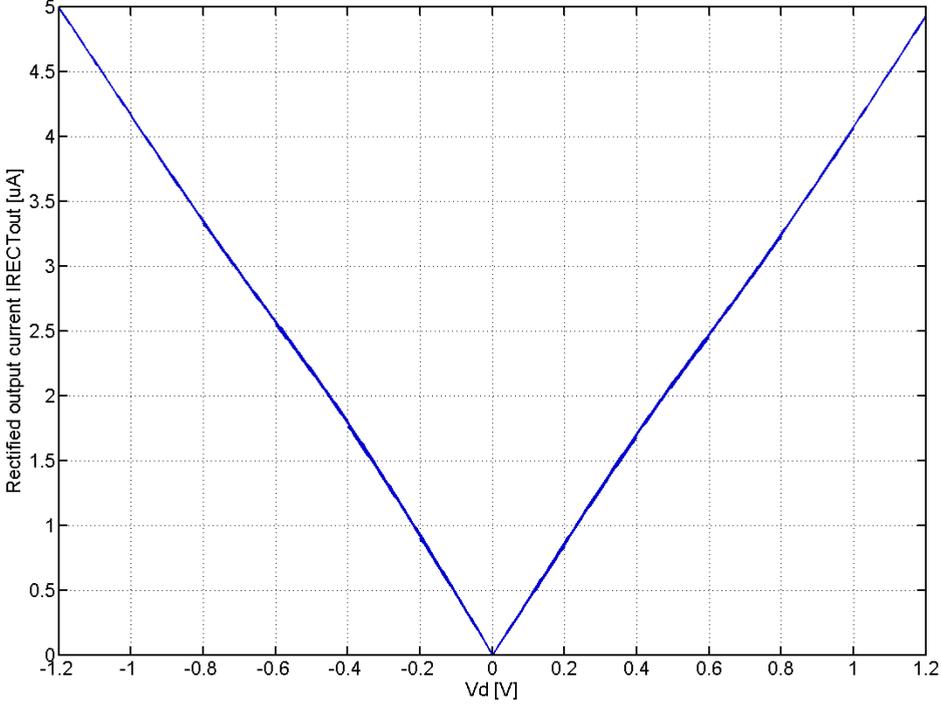
Figure 3.9. Current rectifier's output current `IRECTout`. A total of 11 curves are plotted for different `Vc` values ranging from 1.4 V to 2.6 V with a 0.1 V step. Simulations were performed with `VOTApol`=0.9 V and `VRECTout`=Vdd.

Several `IRECTout` curves are plotted in Fig. 3.9. In these simulations, performed with `VRECTout`=Vdd, the rectifying effect of the stage is perfectly clear. Fig. 3.10 shows the relative percentual error of the rectified output current `IRECTout` between positive and negative `Vd` values expressed as:

$$\epsilon_{r\%\text{IRECTout}} = 100\frac{\text{IRECTout(Vd)}-\text{IRECTout(-Vd)}}{\text{IRECTout(Vd)}}, \qquad (3.5)$$

where $\epsilon_{r\%\text{IRECTout}}$ is the relative percentual error of `IRECTout`. In Fig. 3.10, a detail of the percentual error for $0.5\ V\ <\ \text{Vd}\ <\ 1.2\ V$ is plotted showing that for $\text{Vd}\ >\ 0.5\ V$, the percentual error is smaller than 4.5%. For

Vd $<$ 0.5 V, the percentual relative error increases, but, since small Vd values are not interesting from the SCD Vision point of view, this phenomena does not contradict the objectives of our design.
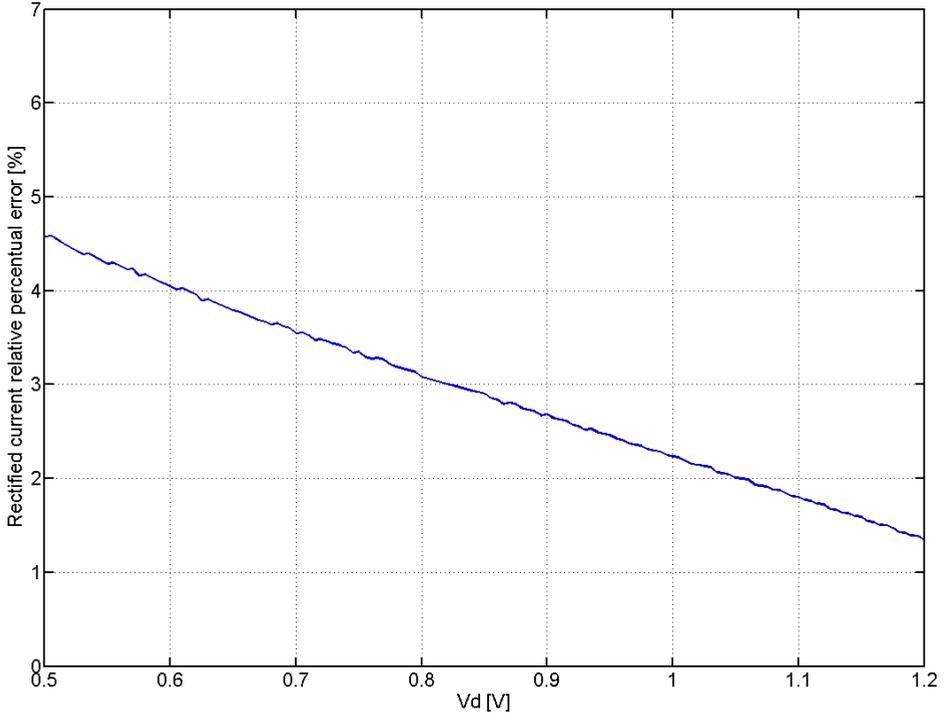
Figure 3.10. IRECTout current rectifier's output relative percentual error expressed as stated in Eq. 3.5. Simulations were performed with Vc = 2 V, VOTApol = 0.9 V and VRECTout = Vdd.

Output transistors MNRECT2-5 are cascoded to attenuate IRECTout fluctuations due to variations in VRECTout voltage. This configuration works properly when the cascoded transitor (MNRECT3) has a high output resistance and the gain transistor (MNRECT5) has a high transconductance. This means that transistor MNRECT3 should be large, and that MNRECT5 should have a large W/L ratio. The sizes of the transistors MNRECT3 and MNRECT5 detailed in Table 3.2 clearly show that the final implementation of this stage is against this criterion. This is because an unwitting mistake was made at the moment of filling in the final data in the Cadence schematic. This

error can be easily corrected, and should be taken into account, in future versions of this sensor.

3.2.3   Winner-takes-all stage

Since their introduction by Lazzaro *et al.* [34], WTA circuits have undergone several modifications (see [35, 36, 37, 38, 39, 40, 81, 82] amongst many others), some focused on improving resolution, time response and power consumption of the circuit, others on controlling its excitatory, inhibitory or hysteretic behaviour. Common applications can be found in neural networks such as [75], or neuromorphic analogue VLSI hardware implementations such as [14, 74, 76, 83, 84, 85, 86].

As explained at the beginning of this chapter, in this thesis the WTA subcircuit followed a mixed analogue-digital design approach [77]. Firstly, the analogue stage discriminates the set of pixels that exhibit the largest illumination changes in the matrix. In a second step, a digital logic selects one single pixel amongst this winning group. These two subcircuits will be explained in this subsection. The simulations presented in Subsection 3.2.3.3 show that the designed circuit meets the necessary requirements to be used as the decision stage in our sensor.

3.2.3.1   Analogue WTA

A basic WTA circuit [34] consisting of two competing cells is depicted in Fig. 3.11(a). `Iin1` and `Iin2` are the input currents to cells 1 and 2 respectively, while $VWTAout_1$ and $VWTAout_2$ are their corresponding output voltages. As $V_{SG1}=V_{SG2}$, when `Iin1=Iin2`, then $V_{SD1}=V_{SD2}$ and $V_{SG3}=V_{SG4}$. Therefore $Id_3 = Id_4 = IWTAtail/2$ and $VWTAout_1=VWTAout_2$. If, for example, `Iin1` rises so that `Iin1 > Iin2`, then $V_{SG1}=V_{SG2}$ also increases. As `Iin2` remains constant, V_{SD2} is forced to decrease moving transistor `MPWTA2` into the triode region. In this way, Id_4 is negligible, $Id_3 \cong IWTAtail$ and $VWTAout_2$ rises to a voltage value close to `Vdd`. $VWTAout_1$ approaches as closely as possible to 0 V, nevertheless keeping `Iin1` current source output transistors in saturation.

In this particular design, as the number of cells is very large, our efforts were mainly focussed on improving the resolution of the circuit. The resolution ability of the WTA circuit is directly related to the gain of each cell's amplifier. The higher the gain, the more accurate the resolution. The gain
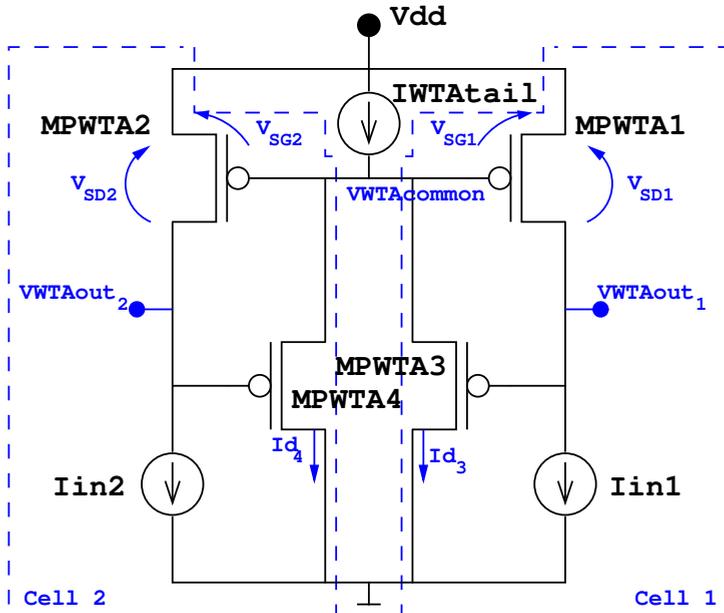
(a) Basic WTA circuit with two competing input currents, Iin1 and Iin2.
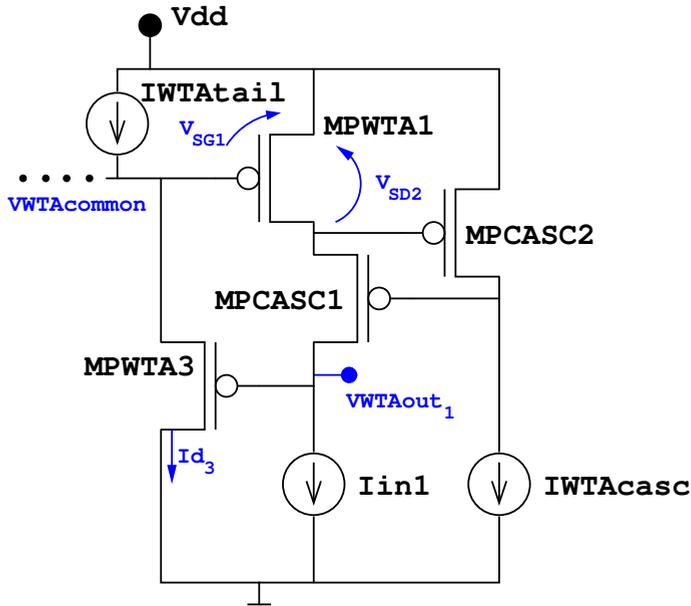
(b) Single WTA amplifier modified with a gain-boosted regulated-cascode stage.

Figure 3.11. WTA amplifiers.

of an amplifier is, classically, its transconductance multiplied by the output impedance. Therefore, an increase in the impedance seen at the WTA's output node would lead to an improvement in its discrimination capability [36, 38]. In [38] the output impedance is increased by means of a basic cascode current source [87]. A more interesting modification to the basic WTA amplifier is the one depicted in Fig. 3.11(b) proposed by Sekerkiran *et al.* in [36]. In Fig. 3.11(b) only one amplifier is represented, this being the same as `cell 1` of Fig. 3.11(a) but with the addition of transistors `MPCASC1-2` and polarization current source `IWTAcasc`.

Transistors `MPWTA1` and `MPCASC1` form a basic cascode stage with `VWTAout1` being its output node. Common emitter transistor `MPCASC2` is polarized by `IWTAcasc` adding an additional gain-path to the signal. `MPWTA1`, `MPCASC1-2` and `IWTAcasc` form a gain-boosted regulated-cascode configuration [88], whose output resistance, $r_{o_{VWTAout}}$, can be expressed as [36, 88]:

$$r_{o_{VWTAout}} = \frac{1}{g_{ds_{WTA1}}} \frac{g_{m_{CASC1}}}{g_{ds_{CASC1}}} \frac{g_{m_{CASC2}}}{(g_{ds_{CASC2}} + g_{o_{IWTAcasc}})}, \qquad (3.6)$$

where $g_{ds_{WTA1}}$, $g_{ds_{CASC1}}$, $g_{ds_{CASC2}}$ and $g_{o_{IWTAcasc}}$ are the output conductances of `MPWTA1`, `MPCASC1`, `MPCASC2` and `IWTAcasc` respectively, and $g_{m_{CASC1}}$ and $g_{m_{CASC2}}$ are the gate transconductances of `MPCASC1` and `MPCASC2` respectively. It can be demonstrated [36] that the gain-boosted regulated-cascode exhibits an output impedance $g_{m_{CASC2}}/(g_{ds_{CASC2}} + g_{o_{IWTAcasc}})$ times larger than the output impedance of a basic cascode stage.

The basic cell depicted in Fig. 3.11(b) is the analogue subcircuit of the decision block implemented in the designed SCD pixel.

3.2.3.2   WTA digital logic

The subcircuit labelled as `WTA digital logic` in Fig. 3.12, generates horizontal inhibition signal that follows the propagation path depicted in Fig. 3.13. Assuming that logic "1" is a voltage close to `Vdd`, logic "0" is close to 0 V, and that the value of the `permit` signal is logic "0", all `InhibitForward` signals in the path are sequentially asserted to "1" until the first winning cell is found. The first winning cell pulls `Verdict` signal to "0", and propagates an `InhibitForward` signal of "0". The logic gates that follow in the propagation path have inputs `InhibitPrevious`=0 and output `Verdict` set to "1". Only one single cell, the first winner found in the propagation path, has its `Verdict` signal in a low state.

Transistor	W/L [$\mu m/\mu m$]
MPWTA1, MPCASC1	10/5
MPCASC2	15/5
MPWTA3	5/5
IWTAtail	1/2
IWTAcasc	1/2
MPinv	0.4/10
MNinv	0.4/0.35
MNstarved	0.4/0.35
MNWORr_r_c, MNWORc_r_c	0.7/0.35
MNbitnc, MNbitnr	0.4/0.35
Address-codifier transistors of row/column r/c	1.2/0.35

Table 3.3. Transistor sizes for the analogue/digital WTA circuit of Fig. 3.12 and Fig. 3.15 . `IWTAtail` and `IWTAcasc` have been implemented with single externally polarized PMOS transistors. For a 32x32 pixel matrix, the range of subscript n is $0\ldots4$, as well as the range of subscripts r and c is $0\ldots31$.

Transistors `MPinv`, `MNinv` and `MNstarved` form a starved inverter that acts as a single bit analogue/digital interface presenting high impedance to the analogue output.

Care must be taken so the charge at `Cprevious` capacitor is updated for the winning pixel only. Transmission gates `TXG1` and `TXG2`, and inverters `INVA` and `INVB`, form a D latch whose objective is to feed a stable `WTAfeedback` signal into the photocircuit.

When `permit` signal is in high state, `InhibitForward` is also high, so the inhibiting mechanism of the digital logic is disabled. The purpose of the `permit` input is to test the WTA winner selection with and without the digital logic. In this way, the necessity of the logic circuit in the winner selection can be experimentally determined.

3.2.3.3 WTA simulations

Simulations published in [77, 89] show that the WTA circuit with improved resolution using a gain-boosted regulated-cascode configuration is capable of handling a matrix of 32x32 cells. In that case the input currents used
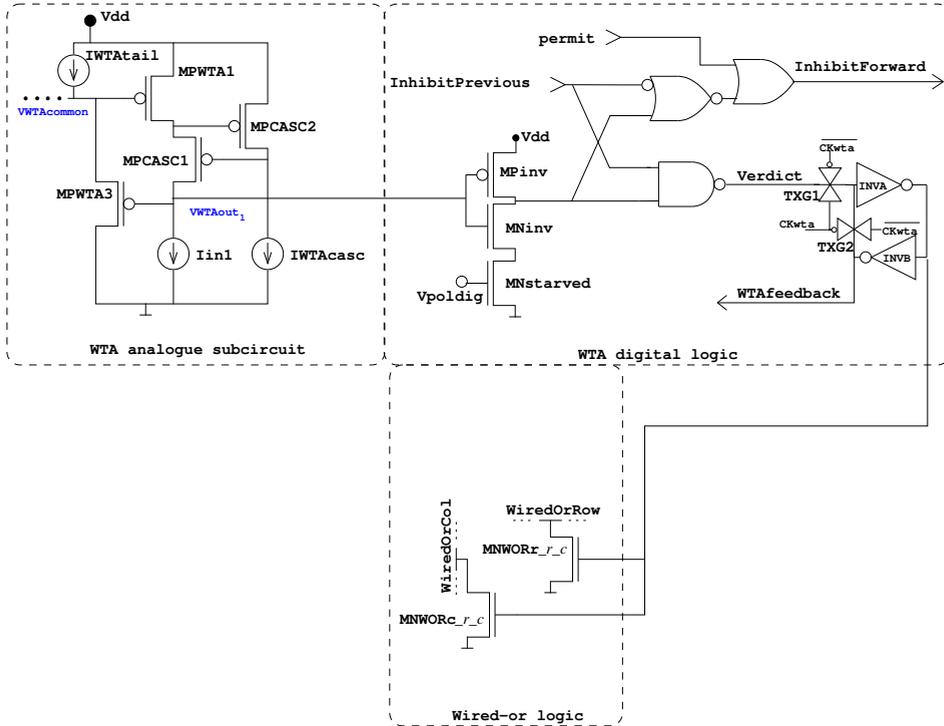
Figure 3.12. WTA analogue/digital circuit for single winner selection. Subscript notation _r_c indicates row r and column c.

to stimulate the circuit were linearly spaced between 0 and 1.25 μA. The combined analogue-digital subcircuits were able to find a single winner with a delay of 1 μs. For the particular objectives pursued in this thesis, these simulations are not very realistic. In real applications of the SCD sensor it is expected that most of the time only a small portion of the pixels undergo relevant changes. It is also expected that the changes observed in this small set of pixels are much larger than the changes observed in the rest of the matrix. In this section new simulations reflecting this fact are presented.

In Fig. 3.14 the results of a simulation of a WTA matrix compatible with the decision stage included in the SCD pixel matrix is presented. Only the OTA, current rectifier, WTA and starved inverter acting as an analogue-digital interface (transistors MPinv, MNinv and MNstarved of Fig. 3.12) are included. The OTA-rectifier and the starved inverter are, respectively, the

Figure 3.13. Propagation path for WTA digital inhibition signals.

input and output of the decision block, while the WTA is the decision block itself. The common (Vc) and differential (Vd) voltages at the input terminals of the OTAs are used to stimulate the simu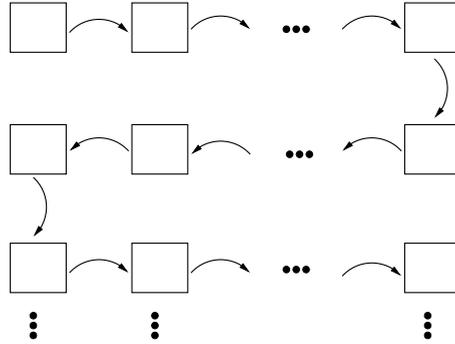lated matrix. Its values are set in concordance with the streaked area of Fig. 3.4(b). Current source IWTAcasc is implemented with a NMOS transistor with a gate voltage equal to 0.9 V, resulting in a drain current of 4.5 μA. Current source IWTAtail is implemented with a PMOS transistor polarized with a gate voltage of 2.0 V, resulting in a source current of 3 μA.

For the first 10 μs of the .TRAN simulation, Vc = 2 V and Vd = 0 V for every OTA in the matrix. After that instant, Vc remains at 2 V, but Vd changes in the following manner:

$$\text{Vd}_k(10\mu s < t < 22\mu s) = \begin{cases} 1 + k \cdot \frac{0.2}{1024} & k = 0 \ldots 63, \\ k \cdot \frac{0.2}{1024} & k = 64 \ldots 1023, \end{cases} \tag{3.7}$$

$$\text{Vd}_k(t > 22\mu s) = \begin{cases} 0 & k = 0, \\ 1 + k \cdot \frac{0.2}{1024} & k = 1 \ldots 63, \\ k \cdot \frac{0.2}{1024} & k = 64 \ldots 1023, \end{cases} \tag{3.8}$$

where Vd_k is the differential input voltage to cell k, with k calculated as $k = 32 \cdot row + col$, $row = 0 \ldots 31$ and $col = 0 \ldots 31$, and t represents time. According to Eqs. 3.7 and 3.8, at $t = 10$ μs the differential inputs of first 64 cells are set to values much larger than the other cells in the matrix. Therefore, these cells should win the competition and the output of their

starved inverters should be set to logic "1". This is accomplished with a delay of 11.4 μs. As can be seen in Fig. 3.14(b), at $t = 21.4\ \mu s$ the outputs of the starved inverters of the first 64 cells are asserted to "1", while the output of the other inverters remain at low logic level.

At $t = 22\ \mu s$, when the outputs of the inverters are stable, the differential input of the first cell in the matrix, Vd_0, is changed from 1 V to $Vd_0 = 0\ V$. This is what would happen if the capacitor Cprevious had been updated to have the same voltage as Cpresent. The simulations show that 700 ns after the input changes, the output of the starved inverter is asserted to "0".

According to the simulations in Figs. 3.14(b) and 3.14(a), when the number of cells that change their input values is large, it takes approximately 12 μs to obtain stable outputs that reflect the state of the Vd inputs. This is similar to what would happen at the end of a frame period. The voltages at Cpresent capacitors are updated with new illumination values coming from the environment, so the Vd values of all the cells in the matrix suddenly change to new levels. After this initial step, the charge at the Cpresent capacitors remain the same. Only the voltage at some of the Cprevious capacitors change one by one. This situation is reflected by the change of Vd_0 at $t = 22\ \mu s$. In this case, it can be seen that the delay in the decision stage is reduced to only 700 ns.
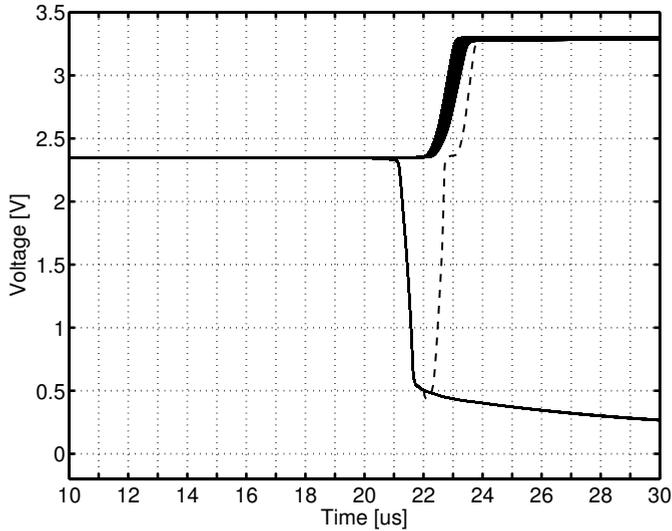
(a) VWTAout$_k$, for $k = 0 \ldots 1023$. The signals close to Vdd correspond to the cells with the lowest differential input values. The signals below 0.5 V correspond to the cells with the largest differential inputs. The dashed line corresponds to cell 0.
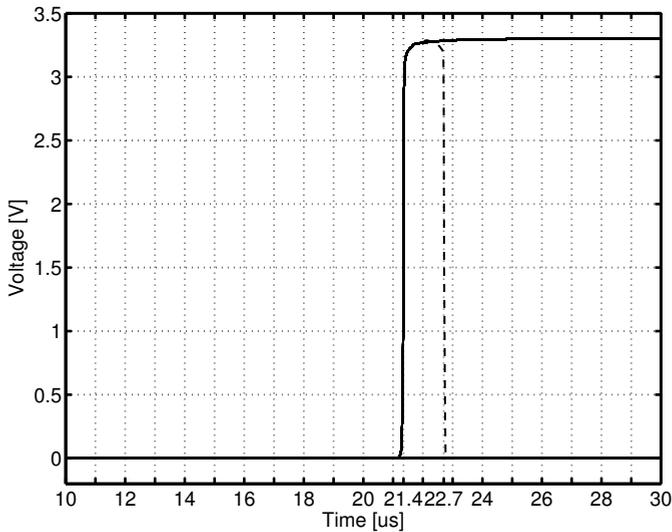
(b) Starved inverter outputs. The signals close to Vdd correspond to the cells with the largest differential input values. The signals close to 0 V correspond to the cells with the lowest differential inputs. The dashed line corresponds to cell 0.

Figure 3.14. Simulation of the decision stage of the pixel matrix. Only the OTA, rectifier, WTA and starved inverter blocks were simulated. The matrix was excited with the Vd distribution detailed in Eqs. 3.7 and 3.8.

3.2.4   Wired-or logic and address codifiers

Whenever signal `Verdict` is pulled low, wired-or vertical and horizontal transistors `MNWORr_r_c` and `MNWORc_r_c`, are activated signalling the pixel of row `r` and column `c` as the winner of the `WTA` competition stage. According to the mixed analogue-digital WTA subcircuit behaviour, only one single pixel activates its wired-or outputs at a time. Therefore, these row and column signals can be easily codified to output a binary representation of the `r-c` matrix-coordinates of the winning pixel. An example of such a codifying scheme is depicted in Fig. 3.15 for a pixel located in row 19 and column 9. Transistors MNbitnr and MNbitnc, with $n = 0 \ldots 4$ for a 32x32 pixel matrix, are common to all the columns and rows respectively.

An image of the layout of address-codifiers for row 19 together with pull-up transistor `MPWORr_19` can be seen in Fig. 3.16(a) and a micropho-tograph of the row codifiers of rows 30 and 31 is shown in Fig. 3.16(b).

3.2.5   Complete schematic of the pixel

In Fig. 3.17 a complete schematic of the pixel circuit is represented with all the blocks described in this chapter interconnected together.
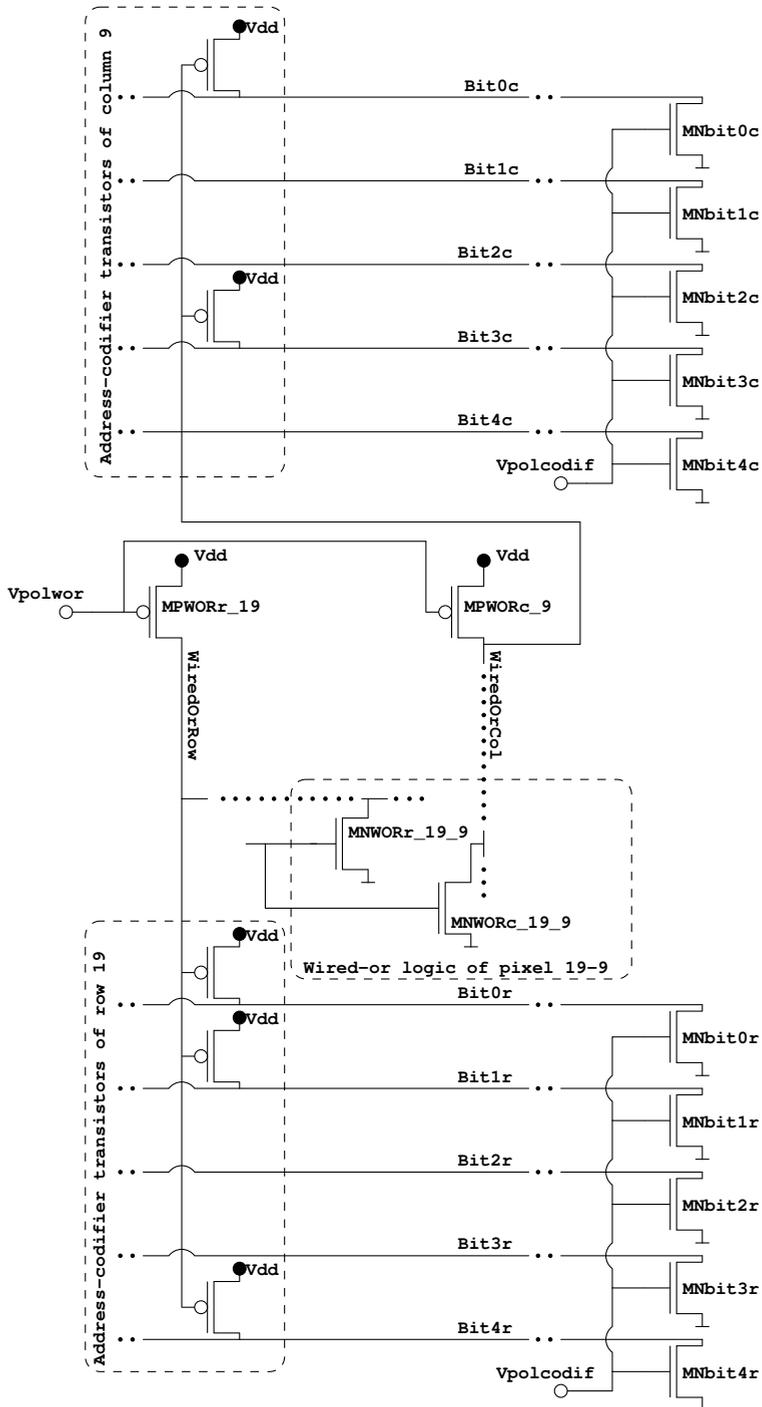
Figure 3.15. Wired-or and address-codifier transistors of the pixel in row 19 and column 9.
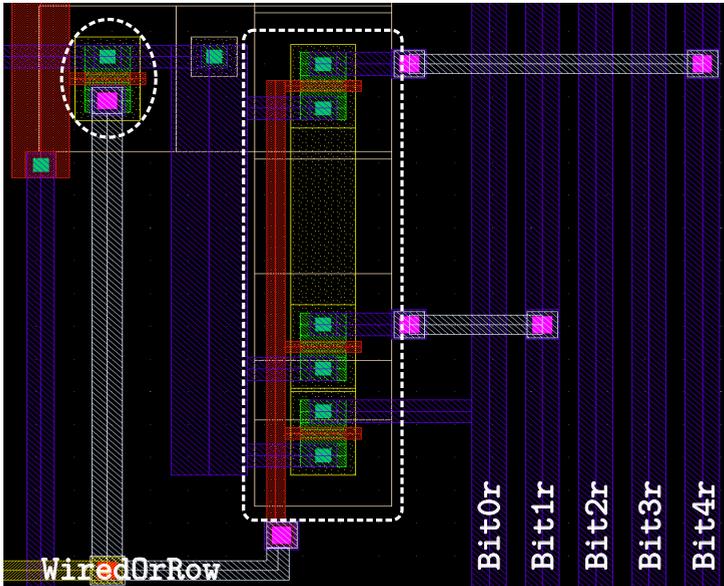
(a) Address-codifier transistors of row 19 (white dashed rectangle) together with pull-up transistor `MPWORr_19` (white dashed ellipse).
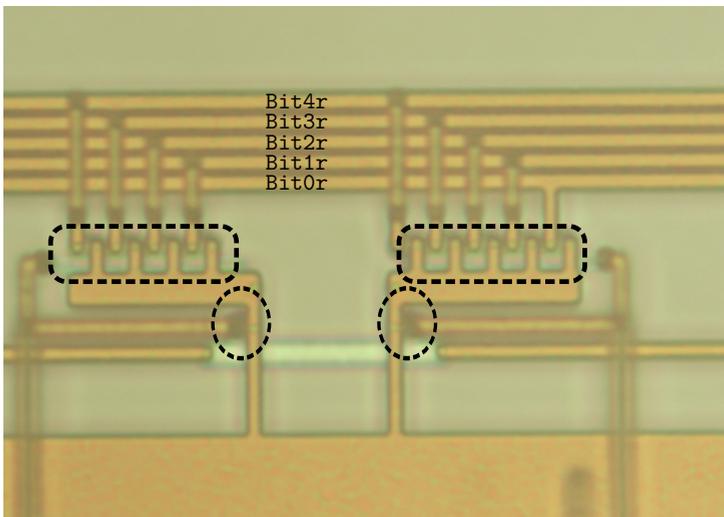
(b) Microphotograph of the address-codifier transistors of rows 30 and 31 (black dashed rectangles) together with pull-up transistors `MPWORr_30-31` (black dashed ellipses).

Figure 3.16. Examples of the layout of the address-codifiers.
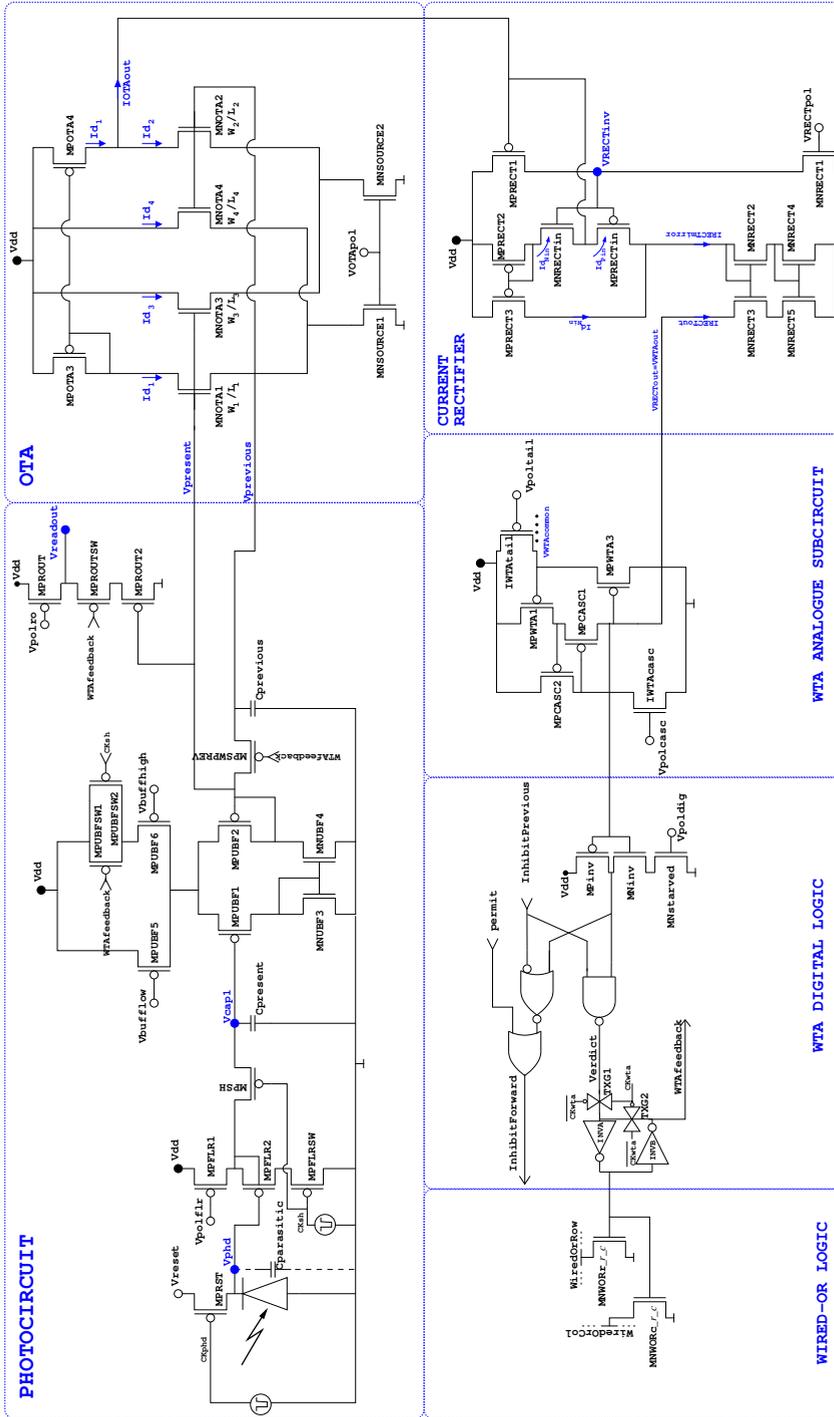
Figure 3.17. Full schematic of one single pixel. For a detailed explanation of every subcircuit, see Chapter 3.

3.2.6   Pixel layout

The previously described circuit has been implemented in austriamicrosystems' CMOS 0.35 μm 4-metal 2-poly process. To form the pixel matrix, the basic reproducible layout unit is composed of four pixels. An image of such a structure is depicted in Fig. 3.18. The main advantage of this 4-pixel layout is that the analogue subcircuits are very well separated from the digital ones. In order to minimize parasitic photocurrents, the whole pixel layout, except for the photodiode, is covered with the top metal layer.

Cpresent and Cprevious have been implemented as irregularly shaped polySi-insulator-polySi (PIP) capacitors with an approximate capacitance of 158 fF and 585 fF respectively. The size of the photodiode is 144 μm^2. Pixel pitch and fill factor are 55 μm and 4.8% respectively. The node VWTAcommon is common to the whole pixel matrix, so a common metal path was routed reaching all the pixels in the matrix.

The VWTAcommon voltage is not stationary, but changes every time the WTA decision block needs to find a new winner. To avoid unwanted resonant feedback loops, the VWTAcommon signal was routed in such a way that no closed loops were formed. A simplified representation of the routing of this signal is depicted in Fig. 3.19.

The size of the whole chip, including pads, is approximately 2.8 x 2.8 mm^2. An image captured from the Cadence layout tool can be seen in Fig. 3.20.

The bonding diagram is depicted in Fig. 3.21. As the die is attached to the cavity with conductive glue, the ground terminal (pin 20) has been connected to the cavity for better substrate polarization. The pixel matrix of Fig. 3.20 and the bonding diagram of Fig. 3.21 have the same orientation. The pinout of the chip is described in Table 3.4.

Microphotographs of the silicon implementation can be seen in Figs. 3.22 and 3.23.
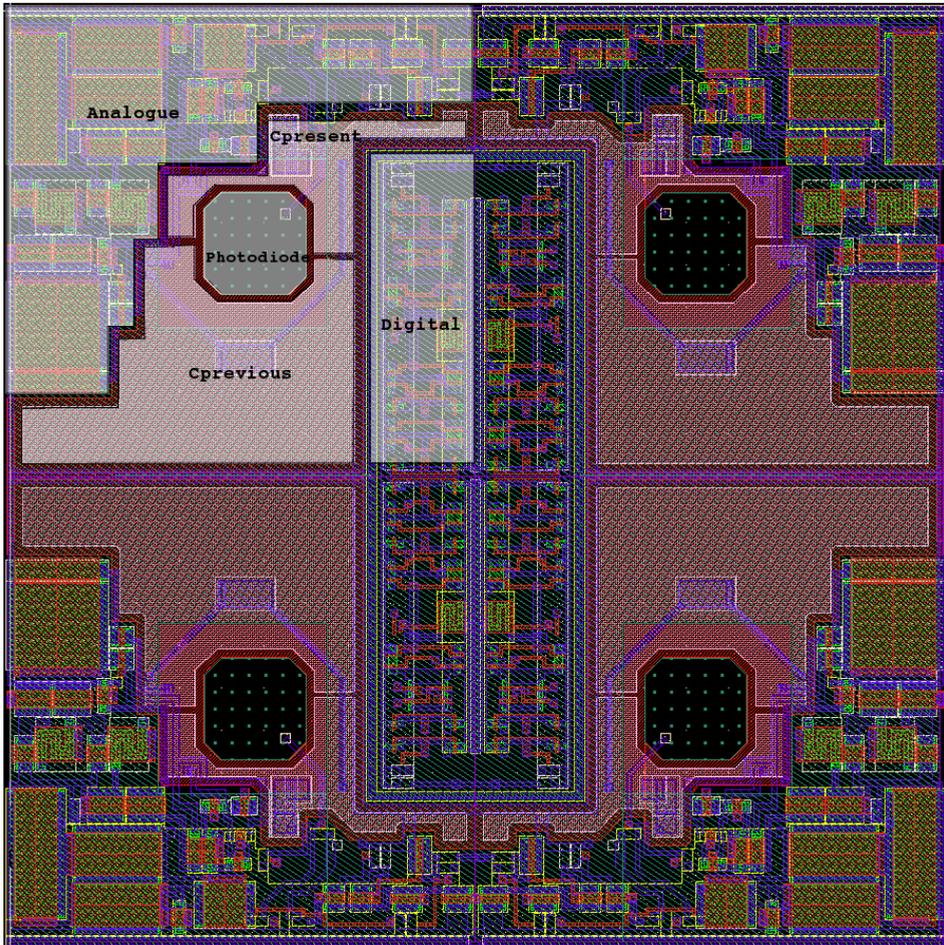
Figure 3.18. Basic 4-pixel reproducible layout unit. The size of one single pixel is 55μm x 55μm.
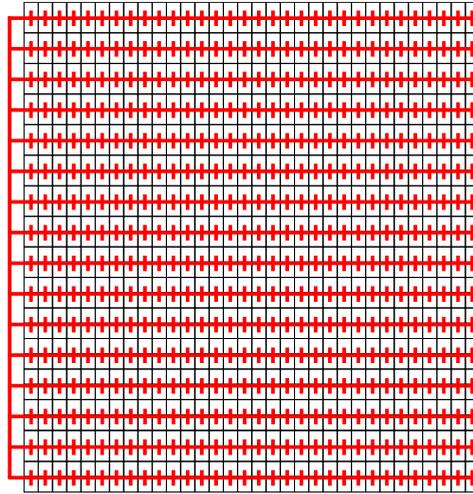
(a) Routing of the `VWTAcommon` node across the pixel matrix.
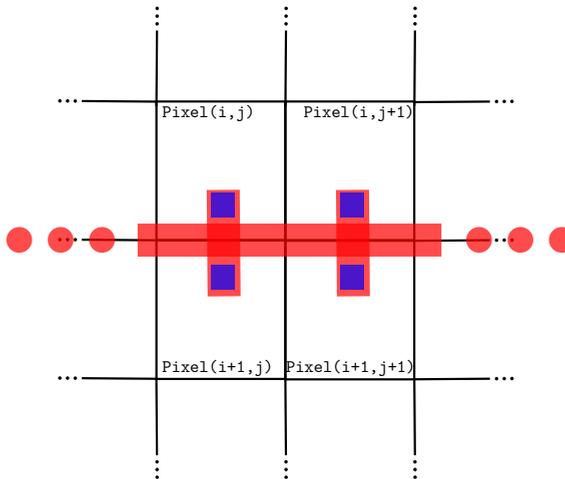
(b) Detail of the routing of the `VWTAcommon` node for a four-pixel area.

Figure 3.19. Simplified representation of the routing of the `VWTAcommon` node. Red lines and blue squares represent the metal-3 path of the `VWTAcommon` node and vias to lower metal layers respectively. This node was routed without any closed loops. Dimensions are not scaled.

Name	Pin number	Type	Direction	Description
Bit<0:4>c	31:35	Digital	Output	Binary-codified column address of the pixel selected to be read out (see Subsection 3.2.4).
Bit<0:4>r	26:30	Digital	Output	Binary-codified row address of the pixel selected to be read out (see Subsection 3.2.4).
CKphd	22	Digital	Input	Clock signal for the reset of the photodiode (see Subsection 3.2.1).
CKsh	23	Digital	Input	Clock signal for the sample & hold transistor of the photocircuit (see Subsection 3.2.1).
CKwta	37	Digital	Input	Clock signal for the latch of the WTA digital logic (see Subsection 3.2.3.2).
gnd	20, 21	Power	Input/Output	Ground.
permit	36	Digital	Input	Signal to enable (permit = "0") or disable (permit = "1") the inhibiting logic mechanism of the WTA subcircuit (see Subsection 3.2.3.2).
Vbuffhigh	8	Analogue	Input	High-current polarization of unity gain buffer of the photocircuit (see Subsection 3.2.1).
Vbufflow	12	Analogue	Input	Low-current polarization of unity gain buffer of the photocircuit (see Subsection 3.2.1).

Table 3.4. Pinout of the SCD sensor.

Name	Pin number	Type	Direction	Description
vdd	1, 40	Power	Input/Output	Power supply.
VOTApol	13	Analogue	Input	Polarization for the MNSOURCE1-2 current sources of the OTA subcircuit (see Subsection 3.2.2).
Vpolcasc	14	Analogue	Input	Polarization for the IWTAcasc current source of the analogue WTA subcircuit (see Subsection 3.2.3.1).
Vpolcodif	6	Analogue	Input	Polarization for the row/column codifiers (see Subsection 3.2.4).
Vpoldig	15	Analogue	Input	Polarization for the starved inverter of the digital subcircuit (see Subsection 3.2.3.2).
Vpolflr	9	Analogue	Input	Polarization for the source follower of the photocircuit (see Subsection 3.2.1).
Vpolro	4	Analogue	Input	Polarization for the readout subcircuit (see Subsection 3.2.1).
Vpoltail	10	Analogue	Input	Polarization for the IWTAtail current source of the analogue WTA subcircuit (see Subsection 3.2.3.1).
Vreadout	5	Analogue	Output	Analogue grey level output (see Subsection 3.2.1).
Vpolwor	7	Analogue	Input	Polarization for the row/column wired-or logic (see Subsection 3.2.4).
VRECTpol	19	Analogue	Input	Polarization for the MNRECT1 transistor of the current rectifier subcircuit (see Subsection 3.2.2).
Vreset	11	Analogue	Input	Reset level for the photodiode (see Subsection 3.2.1).

Table 3.4. Pinout of the SCD sensor.
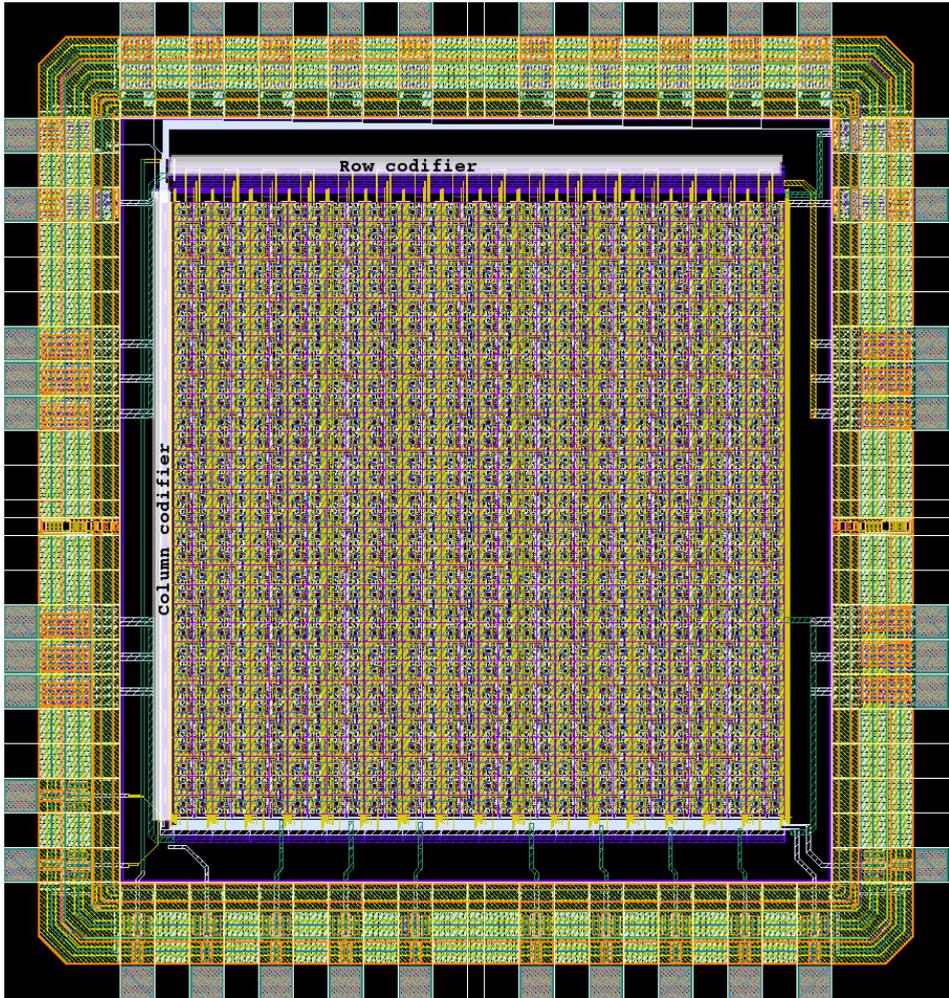
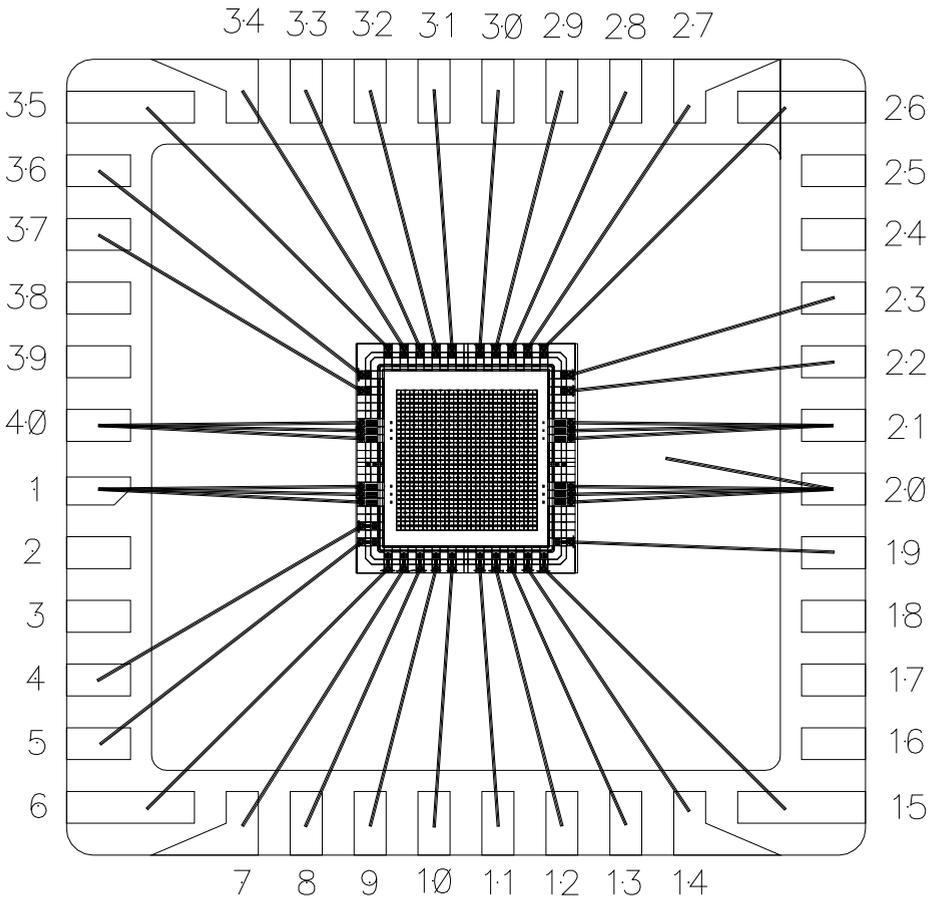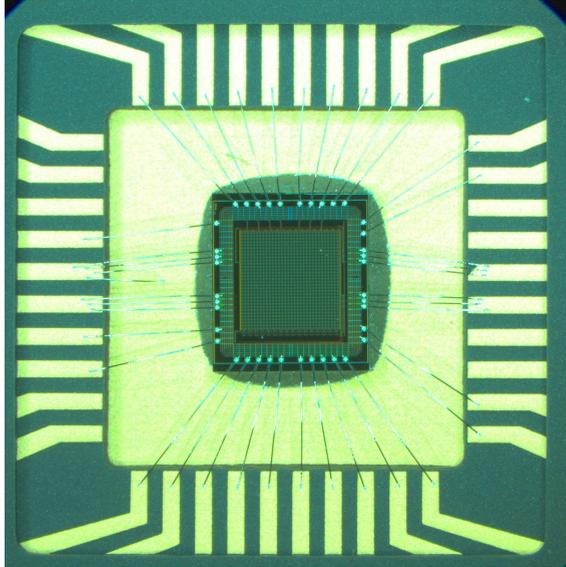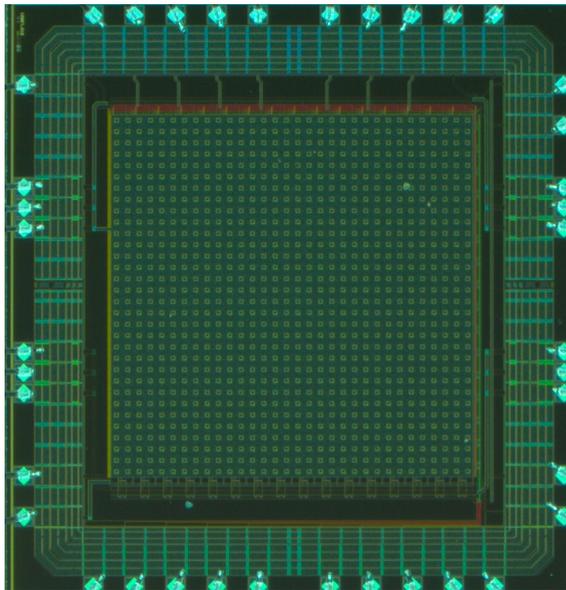**Figure 3.20.** Pixel matrix. 2.8 x 2.8 mm^2.

Figure 3.21. Bonding diagram.

(a) Bonding.

(b) Full pixel-matrix.

Figure 3.22. Microphotographs of the chip.
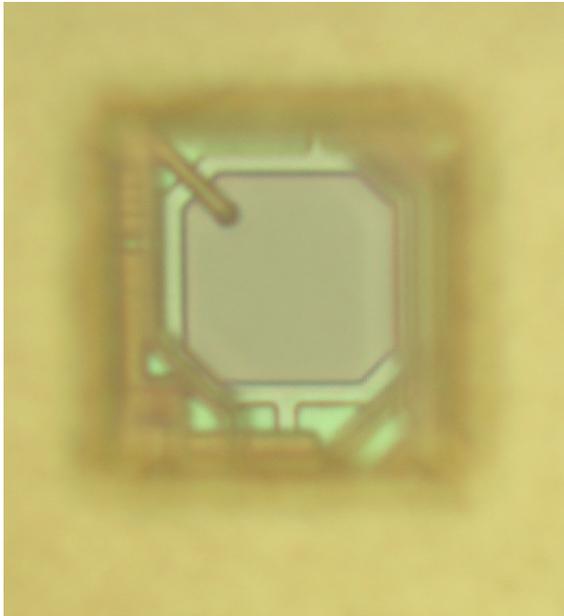
(a) Detail of the photosensitive area.

(b) Detail of the fourth metal layer covering the whole pixel matrix. Only the photodiodes are exposed to incident light.

Figure 3.23. Microphotographs of the chip.

3.3 Conclusions

In this chapter, the design of the first pixel following SCD Vision strategy has been presented. Using this pixel as the basic cell, a 32x32 matrix has been implemented. According to the SCD strategy, the pixel showing the largest change in illumination should be selected amongst the whole matrix. A mixed analogue-digital WTA subcircuit is used to accomplish this objective. Since our sensor requires a large number of WTA amplifiers, and until the beginning of this thesis we could not find any implementation with such a large number of cells, the resolution of this subcircuit has been one of the main challenges in this design. Any WTA implementation has finite resolution capabilities, so a multiple winner situation could easily arise. An improved resolution analogue WTA amplifier [36], together with the digital selection logic published in [77] as part of the original work done during this research, have shown to successfully cope with the multiple winner problem. In our circuit, following a predefined path, a selection logic looks for the first winning pixel in the matrix; as soon as this first winning pixel is found, an inhibiting signal is propagated forcing the rest of the pixels to be losers.

The main drawback of this solution is that the path of the selection logic and the inhibiting signals is predefined, thus biasing the solution according to the location of the pixels in the matrix: for pixels exhibiting similar changes in illumination, the first ones in the path are more suitable to win the competition. In a future version of the sensor, this problem can be solved by changing, according to a certain criteria, the position of the first pixel in the search path from one pixel-period, or frame-period, to the next one. Other topologies for the inhibiting signal path can also be explored. In Fig. 3.24 the inhibiting path starts at the centre of the chip and propagates following a spiral to the periphery. The closer to the centre of the chip that a change is produced, the more relevant it will be considered.

The chip was fabricated with 4-metal 2-poly 0.35 μm technology from austriamicrosystems. From the layout point of view, the pitch of the pixel and the fill factor are 55 μm and 4.8% respectively. The majority of the area of the pixel is dedicated to the PIP capacitors that store the present and previous illumination values. By using other fabrication technologies that incorporate other types of capacitors with increased capacitance per unit area, i.e. metal-insulator-metal (MIM) capacitors in austriamicrosys-

tems' 180 nm, or LFOUNDRY 150 nm technologies, the area of the pixel can be further reduced. Moreover, the minimum transistor size in these technologies is much smaller than in the technology used in this thesis. If these technological options are used in future developments, considerable improvements in important parameters such as fill factor and total area of the chip could be achieved.
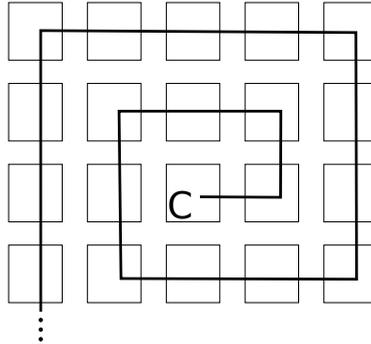
Figure 3.24. Proposal for a different topology for the inhibiting signal path. The small squares represent the pixels in the matrix. The pixel labelled with letter C is in the centre of the matrix. Its precise address depends on the matrix size.

Measurements and results 4

Measurements and results 4

Silicon is so painfully truthful.

———————————————————————

T. Delbrück.

4.1   Experimental set-up and sensor operation

A resource-limited system has been developed in order to demonstrate the advantages of using SCD Vision and the SCD sensor in these kinds of systems. The system has been designed to have restrictions, especially on power consumption and processing capabilities. The system is already small, but due to it being a prototype and for demonstration purposes, size has not been a design parameter. A standard development board has been used as the processing hardware, while a customized PCB was developed for the SCD camera. Power consumption is limited, since all the power comes from the Universal Serial Bus (USB) cable that also serves as a communication link with a computer; this initially limits the current to 100 mA (5 V). The processing power is also limited by a 32-bit 80 MHz microcontroller that drives the SCD sensor and communicates with a computer by means of the USB interface.

Figs. 4.1 and 4.2 show the building blocks and several photographs of the system. The core is a Microchip PIC32MX795F512L microcontroller running at 80 MHz. It has very low processing power, but it includes direct interfaces to USB and Inter-Integrated Circuit (I2C) buses, and many programmable digital input/output (I/O) pins, utilized in this case for generating the clocks of the SCD sensor and for reading the pixel grey levels and addresses coming from the sensor. The constant analogue voltages necessary for the sensor are generated by a set of 12 digital-to-analogue converters programmed by the microcontroller using the I2C interface.

Sensor readout and synchronization are depicted in Fig. 4.3. The three synchronization clocks, grey level signal, and address buses are represented. Every integration time a new frame acquisition starts, the end of the integration time is ordered with a negative pulse (1 μs) of the CKsh signal. At this moment, every pixel stores the illumination value of the photodiode, so a new acquisition can start while the information is read out. Immediately after CKsh is again high, the CKphd signal (photodiode reset) is driven low for a short period (1 μs) initiating a new image acquisition. The pixel flow can start after the rising edge of CKsh. Every new pixel read is initiated after the falling edge of the CKwta. The CKwta pulse must be shorter than 100 ns to avoid undesired WTA feedback. At every falling edge of CKwta, the winning pixel is latched and read out, and a new competition starts again. Since the last read out pixel updates its Cprevious capacitor with the current illumination level, the analogue-digital selection mechanism will consider that this pixel is not interesting any more and it will probably lose all remaining competitions until the next frame is captured. A new winning pixel is always obtained at every falling edge of CKwta with decreasing interest (from the movement analysis point of view) as the time advances and until a new frame is acquired again. The sensor delivers the analogue grey level values and addresses immediately after the falling edge of CKwta.

The internal analog-to-digital converter (ADC) of the microprocessor requires about 1-2 μs to convert the analogue signal, but two other time constraints must be considered to determine the period of the CKwta signal: the time that the WTA circuit needs to determine a winning pixel, and the settling time of the grey level signal. According to the simulations presented in Subsection 3.2.3.3, the delay of the WTA is in the order of 12 μs when a large number of signals changes at the same time and around 1 μs when only one signal changes. The former is similar to what happens immediately

after the rising edge of the `CKphd` pulse, whereas the latter corresponds to the subsequent competitions until the next falling edge of `CKsh`. The results presented in the next sections of the chapter show that about 10 μs are needed to obtain a clear analogue grey level signal (see Subsection 4.2.1). The microcontroller drives the three clocks of the readout scheme, so it can fix the pixel rate at any time depending on the workload.

In concordance with the simulations of the different subcircuits presented in Chapter 3, the most relevant polarization values used in the measurements presented in this chapter are detailed in Table 4.1.
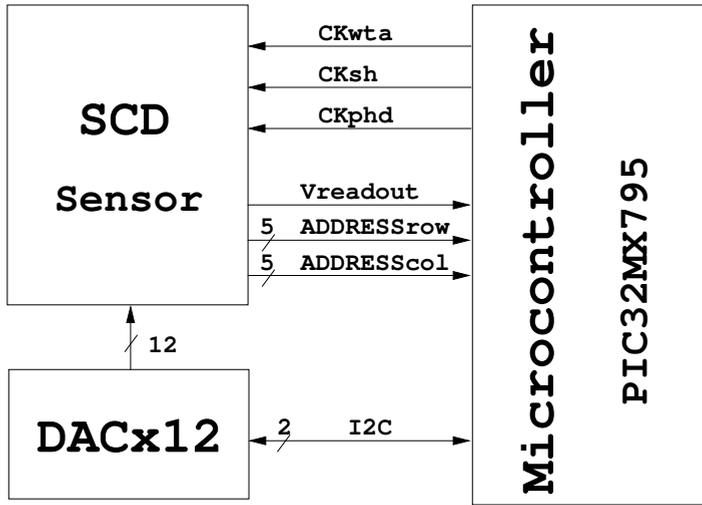
Figure 4.1. SCD camera and processing system. `Vreadout` is the analogue grey level signal (see Fig. 3.2), `ADDRESSrow` and `ADDRESScol` are the row and column address bus, respectively. `CKsh` and `CKphd` are the photocircuit clock signals (see Fig. 3.2), and `CKwta` is the clock signal for the WTA digital logic (see Fig. 3.12).

Polarization	Voltage [V]	Polarization	Voltage [V]	Polarization	Voltage [V]
`Vdd`	3.3	`Vreset`	2.0	`Vpolflr`	2.3
`Vbuffhigh`	1.6	`Vbufflow`	1.8	`Vpolro`	2.1
`VOTApol`	0.9	`VRECTpol`	0.3	`Vpoltail`	2.0
`Vpolcasc`	0.9	`Vpoldig`	0.5		

Table 4.1. Polarization values used in the testing and measurement process.
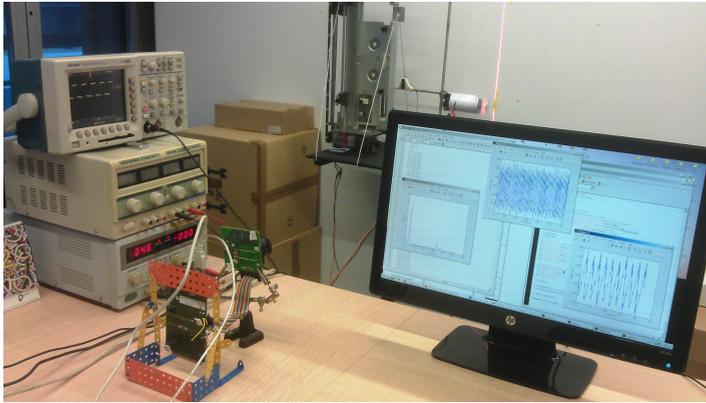
(a)
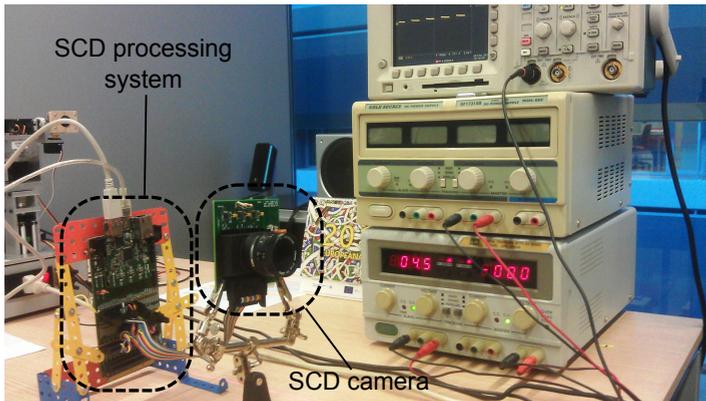
(b)
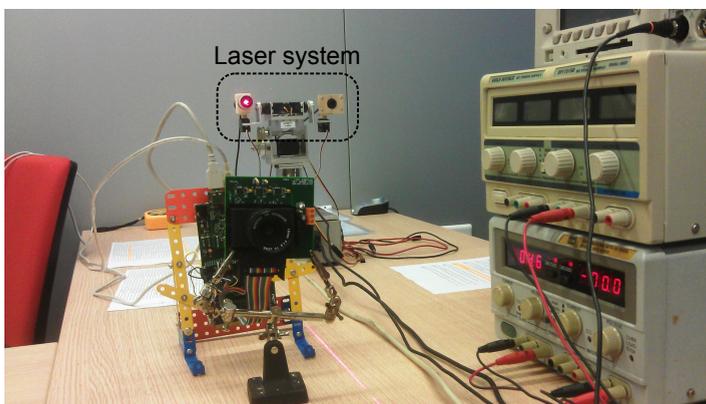
(c)

Figure 4.2. Experimental set-up.
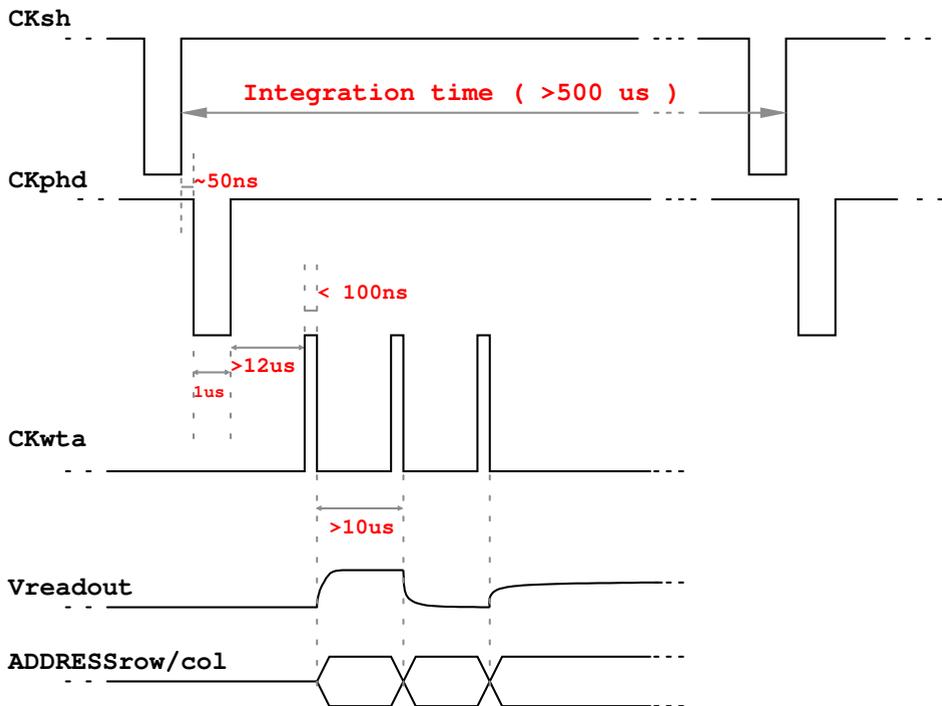
Figure 4.3. SCD sensor synchronization. In this example only 3 pixels per integration time are transmitted out of the sensor.

4.2 Sensor characterization

4.2.1 Grey level signal time response

Fig. 4.4 shows the waveform measured with an oscilloscope directly at node `Vreadout`. Time intervals between readouts have been set to 20 μs. It can be observed that the transition time between 90% and 10% of the step height of the grey level signal is approximately 5 μs. `CKwta` clock pulse width is less than 100 ns. In this figure it is clear that the `CKwta` period should last at least 10 μs to properly read the analogue grey level signal. This transitory is too long and is attributable to the design of the output stage of the analogue grey level value. This stage should be redesigned in future versions of the SCD sensor.
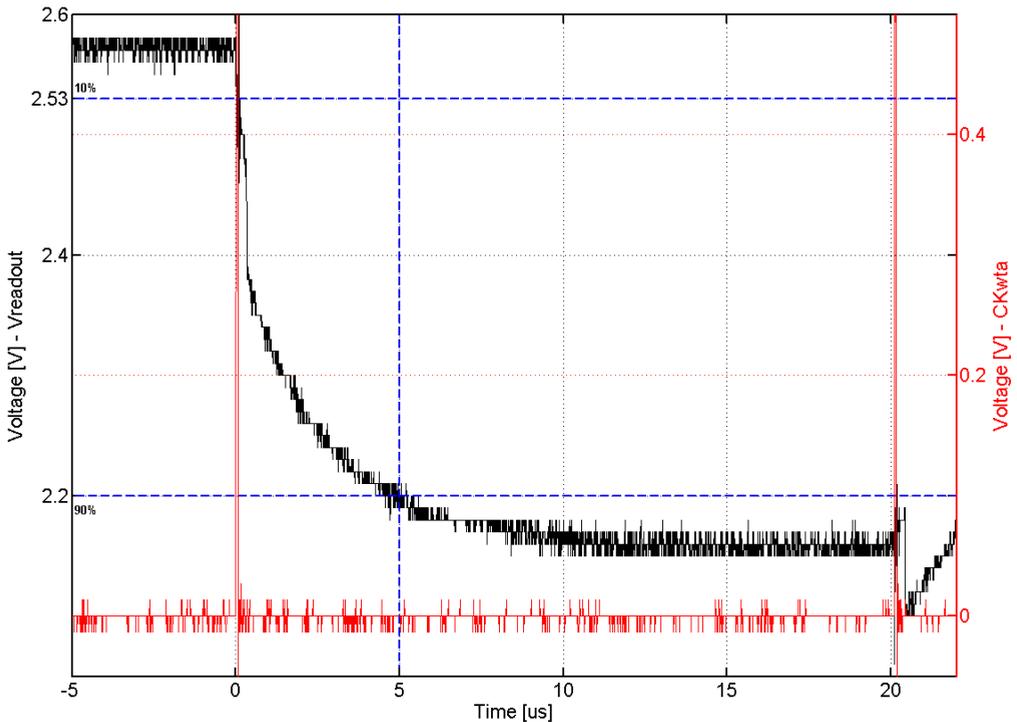
Figure 4.4. Gray level output signal (black line, node `Vreadout`) synchronized with `CKwta` clock (red line).

4.2.2 Dynamic range

The measurements presented in this subsection show the relationship between input light intensity and output voltage at `Vreadout` node. For this experiment, a light source was calibrated using a photometer. After the calibration process, the photometer was replaced with the sensor without a lens so that light could directly reach the pixel matrix. The output voltage at `Vreadout` node was measured with an oscilloscope for integration times of 20 ms, 10 ms, 5 ms, 2 ms, 1 ms and 0.5 ms. During the measurement process, the sensor and the photometer were exposed only to the calibrated light source and they were isolated from any other light sources. The numerical results of this experiment are shown in Table 4.2 and Fig. 4.5.

A set of linear models, each one corresponding to different integration times are plotted in Fig. 4.5 together with the raw measured data. The linear coefficients of these models have been obtained by means of a least-squares fitting procedure. The goodness of fit of these models can be calculated with the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(aI_i + b - V_i)^2}{\sum_{i=1}^{N}(V_i - \mu_V)^2} \tag{4.1}$$

where N is the size of the data set, I_i is the i-th illumination level, V_i is the measured output voltage when the sensor is exposed to the i-th illumination level, a and b are the linear coefficients obtained by linear regression, μ_V is the mean value of the measured output voltages and R^2 is the so called coefficient of determination that represents the goodness of fitting of the linear model to the measured data. The closer R^2 is to 1, the better the linear model fits the measured data. In Eq. 4.1, the denominator is the variance of the measured data with respect to its mean value (μ_V) and the numerator is the variance with respect to a linear model with coefficients a and b.

The R^2 values for the data set of Table 4.2 and the linear models plotted in Fig. 4.5 are shown in Table 4.3. These values clearly show that the linear models can be considered an accurate representation of the measured data. For integration times equal to 20 ms, 10 ms and 5 ms the linear models describe only the linear part of the response before saturation. This saturation voltage was experimentally determined to be between 2.01 and 2.04 V.

The dynamic range values specified in Table 4.2 were calculated according to the following equation:

$$\text{DR} = 20log_{10}(I_{max}/I_{min}), \tag{4.2}$$

where dynamic range (DR) is the dynamic range, I_{max} is the illumination value at which the linear models intersect with the saturation voltage and I_{min} is the minimum illumination value that the pixel is sensitive to.

Voltage at readout node[V]		Integration time [ms]					
		20	10	5	2	1	0.5
	0	2.8	2.8	2.8	2.84	2.84	2.84
	1.4	2.56	2.69	2.76	2.83	2.84	2.83
	2.4	2.48	2.65	2.74	2.82	2.83	2.83
	4.4	2.27	2.53	2.69	2.79	2.82	2.83
	7.3	2.09	2.38	2.61	2.76	2.8	2.82
Light intensity [lux]	11.3	2.04	2.21	2.49	2.73	2.78	2.81
	15.8	2.04	2.07	2.38	2.69	2.75	2.8
	22.0	2.04	2.01	2.27	2.61	2.73	2.79
	31.0	2.04	2.01	2.13	2.54	2.70	2.73
	40.5	2.04	2.01	2.03	2.45	2.65	2.71
	49.0	2.04	2.01	2.01	2.36	2.59	2.70
	58.4	2.04	2.01	2.01	2.29	2.55	2.68

Table 4.2. Readout voltages obtained for different integration times and light intensities.
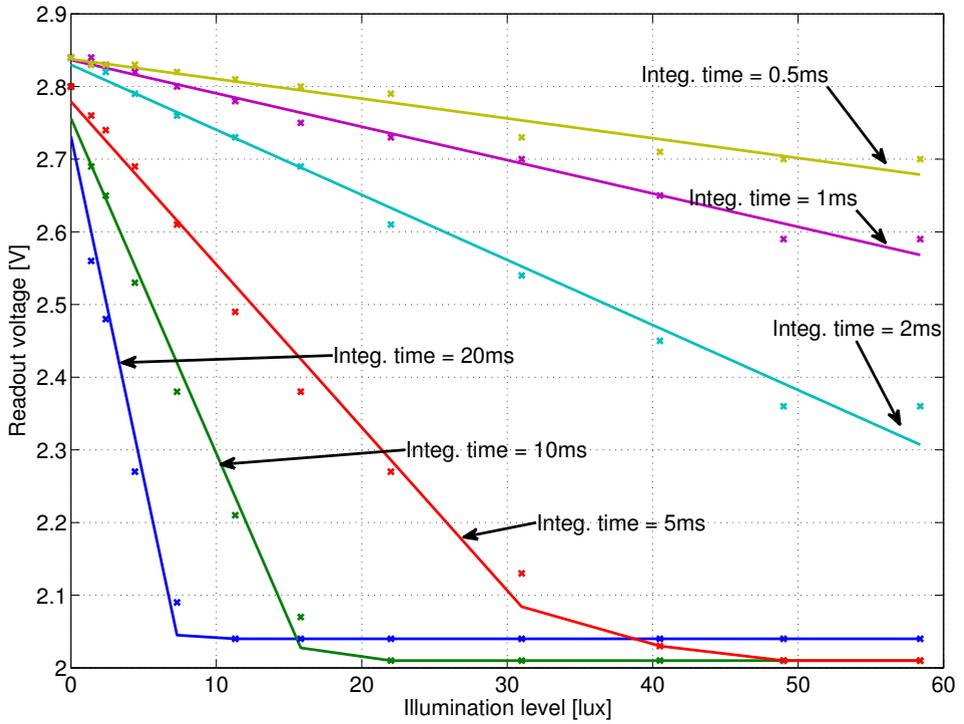
Figure 4.5. Sensor output voltage against light intensity for different integration times. Continuous curves are the linear regression fitting of data in Table 4.2. The crosses (x) correspond to the raw data of Table 4.2.

Integration time [ms]	R^2	a	b	DR
0.5	0.9547	-0.0027	2.8377	15 dB
1	0.9855	-0.0046	2.8365	21 dB
2	0.9856	-0.009	2.83	28 dB
5	0.9855	-0.0224	2.7792	36 dB
10	0.9849	-0.0461	2.7564	37 dB
20	0.9621	-0.0940	2.7313	37.5 dB

Table 4.3. Goodness of fit (R^2) and linear coefficients a and b (see Eq. 4.1) for the linear models plotted in Fig. 4.5. The DR was calculated taking into account the intersection of the linear model of coefficients a and b with the minimum output voltage (2.04 V).

4.2.3 WTA-matrix selection-logic experiments

The `permit` input signal (see Fig. 3.12 in Section 3.2.3.2 and Fig. 3.17) acts as a digital switch enabling and disabling the digital selection logic of the WTA subcircuit. If `permit` input is in high state, the digital logic is disabled, meaning that the analogue WTA is the only discriminant circuit. Experiments show that, in this case, all the address lines are constantly set to logic "1". This is because many pixels are winning the competition in the same time interval. The resulting output address is the wired-or of the addresses of all the pixels that just won the competition. Fig. 4.6(a) depicts the visual information output by the SCD system when `permit` = "0": only pixel (31,31) shows changes in its grey level value. As soon as `permit` is set to logic low, all the pixels in the image rapidly change to different grey level values, as in Fig. 4.6(b), indicating that the pixel flow at the output of the SCD system is now coherent with the visual scene captured by the sensor.

These results indicate that the digital logic circuit arbitrating the WTA network proposed in this research work, and published in [77], is essential to the proper operation of the implemented SCD sensor.
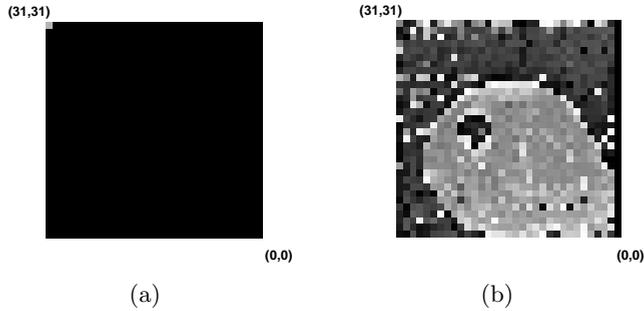
Figure 4.6. Information transmitted out of the chip with (a) `permit` = "0" and (b) `permit` = "1".

4.2.4 WTA-stage errors

4.2.4.1 WTA-stage not being able to choose a winner

According to the measurements detailed in Section 4.2.2 the output voltage at `Vreadout` node cannot be larger than 2.85 V. However, during the experiments we could observe that sometimes the analogue grey level signal reached values higher than this threshold, even larger than 3.0 V. The only possible reason for this behaviour is that the transistor `MPROUTSW` is turned-off for all the pixels in the matrix. This would mean that the WTA stage was not able to choose a winner. An oscilloscope trace of this phenomena can be seen Fig. 4.7.

In order to gain some insight into this non-desirable phenomena the following experiment was carried-out: the pixel flow generated by the SCD sensor for two different scenes was recorded; while the first scene was dynamic with some objects permanently moving at a certain distance in front of the sensor, the other scene was completely static. The integration time was set to 5 ms in concordance with the ambient light illumination, and 100 pixels per frame-period were delivered out of the sensor. The read-out was monitored with an oscilloscope observing that according to the illumination level, in both scenes, the output voltage was around 2.65 V, occasionally reaching 2.8 V when transmitting pixels that were capturing the darkest parts of the scene.

A total of 199 buffers of 100 pixels each were processed for each type of

scene. This information was processed off-line. A threshold of $V_{th} = 2.9\ V$ was set to distinguish the valid grey level values (below the threshold) from the non-valid ones (above the threshold). Fig. 4.8(a) shows the histogram of occurrence of failed pixels as a function of the position in the readout sequence for the dynamic scene. In this case the error rarely appears: only a total error rate of 3/199,000 failures. Fig. 4.8(b) shows the results for the static scene. In this latter case, the error rate increases to 116/199,000 (approximately 0.06% of the total pixels). The results clearly show that the WTA decision block exhibits much better performance under dynamic conditions, with a subset of pixels undergoing changes substantially larger than the rest of the pixels in the matrix. This being said, in both cases, the error rate confirms that this type of unwanted behaviour is not really significant under normal operation of the circuit, at least for a 32x32 matrix.
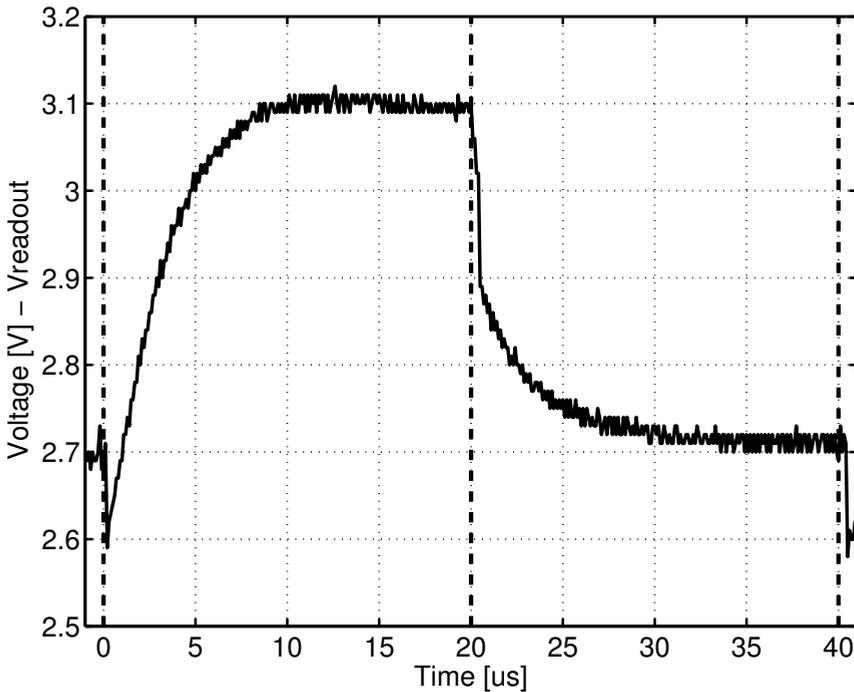
Figure 4.7. Oscilloscope trace of the readout signal reaching values higher than 3.0 V. `CKwta` period is 20 μs. Vertical dashed lines show winner selection timing.
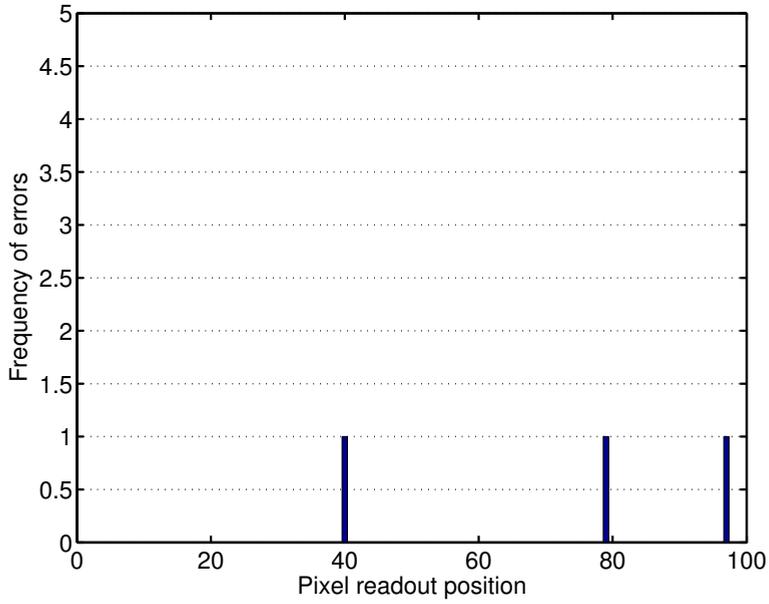
(a) Dynamic scene.
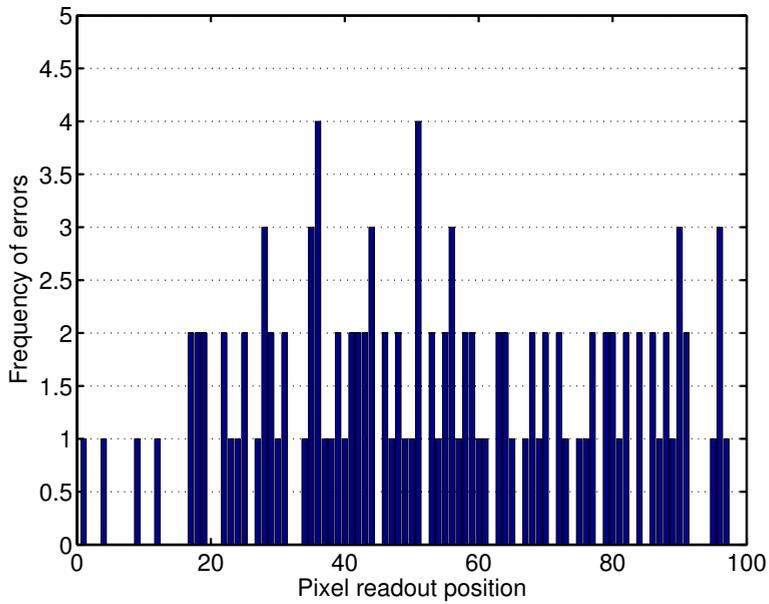
(b) Static scene.

Figure 4.8. Histogram of WTA failures (WTA stage not being able to choose a winner) as a function of the position in the readout sequence for two different types of scenes: dynamic and static.

4.2.4.2   WTA sequentially selecting the same winner

During the experiments carries out for thesis it was observed that, sometimes, the same pixel could be selected as the winner several times during the same frame period. Although not desirable, this behaviour is inherent in the pixel-circuit in its present form. Once a pixel is selected as the winner, the feedback mechanism of the digital logic circuit equals the voltages at `Cpresent` and `Cprevious` capacitors, but the pixel can still compete in the forthcoming pixel-periods. If the other pixels in the matrix show very small changes, such as in a static scene, or when the largest changes in a dynamic scene have already been transmitted, the pixels that have already won the competition are best suited to win it again. Strictly speaking, this is not an analogue WTA subcircuit error, but something that should be taken into account in the digital logic by adding a mechanism to avoid pixels being able to win twice in the same frame period.

4.2.5   Noise

This section deals with the characterization of the two most common types of noise: FPN and random temporal noise. FPN is the spatial variation of pixel output values under uniform illumination conditions. It is produced by mismatching in the parameters of the circuit components, and is supposed to remain constant over time. FPN has offset and gain components. On the other hand, random temporal noise is the fluctuation in the pixel output over time for a stationary stimulus.

To measure the FPN, the sensor must be exposed to a uniform surface so that all the pixels in the matrix are stimulated with the same illumination level. Under these static conditions, the grey level value from all the pixels must be read out. These would give the information necessary for analyzing the spatial variation in the output for a uniform spatial stimulus. Our SCD sensor does not provide a mechanism for capturing the whole pixel-matrix at a certain moment of time; it only delivers a pixel flow based on illumination changes. Moreover, as the path for the selection logic is fixed, the selection of the winning pixels under a static scene is biased: the first pixels in the path will be selected most of the time, so it is almost impossible to collect the necessary information to perform a complete FPN analysis. The attempts made to measure the FPN of the sensor showed that a small subset of pixels were constantly and repeatedly read out. This

behaviour is consistent with the problem explained in Subsection 4.2.4.2. As the results presented in this section are based on this small subset of pixels, they do not comprise a complete FPN analysis. Nevertheless, these measurements give some insight into this characteristic of the sensor.

The random temporal noise can be measured by observing the output of a particular pixel over time. This can be measured easily by properly setting the CKphd, CKsh and CKwta clock signals.

4.2.5.1 Fixed-pattern noise measurements

With the intention of measuring the FPN, the pixel matrix was excited with a uniform illumination level. The clock signals were generated in such a way that switches MPFLRSW and MPSH remained permanently closed (CKsh = "0") and CKphd was used to set the integration period. One single CKwta pulse per integration period was generated, meaning that only one pixel was readout. As the MPFLRSW and MPSH switches were closed, the ramp of the integration signal could be clearly observed at Vreadout node for the pixel selected by the WTA decision block. A timing diagram of the signals involved in the measurement process can be seen in Fig. 4.9. For every pulse of the CKwta clock, a different pixel is selected to be read out, so the offset of the reset voltage of the photodiode can be clearly observed in the oscilloscope. In Fig. 4.9 the offset due to the FPN in the reset voltage between two consecutive pixels is labelled as Voffset. The gain FPN is an indicator of the difference between the slopes of the integration signals under constant illumination conditions.

In Fig. 4.10 an oscilloscope capture of the waveform at the Vreadout node for an integration time of 500 μs is shown. Both the offset and gain FPN are clearly noticeable.

The behaviour of the WTA circuit described in Subsection 4.2.4.1 can be clearly observed at 10 and 12.5 ms. For these integration periods, the WTA decision block is not able to select a pixel to be read out, so the output voltage remains close to Vdd during the whole sampling interval.

The offset FPN, FPN$_{\text{offset}}$, is calculated as:

$$\text{FPN}_{\text{offset}} = 100 \cdot \frac{\sigma_{\text{offset}}}{\Delta_{\text{Vout}}}, \tag{4.3}$$

where σ_{offset} is the standard deviation of the measured reset voltages and Δ_{Vout} is the maximum voltage variation that can be observed at Vreadout

node. According to Table 4.2, $\Delta_{\text{Vout}} \approx 0.8\ V$.

The gain FPN, FPN_{gain}, is calculated with the following equation:

$$\text{FPN}_{\text{gain}} = 100 \cdot \frac{\sigma_{\text{slope}}}{\mu_{\text{slope}}}, \tag{4.4}$$

where σ_{slope} is the standard deviation of the slopes of Vreadout voltages and μ_{slope} their mean value.

Table 4.4 shows the results of the offset and gain FPN measurements performed with data collected from a subset of 70 different pixels and two different integration times.

Integ. time [ms]	$\text{FPN}_{\text{offset}}$	FPN_{gain}
500 μs	3.45%	19%
5 ms	3.39%	7.29%

Table 4.4. FPN measurements calculated with a sample size of 70 pixels.
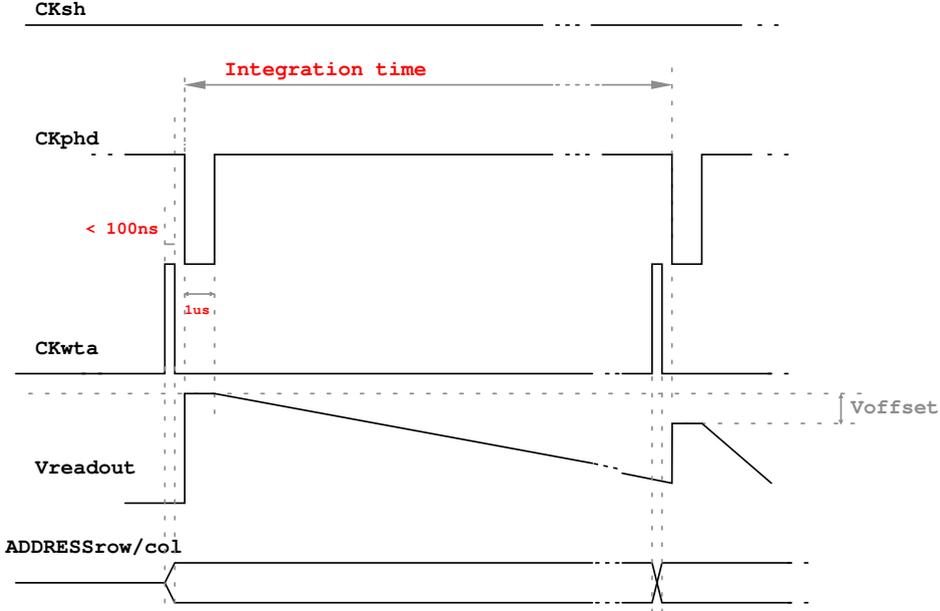
Figure 4.9. Time diagram for noise measurements. CKsh = 0 V.
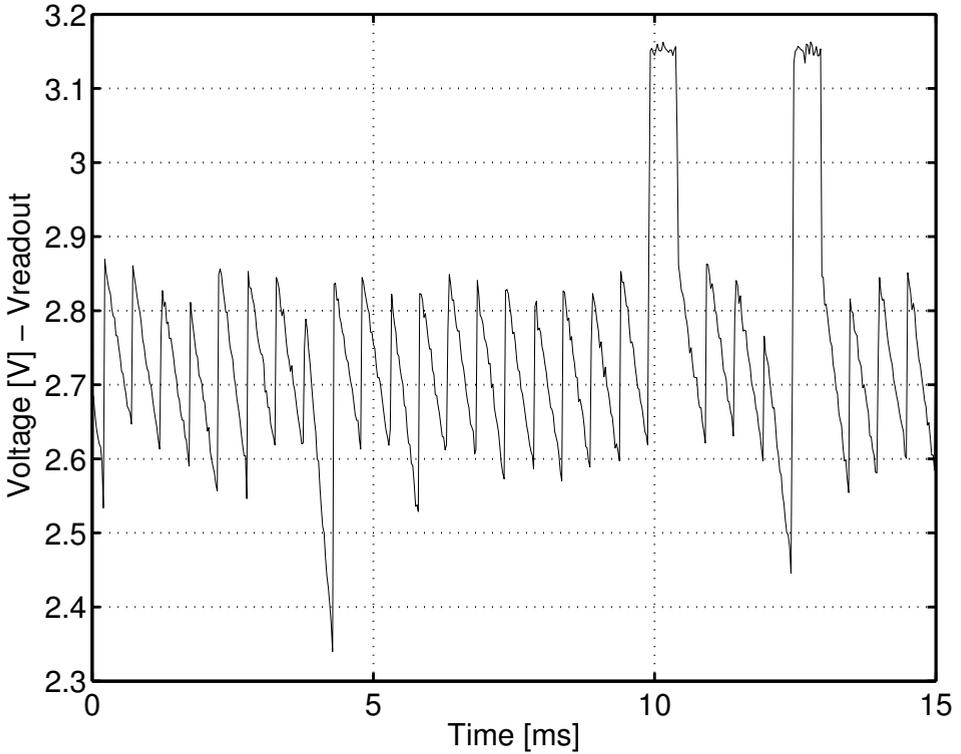
Figure 4.10. Oscilloscope capture of the `Vreadout` noise during FPN measurements. Integration time is 500 μs.

The measured FPN due to offset variations in the reset voltage level is high, though it is still within acceptable levels, but the dispersion in the gain of the photocircuit is high. Suitable sources of this type of noise are the mismatching in the parasitic capacitance of the photodiode, and in the gain of the unity gain and source follower buffers. In the chip fabricated in the context of this thesis, it is impossible to identify the precise source of gain FPN, due to the difficulties of reading the signal at intermediate stages of the photocircuits. A test vehicle, where the different internal nodes of the photocircuit are connected to the output pads, should be fabricated to perform the necessary measurements.

4.2.5.2   Random temporal noise measurements

To perform this measurement the `CKwta` signal is set permanently to 0 V. In this way the address of the pixel being read out does not vary over time. If `CKphd` and `CKsh` are operated in the same way as in the FPN measurement, the evolution of the integration signal of one particular pixel can be traced over time. The only limitation is that it is not possible to choose the particular pixel to be analyzed. Instead, this is determined by the last winner selected by the WTA stage with the last `CKwta` pulse just before starting the measurement. In Fig. 4.11 a capture from the oscilloscope shows the evolution of the voltage signal at the `Vreadout` node. From this figure it is clear that the temporal noise is quite low compared to the FPN. The noise in the slope of the integration voltage is in this case 2.22%.
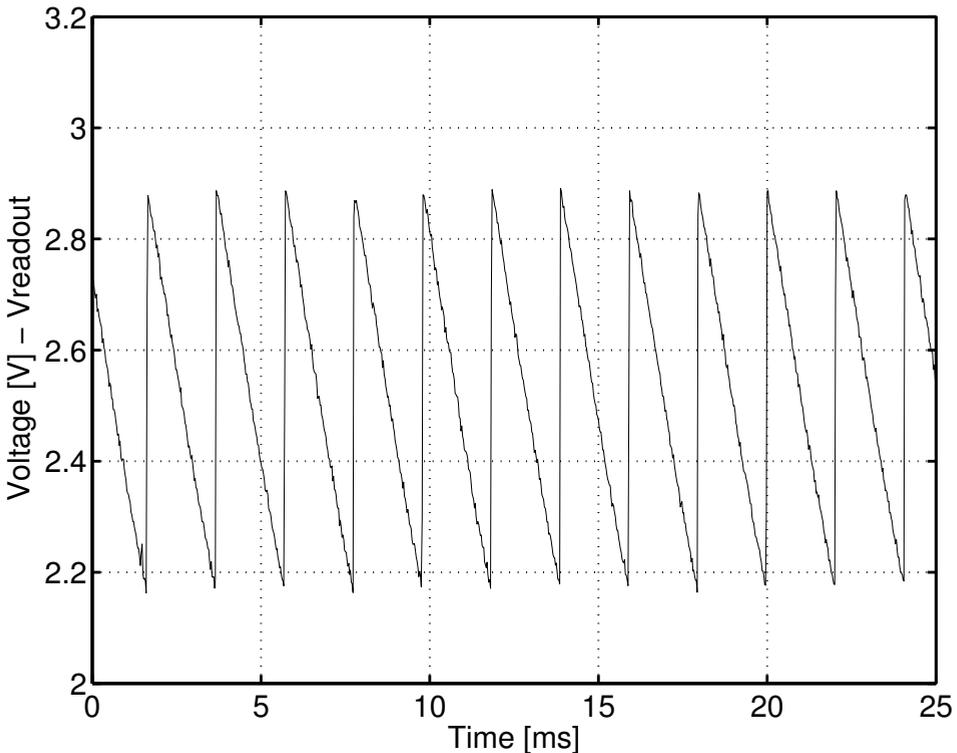
Figure 4.11. Oscilloscope capture of the `Vreadout` noise during temporal noise measurements.

4.2.6   Images

Although obtaining the complete information of the pixel matrix under static conditions is very difficult, it is possible to obtain an almost updated version of the pixel matrix by slowly moving some objects in front of the sensor. This forces the WTA to select the borders of the moving objects to be read out. In this way, by collecting the pixel flow, something similar to an image can be shown. Fig. 4.12 shows several of these pseudo-snapshots taken with the SCD sensor. These figures give an idea of how noisy the information delivered by the sensor is. These images were captured with an integration time of 5 ms. For smaller integration times, the noise increases. With 500 μs the only possibility is to binarize the scene. Under this last condition, tracking algorithms can be successfully applied if the moving object shows enough contrast with respect to the background [69, 68].

(a) Small key.   (b) Letter T.

(c) Kid.   (d) Sensor itself.
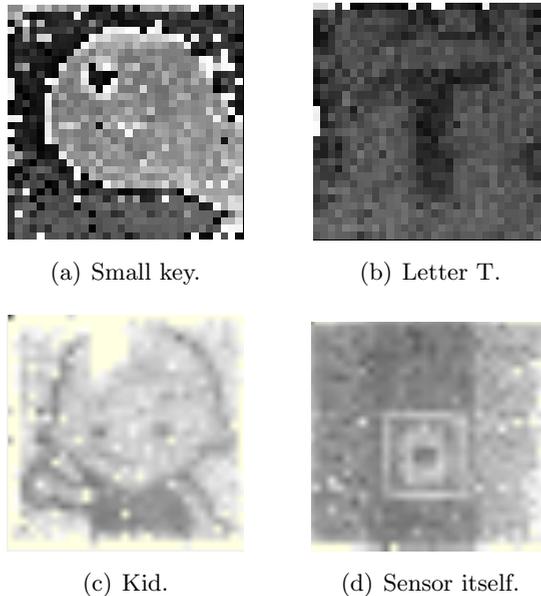
Figure 4.12. Images reconstructed with the pixel-flow delivered by the SCD sensor.

4.2.7   Power consumption

In order to measure the power consumption of the sensor a 1.3 Ω resistor was connected in series between the `Vdd` node and the sensor. The voltage drop across the resistor was measured for several integration times and transmitted pixels per frame. Table 4.5 shows the results of these measurements.

Power consumption remains almost constant for all the measured configurations, meaning that consumption in the sensor is mainly static and due to the constant current sources. Only a very small portion of the power consumed by the chip is due to dynamic activity. The total current of 32 mA is high compared to other state-of-the-art vision sensors [22, 4]. This characteristic of the sensor should be improved on in future VLSI implementations of the SCD strategy.

Integ. time [ms]	Num. of transmitted pixels per integ. time	Total current [mA]	Average current/pixel [μA]
	50	32	31.25
20 ms	100	32	31.25
	200	32	31.25
	50	32	31.25
10 ms	100	32	31.25
	200	32	31.25
	50	32	31.25
5 ms	100	32	31.25
	200	32	31.25
2 ms	50	32.15	31.39
	90	32.15	31.39
1 ms	40	32.15	31.39
0.5 ms	20	32.15	31.39

Table 4.5. Power consumption for several integration times and number of transmitted pixels. Vdd = 3.3 V.

4.2.8 Summary of characteristics

Table 4.6 summarizes the characteristics of the sensor.

Parameter	Value
Fabrication process	austriamicrosystems 0.35 μm 2-poly 4-metal
Array size	32x32 pixels
Pixel dimensions	55x55 μm^2
Chip size	2.8x2.8 mm^2
Fill factor	4.8%
Supply voltage	3.3 V
Power consumption total / average per pixel	Approx. 32 mA / 31 μA (@3.3 V)
DR	15 dB (Integ. time = 500 μs)
	37.5 dB (Integ. time = 20 ms)
Frame rate	Tested at 2k, 1k, 500, 200, 100 and 50 frames/sec.
FPN	3.45% (offset) and 19% (gain), integ. time = 500 μs.
	3.39% (offset) and 7.29% (gain), integ. time = 5 ms.
Pixel timing	> 10 μs

Table 4.6. Summary of characteristics. FPN was estimated with a sample size of 70 pixels (see Subsection 4.2.5).

4.2.9 Discussion and proposals

The characterization of the sensor revealed several drawbacks that should be taken into account in future iterations of the design process. Noise is a severe problem in this first SCD sensor. Likely sources of FPN can be found in the mismatching at the parasitic capacitor of the photodiode, or in the subsequent amplifying stages of the photocircuit or readout buffer. The FPN in the parasitic capacitor of the photodiode can be solved by using a capacitive transimpedance amplifier (CTIA) and PIP capacitor for the integration of the photocurrent [5, 41]. CDS (correlated double sampling) techniques could reduce other sources of mismatching in the photocircuit [78]. The quality of the images is very poor at sampling frequencies of 5 ms, and for higher rates, tracking algorithms can only be applied when binarizing the pixel-flow transmitted by the sensor. The first step in the design process should have been the fabrication of a test vehicle including separate versions of all the building blocks that comprise the pixel with their inputs and outputs connected to the pads of the chip. In this way, all the blocks could have been studied and characterized separately, identifying

likely sources of mismatching, noise and other unwanted effects. As this step was omitted, the sources of noise can not, at the end of this work, be clearly identified.

The dynamic range is low and should be increased in a future design. If a linear photocircuit is to be used, the dynamic range could be increased by enlarging the area of the photodiode, or by using other linear integration techniques. Dual-photodiode architectures [91] could be incorporated for generating long and short exposure images. The information coming from these two photodiodes can be properly processed to extract the information from the darkest and brightest parts of the scene. Photocircuits that integrate up to a fixed voltage value instead of integrating for a fixed time interval could also be a solution [52, 5]. In the latter case, the integration time can not be handled to operate in fast capture mode. The time between frames is determined by the illumination level. The darker the scene, the lower the frame-rate.

Measurements show that power consumption should be reduced. According to the simulations shown in Chapter 3, the main sources of power consumption are the OTA and the rectifier. To cover the necessary input range, the OTA was implemented with two asymmetrically-sized differential pair transistors. In a future implementation, the rectifier could be removed by simply using two separate OTAs with crossed inputs. Positive and negative differences would produce two separate positive currents that can be added to feed the WTA decision block. Another advantage of this design would be that some circuitry could be added to block the response of the sensor to positive or negative changes. This feature could be very useful in many applications of the sensor such as the tracking of very fast moving objects.

The behaviour of the WTA decision block was also studied. It was observed that sometimes the WTA is not capable of selecting a winner amongst the pixels in the matrix. Although not desirable, this phenomenon rarely happens. The statistics show that in the vast majority of the cases the designed analogue-digital WTA stage selects a winner amongst the set of pixels that underwent the largest changes in illumination according to SCD principles.

One more problem related to the WTA stage is that, sometimes, the same pixel is read out several times during the same frame period. This is because once a pixel is transmitted out of the chip, it can still participate

in the selection process in the next forthcoming pixel-periods. If the scene is static, or no more relevant changes are left to process in a dynamic scene, a pixel that was already transmitted can be transmitted again. The most straightforward solution would be to add some additional circuitry to the digital part of the pixel, so that once a pixel is transmitted the output of the starved inverter is forced to be in low a state until the next reset of the photodiode arrives.

4.3 Tracking experiment

A simple tracking experiment has been designed to experimentally explore the abilities of our SCD system. The set-up consists of a small laser mounted in a motor axis balancing at a constant frequency. The laser has a special lens that generates a vertical bar projected onto a wall. The bar swings quickly from one side of the wall to the other. The system should be able to accurately calculate the bar's centre of mass in the image.

4.3.1 Object detection and centre of mass calculation algorithm

The experiment has been designed to detect the object (laser bar projection) binarizing the image using a threshold of the pixel grey level. Once the bar's pixels have been selected by the image threshold, it is easy to calculate the coordinates (x, y) of the centre of mass using the following equations (classical approach):

$$x = \frac{1}{n} \sum_{i=0}^{n-1} x_i, \tag{4.5}$$

$$y = \frac{1}{n} \sum_{i=0}^{n-1} y_i, \tag{4.6}$$

where n is the number of selected pixels and (x_i, y_i) are the coordinates of the i-th pixel. In this equation we have supposed that all pixels have the same mass with a normalized value of 1: $m_i = 1$, with $0 \leq i < n$.

In normal image or video processing, these two operations (binarization and centre of mass calculation) would have to be done for every pixel in every incoming image. In SCD Vision, these operations are performed for

every incoming pixel, not image, because SCD processing does not work
with images but just with pixels. So, for every incoming pixel a threshold
must be applied to detect if it belongs to the object (projected bar) being
detected, if so, the contribution of that pixel to the centre of mass must be
calculated.

As there is no snapshot of the scene, it is necessary to keep track of
those pixels that belong to the object. This can be implemented using a
Belonging List, so any incoming pixel with grey level above the threshold is
added to the list, if it was not in the list already. Also, if a pixel is below the
grey level threshold, it must be removed from the list. The bar's centre of
mass is calculated using the pixels in the list. For practical reasons, there
is not a list of pixels but a binary temporal image constructed with the
pixels received and updated at every new incoming pixel. To calculate the
centre of mass using all the pixels in this temporal image would have been
a waste of resources. Instead, this temporal image is only used to see if the
just-received pixel changed its value with respect to its last stored value in
the temporal image. If it changed, the centre of mass must be updated and
there are two possibilities: first the pixel does not belong to the object (it
has a grey value below the threshold), or it now belongs to the object. In
both cases, the centre of mass must be updated with just the information
of this change. If the pixel is above the threshold, the number of pixels
belonging to the object n, and the centre of mass coordinates (x, y) are
updated as follows:

$$n = n + 1, \tag{4.7}$$

$$x = \frac{(n-1)x + x_i}{n}, \tag{4.8}$$

$$y = \frac{(n-1)y + y_i}{n}, \tag{4.9}$$

where (x_i, y_i) are the coordinates of the incoming pixel. In the same man-
ner, if the incoming pixel changed and it is now below the threshold, the
following operations must be issued instead:

$$n = n - 1, \tag{4.10}$$

$$x = \frac{(n+1)x - x_i}{n}, \tag{4.11}$$

$$y = \frac{(n+1)y - y_i}{n}. \tag{4.12}$$

The operations per processed pixel are roughly similar compared to the traditional image sequence approach, but SCD processing has the advantage of obtaining a new centre of mass at every new incoming pixel, at almost no extra cost. The other advantage is that not all image pixels are processed, only those that are the most interesting.

For practical reasons, and taking into account the limited capacity of the processing device, it is more interesting to introduce a couple of new variables (sum_x, sum_y) that store the sum of the pixel coordinates. It is also possible to introduce a new variable s (sign) that indicates if the processed pixel is introduced in the centre of mass calculation $(s = +1)$ or removed $(s = -1)$. With these assumptions, the new equations for the centre of mass calculation finally implemented are:

$$sum_x = sum_x + s \times x_i \tag{4.13}$$
$$sum_y = sum_y + s \times y_i \tag{4.14}$$
$$n = n + s \tag{4.15}$$
$$x = \frac{sum_x}{n} \tag{4.16}$$
$$y = \frac{sum_y}{n} \tag{4.17}$$

These changes, and the new formulation, allow the implementation of these operations using integer arithmetic, in fixed point notation, instead of floating point arithmetic.

It is interesting to note that pixel flow processing opens the door to other tracking approaches that speed-up object detection even further. For example, it is possible to only take into account those pixels for which the illumination level changes from below to above the threshold. The centre of mass is then calculated with a fixed number of last received pixels, or even just one. The result in this case would not be very different from that calculated by the method explained earlier, but it could be run much faster, especially when thinking about very limited hardware. This new approach would make no sense in a traditional image sequence processing. It is just an example to show that the SCD Vision approach may introduce new ways of solving problems not addressed before.
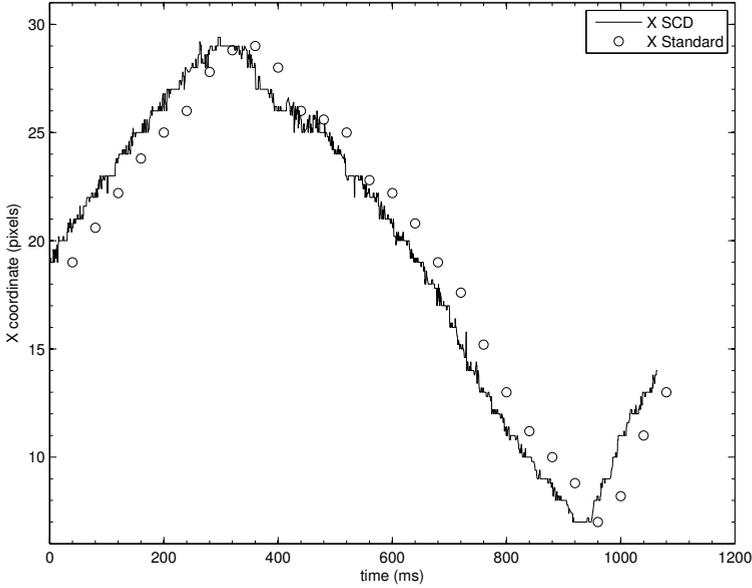
Figure 4.13. x coordinate of the object oscillating at 0.8 Hz

4.3.2   Results

To implement the experiment explained in the previous section, the laser source is mounted in a moving axis directly driven by a motor that swings at desired frequencies. The SCD system stores and calculates the laser projected bar's centre of mass for roughly 1 second, and then it sends the results to the host computer for further processing.

Despite the simplicity of the centre of mass calculation, the 80 MHz microcontroller takes 10 μs to implement the additions and integer divisions it requires. The SCD sensor may deliver pixels every 10 μs, so both times are roughly the same and the performance is not constrained by the microcontroller. The bar being detected in our experiment measures several tenths of pixels. For this experiment, we have found that it was necessary to process at least 10 pixels per frame (10 pixels every 500 μs). This rate of 20,000 pixels per second is still much lower than the maximum SCD delivery rate or processing capabilities (100,000 pixels per second).

Fig. 4.13 shows the x coordinate, as a function of time, of the trajectory of the tracked bar in the focal plane for roughly one second. The swing

period of the movement was 1.25 s (0.8 Hz). The y coordinate is not shown since it remains constant for all the movement. The trajectory shown in this figure matches a linear movement from one side to the other. It is interesting to note that the position is calculated at subpixel level, since many pixels contribute to the calculation. This same figure shows the points at which a standard system (@25 fps 32×32) would have calculated the centre of mass (these points are marked with a small circumference in the figure). It is interesting to see that the standard camera has calculated just 27 positions while the SCD system yielded more than 2,000.

In this first experiment, the object is not moving very fast so we could consider the results from the standard camera acceptable, since the trajectory can be traced from the calculated points. The standard camera runs at 25 fps which implies a time of 40 ms between two consecutive images. Thus, even considering that there is no latency in the image processing, there is a delay of 40 ms between the snapshot and the object position delivery. This delay, and the difference in the position estimation, can be seen in the experiment figures, where the small circumferences represent the position calculation (only the x component) delivered by the standard system. SCD Vision does not have this delay problem, since every new position is calculated right after the capture integration time that, for this experiment, is just 500 μs. This is probably one of the main advantages of SCD vision when compared to other camera systems.

Fig. 4.14 shows the same coordinates but the object is now moving at 4 Hz. Again the object position is accurately calculated. It is possible to see some glitches that are below one pixel in most of the cases. As the movement speed increases, it is more and more difficult to calculate the real trajectory from the positions calculated by the standard system. At 4 Hz it is still possible to see a kind of oscillation in the horizontal movement, with a linear connection between sides, but nothing else.

Finally, Fig. 4.15 shows the object position when it moves at 10 Hz. The presence of glitches is more noticeable in this last experiment, especially when the motor stops changing its rotating sense. The standard system in this case delivers an almost random set of positions that makes it almost impossible to guess the real trajectory of the object. If the object is moving fast, as in the experiment, the 40 ms latency of the standard system can be considered very large since the position of the object has changed several pixels during this time.
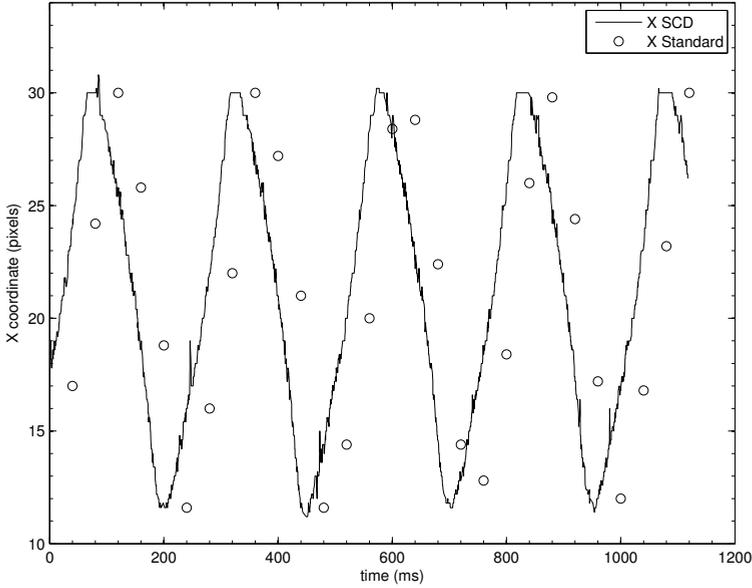
Figure 4.14. x coordinate of the object oscillating at 4 Hz

4.3.3 Discussion

The solution to this problem with a standard system requires a standard camera and a processing system. Supposing the standard camera has 32×32 pixels and runs at 25 fps, the required bandwidth is $32 \times 32 \times 25 = 25,600$ pixels per second. This is a very low bandwidth compared to a standard video rate, so any PC system could handle it without any problems. Nevertheless, even with such a low resolution, a system that does not take more than 39 μs to process each pixel is necessary. This is roughly the time required by a microcontroller to perform simple centre of mass calculations, but there is little time for doing any other tasks and calculations.

With the SCD Vision system, it was possible to accurately calculate the centre of mass of an object, with only a few pixels for each 500 μs, at a rate of just 20,000 pixels per second, which is similar to the pixel rate required by a standard vision system. The big difference is in the time resolution and accuracy: whereas the SCD system was giving a new accurate object position every 500 μs, the standard system can only deliver a new result every 40 ms (80x less time resolution than SCD Vision). Along with the
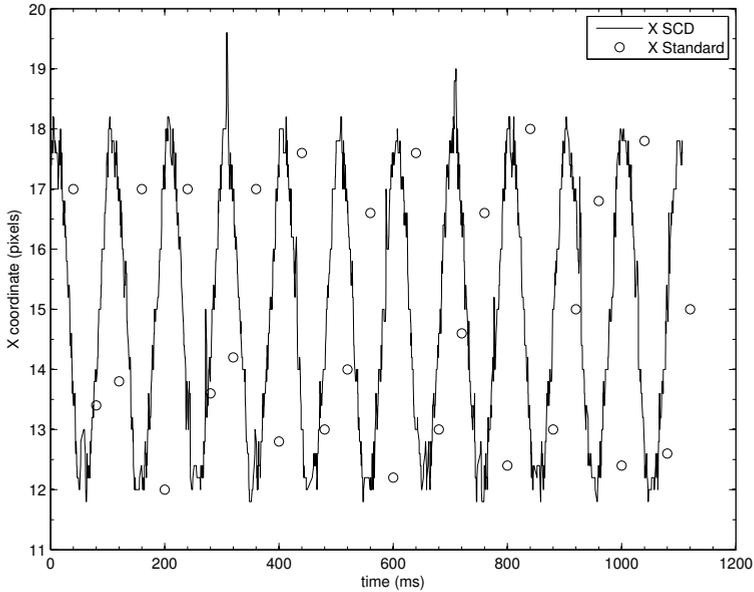
Figure 4.15. x coordinate of the object oscillating at 10 Hz

resolution, the other important parameter is latency: SCD Vision has a latency of just 500 μs, whereas a standard system at 25 fps has 80 times more (40 ms). This is an almost two orders of magnitude gain using the same resources.

SCD Vision improves pixel rate requirements, time resolution and latency for this specific experiment compared to a standard camera. It is possible to use a high speed camera, yet with a standard vision system based on image processing, to try to obtain similar results in terms of latency and time resolution this camera would need to run at 2,000 fps. The pixel rate would be $32 \times 32 \times 2,000 = 2,048,000$ pixels per second. There are only 488 ns to process each pixel. This processing time is still possible with standard processors running at GHz consuming tenths of Watts, but the results would have been the same as those obtained with a simple, very small, SCD system consuming less than 1 W, at a pixel rate 100 times lower (two orders of magnitude). Table 4.7 summarizes the comparison of these three systems. This table shows that SCD Vision offers similar results to traditional systems with several orders of magnitude fewer resource requirements. It also shows that with similar resources, SCD Vision offers

two orders of magnitude more accurate results in terms of time resolution and latency.

	SCD	Std. at 25 fps	Std. at 2000 fps
Pixel rate (kb/s)	20	26.6	2048
Time resolution (ms)	0.5	40	0.5
Processing time (μs/pixel)	2.3	39	0.488

Table 4.7. Comparison of three vision systems for simple object tracking.

4.4   Conclusions and improvement proposals

In this chapter the results of the experiments with the designed SCD sensor were presented. The sensor was embedded in an autonomous system together with a 32-bit microcontroller using a USB cable as the only necessary external connection. The USB was used as the communication link for transmitting/receiving information as well as the only power source for the system. This system was used to characterize the basic features of the sensor, and to carry out a tracking experiment that clearly showed the benefits of SCD Vision when dealing with fast moving objects under the constraints of limited bandwidth and processing power.

In the tracking experiment, the position of a target moving in front of the SCD camera has been calculated in real-time at very high speed (2000 fps) by reading-out only a very small number of pixels. The results of the experiment clearly show that a vision system based on an SCD camera outperforms a vision system based on a standard camera (25 fps; 32×32 resolution) in time resolution and processing time by two and one orders of magnitude respectively, while keeping similar pixel rates. The SCD system was also compared to a vision system based on a traditional-readout frame-based high-speed camera (2000 fps; 32×32 resolution). In this case, both systems exhibit similar performance in terms of time resolution and processing time, but the SCD system requires two orders of magnitude less bandwidth than the traditional high-speed camera system.

The characterization of the sensor revealed several drawbacks and problems in this first version of the sensor. Noise, power consumption and dy-

namic range are aspects to be improved in future iterations of the design process. The WTA decision stage can also be modified to solve some of the problems detailed in Subsection 4.2.4.

As a general conclusion to this chapter it can be said that although much work needs to be done to reach a final design, the sensor and SCD system implemented in this thesis can still be successfully applied in real situations where the benefits of SCD Vision over other artificial vision systems based on a standard-readout frame-based sensor is clearly demonstrated. In Subsection 4.2.9 ideas and proposals for improving the performance of the sensor in many of its aspects were given.

Conclusions and future work

<div style="text-align: right; font-size: 2em;">5</div>

Conclusions and future work 5

Al final, todo tiende a salir bien.

M. Cobos

Traditional frame-based imagers present some dynamic limitations when applied to artificial vision systems. One of them is that they constantly transmit information even if no changes have been produced in the scene under analysis, therefore wasting valuable resources such as bandwidth and processing power. Depending on the particular problem and hardware configuration, the processing hardware may even not be able to cope with all the generated data. Many of these problems could be overcome with the design of new sensing and readout strategies focused on relevant changing information. Over the last decade many relevant improvements have been achieved in this direction (see Section 2.1). Taking the biological vision system as general guide and inspiration, an increasing number of VLSI vision sensors are being designed where the sparcity, asynchrony and event-driven generation of the information coming from the visual field is taken into account.

It is within this framework that SCD Vision emerges as an innovative and original proposal (see Section 2.2). SCD Vision relies on the idea that a pixel showing a large change in intensity is an indicator of fast movements

and object edges around it. An SCD sensor is frame-based in the sense that successive frames are captured at a very high rate, but pixel readout is performed in an entirely different manner. The pixels are read out in order of relevance. The larger the change in illumination, the more relevant the pixel is considered. Not all the pixels in the sensing matrix need to be transmitted. As the pixels showing relevant changing information are transmitted first, only a small set of pixels may be read out, these being the ones conveying the most important information of the scene under analysis. In simulations published prior to this thesis it has been clearly shown that processing power needs and bandwidth occupancy would be substantially reduced in an SCD-based artificial vision system.

As stated in the first chapter of this document, the main objectives of this work were: to design and fabricate the first vision chip following SCD readout strategy, to build an SCD vision system based on this chip, and to show, by proper experimentation with this newly developed system, the benefits of SCD Vision in practical applications. All these objectives have been achieved. The first SCD chip has been designed, fabricated and embedded into a small, but powerful artificial vision system. This system has been used to implement tracking algorithms as well as to characterize the main basic features of this sensor.

The design of a pixel-matrix following SCD principles, implies the treatment of the problem of a global comparison of analogue quantities amongst all the pixels in the matrix. This problem is not easy to solve. An attempt to solve it has been addressed by means of a WTA circuit. A large WTA network together with a proposal for single winner selection has been designed, implemented and its behaviour characterized. Although during the experimentation several problems were discovered (see Subsection 4.2.4), the implemented WTA circuit behaves as expected when the scene under analysis has relevant changing information to transmit. The tracking experiment carried out during this thesis show that the WTA selects the pixels that underwent the largest changes as the first pixels in the matrix to be transmitted (see Section 4.3). This can be considered as one of the major achievements of this research work.

In this first prototype version, the resolution of the sensor is low. Real-world applications would require the implementation of larger SCD pixel matrices. WTA matrices are not easy to scale-up. In Subsection 5.2, alternatives for the implementation of the decision block in larger SCD sensing

matrices are discussed.

The characterization process reveals several drawbacks and problems that have to be taken into account for future and improved versions of this sensor. Noise seems to be a real problem that degrades the quality of the visual information transmitted by the sensor. The achieved dynamic range is low, thus limiting the applicability of the sensor in many environmental lighting conditions. Power consumption per pixel is larger than other state-of-the-art event-driven vision sensors, and should be substantially reduced in future versions of the SCD pixel. The transitory time of the analogue grey level signal is too high and should also be reduced in future versions of the photocircuit.

CMOS VLSI designs such as the sensor proposed in this thesis are usually hard to develop. Chips that reach market standards are usually obtained after working on several runs. This thesis is the first iteration of the design process of an SCD vision sensor, and it is clear that a lot of work has still to be done to reach a final application. Nevertheless, under proper lighting conditions, successful experiments have been carried out using our newly developed SCD Vision hardware. The tracking experiment detailed in Section 4.3 clearly shows the advantages of an SCD-based vision system over a traditional-readout frame-based one. The system has been able to track objects at very high-speed by means of simple calculations with very few transmitted pixels. In this way, we have experimentally demonstrated how bandwidth and processing power can be reduced by orders of magnitude when SCD hardware is used. These experimental results are one of the major and most original achievements of this thesis.

5.1 Contributions

- Before the beginning of this research work, the advantages of SCD Vision had only been presented in simulated environments. One of the main contributions of this thesis is the design and implementation of the first VLSI CMOS vision sensor following SCD Vision principles. This achievement opens the door for the development and implementation of artificial vision systems where very high pixel rates can be achieved with standard, easily available low-cost processing hardware.

- The main challenge of the microelectronic development carried out

in this thesis was the design of the decision block needed to select the most relevant pixels in the matrix. Analogue WTA circuits are a widely used solution in situations where a single winner has to be selected from a group of current or voltage competing values. When the number of competing cells is large, as in our case, multiple winners can be observed at the output of the analogue WTA stage. One of the original contributions of this thesis is the proposal of a digital subcircuit to cope with multiple-winner situations in analogue WTA circuits. The final result is a mixed analogue-digital stage capable of selecting one single winner in a large array of WTA amplifiers. The necessity of the proposed digital logic has been confirmed with proper experimentation. In Subsection 4.2.3 it is shown that in normal sensor operation, if the digital WTA subcircuit is disabled, the analogue decision block alone is not able to select one single winner amongst the pixels in the sensing matrix.

- Another interesting achievement related to the WTA stage is that, until the moment of writing this document, no documentation has been found concerning the fabrication and testing of the particular WTA cell used in this thesis. Although the design of this cell was proposed by Sekerkiran *et al.* back in 1998, this is the first time this cell has been fabricated and tested. The tracking experiment presented in Section 4.3 shows that the decision block based on this cell successfully selects the pixels undergoing the largest changes in the matrix.

- The implementation of a compact SCD system based on the full-custom VLSI sensor designed in this thesis, and the experimentation carried out with it are also one of the major achievements of this research work. Despite the problems related to noise and dynamic range detailed and discussed in Subsections 4.2.5, 4.2.2 and 4.2.9, the experimentation detailed in Subsection 4.3 demonstrates how a simple SCD system based on the sensor developed in this thesis is able to track fast moving objects with the bandwidth requirements of a 25 fps standard camera, but with the time resolution and processing time of a high-speed camera working at 2000 fps.

5.2   Proposals for improvements and future work

In Subsection 4.2.9 several alternatives have already been proposed to improve the performance of the sensor in many aspects. In this section, the ideas for improvements of the pixel circuit presented in the previous sections of this document are summarized together with new proposals for future research and development work.

- **Noise**: this is a severe problem affecting the quality of the visual information transmitted by the sensor. For sampling periods below 2-1 ms the tracking algorithms can only be applied by binarizing the grey level of the transmitted pixels. FPN could be reduced by implementing a CTIA to integrate the photocurrent over a PIP capacitor instead of discharging the parasitic capacitance of the photodiode. CDS techniques could also help to reduce this source of noise.

- **Dynamic range**: this parameter should be increased so that the sensor can be used in a wider range of lighting environments. If the intention is to linearly integrate the photocurrent, enlarging the sensitive area of the pixel or using dual-photodiode architectures are both suitable alternatives. Another alternative could be to integrate the photocurrent up to a fixed voltage instead of integrating for a fixed time. If the latter is implemented the integration time cannot be set to any desired value, instead, it will depend on the lighting conditions of the environment. This can be considered a serious disadvantage if the sensor is to be operated at very high frame speeds (see Subsections 4.2.9 and 4.2.2).

- **Power consumption**: the power consumption per pixel should definitely be reduced in future versions of the pixel circuit. If positive and negative changes are treated separately, the current-rectifier subcircuit could be omitted, therefore saving power consumption and silicon area. The rectifier is not the only source of power consumption, so a careful analysis of this aspect should be carried-out for future versions of the SCD pixel (see Subsections 4.2.9 and 4.2.7).

- **WTA failures**: two problems affect the WTA subcircuit. The first one is that, occasionally, the WTA block is not able to choose a winner amongst the pixels in the matrix. Statistics show that this problem

rarely happens, so no modifications to the circuit are proposed. The second problem is that sometimes the same pixel is selected to be transmitted within the same frame-period. This happens when the difference between the voltages at the `Cprevious` and `Cpresent` capacitors is too small for all the pixels in the matrix. In this case, all the pixels, including the ones that have just updated the charge of its `Cprevious` capacitor, compete with similar credentials at the WTA decision block. This misbehaviour is the reason why FPN cannot be easily measured. As explained in Subsection 4.2.4.2 this problem is inherent in the current pixel design, but it could be easily solved with simple modifications to the pixel logic.

- **Winner selection biasing**: the selection of the winning pixel is biased by the fixed propagation path of the inhibiting logic signal (see Fig. 3.13 in Subsection 3.2.3.2). This could be solved by changing the starting point of the inhibiting path from one frame (or pixel) period to the next one. An in-pixel codifier could output a logic "1" when the address of the pixel is externally selected, and a logic "0" in any other case. This logic signal could be used to select the starting point of the inhibiting path. Other topologies for the inhibiting signal path can also be considered. Fig. 3.24 describes a more biologically inspired option where the centre of the matrix is considered more relevant than the periphery.

- **Unwanted WTA feedback**: another modification to the circuit that should be incorporated is related to the level-sensitive latch of the digital subcircuit (`INVA-B` and `TXG1-2` in Fig. 3.17). This should be replaced with an edge-sensitive one. This would avoid the unwanted feedback now present in the pixel-circuit, also relaxing the 100 ns pulse-width constraint on the `CKwta` signal.

- **Scalability**: As the power consumption is mainly static due to the constant current sources inside the pixel, this quantity increases linearly with the number of pixels in the matrix. So it is important to carefully analyse and reduce the power consumption at pixel level before enlarging the pixel matrix dimensions. Another important aspect to be considered is the scalability of the decision stage. WTA matrices are difficult to scale-up, so for the future implementation of larger SCD pixel matrices two possibilities arise. The first one is to assemble several 32x32 WTA matrices. The visual field would be sectored,

and the most interesting pixels in each sector would be found and transmitted out of the chip. One of the advantages of this strategy is that the attention of the processing system could be focused only in the sectors of the matrix that show changes above a certain threshold. This would further save on processing power and bandwidth since only a few sectors of the matrix would be fully analysed and processed. The other possibility would be to implement the comparison/decision block with a completely different subcircuit. Variable threshold circuits have already been implemented to accomplish similar goals under similar contexts [5] and could be easily adapted to our particular problem.

- **Global reset**: some global reset signal for `Cprevious` capacitors should be added. This, together with the implementation of a mechanism that prevents a certain pixel being read out twice within the same frame-period, would allow the grey level values of all the pixels in the matrix within one frame-period to be read.

5.3 List of publications and research projects

Papers in the Journal Citation Report

- P. Zuccarello, F. Pardo, A. de la Plaza, and J. A. Boluda. 32x32 winner-take-all matrix with single winner selection. *Electronics Letters*, 46:5 (Mar. 2010), pp. 333 - 335.

- F. Pardo, P. Zuccarello, J. A. Boluda, and F. Vegara. Advantages of Selective Change Driven Vision for resource-limited systems. *IEEE Trans. on Circuits and Systems for Video Tech.* 21:10 (Oct. 2011). *Special Issue on Video Analysis on Resourced-Limited Systems*, pp. 1415 - 1423.

- J. A. Boluda, P. Zuccarello, F. Pardo, and F. Vegara. Selective Change-Driven Imaging: A Biomimetic Visual Sensing Strategy. *Sensors*, 11:11 (Nov. 2011). *Special Issue on Biomimetic Sensors, Actuators and Integrated Systems*, pp. 11000 - 11020.

Papers in Lecture Notes

- F. Pardo, J. A. Boluda, F. Vegara, and P. Zuccarello. *On the Advantages of Asynchronous Pixel Reading and Processing for High-Speed Motion Estimation.* 4th International Symposium on Advances in Visual Computing (ISVC 2008). Dec. 1 - 3, 2008. Las Vegas, NV, USA. In: *Lecture Notes in Computer Science.* Vol. 5358/2008: *Advances in Visual Computing.* Springer-Verlag Berlin, Heidelberg, 2008, pp. 205 - 215.

- J. A. Boluda, F. Vegara, F. Pardo, and P. Zuccarello. *Selective Change-Driven Image Processing: a Speeding-up Strategy.* 14th Iberoamerican Conference on Pattern Recognition (CIARP 2009). Nov. 15 - 18, 2009. Guadalajara, Jalisco, Mexico. In: *Lecture Notes in Computer Science.* Vol. 5856/2009: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications.* Springer-Verlag Berlin, Heidelberg, 2009, pp. 3744.

Papers in Peer-Reviewed International Conferences

- P. Zuccarello, F. Pardo, and F. Vegara. Silicon implementation of a 32x32-pixel vision sensor for Selective Change Driven read-out strategy. *Proceedings of the 8th Conference on PhD. Research in Microelectronics and Electronics (PRIME 2012).* June 12 - 15, 2012. Aachen, Germany.

- P. Zuccarello, F. Pardo, A. de la Plaza, and J. A. Boluda. A 32x32 pixels vision sensor for Selective Change Driven readout strategy. *36th European Solid State Circuit Research Conference (ESSCIRC 2010)-Poster session.* Sept. 13 - 17, 2010. Seville, Spain.

Research projects

Title: Development of sensor and techniques for asynchronous change-driven vision for the analysis of very high speed movement.
Funded by: Ministry of Science and Technology of Spain - TEC2006-08130
Responsible researcher: Fernando Pardo Carpio
Duration from: January-2007   To: December-2009
Funds: €49,900.00   Number of researchers: 7

Title: Development of techniques, sensor and equipment for selective change-driven vision.
Funded by: Ministry of Science and Technology of Spain - TEC2009-12980
Responsible researcher: Fernando Pardo Carpio
Duration from: January-2010   To: December-2013
Funds: €50,700.00   Number of researchers: 7

Bibliography

[1] C. Mead and M. Mahowald. "A silicon model of early visual processing." *Neural Networks*, 1:1 (1988), pp. 91–97.

[2] M. A. Mahowald. "VLSI analogs of neural visual processing: a synthesis of form and function". PhD thesis. Pasadena, California: California Institute of Technology, May 1992.

[3] K. Boahen. "The Retinomorphic Approach: Pixel-Parallel Adaptive Amplification, Filtering, and Quantization". *Analog Integrated Circuits and Signal Processing*, 13:1-2 (May 1997), pp. 53–68.

[4] P. Lichtsteiner, C. Posch, and T. Delbrück. "A 128x128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor". *IEEE Journal of Solid-State Circuits*, 43:2 (Feb. 2008), pp. 566–576.

[5] P.-F. Rüedi et al. "A 128x128 Pixel 120-dB Dynamic-Range Vision-Sensor Chip for Image Contrast and Orientation Extraction". *IEEE Journal of Solid-State Circuits*, 38:12 (Dec. 2003), pp. 2325–2333.

[6] Y. M. Chi, U. Mallik, M. A. Clapp, E. Choi, G. Cauwenberghs, and R. Etienne-Cummings. "CMOS Camera With In-Pixel Temporal Change Detection and ADC". *IEEE Journal of Solid-State Circuits*, 42:10 (Oct. 2007), pp. 2187–2196.

[7] T. Delbrück and S.-C. Liu. "A silicon early visual system as a model animal". *Vision Research*, 44:17 (Aug. 2004), pp. 2083–2089.

[8] V. Chan, S.-C. Liu, and A. van Schaik. "AER EAR: A Matched Silicon Cochela pair With Address Event Representation Interface". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 54:1 (Jan. 2007), pp. 48–59.

[9] Y. Wang and S.-C. Liu. "A Two-Dimensional Configurable Active Silicon Dendritic Neuron Array". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58:9 (Sept. 2011), pp. 2159–2171.

[10] L. Camuñas-Mesa, C. Zamarreño-Ramos, A. Linares-Barranco, A. Acosta-Jiménez, T. Serrano-Gotarredona, and B. Linares-Barranco. "An Event-Driven Multi-Kernel Convolution Processor Module for Event-Driven Vision Sensors". *IEEE Journal of Solid-State Circuits*, 47:2 (Feb. 2012), pp. 504–517.

[11] L. Camuñas-Mesa, A. Acosta-Jiménez, C. Zamarreño-Ramos, T. Serrano-Gotarredona, and B. Linares-Barranco. "A 32x32 Pixel Convolution Processor Chip for Address Event Vision Sensors With 155 ns Event Latency and 20 Meps Throughput". *IEEE Journal of Solid-State Circuits*, 47:2 (Feb. 2012), pp. 504–517.

[12] K. Zaghloul and K. Boahen. "Optic nerve signals in a neuromorphic chip I: outer and inner retina models". *IEEE Transactions on Biomedical Engineering*, 51:4 (Apr. 2004), pp. 657–666.

[13] K. Zaghloul and K. Boahen. "Optic nerve signals in a neuromorphic chip II: testing and results". *IEEE Transactions on Biomedical Engineering*, 51:4 (Apr. 2004), pp. 667–675.

[14] C. Bartolozzi and G. Indiveri. "Selective Attention in Multi-Chip Address-Event Systems". *Sensors*, 9 (June 2009), pp. 5076–5098.

[15] G. Indiveri, E. Chicca, and R. Douglas. "Artificial cognitive systems: From VLSI networks of spiking neurons to neuromorphic cognition". *Cognitive Computation*, 1:2 (2009), pp. 119–127.

[16] E. Culurciello, R. Etienne-Cummings, and K. Boahen. "A Biomorphic Digital Image Sensor". *IEEE Journal of Solid-State Circuits*, 38:2 (Feb. 2003), pp. 281–294.

[17] K. Boahen. "Neuromorphic microchips". *Scientific American*, 292:5 (May 2005), pp. 56–63.

[18] K. Boahen. "A Burst-Mode Word-Serial Address-Event Link-I: Transmitter Design". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 51:7 (July 2004), pp. 1269–1280.

[19] K. Boahen. "A Burst-Mode Word-Serial Address-Event Link-II: Receiver Design". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 51:7 (July 2004), pp. 1281–1291.

[20] K. Boahen. "A Burst-Mode Word-Serial Address-Event Link-III: Analisys and Testing". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 51:7 (July 2004), pp. 1292–1300.

[21] P. Rüedi et al. "An SoC Combining a 132 dB QVGA Pixel Array and a 32 dB DSP/MCU Processor for Vision Applications". *Proceedings of the 2009 IEEE International Solid-State Circuits Conference (ISSCC '09)*, pp. 46–47. Feb. 8–12, 2009. San Francisco, CA, USA.

[22] M. Gottardi, N. Massari, and S.-A. Jawed. "A 100 μW 128x64 Pixels Contrast-Based Asynchronous Binary Vision Sensor for Sensor Networks Applications". *IEEE Journal of Solid-State Circuits*, 44:5 (May 2009), pp. 1582–1592.

[23] X. Guo, X. Qi, and J. G. Harris. "A Time-to-First-Spike CMOS Image Sensor". *IEEE Sensors Journal*, 7:8 (Aug. 2007), pp. 1165–1175.

[24] S. Chen, W. Tang, X. Zhang, and E. Culurciello. "A 64x64 UWB Wireless Temporal-Difference Digital Image Sensor". *IEEE Transactions on Very Large Scale Integration Systems*, 20:12 (Dec. 2012), pp. 2232–2240.

[25] D. Kim and E. Culurciello. "A Compact-pixel Tri-mode Vision Sensor". *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems (ISCAS '10)*, pp. 2434–2437. May 30–June 2, 2010. Paris, France.

[26] D. Kim, Z. Fu, J. H. Park, and E. Culurciello. "A 1 mW CMOS Temporal-Difference AER Sensor for Wireless Sensor Networks". *IEEE Transactions on Electron Devices*, 56:11 (Nov. 2009), pp. 2586–2293.

[27]   F. Pardo, J. A. Boluda, F. Vegara, and P. Zuccarello. *On the Advan-
       tages of Asynchronous Pixel Reading and Processing for High-Speed
       Motion Estimation.* 4th International Symposium on Advances in Vi-
       sual Computing (ISVC 2008). Dec. 1–3, 2008. Las Vegas, NV, USA.
       In: *Lecture Notes in Computer Science.* Vol. 5358/2008: *Advances in
       Visual Computing.* Springer-Verlag Berlin, Heidelberg, 2008, pp. 205–
       215.

[28]   J. A. Boluda, F. Vegara, F. Pardo, and P. Zuccarello. *Selective Change-
       Driven Image Processing: a Speeding-up Strategy.* 14th Iberoamerican
       Conference on Pattern Recognition (CIARP 2009). Nov. 15–18, 2009.
       Guadalajara, Jalisco, Mexico. In: *Lecture Notes in Computer Sci-
       ence.* Vol. 5856/2009: *Progress in Pattern Recognition, Image Analy-
       sis, Computer Vision, and Applications.* Springer-Verlag Berlin, Hei-
       delberg, 2009, pp. 37–44.

[29]   F. Pardo, X. Benavent, J. Boluda, and F. Vegara. "Selective Change-
       Driven image processing for high-speed motion estimation". *Proceed-
       ings of the 13th International Conference on Systems, Signals and
       Image Processing (IWSSIP '06)*, pp. 163–166. Sept. 21–23, 2006. Bu-
       dapest, Hungary.

[30]   F. Pardo, J. Boluda, X. Benavent, J. Domingo, and J. Sosa. "Cir-
       cle detection and tracking speed-up based on change-driven image
       processing". *Proceedings of the first ICGST International Conference
       on Graphics, Vision and Image Processing (GVIP '05)*, pp. 131–136.
       Dec. 19–21, 2005. Cairo, Egypt.

[31]   J. P. Shen and M. H. Lipasti. *Modern Processor Degin: Fundamentals
       of Superscalar Processors.* McGraw Hill Higher Education, 2004. ISBN:
       978-0070570641.

[32]   J. A. Boluda and F. Pardo. *Speeding-up Differential Motion Detection
       Algorithms Using a Change-Driven Data Flow Processing Strategy.*
       12th International Conference on Computer Analysis of Images and
       Patterns (CAIP '07). Aug. 27–29, 2007. Vienna, Austria. In: *Lecture
       Notes in Computer Science.* Vol. 4673: *Computer Analysis of Images
       and Patterns.* Springer Berlin, Heidelberg, 2007, pp. 77–84.

[33] J. Sosa, J. Boluda, F. Pardo, and R. Gómez-Fabela. "Change-Driven data flow image processing architecture for optical flow computation". *Journal of Real-Time Image Processing*, 2:4 (Dec. 2007). *Special Issue on Field-Programmable Technology for Real-Time Image Processing*, pp. 259–270.

[34] J. Lazzaro, S. Rayckebusch, M. A. Mahowald, and C. A. Mead. "Winner-take-all networks of O(n) complexity". In: *Advances in Neural Information Processing Systems*. Ed. by D. S. Touretzky. Vol. 1. San Mateo, CA: Morgan Kaufmann, 1989, pp. 703–711.

[35] J. Starzyk and X. Fang. "CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback". *Electronics Letters*, 29:10 (May 1993), pp. 908–910.

[36] B. Sekerkiran and U. Cilingiroglu. "Precision improvement in current-mode winner-take-all circuits using gain-boosted regulated-cascode CMOS stages". *The IEEE International Joint Conference on Neural Networks Proceedings*. Vol. 1, pp. 553–556. May 4–9, 1998. Anchorage, Alaska, USA.

[37] A. Demosthenous, S. Smedley, and J. Taylor. "A CMOS analog winner-take-all network for large-scale applications". *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 45:3 (Mar. 1998), pp. 300–304.

[38] N. Donckers, C. Dualibe, and M. Verleysen. "Design of Complementary Low-Power CMOS Architecures for Looser-take-all and Winner-take-all". *Proceedings of the 7th International Conference on Microelectronics for Neural, Fussy and Bio-inspired Systems*, pp. 360–365. Apr. 7–9, 1999. Granada, Spain.

[39] G. Indiveri. "A Current-Mode Hysteretic Winner-take-all Network, with Excitatory and Inhibitory Coupling". *Analog Integrated Circuits and Signal Processing*, 28:3 (Sept. 2001), pp. 279–291.

[40] G. Indiveri. "Winner-Take-All Networks with Lateral Excitation". *Analog Integrated Circuits and Signal Processing*, 13:1/2 (May 1997), pp. 185–193.

[41]   K. Murari, R. Etienne-Cummings, N. Thakor, and G. Cauwenberghs. "Which photodiode to use: A comparison of CMOS-compatible structures". *IEEE Sensors Journal*, 9:7 (July 2009), pp. 752–760.

[42]   J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona, and B. Linares-Barranco. "A Spatial Contrast Retina With On-Chip Calibration for Neuromorphic Spike-Based AER Vision Systems". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 54:7 (July 2007), pp. 1444–1458.

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

[44]   J. A. Leñero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco. "A Five-Decade Dynamic-Range Ambient-Light-Independent Calibrated Signed-Spatial-Contrast AER Retina With 0.1 ms Latency and Optional Time-to-First-Spike Mode". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 57:10 (Oct. 2010), pp. 2632–2643.

[45]   R. Serrano-Gotarredona et al. "CAVIAR: A 45k Neuron, 5M Synapse, 12G Connects/s AER Hardware Sensory-Processing-Learning-Actuating System for High-Speed Visual Object Recognition and Tracking". *IEEE Transactions on Neural Networks*, 20:9 (Sept. 2009), pp. 1417–1438.

[46]   Z. Fu, T. Delbrück, P. Lichtsteiner, and E. Culurciello. "An Address-Event Fall Detector for Assisted Living Applications". *IEEE Transactions on Biomedical Circuits and Systems*, 2:2 (June 2008), pp. 88–96.

[47]   D. Drazen, P. Lichtsteiner, P. Häfliger, T. Delbrück, and A. Jensen. "Toward real-time particle tracking using an event-based dynamic vision sensor". *Experiments in Fluids*, 51:5 (2011), pp. 1465–1469.

[48]   J. Conradt, M. Cook, R. Berner, P. Lichtsteiner, and T. Delbrück. "Live Demonstration: A Pencil Balancing Robot using a Pair of AER Dynamic Vision Sensors". *Proceedings of the 2009 IEEE Interna-*

tional Symposium on Circuits and Systems (ISCAS '09), pp. 781–785. May 24–27, 2009. Taipei, Taiwan.

[49] J. Lee, T. Delbrück, P. K. Park, M. Pfeiffer, C.-W. Shin, H. Ryu, and B. chang Kang. "Live Demonstration: Gesture-Based remote control using stereo pair of dynamic vision sensors". *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems (IS-CAS '12)*, pp. 741–745. May 20–23, 2012. Seoul, Korea.

[50] T. Delbrück and R. Berner. "Temporal Contrast AER Pixel with 0.3%-Contrast Event Threshold". *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems (ISCAS '10)*, pp. 2442–2445. May 30–June 2, 2010. Paris, France.

[51] J. A. Leñero-Bardallo, T. Serrano-Gotarredona, and B. Linares-Barranco. "A 3.6 μs Latency Asynchronous Frame-Free Event-Driven Dynamic Vision Sensor". *IEEE Journal of Solid-State Circuits*, 46:6 (June 2011), pp. 1443–1455.

[52] C. Posch, D. Matolin, and rainer Wohlgenannt. "A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS". *IEEE Journal of Solid-State Circuits*, 46:1 (Jan. 2011), pp. 259–275.

[53] D. Matolin, C. Posch, and R. Wohlgenannt. "True correlated double sampling and comparator design for time-based image sensors". *Proceedings of the 2009 IEEE International Symposium on Circuits and Systems (ISCAS '09)*, pp. 1269–1272. May 24–27, 2009. Taipei, Taiwan.

[54] M. Azadmehr, H. Abrahamsen, and P. Häfliger. "A foveated AER imager chip". *Proceedings of the 2005 IEEE International Symposium on Circuits and Systems (ISCAS '05)*, pp. 2751–2754. May 23–26, 2005. Kobe, Japan.

[55] J. Kramer. "An integrated optical transient sensor." *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 49:9 (Sept. 2002), pp. 612–628.

[56] R. Berner and T. Delbrück. "Event-Based Pixel Sensitive to Changes of Color and Brightness". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58:7 (July 2011), pp. 1581–1590.

[57] J. Leñero-Bardallo, D. Bryn, and P. Häfliger. "Bio-inspired Asynchronous Pixel Event Tri-color Vision Sensor". *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS 2011)*, pp. 253–256. Nov. 10–12, 2011. San Diego, California, USA.

[58] K. Aizawa, H. Ohno, Y. Egi, T. Hamamoto, M. Hatori, H. Maruyama, and J. Yamazaki. "On Sensor Image Compression". *IEEE Trans. on Circuits and Systems for Video Tech.* 7:3 (June 1997), pp. 543–548.

[59] K. Aizawa, Y. Egi, T. Hamamoto, M. Hatori, M. Abe, H. Maruyama, and H. Otake. "On Sensor Image Compression". *IEEE Transactions on Electron Devices*, 44:10 (Oct. 1997), pp. 1724–1730.

[60] E. Grenet, S. Gyger, P. Heim, F. Heitger, F. Kaess, and P.-F. Rüedi. "High dynamic range vision sensor for automotive applications". *Proceedings of the SPIE 5663 - Photonics in the automobile*, pp. 246–253. Feb. 28, 2005.

[61] N. Massari, M. Gottardi, L. Gonzo, D. Stoppa, and A. Simoni. "A CMOS Image Sensor With Programmable Pixel-Level Analog Processing". *IEEE Transactions on Neural Networks*, 16:6 (Nov. 2005), pp. 1673–1684.

[62] N. Massari and M. Gottardi. "A 100-dB Dynamic-Range CMOS Vision Sensor With Programmable Image Processing and Global Feature Extraction". *IEEE Journal of Solid-State Circuits*, 42:3 (Mar. 2007), pp. 647–657.

[63] L. Gasparini, R. Manduchi, M. Gottardi, and D. Petri. "An Ultralow-Power Wireless Camera Node: Development and Performance Analysis". *IEEE Transactions on Instrumentation and Measurement*, 60:12 (Dec. 2011), pp. 3824–3832.

[64] N. Cottini, L. Gasparini, M. De-Nicola, N. Massari, and M. Gottardi. "A CMOS Ultra-Low Power Vision Sensor With Image Compression and Embedded Event-Driven Energy-Management". *IEEE Journal*

on Emerging and Selected Topics in Circuits and Systems, 1:3 (Sept. 2011), pp. 299–307.

[65] F. Vegara, J. Boluda, J. Domingo, F. Pardo, and X. Benavent. "Accelerating Motion Analysis Algorithms with a Pixel Change-Driven Scheme". *Proceedings of the 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV '09)*, pp. 895–900. July 13–16, 2009. Las Vegas, USA.

[66] A. Yilmaz, O. Javed, and M. Shah. "Object tracking: A survey." *ACM Comput. Surv.* 38:4 (Dec. 2006), 1–45–Article 13.

[67] T. Delbrück, B. Linares-Barranco, E. Culurciello, and C. Posch. "Activity-Driven Event-Based Vision Sensors". *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems (ISCAS '10)*, pp. 2426–2429. May 30–June 2, 2010. Paris, France.

[68] P. Zuccarello, F. Pardo, and F. Vegara. "Silicon implementation of a 32x32-pixel vision sensor for Selective Change Driven read-out strategy". *Proceedings of the 8th Conference on PhD. Research in Microelectronics and Electronics (PRIME 2012)*. June 12–15, 2012. Aachen, Germany.

[69] F. Pardo, P. Zuccarello, J. A. Boluda, and F. Vegara. "Advantages of Selective Change Driven Vision for resource-limited systems". *IEEE Trans. on Circuits and Systems for Video Tech.* 21:10 (Oct. 2011). *Special Issue on Video Analysis on Resourced-Limited Systems*, pp. 1415–1423.

[70] J. A. Boluda, P. Zuccarello, F. Pardo, and F. Vegara. "Selective Change-Driven Imaging: A Biomimetic Visual Sensing Strategy". *Sensors*, 11:11 (Nov. 2011). *Special Issue on Biomimetic Sensors, Actuators and Integrated Systems*, pp. 11000–11020.

[71] R. Serrano-Gotarredona et al. "On Real-Time AER 2-D Convolutions Hardware for Neuromorphic Spike-Based Cortical Processing". *IEEE Transactions on Neural Networks*, 19:7 (July 2008), pp. 1196–1219.

[72]   R. Serrano-Gotarredona, T. Serrano-Gotarredona, A. Acosta-Jiménez, and B. Linares-Barranco. "A Neuromorphic Cortical-Layer Microchip for Spike-Based Processing Vision Systems". *IEEE Transactions on Circuits and Systems I: Regular Papers*, 53:12 (Dec. 2006), pp. 2548–2566.

[73]   A. Nedungadi and T. Viswanathan. "Design of Linear CMOS Transconductance Elements". *IEEE Transactions on Circuits and Systems*, CAS-31:10 (Oct. 1984), pp. 891–894.

[74]   A. Fish, D. Turchin, and O. Yadid-Pecht. "An APS With 2-D Winner-Take-All Selection Employing Adaptive Spatial Filtering and False Alarm Reduction". *IEEE Transactions on Electron Devices*, 50:1 (Jan. 2003), pp. 159–165.

[75]   J. Choi and B. J. Sheu. "A High-Precision VLSI Winner-Take-All Circuit for Self-Organizing neural Networks". *IEEE Journal of Solid-State Circuits*, 28:5 (May 1993), pp. 576–584.

[76]   S.-C. Liu. "A silicon retina with controllable winner-take-all properties". *Proceedings of the 2003 International Symposium on Circuits and Systems (ISCAS '03)*. Vol. 4, pp. 804–807. May 25–28, 2003. Bangkok, Thailand.

[77]   P. Zuccarello, F. Pardo, A. de la Plaza, and J. A. Boluda. "32x32 winner-take-all matrix with single winner selection". *Electronics Letters*, 46:5 (Mar. 2010), pp. 333–335.

[78]   A. Moini. *Vision Chips*. Kluwer Academic Publishers, 2000. ISBN: 0-7923-8664-7.

[79]   M. Bigas, E. Cabruja, J. Forest, and J. Salvi. "Review of CMOS image sensors". *Microelectronics Journal*, 37:5 (May 2006), pp. 433–451.

[80]   Z. Wang. "Novel Pseudo RMS Current Converter for Sinusoidal Signals Using a CMOS Precision Current Rectifier". *IEEE Transactions on Instrumentation and Measurement*, 39:4 (Aug. 1990), pp. 670–671.

[81]   B. Sekerkiran and U. Cilingiroglu. "Improving the Resolution of Lazzaro Winner-Take-All Circuit". *International Conference on Neural Networks - 1997*. Vol. 2, pp. 1005–1008. June 9–12, 1997. Houston, TX, USA.

[82] N. Massari and M. Gottardi. "Low power WTA circuit for optical position detector". *Electronics Letters*, 42:24 (Nov. 2006), pp. 1373–1374.

[83] Z. Kalayjian and A. G. Andreou. "Asynchronous Communication of 2D Motion Information Using Winner-Takes-All Arbitration". *Analog Integrated Circuits and Signal Processing*, 13:1/2 (May 1997), pp. 103–109.

[84] G. Indiveri, P. Oswald, and J. Kramer. "An adaptive visual tracking sensor with hysteretic winner-take-all network". *Proceedings of the IEEE 2002 International Symposium on Circuits and Systems - ISCAS 2002*. Vol. 2, pp. II324–II327. May 26–29, 2002. Phoenix-Scottsdale, AZ, USA.

[85] P. Häfliger. "Adaptive WTA With an Analog VLSI Neuromorphic Learning Chip". *IEEE Transactions on Neural Networks*, 18:2 (Mar. 2007), pp. 551–572.

[86] E. Özalevli, P. Hasler, and C. M. Higgins. "Winner-Take-All-Based Visual Motion Sensors". *IEEE Transactions on Circuits and Systems II: Express Briefs*, 53:8 (Aug. 2006), pp. 717–721.

[87] B. Razavi. *Design of Analog CMOS Integrated Circuits*. McGraw-Hill Education (Asia) Co. and Tsinghua University Press, 2005. ISBN: 7-302-10886-2.

[88] B. J. Hosticka. "Improvement of the Gain of MOS Amplifiers". *IEEE Journal of Solid-State Circuits*, 14:6 (Dec. 1979), pp. 1111–1114.

[89] P. Zuccarello, F. Pardo, A. de la Plaza, and J. A. Boluda. "A 32x32 pixels vision sensor for Selective Change Driven readout strategy". *36th European Solid State Circuit Research Conference (ESSCIRC 2010)- Poster session*. Sept. 13–17, 2010. Seville, Spain.

[90] R. Figueras, J. Sabadell, F. Serra-Graells, and L. Terés. "A 70-μm 8-μW Self-Biased Charge-Integration Active Pixel for Digital Mamomgraphy". *IEEE Transactions on Biomedical Circuits and Systems*, 5:5 (Oct. 2011), pp. 481–489.

[91]   J.-B. Chun, H. Jung, and C.-M. Kyung. "Dynamic-Range Widening
        in a CMOS Image Sensor Through Exposure Control Over a Dual-
        Photodiode Pixel". *IEEE Transactions on Electron Devices*, 56:12
        (Dec. 2009), pp. 3000–3008.