Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Diplom-Physicist:	Thomas Nirmaier
born in:	Ludwigshafen am Rhein
	Oral examination: July 15, 2003

A CMOS-based Hartmann-Shack Sensor for Real-Time Adaptive Optical Applications

Referees: Prof. Dr. Josef Bille
Prof. Dr. Reinhard Männer

Zusammenfassung:
Die wachsende Anzahl von Anwendungen für adaptiv optische Systeme in der Ophthalmologie motiviert die Entwicklung von geeigneten Sensoren und Aktuatoren. Als Wellenfrontsensoren werden meist Hartmann-Shack-Sensoren verwendet, die die laterale Verschiebung von Fokalpunkten einer Mikrolinsenmatrix in der Fokalebene zur Berechnung der Wellenfront verwenden. Die Bandbreite heutiger Systeme ist durch die Bildverarbeitung des Spotmusters auf einige Hz begrenzt, was ihren Nutzen für schnelle Echtzeitsysteme stark beeinträchtigt.
Im Rahmen dieser Arbeit wurde ein schneller Wellenfrontsensor auf der Basis eines applikationsspezifischen integrierten Schaltkreises entwickelt und getestet, der Wellenfrontmessungen bis zu 6 kHz Bandbreite erlaubt. Auf diesem Sensorchip wurden Photodetektoren mit 40% Quanteneffizienz bei 680 nm Wellenlänge sowie die Signalverarbeitung zur Bestimmung der Spotpositionen integriert. Diese basiert auf einer neuartigen Schaltkreistopologie zur Reduktion der Effekte von in CMOS-Sensoren (Complementary Metal Oxide Semiconductor) häufig auftretenden Parameterschwankungen. Für ophthalmologische Anwendungen musste insbesondere eine sehr große Sensitivität bei Strahlungsleistungen der Fokalpunkte von weniger als 1 nW erreicht werden.
Mit dem gefertigten Sensorsystem konnten Wellenfrontmessungen mit einer Genauigkeit von 0.16 dpt Defokus bei lediglich 160 pW Strahlungsleistung durchgeführt werden. Erstmals war es dadurch möglich, Wellenfrontaberrationen am menschlichen Auge mit 300 Hz Wiederholrate durchzuführen und deren Leistungsspektrum zu bestimmen.

Abstract:
Adaptive optical systems have a growing field of applications in opthalmology. In every adaptive system there is the need for a sensor and an actuator. The Hartmann-Shack wavefront sensor uses the displacement of spots in the focal plane of a lenslet array for subsequent calculation of the wavefront. The bandwidth of current sensors is mostly limited by software processing of the focal plane image to some tens of Hz, which makes it unsuitable for real-time adaptive optical systems.
To overcome the current bandwidth limitations a fast Hartmann-Shack sensor based on an application specific integrated circuit has been developed and tested, that reaches a bandwidth of up to 6 kHz. The sensor includes photodetectors with 40% quantum efficiency at 680 nm wavelength and an image processing, that is especially suitable to reduce the effects of the common mismatching of process parameters in CMOS-based sensors (Complementary Metal Oxide Semiconductor). A special problem in ophthalmic applications is the low available spot power below 1 nW.
The developed Hartmann-Shack sensor allowed wavefront measurements with an accuracy of 0.16 dpt defocus at 160 pW spot power. It has been possible for the first time, to measure wavefront aberrations at the living human eye with 300 Hz repetition rate and to calculate the power spectral density of these aberrations.

Contents

Introduction **5**

1 Adaptive Optics in Opthalmology **7**
1.1 Refractive Errors of the Human Eye 7
1.2 Adaptive Optics in Opthalmology 9
1.3 Wavefront Sensors 10
1.4 Hartmann-Shack Sensor 12
1.4.1 Spot Irradiance Distribution 13
1.4.2 Zernike polynomials 14
1.4.3 Wavefront Reconstruction 15

2 CMOS Process Technology **21**
2.1 Introduction 21
2.2 The MOSFET 22
2.2.1 The MOSFET as an Active Resistor 24
2.2.2 Matching of MOS Transistors 25
2.3 Passive Devices 26
2.4 Silicon Photodetectors 26
2.4.1 Junction Photodiodes 28
2.4.2 Photodiode Noise Sources 31

3 Position Sensitive Detectors **33**
3.1 Introduction 33
3.2 Spatial Estimators of Location 34
3.2.1 Influence of Impulsive Noise and Background Illumination . . 37
3.2.2 Estimators of Spread 38
3.3 Lateral-Effect Photodiode 39
3.4 Quad-Cell 40
3.5 Multipixel Position Sensitive Detectors 41
3.5.1 Resistive-Line Networks 42
3.5.2 Winner-Take-All Circuit 42
3.5.3 Transient Response 46
3.5.4 Local Inhibition Winner-Take-All Circuit 47
3.5.5 Resistive-Ring Network of Winner-Take-All Circuits 49

3.6 Position Uncertainty 52

4 The Centroid Hartmann-Shack Sensor 59
4.1 General Architecture 59
4.2 Position Sensitive Detector 60
4.3 Pixel Array 61
4.4 Resistive-Ring Network of WTA Circuits 62
4.4.1 Acceleration of the Transient Response 63
Decoupling of the Junction Capacitance 64
Current Mirror 68
4.4.2 Output Amplifier 69
4.4.3 Detector Layout 70
Mismatching Considerations 71
Resistive-Ring Network and Current Sinks 73
4.5 Position-Data Read-Out 75
4.5.1 Address Decoder 75
4.5.2 Data Read-Out 76
4.5.3 Communication Test 78
4.6 The Prototype 80
4.6.1 Quantum Efficiency and Spot-Size Measurements 80
4.6.2 Position Sensitive Detector 82
4.6.3 Resistive-Line Network 83
4.7 Conclusions from the Prototype 84

5 External System and Data Analysis 85
5.1 Software and Custom Device Driver 85
5.2 Bit-Vector Analyses 86
5.2.1 Spot Tracking 86
5.2.2 Spot Detection 88

6 Measurement Results 91
6.1 Optical Setup 91
6.2 Spot Tracking Results 91
6.2.1 Resistive-Ring Measurements and Sub-Pixel Resolution . . . 95
6.2.2 Spot Tracking with the Resistive-Line Network 97
6.3 Spot Detection 99
6.4 Static Wavefront Measurements 100
6.5 Dynamic Wavefront Measurements 102
6.5.1 Astigmatism Measurements 102
6.5.2 Wavefront Measurements at the Human Eye 105

Discussion 113

A Least-Squares Wavefront Reconstruction **117**

B Pin Description of the CeHSSA **119**

C Pin Description of the Prototype **121**

D Comparison between different CMOS-based Wavefront Sensors **123**

E List of Symbols **125**

Bibliography **129**

Introduction

Adaptive optical systems can help to reach the diffraction limit in imaging systems through a dynamic cancellation of the wavefront aberrations, that are introduced by a medium with a time-space variant index of refraction. The human eye is not a rigid structure and contains fluids that introduce aberrations to the wavefront. These aberrations limit the Visual Acuity, but also the resolution of retinal imaging system, which are often used as diagnostic tools.

Adaptive optical system consist at least of a wavefront sensor, a processor to calculate the wavefront and an actuator, which can be an adaptive micromirror. The Hartmann-Shack sensor is a common non-interferometric wavefront sensor, suitable for real-time applications, because it does not contain any movable parts. It consists of a two-dimensional microlens array for the spatial sampling of the wavefront and an image sensor in the focal plane. The spots from the lenslets move sideways in the focal plane, when the wavefront contains a local distortion. The wavefront can be reconstructed from these spot displacements. A global description of the wavefront with orthogonal Zernike polynomials is often used for this purpose.

In many applications the bandwidth of the adaptive optical system is limited by the software image processing of the spot pattern to a few Hz. This thesis presents a CMOS-based Hartmann-Shack sensor, that includes photodetectors and image processing on a single chip and thus reaches bandwidths in the kHz-range. A special difficulty in the design of a custom CMOS-based sensor is the very low available spot intensity, which only amounts to some nW, due to security restrictions on the maximum applicable laser power to the human eye. Special attention has to be paid to the quantum efficiency of the photodetectors and the signal processing has to deal with photocurrents of only a few tens of pA.

The thesis is structured as follows. After this introduction, the 1st chapter gives a short overview of the optical qualities of the human eye and introduces sensor concepts for wavefront measurements. The Hartmann-Shack sensor is described in detail and how a wavefront can be reconstructed from the spot pattern. Of special interest for the tracking of the spots is the irradiance distribution in the focal plane.

The 2nd chapter deals with the available photodetectors in the CMOS process technology, their quantum efficiency and noise characteristics. The regions of operation of the MOSFET are introduced, with special attention to the mismatching effects and how they can be avoided.

Different concepts of position sensitive detectors and their characteristics are ex-

plained in the 3rd chapter. After introducing the mathematical concept of position estimation, different physical realizations for position sensitive detectors are introduced, including the resistive-line network and the Winner-Take-All circuit. In this thesis a new topology for this circuit has been developed, the resistive-ring network of Winner-Take-All circuits, which has several advantages with respect to mismatching sensitivity and transient response time.

The 4th and the 5th chapter explain the design of the prototype and final version of the Centroid Hartmann-Shack Sensor ASIC (CeHSSA) and the external data acquisition system. The decoupling of the photodiodes from the neurons of the WTA circuit to achieve a larger bandwidth is introduced and the digital read-out scheme for random access to the spot detectors.

The 6th chapter introduces spot tracking and wavefront measurements. In the first part, the spot tracking performance has been evaluated at different spot powers and bandwidths. The second part of this chapter deals with static and dynamic wavefront measurements and closes with results from the first human eye wavefront measurements at several hundreds of Hz, including a temporal spectral analyses of the aberrations.

In the last chapter the results of the measurements are discussed and an outlook for further improvements of the system is given.

1 Adaptive Optics in Opthalmology

In this chapter the refractive errors of the human eye are introduced, how they can be canceled out with adaptive optical systems and thus, increase the resolution of diagnostic tools in opthalmology. The components of adaptive optical systems are presented, with special attention to the Hartmann-Shack wavefront sensor. The wavefront reconstruction with this special kind of wavefront sensor will be addressed and mathematical concepts for the wavefront description will be presented.

1.1 Refractive Errors of the Human Eye

Many opthalmic applications use the retinal reflection of a laser beam for diagnostics purposes. Figure 1.1 shows a cross-section of the human eye. Laser beams with wavelengths between 400 nm and 1400 nm can pass the optical apparatus through the *cornea* and the *lens*, and reach the *retina* [BW62]. The reflectivity of the retina depends on a large scale on the wavelength of the incident light.

The human eye is not an ideal optical system in the sense, that the quality of vision changes with age and often suffers from deficiencies, like refraction errors and opacity of the lens. Classical refraction errors are *myopia, hyperopia* and *astigmatism*. Superimposed to them are higher order aberrations like *coma* and *spherical aberrations*. Wavefront aberrations generally augment with a rising distance from the center of the pupil.

Refractive errors directly affect the *Visual Acuity*, which is a measure for the spatial resolving capacity of the optical system. Figure 1.2 shows the Visual Acuity with pupil size. For small pupil sizes the Visual Acuity is limited by diffraction. The Visual Acuity augments with pupil size up to about 3 mm [Cha91]. For even larger pupils the Visual Acuity is reduced again by the rising effect of refractive errors.

Any optical system for the imaging of the retina suffers from the same errors, as the Visual Acuity does. *Adaptive optical (AO)* systems for ophthalmic applications try to reach the diffraction limit for all pupil sizes by canceling these refractive errors out.

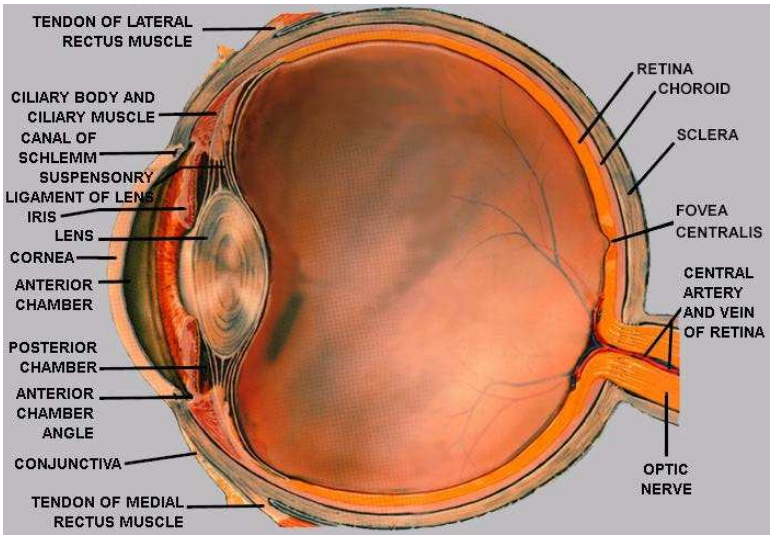

Figure 1.1: Cross-section of the human eye [Chu03].

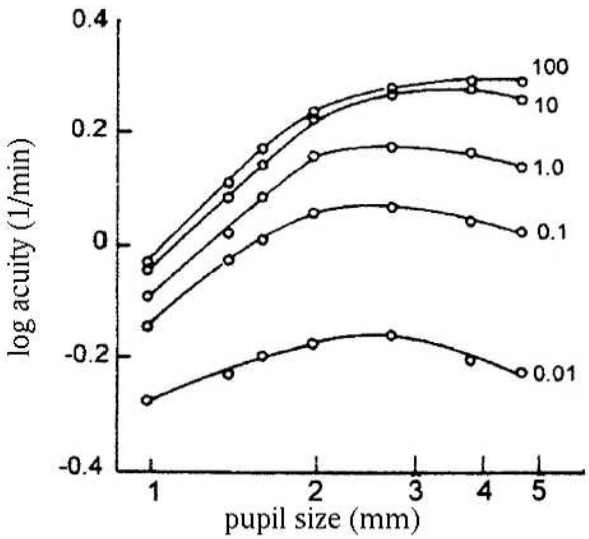

Figure 1.2: Visual Acuity with pupil size for five different illuminations (in Millilambert) [Cha91].

The bandwidth of such systems has to be larger then the bandwidth of the non-static wavefront aberrations in order to achieve the diffraction limit.

The human eye performs transient movements for fixation of the target, to avoid adaption of the photoreceptors and to adapt to different ambient light levels. Transient movements for the fixation of a target are *saccades* and *microsaccades*. They can reach angular velocities of up to 500° per second. The microsaccades only have a duration of up to 20 ms. Besides these transient movements a constant *microtremor* with a frequency of up to 90 Hz constantly moves the eye to avoid adaptation of the photoreceptors. Part of the normal focusing to an object is the *accommodation*. Besides this, there is also an *adaptation* to different ambient light intensities, where the eye shows an amazing dynamic range of about 11 decades of intensity [ST97].

1.2 Adaptive Optics in Opthalmology

The resolution of an ideal optical system is only limited by the size of its aperture. A system which reaches this limit, is said to be diffraction limited. When the light emanated from the object of interest passes through a medium with inhomogeneous index of refraction, different rays propagate with different velocities. A plane wavefront, which passes through such an inhomogeneous medium, leaves the medium in distorted form and reduces the resolution of the imaging system.

Adaptive optical systems cancel these aberrations out and allow diffraction limited imaging. These systems consist of a *wavefront sensor*, a processing unit and an actuator. Adaptive optical systems were introduced in astronomy [Har78] to cancel-out the effect of the turbulent atmosphere. Strong efforts have been made to achieve the diffraction limit for the imaging of astronomical objects. These systems are rather large scale, as mirrors and optics in astronomy are in the range of meters. In astronomy bright stars or artifical light sources, so called sodium-layer laser guide stars, are used as probes for the aberrations of the atmosphere. The light from the star enters a wavefront sensor and the reconstructed wavefront is used to drive an actuator, usually an *adaptive mirror*. With such enhanced resolution, weak astronomical objects can be imaged. See also [fAO01] for an overview of AO applications.

In opthalmology strong efforts have been made to achieve the diffraction limit also for the imaging of the human retina. A number of diseases, like the common glaucoma, can be diagnosed by scanning the retina, e.g. with a laser-scanning confocal microscope [Mül01], but the imaging system suffers from the same aberrations as the visual acuity does. Large effort goes in the design of adaptive optical systems to cancel these aberrations out, allow diffraction limited imaging and thus allow earlier detection of ocular diseases, but also an improvement of the human vision itself seems to be possible [App00, vP02]. Ophthalmic AO systems for retinal imaging have been reported [Tür00, HCY$^+$01, FIA01] with up to 30 Hz closed-loop bandwidth. The dynamics of the human eye‘s wavefront aberrations suggest, that the power spectral density of the higher-order aberrations drops with approximately 4 dB per octave

and that most aberrations can be canceled out with a closed-loop bandwidth of 12 Hz [HAS+01]. The wavefront sensor should provide at least a bandwidth of 100 Hz to avoid undersampling and allow averaging of the wavefront data.

AO systems in opthalmology should usually be microsystems in the mm-range. Strong efforts are going in the design of compact systems for implementation in cost-effective ophthalmic diagnostic tools. This implies the need to develop small and integrated systems on a microchip basis. These systems may contain the wavefront sensor and the processor on a single chip. These sensors have to be very sensitive due to security guidelines for the application of laser energy to the human eye. The maximum applicable laser power is strongly wavelength dependent, but must not exceed 0.01 W/m in general in the red spectrum.

Lots of devices have been proposed and build as actuators for AO systems. Among them are *Membrane Mirrors* [VMP97], *Micro-Electro-Mechanical-Systems Deformable Mirrors (MEMS)* [Geh02], see Figure 1.3 for an example, and *Spatial Light Modulators (SLM)* [NV98] on the basis of liquid crystal displays. Deformable mirrors form a complete field of investigation themselves.

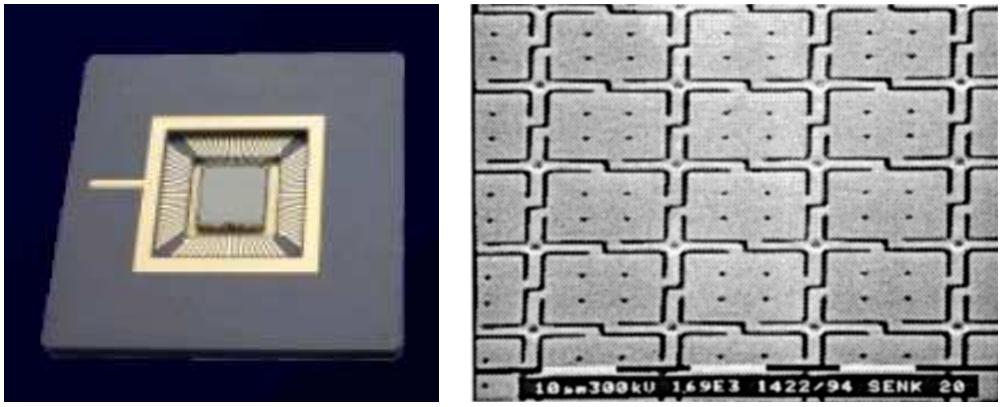

Figure 1.3: Images of a MEMS micro-machined mirror [Geh02]. Packaged device (left) and scanning-electron microscopy image (right).

1.3 Wavefront Sensors

Methods for wavefront measurements can be subdivided into the three classes interferometric, irradiance and geometrical.

Interferometric methods like the Twyman-Green interferometer are the most widely used and often feature high accuracy. An interferogram is obtained from the interference or fringe pattern of a test beam and a reference beam. The wavefront can be obtained by a thorough evaluation of the fringe pattern. The basic disadvantages of interferometric methods are the need for a coherent source of light, the necessary movable parts and the sensitivity of the setup to vibration.

Irradiance methods, like phase retrieval and the curvature sensor, rely on the measurement of the irradiance on two different planes and the reconstruction of the wavefront from these irradiance distributions. Besides computational problems associated with irradiance methods, the setup also requires movable parts.

Geometrical methods are based on the simple approach of ray optics. Diffraction is only a secondary effect in these sensors. Yet in the late 19th century Foucault has used the *knife-edge test* to test spherical surfaces. A sharp object, like a razor blade, is moved inside the focus of the spherical surface which is to be tested. If the surface is perfect, the razor will block all the light at the moment, when it crosses the focal point. Another geometrical method is *laser ray tracing*, which is a scanning technique. A single collimated laser beam scans a grid of points on the aberration plane and the spot position is recorded for each step of the scan. The wavefront can be calculated from the displacements of the spots. Laser ray tracing is not suitable for real-time applications due to the long scanning time. An interesting new sensor based on geometrical optics is the *pyramidal sensor*, see Figure 1.4 for a schematic, which is basically a four-fold knife-edge test. The beam is focused on the central vertex of a transparent pyramidal plate and thus split into four separate images, which are imaged on a screen. A plane wavefront results in four spots, with all spots having the same intensity. An aberrated wavefront results in blurred spots of different intensities, from which the local tilt of the wavefront can be calculated.

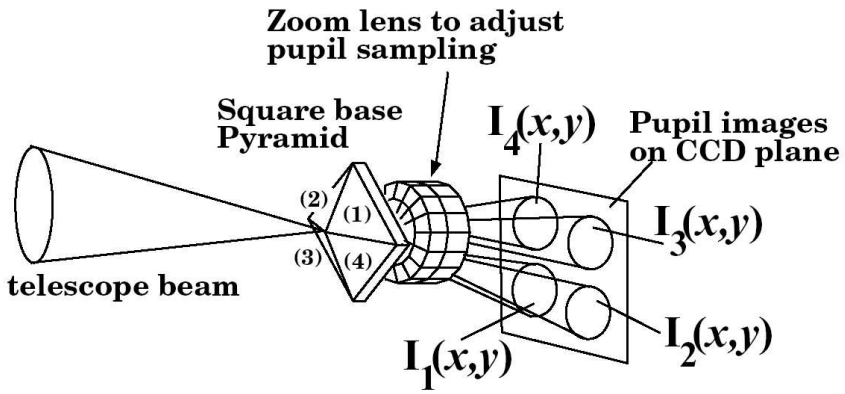

Figure 1.4: Schematic of a pyramidal wavefront sensor [ED01].

The most common wavefront sensor based on geometrical optics is the *Hartmann-Shack sensor*[1], which originates from the *Hartmann test*. Hartmann used an opaque mask in the aperture with a grid of holes for wavefront sensing. The light transmitted through the mask is imaged on a screen behind the mask. Local tilts of the wavefront result in lateral shifts of the spots on the screen, which is placed at a certain distance from the mask. The wavefront can be reconstructed from the lateral shifts of

[1] Also referred to as Shack-Hartmann sensor by other authors.

the spots.

The Hartmann-Shack sensor has been suggested by Shack [PS71] and used by Liang [LGGB94, Lia98], see Figure 1.5 for a schematic. It uses a lenslet array instead of an opaque mask. The spot displacements are detected in the focal plane of the lenslets. In the Hartmann-Shack sensor all the light of the incident beam is used for detection and nothing is lost on the opaque mask. The Hartmann-Shack sensor is also favorable, because the displacement of the small focused spots is much easier to detect, than the displacement of the large spots from the holes in the opaque mask and the dense lenslet array allows a more accurate spatial sampling of the wavefront. Of all wavefront sensing methods the Hartmann and the Hartmann-Shack sensor are the most promising for real-time applications, because it works without movable parts and with non-coherent light sources.

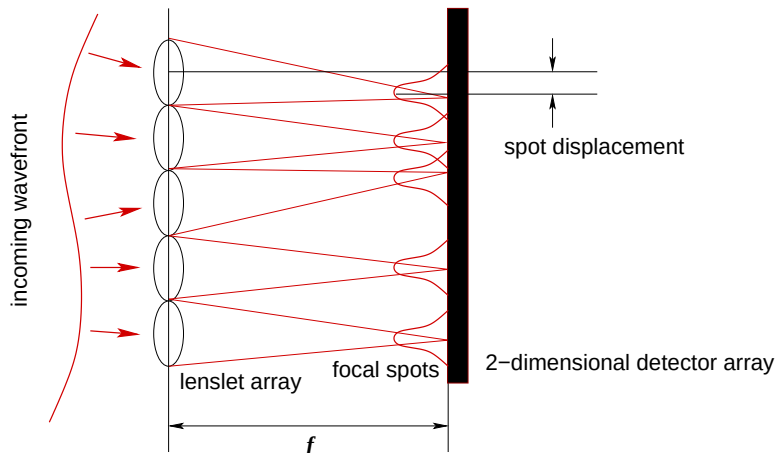

Figure 1.5: Schematic of a Hartmann-Shack wavefront sensor.

1.4 Hartmann-Shack Sensor

The first step to reconstruct a wavefront with a Hartmann-Shack sensor is to determine the displacement $(\Delta x_{i,j}, \Delta y_{i,j})$ of the spots at the grid positions $(x_{i,j}, y_{i,j})$ for all lenslets in the array. The local gradient $\nabla W(x_{i,j}, y_{i,j})$ of the wavefront is proportional to the displacement of the associated spot $(\Delta x_{i,j}, \Delta y_{i,j})$

$$\nabla W(x_{i,j}, y_{i,j}) = \frac{(\Delta x_{i,j}, \Delta y_{i,j})}{f}, \tag{1.1}$$

where f is the focal length of the lenslets. The Hartmann-Shack sensor has a limited number of subapertures and the largest resolvable spatial frequency of the wavefront is therefore limited to half the pitch of the lenslets.

The second step to reconstruct the wavefront consists of finding an appropriate description of the wavefront, which can be either zonal or modal. The *zonal approach* describes the wavefront by the values $W(x_{i,j}, y_{i,j})$ at a discrete grid of data points. Although all the measured information is present in this kind of description, the wavefront is often expanded as a superposition of spatial modes

$$W(x,y) = \sum_{i=1}^{\infty} c_i f_i(x,y), \tag{1.2}$$

where the spatial modes $f_i(x,y)$ have well known characteristics. This kind of wavefront description is called *modal approach*. A special set of spatial modes are the *Zernike polynomials*, which describe common aberration like coma and astigmatism.

1.4.1 Spot Irradiance Distribution

The achievable resolution of the spot displacement depends on a large scale on the optical quality of the lenslet array. An ideal lenslet leads to a diffraction-limited spot size, without any additional wavefront aberrations induced by the lenslet. The irradiance distribution of a spot in the focal plane can be calculated as the Fraunhofer diffraction [2] pattern of the aperture. A detailed discussion on the spots irradiance distribution has been done by Liang [Lia98]. For rectangular lenslets the irradiance distribution in the focal plane is

$$\begin{aligned} J(x,y) &= J_0 \frac{\sin^2 z_x}{z_x^2} \frac{\sin^2 z_y}{z_y^2}, \\ z_x &= \frac{2\pi D_L}{\lambda f}(x - x_0), \\ z_y &= \frac{2\pi D_L}{\lambda f}(y - y_0) \end{aligned} \tag{1.3}$$

and for circular lenslets

$$\begin{aligned} J(x,y) &= 4J_0 \frac{J_1^2(\sqrt{z_x^2 + z_y^2})}{z_x^2 + z_y^2}, \\ z_x &= \frac{2\pi D_L}{\lambda f}(x - x_0), \\ z_y &= \frac{2\pi D_L}{\lambda f}(y - y_0). \end{aligned} \tag{1.4}$$

[2] Fraunhofer diffraction predominates, when the *Fresnel number* $F = D_L^2/4\lambda f$ is small $F \ll 1$. For large Fresnel numbers $F \geq 1$ Fresnel diffraction dominates.

with the lens diameter D_L and the wavelength λ. The coordinate (x_0, y_0) is the point, where a ray intersects the focal plane according to ray optics.

The first minima in the circular case, which is the minima of the first order Bessel polynomial J_1, is called *Airy disc* and often used as a definition of the spot size. Another indicator is the *Full Width at Half Maximum (FWHM)*, the width of the irradiance distribution at half of the maximum of intensity. For a circular aperture the diameter of the Airy disc is

$$\Delta d = 1.22 \frac{f\lambda}{D_L} \tag{1.5}$$

and the FWHM is

$$\mathrm{FWHM} = 1.04 \frac{f\lambda}{D_L} \tag{1.6}$$

respectively. For a lenslet array with 400 μm lens diameter and 35 mm focal length at a wavelength of 680 nm the diameter of the Airy disc is 72.6 μm and the FWHM is 61.9 μm.

1.4.2 Zernike polynomials

The Zernike polynomials Z_n^m have been introduced by Zernike [Zer34] to describe wavefront aberrations on circular apertures. The most frequent aberrations in optical systems are described by low-order Zernike polynomials. The magnitude of each polynomial is carried by the associated Zernike coefficient C_n^m. In polar coordinates the Zernike polynomials have the form

$$Z_n^m(\rho, \theta) = \begin{cases} R_n^m(\rho)\cos(m\theta) & \text{for } m > 0 \\ R_n^{-m}(\rho)\sin(m\theta) & \text{for } m < 0 \\ R_n^m(\rho) & \text{for } m = 0 \end{cases} \tag{1.7}$$

with

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s!((n+m)/2-s)!((n-m)/2-s)!} \rho^{n-2s}. \tag{1.8}$$

The indices n and m describe the radial order and the azimuthal frequency respectively, where m has to satisfy the conditions $m \leq n$ and $(n-m)$ is an even number. Every wavefront on the unit circle can be decomposed in a set of Zernike polynomials

$$W(\rho, \theta) = \sum_{n=0}^{\infty} \sum_{\forall m} C_n^m Z_n^m(\rho, \theta). \tag{1.9}$$

See all Zernike polynomials up to 4th radial order in Table 1.1. Figure 1.6 shows a graphic visualization of low order Zernike terms. Besides the mentioned characteristics the Zernike polynomials are orthogonal on the unit circle, which means that they have zero mean and that the variances of each Zernike polynomial can be added up quadratically, while mixed terms cancel out.

Z_n^m	j	polar	monomial	name
Z_0^0	0	1	1	piston
Z_1^{-1}	1	$\rho \sin\theta$	x	tilt in x
Z_1^1	2	$\rho \cos\theta$	y	tilt in y
Z_2^{-2}	3	$\rho^2 \sin 2\theta$	$2xy$	astigmatism $\pm 45°$
Z_2^0	4	$2\rho^2 - 1$	$2x^2 + 2y^2 - 1$	defocus
Z_2^2	5	$\rho^2 \cos 2\theta$	$y^2 - x^2$	astigmatism 0°/90°
Z_3^{-3}	6	$\rho^3 \sin 3\theta$	$3xy^2 - x^3$	
Z_3^{-1}	7	$(3\rho^3 - 2\rho) \sin\theta$	$-2x + 3xy^2 + 3x^3$	coma in x
Z_3^1	8	$(3\rho^3 - 2\rho) \cos\theta$	$-2y + 3x^2y + 3y^3$	coma in y
Z_3^3	9	$\rho^3 \cos 3\theta$	$y^3 - 3x^2y$	
Z_4^{-4}	10	$\rho^4 \sin 4\theta$	$4y^3x - 4x^3y$	
Z_4^{-2}	11	$(4\rho^4 - 3\rho^2) \sin 2\theta$	$-6xy + 8y^3x + 8x^3y$	
Z_4^0	12	$6\rho^4 - 6\rho^2 + 1$	$1 - 6y^2 - 6x^2 + 6y^4$ $+12x^2y^2 + 6x^4$	3rd order spherical
Z_4^2	13	$(4\rho^4 - 3\rho^2) \cos 2\theta$	$-3y^2 + 3x^2 + 4y^4 - 4x^4$	
Z_4^4	14	$\rho^4 \cos 4\theta$	$y^4 - 6x^2y^2 + 4x^4$	

Table 1.1: Zernike polynomials up to 4th radial order.

A single indexation scheme for the Zernike coefficients is often useful, but the mapping of m and n to a single index j is ambiguous. The recommended indexation from Thibos [TASW00] uses the relation

$$j = \frac{n(n+2) + m}{2} \tag{1.10}$$

for indexation, which is also used throughout this thesis.

1.4.3 Wavefront Reconstruction

The algorithms for wavefront reconstruction from spot displacements are generally divided into *zonal* and *modal*. Zonal wavefront reconstruction relies on local operations, e.g. numerical integration of local wavefront gradients, and thus gives a clear picture of the wavefront, even when it contains steep local variations.

Modal reconstructors fit the measured spot displacements to a set of known aberrations to extract more global features of the wavefront. This means in general, that a matrix equation of the form

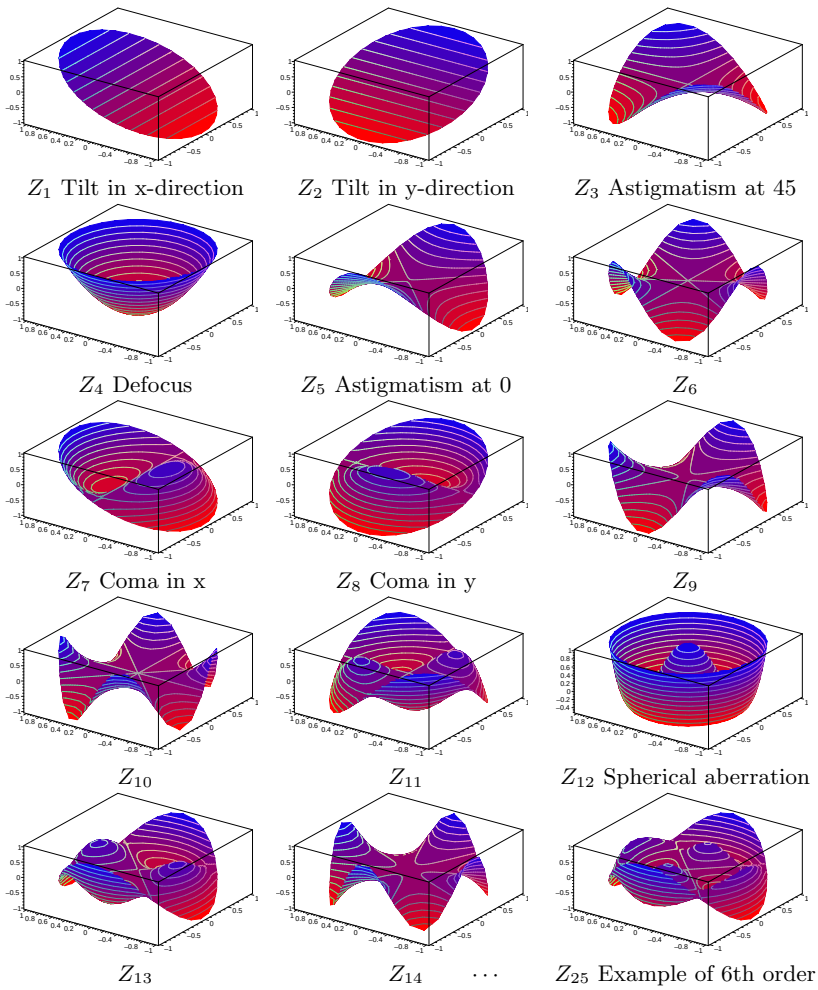

Figure 1.6: Zernike polynomials up to 4th order [BCM02].

$$\mathbf{s} = \mathbf{B}\mathbf{c} \tag{1.11}$$

has to be solved, where $\mathbf{s}$ is a column vector, that contains the spot displacements divided by the focal length f, $\mathbf{c}$ is a vector of spatial mode coefficients and $\mathbf{B}$ is the reconstruction matrix, which relates the two vectors with each other. An explicit form of this equation for N lenslets and M Zernike polynomials is

$$f^{-1}\begin{pmatrix} \Delta x_0 \\ \Delta y_0 \\ \vdots \\ \Delta x_k \\ \Delta y_k \\ \vdots \\ \Delta x_{N-1} \\ \Delta y_{N-1} \end{pmatrix} = \begin{pmatrix} \frac{dZ_0}{dx}\big|_0 & \frac{dZ_1}{dx}\big|_0 & \cdots & \frac{dZ_{M-1}}{dx}\big|_0 \\ \frac{dZ_0}{dy}\big|_0 & \frac{dZ_1}{dy}\big|_0 & \cdots & \frac{dZ_{M-1}}{dy}\big|_0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dZ_0}{dx}\big|_i & \frac{dZ_1}{dx}\big|_i & \cdots & \frac{dZ_{M-1}}{dx}\big|_i \\ \frac{dZ_0}{dy}\big|_i & \frac{dZ_1}{dy}\big|_i & \cdots & \frac{dZ_{M-1}}{dy}\big|_i \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dZ_0}{dx}\big|_{N-1} & \frac{dZ_1}{dx}\big|_{N-1} & \cdots & \frac{dZ_{M-1}}{dx}\big|_{N-1} \\ \frac{dZ_0}{dy}\big|_{N-1} & \frac{dZ_1}{dy}\big|_{N-1} & \cdots & \frac{dZ_{M-1}}{dy}\big|_{N-1} \end{pmatrix} \begin{pmatrix} C_0 \\ C_1 \\ \vdots \\ C_i \\ \vdots \\ C_{M-2} \\ C_{M-1} \end{pmatrix}. \tag{1.12}$$

This matrix equation only has an exact solution for $M = 2N$. In this case the solution can be found from

$$\mathbf{c} = \mathbf{B}^{-1}\mathbf{s}, \tag{1.13}$$

provided that $\mathbf{B}$ is not singular.

If the vector $\mathbf{s}$ has a lower dimension than vector $\mathbf{c}$, the system is underdetermined and the system of equations has either no solution or a number of solution vectors, which can be found, e.g. by *Singular Value Decomposition (SVD)*.

In the third case the vector $\mathbf{s}$ has a larger dimension than vector $\mathbf{c}$ and the system of equations is overdetermined. Although no exact solution exists in general, the vector at the left side of equation 1.11 can fit in a least-squares sense the vector at the right side. This is equivalent to solving equation

$$(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{s} = \mathbf{c}, \tag{1.14}$$

where $(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ is the pseudo-inverse of the reconstruction matrix. If $(\mathbf{B}^T\mathbf{B})$ is singular or almost singular, other methods like SVD have to be applied to find a solution instead.

Another least-squares method has been suggested by Malacara [Mal92], which derives the Zernike coefficients through the calculation of a second set of orthogonal polynomials. A short description of Malacara‘s solution can be found in the Appendix A.

An interesting intermediate approach between a pure zonal and a pure modal reconstruction contains only the most important aberration terms, defocus and astigmatism, on every subset of 2×2 spots repetitively. A power map of the aperture is

calculated containing the power of the defocus $C_Z = C_2^0$, the power of the overall astigmatism

$$C_A = \sqrt{C_2^2 + C_{-2}^{-2}} \tag{1.15}$$

and the angle of astigmatism

$$\alpha_A = \frac{1}{2} \arctan \frac{C_{-2}^{-2}}{C_2^2}. \tag{1.16}$$

These aberration terms are calculated for each subset of 2×2 spots on the aperture, by explicitly solving equation 1.14. In this case the reconstruction matrix is

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & -D_L & -2D_L & D_L \\ 0 & 1 & -D_L & -2D_L & -D_L \\ 1 & 0 & -D_L & 2D_L & -D_L \\ 0 & 1 & D_L & -2D_L & -D_L \\ 1 & 0 & D_L & -2D_L & D_L \\ 0 & 1 & -D_L & 2D_L & D_L \\ 1 & 0 & D_L & 2D_L & -D_L \\ 0 & 1 & D_L & 2D_L & D_L \end{pmatrix}. \tag{1.17}$$

Here, D_L has to be normalized with the unit aperture and is therefore dimensionless. Calculation of the pseudo-inverse of matrix $\mathbf{B}$, multiplication with vector $\mathbf{s}$ and comparison with the coefficients of vector $\mathbf{c}$ yields the Zernike coefficients

$$\begin{aligned}
C_1^{-1} &= \tfrac{1}{4f}(\Delta x_{0,0} + \Delta x_{0,1} + \Delta x_{1,0} + \Delta x_{1,1}) \\
C_1^{1} &= \tfrac{1}{4f}(\Delta y_{0,0} + \Delta y_{0,1} + \Delta y_{1,0} + \Delta y_{1,1}) \\
C_{-2}^{-2} &= \tfrac{1}{8fD_L}(-\Delta x_{0,0} - \Delta y_{0,0} - \Delta x_{1,0} + \Delta y_{1,0} + \Delta x_{0,1} - \Delta y_{0,1} + \Delta x_{1,1} + \Delta y_{1,1}) \\
C_2^{0} &= \tfrac{1}{16fD_L}(-\Delta x_{0,0} - \Delta y_{0,0} + \Delta x_{1,0} - \Delta y_{1,0} - \Delta x_{0,1} - \Delta y_{0,1} + \Delta x_{1,1} + \Delta y_{1,1}) \\
C_2^{2} &= \tfrac{1}{8fD_L}(\Delta x_{0,0} - \Delta y_{0,0} - \Delta x_{1,0} - \Delta y_{1,0} + \Delta x_{0,1} + \Delta y_{0,1} + \Delta x_{1,1} + \Delta y_{1,1})
\end{aligned} \tag{1.18}$$

Local deviations from the global defocus and astigmatism values, i.e. at the border of the aperture, are easier to detect from the power map. Figure 1.7 shows an example.

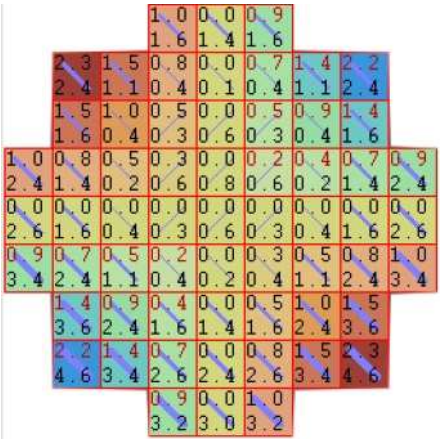

Figure 1.7: Exemplary power map with defocus and astigmatism values, and vectors representing the angle of astigmatism [Sch02].

2 CMOS Process Technology

This chapter introduces the CMOS process technology and the most important building blocks for CMOS-based vision chips, which are the MOSFET and the junction photodiode, after briefly introducing other available process technologies. Operating the MOSFET in weak inversion and as an active resistor is described subsequently. The effects of mismatching in the CMOS process technology are explained and active and passive measures against them. The last part addresses the quantum efficiency of junction photodiodes and their noise characteristics.

2.1 Introduction

A number of standard industrial process technologies include the possibility to implement optoelectronic devices on chip; the most important ones being *Charged Coupled Device (CCD)*, *Complementary Metal Oxide Semiconductor (CMOS)* and *Bipolar CMOS (BiCMOS)*. BiCMOS technology is similar to CMOS process technology, but offers additional bipolar devices. CCD technology has been optimized since long for imaging purposes, but has the disadvantage of not offering optimized devices for image processing on the same *die* [1]. Their use for on-chip signal processing is therefore limited.

The CMOS process technology has been originally developed for *the Very Large Scale Integration (VLSI)* of digital circuits. These digital chips consist of simple building blocks, e.g. logic gates, taken from *standard cell libraries* provided by the silicon foundry. When CMOS process technology grew mature, analog applications became popular too. The availability of photodetectors in standard CMOS process technologies soon led to the first integration of photodetectors and image processors on the same die, and for these chips the term *vision chip* was formed. A disadvantage of CMOS imagers with respect to CCD imagers is the missing characterization of photodetectors by the silicon foundry and the poor matching of CMOS devices.

[1] The *die* is the piece of silicon, which contains all structures on its surface.

CMOS process technology is based on either a p-type or n-type silicon substrate (also called *bulk* or *body*). The n-type substrate is doped with donator atoms like phosphorus or arsenic, which provide an extra electron for conduction. Acceptor atoms like boron are p-type dopants and provide a conducting hole. The CMOS process technologies offers an implanted well of the complementary charge carrier type and thus allows the implementation of the complementary device type. All devices, like transistors and resistors, are formed by consecutive process steps on top of the substrate, including ion implantation, oxidation, etching and nitride, polysilicon and metal deposition [GHLM00].

2.2 The MOSFET

The basic electronic device in CMOS technology is the *Metal Oxide Semiconductor Field Effect Transistor (MOSFET)*. In contrast to bipolar transistors, where additional charge carriers have to be introduced through the *base*, the field effect between the *gate* electrode and the substrate modulates the current here.

The enhancement n*-channel MOS (NMOS)* transistor is physically formed in a p-type substrate by two separated $n+$ regions, which serve as *source* and *drain* of the transistor. A polysilicon layer is deposited on a thin gate oxide between the two contacts of the transistor to separate the gate from the substrate. The complementary p-channel transistor *(PMOS)* is formed in the same way, but in a slightly doped n-well region instead, with two $p+$ regions forming the source and the gate. Figure 2.1 shows a cross section of the two transistor types.

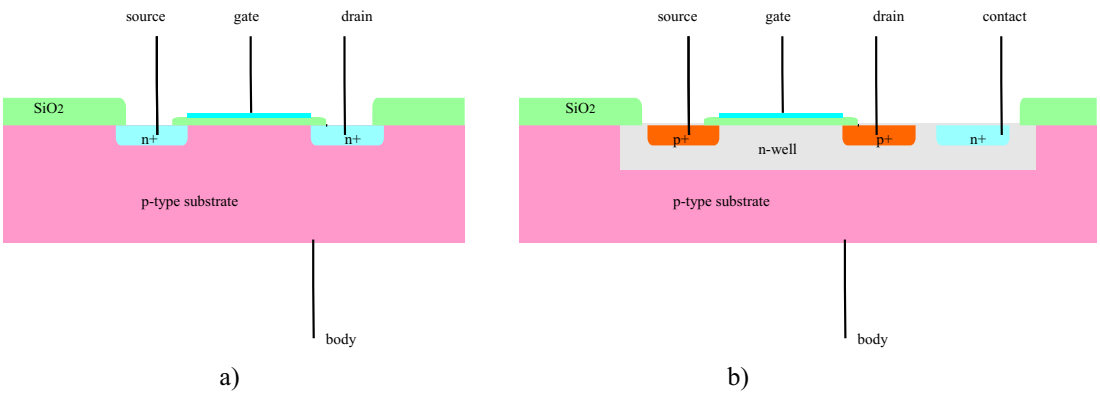

Figure 2.1: Typical NMOSFET (a) and PMOSFET (b) structures in an n-well process technology.

The drain current I_D of the MOSFET is a function of both the *gate-source voltage* V_{GS} and the *drain-source voltage* V_{DS}. When a voltage between the gate and the bulk of the transistor is applied, the electric field between the gate and the substrate induces a conducting channel in the substrate under the gate. The *gate-bulk voltage*

V_{GB} modifies the conductance of the region under the gate and modulates the current flow. The bulk of the NMOSFET is normally connected to ground and has no AC component. In this case the gate-bulk voltage is equal to the gate-source voltage V_{GS} [2].

A conducting channel forms for gate-source voltages above a process-dependent *threshold voltage* V_t. The MOSFET works in the region of *strong inversion*. For smaller gate-source voltages a depletion region only exists at the surface of the substrate. This region of operation is called *subthreshold* or *weak inversion*.

The strong-inversion region is further divided into the *linear region*, when $V_{DS} < (V_{GS} - V_t)$, and the *saturation region*, when $V_{DS} > (V_{GS} - V_t)$. In this region the conducting channel is no longer connected to the drain and the drain current therefore almost independent of the drain-source voltage. Figure 2.2 shows the I_D/V_{DS} characteristics of the MOSFET in strong inversion.

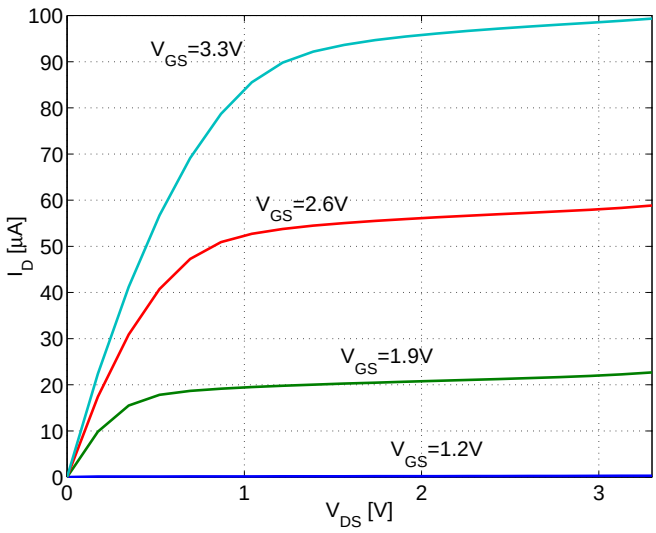

Figure 2.2: NMOS drain current example with drain-source voltage for different gate-source voltages in the AMS 0.35 μm process technology with $V_{dd} = 3.3$ V. The linear and the saturation region are clearly distinguishable. Also the channel-length modulation is visible through the slope in the saturation region.

The remaining dependence of the drain current from the drain-source voltage in the saturation region comes from the *channel-length modulation*. It is present in all regions of operation due to the modulation of the channel length with the drain-source voltage. The channel-length modulation is characterized by the *channel-length-modulation factor* λ and normally neglected in the linear and subthreshold region.

The drain current is modeled in different operation regions for an NMOSFET as follows [GHLM00]. In the saturation region it is

[2] The modulation of the drain current through the *source-bulk voltage* V_{SB} is called *body-effect*.

$$I_D = \frac{k'}{2}\frac{W}{L}(V_{GS} - V_t)^2(1 + \lambda V_{DS}) \tag{2.1}$$

and in the linear region it is

$$I_D = \frac{k'}{2}\frac{W}{L}\left(2(V_{GS} - V_t)V_{DS} - V_{DS}^2\right). \tag{2.2}$$

The parameter $k' = \mu C_{ox}$ is process technology dependent through the channel mobility μ and the gate-oxide capacitance C_{ox}. W/L is the geometrical width-to-length ratio of the MOSFET. In the subthreshold region the drain current can be modeled as

$$I_D = \frac{W}{L} I_{D0} e^{\frac{V_{GS}}{nV_T}} \left(e^{\frac{V_{SB}}{V_T}} - e^{-\frac{V_{DS}}{V_T}}\right)(1 + \lambda V_{DS}), \tag{2.3}$$

where $V_T = kT/q$ is the *thermal voltage* (25 mV at room temperature) and n is the *subthreshold parameter*

$$n = 1 + \frac{C_{js}}{C_{ox}} \tag{2.4}$$

with the *depletion-region capacitance* C_{js} [3]. The current I_{D0} is modeled [GAS90]

$$I_{D0} = \mu C_{ox} \frac{2(nV_T)^2}{e^2}. \tag{2.5}$$

The body effect in Equation 2.3 is often neglected and $e^{\frac{V_{SB}}{V_T}}$ replaced by 1. Equations 2.1, 2.2 and 2.3 describe the operating point of the MOSFET over the complete range of drain-source and gate-source voltages. The subthreshold region is of special interest for the design of CMOS imagers and processors, because the photocurrents are often very small and the MOSFETs work in the subthreshold region then.

2.2.1 The MOSFET as an Active Resistor

Large resistances on a small area are available in the CMOS process technology through active MOSFET resistors. The disadvantage of this kind of active resistor is the non-linearity of the transconductance $G = R^{-1}$ due to the dependence of the resistance from the drain-source voltage. The resistance at the operating point is calculated by differentiation of either Equation 2.1, 2.2 or 2.3

$$R = \left(\frac{\partial I_D}{\partial V_{DS}}\right)^{-1}, \tag{2.6}$$

for the different regions of operation.

[3] The parameters have been derived from simulations with the Spectre circuit simulator for the AMS 0.35 μm process technology to $n = 1.37$ and $I_{D0} = 150$ fA .

Neglecting the channel-length modulation, the resistance in the saturation region is infinite, which makes the MOSFET suitable for current sources. In the subthreshold region the resistance from Equation 2.3 becomes

$$R = \frac{L}{W} \frac{V_T}{I_{D0} e^{\frac{V_{GS}}{nV_T}} e^{-\frac{V_{DS}}{V_T}}}, \tag{2.7}$$

when the channel-length modulation and the body-effect are neglected.

2.2.2 Matching of MOS Transistors

Industrial standard CMOS process technologies are normally optimized for digital circuits, where the digital signals have an inherently large *Signal-to-Noise Ratio (SNR)*. Analog circuits instead have to rely to a larger extend on the spatial homogeneity of process parameters, which otherwise result in *mismatching* of devices with the same layout, but at different locations on the die. Various steps in the photolithographic process lead to inhomogeneities, i.e. the limited imaging quality in photolithography results in variations of the geometrical dimensions of the devices, process parameter variations lead to spatially inhomogeneous sheet resistances and threshold voltages.

The variation of the drain current of the MOS transistor is regarded to quantify the drain current variations $\sigma(I_D)$. Although several parameters in Equations 2.1,2.2 and 2.3 suffer from inhomogeneities, like the channel mobility μ and the gate-oxide capacitance C_{ox}, the predominant mismatching sources are geometrical mismatch of the W/L ratios, $\sigma(W/L)$ and the mismatch in the threshold voltage $\sigma(V_t)$ [Loo99]. The relative drain current mismatch is the quadratic sum of these two contributions and can be calculated for the MOS transistor in strong inversion

$$\frac{\sigma(I_D)}{I_D} = \sqrt{\frac{\sigma^2(W/L)}{(W/L)^2} + 4\frac{\sigma^2(V_t)}{(V_{GS} - V_t)^2}} \tag{2.8}$$

and in weak inversion

$$\frac{\sigma(I_D)}{I_D} = \sqrt{\frac{\sigma^2(W/L)}{(W/L)^2} + \frac{\sigma^2(V_t)}{(nV_T)^2}}. \tag{2.9}$$

Because the threshold voltage variance is only divided by the small term $(nV_T)^2$, rather small threshold voltage variations can lead to large drain current variations in weak inversion. Mismatching has been measured by Droste [Dro99] for the AMS 0.6 μm process technology in weak inversion and turned out to be one of the major sources for position uncertainty in the implemented position sensitive detectors.

A number of passive and active measures to reduce the effect of mismatch have been proposed. The passive measures refer to layout techniques, while active measures refer to additional signal processing circuitry, e.g. the *correlated double sampling* in integrating CMOS image sensors [KDS$^+$00]. A sensitivity analysis of different circuits

with the same functionality can also be seen as a passive measure against mismatching effects. A good example is the analysis of the mismatching characteristics of different types of current mirrors, see Moini [Moi00].

Almost all layout techniques go to the expense of area and wiring. They include the *common-centroid technique*, the placement of *dummy devices* and *orientation matching* in sensitive analog circuits [BLB97]. The common-centroid technique splits devices into two separate parts and places them around the same common centroid. Linear gradients of the sheet resistance are thus canceled out.

Edge effects in multiple adjacent copies of the same layout can be avoided by surrounding them with dummy devices. These dummy elements are not connected, but serve as a homogeneous environment for the devices lying inside. Another general rule to reduce mismatch, is the placement of sensitive devices as close together as possible and to orient them, such that the current flow has the same direction.

In the circuits implemented in this thesis only the mismatching of the MOSFETs is of importance, as passive resistors have not been used. The geometrical mismatch of the photodiodes is neglected due to their large area.

2.3 Passive Devices

Passive resistors are formed from polysilicon layers, diffusion layers or from n-wells; the resistance of the device depends on the *sheet resistance* $R_\square$ of the layer and the geometry, according to

$$R = R_\square \frac{L}{W}. \tag{2.10}$$

Sheet resistances are in general rather small ($< 10\,k\Omega/\square$) [4] and matching characteristics are poor. Large resistors therefore have to occupy a large area. Special layout techniques like the common-centroid technique and the placement of dummy devices adjacent to sensitive devices have to be applied, when matching is the crucial point for the circuitry. Active MOSFET resistors are preferred, when large resistances have to be provided.

Capacitors can be formed from two polysilicon layers or parasitic capacitances are used instead, e.g. the parasitic capacitance between the gate and the drain of a MOSFET.

2.4 Silicon Photodetectors

There are three fundamental interactions between photons and matter, *Compton scattering*, *pair production* and *photoelectric absorption*. In the energy range of visible and near-infrared light, silicon shows a large photoelectric absorption, which can be

[4] The symbol $\square$ refers to the smallest available square area in a given process technology.

used for photodetection. A photon is absorbed in the bulk semiconductor by producing an electron-hole pair, when the photon energy hc/λ is larger than the band gap. The thus separated charge carriers either recombine after a certain drift time or they are driven away from each other in the presence of an electric field.

The total incident *photon flux* J_0 is partially transmitted to the bulk and partially reflected at the silicon surface, depending on the *spectral reflectivity* $\sigma(\lambda)$. An anti-reflective layer of SiO_2 can increase or decrease the transmittance to the bulk, depending on the wavelength and the thickness of the layer, see Figure 2.3. Such a layer is always present as a *depassivation* of the die.

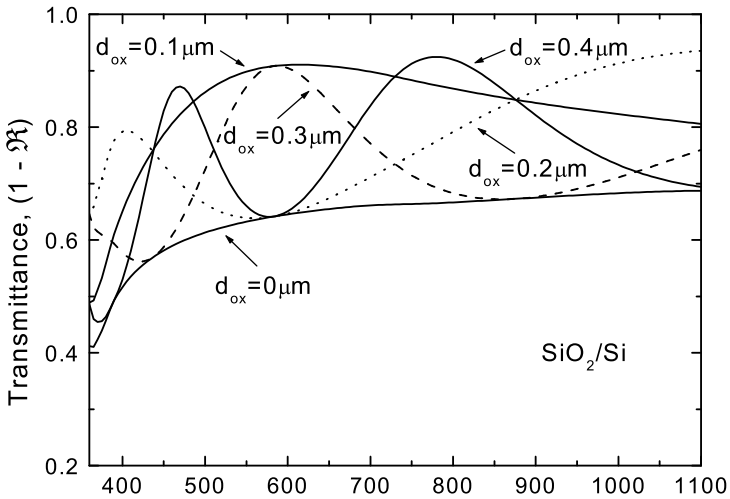

Figure 2.3: Transmittance from air to silicon with wavelength (in nm) for different SiO_2 layer thicknesses [dLM02]. Note, that in the used process technology the passivation layer has a thickness of 0.9 μm [AMSb].

The transmitted photon flux

$$J_T = J_0(1 - \sigma(\lambda)) \tag{2.11}$$

gets absorbed according to Lambert's law

$$J(z) = J_T e^{-\alpha(\lambda)z} \tag{2.12}$$

in a homogeneous material with *absorption coefficient* $\alpha(\lambda)$. Figure 2.4 shows the wavelength dependent *penetration depth* $1/\alpha(\lambda)$ of silicon.

In photodetectors the photogenerated charge carriers are separated by an electric field, that originates from a *depletion zone* and/or an external applied voltage, and that prevents them from recombining. Photodetectors are basically characterized by the *spectral quantum efficiency* $\eta(\lambda)$

$$\eta(\lambda) = \frac{I_p/e}{J_T/E} = \frac{I_p hc}{J_T \lambda e}, \tag{2.13}$$

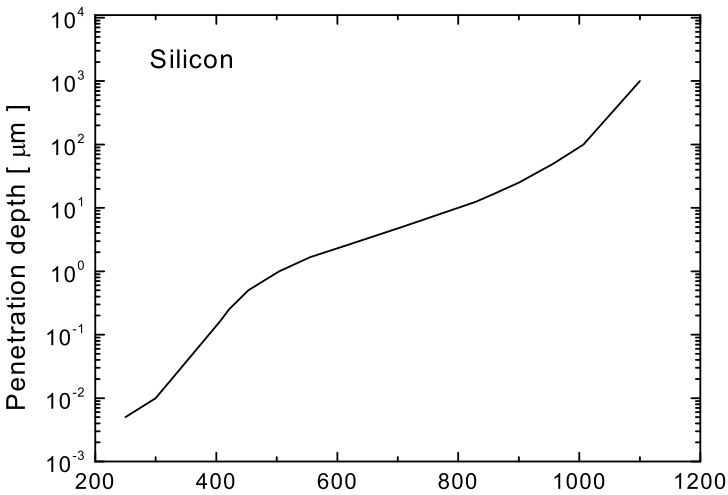

Figure 2.4: Penetration depth $1/\alpha(\lambda)$ in silicon with wavelength (in nm) [dLM02].

which describes the photocurrent I_p, generated by the transmitted photon flux J_T with photon energy E, see Chapter 4.6.1 for measured quantum efficiencies.

A number of silicon photodetectors can be formed at metal-semiconductor and semiconductor-semiconductor junctions, which will be briefly introduced here. The *Schottky photodiode* is a metal-semiconductor photodiode, suitable only for small wavelengths. All other photodetectors are based on semiconductor-semiconductor junctions. The *junction photodiode* is the most common device, because standard CMOS process technologies offer at least three different types. Other photodetectors, like the p-i-n *photodiode* and the *avalanche photodiode*, require an additional layer of lightly doped silicon between the p and the n regions and thus an additional process step. In avalanche photodiodes, the photogenerated electrons and/or holes are accelerated by a large reverse bias voltage and generate secondary electron-hole pairs. See Figure 2.5 for schematic drawings of the most common photodetectors.

Photodetectors with controllable amplification are the phototransistors, also available in standard CMOS process technologies, which are either vertical or horizontal. The internal amplification of the photocurrent in these devices results in effective quantum efficiencies larger than one. Despite this advantage phototransistors are rarely used photodetectors, due to their slow transient response, a nonlinear response and disadvantagous noise characteristics.

2.4.1 Junction Photodiodes

The spectral quantum efficiency of the junction photodiode is determined by the vertical doping profiles. At the junction of the n- and the p-type semiconductor, a drift current of electrons from the n-type semiconductor to the p-type and a drift current of

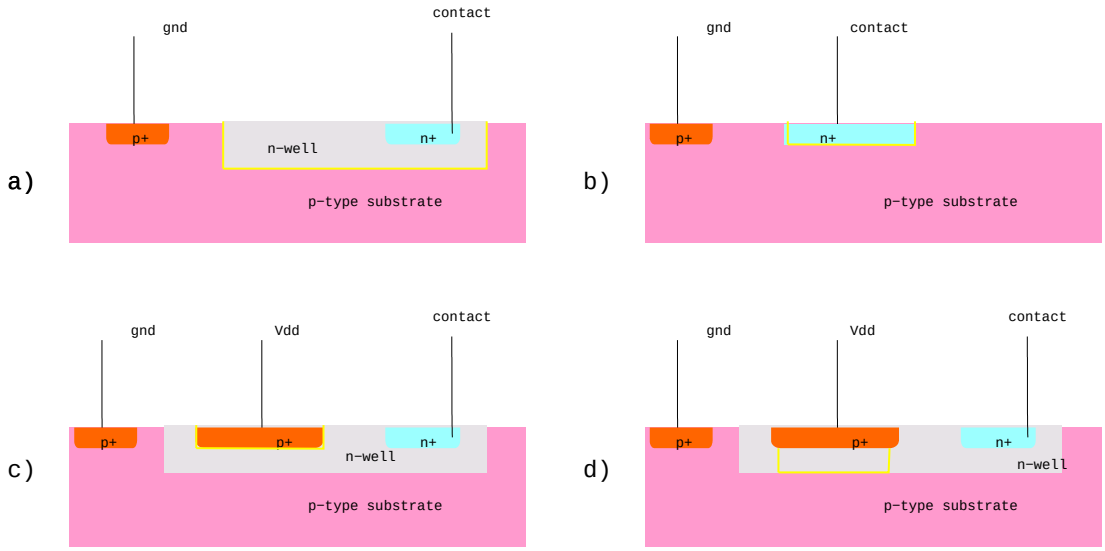

Figure 2.5: Photodetectors available in most standard CMOS process technologies; junction photodiodes n-well/substrate (a), $n+$/substrate (b), $p+/n$-well (c) and vertical phototransistor (d). The yellow region is the area of photogeneration.

holes from the acceptors in the p-type semiconductor in the opposite direction occurs. An electric field therefore forms in the depletion region due to the separated charges.

As a general rule, the broader the junction and the closer to the surface the junction is, the larger is the quantum efficiency of the photodiode.

The *junction width*

$$W_J = \sqrt{\frac{2\epsilon_{\text{Si}}}{e}\frac{N_A + N_D}{N_A N_D}(V_D - V)} \tag{2.14}$$

is a function of the applied *reverse bias voltage* V. The *built-in junction potential* is

$$V_D = \frac{kT}{e}\ln\frac{N_A N_D}{n_i^2} \tag{2.15}$$

with the concentration of donators and acceptors N_A and N_D, and the intrinsic charge carrier density n_i. The reverse bias voltage alters the spectral quantum efficiency through the modulation of the junction width.

The junction photodiode can be modeled as a current source with photocurrent I_p and several parasitics and noise contributions, see the equivalent circuit in Figure 2.6. Besides the photocurrent I_p, the *dark current* [5] I_s and noise sources i_n and i_r contribute

[5] The reverse leakage current is 0.03 fA/μm^2 for large n-well diodes in the AMS 0.35 μm process technology [AMSb].

to the overall current. The dark current

$$I_s = qWLn_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right) \tag{2.16}$$

is caused by the thermal generation of electron-hole pairs, where L_p and L_n are the diffusion lengths of the minority carriers and D_p and D_n are the diffusivity. The parallel resistance R_j is usually large (> 10 MΩ) and the series resistance R_s rather small (< 100 Ω). R_L is the external load resistance.

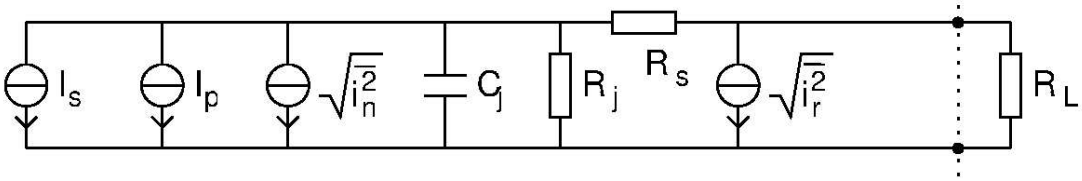

Figure 2.6: Equivalent circuit diagram of a junction photodiode.

The parasitic capacitance [AMSb]

$$C_j = \frac{WLc_j}{1 + \frac{V^{M_j}}{V_D}} \tag{2.17}$$

associated with the junction, can be reduced by applying a reverse bias voltage V and thus enlarge the width of the junction. Process dependent parameters are the (unbiased) area junction capacitance c_j, the area junction grading coefficient M_j and the junction potential V_D. Figure 2.7 shows an example of the parasitic capacitance with reverse bias voltage.

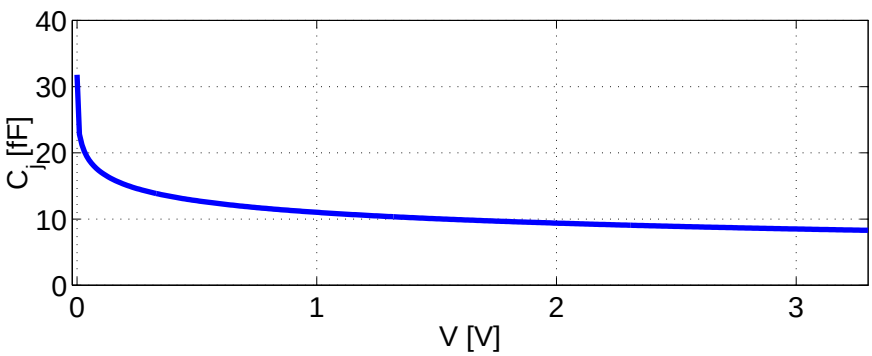

Figure 2.7: Parasitic capacitance of a 17 × 17 μm^2 n-well photodiode with applied reverse bias voltage in the AMS 0.35 μm process technology.

2.4.2 Photodiode Noise Sources

Noise in photodetectors has its origin in a variety of physical generation mechanisms. Major noise sources are *photon shot noise*, *flicker noise*, *dark-current shot noise* and *thermal noise*.

The contributions of photon shot noise and dark-current shot noise are

$$\overline{i_n^2} = 2q(I_p + I_s)\Delta f, \tag{2.18}$$

with bandwidth Δf. Shot noise has its origin in statistical fluctuations of the number of generated electron-hole pairs. Thermal noise current

$$\overline{i_r^2} = 4kT\frac{1}{R}\Delta f \tag{2.19}$$

arises from the random movement of charge carriers in resistive material. Therefore the parallel and series resistors R_j and R_S contribute to this noise, but also the load resistance R_L affects the noise. Flicker noise is supposed to have its origin in mobility fluctuations of the charge carriers; its power spectrum is inversely proportional to the frequency. The overall noise current $\sqrt{\overline{i^2}}$ is the quadratic sum of all noise contributions, because the generation mechanisms are independent from each other and therefore uncorrelated.

It is obvious from Equations 2.18 and 2.19 that noise can be decreased by decreasing the bandwidth within the specifications and by increasing the parallel resistance R_j. Thermal noise can also be reduced by decreasing the temperature, when cooling of the device is posible.

3 Position Sensitive Detectors

Different mathematical concepts for the estimation of the position of an object are presented in this chapter . The second part describes physical realizations of position sensitive detectors and their properties, with special attention to the Winner-Take-All circuit and the transient behavior. A new topology of this circuit for spot position measurements, the resistive-ring network of Winner-Take-All circuits is introduced. The last part is dedicated to the effects of noise and mismatching on the position uncertainty.

3.1 Introduction

Image processing is often performed on a pure software basis, with the disadvantage of a large processing load and a slow frame rate. If an optimized processing on a dedicated hardware basis is used instead, frame rates which are orders of magnitude larger, can be achieved. This hardware can be based for example on *Field Programmable Gate Arrays (FPGAs)* or microprocessors. The largest frame rates though are available through dedicated integrated circuits, also called *Application Specific Integrated Circuits (ASICs)*. If these ASICs include photodetectors and processors on the same chip, they are named *vision chips*.

Image processing in vision chips often relies on biological signal processing models, like the ones in insect vision [Hor75, HS91], i.e. simple building blocks are used and device characteristics are exploited to solve a specific task with minimum space and energy consumption, and maximum speed of operation. This often implies a high grade of parallelism. A large number of vision chips have been designed for spatial and temporal processing from the late 80's on, e.g. silicon retinas [AB94, BNDZ93, BZD93, FNQS95, Mea89], orientation detection chips [Sta91], centroid computation chips [Dew92], motion detection chips [Lyo81] and many more. An extensive overview over present vision chip concepts has been done by Moini [Moi00].

A special kind of image processor is the *Position Sensitive Detector (PSD)*, which generates an analog or digital signal, which is a function of the position of an object in one ore more dimensions. In a Hartmann-Shack sensor the objects to be detected

are spots from a lenslet array and the two-dimensional position of the spots has to be measured. As the focal spot from a lenslet array is an extended object, the position of the spot is defined as the location, where a ray would intersect the focal plane according to the laws of ray optics.

3.2 Spatial Estimators of Location

Closely related to the position sensitive detector is the mathematical concept of an *Estimator of Location*, although the term location refers to an estimation of a desired parameter $\hat{x}$ in a discrete data set J_i or on a continuous function $J(x)$, which is not necessarily a physical location. The estimator $\mathfrak{F}$ itself can be a mathematical function or an algorithm operating on the continuous function or the discrete data set

$$\hat{x} = \mathfrak{F}(J_i). \tag{3.1}$$

The *bias* is the difference between the estimated parameter and the *expectation value* $\langle x \rangle$,

$$\mathrm{bias}(x) = \langle x \rangle - \hat{x}. \tag{3.2}$$

An estimator of location is *unbiased*, when the expected value of the parameter equals the estimated value after averaging an infinite number of measurements. The concept of *robustness* applies to a specific estimator, when the estimated parameter is not affected by a certain number of outlying values in the data set [PV90]. The most common estimator of location, the arithmetic mean, is not a robust estimator of location in this sense, because every value in the data has an influence on the estimated parameter. Although the concept of location estimation is normally applied to random variables, it also offers useful concepts for estimation of location, when the data is not a random variable, but a spatially sampled physical quantity.

In spatial estimation of location the most simple estimator is the *maximum*. The spatial estimator of location related to the arithmetic mean is the *centroid*. The *median* is also applicable for spatial estimates. In contrast to random variables it is useful to define the spatial position estimators first for continuous functions $J(x, y)$ and then for discrete spatially sampled data sets $J(x_i, y_j)$.

Estimation on Continuous Functions

The following definitions are for the estimation of a position on a continuous function. The *maximum* of a two-dimensional function $J(x, y)$ is the location $(x_{\max}, y_{\max})$ of the largest global value $J_0 = J(x_{\max}, y_{\max})$ with

$$x_{\max} = \arg\max_x (J(x, y)), \tag{3.3}$$

$$y_{\max} = \arg\max_y (J(x, y)). \tag{3.4}$$

For the definition of the *centroid*, the (n, m)-th moment of a two-dimensional function has to be defined first

$$J_{x^n,y^m} = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} x^n y^m J(x,y) dx dy. \tag{3.5}$$

The 0th moment thus is

$$J_{0,0} = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} J(x,y) dx dy \tag{3.6}$$

and the 1st moments are

$$J_{x,0} = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} x J(x,y) dx dy, \tag{3.7}$$

$$J_{0,y} = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} y J(x,y) dx dy. \tag{3.8}$$

The centroid, also often named *center of gravity*, is given by the normalized 1st moments

$$\overline{x} = J_{x,0}/J_{0,0} \tag{3.9}$$

$$\overline{y} = J_{0,y}/J_{0,0}. \tag{3.10}$$

Another estimator of location often used on random variables, the *median*, is the 50% quantile [1] of the random variable. When the concept of quantiles is applied to the estimation of location on a continuous function $J(x, y)$, instead of a random variable, the 50% quantile $H_{0.5}$ can be derived from the normalized integrals

$$F(X) = \int_{x=-\infty}^{X} \int_{y=-\infty}^{\infty} J(x,y) dx dy / J_{0,0}, \tag{3.11}$$

$$G(Y) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{Y} J(x,y) dx dy / J_{0,0}. \tag{3.12}$$

The median $(x_{\mathrm{med}}, y_{\mathrm{med}})$ is the point that satisfies the relations

$$F(x_{\mathrm{med}}) = \frac{1}{2}, \tag{3.13}$$

$$G(y_{\mathrm{med}}) = \frac{1}{2} \tag{3.14}$$

[1] The *q-quantile* of a random variable X is the value x, such that the probability density function is $H(X \leq x) = q$.

then. Centroid and median can also be applied to the function after performing a *threshold operation*

$$T(x,y) = \begin{cases} 1: & J(x,y) \geq J_T \\ 0: & \text{else}, \end{cases} \tag{3.15}$$

for a given threshold value J_T, which should be in the range $0 < J_T < J_0$ to get a non-trivial thresholded function. For symmetric functions $J(x,y)$, like the intensity distribution of a focal spot according to Equation 1.3 or 1.4, the centroid and the median of the intensity distribution are the same, if applied either to $J(x,y)$ or $T(x,y)$. The threshold value can be chosen such, that it equals the value at the FWHM of the intensity distribution

$$J_T = \frac{J_0}{2}. \tag{3.16}$$

In this case the thresholded function changes its value at $x = x_0 - \mathrm{FWHM}/2$ and $x = x_0 + \mathrm{FWHM}/2$. The same applies for the y-coordinate.

Maximum, centroid and median are unbiased estimators of location for intensity distributions from rectangular or circular lenslets, i.e. they can be used to estimate the position of Hartmann-Shack spots $(\hat{x}, \hat{y}) = (x_0, y_0)$. This can be shown by substituting $J(x,y)$ with the irradiance distributions from Equations 1.3 or 1.4. This is no longer true for all mentioned estimators of location, when the irradiance distribution is asymmetric. If the lenslets themselves introduce a large astigmatism, the maximum is biased and no longer equal to the position of the spot $(x_{\max}, y_{\max}) \neq (x_0, y_0)$. The centroid and median though still indicate the position of the spot and are thus better suited for the position estimate.

Estimation on a Discrete Data Set

Spatial sampling at $N \times M$ discrete data points, e.g. from pixelized sensors, yields a discrete data set $J(x_i, y_j)$ instead of a continuous function. To transform the continuous centroid to its discrete counterpart on a finite area, the infinite integrals have to be converted into sums over a finite range. Thus the discrete counterpart of the 1st and 0th moment are

$$J_{x,0} = \sum_{i=1}^{N} \sum_{j=1}^{M} x_i J(x_i, y_j), \tag{3.17}$$

$$J_{0,y} = \sum_{i=1}^{N} \sum_{j=1}^{M} y_j J(x_i, y_j) \tag{3.18}$$

and

$$J_{0,0} = \sum_{i=1}^{N} \sum_{j=1}^{M} J(x_i, y_j). \tag{3.19}$$

Now Equations 3.17 and Equation 3.19 can be used to calculate the centroid on the discretized function. For the median, the definition on a discrete data set can be derived from a discretization of Equation 3.11 and 3.12 or directly from the classical definition of the median in order statistics.

3.2.1 Influence of Impulsive Noise and Background Illumination

One major noise source in the measured photocurrents is impulsive, i.e. outliers due to defective pixels or circuitry, and a second one uniformly distributed as a result of a constant background illumination. The chosen estimator should be robust against impulsive noise, but also unbiased in the presence of the background illumination.

First, the effect of impulsive noise on the median, the centroid and the maximum will be addressed. The median is a robust estimator of location and is often used in image processing for the supression of impulsive noise [Jäh97]. The number of outlying values n in the data set with w values has to be larger than $w/2$, before the median gets affected by the outliers and the estimation is no longer unbiased [PV90]. Every outlier J_o has an influence of J_o/w on the centroid though, which makes it sensitive to impulsive noise. Note, that the implementation of the median is also possible on a circuit level in VLSI [OK97]. With respect to the maximum, it gets affected by large impulsive noise values.

A constant background illumination level J_B on the irradiance distribution

$$J'(x, y) = J(x, y) + J_B. \tag{3.20}$$

affects both the median and the centroid through the addition in the integrals of the moments calculation for the centroid and the integrals of the median calculation. This will be shown for the centroid here. Without loss of generality the range for x_i and y_j is assumed to be symmetrical around $x = 0$ and $y = 0$. From Equation 3.19 the 0th moment in the presence of a constant background follows

$$J'_{0,0} = J_{0,0} + NMJ_B \tag{3.21}$$

and from Equation 3.17 the 1st moment follows

$$J'_{x,0} = J_{x,0} + \sum_{i=1}^{N} \sum_{j=1}^{M} x_i M.$$

The remaining sum is zero for x_i symmetrical around $x = 0$. The centroid from Equation 3.9 is thus

$$\overline{x} = \frac{J_{x,0}}{J_{0,0} + NMJ_B}. \tag{3.22}$$

The same calculation applies for the y-coordinate. The larger the background level J_B, the more will the centroid be shifted to the origin $x = 0$ and the estimation is more and more biased. The effect on the median is similar. In the presence of a constant background illumination the maximum seems to be favorable, because it is not affected as long as $J_B < J_0$.

To make the centroid and the median unbiased in the presence of the background illumination a threshold can be applied to the intensity distribution first. The thresholded function of an intensity distribution without constant background is the same as the thresholded function with background, as long as the threshold is larger than the background $J_B < J_T$. To get a got performance in the presence of both impulsive and constant background noise, the median taken of the thresholded function seems to be the best tradeoff.

3.2.2 Estimators of Spread

Not only the value of an estimated parameter is of importance, but also the spread of the parameter itself, e.g. to reject outlying position measurements. Common estimators of spread (for a single random variable x) are the *standard deviation*

$$\sigma(x) = \sqrt{\sum_{i=1}^{N} (x_i - \langle x \rangle)^2 / (N-1)} \tag{3.23}$$

the *mean absolute deviation from the median*

$$\mathrm{mad}(x) = \sum_{i=1}^{N} |x_i - \mathrm{med}(x)| / N, \tag{3.24}$$

and the *interquartile range*

$$IQR(x) = H_{0.75}(x) - H_{0.25}(x).$$

The two last estimators are closely related to order statistics and therefore more robust against outliers, than the standard deviation. Note, that for biased data

$$\mathrm{bias}(x) = \langle x \rangle - \overline{x} \neq 0 \tag{3.25}$$

the *root-mean-square*

$$\mathrm{rms}(x) = \sqrt{\sum_{i=1}^{N} (x_i - \overline{x})^2 / (N-1)} \tag{3.26}$$

is not identical to the standard deviation. Both are related through

$$\mathrm{rms}^2(x) = \sigma^2(x) + \mathrm{bias}^2(x). \tag{3.27}$$

instead.

3.3 Lateral-Effect Photodiode

Position Sensitive Detectors (PSDs) are a physical realization of an estimator of location, where the estimated parameter is the spatial location of an object. Although any spatial distribution of a physical quantity, e.g. temperature, magnetic field strength, can be the input of the PSD, only opto-electronic PSDs are considered here, where the data set is the photocurrent. Two basic categories of PSDs exist, the first being *lateral effect PSDs* and the second being *pixelized PSDs*. While the lateral effect PSDs exploit the geometry of the photosensitive area and the effect of current redistribution, the pixelized PSDs sample the physical quantity spatially and do the processing on a circuit level. The operations different PSDs perform on the input photocurrents are related to their mathematical equivalents, e.g. the maximum or the centroid, within certain limits.

Probably the most simple PSD is the *lateral-effect photodiode*, which is a large photodiode with sheet resistance ρ and two (1D) or four (2D) contacts at each side. Photogenerated charges create a potential difference and drift to one of the sides, where they are collected by the contacts. The operation performed by the 1D-lateral-effect photodiode resembles the centroid operation, what will be shown hereafter.

A photocurrent I_p is generated at point x in the resistive material and divided into a current flowing to the left I_L and a current I_R flowing to the right contact. These currents can be found from Kirchhoff laws

$$I_L x\rho = I_R(l - x)\rho \tag{3.28}$$

and

$$I_R l = x(I_L + I_R) = xI_p(x), \tag{3.29}$$

where l is the length of the diode. Integration over all current contributions $I_p(x)$ yields

$$I_R l = \int_{x=0}^{l} xI_p(x)dx = I_1, \tag{3.30}$$

which is equal to the 1st moment I_1 of the photocurrent distribution. The 0th moment is the sum of the two currents $I_0 = I_L + I_R$ and the centroid is found from the output currents according to

$$\overline{x} = \frac{I_1}{I_0} = l\frac{I_R}{I_R + I_L}. \tag{3.31}$$

It should be noted, that the centroid is a linear function of the left (and the right) output current, because the sum of both currents is constant.

Although the lateral-effect photodiode directly provides the centroid in one dimension, it is not suitable for the two-dimensional case, because the position measurements become interdependent, which results in position nonlinearity. The large sheet resistance required, also defavours the lateral-effect photodiode.

3.4 Quad-Cell

The limitations of the lateral-effect PSDs in two dimensions case may be overcome by separating photocurrents for the x- and for the y-coordinate. This implies a division of the photosensitive area into four separate photodiodes, see Figure 3.1 for a schematic. The position estimates for x and y from the quad-cell are independent from each other, but in general no linear function of the left and right or upper and lower output current any more. Instead a pseudo-position function has to be defined, which is a functional of the irradiance distribution.

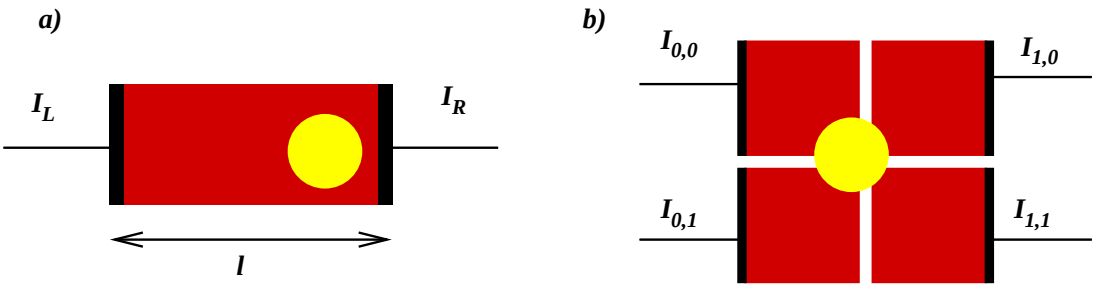

Figure 3.1: Lateral-effect photodiode (a) and quad-cell (b). In the quad-cell, the relation between the four photocurrents define the pseudo-position of the spot.

The four photodiodes with the photosensitive area $A_{0,0}, A_{1,0}, A_{1,1}$ and $A_{1,1}$ with respective photocurrents $I_{0,0}, I_{1,0}, I_{0,1}$and $I_{1,1}$ define the pseudo-position functions for x and y

$$x_{\mathrm{QC}} = \frac{(I_{1,1} + I_{0,1}) - (I_{1,0} + I_{0,0})}{I_{0,0} + I_{0,1} + I_{1,0} + I_{1,1}}, \tag{3.32}$$

$$y_{\mathrm{QC}} = \frac{(I_{1,0} + I_{1,1}) - (I_{0,1} + I_{0,0})}{I_{0,0} + I_{0,1} + I_{1,0} + I_{1,1}}, \tag{3.33}$$

with x_{QC} and y_{QC} ranging from -1 to 1. It is obvious, that the pseudo-position function is a function of the irradiance distribution, because the photocurrents

$$I_{i,j} \sim \int^{A_{i,j}} J(x,y)dxdy \tag{3.34}$$

from photodiode $A_{i,j}$ are a function of the photocurrent distribution. A CMOS integrated wavefront sensor based on quad-cells has been designed and analyzed by Lima Monteiro [dLM02], see an exemplary quad-cell in Figure 3.2.

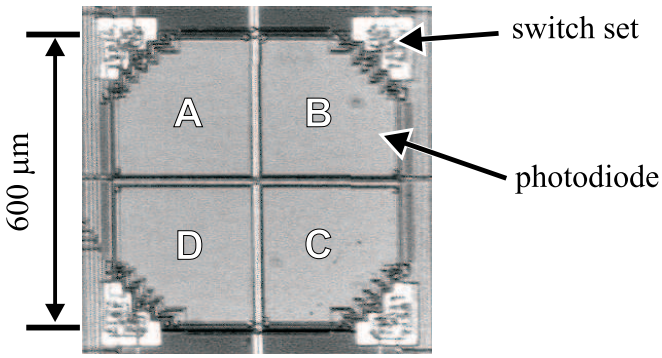

Figure 3.2: Micro-photograph of a quad-cell from a CMOS-based wavefront sensor implemented by Lima Monteiro [dLM02].

3.5 Multipixel Position Sensitive Detectors

In *multipixel PSDs* the photocurrent distribution is first sampled at a large number of points and then processed in a second step. In many applications the imaging processing task can be reduced from real 2D to 1D processing, when the operations on the data for the x- and the y-coordinate are independent from another. In this case pixels can be connected together in rows and columns and the processing can be performed spatially separated from the pixel array. This implies, that the photodiodes in a pixel are either connected to a column or a row, or that the pixel contains two photodiodes. A number of possible layouts has been proposed to divide pixels and photodiodes into two groups for columns and rows each, ranging from chessboard-like patterns to spiral structures within each pixel [dLMVS99].

Yet another possibility would be to alternate between x and y in time, by switching first column and then row photocurrents to the same processor. A number of multipixel PSDs will be introduced subsequently, including the ones realized in this thesis.

3.5.1 Resistive-Line Networks

The *resistive-line network* is the discrete version of the lateral-effect photodiode. Photocurrents for the x- and for the y-coordinate are separated by dividing the pixels into two groups. The photocurrents from pixel rows and columns are collected and fed into a resistive-line network, like the one depicted in Figure 3.3.

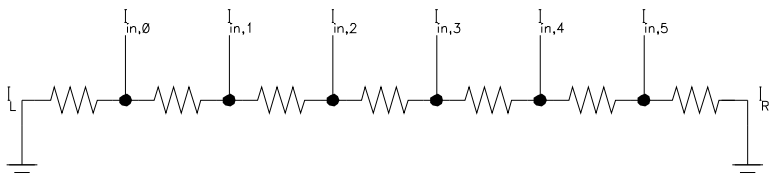

Figure 3.3: Schematic of a one-dimensional resistive-line network with six current inputs.

The resistive-line network avoids the nonlinearity associated with the lateral-effect photodiode in two dimensions and the quad-cell, by separating photodetection and current redistribution from one another.

The photocurrents at the edges of the grid $I_{L,x}, I_{R,x}, I_{L,y}$ and $I_{R,y}$ with N input currents are the weighed sum of the row photocurrents

$$I_{R,x}N = \sum_{k=1}^{N} kI_{\text{in},k} \tag{3.35}$$

and the column photocurrents $I_{R,y}$ respectively. For a large number of input currents the centroid is approximated here

$$\overline{x} = \frac{I_{R,x}}{I_{L,x} + I_{R,x}} \quad \text{and} \quad \overline{y} = \frac{I_{R,y}}{I_{L,y} + I_{R,y}} \tag{3.36}$$

from Equation 3.31. Besides this straightforward 2D expansion of the lateral-effect photodiode, which reduces the 2D to a 1D problem, real 2D resistive-grid networks have been built by Standley [Sta91]. The photocurrents in Standley's scheme are distributed to all four nearest neighbors of each pixel.

3.5.2 Winner-Take-All Circuit

The *Winner-Take-All (WTA)* circuit in its most simple form detects the global maximum of a number of input currents and thus performs the maximum operation from Equation 3.3. Its basic advantage is the digital signal it provides as an output. This means that the WTA circuit can be directly connected to digital standard cells; no further amplification, A/D-conversion or calculation is necessary.

The WTA circuit is a typical example of the smart exploitation of the device characteristic of the MOS transistor and often referred to as an example for bio-inspired signal processing. The current-mode WTA circuit has been introduced by Lazzaro [LRMM89]. Later the circuit has been modified to improve resolution and speed [SF93, SGLB98]. The circuit is often used in conjunction with neural analog networks [UN95]. Other modifications allow the circuit to use feedback and local inhibition [DM95]. Besides WTA circuits also Loser-take-all circuits have been proposed [WJK00].

A new type of topology for the WTA circuit has been developed in this thesis, which is especially suitable for spot detection and tracking, therefore the operating point of the circuit will be derived in detail hereafter. Figure 3.4 shows a schematic of the basic WTA circuit. It consists of N *neurons*, which sense the N input currents $I_{\mathrm{in},k}$. All neurons are interconnected through a common net with a single current sink I_{src}. Each neuron contains a sense-transistor $M_{S,k}$ and a flow-transistor $M_{F,k}$, through which the output current $I_{\mathrm{out},k}$ flows. The sense-transistor with the largest input current has the highest drain potential. The adjacent flow-transistor therefore has the highest gate-source potential and drains all the available current from the current sink. It thereby indicates the position of the largest input current and the winning neuron. Almost no current will flow through the other flow-transistors. The WTA has a binary spatial impulse response in form of the bit-vector

$$\vec{I}_{\mathrm{out}} = I_{\mathrm{src}}[0, \ldots, 1, \ldots, 0]. \tag{3.37}$$

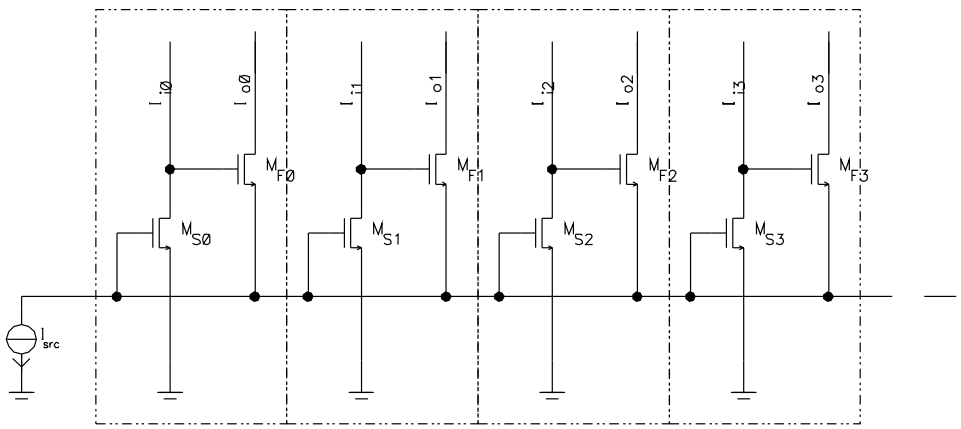

Figure 3.4: Schematic of the simple Winner-Take-All circuit topology. For each input current $I_{i,k}$, a binary output current $I_{o,k}$ is established, that denotes the largest of all input currents.

The DC operating point for a two-neuron WTA circuit will now be derived in detail. The drain-source voltages of the sense-transistors are found from Kirchhoff

laws

$$V_{GS}(M_{S,0}) + V_{GS}(M_{F,0}) = V_{DS}(M_{S,0}), \tag{3.38}$$
$$V_{GS}(M_{S,1}) + V_{GS}(M_{F,1}) = V_{DS}(M_{S,1}). \tag{3.39}$$

The gate-source voltages of the sense-transistors according to Equation 2.3 are

$$V_{GS}(M_{S,1}) = nV_T \ln \frac{I_{\text{in},1}}{SI_{D0}(1-\exp(-V_{DS}(M_{S,1})/V_T))(1+\lambda V_{DS}(M_{S,1}))}, \tag{3.40}$$
$$V_{GS}(M_{S,0}) = nV_T \ln \frac{I_{\text{in},0}}{SI_{D0}(1-\exp(-V_{DS}(M_{S,0})/V_T))(1+\lambda V_{DS}(M_{S,0}))}, \tag{3.41}$$

when the Body effect is neglected. The gate-source voltages of the flow-transistors are

$$V_{GS}(M_{F,1}) = nV_T \ln \frac{I_D(M_{F,1})}{SI_{D0}(1-\exp(-V_{DS}(M_{F,1})/V_T))(1+\lambda V_{DS}(M_{F,1}))}, \tag{3.42}$$
$$V_{GS}(M_{F,0}) = nV_T \ln \frac{I_D(M_{F,0})}{SI_{D0}(1-\exp(-V_{DS}(M_{F,0})/V_T))(1+\lambda V_{DS}(M_{F,0}))}. \tag{3.43}$$

With the additional relations from Kirchhoff laws

$$I_{\text{src}} - I_D(M_{F,0}) = I_D(M_{F,1}), \tag{3.44}$$

$$V_{DS}(M_{F,0}) = V_{dd} - V_{GS}(M_{S,0}), \tag{3.45}$$

$$V_{DS}(M_{F,1}) = V_{dd} - V_{GS}(M_{S,1}) \tag{3.46}$$

and Equations 3.42 and 3.43 the DC operating point of the circuit can be calculated numerically, see Figure 3.5 for the operating point of the sense-transistor. Figure 3.6 shows the output currents as a function of the ratio of input currents for a typical parameter set from the AMS 0.35 μm process technology. The circuit is very sensitive to differences in the input currents; a difference of 1% is enough to establish the winner in this example.

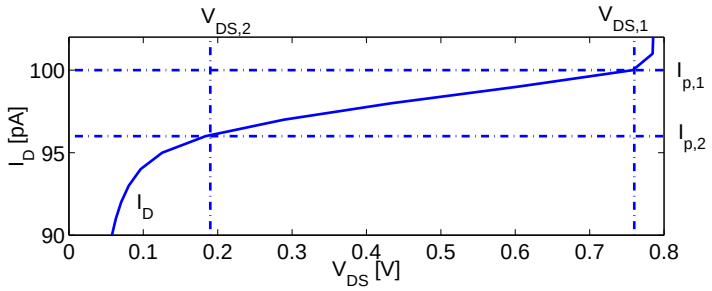

Figure 3.5: Operating point of the sense-transistor; $I_D(M_{S,1})$ versus drain-source voltage $V_{DS}(M_{S,1})$ for two slightly different input photocurrents $I_{p,1}$ and $I_{p,2}$. In this example a photocurrent difference of only 4% changes the drain-source voltage from 0.19 V to 0.76 V.

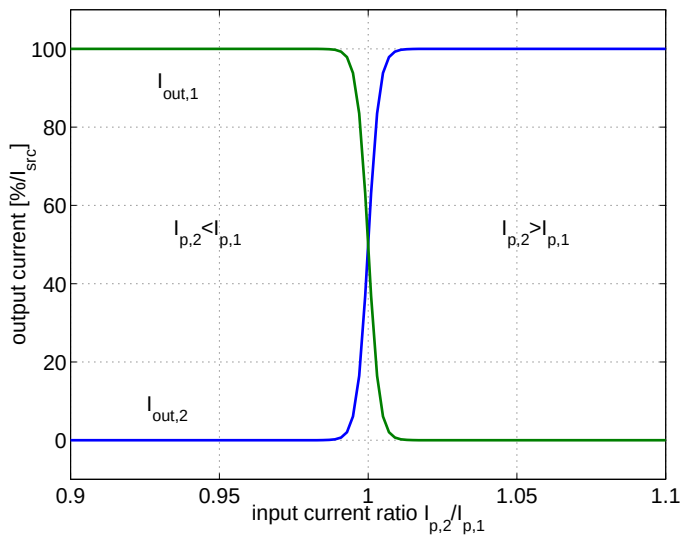

Figure 3.6: Output currents of a two-neuron WTA circuit as a function of the ratio of input currents (percentage of current I_{src}). Current differences as low as 1% are still resolvable.

3.5.3 Transient Response

The transient behavior of the WTA circuit is basically determined by the parasitic capacitances $C_{\Psi,n}$ at the drain nodes of the sense-transistors $M_{S,n}$. Three time steps occur, when the WTA circuit changes from one equilibrium state to another [DB02], e.g. from the state $I_{\text{in},2} < I_{\text{in},1}$ to $I_{\text{in},2} > I_{\text{in},1}$ at $t = 0$ and constant $I_{\text{in},1}$. These steps are shown in Figure 3.7. In the first step $0 < t < t_1$ the gate potential of $M_{S,2}$ remains constant, because it is still defined by $M_{S,1}$ and $M_{F,1}$. The parasitic capacitor $C_{\Psi,2}$ is charged with the difference of the input current and the drain current $\Delta I_2 = I_{\text{in},2} - I_D(M_{S,2})$. When $V_{DS}(M_{S,2})$ is large enough to affect the current through $M_{F,2}$, the second step $t_1 < t < t_2$ occurs; the flow-transistor $M_{F,2}$ inhibits the current flow through $M_{F,1}$ by increasing its own drain current to almost I_{src}. When the drain current of $M_{F,1}$ is almost zero, the parasitic capacitor $C_{\Psi,2}$ discharges through the sense-transistor $M_{S,1}$ in the third step $t > t_2$ and the new equilibrium state is reached.

In step two of the transient response the drain currents of both flow-transistors will be non-zero; which is the winning neuron is ambiguous during this transition time $t_1 < t < t_2$. To decrease the transient response time, a number of methods have been investigated by Droste [DB02], like the decoupling of the parasitic capacitance from the drain of the sense-transistors $M_{S,n}$ or a reset of all drain potentials $V_{DS}(M_{S,n})$ to a predefined potential and thus bypass step one of the transition phase. These measures together reduced the transient response time more than three orders of magnitude.

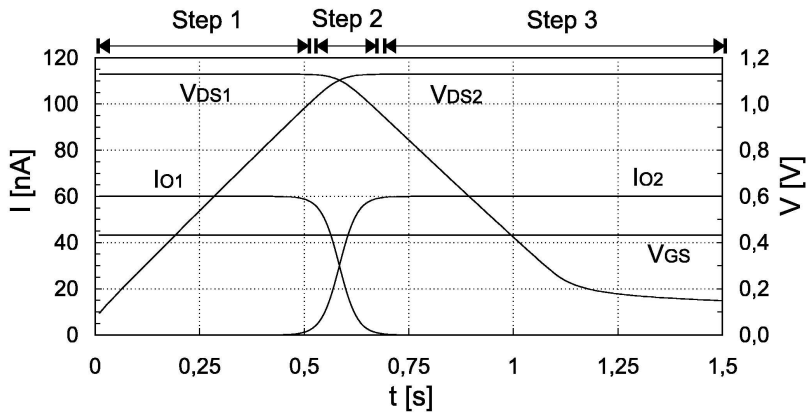

Figure 3.7: Time steps for the output current and the drain-source voltages of the sense-transistors when changing the state of the WTA circuit [DB02] from neuron 1 being the winner to neuron 2 being the winner. Step 1 is dominated by the rising $V_{DS}(M_{S,2})$ due to the charging of the parasitic capacitance at the drain node of $M_{S,2}$. In step 2 $M_{F,2}$ inhibits the current flow through $M_{F,1}$. Step 3 is dominated by the discharging of the parasitic capacitance at the drain node of $M_{S,1}$.

3.5.4 Local Inhibition Winner-Take-All Circuit

The *Local Inhibitory WTA Circuit* has been proposed to find local winners (or maxima) instead of global ones [LRMM89]. Instead of having one global current sink, each neuron contains its own sink. Adjacent neurons are not interconnected by a global net, but through resistances R, see Figure 3.8 for a schematic. A winning neuron sinks the current from its own current sink and from some of its neighbors and therefore inhibits the current flow through their flow-transistors. The spatial width of the influence of the winner can be controlled by varying the resistance R between global ($R = 0$) and local ($0 < R < \infty$).

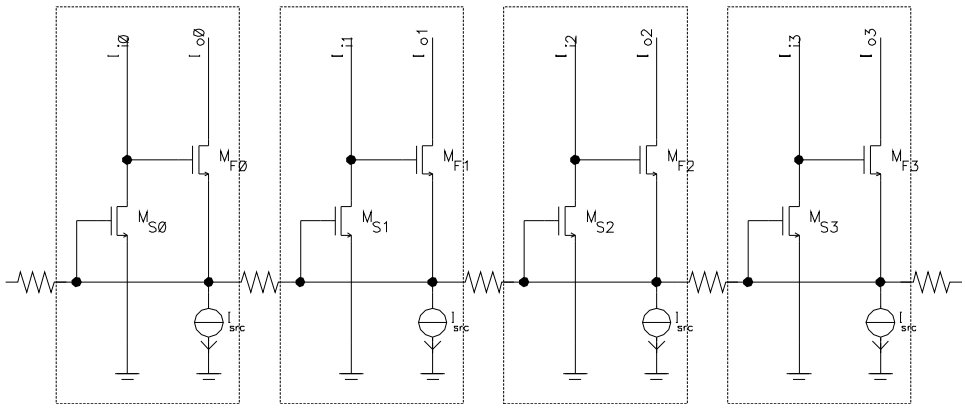

Figure 3.8: Schematic of the local inhibition Winner-Take-All circuit. The sense-nets of the neurons are interconnected through resistors. The output currents thus depend on the resistance value.

The local inhibition WTAs are interesting PSDs, because the performed operation can resemble the maximum operation if set to global or can approximate the centroid operation, if set to local. This operation is only pseudo-centroid, because the estimated position is not derived from the linearly weighed sum of the input currents, but from nonlinear weighing instead. The disadvantage for spot detection is, that the local inhibition WTA may find multiple local winners, although only one spot can be present in each subaperture.

The spatial impulse response is also not binary. It will be derived here for a continuous approximation of the discrete network. Without loss of generality the largest photocurrent is at $x = 0$. The winning neuron sinks the current from its own current sink and a lateral current from its neighbors $I_{\text{src}} + 2I_{\text{lat}}$. Their flow-transistors are thus inhibited by the reduced current flow through their own flow-transistors.

To derive the spatial impulse response, the voltage drop from one neuron to the next through the resistance ρ is calculated first

$$V(x + dx) - V(x) = -\rho I(x + dx)dx. \tag{3.47}$$

The current flow through the horizontal resistors is given by Kirchhoff laws

$$I(x+dx) - I(x) = (I_D(x) - I_{\mathrm{src}})dx = (g(V_{dd} - V(x)) - I_{\mathrm{src}})dx, \tag{3.48}$$

where g is the transconductance of the flow-transistors. Equations 3.47 and 3.48 are combined to

$$\frac{dV(x)}{dx} = -\rho I(x), \tag{3.49}$$

$$\frac{dI(x)}{dx} = -g(V_{dd} - V(x)) - I_{\mathrm{src}}. \tag{3.50}$$

Differentiation of Equation 3.49 and combination with 3.50 leads to the differential equation

$$\frac{d^2V(x)}{dx^2} = \rho(I_{\mathrm{src}} - g(V_{dd} - V(x))), \tag{3.51}$$

which is solved by the exponential function with offset

$$V(x) = V_{dd} - \frac{I_{\mathrm{src}}}{g} + V_0 e^{-\sqrt{\rho g}|x|}. \tag{3.52}$$

It is convenient to define the *diffusion length* $l = 1/\sqrt{\rho g}$. The lateral currents then are

$$I(x) = -\frac{1}{\rho}\frac{V(x)}{dx} = I_0 e^{-|x|/l} \tag{3.53}$$

for $x \neq 0$. Note, that the vertical quantities are current densities instead of currents in this continuous approximation of the discrete network. The lateral current will approximate zero for neurons farther away from the winner than the diffusion length. These neurons maintain their function and other local maxima are detectable there.

The total drain current of the winning neuron is $I_D(0) = I_{\mathrm{src}} + 2I_{\mathrm{lat}}$. The drain current density through the other flow-transistors is $j_D(x)dx = j_{\mathrm{src}}dx - I(x)$ and the spatial distribution of drain currents therefore

$$j_D(x) = j_{\mathrm{src}} - j_0 e^{-|x|/l}, \tag{3.54}$$

for $x \neq 0$. The output current logarithmically approaches j_{src} with increasing distance from the winner. Integration of all drained currents leads to the current density j_0, which must satisfy the Equation

$$\frac{I_D(0)}{2} = \frac{I_{\mathrm{src}}}{2} + \int_{x=0}^{\infty} I(x)dx = \frac{I_{\mathrm{src}}}{2} + j_0 l. \tag{3.55}$$

Solving for j_0 gives

$$j_0 = \frac{I_D(0) - I_{\text{src}}}{2l}. \tag{3.56}$$

The spatial current distribution calculation for a discrete resistive line is more tedious, but the solution differs little from the continuous case, if $l \geq 1$. A comparison between the analytical result for the continuous case and an analog Spice [2] simulation can be found in Figure 3.9.

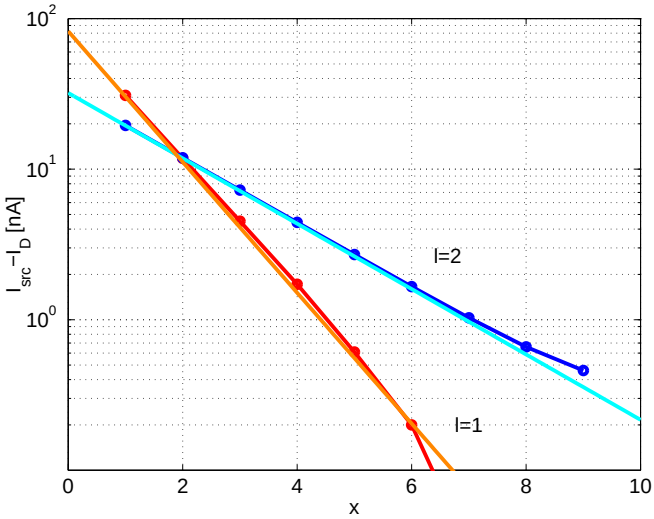

Figure 3.9: Comparison between a discrete resistive line with diffusion length $l = 1$ and $l = 2$ for I_{src}=50 nA and $I_D(0)$=100 nA from a Spice simulation (line with circles), and the analytical result for the continuous case (straight lines). Discrete and continuous case show the same exponential decay with distance from the winning neuron. The coordinate x is either the spatial dimension in the continuous or the index of the neuron in the discrete case.

Some approximations have been made in this derivation, e.g. the nonlinearity of the transconductance of the flow-transistors has been neglected. The horizontal resistors are generally substituted by active resistors, which induce further nonlinearities.

3.5.5 Resistive-Ring Network of Winner-Take-All Circuits

In this thesis a new topology for WTA circuits is presented, which is especially suitable to estimate the location of a single maxima on a fixed number of input currents, i.e. the

[2] *SPICE* stands for *Simulation Program Integrated Circuits Especially*. The program originates from the University of California, Berkeley. Spice is used to provide a reasonably detailed analysis of circuits containing active components such as bipolar transistors, field effect transistors, diodes and passive components such as resistors, capacitors and inductors.

position of a focal spot from a number of input photocurrents. The performed operation approximates the threshold operation on the photocurrent distribution in both dimensions for a given spot width. The thresholded photocurrent distribution is thus binary and the position of the spot can be found from the centroid or the median of the thresholded photocurrent distribution. It will be shown, that for an intensity distribution as expected for a lenslet from a lenslet array, the task of thresholding the photocurrent distribution is equal to the task of finding the w adjacent largest photocurrents.

Interdigital WTA circuit

In order to find all photocurrents that are larger than the photocurrent at the FWHM, a single WTA circuit with n neurons is split into w groups, having n/w neurons each. The number of groups w is chosen such, that its number multiplied by the pixel pitch p equals the FWHM of the spot $w = \mathrm{FWHM}/p$. The drains of the flow-transistors are connected in an interdigital manner to w current sinks

$$I_D(M_{F,i}) \rightarrow I_{\mathrm{src},n} \;\; \text{with} \;\; n = i \mod w. \tag{3.57}$$

This circuit establishes an output current vector of the form

$$\vec{I}_{\mathrm{out}} = I_{\mathrm{src}}[0, \ldots, 0, 1, \ldots 1, 0 \ldots, 0], \tag{3.58}$$

with w winning neurons and $n - w$ losing neurons. It can be seen from the output current vector, that the output is equal to a thresholding of the input photocurrents, where the threshold itself is set to the value at the FWHM of the photocurrent distribution. Although the output of the circuit is not the position of the spot, this can be easily found from the mean or median of the winning neurons in the output current vector.

The performance of the circuit is not limited to a fixed spot size. If the spot is much larger than expected, the w largest photocurrents lie closer to the maximum of the photocurrent distribution and the operation the circuit performs becomes comparable to the maximum operation of Equation 3.3. On the other side, the FWHM should not be much smaller than the width $p \cdot w$, because in this case some of the now independent WTA circuits will have no photocurrent inputs and will thus show random winners.

To circumvent these spot size limitations of the interdigital WTA circuit, the resistive ring network of WTA circuits has been implemented in the sensor developed in this thesis, which will now be described.

Resistive-Ring Network of Winner-Take-All Circuits

The limitations with respect to spot size of the interdigital WTA circuit may be overcome by connecting the current sinks through a ring of active resistors, see Figure 3.10 for a schematic of the ring. By varying the resistance R the number of winners $w(R)$ can be varied between the number w and a single one. A ring of resistors must be used, instead of a line of resistors, to maintain the symmetry between adjacent WTA circuits.

When the FWHM of the spot is much smaller than the width $p \cdot w$, the resistance can be reduced and a fewer number of neurons wins. When the resistance is reduced to zero, the performance of the circuit is equal to the performance of the simple WTA circuit, i.e. it performs the maximum operation.

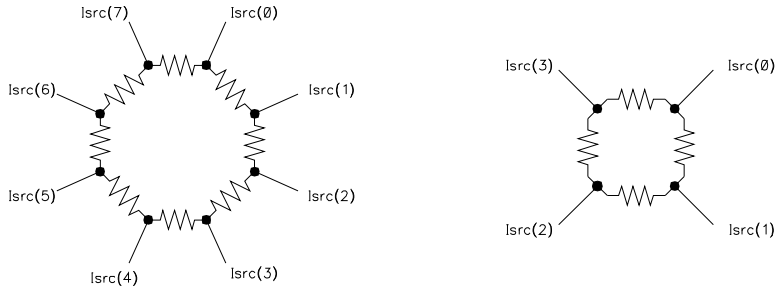

Figure 3.10: Resistive rings with $w = 8$ and $w = 4$ resistors. The nodes are connected to the current sinks of w individual WTA circuits. When the resistance is large, w winners will be detected. When the resistance approaches zero, only one global winner will be selected. Intermediate resistance values lead to a number of winners between one and w.

The spatial influence of a WTA circuit connected to the resistive ring, can be fully understood by taking only the largest input photocurrent in the respective WTA circuits into account, as the other losing WTA neurons are already inhibited by the winning neuron in their own WTA circuit. The spatial impulse response calculation of the resistive ring network follows a similar way, as the one for the resistive line network. Instead of an infinite resistive line, the periodicity on the circumference U of the ring has to be taken into account. To calculate the node voltages $V(x)$ with $x \in [0, U)$ along the circumference of the ring, the largest photocurrent is assumed to be at $x = 0$ without loss of generality. All neurons notice the influence of the winner from both neurons in clockwise and in counterclockwise direction. A mirrored image of the winning neuron is placed at $x = U$ to take this U-periodicity into account. The voltages along the circumference x of the ring are

$$\begin{aligned} V(x) &= V_{dd} - \frac{I_{\text{src}}}{g} + V_0 \left(e^{-x/l} + e^{-(U-x)/l}\right) \\ &= V_{dd} - \frac{I_{\text{src}}}{g} + V_0 \left(k_1 \left(\cosh(l(x - U/2)) - 1\right) + k_0\right), \end{aligned} \tag{3.59}$$

with

$$k_0 = 2e^{Ul/2} \tag{3.60}$$

and

$$k_1 = \frac{1 + e^{-Ul}}{\cosh(Ul/2)}. \tag{3.61}$$

The voltage and thereby the drain current distribution follows a hyperbolic cosine law according to Equation 3.59. The lateral currents are

$$I(x) = -\frac{1}{\rho}\frac{V(x)}{dx} = I_0\left(e^{-x/l} + e^{-(U-x)/l}\right) \tag{3.62}$$

and the spatial drain current distribution is

$$j_D(x) \quad = \quad j_{\mathrm{src}} - j_0\left(e^{-x/l} + e^{-(U-x)/l}\right). \tag{3.63}$$

The current density j_0 must satisfy the equation

$$\frac{I_D(0)}{2} = \frac{I_{\mathrm{src}}}{2} + \int_{x=0}^{U} I(x)dx = \frac{I_{\mathrm{src}}}{2} + j_0 l\left(1 + e^{-U/l} - 2e^{-U/2l}\right). \tag{3.64}$$

Solving for j_0 yields

$$j_0 = \frac{I_D(0) - I_{\mathrm{src}}/2}{l\left(1 + e^{-U/l} - 2e^{-U/2l}\right)}. \tag{3.65}$$

In the discrete case the variable x is an index and the current distribution can be replaced by discrete currents

$$I_D(x) \quad = \quad I_{\mathrm{src}} - I_0\left(e^{-x/l} + e^{-(U-x)/l}\right) \tag{3.66}$$

for sufficiently large diffusion lengths $l \geq 1$. A comparison between the discrete case and the continuous case can be found in Figure 3.11.

3.6 Position Uncertainty

Sources for position uncertainty in PSDs can be subdivided into spatial and temporal, although temporal noise is transduced into spatial noise through the spatial processing steps. A number of noise sources in photodetectors, like shot noise, are discussed in Chapter 2.4.2. Temporal noise $\sigma_t(x)$ is the noise from the photodetectors and the circuitry. It normally adds no bias to the position estimate and the variance can be reduced by temporal averaging of a number of subsequent position estimates. Different kinds of temporal filtering techniques, e.g. running average or running median [PV90], are available for this purpose in time series analysis.

Spatial position uncertainty $\sigma_{\mathrm{FPN}}(x)$ results in a biased position estimate, even if there is no temporal noise present. It has its origin mostly in mismatching of the

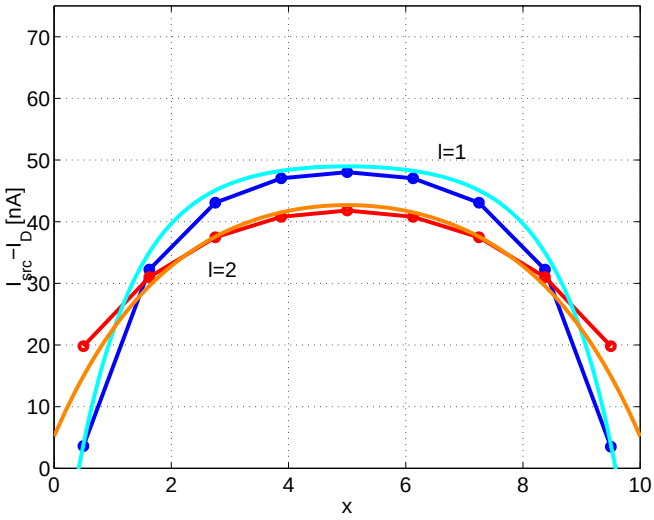

Figure 3.11: Comparison between a discrete resistive ring with 10 elements at $l = 1$ and $l = 2$ for I_{src}=50 nA and $I_D(0)$=100 nA from a Spice simulation (lines with circles), and analytical result for the continuous case (straight lines). Discrete and continuous case show the same hyperbolic cosine behavior. The coordinate x is either the spatial dimension or the index of the neuron in the discrete case.

devices of the PSD. Also the uncertainty $\sigma_{\text{res}}(x) = p/4$ of the object's position due to the finite resolution of the detector can be seen as spatial position uncertainty. All noise contributions add up quadratically to

$$\sigma^2(x) = \sigma_t^2(x) + \sigma_{\text{FPN}}^2(x) + \sigma_{\text{res}}^2(x). \tag{3.67}$$

The spatial effect of mismatching in image sensors is called *Fixed Pattern Noise (FPN)*. CMOS image sensors with active pixels suffer from the mismatching of the devices, which cannot always be reduced by making the devices larger. This is a crucial point, due to the fact, that geometrical mismatching reduces only with the square root of the W/L ratio of the devices. Transistors cannot always be made as large as desired, because this increases the parasitic capacitance and decreases the bandwidth of the sensor. The resolution is also limited by the size of the transistors in active pixels. In CMOS imagers an active FPN reduction is often used, which is the *correlated double sampling* technique [KDS+00]. This technique only works with integration based imagers. It takes two subsequent samples of the signal from the photodetector, one in the reset state of the pixel and one after integration. These signals are subtracted prior to read-out and a large part of the FPN is thus reduced. In continuously working sensors there is no reset state and therefore no reference for subtraction available.

Some position uncertainty characteristics of the WTA circuit and the ring of WTA circuits will now be discussed. Noise characteristics and measurements of the quad-

cell can be found in [dLM02]. Except from the finite resolution, position uncertainty

$$\sigma^2(I_D) = \sigma_t^2(I_D) + \sigma_{\mathrm{FPN}}^2(I_D) \tag{3.68}$$

results from temporal $\sigma_t^2(I_D)$ and spatial $\sigma_{\mathrm{FPN}}^2(I_D)$ variations of the drain current of the sense-transistors. While the direct spatial noise is time-independent, the temporal variations are transfered to temporal noise in the position estimate.

Effects of Temporal Noise

The ideal WTA circuit has an infinite bandwidth, i.e. its output currents change immediately, when one input current surpasses another formerly larger input current. Assumed that $I_{D,1}(t) > I_{D,2}(t)$ at a certain time t and that both input currents $I_{D,1}$ and $I_{D,2}$ contain the noise components $i_{D,1}$ and $i_{D,2}$, the WTA output is erroneous, when $I_{D,1} + i_{D,1} < I_{D,2} + i_{D,2}$ or expressed with the photocurrent difference $\Delta I_D = I_{D,1} - I_{D,2}$,

$$\Delta I_D < i_{D,2} - i_{D,1}. \tag{3.69}$$

The number of erroneous measurements depends on the spectral density of the noise, which can be assumed to be Gaussian, because it contains several independent sources. According to the *central-limit theorem* the quadratic sum of independent noise sources with arbitrary spectrum is normal distributed

$$p(I_{D,2}) = \frac{1}{\sqrt{2\pi\overline{i_{D,2}^2}}} e^{-I_{D,2}^2/(2\overline{i_{D,2}^2})}. \tag{3.70}$$

In an ideal 2-neuron WTA circuit the number of erroneous outputs p_{err} can be calculated from the probability of Condition 3.69

$$p_{\mathrm{err}} = \int_{I_{D,2}=\Delta I_D}^{\infty} \frac{1}{\sqrt{2\pi\overline{i_{D,2}^2}}} e^{-I_{D,2}^2/(2\overline{i_{D,2}^2}^2)} = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\frac{\Delta I_D}{\sqrt{2\overline{i_{D,2}^2}}}, \tag{3.71}$$

which is determined by the error function. Figure 3.12 shows a schematic of the probability density. The ratio $\Delta I_D/\sqrt{\overline{i_{D,2}^2}}$ can be interpreted as a signal-to-noise ratio SNR. At about $SNR = 2$ the error is negligible for most practical cases. When both currents are identical, half of the data has to be erroneous, which is consistent with Equation 3.71. These considerations apply for the ideal WTA circuit, which has zero transient response time. The real WTA circuit has a limited bandwidth and only spectral noise components smaller than the bandwidth of the circuit will produce erroneous output states. A detailed analyses of the temporal noise contributions in Droste's WTA implementation can be found in [Dro99, DB02].

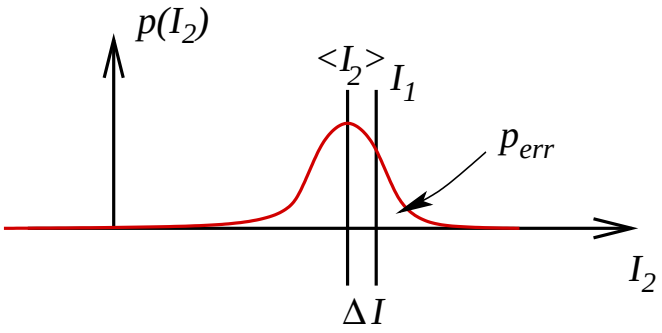

Figure 3.12: Probability density for the current with smaller mean I_2, when Gaussian noise is added. The probability to detect the wrong winner of both currents is the area above the current with largest mean I_1. This probability diminishes fast with a rising difference ΔI between the two currents.

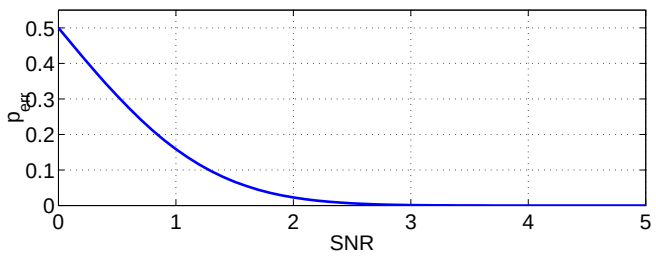

Figure 3.13: Probability of selecting the wrong winner with SNR. In most practical cases the probability is negligible above $SNR > 2$.

Effects of Mismatching

The mismatched quantity is mostly the drain current of the sense-transistor $\sigma(I_D)$, which is a function of the W/L ratio and the threshold voltage variance, according to Equation 2.8. In contrast to the temporal noise it cannot be assumed, that the drain current mismatch is normal distributed. The nonlinearity of the drain current in the subthreshold region results in an asymmetric probability density function for the drain current mismatch.

In a simple model for small mismatch of the drain currents it is assumed nevertheless, that the actual drain current is the sum of the ideal current I_D and a normal distributed current with standard deviation $\sigma(I_D)$. The spatial noise, that results from the mismatched drain currents, can be approximated for the intensity distribution from Equation 1.4 in the region between the maximum and the FWHM of the spot, where $|x| < f\lambda/D_L$. A second order polynomial expansion of the photocurrent distribution around the origin $x = 0$ yields

$$I_D(x) = I_0(1 - \frac{\pi^2 D_L^2}{\lambda^2 f^2} x^2 + O(x^4)). \tag{3.72}$$

The spatial noise due to mismatched drain currents therefore is approximately

$$\sigma^2_{\mathrm{FPN}}(x) = \frac{\sigma(I_D)}{I_0} \frac{\lambda^2 f^2}{\pi^2 D_L^2}. \tag{3.73}$$

This important result shows, that the use of a larger wavelength or a lenslet array with larger diameter or smaller focal length can reduce the effects of mismatched drain currents on the position estimate.

The interdigital WTA circuit with w independent WTA circuit groups performs w independent measurements on the photocurrent distribution. The spatial uncertainty or mismatching effects should thus be reduced by

$$\sigma^2_{\mathrm{IWTA}}(x) = \sigma^2(x)/w, \tag{3.74}$$

with respect to the simple WTA circuit. In the ring of WTA circuits the same behavior can be expected, where $w(R)$ is the number of winning neurons in the resistive-ring network. Figure 3.14 shows a Monte-Carlo simulation for mismatched drain currents. At small drain current variations the position uncertainty of the IWTA circuit is $\sqrt{5}$ times smaller, than the simple WTA circuit, as expected from Equation 3.74.

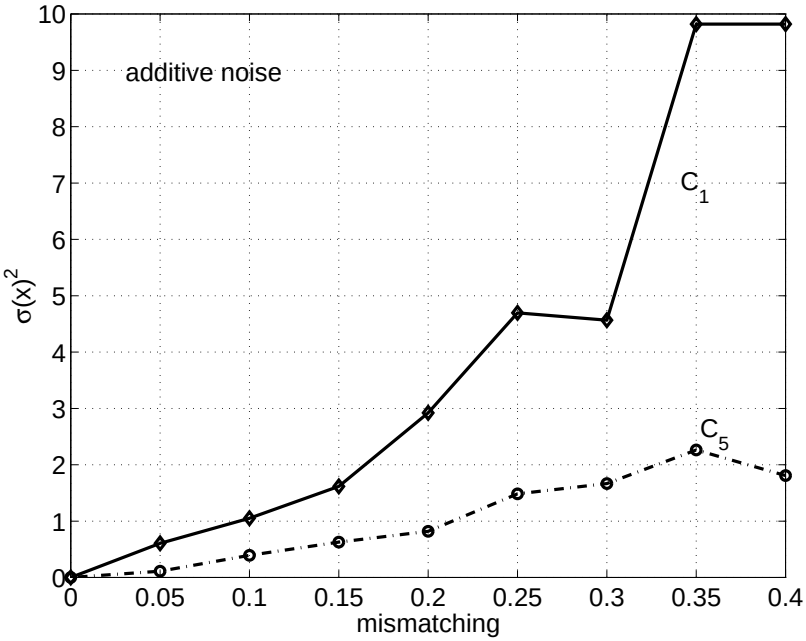

Figure 3.14: Position uncertainty in pixels due to drain current mismatch $\sigma(I_D)/I_D$. In this Monte-Carlo simulation a spot with a FWHM of five pixels has been tracked with the simple WTA circuit (C_1) and the interdigital WTA circuit with five WTA groups (C_5).

4 The Centroid Hartmann-Shack Sensor

This chapter describes the development of the CeHSSA chip, including details about its schematic and layout. In the first part the pixel matrix and the analog circuitry is introduced and the measures to decrease the transient response time of the position sensitive detectors. The second part describes the digital part of the chip for the addressing of a specific detector and the read-out scheme. In the last part the prototype version of the chip is described and the conclusions drawn from the quantum efficiency and spot tracking measurements.

4.1 General Architecture

The *Centroid Hartmann-Shack Sensor ASIC (CeHSSA)* has been developed for spot tracking at very low spot intensities, which is a necessary demand for ophthalmic applications. Special attention had to be paid to the photodetector quantum efficiency, read-out noise, linearity and bandwidth. Besides these demands other aspects had to be taken into account, e.g. simple read-out, easy scalability to a larger number of spots and also the use of an industrial submicron process technology, where there is no need for further process steps.

The chip has been realized in the AMS 0.35 μm CMOS process technology, which offers three metal layers, with $n+$/substrate photodiodes as photodetectors. The core of the chip contains 8$\times$8 spot detectors with 21$\times$21 passive pixels each, forming a total number of about 28.000 pixels. A resistive ring of WTA circuits has been chosen as a PSD for spot tracking, such that the position of the spot is derived from the largest photocurrent or the approximate centroid of photocurrents, referred to as *maximum mode* and *centroid mode* respectively. The data of the addressed detector is written to a tristate bus for subsequent read-out through several multiplexer read-out schemes. See Figure 4.1 for a general architecture of the chip. An external test-vector can be written to the the tristate bus as a basic communication test with the chip. Dur-

ing this test the read-out of the detector array is disabled. The components of the chip are described in detail in the following chapters.

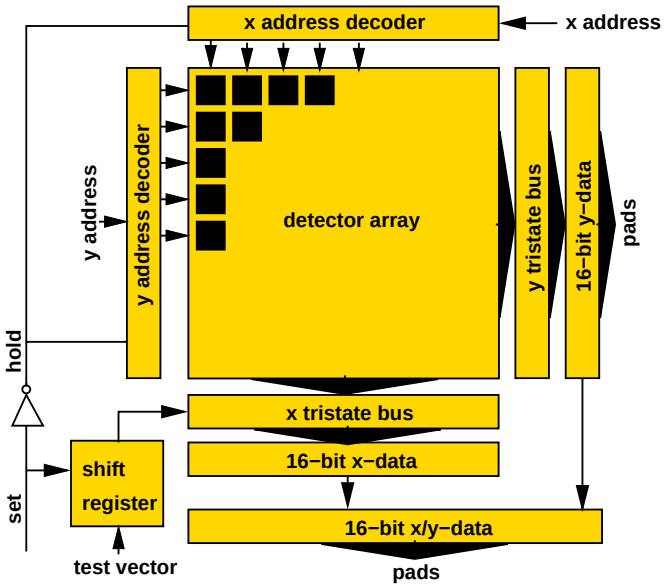

Figure 4.1: General architecture of the CeHSSA.

4.2 Position Sensitive Detector

A resistive-ring network of WTA circuits, see Chapter 3.5.5, has been chosen as a PSD for the CeHSSA. Also a resistive-line network, see Chapter 3.5.1, has been evaluated in a prototype version. Each PSD matches the size of the microlens array of 400×400 μm^2. It contains a pixel array, the resistive-ring network of WTA circuits and digital multiplexers for read-out. See Figure 4.2 for an overview of the general architecture of the PSDs.

A *fill factor* of 80% has been achieved by optimizing the layout for minimum area consumption. The PSD allows a tracking of the spots within a dynamic range of 357 μm in maximum mode and 289 μm in centroid mode. The defocus-only dynamic range[1] of the PSD is explained in Table 4.1. The pixel resolution of 17 μm results in a defocus resolution of 0.087 dpt. A general analysis of the wavefront error as a function of the resolution has been done by Lima de Monteiro, see Figure 4.3. A wavefront error

[1] All Zernike terms add up independently to the wavefront, therefore a general dynamic range can only be given for single terms. It is convenient to use the defocus term to calculate the dynamic range, which has a partial derivative of $4Cx$ in x-direction. The defocus at the full dynamic range can be calculated to be $C = \Delta x/(4xf)$, with x being the distance of the center of the outermost detector from the center of the detector array. The coefficients in $[\mathrm{dpt}] = [1/\mathrm{m}]$ are transformed to the wavefront in μm by $C\,[\mathrm{dpt}] = C\,[\mu\mathrm{m}]/A^2$, where A is the aperture.

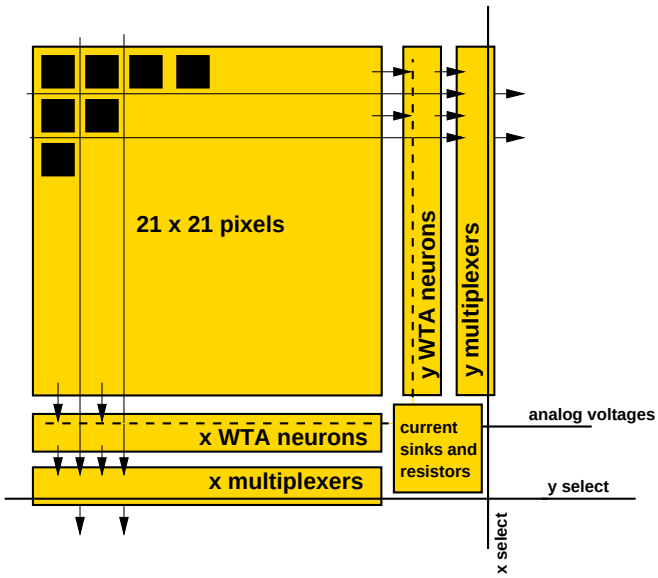

Figure 4.2: General architecture of the spot detectors.

of approximately 0.2λ at 633 nm and nine Zernike terms can be expected for the given resolution.

	8×8	4×4	2×2
centroid	1.47 dpt	3.44 dpt	10.32 dpt
maximum	1.82 dpt	4.25 dpt	12.75 dpt

Table 4.1: Dynamic range of defocus only, for different numbers of detectors in centroid and maximum mode at 53 mm focal length.

4.3 Pixel Array

The pixel array contains 21×21 passive pixels with 17 μm pitch. This pitch has been adjusted to track spots with a FWHM of 85 μm at a wavelength of 680 nm. The pixels are arranged in a chessboard-like manner, see Figure 4.4 for the layout. The photodetector has been chosen with respect to quantum efficiency, junction capacitance and layout restrictions[2]. Table 4.2 shows a comparison of the achievable exemplary photocurrents, when solving Equation 2.13 for I_p and junction capacitances from Equa-

[2]Identical layers must have a minimum spacing between each other due the limited resolution of the photolithographic process. The minimum allowed spacing between diffusion layers is 600 nm and between n-wells 3000 nm in the AMS 0.35 μm process technology [AMSa]. This reduces the fill-factor for pixels with n-well/substrate-diodes with respect to pixels with diffusion diodes.

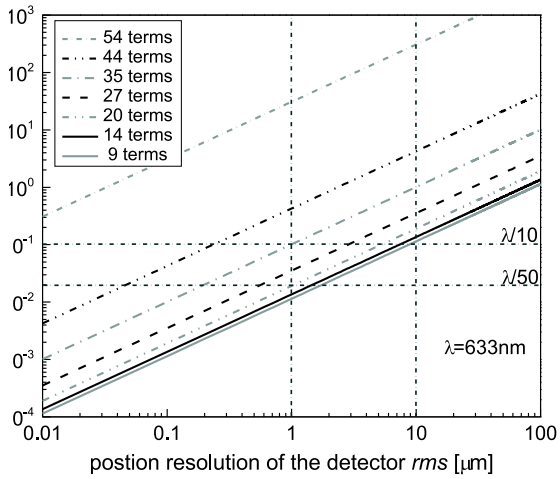

Figure 4.3: Wavefront RMS error with detector resolution for different reconstruction orders from Lima Monteiro [dLM02].

tion 2.17 for the three types of available junction photodiodes. The large quantum efficiency and the small parasitic capacitance strongly favor the $n+$/substrate diode.

	$n+$/substrate	$p+/n$-well	n-well/substrate
C_j	145.2 fF	172.7 fF	11.0 fF
I_p	80.1 pA	24.1 pA	219.4 pA

Table 4.2: Junction-photodiode parasitic capacitance and photocurrent for a pixel with 17×17 μm^2 at 3.3 V reverse bias voltage, 1 nW incident power at 680 nm wavelength for the three different types of junction photodiodes.

4.4 Resistive-Ring Network of WTA Circuits

A resistive ring of WTA circuits receives the photocurrents from the pixel array and provides a digital output of the spot position in form of a bit-vector, when the detector is addressed through its x- and y-address lines. The ring consists of five WTA circuits, four of the them have four neurons and one of them has five neurons. All together they receive 21 input photocurrents from the rows and columns of the pixel array. Figure 4.5 shows a schematic of the implemented circuitry. The number of WTA circuits connected to the resistive ring has been adjusted to the minimum FWHM of the spot, which is 85 μm/17 μm$= 5$. All neurons $n = [0 \ldots 20]$ are identical, except with respect to their connection to one of the five current sinks, which alternates in the form

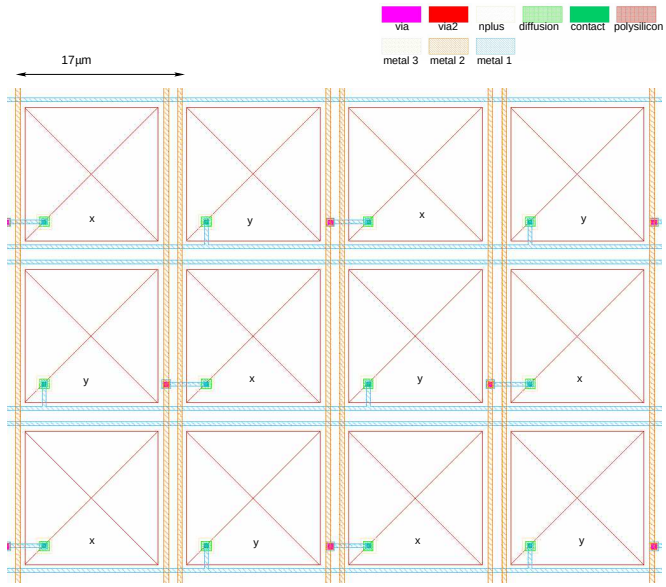

Figure 4.4: Chessboard-like layout of the pixel array. The n-well/substrate photodiodes are represented by crosses.

$$I_D(M_{F,i}) \rightarrow I_{\mathrm{src},n} \;\; \text{with} \;\; n = i \mod w. \tag{4.1}$$

Figure 4.6 shows a schematic of the implemented neuron. Due to the large parasitic capacitance of the photodiodes, a decoupling from the WTA circuit has been necessary to achieve transient response times in the ms-range. The measures to accelerate the transient response are described in detail below. An output amplifier converts the binary output current into a binary output voltage. Care had to be taken, to realize the desired functionality with a minimum number of transistors and thereby maximize the photosensitive area.

4.4.1 Acceleration of the Transient Response

The parasitic capacitance in a row of pixels amounts to approximately 110 fF, which would result in a transient response time of approximately 20 ms at 20 pA photocurrent difference with the basic WTA circuit from Chapter 3.5.2. This transient response time is far above the specified response time of at most 1 ms. Different measures of acceleration of the WTA circuit have been evaluated by Droste [DB02], among them are a decoupling of the photodiode's parasitic capacitance from the WTA circuit, an initialization of the WTA circuit by a reset mechanism and a positive feedback with controllable gain to boost the charging of the parasitic gate-source capacitance of the flow-transistor. Droste showed that the reset mechanism accelerates the transient response, but also introduces noise to the circuit. The positive feedback has only been

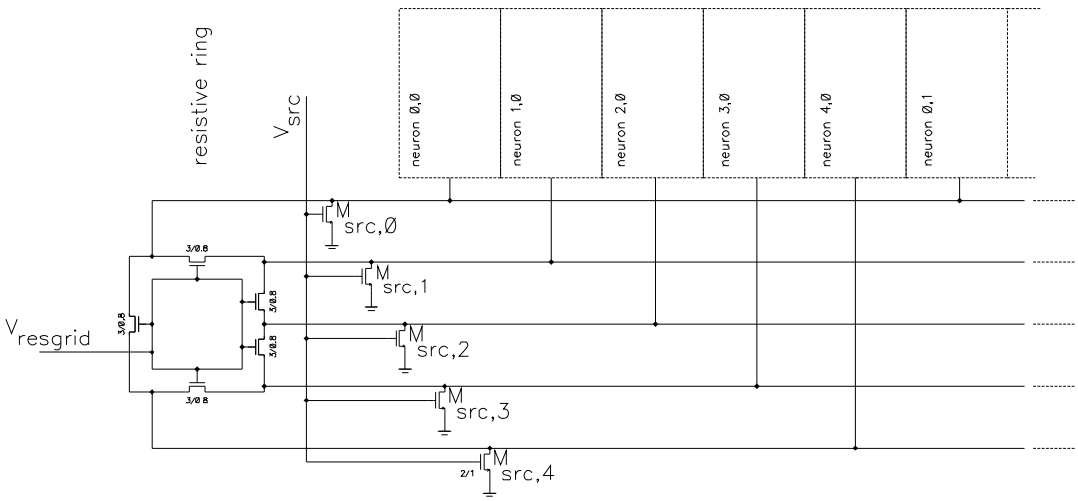

Figure 4.5: Schematic of the resistive-ring network of WTA circuits. The sense-nets of the individual neurons are connected inter-digitally to five current sinks and the ring of MOS transistors, which act as active resistors. When the transconductance of the transistors is large, the individual sense-nets almost form a single sense-net. Note, that only the first six out of 21 neurons are shown here.

possible to implement together with $p+/n$-well diodes, which are unfavorable due to their small quantum efficiencies and large parasitic capacitances.

In the CeHSSA, transient response times in the ms-scale have been achieved through the use of n-well/substrate photodiodes with an approximately ten times smaller parasitic capacitance than diffusion photodiodes, and a decoupling of the WTA circuit from the photodiodes through a single-transistor cascode and a cascoded current mirror. The use of the AMS 0.35 μm process technology instead of AMS 0.6 μm reduced the parasitic gate-source capacitances of the transistors in addition. All these measures together achieved an acceleration of the transient response of several orders of magnitude, with respect to the simple WTA circuit.

The complete schematic of a neuron is shown in Figure 4.6. The single-transistor cascode M_{cas} decouples the cathode of the row of photodiodes from the source of transistor M_{c3} of the cascoded p-type current mirror. The photocurrent I_p is mirrored to the input current of the WTA circuit.

Decoupling of the Junction Capacitance

The simple cascode consists of the transistor M_{cas}, which is held at a constant gate potential V_{cas}. The potential at node Ψ is defined by the gate-source voltage $v_{GS}(M_{\text{cas}})$. A change of the drain-source voltage $v_{DS}(M_{\text{cas}})$ leads to a change of the potential at node ξ

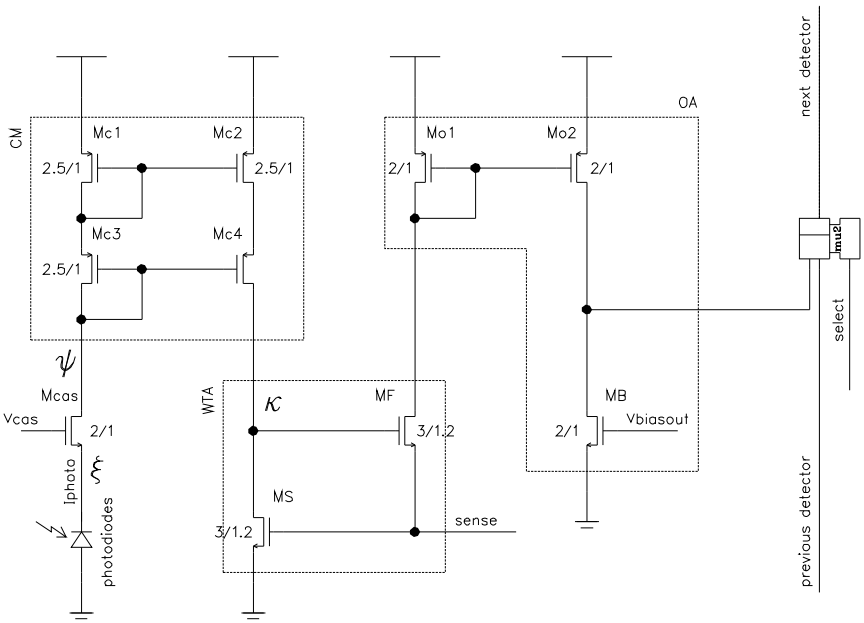

Figure 4.6: Schematic of the implemented WTA neuron. It contains a single-transistor cascode M_{cas} connected to the photodiode row, a cascoded current mirror (CM) and an output amplifier (OA) besides the sense- and flow-transistor of the WTA circuit. The digital output of the amplifier is connected to the multiplexers of the digital data bus.

$$\begin{aligned}\frac{dv(\xi)}{dv(\Psi)} &= \frac{dv_{GS}(M_{\text{cas}})}{dv(\Psi)} \\ &= \frac{\lambda n V_t}{1+\lambda(dv(\Psi)-dv(\xi))}\end{aligned}$$

or approximately

$$dv(\xi) = \lambda n V_t dv(\Psi) \tag{4.2}$$

for $\lambda \ll 1$. The gate-source voltage here has been derived from Equation 2.3

$$\begin{aligned}dv_{GS}(M_{\text{cas}}) &= nV_t \ln \frac{I_p}{I_{D0}\frac{W}{L}(1+\lambda dv_{DS}(M_C))} \\ &= nV_t \ln \frac{I_p}{I_{D0}\frac{W}{L}(1+\lambda(dv(\Psi)-dv(\xi))}.\end{aligned} \tag{4.3}$$

A numeric calculation for the implemented single-transistor cascode with $W/L = 2/1$ yields a reduction of the effective capacitance

$$C_j' = \lambda n V_t C_j \tag{4.4}$$

at node Ψ of C_j'/C_j =0.051, i.e. the effective parasitic capacitance is reduced 28.1 times from 110.0 fF to 5.6 fF. The parameters for the calculation have been derived from Spectre simulations in the subthreshold region to be $n = 1.53$ and $\lambda = 0.051$ V^{-1}. The subthreshold voltage $V_t = 0.51$ V has been taken from the manufacturer's process parameters [AMSb]. A change of the photocurrent ΔI of 20 pA only leads to a change of $\Delta V = 0.10$ V of the potential at node ξ. The charging time Δt of the parasitic capacitance in a linear approximation is then

$$\Delta t = \Delta V C_j / \Delta I, \tag{4.5}$$

which evaluates to 0.55 ms. An even larger decoupling through a regulated cascode would achieve a reduction of the effective capacitance of $C_j'' = (\lambda n V_t)^2 C_j$ [Dro99], but is not necessary in our case, as the charging time is already below 1 ms. Figure 4.7 shows the achievable bandwidth with the parasitic capacitance at node ξ and Figure 4.8 with the photocurrent difference. The Spice simulation results of the circuit follow the expected behavior from Equation 4.5.

The parasitic junction capacitance is not the only factor, that limits the bandwidth. An upper limit is also introduced by the parasitic gate-oxide capacitance $C_G(M_F) = 16.5$ fF of the flow-transistor at node κ, which is charged or discharged, when the neuron changes its state. It has a relatively large effect, because the voltage swing at node κ is large. The parasitic drain-source capacitances $C_D(M_S)$ and $C_D(M_{C4})$

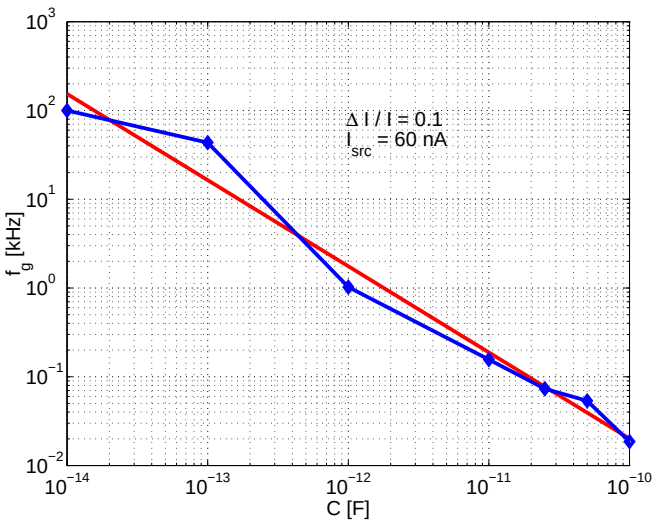

Figure 4.7: Bandwidth with parasitic capacitance at 10% photocurrent difference from a Spice simulation (marked line) and analytical slope (straight line).

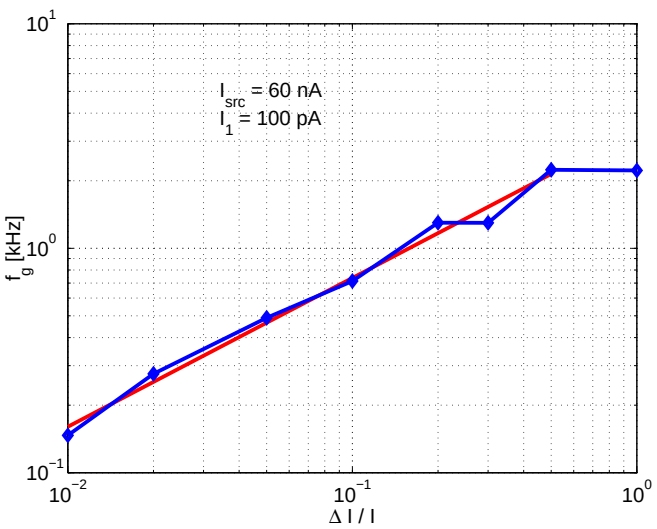

Figure 4.8: Bandwidth with photocurrent difference from a Spice simulation (marked line) and analytical slope (straight line).

can be neglected here. From the voltage swing of 696 mV from the DC operation point calculation results a charging time $\Delta t(C_G) = 0.58$ ms at 20 pA photocurrent difference. The total transient response time

$$\Delta t = \Delta t(C_j) + \Delta t(C_G) \tag{4.6}$$

to charge the parasitic capacitance of the photodiodes and the parasitic gate-source capacitance is 1.13 ms, which corresponds to a bandwidth of 0.89 kHz. A bandwidth larger than 1 kHz is thus obtained for photocurrent differences larger than approximately 22.6 pA.

Current Mirror

Current mirrors produce one or multiple copies of an input current. Figure 4.9 shows three common current-mirror types. A cascoded current mirror has been implemented to further decouple the drain of the sense-transistor from the photodiodes. Ideal current mirrors have an output current, which is independent of the potential at the output terminal and thus act as ideal current sources. Real current mirrors instead, have a finite output resistance due to the modulation of the channel length.

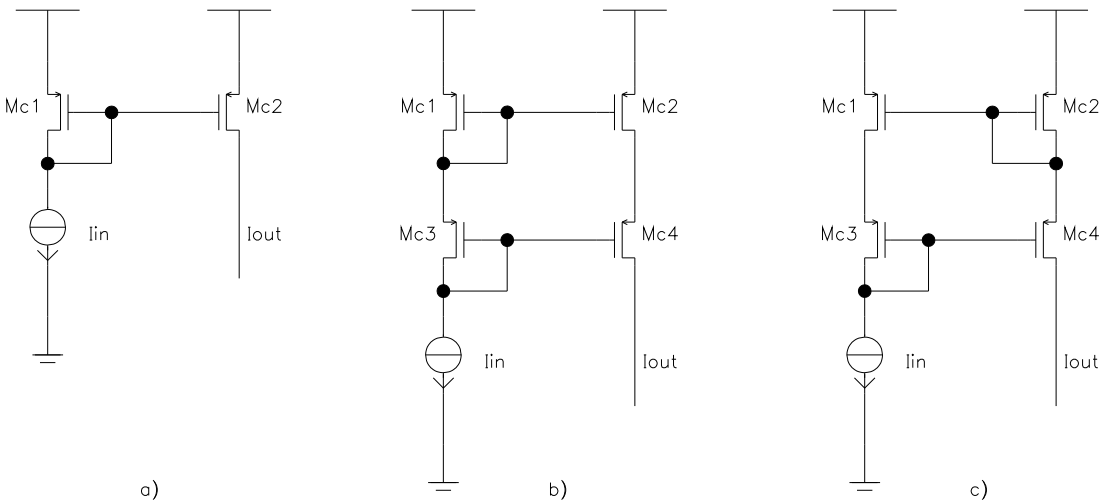

Figure 4.9: Common p-type current mirrors, a) simple current mirror, b) cascode current mirror, c) Wilson current mirror.

The simple current mirror consists of the two transistors M_{c1} and M_{c2}. The drain current $I_D(M_{c1})$ defines the gate-source voltage $V_{GS}(M_{c1})$. In case of identical transistors the same gate-source voltage $V_{GS}(M_{c2}) = V_{GS}(M_{c1})$ is present at M_{c2}, which therefore has the same drain current $I_D(M_{c2}) = I_D(M_{c1})$ and the current gain is unity. The current gain can be controlled arbitrarily through the choice of the (W/L) ratios

of the transistor pair

$$\frac{I_2}{I_1} = \frac{(W_{c2}/L_{c2})}{(W_{c1}/L_{c1})}. \tag{4.7}$$

The small-signal output resistance r_0 of M_{c2} is not independent of the source potential due to the channel-length modulation. It can be calculated as the reciprocal of the slope of the output characteristics, which is

$$r_0 = \frac{1}{\lambda I_D(M_{c2})} \tag{4.8}$$

at the operating point. The cascode current mirror achieves an even larger output resistance through the addition of two transistors M_{c3} and M_{c4}, which form a cascode with the first transistor pair. The output resistance is given by

$$R_0 \simeq g(M_{c4}) r_0^2 \tag{4.9}$$

with the transconductance $g(M_{c4})$. For the CeHSSA's neurons the small-signal resistance r_0 at 1 nA evaluates to 19.6 GΩ, calculated from Equation 4.8 with $\lambda = 0.051$ V^{-1}. The transconductance evaluates to $g = R^{-1}$ =29.9 nS from Equation 2.7 and the implemented cascode current mirror with unity gain has an output resistance of 11.5 TΩ, according to Formula 4.9. These very large resistances result from the operation in the subthreshold region and are not that large for larger drain currents.

4.4.2 Output Amplifier

Although the output of the WTA circuit is binary in terms of current, the output has to be converted to a binary voltage, suitable as an input for digital logic cells. A p-type common source amplifier, which uses the mirrored current through the flow transistor M_F as the current load, is used as an output stage. Transistor M_B has to be in saturation, when $I_D(M_B)$ reaches I_{src} for the winning neuron, i.e. the bias gate-source voltage $V_{GS}(M_B) = V_{\text{biasout}}$ has to be chosen such, that

$$I_{\text{src}} > \frac{k'}{2}\frac{W}{L}(V_{GS} - V_t)^2. \tag{4.10}$$

The right side in the above equation is the drain current in the saturation region from Equation 2.1, when the channel-length modulation is neglected. For the loosing neuron, transistor M_B has to be in the subthreshold region. Figure 4.10 shows the amplifier output with $I_D(M_B) = I_{\text{src}}$ for two different bias voltages $V_{GS}(M_B)$. The steep transition from $V_{DS}(M_B) = 0$ to $V_{DS}(M_B) = V_{dd}$ is due to the large output resistance in the saturation region. For the desired current of $I_{\text{src}} = 100$ nA in the CeHSSA, a bias voltage of 400 mV$\pm$100 mV has been a reasonable value.

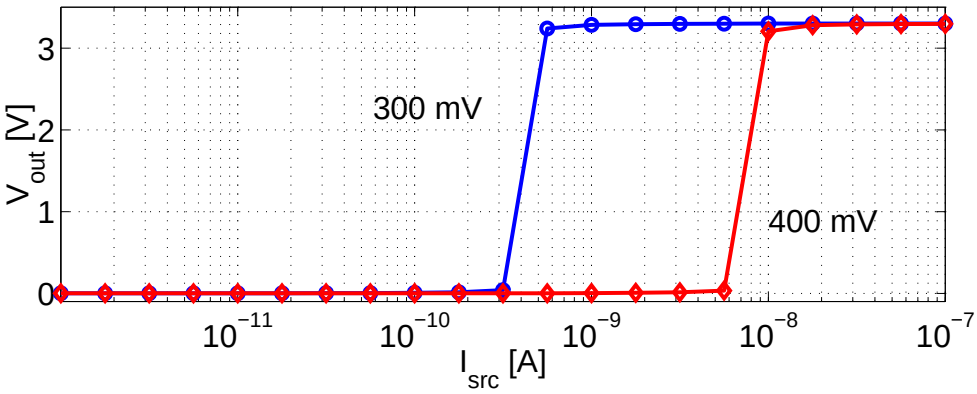

Figure 4.10: Amplifier output voltage with input current for two different bias voltages (Spice simulation).

4.4.3 Detector Layout

Each of the 21 neurons of a WTA ring receives the photocurrent from the pixel array and provides an output in form of a binary voltage. Besides the analog circuitry, the neuron contains a digital 2-to-1 multiplexer (MUX) to switch between the state of the neuron of the previous detector or the state of the neuron, when addressed by the row or column select line. Analog and digital circuitry have separated ground connections to avoid cross-talk between analog circuits and digital logic cells. The external bias voltages V_{cas} and V_{biasout} control the operation point of the cascode and the output amplifier.

Layout views of the neuron can be found in Figure 4.11 and a larger overview in Figure 4.12. Figure 4.13 shows micro-photographs of the chip. Two large separate n-wells extend through all neurons in a row, one for the analog PMOS transistors and one for the PMOS transistors of the multiplexers. Besides analog ground $gnda$ and digital ground gnd the positive supply voltage V_{dd} has to be distributed across the chip, for which the third metal layer (MET3) is used.

This layer is also used to shield the underlying transistors from incident light. The effect of an aluminum metal shielding has been discussed and measured by Teodori [Teo97] for the AMS 0.8 μm process. Transmittance and attenuation have been measured to be 1.14·10^{-3} and 1.78 μm^{-1} respectively at 575 nm. For the AMS 0.35 μm process, where the thickness of the MET3 layer is 925 nm [AMSb], the total attenuation through the MET3 layer is then

$$\frac{J}{J_0} = Te^{-\alpha z} = 0.67 \cdot 10^{-3}. \tag{4.11}$$

Absorption is negligible here, but the transmittance is low enough for the circuit to be sufficiently shielded from the incident light.

Mismatching Considerations

Some layout details to reduce mismatching effects are described here, see also Chapter 2.2.2 for some general aspects of mismatch. Transistors, which are sensitive to mismatch in the layout of the neurons, are the cascode transistor M_{cas}, the transistors of the cascoded current mirror and the sense- and the flow-transistors M_S and M_F of the WTA circuit, because they operate in weak inversion. The transistors of the output amplifier and the digital cells work in strong inversion where mismatching effects are negligible.

Mismatching effects have been reduced by the use of large transistors in the analog circuitry with a W/L of at least 2/1. The mismatch of the drain currents therefore is at least reduced by a factor 3.37 with respect to a transistor with minimum gate length. Even larger devices would have reduced the bandwidth, because the parasitic capacitances increase with the gate area. Not all possible layout techniques for mismatching reduction have been implemented in this layout due to the limited available area, i.e. dummy cells have not been placed. Guard rings have been placed, wherever space requirements allowed it and the current mirrors have been placed such, that the current flows in the same direction.

Nevertheless, a quantitative analysis has been done with the Monte-Carlo (MC) simulation tool of the Cadence design framework to assure, that mismatching effects can indeed be neglected. Common process parameter variations have been statistically varied and a simulation then repeated for a large number of runs. The variations in the simulation results are quantitative measures for the expected mismatching effects. These simulations have been conducted in centroid and maximum mode for a 100 μm FWHM spot at different spot intensities. The frequency of bit-errors n_f in this simulation is shown in Table 4.3. As a result, mismatching is negligible in centroid mode for spot powers in the 100 pW range and above. In maximum mode mismatching effects only disappear at two orders of magnitude larger spot powers. The MC simulations clearly show the advantage of the resistive-ring network of WTA circuits over the simple WTA circuit topology.

P/spot	$\max(I_p)$	n_f(max)	n_f(cent)
2.8 pW	100 fA	46%	48%
28 pW	1 pA	5%	13%
280 pW	10 pA	9%	0%
2.8 nW	100 pA	5%	0%
28 nW	1 nA	0%	0%

Table 4.3: MC simulation (100 runs) of the frequency of bit-errors for a photocurrent distribution from a spot with 100 μm FWHM at different spot powers in centroid and maximum mode.

Figure 4.12 shows four detectors, with the circuitry in between them. The pixel

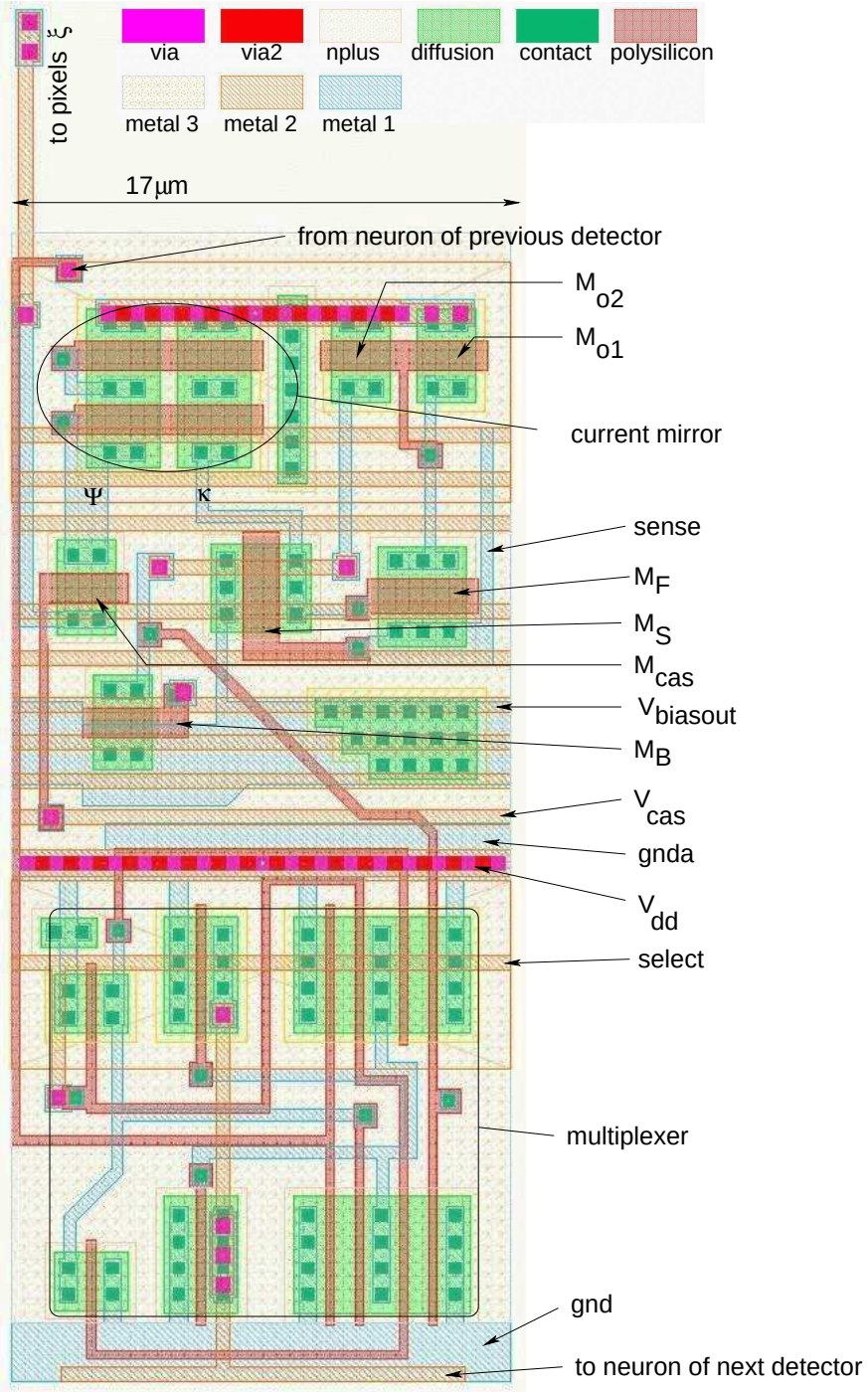

Figure 4.11: Layout of a neuron, including analog circuitry and multiplexer.

array are completely surrounded by guard-rings to avoid, that free charge carriers drift to the analog circuitry. The space in between four detectors has been used to place the current sinks and the transistors for the resistive-ring network. These transistors are placed as close together as possible.

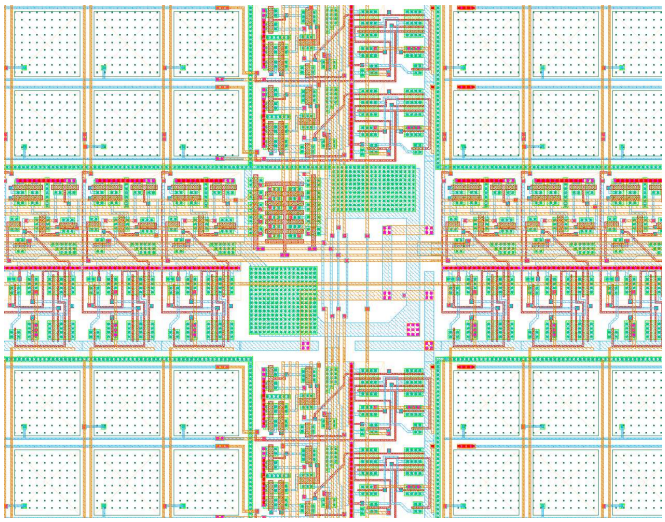

Figure 4.12: Layout of four adjacent detectors. The third metal layer (MET3), which covers all the area except the photodiodes, is not shown here.

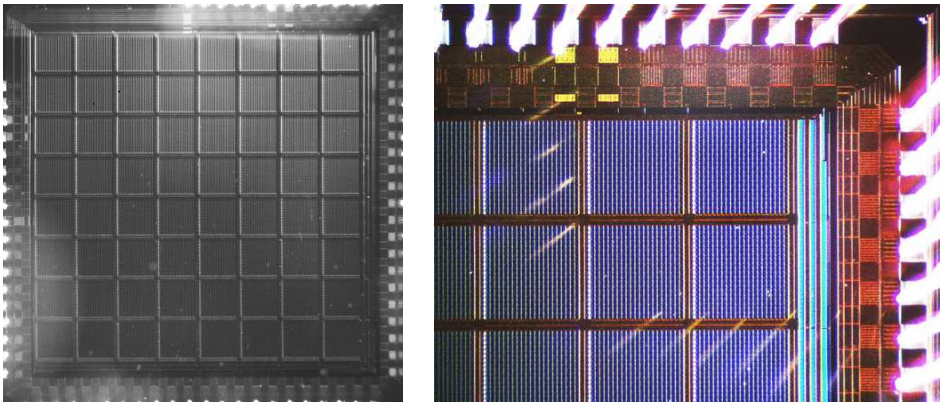

Figure 4.13: Micro-photographs of the complete CeHSSA chip (left) and upper right corner (right). The die has an overall size of 4.08 mm$\times$4.08 mm.

Resistive-Ring Network and Current Sinks

The 21 neurons are connected to the five current sinks according to the mapping

$$I_D(M_{F,i}) \rightarrow I_{\text{src},n} \text{ with } n = i \mod 5.$$

The current sinks themselves are interconnected through five active resistors in form of n-type MOSFETs. The sunk currents are distributed between the winning neurons according to the relation

$$I_D(M_F) = \frac{n}{m} I_{\text{src}}, \tag{4.12}$$

where n is the number of current sinks (five in this case) and m is the (average) number of winning neurons. The resistance of the active resistors is controlled by the external bias voltage V_{resgrid}, according to Formula 2.6. When V_{resgrid} approaches zero, the sense-nets are independent from another and the PSD works in centroid mode ($l \rightarrow \infty$). For V_{resgrid} approaching V_{dd} the diffusion length approaches zero ($l \rightarrow 0$) and the PSD works in maximum mode. For intermediate resistances the diffusion length is finite $(0 < l < \infty)$ and the number of winning neurons lies between one and five.

The current I_{src} is controlled by the external bias voltage V_{src}. If it is smaller than the threshold voltage, the current sinks work in the subthreshold region and Equation 2.3 applies

$$I_{\text{src}} = \frac{W}{L} I_{D0} e^{\frac{V_{\text{src}}}{nV_T}} \left(1 - e^{-\frac{V_{dd} - V_t - V_{DS}(M_F) - V_{DS}(M_{o1})}{V_T}}\right). \tag{4.13}$$

The average overall DC power consumption of the chip is dominated by the current $I_{\text{tot}} = 8 \cdot 8 \cdot 2 \cdot (5 + 5) \cdot I_{\text{src}}$ through the current sinks and the current through the output amplifiers, therefore I_{src} should not exceed approximately 1 μA. Each winning neuron causes a current I_{src} to flow through its output amplifier, therefore factor 2 has to be included in the calculation. I_{src} should be chosen to be about 100 nA and the average power consumption then is 0.42 mW. The worst case power consumption with $P_{\text{tot}} = 8 \cdot 8 \cdot 2 \cdot (5 + 21) \cdot I_{\text{src}} \cdot V_{dd} = 10.98$ mW, which should not appear in normal mode of operation, occurs, when all output amplifiers are high and additionally I_{src} is chosen to be 10 times larger than in normal operation. Although operating the sinks in the linear region would reduce the effect of mismatching, the subthreshold region is preferable to reduce the overall power consumption.

Figure 4.14 shows the layout of the current sinks and the active resistors, which are located in between the x- and y-WTA circuit blocks. Extra dummy polysilicon gates have been placed besides the active resistors. Note, that matching of all transistors only plays a role for intermediate diffusion lengths through the variation in sunk current $\sigma(I_{\text{src},n})$ and resistance $\sigma(R_n)$, but not in pure maximum or centroid mode.

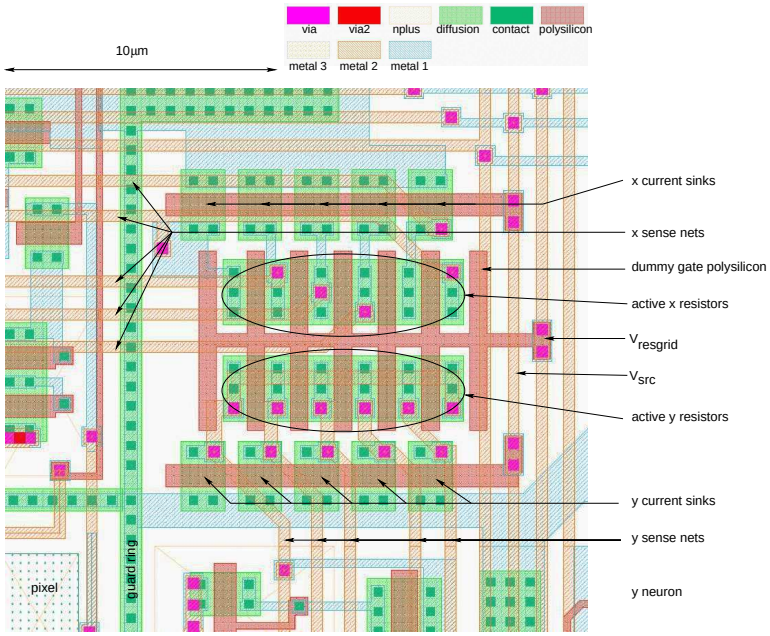

Figure 4.14: Layout of the current sinks and the resistive ring. The third metal layer (MET3), which covers all the area except the photodiodes, is not shown here for better visibility.

4.5 Position-Data Read-Out

4.5.1 Address Decoder

A random read-out scheme with random access to the detectors has been implemented, because in many cases it is not necessary to read the positions of all spots out, e.g. when only a circular subaperture is evaluated. By decreasing the number of read-out detectors the frame rate can thus be increased. Therefore an addressing of the individual detector through two 3-bit column and row address decoders has been implemented, instead of the second possibility for addressing, which would have been a shift register.

A specific detector is selected by two 3-to-8 decoders, one for the x- and one for y-coordinate, see Figure 4.15 for a schematic. For each address bit A_i with $i = 0 \ldots 2$ the inverted bit $\overline{A_i}$ is generated and these signals are connected to the inputs of eight AND-gates, such that the address lines have the logic state

$$S_{x,m} = \delta_{m,s}\overline{H} \;\text{ with }\; s = (A_{x,0}2^0 + A_{x,1}2^1 + A_{x,2}2^2) \tag{4.14}$$

and

$$S_{y,n} = \delta_{n,s}\overline{H} \;\text{ with }\; s = (A_{y,0}2^0 + A_{y,1}2^1 + A_{y,2}2^2) \tag{4.15}$$

with the column index $m = 0 \ldots 7$ and the row index $n = 0 \ldots 7$. The symbol δ_{ij} denotes Kronecker's delta function. The additional hold signal H has been introduced to disable all multiplexers in the detectors and thus conserve the logic state of the tristate bus.

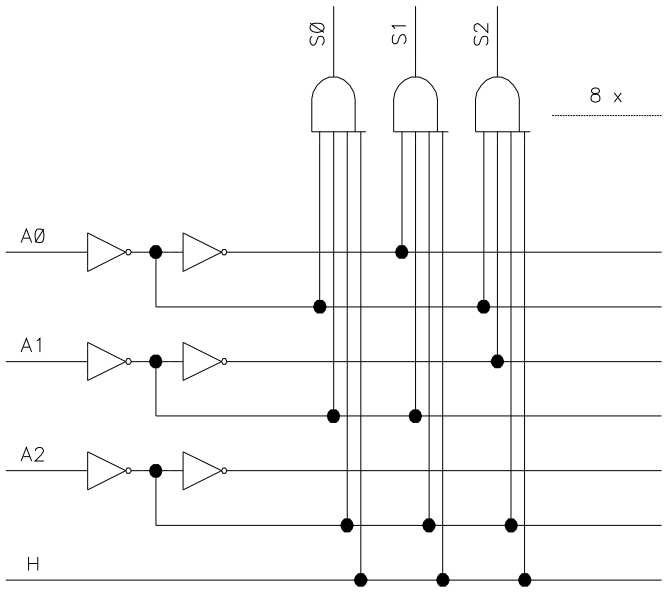

Figure 4.15: Implemented 3-bit address decoder with additional hold signal. Only the first three out of eight AND-gates are shown here.

4.5.2 Data Read-Out

Each neuron contains a 2-to-1 multiplexer to switch the logic state of the neuron from the previous detector $X_{\mathrm{out},m-1}$ or its own state X_m to the input of the multiplexer of the next detector with status X_{m+1}, when m is addressed through $S_{m,n}$. Figure 4.16 shows the implemented multiplexer, which is a redesigned version of an AMS standard cell [AMSc]. It performs the logic operation

$$Y_{\mathrm{out},m} = Y_m S_{x,m} + Y_{\mathrm{out},m-1}\overline{S_{x,m}} \tag{4.16}$$

or

$$X_{\mathrm{out},n} = X_n S_{y,n} + X_{\mathrm{out},n-1}\overline{S_{y,n}} \tag{4.17}$$

for the y-coordinate. The additional indexation of the 8·21 neurons has been left out here for better readability. The neurons of the eight detectors form a chain of multiplexers, with eight elements each, see Figure 4.17 for a schematic. The 8×21

outputs of the last multiplexers of the chains are connected to the inputs of *tristate buffers*, which set the state of two 21-bit *tristate buses*[3] T_x and T_y with

$$T_x = X_{\mathrm{out},7} S_{x,n} \tag{4.18}$$

and

$$T_y = Y_{\mathrm{out},7} S_{y,m} \tag{4.19}$$

for the addressed detector (m, n). When the hold signal H is high, the state of the bus is preserved, because all S_x and S_y are low then. Thus the data of a the selected detector is available within 42 bits at the periphery of the detector array.

Although the chip has been placed within a ceramic chip carrier with 84 pins (CLC84) with sufficient pins for parallel read-out, it is much more convenient to read 32 bits or 16 bits out in parallel. Most data acquisition hardware only allows 16 bits width, including the data acquisition card used in this thesis. To allow read-out with less bit count, the 42 data bits can be multiplexed to either 32 bits or 16 bits width. In 32-bit read-out mode the lowest 16 bits (0. . . 15) of the data are multiplexed to the output pins and afterwards the uppermost 16 data bits (5. . . 21). This example shows how a 21-bit word

0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0

is split into the two 16-bit words

0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0

and

0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0

.

The external signal HLM is used to switch between the bits, according to

$$P_{x,u} = T_{x,u} \overline{HLM} + T_{x,u+5} HLM \tag{4.20}$$

and

[3]Tristate buffers are logic cells that provide a high impedance output state Z in addition to the logic states high H and low L. Several tristate buffers can be connected to the same output line of a bus, as long as only one of them is not tristated at the same time. If all buffers are tristated, the bus conserves the previous state through an additional element, which is a *signal-hold*. It consists of two cross-coupled inverters. The drive strength of the tristate buffers has to be large enough to alter the state of the signal-hold. It has to be secured, that only one buffer is not tristated at the same time. Otherwise the state of the bus would be undefined.

$$P_{y,v} = T_{y,v}\overline{HLM} + T_{y,v+5}HLM \tag{4.21}$$

with bit indices $u = 0 \ldots 15$ and $v = 0 \ldots 15$. A reduction from 32 bits to 16 bits width is available through another multiplexer level, that switches the x- and the y-data to the same output pads through the external signal XYM

$$Q_w = P_{y,w}\overline{XYM} + P_{x,w}XYM \tag{4.22}$$

with the bit index w. The output state at the pins is thus a static logic state of the state of the neurons of the addressed detector as shown in Table 4.4.

HLM	XYM	data at pins
L	L	bits 0. . . 15, x-data
L	H	bits 5. . . 20, x-data
H	L	bits 0. . . 15, y-data
H	H	bits 5. . . 20, y-data

Table 4.4: Multiplexer scheme for read-out in 16-bit mode.

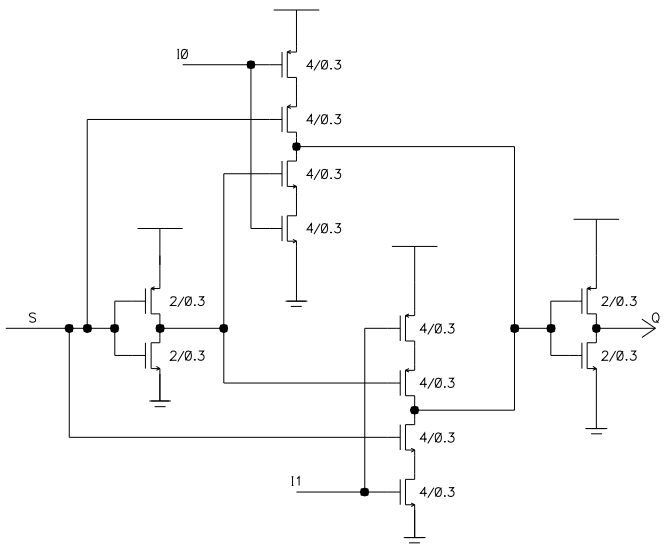

Figure 4.16: Implemented 2-to-1 multiplexer. The multiplexer performs the logic operation $Q = I_0S + I_1\overline{S}$.

4.5.3 Communication Test

A test bit-pattern can be loaded to the x-tristate bus through an external serial signal. When the pattern read out subsequently is the same as the written pattern, communi-

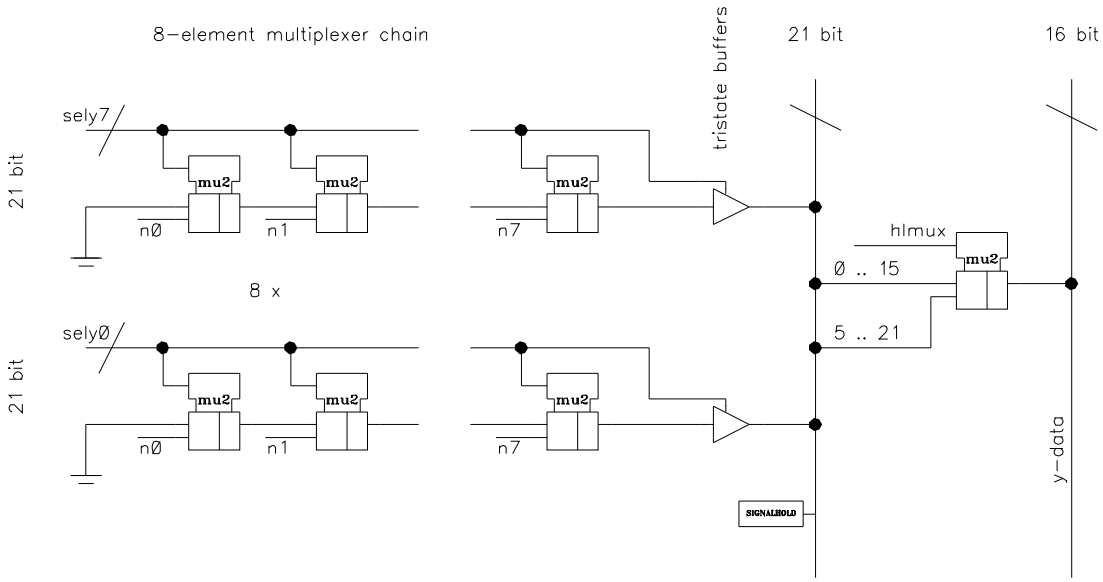

Figure 4.17: Read-out scheme of the y-position data. The output of each neuron is connected to an 8-element multiplexer chain. All bits from the selected row (logic high on one of the S_y address lines) are written to a 21-bit tristate data bus through tristate buffers. These 21-bits are multiplexed to another 16-bit data bus for read-out or further multiplexing with the x-position data. A logic low transmits bits 0 to 15, a logic high transmits bits 5 to 21.

cation with the chip is assured to work.

The bit-pattern is loaded serially to a 21-bit D-flip-flop shift register, which is driven by an external clock signal $TBClk$, see Figure 4.18 for a timing diagram. With each rising edge of the clock the state of the input signal $TBIn$ is loaded to the first shift register. After 21 clock cycles the bit-pattern is written to the x-tristate bus by setting the external signal $TBSelect$ to high. Additional AND-gates have been placed between the output of the 3-bit x-decoders and the eight address lines to the detector array to avoid, that data from the shift register and the sensors is written to the tristate bus at the same time. The hold signal H is set, when $TBSelect$ is high or $BHold$ is high. The data transfer from the detector array to the tristate bus is disabled then.

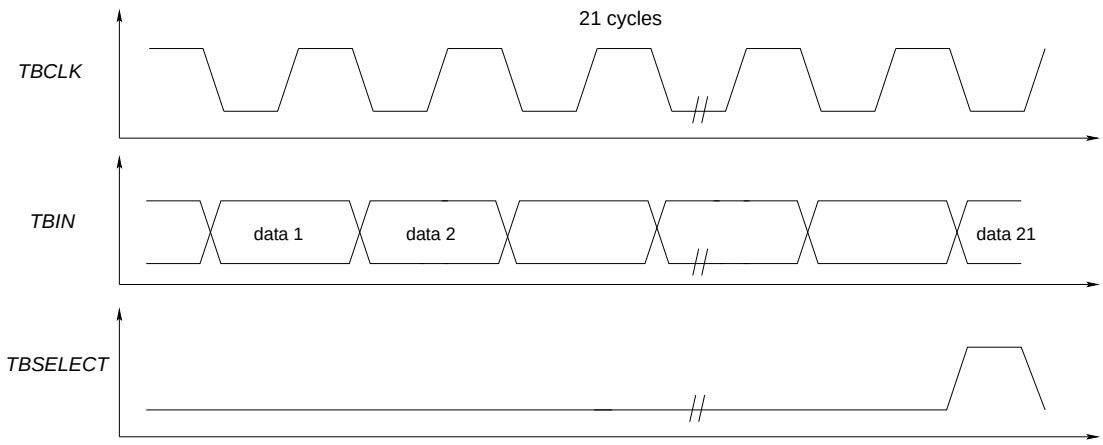

Figure 4.18: Timing diagram for loading the test bit-pattern to the tristate bus. With each rising edge of $TBClk$ the state of $TBIn$ is loaded to the shift register. After 21 cycles it is written to the tristate bus through $TBSelect$.

4.6 The Prototype

A prototype ASIC has been designed and tested in AMS 0.35 μm technology, for the purpose of evaluating the quantum efficiency of different junction photodiodes, evaluating the resistive-line network and the resistive-ring network of WTA circuits for spot tracking. The results of this evaluation have been used to optimize the final sensor chip.

4.6.1 Quantum Efficiency and Spot-Size Measurements

Because the red and near-infrared spectrum is of special interest for ophthalmic applications, the quantum efficiency at 680 nm has been measured with a laser spot from a diode-laser on large area photodiodes (300 μm$\times$500 μm). They have been surrounded

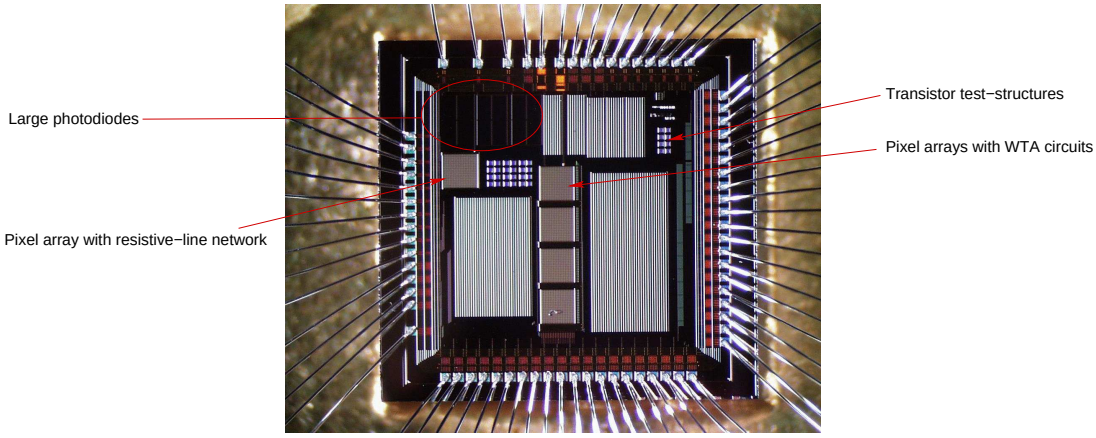

Figure 4.19: Micro-photograph of the prototype with bond wires. The die has an overall size of 3.25 × 3.25 mm^2.

by guard-rings, so that free charge carriers from outside the diode's area cannot contribute to the measured photocurrent. To allow the measurement of the spot diameter, stripes of 10 μm wide metal layer (MET1) have been placed on the diodes. The chip itself has been mounted on a PC controlled x-y-translation stage with μm accuracy for this purpose. The cathode of the n+/substrate and n-well/substrate photodiode, and the anode of the p+/n-well photodiode are directly connected to analog output pins for external current measurements.

Table 4.5 shows a comparison between the measured quantum efficiencies for the scaled processes AMS 0.35 μm and AMS 0.6 μm. The scaling effect, i.e. the shift of the maximum of spectral quantum efficiency to shorter wavelengths, is clearly visible from the data. In the chosen AMS 0.35 μm process only the n-well/substrate photodiode with the deepest junction[4] has a sufficient quantum efficiency.

photodiode type	η(0.6 μm)	η(0.35 μm)
n+/substrate	0.49	0.15
p+/n-well	0.14	0.04
n-well/substrate	0.54	0.40

Table 4.5: Measured quantum efficiencies for AMS 0.6 μm [Dro99] and AMS 0.35 μm for different junction photodiodes at 680 nm and 50 μW spot power.

The photodiode structures can also be used to measure the spot size. When a spot moves with constant velocity v over the photodiode, the photocurrent from the

[4]The n-well junction depth is 1200 nm, the p+ and n+ junction depths are 200 nm. The passivation layer thickness is 900 nm [AMSb].

photodiode can be calculated as the convolution of the spot's intensity distribution $J(x) = J(vt)$ and the spatial geometry $\eta(x)$ of the photodiode

$$I_p(x) \sim \int_{x_0=-\infty}^{\infty} J(x - x_0)\eta(x_0)dx_0. \tag{4.23}$$

A least-squares-fit of the intensity distribution from Equation 1.4 and the measured photocurrent yields the spot size, see Figure 4.20 for an example.

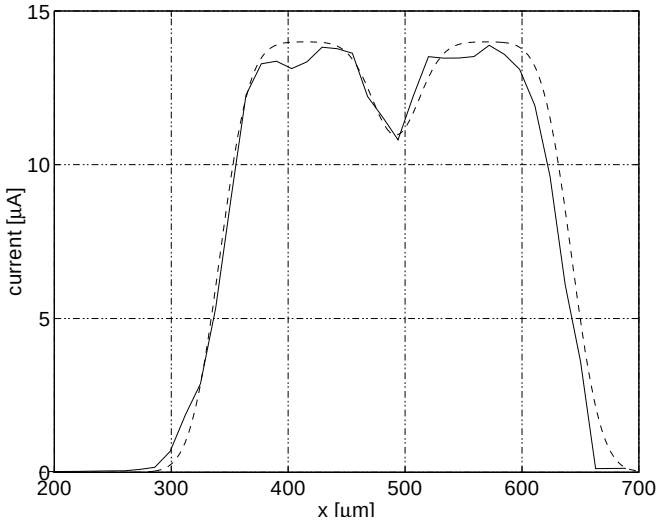

Figure 4.20: Measured photocurrent (solid line) from a spot moving with 100 μm/s and convolution of the geometry with an intensity distribution of a spot with 20 μm FWHM (dashed line). The trench in the middle is a result of the metal stripe, shielding a part of the incident light.

4.6.2 Position Sensitive Detector

Unlike the CeHSSA the prototype's 20×20 pixels contained two photodiodes, one for row and one for column currents, see Figure 4.21 for the layout. As photodetectors $n+$/substrate photodiodes have been used.

Only five detectors have been implemented in the prototype for evaluation purposes of spot tracking accuracy. One of the detectors contains a resistive-ring of WTA circuits with five elements. An additional dummy neuron has been added to the WTA circuit of this detector. The input current of this dummy neuron is provided by an additional current sink, which is controlled by the external voltage V_{dummy}. When the dummy current is larger than the largest photocurrent from the photodetector array, the dummy neuron wins. This concept has been used in the HSSX to reject missing spots [Dro99]. The other three pixel arrays are connected to an inter-digital WTA circuit, with no connection through a resistive-ring network.

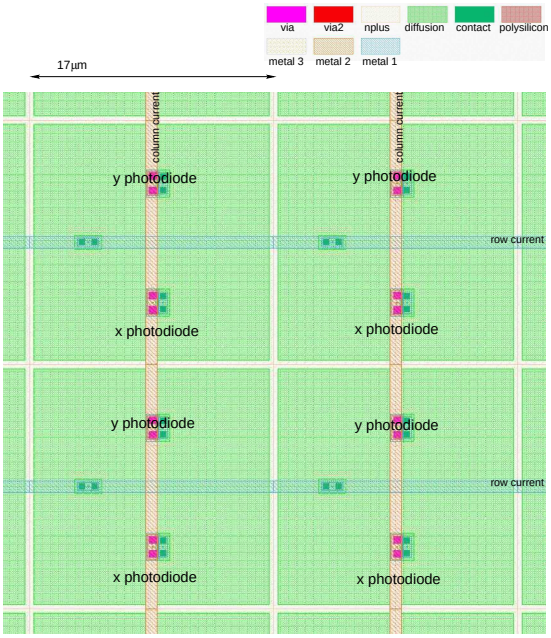

Figure 4.21: Layout of the prototype's pixel array. Each pixel contains two $n+$/substrate photodiodes.

4.6.3 Resistive-Line Network

Spot tracking evaluation with a resistive-line network has been possible through a 5th pixel array and a resistive-line network of active MOS transistors, see Chapter 3.5.1 for details about the concept of the resistive-line network. Figure 4.22 shows a schematic of the resistive-line network. The resistance is controlled by the external voltage V_{resgrid}. Rather large transistors $W/L = 2/4$ have been used here to reduce mismatching effects. The four output currents I_L, I_R, I_U and I_D of the resistive line are connected to analog output pins, where they are measured with a semiconductor parameter analyzer for off-line spot position calculation according to Equation 3.36.

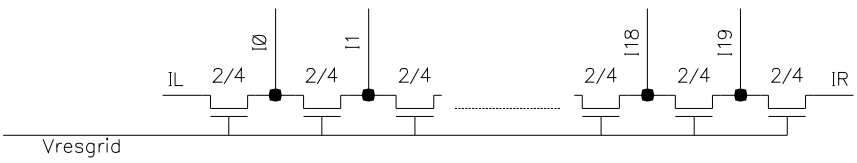

Figure 4.22: Schematic of the resistive-line network of active MOSFET resistors.

4.7 Conclusions from the Prototype

The evaluation results of different types of photodetectors and different types of PSDs have been used to optimize the final Hartmann-Shack-Sensor ASIC. First of all, the quantum efficiencies in the 0.35 μm process technology differed to a large scale from the ones in the 0.6 μm process technology. Scaling effects are clearly seen and as a result the n-well/substrate photodiode is the only adequate choice for the photodetectors in the red and near infrared spectrum.

Second, it has been shown, that both the resistive-line network and the WTA circuits are capable of tracking the spots, see Chapter 6.2 for details about the spot tracking performance. Although the resistive-line network offers theoretically an infinite resolution and basic spot tracking has been successfully conducted, the ambiguity of the pure centroid operation and the sensitivity to background noise, the need for on-chip current amplification and A/D conversion defavor the resistive-line network. The direct digital output, linearity and easy read-out strongly favor the WTA circuit or one of its derivatives for spot tracking. Finally, the resistive-ring of WTA circuits has been chosen as a PSD, because it allows adaptation to different spot sizes with little extra circuitry.

5 External System and Data Analysis

In this Chapter the software and the custom Linux device driver are shortly described first. In the second part the analysis of the bit-vector from the detectors is described and how unuseful measurements can be rejected from the wavefront calculation.

The CeHSSA input and output is CMOS compatible and can be connected with any other CMOS compatible hardware. A fast closed-loop AO system could be realized with an FPGA or a microcontroller, that reads the spot position data directly from the sensor and calculates the signals to drive an adaptive mirror. For the evaluation of the chip performance a simpler system has been built up with a commercial PCI-based *Data Acquisition (DAQ)* card[1] though. Besides the digital address words and the two signals for the multiplexers, the chip needs to be provided with four analog bias voltages and the supply voltage of 3.3 V, all available with the DAQ card except the supply voltage. Analog and digital outputs are controlled by software, described in detail below.

5.1 Software and Custom Device Driver

The operating system Linux[2] and the object oriented programming language C++ have been chosen to drive the DAQ-card, process the data and calculate the wavefront. Software in Linux is in general open-source and available under the GNU public license, which means that the source-code is published and the programmer can use and modify it for his own purposes. This was an essential point in the design of the readout system, because the driver provided by the manufacturer of the DAQ-card had to be modified to achieve the desired bandwidth of some hundreds of Hz.

The software is organized in three levels, of which the uppermost is a *Graphical User Interface (GUI)*, based on the portable graphics library Qt [3]. Figure 5.1 shows

[1] Meilhaus ME-2600. The PCI DAQ-card contains four D/A converters, four A/D converters and two 16-bit digital ports, which can be programmed as input or output ports separately.

[2] SUSE Linux 7.3 with kernel version 2.4.10.

[3] Qt is a trademark of Trolltech.

a snapshot of the GUI, while measuring. The second computational level calculates the spot positions from the read-in data and the Zernike coefficients. It also decides, whether a spot is detected or not, by analyzing the bit-pattern of the PSD. The GUI allows relative measurements by subtracting the spot deviations from a reference measurement from the actually measured spot positions. The first two levels are system independent and may be ported easily to other operating systems. The third level consists of the hardware dependent I/O calls to the DAQ-card. For every read-out step the device driver [Kof00] of the DAQ-card has to be addressed and 512 I/O calls have to be made to read all data out, which takes 41.05 ms on average [4] , which is equivalent to a bandwidth of only 24 Hz.

To achieve the desired frame rate of several hundreds of Hz, a custom device driver has been implemented, which reads out all 64 detectors with only one I/O call through the device driver. The read-out sequence of the complete data is included into the device driver and thus reduces the read-out time to 2.76 ms, including already 0.22 ms for the calculation of the Zernike coefficients from the spot positions. This is equivalent to a final frame rate of 362 Hz, which is fifteen times faster with the custom device driver. Read-out of a subset of 4x4 detectors takes 0.84 ms, which is equal to a frame rate of 1.2 kHz. The fastest measurements have been conducted though with only one detector at 19.2 kHz. Although Zernike evaluations other than tilt are not possible in this mode, it has been used for fast spot tracking and it was demonstrated, that the sensor can follow spot movements at 4 kHz bandwidth equivalent.

5.2 Bit-Vector Analyses

5.2.1 Spot Tracking

The bit-vector read in from the PSDs has to be analyzed to decide, whether it contains useful information or the measurement must be rejected. This can be the case, when the spot leaves the dynamic range of the detector or no spot is present on the detector.

The median of the set bits in the bit-vector, which correspond to the coordinates x_i and y_j, is used as a robust estimator of the spot position

$$(\hat{x}, \hat{y}) \quad = \quad (\mathrm{med}(x_i), \mathrm{med}(y_j)). \tag{5.1}$$

Two of the five bits from the bit-vector can be outliers without influencing the median, see the following example, where the position derived with the median from the bit-vector

...	0	0	0	0	0	0	0	1	1	1	1	1	0	0	...

and

[4] AMD Athlon processor with 800 MHz CPU clock and 33 MHz PCI bus interface.

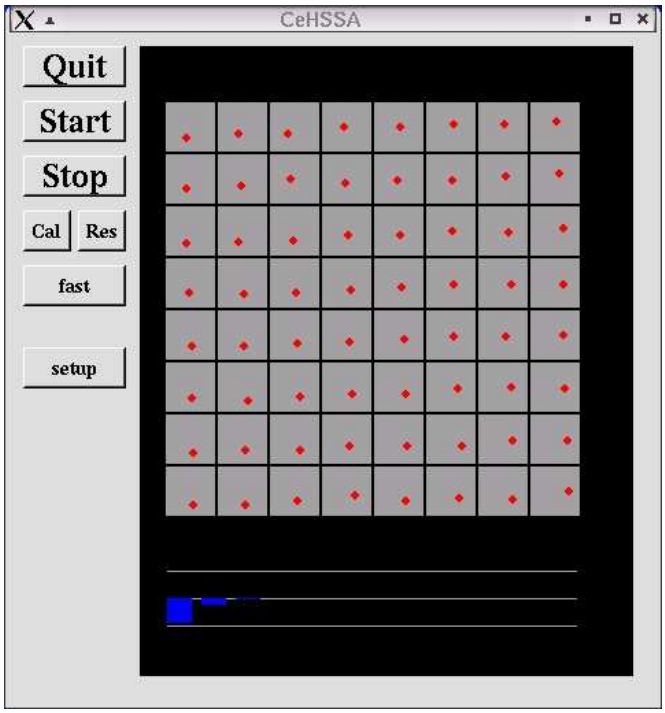

Figure 5.1: Snapshot of the GUI while measuring. Different Zernike terms are visualized as blue bars below the detector array. Detected spots are shown as red dots.

...	0	1	1	0	0	0	0	0	0	1	1	1	0	0	...

is the same. In centroid mode the number of winners can vary slightly during the transition of the spot from one pixel row to another. This is not an undesired feature, because it can slightly increase the resolution for spot tracking. The following example with a moving spot explains this behavior. The maximum of the spot is first on the mid pixel row and three winners are selected:

0	0	0	0	0	0	0	1	1	1	0	0	0	0	...

When the spot has moved and the maximum is between two pixel rows four winners are selected for a transient time

0	0	0	0	0	0	0	1	1	1	1	0	0	0	...

and finally the spot has further moved to

0	0	0	0	0	0	0	0	1	1	1	0	0	0	...

The derived spot positions in this example are 8.0, 8.5 and 9.0 and the resolution is thus increased to one half of the pixel pitch. This only works for intermediate voltages on the resistive ring, where the transconductance of the NMOS resistors is neither zero nor infinite.

5.2.2 Spot Detection

When there is no spot present on the detector or the spot is outside the photosensitive area, the bit-vector contains winners at random positions and the data is not used for the wavefront calculation.

A simple criterion of randomness is, when not all bits in the bit-vector are adjacent to each other. The probability, that a random pattern of winners are adjacent can be calculated, assuming a uniform background noise distribution. The probability that a certain neuron in a WTA circuit with k neurons wins is then $p_0 = 1/k$. The probability for a specific combination of b adjacent bits therefore is p_0^b. In an inter-digital WTA circuit with w independent circuits a total number of $kw - b + 1$ combinations exists, that realize this combination and the probability to detect such a pattern is

$$p(b) = \frac{1}{k^b}(kw - b + 1), \tag{5.2}$$

from which the *Probability of False Detection (PFD)* can be calculated.

A more flexible criterion should allow some outliers, such that only $b - 1$ or $b - 2$ adjacent winners have to be present in the bit-vector, in order to accept it. When calculating the PFD for $b - 1$ adjacent winners it has to be taken into account, that the bit-vector with $b - 1$ adjacent winners is included in the PFD for b adjacent winners.

The same applies for $b-2$, where the bit-vectors are already included in $b-1$. Therefore the probability to detect a bit-vector with a larger number of adjacent bits has to be subtracted first and the final PFD is

$$p_F(b) = p(b) - \sum_{i=b+1}^{l} p(i). \tag{5.3}$$

The implemented PSD with five WTA circuits $(w = 5)$ and four neurons each $(k = 4)$ has a PFD of only 1.56%, that a random bit-pattern is mistaken as a detected spot[5]. Table 5.1 shows a comparison between the PFD from Equation 5.3 and from a Monte-Carlo simulation.

b	5	4	3
$p_F(b)$ simulation	1.72%	5.15%	22.35%
$p_F(b)$ analytical	1.56%	5.08%	19.92%

Table 5.1: False detection probability for the criterion, that b adjacent winners have to be present in five separate WTA circuits from Equation 5.3 and from a MC simulation. The results agree within the 2σ-interval, which is 1.6% in this simulation.

Yet another criterion for randomness uses the *normalized mean absolute deviation from the median (mad)* of the coordinates of the winning neurons

$$M_x = \frac{\mathrm{mad}(x_i)}{a(w)} - 1 \tag{5.4}$$

and

$$M_y = \frac{\mathrm{mad}(y_i)}{a(w)} - 1 \tag{5.5}$$

to decide, whether the data from a PSD has to be rejected or not. The data is rejected, in the case that

$$M_x > M_L \quad \text{or} \quad M_y > M_L \tag{5.6}$$

with M_L being a value adapted to the actual rate of expected false data. The normalization factor $a(w)$ is chosen such, that for w adjacent winning neurons the values M_x and M_y are zero. Table 5.2 shows the normalization factor for some values of w.

[5]The CeHSSA results are slightly smaller, because one of the WTA circuits has five neurons instead of four.

w	2	3	4	5	6	7	8
$a(w)$	$\frac{1}{2}$	$\frac{2}{3}$	1	$\frac{6}{5}$	$\frac{6}{4}$	$\frac{12}{7}$	2

Table 5.2: Normalization factor for different numbers of adjacent winners.

The criterion M_L has to be chosen such, that bit-vectors with only slight, still acceptable deviations are not rejected. A useful measure for the efficiency of the detection algorithm is the *Probability of False Rejection (FRP)* p_R, which is the number of rejected spots for a given M_L, divided by the total number of spots, in case all spots are present and should be detected. In the spot measurements with the CeHSSA a value of $M_L = 1$ has been a good trade-off between a large probability of false rejection and a small probability of false detection.

6 Measurement Results

In this chapter the measurement results from the prototype and the final version of the chip are presented. The optical setup is explained. In the first part the spot tracking performance and the achievable bandwidth will be presented for different spot powers and operation modes. The second part is dedicated to wavefront measurements, including the first measurements of human eye‘s wavefront errors with several hundred Hz.

6.1 Optical Setup

All spot tracking and detection measurements have been conducted with an optical setup and a motorized x-y-translation stage. The first optical system consists of a collimated laser beam from a 680 nm diode laser. The beam passes through a neutral density filter to achieve the desired spot power. After collimation the beam is widened up by a telescope and focused into the focal plane of the sensor by a microscope lens. This setup has been used to produce single spots with minimum 20 μm spot FWHM.

In the second setup the microscope lens is replaced by microlens arrays and thus forms a Hartmann-Shack wavefront sensor together with the CeHSSA. Gradient index arrays, which match the detector pitch of 400 μm, have been available with 53 mm and 30 mm focal length[1]. In both setups the sensor is mounted on a PC-controlled x-y-translation stage for positioning and spot tracking measurements. Figure 6.1 shows schematics of optical setup 1 and 2.

6.2 Spot Tracking Results

Single Focused Spots

The tracked spot position has been evaluated with special interest in the lowest possible spot power and with respect to the achievable position noise at different spot powers and in different operation modes.

[1]The microlens arrays have been manufactured by Adaptive Optics Associates (AOA).

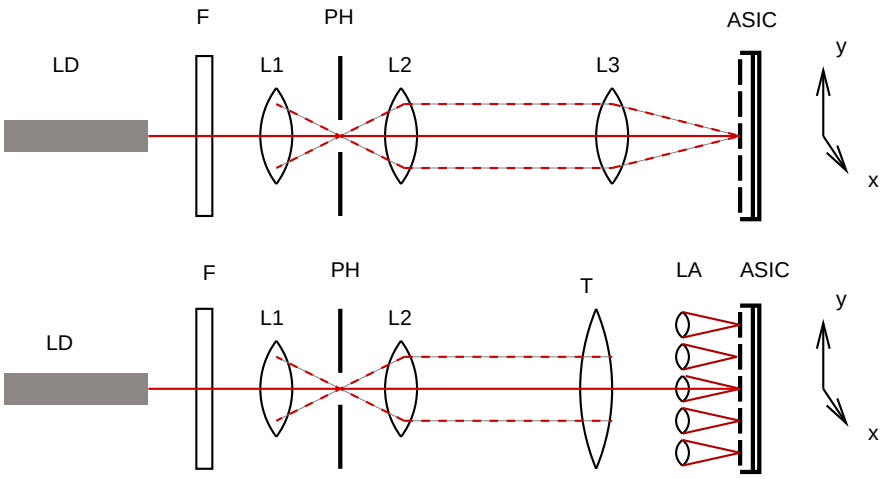

Figure 6.1: Optical setup 1 (above) and optical setup 2 (below). The beam from the laser diode LD passes a neutral density filter F and is widened up and collimated by lenses L1 , L2 and the pinhole PH. In setup 1 the beam is focused by lens L3 onto the sensor ASIC. In setup 2 the lens is replaced by the microlens array LA and the aberrations of test object T can be measured. The ASIC is mounted on a x-y-translation stage.

Figure 6.2 and Figure 6.3 show spot tracking results in maximum and centroid mode respectively from optical setup 1 with the prototype. The dynamic range of the prototype was limited to 272 μm.

The measurements show, that for an incident power as low as 300 pW/spot, the position is tracked within the resolution of the sensor in centroid mode, while in maximum mode spot tracking performance is poor. At 30 pW a spot can still be tracked in centroid mode, although the tracking error is now above the resolution of the sensor. As the SNR gets smaller, the effects of fixed-pattern noise increase. Table 6.1 shows the standard deviation from the measured data and the linear fitted data in both modes of operation at different spot powers. The upper limit of a random bit-vector is given by 20/4=5 pixels, which agrees reasonably well with a MC simulation result of 5.8 pixels.

The data from the prototype showed, that measurements in the sub-nW range are only feasible in centroid mode, because the maximum mode is much too susceptible to fixed pattern noise. This result agreed well with the expectation and motivated the use of the resistive-ring network in the CeHSSA.

Spots from a Lenslet Array

In contrast to the single laser spots with about 20 μm FWHM, the spots from the used lenslet arrays are about four times larger and have a poorer optical quality. Tracking of these spots is thus more susceptible to FPN due to their broader peak, which causes

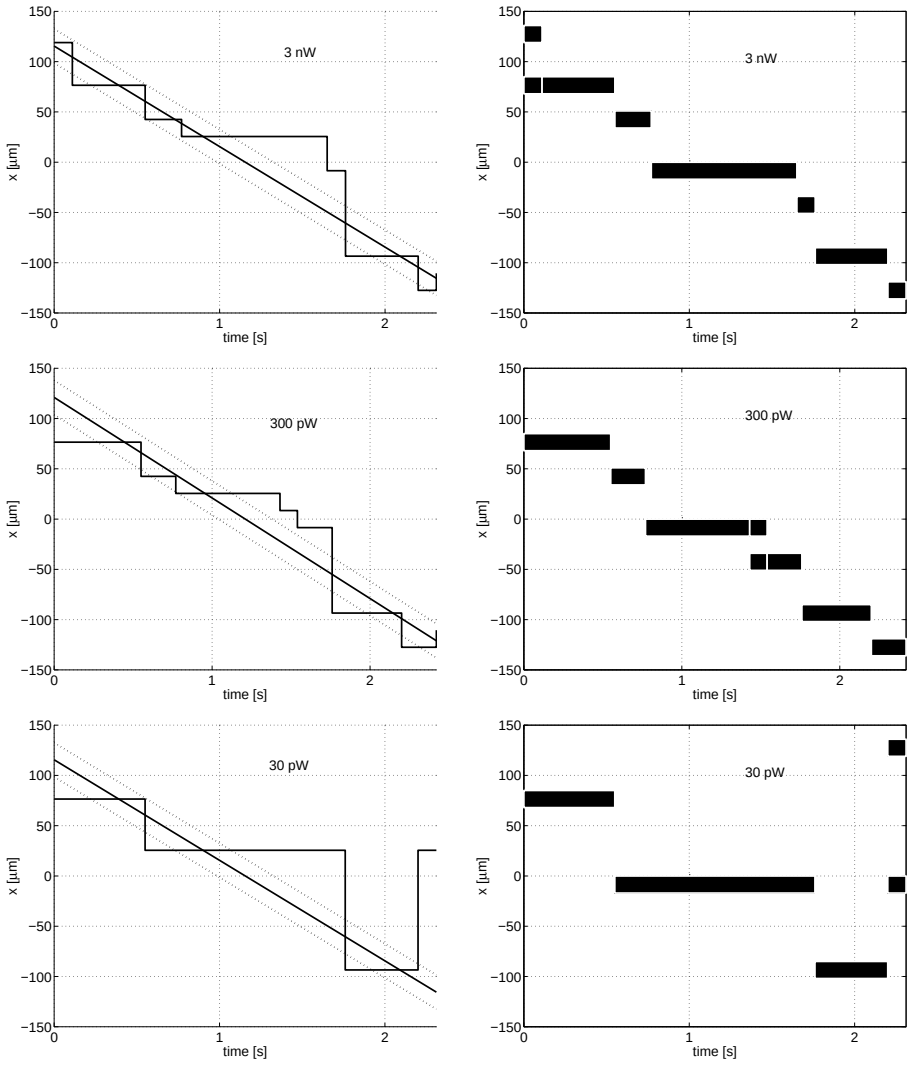

Figure 6.2: Spot tracking at different spot powers in maximum mode. Raw bit-vector (right column) and spot position (left column), with the actual spot position (straight line) and the one-pixel error (dotted lines).

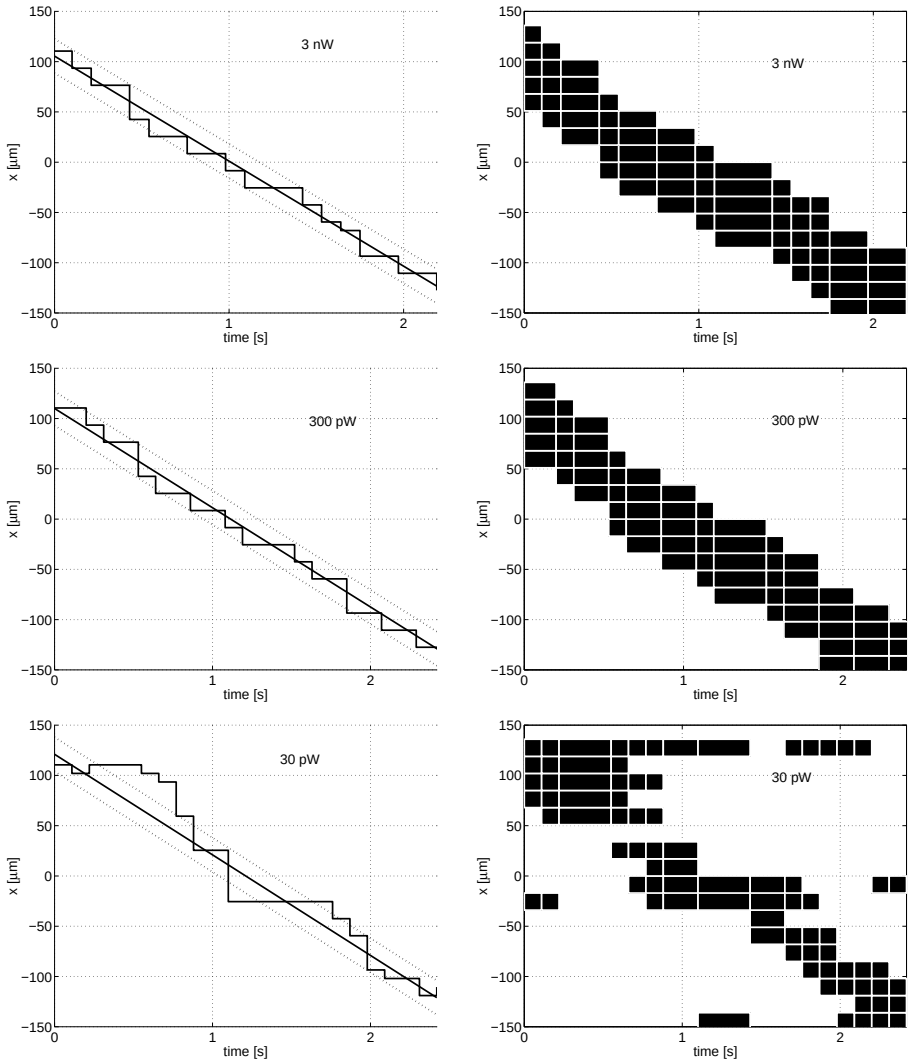

Figure 6.3: Spot tracking at different spot powers in centroid mode. Raw bit-vector (right column) and spot position (left column), with the actual spot position (straight line) and the one-pixel error (dotted lines).

spot power	σ_{max}	$\sigma_{\mathrm{centroid}}$
300 nW	1.19	0.42
30 nW	1.06	0.41
3 nW	1.40	0.43
300 pW	1.58	0.59
30 pW	3.41	1.25

Table 6.1: Standard deviation of the spot position in pixels in maximum and centroid mode. Note, that a completely random bit-vector has a standard deviation of $\sigma = 5$.

a smaller difference between the input photocurrents of adjacent neurons. The diffraction limit of the spot's FWHM for the used microlens array with 400 μm aperture and 53 mm focal length is 85 μm according to Equation 1.6.

While spot measurements with setup 1 and the prototype have been conducted at slow spot velocities, the measurements with the CeHSSA have been done with fast moving spots at the limit of the bandwidth of the PSDs. Although the x-y-translation stage only allows linear movements, a spot moving with constant velocity v can be related to a bandwidth B through the number of pixels with pitch p, the spot crosses per second $B = v/p$. Figure 6.4 shows a number of measurements for equivalent bandwidths ranging from 60 Hz up to 6 kHz. Only 4×4 detectors have been read out for these measurements to achieve the necessary large frame rate of 1.2 kHz. The fastest measurement has been done with only one detector and a frame rate of 19.2 kHz. Spot velocities have been chosen between 1 mm/s and 100 mm/s.

Tracking of the spots from a lenslet array leads to a larger standard deviation of the spot position due to the poorer optical quality and the broader peak of these spots. Nevertheless, the standard deviation is below 1 pixel in general, even for very fast spot movements with large equivalent bandwidths. The CeHSSA fulfills the requirements with respect to linearity, bandwidth at low spot powers.

6.2.1 Resistive-Ring Measurements and Sub-Pixel Resolution

The spot tracking measurements in the previous chapter have been conducted in pure maximum mode ($V_{\mathrm{resgrid}} = V_{dd}$) or pure centroid mode ($V_{\mathrm{resgrid}} = 0$), where there is no current distribution through the resistive ring. When operated in the intermediate region ($0 < V_{\mathrm{resgrid}} < V_{dd}$), the current through the current sinks is redistributed between the different WTA circuits and they interact. This can result in an increase of the effective resolution of the PSD.

Although the resolution of the simple WTA circuit is equal to the pixel pitch p, a sub-pixel resolution of almost 20% has been achieved, when the PSD is not operated in pure maximum or pure centroid mode, but with intermediate resistances at the resistive-ring network instead. This can be explained by the current redistribution in

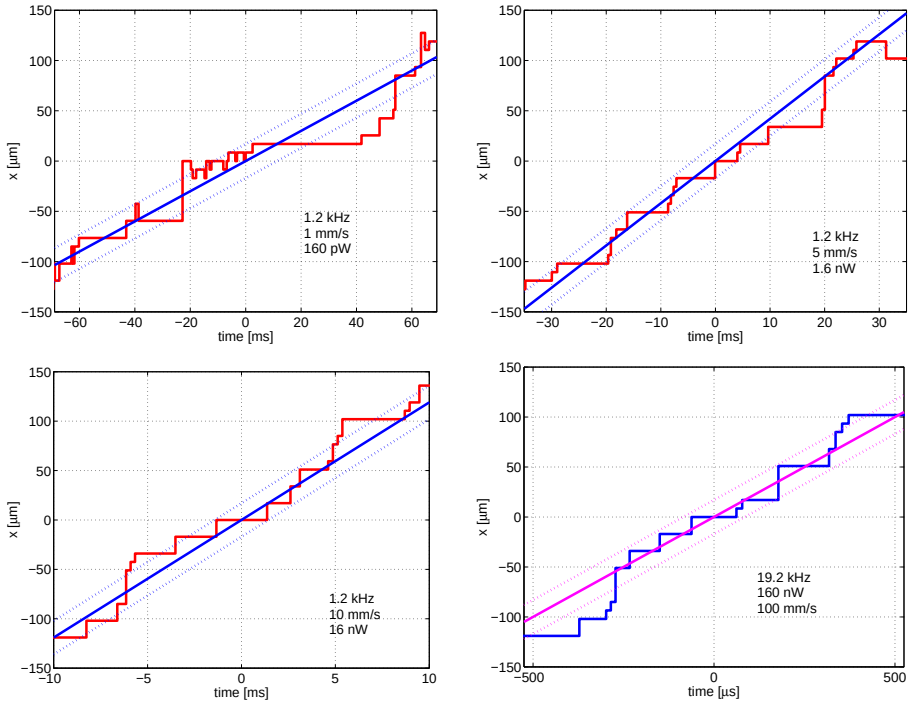

Figure 6.4: Fast spot tracking at different spot velocities, which have an equivalent bandwidth of 59 Hz (upper left), 294 Hz (upper right), 590 Hz (lower left) and 5.9 kHz (lower right). The frame rate varied between 1.2 kHz and 19.2 kHz.

the resistive-ring network, which means, that the number of winning neurons is not fixed, but differs by one winner more or one winner less from the average number of winning neurons, depending on the photocurrent distribution at a certain time. This effect only occurs, when the voltage at the resistive ring V_{resgrid} is neither zero nor V_{dd}. In the example shown in Table 6.2 at $V_{\text{resgrid}} = 1.5\,V$, the number of winning neurons is 2±1. Figure 6.5 shows spot tracking results at different values of V_{resgrid}.

no. winners	1	2	3	4	5
frequency	30%	50%	20%	0%	0%

Table 6.2: Frequency of the number of winning neurons at $V_{\text{resgrid}} = 1.5\,V$.

Table 6.3 shows the average number of winners $\overline{w}$ and the effective resolution $\overline{p}$ as a function of the bias voltage at the resistive-ring network.

V_{resgrid}	$\overline{w}$	$\overline{p}/p$
0 V	4.8	1.02
1.25 V	3.6	0.88
1.5 V	1.9	0.81
2.0 V	1.6	0.87
3.3 V	1.0	1.01

Table 6.3: Average number of winners $\overline{w}$ at different bias voltages V_{resgrid} at the resistive-ring network and ratio of effective resolution and pixel pitch.

Almost 20% sub-pixel resolution have been achieved, with the PSD operated in the intermediate region at a bias resistive-ring network voltage of 1.5 V.

Although this was a promising result, most consecutive measurements have been conducted in centroid mode, because the spot detection algorithm needs at least three winning neurons to work properly. The bias voltage should therfore be chosen at least 1.25 V.

6.2.2 Spot Tracking with the Resistive-Line Network

The spot tracking performance of the resistive-line network from Chapter 4.6.2 has been evaluated with optical setup 1 and the prototype. The spot with about 40 μW spot power moved with 75 μm/s over the pixel array, which is equivalent to a velocity of 4.4 pixels/s. A voltage of 0.3 V has been chosen for the bias voltage of the active resistors. Figure 6.6 shows the recorded output currents of the resistive-line network to ground, as the spot moves over the array.

Figure 6.7 shows the position of the spot calculated with Equation 3.36, which shows a very linear response and very good resolution at the innermost 200 μm of the

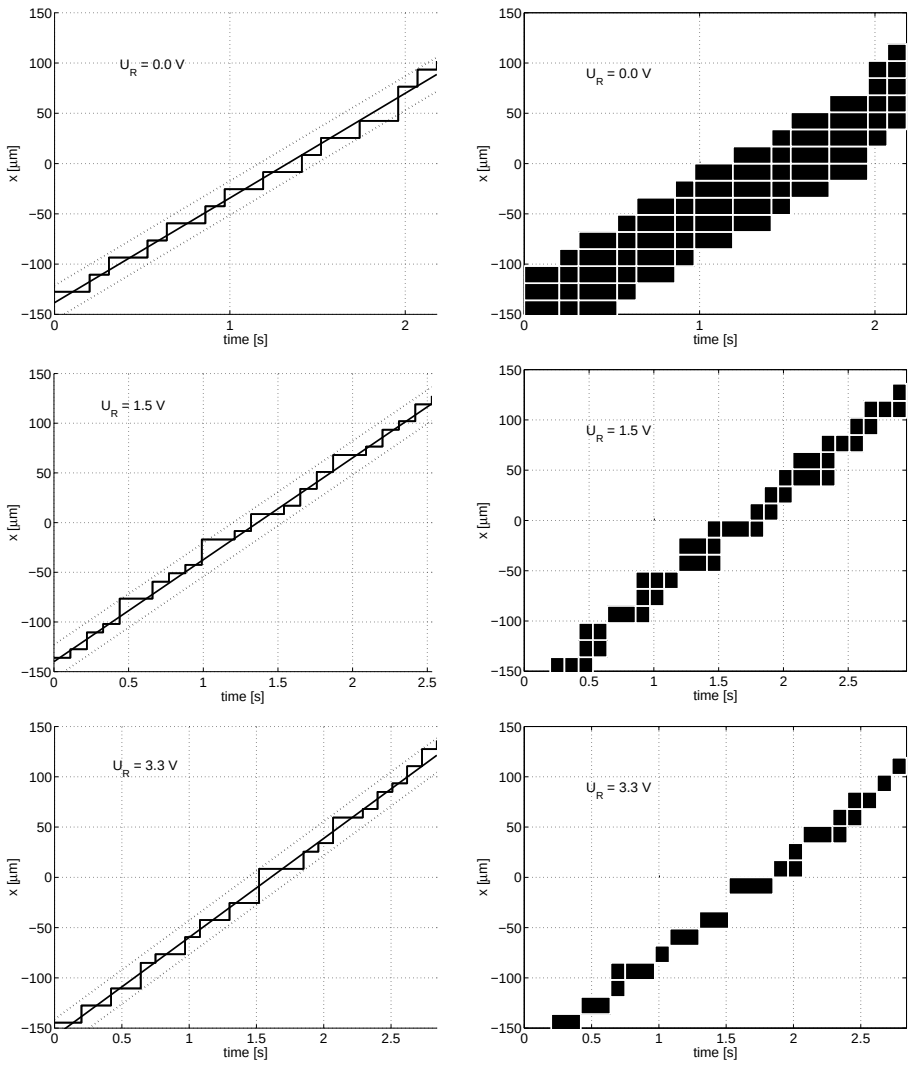

Figure 6.5: Spot tracking at different bias voltages V_{resgrid} at the resistive-ring network. Raw bit-vector (right column) and spot position (left column), with the actual spot position (straight line) and the one-pixel error (dotted lines). It can be seen, how the number of winning neurons in the bit-vector gets smaller with rising V_{resgrid}.

detector. When the spot begins to leave the array, the measured photocurrents have no relation to the position of the spot and the output is ambiguous. Although these measurements showed the promising linearity and the good resolution of the resistive-line network, it has been rejected for use in the CeHSSA due to the manyfold reasons explained in Chapter 3. Nevertheless, future work could overcome the mentioned limitations and allow tracking at high resolution with this kind of PSD.

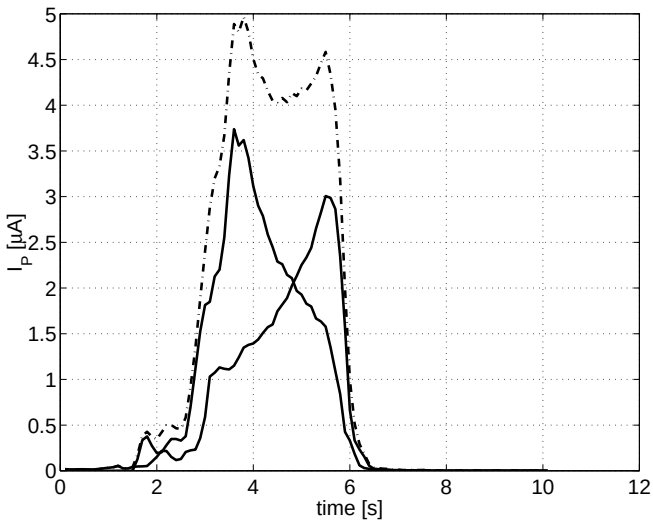

Figure 6.6: Left and right output current from the resistive-line network to ground (solid lines) and total current (dashed line) of a focused laser spot moving over the pixel array.

6.3 Spot Detection

The data from the detector is rejected, when the normalized mean deviation from the median surpasses a predefined confidence limit. Figure 6.8 shows the effectiveness of the detection filter for a single spot, that moves over four adjacent detectors. The filter nearly rejects all data, when no spot is present on the detector. This filtering works best in centroid mode, while no filtering is possible in maximum mode.

False detection probability p_F and false rejection probability p_R have been evaluated for the CeHSSA at different confidence levels. Table 6.4 shows the measured probabilities for different values of the confidence limit M_L.

The false detection probability has been derived from the number of detections, when no spot pattern was present, while the false rejection probability has been evaluated from the number of rejected spots from a static pattern at a large spot power, which are falsely rejected. Although the ideal case would be, that p_F is zero at $M_L = 0$, a weaker criterion like $M_L = 1$ should be used for spot rejection, where both p_F and p_R

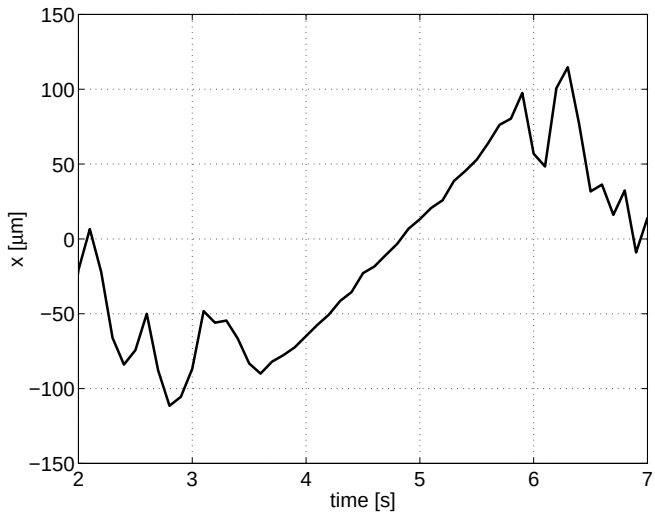

Figure 6.7: Spot position from the resistive-line network. A spot moved with constant velocity over the pixel array between t =3.5 s and t =5.9 s. When no spot is present, the result of the calculation is arbitrary.

M_L	p_F	p_R
0.2	0%	42%
0.5	0%	12%
1.0	0%	0%
2.0	2%	0%

Table 6.4: Probability of false detection p_F and probability of false rejection p_R for different confidence limits M_L.

are zero, which takes the poorer optical quality of the spots from the lenslet array into account.

6.4 Static Wavefront Measurements

In the following chapters Zernike coefficients have been calculated from the spot position data, according to the least-squares solution from Appendix A. Only spot position data that passed the detection filter has been used for the calculation. For some fast measurements only a subset of detectors has been used, in order to reach a bandwidth above 1 kHz. All other measurements use all 8×8 spots for the evaluation.

In optical setup 2 a telescope has been added for the evaluation of static defocus. Such a setup is often used to pre-compensate large defocus values for subsequent measurement of higher order aberrations [Sch02], but it can also be used vive versa to add

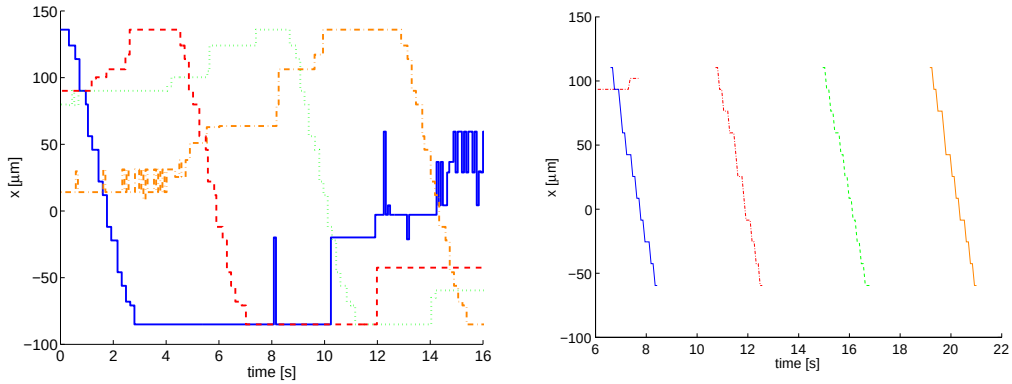

Figure 6.8: Detection filter. A single focused laser spot passes over four adjacent detectors at constant velocity. The raw position data (left side) is filtered and almost all data is rejected (right side), when no spot is present on the detector (different line styles indicate the data from different detectors).

a controllable amount of defocus to a plane wavefront. The linearity and accuracy of a Hartmann-Shack sensor can thus be evaluated.

Defocus C_2^0 can be added with a telescope with two spherical lenses of focal length f_1 and f_2, which are extended from their normal distance (f_1+f_2) by a displacement Δz. The effective power g_D can be calculated from the lens equation [Sch02]

$$\frac{1}{f_1} = \frac{1}{f_1 + \Delta z} + \frac{1}{f_1 - g_D} \tag{6.1}$$

to

$$C_2^0 = \frac{1}{g_D} = \frac{\Delta z}{f_1^2}. \tag{6.2}$$

Figure 6.9 shows the optical setup 2 with the movable first lens (f_1 =14 cm). This means, that every 1.96 cm one diopter of defocus is added to the wavefront. The second lens of the collimator has been used here as the second lens of the telescope. In Figure 6.12 the measured defocus C_2^0 as a function of the position of the second telescope lens is shown in maximum and centroid mode for 8×8 and 4×4 detectors at low spot powers. The measured defocus shows the expected linear dependence from the position of the second telescope lens. The standard deviations are 0.16 dpt in centroid mode and 0.36 dpt in maximum mode, when all spots are evaluated. These results agree well with the assumption, that in centroid mode the standard deviation of the tracked position due to fixed-pattern noise is reduced by the square root of the number of winning neurons, or $\sqrt{5}$ in this case.

When only a sub-window of 4×4 detectors is used for the wavefront reconstruction, the standard deviation is 0.28 dpt in centroid mode. This agrees well with the as-

sumption, that the standard deviation of the wavefront σ_C is proportional to the square root of the number N of spots used for the wavefront reconstruction

$$\frac{\sigma_C^2(N_1)}{\sigma_C^2(N_2)} = \frac{N_1}{N_2}. \tag{6.3}$$

A number of measured standard deviations for different configurations can be found in Table 6.5.

configuration	centroid, 8×8	maximum, 8×8	centroid, 4×4
measured σ_C	0.16 dpt	0.36 dpt	0.28 dpt
expected σ_C		0.36 dpt	0.32 dpt

Table 6.5: Measured and expected standard deviation of defocus in different configurations. The expected standard deviation has been normalized to the measured value in centroid mode with 8×8 detectors.

The standard deviation of different Zernike terms at different spot intensities is of great importance to estimate the expected wavefront errors. A static wavefront has been measured with 120 wavefront snapshots. Figure 6.10 shows the standard deviation of the first Zernike terms up to 4th radial order for 300 pW/spot and 300 nW/spot and Figure 6.11 a time series of the same measurement. The measured data suggests, that the wavefront error σ_C is proportional to the square root of the spot intensity J or

$$\frac{\sigma_C^2(J_1)}{\sigma_C^2(J_2)} = \frac{J_1}{J_2}. \tag{6.4}$$

The larger the azimuthal order of the Zernike term is, the larger is also the standard deviation, which could have its origin in the reconstruction algorithm.

Absolute defocus measurements with spherical test-lenses are shown in Figure 6.13. The measured values agreed well within the theoretical resolution limit of about 0.1 dpt, except for the lowest and the largest defocus values. In the case of the largest value, this can be contributed to the already exceeded dynamic range. The deviation of the lens with 0.25 dpt may be due to the poor optical quality of this test-lenses.

6.5 Dynamic Wavefront Measurements

6.5.1 Astigmatism Measurements

A cylinder compensation system is used to pre-compensate large cylinder values, prior to measurements of higher order aberrations. On the other side, such a system can be used to add arbitrary cylinder values to a plane wavefront. This system consists of two cylinder lenses of equal focal length f_1 and f_2, but different sign, which are rotatable around their central axis. Figure 6.14 shows optical system 2 with the two cylinder

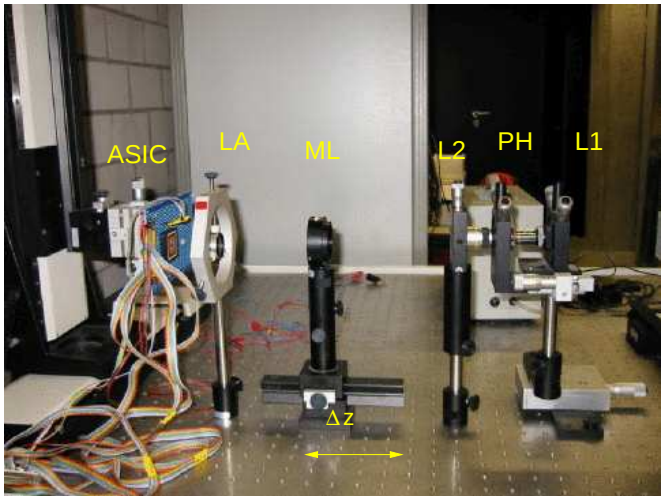

Figure 6.9: Picture of optical setup 2 with movable first telescope lens ML. The neutral density filter and the laser diode are not shown in this picture.

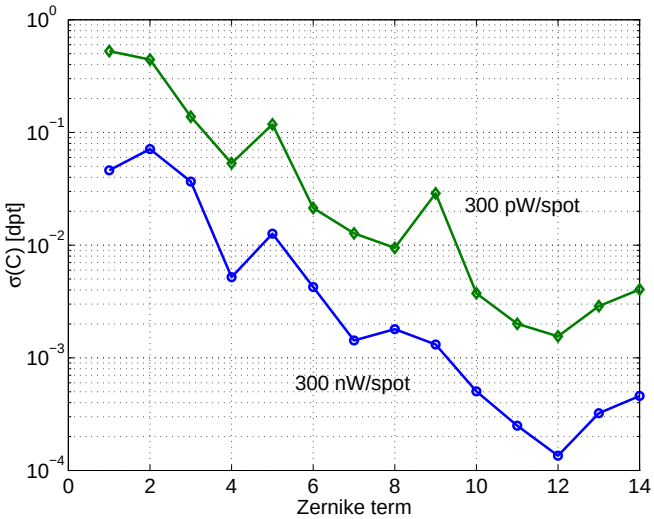

Figure 6.10: Zernike-term standard deviation for large and small spot powers.

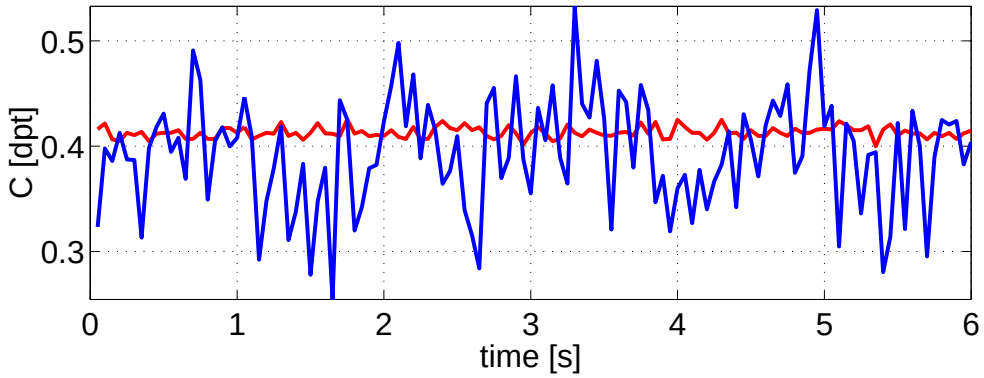

Figure 6.11: Time series of the C_2^{-2} Zernike-term at 300 pW/spot (blue) and 300 nW/spot (line).

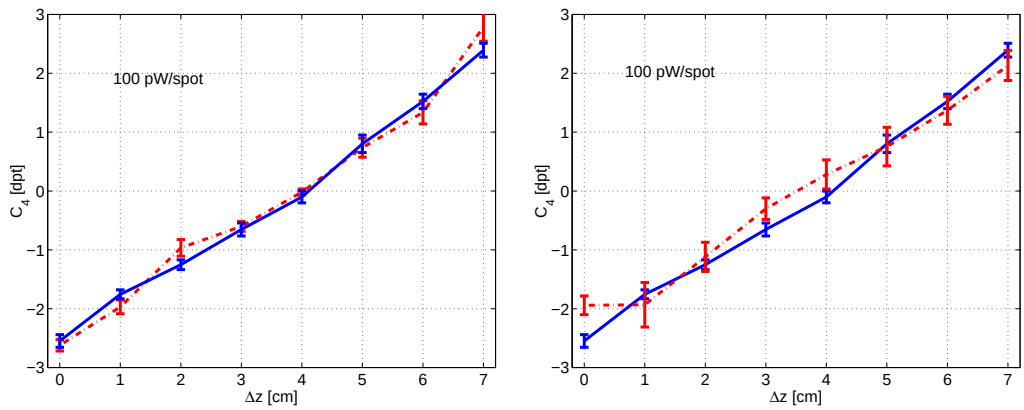

Figure 6.12: Static defocus measurements with 8×8 spots (left side) and 4×4 spots (right side) in maximum mode (dashed lines) and centroid mode (solid lines). The Zernike coefficient C_2^0 is plotted as a function of the extension of the telescope. The error bars show the standard deviation of 100 measurement.

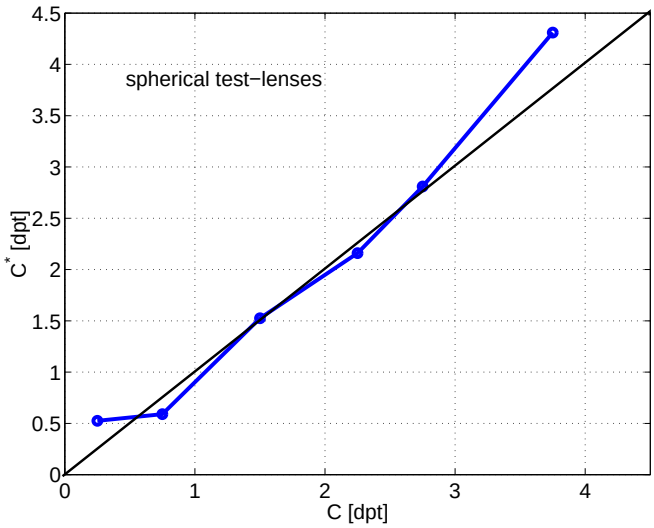

Figure 6.13: Defocus of spherical test-lenses. The measured defocus C_2^{0*} is plotted versus the denoted value of the lenses C_2^0.

lenses CL1 and CL2. A cylinder compensation module from Schottner [Sch02] with two rotatable cylinder lenses ($f_1 = -f_2$ =500 mm) has been used for this purpose. The maximum astigmatism, which can be added to a plane wavefront is

$$C_{A,\max} = -\frac{1}{f_1} - \frac{1}{f_2} = 4\,\mathrm{dpt}. \tag{6.5}$$

The two rotation angles α_1 and α_2 of the cylinder lenses define the effective power

$$C_A = 2C_{A,\max}|\sin(\alpha_1 - \alpha_2)| \tag{6.6}$$

and the effective angle

$$\alpha_A = \frac{\alpha_1 + \alpha_2 - \pi}{4} \tag{6.7}$$

of the astigmatism. Figure 6.15 shows exemplary time series of the effective angle α, when one of the cylinder lenses is moved by a manual impulse. Due to this manual stimulus, a linear response can not be expected. Nevertheless an averaging with an 8th-order running median filter already gave a very good response with very few noise left. Figure 6.16 shows a contour-plot of the measured wavefront before and after rotation of one of the lenses.

6.5.2 Wavefront Measurements at the Human Eye

First successful wavefront measurements at the human eye are the most important step towards an application of the CMOS-based sensor in opthalmology. A Hartmann-

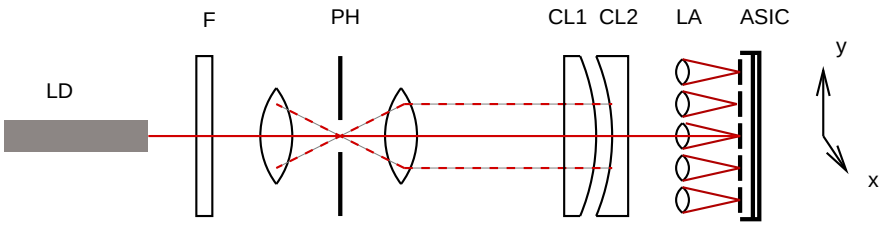

Figure 6.14: Optical setup 2 with module for adding cylinder aberrations. The cylinder lenses CL1 and CL2 are rotatable around their central axis.

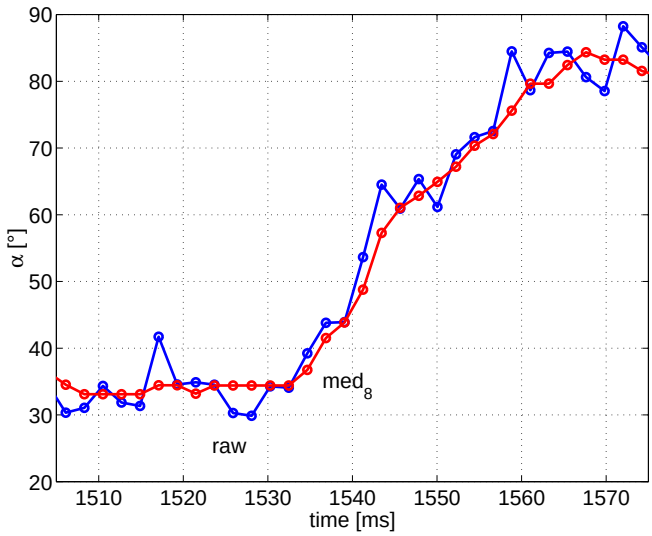

Figure 6.15: Time series of astigmatic angle from a rotating cylindrical lens after transient manual rotation at t =1530 ms, raw data (blue line) and data filtered with 8th order running median filter (red line). This data has been evaluated with 6×6 spots.

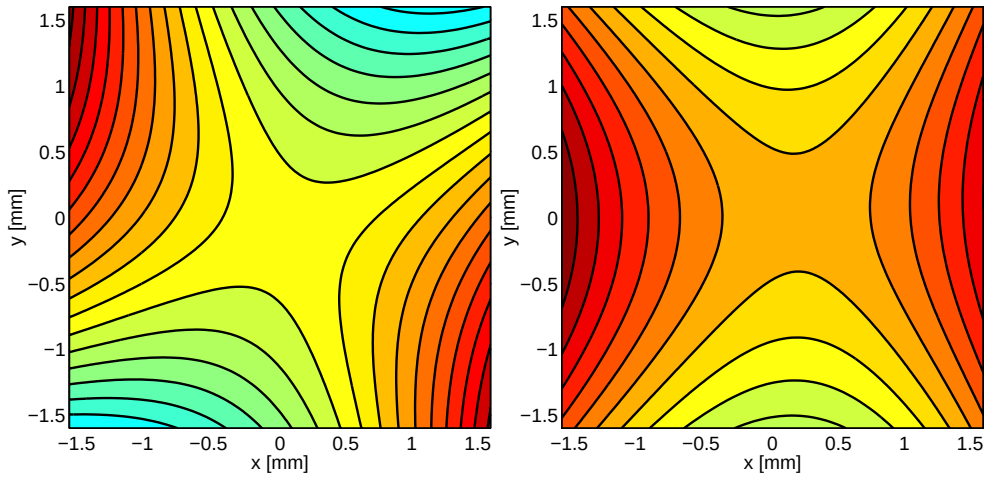

Figure 6.16: Wavefront contour plot at t =1530 ms (left) and t =1570 ms (right). The contour lines have 50 nm spacing.

Shack system has been set up to show the proof of principle. Figure 6.17 shows a schematic of optical setup 3 for wavefront measurements at the human eye and Figure 6.18 a picture of the setup with a test person. Besides the Hartmann-Shack sensor, the setup consists of a telescope and a beam-splitter to couple a laser beam into the eye. In this setup a 720 nm diode laser has been used and a lenslet array with 30 mm focal length. Instead of the custom based Hartmann-Shack sensor, a CCD camera can be mounted for subsequent evaluation with a commercial software for wavefront analysis [2]. This step is of great importance to assure that the correct reflex is imaged on the sensor. As the CMOS-based sensor provides no real image, this is a necessary control prior to measurements with the CMOS-based sensor.

First wavefront measurements with a test person showed the necessity for the rejection of false data, because the human eye constantly moves and some detectors may temporarily be not illuminated. Figure 6.20 shows a time series of measured astigmatism, defocus and 3rd order spherical aberration at 300 Hz repetition rate. The measured wavefront is also visualized in the time series of contour plots of Figure 6.22. Faster measurements can be made to the expense of a larger SNR, when only a subset of detectors is read out with each frame. Figure 6.19 shows the standard deviation of the Zernike terms in this measurement. Note the similarity with the standard deviation from the static wavefront measurement in Figure 6.10.

The data suggest the interesting fact, that the standard deviation is lowest for the two terms with radial order zero, defocus C_2^0 and 3rd order spherical aberration C_4^0. On the other side, the terms with large azimuthal order suffer from rather large temporal variations, like astigmatism C_2^2 and C_2^{-2}, C_4^{-4} and C_4^4. This is especially noticeable in

[2]The WaveScan Wavefront Analyzer is a trademark of the company 20/10 Perfect Vision.

the series C_4^0, C_4^2 and C_4^4. This behavior has its origin probably in the reconstruction algorithm, which may be numericaly sensitive to noise. Further analysis have to be conducted to get rid of the large noise in the measured astigmatism values.

A Fourier analyses of the time series yields the temporal spectrum of the human eye‘s wavefront aberrations. According to Hofer et al [HAS+01] the power spectrum drops with approximately 4 dB per octave. These results have been confirmed and extended to a larger bandwidth with the CeHSSA through a power spectral analyses of a time series of human eye measurements. Figure 6.21 shows an analyses and a comparison with the power spectral density of a static wavefront. As the power spectral density of the noise from a static wavefront is equally distributed, while the wavefront measurements at the human eye show a clear dropping at a rate of 3.6 dB/octave for defocus, it is assured that indeed the spectrum of the human eye‘s wavefront errors are measured and not only a static noise spectrum. While the data from Hofer et al has been measured with 60 Hz, the CeHSSA measurements have been conducted with more than 300 Hz.

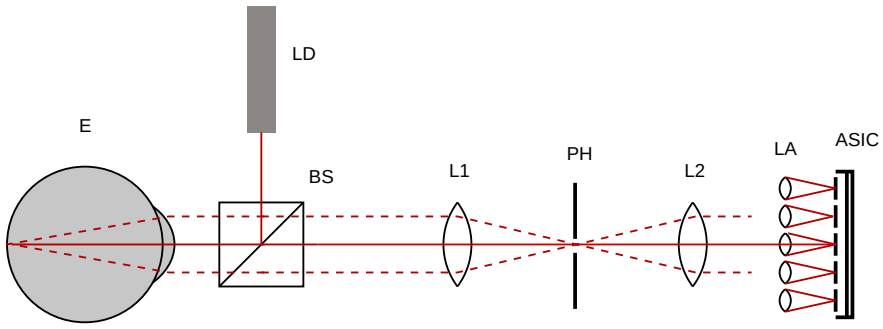

Figure 6.17: Optical setup 3 for aberration measurements of the human eye. The beam from the laser diode (720 nm) is reflected into the optical axis by the beam-splitter cube BS and gets reflected on the retina of the eye E. The reflected beam passes a telescope with lenses L1, L2 and the pinhole PH. The microlens array and the ASIC form the wavefront sensor.

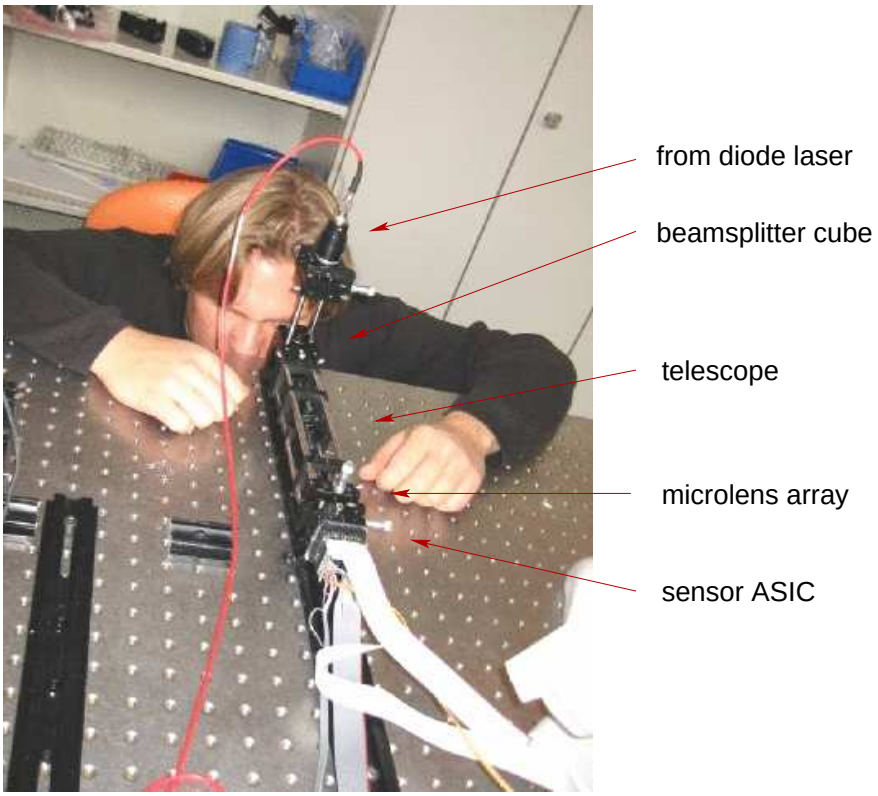

Figure 6.18: Optical setup 3 for wavefront measurements at the human eye with test person UP.

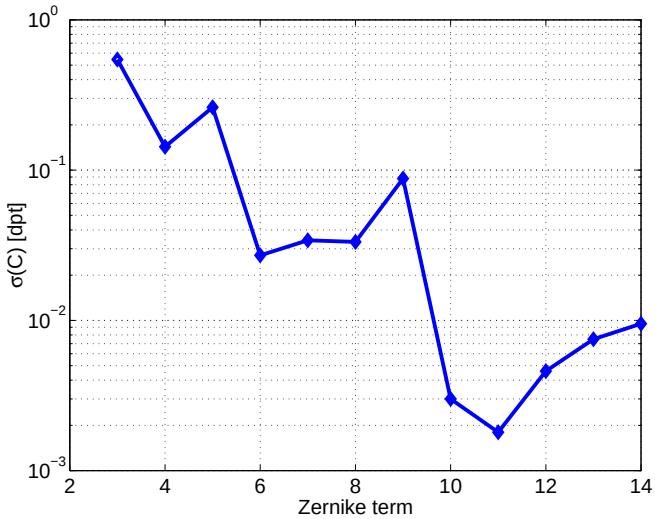

Figure 6.19: Zernike-term standard deviation. Note the similarity with Figure 6.10.

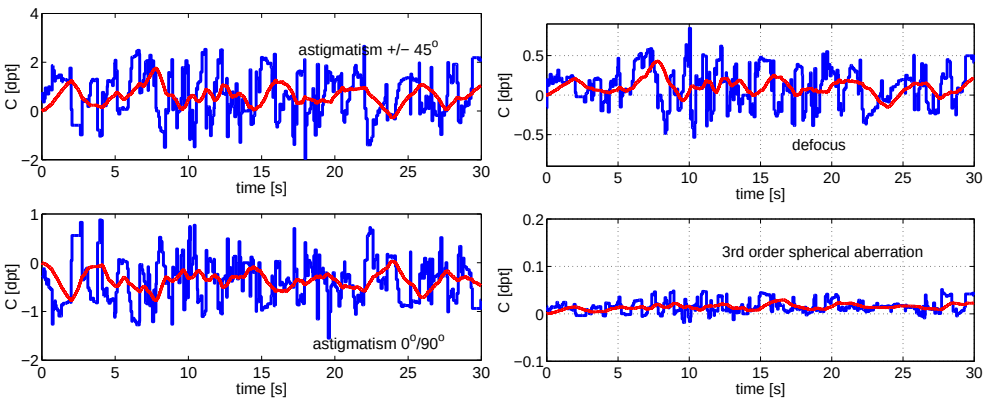

Figure 6.20: Time series of astigmatism (left column), defocus and 3rd order spherical aberration (right column) from test person TK. The raw data is shown (blue lines) and the low-pass filtered data (red lines) with $f_g = 1$ Hz.

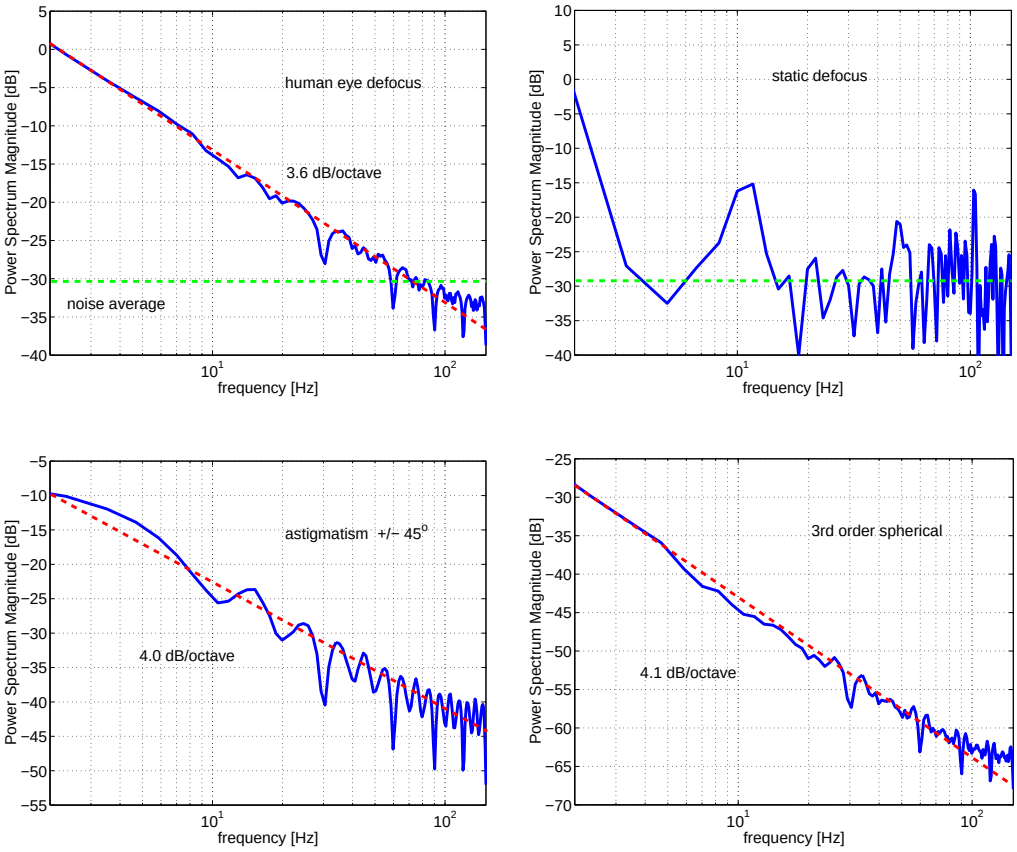

Figure 6.21: Power spectral density of human eye‘s defocus (upper left) and of a static wavefront (upper right) at the lowest possible spot power. Below approximately 70 Hz the spectral power of the human eye‘s defocus is above the noise level (dashed green line). The power spectral density of astigmatism and 3rd order spherical aberration (lower row) show a decrease of about 4 dB/octave. Note, that the power spectral density estimate of the static wavefront is noisier, because it was taken at the lowest possible spot power.

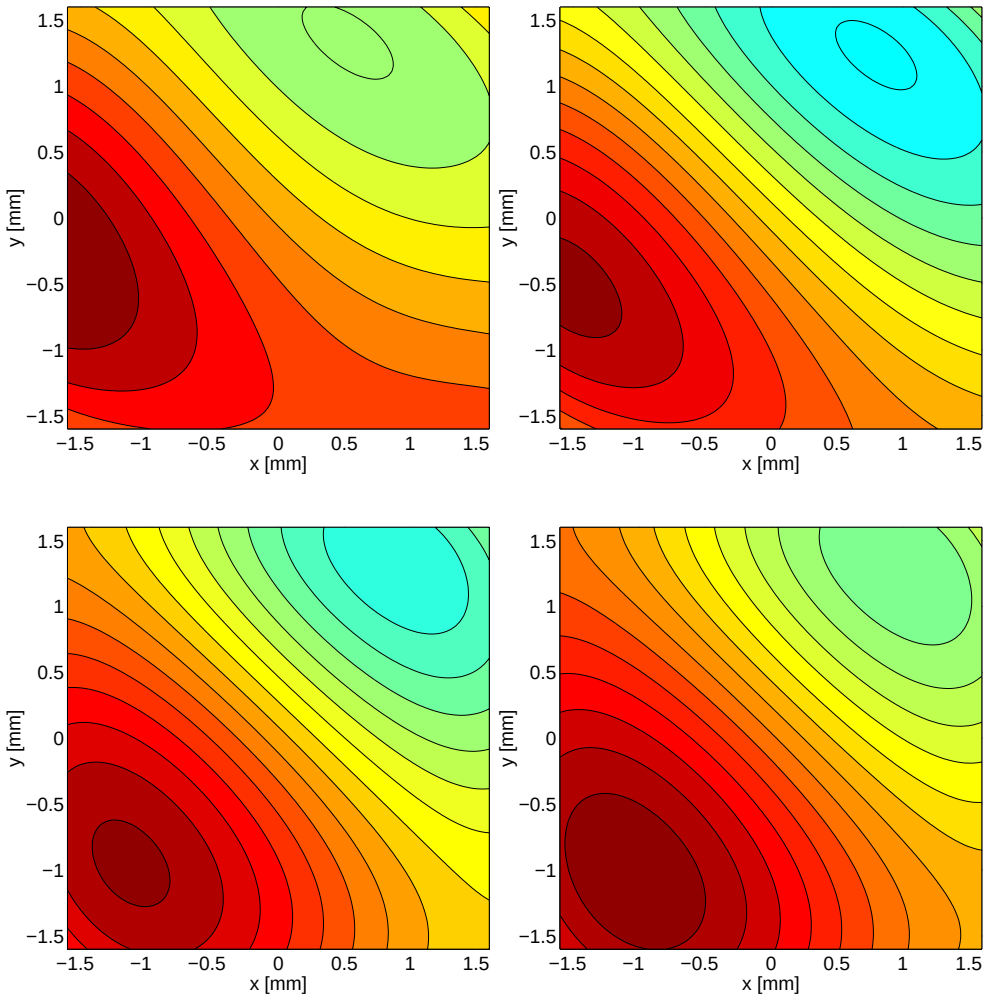

Figure 6.22: Wavefront contour plot. Time series ($\Delta t = 30$ ms) of higher order aberrations (C_4^{-4} and C_4^4 also removed). The contour lines have 20 nm spacing.

Discussion

Within the scope of this thesis, a Hartmann-Shack wavefront sensor based on an application specific integrated circuit, the CeHSSA[3] chip, has been developed and tested. This sensor is designed to be part of a fast adaptive optical system, that aims to reach the diffraction limit in imaging systems by an adaptive compensation of time-space variant wavefront aberrations.

Hartmann-Shack sensors are common non-interferometric wavefront sensors, suitable for real-time applications, because they do not contain any movable parts. The wavefront aberrations are measured by splitting the aperture into subapertures by a microlens array and determining the position of the focal spots. A local tilt in the wavefront causes a lateral displacement of the associated focal spot and the wavefront can be reconstructed from the spot pattern. Current wavefront sensors usually use cameras to grab the focal plane image and determine the spot position by software image processing. The bandwidth of sensors reported so far only amounts to some tens of Hz, because of the slow image acquisition and the time consuming software image processing, while a bandwidth of 1 kHz would be desirable to get an appropriate sampling of the wavefront.

These bandwidth limitations may be overcome by a dedicated sensor, that includes photodetectors and image processing, determines the spot positions directly and makes them available to drive an actuator for wavefront correction. As the CMOS[4] process technology offers to include photodetectors, analog and digital signal processing on a single silicon die, integrated wavefront sensors [Dro99, dLM02, NDB02] have been developed in the last years, but none of these sensors reached the sensitivity for a direct measurement of human eye wavefronts yet (see Appendix D).

After first evaluating the photodetectors with the largest quantum efficiency for the desired wavelength region and second, by integrating a position sensitive detector concept, that reduces the effects of the mismatching of process parameters and thus fixed pattern noise, a sensor has been developed in this thesis, that reaches the desired sensitivity. In a prototype version the quantum efficiency of all available passive photodiode types in the used 0.35 μm process technology has been evaluated. The passive photodiode formed by an n-well and the silicon substrate reached the largest quantum efficiency, which is 40% at a wavelength of 680 nm.

[3]Centroid Hartmann-Shack Sensor Application Specific Integrated Circuit
[4]Complementary Metal Oxide Semiconductor

Different concepts for position sensitive detectors can be used for the determination of the spot position, e.g. the quad-cell concept, the resistive-line network or the Winner-Take-All circuit. For very low spot powers the Winner-Take-All circuit shows several advantages over other position sensitive detectors, with respect to linearity, noise sensitivity and complexity of implementation on chip. The Winner-Take-All circuit offers a direct digital representation of the spot position, while in other position sensitive detectors, there is always the need for an analog-to-digital conversion of the photocurrents and further processing of this data.

Nevertheless the simple Winner-Take-All circuit has some limitations, which basically originate in the large sensitivity to parameter variations in the CMOS-process technology, the mismatching that leads to fixed-pattern noise and position uncertainty. A new topology for the Winner-Take-All circuit has been developed, that reduced the position uncertainty from 1.56 pixels to 0.56 pixels at 300 pW spot power and even allowed the tracking of spots at spot powers down to 30 pW. This topology splits the Winner-Take-All circuit in a first step into separate circuit groups, that are interdigitally connected to the input photocurrents. These groups by themselves are connected through a resistive-ring network of active resistors. It can be shown, that each group has a spatial influence that follows a hyperbolic cosine law on the circumference of the resistive-ring network. As a result of this topology, the position uncertainty has been reduced 2.8 times and the bandwidth could be extended to 6 kHz at 1 nW spot power [NPB03].

The final version of the sensor has been implemented with 28.224 pixels to detect 8×8 focal spots. The read-out scheme allows random access to the detectors. Each spot can be tracked within a dynamic range of 357 μm with 17 μm resolution, which results in a defocus dynamic range of 1.82 dpt, when a lenslet array with 53 mm focal length is used. A fill-factor of 80% has been reached within a detector area of 400×400 μm^2.

The final frame rate was limited by the external read-out system, based on a commercial data acquisition card, to 300 Hz, when all 8×8 spots are used for the wavefront reconstruction. If only one detector is read out, spot tracking measurements could be conducted with 19.2 kHz sampling rate. The spot tracking measurements with the final CeHSSA were focused on the achievable bandwidth at different spot powers. Spot movements with 300 Hz bandwidth have been tracked at 1.6 nW spot power, which is a large enough bandwidth for most ophthalmic applications.

A number of optical setups have been used for static and dynamic wavefront measurements. For this purpose, defocus could be added to a plane wavefront with a movable spherical lens and astigmatism could be added with two rotatable cylinder lenses. At a spot power of 160 pW the defocus added to a plane wavefront has been measured with an accuracy of 0.16 dpt. The standard deviation is inverse proportional to the square root of the number of spots, used for the reconstruction of the wavefront. Spherical test-lenses with different defocus values have been measured with an accuracy of 0.1 dpt.

Finally, the sensor has been used to measure human eye‘s wavefront aberrations at

300 Hz frame rate with a suitable optical setup. The spot pattern from a retinal reflection of a diode laser with 720 nm wavelength was successfully detected, although the constant movement of the eye often let the retina reflection move out of the aperture.

A final comparison between the power spectral density of static wavefront measurements at the lower limit of the sensitivity of the sensor and the human eye‘s measurements with 300 Hz frame rate showed, that the power spectral density below about 70 Hz is above the static defocus noise level. This drop in the power spectral density of the wavefront aberrations is consistent with previously reported dynamic wavefront measurements [HAS$^+$01], which state a drop rate of approximately 4 dB per octave. These successful measurements at the human eye showed, that the sensor is sensitive enough for measuring the laser reflex from a human retina and applications in ophthalmic instruments become feasible.

The CMOS-based Hartmann-Shack wavefront sensor presented in this thesis fulfilled the specified requirements with respect to bandwidth, sensitivity and resolution. Several aspects can be subject to future improvements of the performance, of which the most important one is an improvement of the resolution of the sensor. Other aspects are to embed the sensor in a real-time environment, to make the wavefront reconstruction less sensitive to noise and to include further processing steps on chip.

Currently the resolution is limited to 17 μm, which also limits the resolution of the wavefront reconstruction, e.g. about 0.1 dpt for the defocus term. The first idea to increase the resolution would be a decrease in pixel size, as there still is some room with respect to the minimum dimensions in the used sub-micron process. A practical limit would be a pixel size of about 10 μm, because the design rules imply a minimum spacing of 3 μm between adjacent photodiodes. On the other side, a decrease in pixel size also decreases the photocurrent and the signal-to-noise ratio. Another approach would be, to use a mixed type of position sensitive detector, with the resistive-ring network of Winner-Take-All circuits first selecting the two or three largest currents, having them analog-to-digital converted and calculate the spot position from this data. Such a system could be based on the existing concept with additional circuitry at the periphery of the pixel array.

With respect to a real-time adaptive optical system, a first step would be to incorporate the sensor in an appropriate hardware environment, that could also directly drive an adaptive mirror. A host computer would only be necessary to display and store the data, but not for processing.

After the first human eye wavefront measurements with a large bandwidth have been successfully conducted, the sensor should now be used in a variety of applications for long term studies.

A Least-Squares Wavefront Reconstruction

Malacara [Mal92] derived a least-squares solution for the Zernike coefficients through the calculation of a second set of orthogonal polynomials P_l. The following equations show, how the coefficients can be calculated from N spot displacements up to 4th radial order. The set of orthogonal polynomials is

$$\begin{aligned}
P_1 &= 1, \\
P_2 &= x, \\
P_3 &= y, \\
P_4 &= xy, \\
P_5 &= x^2 - y^2, \\
P_6 &= x^2 + y^2 - \alpha, \\
P_7 &= x^3 - 2x\beta, \\
P_8 &= -y3 + 2y\beta, \\
P_9 &= xy^2 - 2x\tau, \\
P_{10} &= x^2y - 2y\tau,
\end{aligned} \tag{A.1}$$

with the constants

$$\begin{aligned}
\alpha &= 2\sum_{i=1}^{N} x_i/N, \\
\beta &= \sum_{i=1}^{N} x_i^4 / \sum_{i=1}^{N} x_i^2, \\
\tau &= \sum_{i=1}^{N} x_i^2 y_i^2 / \sum_{i=1}^{N} x_i^2,
\end{aligned} \tag{A.2}$$

which have to be calculated first. The derived coefficients

$$\begin{aligned}
r_l &= \sum_{i=1}^{N} \frac{\Delta x_i}{f} P_l(x_i, y_i) / \sum_{i=1}^{N} P_l^2(x_i, y_i) \\
s_l &= \sum_{i=1}^{N} \frac{\Delta y_i}{f} P_l(x_i, y_i) / \sum_{i=1}^{N} P_l^2(x_i, y_i)
\end{aligned} \tag{A.3}$$

are used to calculate the Zernike coefficients in the indexation of Malacara

$$
\begin{aligned}
C_1 &= r_1 + (\tfrac{1}{3} - \alpha) r_6, \\
C_2 &= s_1 + \tfrac{1}{12} r_4 + \tfrac{1}{6} s_5 + (\tfrac{1}{6} - \alpha) s_6, \\
C_3 &= \tfrac{1}{2} r_3 + (\tfrac{1}{2}\beta - \tfrac{3}{16}) r_8 + (\tfrac{1}{16} - \tfrac{1}{2}\tau) r_{10}, \\
C_4 &= \tfrac{1}{8} r_2 + \tfrac{1}{8} s_3 + (\tfrac{3}{64} - \tfrac{1}{8}\beta) r_7 + (\tfrac{1}{32} - \tfrac{1}{8}\tau) r_9 + (\tfrac{1}{8}\beta - \tfrac{3}{64}) s_8 - \tfrac{1}{8} s_{10}\tau, \\
C_5 &= -\tfrac{1}{4} r_2 + \tfrac{1}{4} s_3 + (\tfrac{1}{4}\beta - \tfrac{3}{32}) r_7 + (\tfrac{1}{4}\beta - \tfrac{3}{32}) s_8 + \tfrac{1}{4} r_9\tau - \tfrac{1}{4} s_{10}\tau, \\
C_6 &= -\tfrac{1}{3} r_5 + \tfrac{1}{6} r_6, \\
C_7 &= \tfrac{1}{6} r_6, \\
C_8 &= \tfrac{1}{24} r_4 - \tfrac{1}{12} s_5 + \tfrac{1}{12} s_6, \\
C_9 &= \tfrac{1}{8} r_4 - \tfrac{1}{12} s_5 + \tfrac{1}{12} s_6, \\
C_{10} &= -\tfrac{1}{8} r_8 - \tfrac{1}{24} r_{10}, \\
C_{11} &= -\tfrac{1}{16} r_8 + \tfrac{1}{48} r_{10}, \\
C_{12} &= \tfrac{1}{64} r_7 - \tfrac{1}{64} s_8 + \tfrac{1}{96} r_9, \\
C_{13} &= -\tfrac{1}{32} r_7 + \tfrac{1}{32} s_8, \\
C_{14} &= \tfrac{1}{32} r_7 - \tfrac{1}{32} s_8 - \tfrac{1}{16} r_9.
\end{aligned}
\tag{A.4}
$$

Note, that Malacara‘s indexation is not recommended anymore and has been substituted by Thibos‘ indexation [TASW00].

B Pin Description of the CeHSSA

All digital input pins are of type pull-up. Analog and digital ground pins must be connected together off-chip. See Figure B.1 for a pin diagram and Table B.1 for a description of the I/O pads.

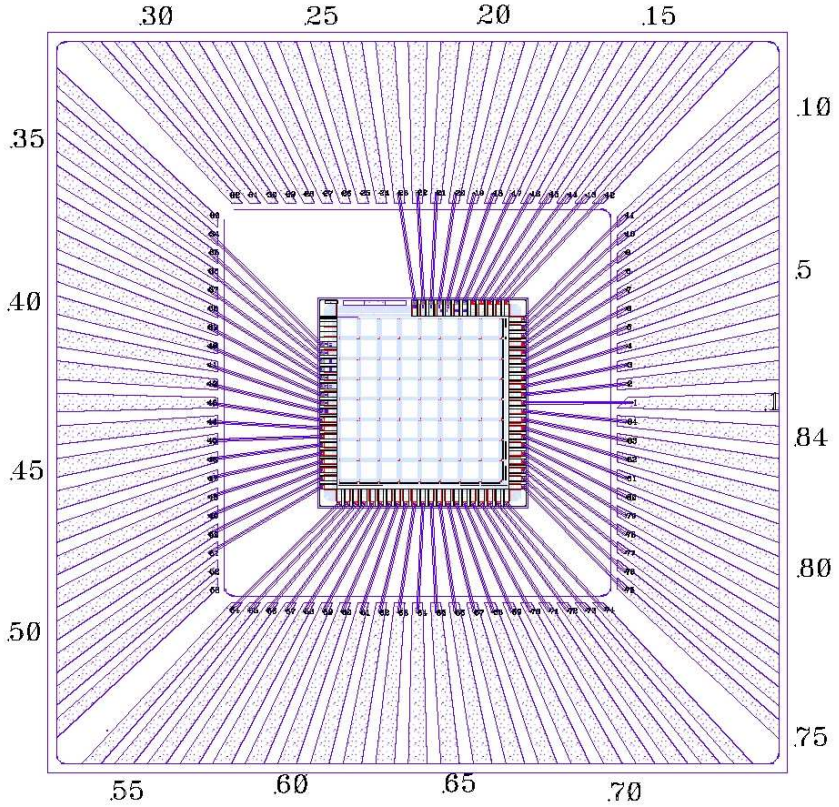

Figure B.1: Pad diagram of the CeHSSA.

pin number	pin name	description
12-14	$YAdr[0-2]$	3-bit decoder y-address
15	$XYMux$	multiplexes y-data to pins 54-74 when high
16	$HLMux$	multiplexes bits 5-20 of the x-bit-vector to the output pins 54-60 and the y-bitvector to pins 75-84,1-6 when high
17,36	$Gnda$	ground potential for the analog components
18,37	Gnd	ground potential for the digital components
19-22	$TD[3-0]$	connections to the three photodiode test-structures n-well/substrate, $p+/n$-well and n-well/substrate sandwich structure, and $n+$/substrate
23,42	V_{dd}	3.3 V supply voltage
38	V_{cas}	analog bias voltage for the cascode MOSFET, typical value 2 V
39	V_{resgrid}	analog bias voltage for the active MOSFET resistors, i.e. connected to ground is centroid mode, connected to V_{dd} is maximum mode, intermediate voltages for smaller spot sizes
40	V_{src}	analog bias voltage for the current sinks, typical value 600 mV
41	V_{biasout}	analog bias voltage for the output amplifier, typical value 400 mV
45	$THold$	holds the state of the tristate bus when high; must be pulled down for read-out and pulled up to write the test pattern to the x-tristate bus
46	$TSelect$	writes the data from the shift register to the tristate bus when also $THold$ is high
47	$TClk$	serial read-in of the test bit-pattern; with each rising edge of $TClk$ the data at TIn is written to a shift register with the test bit-pattern
48	TIn	serial input of the test bit-pattern
49-51	$XAdr[0-2]$	3-bit decoder x-address
54-74	$XData[0-20]$	x bit-vector; in 16-bit read-out mode, the data is available at pins 54-69
75-84,1-11	$YData[0-20]$	y bit-vector; in 32-bit read-out mode, the x-data is available at pins 54-69 and the y-data at pins 75-84,1-6

Table B.1: Description of the I/O pads.

C Pin Description of the Prototype

All digital input pins are of type pull-up. Analog and digital ground pins must be connected together off-chip. See Figure C.1 for a pin diagram and Table C.1 for a description of the I/O pads.

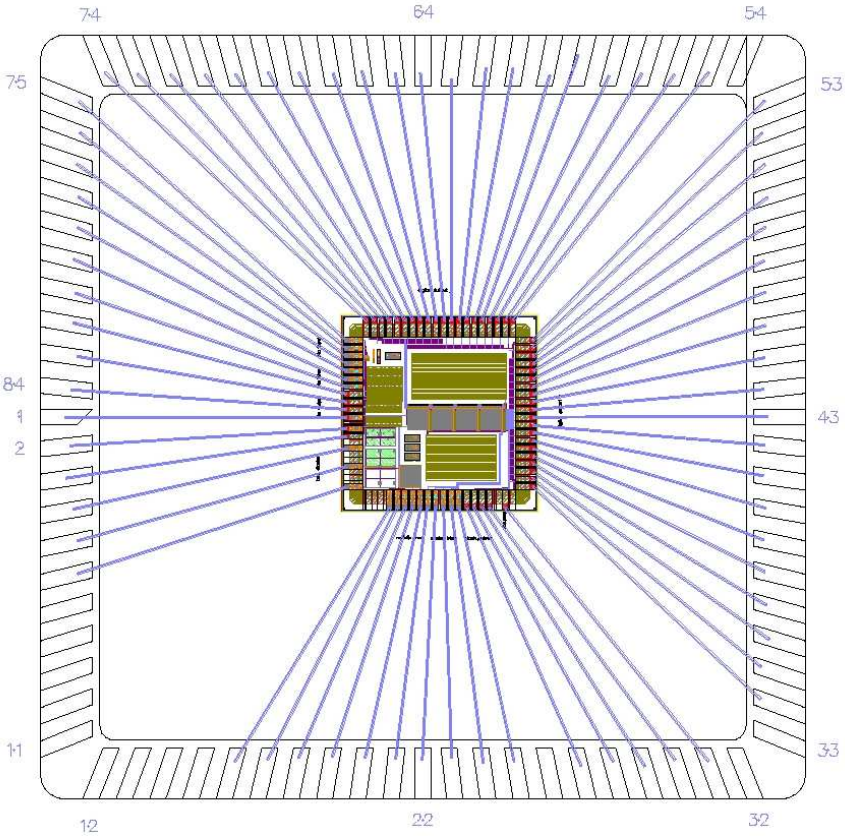

Figure C.1: Pad diagram of the prototype CeHSSA.

pin number	pin name	description
1	V_{dd}	3.3 V supply voltage
2,7	Gnd	ground potential
4-6	$TD[0-2]$	connections to the three photodiode test-structures $n+$/substrate, $p+/n$-well and n-well/substrate
16,17,19,20	I_L,I_R,I_D,I_U	left, right, lower and upper current from the resistive-line network
18	V_{resline}	analog bias voltage for the active resistors in the resistive-line network
21	V_{dummy}	analog bias voltage for current source in the dummy neuron; off, when $V_{\text{dummy}} = 0$
22	V_{cas}	analog bias voltage for the cascode MOSFET, typical value 2 V
23	V_{resgrid}	analog bias voltage for the active MOSFET resistors, i.e. connected to ground is centroid mode, connected to V_{dd} is maximum mode, intermediate voltages for smaller spot sizes
24	V_{src}	analog bias voltage for the current sinks, typical value 600 mV
25	V_{biasout}	analog bias voltage for the output amplifier, typical value 400 mV
27-30	$YAdr[0-2]$	selects detector for output of the y bit-vector; only one must be selected
31	$YDataDummy$	state of the y dummy neuron
34-53	$YData[0-19]$	y bit-vector of the selected detector
55-74	$XData[0-19]$	x bit-vector of the first detector

Table C.1: Description of the I/O pads.

D Comparison between different CMOS-based Wavefront Sensors

At the moment of writing this thesis there have been published two other CMOS-based wavefront sensors to the best knowledge of the author. Table D.1 shows a comparison between the basic parameters of these wavefront sensors.

The HSSX by Droste [Dro99] uses the simple WTA circuit as a PSD and thus performs the maximum operation on the photocurrent distribution. The data is read out serially with an FPGA, that also communicates with a PC via the parallel port. The HSSX can measure the rather large number of 16×16 spots. It has been successfully used as an eye tracker, but the sensitivity was not large enough to capture the spots from a reflection of the retina. Droste implemented a number of innovative ways to get a faster transient response of the WTA circuit and many aspects of his design have been used in the design of the CeHSSA, including the decoupling of the WTA neurons from the parasitic junction capacitance through a cascode and a current mirror.

The wavefront sensor from Lima Monteiro [dLM02] has been tested as a Hartmann Sensor, which detects the beams from an opaque mask with an array of holes, rather than the spots from a lenslet array, but the sensor is also usable in conjunction with a microlens-array as a Hartmann-Shack sensor. This sensor uses quad-cells as PSDs for centroiding and double-junction photodiodes with a rather large parasitic capacitance. Instead of a submicron process technology, the conservative DIMES[1] 1.6 μm process technology has been used. The currents from the quad-cells are read out through an analog data bus and externally converted into digital signals. This sensor has been tested successfully in a closed-loop adaptive optical system. The quad-cells show the typical nonlinear response but offer a very good resolution at spot powers in the μW-range. Spot tracking in the nW-range or below for opthalmic applications is not available yet with this sensor.

All presented wavefront sensors show several advantages, for the quad-cell being the good resolution and for the pixelized sensors being the large sensitivity and linearity. The large sensitivity of the pixelized sensors is also the result of the immediate digitization of the data, directly where the photocurrents are generated, which makes them less sensitive to noise. Future CMOS-based wavefront sensors can reach a good resolution even at very low spot intensities.

[1] Delft Institute of Microelectronics and Submicron Technology

design	T. Nirmaier	D. Droste	D.W. de Lima Monteiro
CMOS technology	AMS 0.35 μm	AMS 0.6 μm	DIMES 1.6 μm
chip area	4.1 $\times$ 4.1 mm^2	7.7 $\times$ 8.2 mm^2	10.0 $\times$ 10.0 mm^2
PSD type	resistive-ring network of WTA circuits	WTA circuit	quad-cells
operation	maximum or centroid	maximum	centroid
detector pitch	400 μm	400 μm	1 mm
number of spots	8$\times$8	16$\times$16	8$\times$8
photodiode type	n-well/substrate	$p+/n$-well	sandwich $p+/n$-well and n-well/substrate
fill-factor	80%	70%	95%
random access	yes	no	yes
quantum efficiency	40% @ 680 nm	10% @ 680 nm	40% @ 633 nm
pixel capacitance	0.01 pF	0.12 pF	50 pF
max. bandwidth	6 kHz	1 kHz	3 kHz
resolution	17.0 μm @ 1 nW/spot	17.6 μm @ 1 nW/spot	0.8 μm @ 12.0 μW/spot 47.1 μm @ 0.2 μW/spot

Table D.1: Comparison between different CMOS-based wavefront sensors.

E List of Symbols

symbol	description	unit
A	area	m^2
b	number of adjacent winning neurons	-
B	bandwidth	Hz
$\mathbf{B}$	reconstruction matrix	-
c	speed of light	m/s
c_i	spatial mode coefficient	-
c_j	photodiode junction capacitance per unit area	F/m^2
$\mathbf{c}$	vector of spatial mode coefficients	-
C_j	photodiode junction capacitance	F
C_{js}	depletion-region capacitance	F
C_n^m	Zernike coefficient	dpt
C_{ox}	gate-oxide capacitance	F
C_A	power of astigmatism	dpt
C_Z	power of defocus	dpt
Δd	diameter of the Airy disc	m
D_n	electron diffusivity	m^2/s
D_p	hole diffusivity	m^2/s
D_L	lens diameter	m
e	elementary charge	C
f	focal length	m
f_i	spatial mode	-
Δf	noise bandwidth	Hz
F	Fresnel number	-
$\mathfrak{F}$	position estimator	-
g,G	transconductance	A/V
g_D	effective power of the telescope	m
h	Planck constant	J·s
i_n	photodiode shot noise	A
i_r	photodiode thermal noise	A
I_{in}	input current	A
I_{out}	output current	A
I_p	photocurrent	A

I_s	photodiode dark current	A
I_{lat}	lateral current	A
I_{src}	current through current sink	A
I_{tot}	total DC current	A
I_D	drain current	A
I_{D0}	parameter for the subthreshold current	A
j_D	current density of drain currents	A/m
j_{src}	current density of current sinks	A/m
J	light intensity (irradiance)	W/m^2
J_0	maximum light intensity	W/m^2
J_B	background light intensity	W/m^2
J_T	transmitted light intensity	W/m^2
J_T	threshold light intensity	W/m^2
k	Boltzmann constant	J/K
k	number of neurons in a WTA circuit	-
l	diffusion length	m
L	transistor channel length	m
L_n	electron diffusion length	m
L_p	hole diffusion length	m
M	number of data points	-
M_L	confidence level for spot detection	-
n	number of input currents	-
n	subthreshold parameter	-
n_f	frequency of false bits	-
n_i	intrinsic charge carrier density	1/m^3
N	number of data points	-
N_A	acceptor impurity density	1/m^3
N_D	donator impurity density	1/m^3
p	pixel pitch	m
$\overline{p}$	effective resolution	m
p_{err}	probability of erroneous WTA outputs	-
p_F	probability of false spot detection	-
p_R	probability of false spot rejection	-
P	spot power	W
q	charge	C
r,R	resistance	Ω
R_j	photodiode parallel junction resistance	Ω
R_s	photodiode series resistance	Ω
R_L	load resistance	Ω
$R_\square$	resistance of minimum size square	Ω
$\mathbf{s}$	spot displacement column vector	-
t	time	s
T	temperature	K

U	circumference of the resistive-ring network	m
v	spot velocity over detector	m/s
V	voltage	V
V_{biasout}	output amplifier bias voltage	V
V_{cas}	cascode bias voltage	V
V_{dd}	positive supply voltage	V
V_{dummy}	dummy-neuron bias voltage	V
V_D	junction potential	V
v_{DS}, V_{DS}	drain-source voltage	V
v_{GS}, V_{GS}	gate-source voltage	V
V_{resgrid}	resistive-network bias voltage	V
V_{SB}	source-bulk voltage	V
V_t	threshold voltage	V
V_T	thermal voltage	V
w	number of WTA circuits	-
W	wavefront	m
W	transistor channel width	m
W_J	junction width	m
x	coordinate	m,-
x_0	spot position in x	m
Δx	spot displacement in x	m
y	coordinate	m,-
y_0	spot position in y	m
Δy	spot displacement in y	m
z	coordinate	m
Z	Zernike polynomial	-
α	optical absorption coefficient	1/m
α	angle	-
α_A	angle of astigmatism	-
ϵ_{Si}	silicon permitivity	C/(V·m)
η	quantum efficiency	-
λ	wavelength	m
λ	channel-length modulation factor	-
μ	channel mobility	m^2/(V·s)
ρ	sheet resistance	Ω/m
ρ	normalized radial distance	m
θ	angle with the x-axis	-
σ	spectral reflectivity	-

Bibliography

[AB94] A.G. Andreou and K.A. Boahen. *Analog VLSI signal and information processing*, chapter Neural information processing II, pages 358–413. McGraw-Hill, 1994.

[AMSa] Austria Mikro Systeme AMS. *0.35 μm CMOS Design Rules*.

[AMSb] Austria Mikro Systeme AMS. *0.35 μm CMOS Process Parameters*.

[AMSc] Austria Mikro Systeme AMS. *0.35 μm Digital Standard Cell Library*.

[App00] R.A. Applegate. Limits to vision: Can we do better than nature. *J. of Refractive Surgery*, 16, 2000.

[BCM02] J.F. Bille, J. Buchler Costa, and F. Mueller. *Aberration-free Refractive Surgery*, chapter Optical Quality of the Human Eye: The Quest for Perfect Vision. Springer, 2002.

[BLB97] R. J. Baker, H.W. Li, and D. E. Boyce. *CMOS Circuit Design, Layout and Simulation*. Wiley, 1997.

[BNDZ93] T.M. Bernard, P.E. NGuyen, F.J. Devos, and B.Y. Zavidovique. A programmable VLSI retina for rough vision. *Machine Vision and Applications*, 7(1):4–11, Winter 1993.

[BW62] E.A. Boerttner and J.R. Wolter. Transmission of the ocular media. *Invest. Opthalmol.*, 1:776–783, 1962.

[BZD93] T.M. Bernard, B.Y. Zavidovique, and F.J. Devos. A programmable artificial retina. *IEEE Journal of Solid State Circuits*, 28(7):789–798, July 1993.

[Cha91] W.N. Charman. *Visual Optics and Instrumentation*, volume 1 of Vision and Visual Dysfuction. MacMillan Reference, Ltd., London, 1st edition edition, 1991.

[Chu03] O. Chutatape. Online Source: Nanyang Technological University Singapor, http://www.ntu.edu.sg/home/eopas/Fundus.html, 2003.

[DB02] D. Droste and J. Bille. An ASIC for Hartmann-Shack wavefront detection. *IEEE Journal of Solid State Circuits*, 37, 2002.

[Dew92] S.P. Deweerth. Analog vlsi circuits for stimulus localization and centroid. *International Journal of Computer Vision*, 8:191–202, 1992.

[dLM02] D.W. de Lima Monteiro. *CMOS-based integrated wavefront sensor*. Delft University Press, 2002.

[dLMVS99] D.W. de Lima Monteiro, G. Vdovin, and P.M. Sarro. Various layouts of analog CMOS optical position-sensitive detectors. In *Proc. SPIE*, volume 3794, pages 134–142, 1999.

[DM95] S.P. DeWerth and T.G. Morris. CMOS Current mode Winner-take-all circuit with distributed hysteresis. *Electronics Letters*, 31(13), 1995.

[Dro99] D. Droste. *Realisierung eines Wellenfrontsensors mit einem ASIC*. PhD thesis, University of Heidelberg, 1999.

[ED01] S. Esposito and N. Devaney. Segmented telescopes co-phasing using pyramid sensor. In *Beyond Conventional Adaptive Optics*, 2001.

[fAO01] Center for Adaptive Optics. Annual report. Technical report, University of California, Santa Cruz, 2001.

[FIA01] E.J. Fernandez, I. Iglesias, and P. Artal. Closed-loop adaptive optics in the human eye. *Optics Letters*, 26(10), May 2001.

[FNQS95] F. Ferrari, J. Nielsen, P. Questa, and G. Sandini. Space variant imaging. *Sensor Review*, 15(2):17–20, 1995.

[GAS90] R.L. Geiger, P.E. Allen, and N.R. Strader. *VLSI design techniques for analog and digital circuits*. McGraw-Hill, Inc., 1990.

[Geh02] A. Gehner. Active-matrix addressed micromirror array for aberration correction of the human eye. Summer School on Adaptive Optics, August 2002. Santa Cruz, CA, USA.

[GHLM00] P.R. Gray, P.J. Hurst, S.H. Lewis, and R.G. Meyer. *Analysis and Design of Analog Integrated Circuits*. Wiley, 2000.

[Har78] J.W. Hardy. Adaptive optics: A new technology for the control of light. In *Proceedings of the IEEE*, volume 66, 1978.

[HAS+01] H. Hofer, P. Artal, B. Singer, J.L. Aragon, and D.R. Williams. Dynamics of the eyes wave aberration. *J. Opt. Soc. Am.*, 18(3):497–506, 2001.

[HCY+01] H. Hofer, L. Chen, G.Y. Yoon, B. Singer, Y. Yamauchi, and D.R. Williams. Improvement in retinal imaging quality with dynamic correction of the eye's aberrations. *Optics express*, 8(11):631–643, 2001.

[Hor75] G.A. Horridge. *The Compound eye and vision of insects*. Clarendon Press, Oxford, 1975.

[HS91] G.A. Horridge and P. Sobey. An artificial seeing system copying insect vision system. *International Journal of Optoelectronics*, 6(1/2):177–193, 1991.

[Jäh97] B. Jähne. *Image Processing for Scientific Applications*. CRC Press LLC, 1997.

[KDS+00] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts. A logarithmic response CMOS image sensor with on-chip calibration. *IEEE J. of Solid State Circuits*, 35(8), August 2000.

[Kof00] M. Kofler. *Linux Installation, Konfiguration, Anwendung*. Addison-Wesley, 2000.

[LGGB94] J. Liang, B. Grimm, S. Goelz, and J. Bille. Objective measurements of wave aberrations of the human eye with the use of a Hartmann-Shack wave-front sensor. *J. Opt. Soc. Am.*, 11(7), July 1994.

[Lia98] J. Liang. *Hartmann-Shack wavefront sensor*. PhD thesis, University of Heidelberg, 1998.

[Loo99] M. Loose. *A Self-Calibrating CMOS Image Sensor with Logarithmic Response*. PhD thesis, University of Heidelberg, 1999.

[LRMM89] J. Lazarro, S. Ryckebusch, M.A. Mahowald, and C.A. Mead. *Advances in neural information processing systems*, volume 2, chapter Winner-Take-All Networks of O(n) Complexity, pages 703–711. D.S. Touretzky, San Mateo- CA, 1989.

[Lyo81] R.F. Lyon. The optical mouse, and an architectural methodology for smart digital sensors. In *CMU Conference on VLSI Structures and Computations*. Computer Science Press, October 1981.

[Mal92] D. Malacara. *Optical Shop Testing*. Wiley and Sons, 1992.

[Mea89] C. Mead. Adaptive retina. In C. Mead and M. Ismail, editors, *Analog VLSI implementation of neural systems*, chapter 10, pages 239–246. Kluwer Academic Publishers, Boston, 1989. Proceedings of a workshop on Analog Integrated Neural Systems.

[Mül01] F. Müller. *Konzeption und Entwicklung eines adaptiv-optisch korrigierten Laser-Scanning Retina-Tomographen*. PhD thesis, University of Heidelberg, 2001.

[Moi00] A. Moini. *Vision Chips*. Kluwer Academic Publishers, 2000.

[NDB02] T. Nirmaier, D. Droste, and J. Bille. Hartmann-Shack sensor ASICs for real-time adaptive optics in biomedical physics. In *The 6th world multiconference on systemics, cybernetics and informatics*, volume 13, pages 280–284. IIIS, July 2002.

[NPB03] T. Nirmaier, G. Pudasaini, and J. Bille. CMOS-based wavefront sensor for real-time adaptive optics. In *Proceedings of the 13th IEEE Real Time Conference*, May 2003. accepted for publication.

[NV98] A.F. Naumov and G.V. Vdovin. Multichannel lc-based wavefront corrector with modal influence functions. *Opt. Lett.*, 23:1550–1552, 1998.

[OK97] I.E. Opris and T.A. Kovacs. A high-speed median circuit. *IEEE J. of Solid State Circuits*, 32(6):905908, June 1997.

[PS71] B. Platt and R.V. Shack. Lenticular Hartmann screen. *Opt. Science Center Newsletter*, 5(1), 1971.

[PV90] I. Pitas and A. Venetsanopoulos. *Nonlinear digital filters*. Kluwer Academic Publisher, 1990.

[Sch02] M. Schottner. *Algorithms for the application of Hartmann-Shack wavefront sensors in opthalmology*. PhD thesis, University of Heidelberg, 2002.

[SF93] J.A. Startzyk and X. Fang. Cmos current mode winner-take-all circuit with both exitatory and inhibitory feedback. *Electronics Lettters*, 29(10), 1993.

[SGLB98] T. Serrano-Gotarredona and B. Linares-Barranco. A high-precision current-mode WTA-MAX circuit with multichip capability. *IEEE J. of Solid State Circuits*, 33(2):280–286, February 1998.

[ST97] R.F. Schmidt and G. Thews. *Physiologie des Menschen*. Springer, 1997.

[Sta91] D.L. Standley. An object position and orientation IC with embedded imager. *IEEE J. Solid State Circuits*, 26(12):1853–1859, December 1991.

[TASW00] L.N. Thibos, R.A. Applegate, J.T. Schwiegerling, and R. Webb. Standards for reporting the optical aberrations of eyes. *Vision Science and its applications. Opt. Soc. Am. Technical Digest*, pages 146–149, 2000.

[Teo97] M. Teodori. *Design und Messung optoelektronischer Strukturen in CMOS Technologie*. University of Heidelberg. Diploma Thesis, 1997.

[Tür00] A. Türpitz. *Entwicklung eines adaptiv-optischen Closed-Loop-Systems zur Kompensation der Aberrationen des menschlichen Auges*. PhD thesis, University of Heidelberg, 2000.

[UN95] K. Urahama and T. Nagao. K-Winner-take-all circuit with O(n) complexity. *IEEE Transactions on Neural Networks*, 6(3), May 1995.

[VMP97] G.V. Vdovin, S. Midelhoek, and P.Sarro. Technology and applications of micromachined silicon adaptive mirrors. *Opt. Eng.*, 36(5):1382–1390, 1997.

[vP02] U. von Pape. *Wavefront Sensing in the Human Eye*. PhD thesis, University of Heidelberg, 2002.

[WJK00] B.M. Wilamowski, D.L. Jordan, and O. Kaynak. Low power current mode Loser-take-all circuit for image compression. In *9th NASA Symposium on VLSI Design*, 2000.

[Zer34] F. Zernike. Beugungstheorie des Schneideverfahrens und seine verbesserte Form der Phasenkontrastmethode. *Physica*, 1:689, 1934.

Acknowledgments

Working on this interdisciplinary project gave me the chance to take a look at the latest developments in microelectronics, but also in biomedical physics and optics. To succeed in areas that different from one another I needed the help of many people, without them this thesis would have not been possible to realize.

First of all I want to thank my parents Anita and Karl-Heinz and my wife Teresa for their continuous support. They gave me the possibility to go through my studies with patience and understanding.

I also want to thank the following people for their support and company:

Prof. Dr. Josef Bille for the interesting project, the possibility to participate in the postgraduate course of lectures at the German Cancer Research Center (DKFZ) and for the examination of this thesis,

Prof. Dr. Reinhard Männer from the University of Mannheim for the interest in this work and taking on the second referee,

Dr. Dirk Droste for support at the beginning of my thesis, for providing me with the layout of the HSSX and the software,

Tobias Kuhn and Dr. Ulrich von Pape for the optical setup, being willing to be the first probands and interesting discussions about wavefront sensors,

Gopal Pudasaini for joining the efforts of making better wavefront sensors and various political discussions while climbing,

Dr. Davies William de Lima Monteiro and Prof. Dr. Gleb from the Technical University of Delft for their hospitality and especially Davies for some of the graphs,

Thorsten Maucher, Dr. Johannes Schemmel, Andreas Breidenassel, Andreas Grübl and various other people from the Vision Group at the Kirchhoff Institute for good ideas, but even more interesting discussions in unofficial meetings,

Klaus Greger, Lisa Kierig, Nina Korablinova and all others at the Biomedical Optics Group for help in many aspects.

I also want to thank the people from the ASIC lab, especially Markus Dorn and Ralf Achenbach for their help with the design tools, and the Kirchhoff Institute for Physics. My thanks are extended to the company 20/10 Perfect Vision, for their support and the lenslet array.