CMOS-BASED INTEGRATED WAVEFRONT SENSOR

CMOS-BASED INTEGRATED WAVEFRONT SENSOR

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. J. T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
maandag 4 november 2002 om 10:30 uur

door

Davies William de LIMA MONTEIRO

Master of Science (Federal University of Espírito Santo, “Brazilië”)
geboren te Volta Redonda - Rio de Janeiro, Brazilië

Dit proefschrift is goedgekeurd door de promotoren:
Prof.dr. P. J. French
Prof.dr.ir. A. J. P. Theuwissen

Toegevoegd promotor:
Dr. G. V. Vdovin

Samenstelling promotiecommissie:

Rector Magnificus	Technische Universiteit Delft, voorzitter
Prof.dr. P.J. French	Technische Universiteit Delft, promotor
Prof.dr.ir. A. Theuwissen	Technische Universiteit Delft, promotor
Dr.G.V. Vdovin	Technische Universiteit Delft, toegevoegd promotor
Prof.dr. J.W. Swart	State University of Campinas, Brazil
Prof. dr. R. Dändliker	University of Neuchatel, Switzerland
Dr.ir. J. H. Correia	University of Minho, Portugal
Dr.ir. F. A. van Goor	Technische Universiteit Twente

Reserve lid:
Prof.dr.ir. G. C. M. Meijer Technische Universiteit Delft

Published and distributed by: DUP Science

DUP Science is an imprint of
Delft University Press
P.O. Box 98
2600 MG Delft
The Netherlands
Phone: (+31)-15-2785678
Fax: (+31)-15-2785706
E-mail: DUP@DUP.TUDelft.NL

ISBN 90-407-2344-3

Key words: wavefront, sensor, adaptive optics, position sensing, silicon, photodetector, CMOS, aberration, Hartmann

Printed in The Netherlands

Contents

Introduction 1

Optical components such as mirrors and lenses are ubiquitous in modern societies; their application ranges from spectacles to cameras, from compact-disc players to satellite guiding systems.

Deviations from the optimal profile of an optical component can severely affect its performance, the consequences of which are experienced on a daily basis by a substantial part of the world population when the natural lenses of their eyes are, for instance, astigmatic or myopic.

Any particular application of an optical component or system has its own requirements in terms of magnitude and types of aberration that can be tolerated. In order to assess the suitability of the optics, adequate tests are necessary to determine whether either component replacement, reshaping or compensation must be chosen.

Over the years, optical profile tests have evolved from craftsmanship to systematic test procedures and the profile analysis ranges from that of components in the micrometric scale to that of components which are several meters large. The demand for a combination of three test features, namely accuracy, real-time capability and low cost, has been steadily increasing. To date several test methods are available [1.1], but there is still a lack of an inexpensive compact real-time device capable of quantitative analysis.

This thesis addresses the proposition, design and implementation of an integrated device, a *wavefront sensor*, suitable for real-time optical-profile analysis. This sensor can also be coupled to a deformable mirror in order to dynamically detect and correct optical wavefronts.

The test method is based on estimating the local tilts of a *wavefront* after it has been altered by the profile of the optical component under test, as explained in section 1.1. A brief review of optical-profile tests is presented in section 1.2. The motivation and goals of the research presented in this thesis are discussed in section 1.3 and a description of the integrated-wavefront-sensor concept is given in section 1.4. Section 1.5 details the organization of further chapters.

1.1 Optical wavefronts

A wavefront is a hypothetical surface defined as the locus of all points of a given wave featuring the same phase and because of the wave nature of light the concept of wavefronts applies to the optical domain. The definition of phase and the relation between wavefronts and component profile will be addressed in this section.

Phase

For an electromagnetic wave, each component of the electric-field vector $\boldsymbol{E}(\boldsymbol{r},t)$, at a point $\boldsymbol{r}$ in the three dimensional space and instant t in time, must satisfy the vector wave equation (1.1) in order to comply with the Maxwell equations in free space [1.2, 1.3]:

$$\nabla^2 \boldsymbol{E} = \frac{1}{c_o^{\,2}} \frac{\partial^2 \boldsymbol{E}}{\partial t^2} \quad . \tag{1.1}$$

One solution of this equation is given in (1.2) and represents a harmonic transverse plane wave, whose direction of propagation is given by the wave vector $\boldsymbol{k}$, perpendicular to the oscillating field $\boldsymbol{E}$. The electric field is further described by its amplitude $\boldsymbol{E}_0$, angular temporal frequency ω, and initial phase ε.

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}_0 \, e^{i\,(\boldsymbol{k}\cdot\boldsymbol{r} \pm \omega t + \varepsilon)} , \tag{1.2}$$

where the modulus k of the wave vector is called wave number and is related to the wavelength λ by

$$k = (k^2_x + k^2_y + k^2_z)^{1/2} = 2\pi/\lambda , \tag{1.3}$$

and the angular frequency is given by $\omega = 2\pi\nu = 2\pi c_o/\lambda$, where ν is the temporal frequency.

Equation (1.2) can be rewritten as

$$\boldsymbol{E}(\boldsymbol{r}, t) = \boldsymbol{E}_0 \, e^{i\,(\boldsymbol{k}\cdot\boldsymbol{r} + \varepsilon)} \, e^{\pm i\,\omega t} , \tag{1.4}$$

which now contains two independent terms: a quickly oscillating temporal term $e^{\pm i\omega t}$ and a slowly oscillating spatial term $e^{i\,(\boldsymbol{k}\cdot\boldsymbol{r} + \varepsilon)}$, whereupon phase is defined as

$$\varphi = \boldsymbol{k}\cdot\boldsymbol{r} + \varepsilon . \tag{1.5}$$

It is known that the phase velocity c, and consequently the wavelength λ and the wave number k are modified as a wave propagates through a medium with a refractive index $n \neq 1$, and a notation as given by the following equations can be used:

$$c_n = c_o/n , \tag{1.6}$$

$$\lambda_n = \lambda_o/n , \text{ and} \tag{1.7}$$

$$k_n = k_o n . \tag{1.8}$$

Therefore, the phase φ is a function of the refractive index n.

Wavefronts

Due to the linear nature of the wave equation (1.1), any three-dimensional wave can be written as a combination of harmonic plane waves having particular propagation directions, amplitudes and spatial frequencies.

Considering an idealized point light source, spherical waves originating at the source can be described as plane waves travelling uniformly outwards in each radial direction. For the sake of simplicity, Figure 1.1 illustrates the two-dimensional case for transverse plane waves in some of the radial directions at a given instant in time.

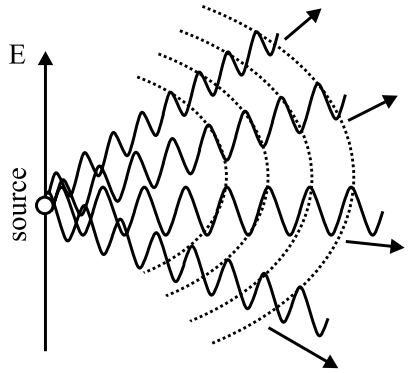

Figure 1.1 *Transverse plane waves originating at a point source.*

Supposing that the wavelength λ and the initial phase ε are equal for all waves originating at the source, then a wavefront is defined as an imaginary line connecting points of every wave featuring the same phase φ. If the propagation medium is linear, homogeneous and isotropic, then the wavefronts can be represented as concentric circles. In three dimensions each wavefront can be thought of as a spherical surface joining all points with equal phase, as depicted in Figure 1.2.

In geometrical optics, a wavefront is defined as a surface having normals parallel to the light rays, where the rays represent the direction of energy propagation. The normals represent the directions of the maximum phase change rate.

Phase, wavefront and optical profile

For two parallel sections of a plane wave emitted by a monochromatic light source, there are two usual situations that cause a delay between the

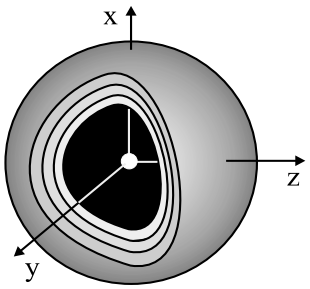

Figure 1.2 *Section of three-dimensional concentric wavefronts propagating out of a point source.*

sections, i.e., a change in either the medium or the path length through which one of the sections propagates. The first case is illustrated in Figure 1.3 for two plane-wave parallel sections, A and B, from point *a* to *a'* and from point *b* to *b'*, respectively. Wave section A passes through a medium with refraction index *n* and length *L*, which introduces a phase delay, $k_o L(n-1)$, with respect to wave section B, which propagates through a medium with *n*=1. At point P_1 both sections feature the same phase, and at P_2 the phase of section B equals the phase of section A at point P'_2. Note that points *a* and *b* are in phase, and points *a'* and *b'* are not. The delay is directly proportional to both the length and the refraction index of the medium. A second case arises when the media are the same but the propagation path of section A is increased by an amount *L* with respect to the path of section B, in which case the phase delay is simply $k_o L$.

The previous example can be extended to as many parallel sections of the plane wave as desired. Therefore, when a wavefront (points with equal phase) that is initially plane propagates though an optical component, such as a lens, local delays are introduced to it according to the lens profile, as in Figure 1.4.

Any mirror that is not flat acts in much the same way by introducing different propagation lengths for each part of the wavefront. Rephrasing the previous comments, an optical component imprints its profile on the incoming wavefront. As a direct consequence, under certain conditions, by measuring the resultant wavefront one is able to determine the component profile.

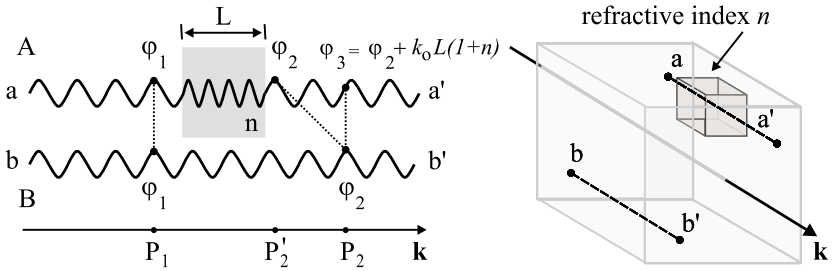

Figure 1.3 *Phase delay between waves A and B introduced by a different medium along the propagation path.*

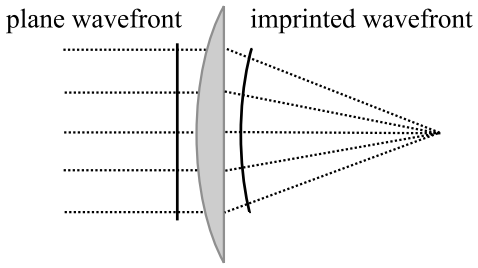

Figure 1.4 *A plane wavefront passing through a lens becomes imprinted with the lens profile.*

1.2 Optical profile tests

As discussed in the previous section a component profile can be determined by measuring the associated imprinted wavefront. This implies that phase should be measured. The existing photodetectors, however, including those of the eye, are only sensitive to the light intensity and not to its phase. Therefore, methods to transduce intensity information into phase information must be used.

The leading solutions in optical-profile testing of smooth macroscopic surfaces are the interferometric methods, which have long been used both qualitatively and quantitatively. The range of devised interferometer configurations dominates almost every aspect of optical shop testing and they often feature high accuracy. In spite of their vast applicability, they do

have inherent limitations that favor the implementation of alternative optical tests based on either irradiance measurements or geometrical optics. The latter are often straightforward in concept and setup, and are usually less costly. A brief overview of each of these categories is presented.

1.2.1 Interferometric methods

Interferometers are instruments that, as the name implies, depend on the interference between light beams. The Twyman-Green interferometer [1.4], shown in Figure 1.5, is a good example of the principles of interferometry. A collimated and expanded laser beam is split into two arms; one heads to a reference mirror and the other to the reflective surface under test. The beams reflected by these two surfaces recombine at the beam splitter and are imaged onto an image plane. The test plane is located so that if a second reference surface - identical to the first - is positioned there, the optical path lengths (*OPL*[1]) of the two beams will be exactly the same on recombination. In practice, the surface to be tested is neither perfect nor identical to the reference one and may have a considerable profile deviation from the reference, which causes local differences in *OPL* on recombination.

If both beams have exactly the same *OPL* from source to screen, or if they differ by an integral multiple of the beam wavelength, then the beams are completely in phase and interact constructively, which results in a bright image. If the *OPL*s differ by an odd integral multiple of half the wavelength, then destructive interference takes place and a dark image will result. A third possibility is when one of the beams has local *OPL*s that differ from those of the other beam. In this case, the result will be a fringe pattern, i.e., a two-dimensional intensity distribution map on the imager.

Although the fringe pattern is not a direct mapping of the phase profile, namely the wavefront, this can be obtained after thorough data evaluation. With basic familiarity with an interferometric system one can qualitatively recognize occurrences of all lower order aberrations by observing an

1. The refractive indices of all traversed media, n_i, multiplied by the corresponding lengths and then summed defines the total *OPL*. More generally: $OPL = \int_S n(\boldsymbol{r})\, d\boldsymbol{r}$, where S is the ray path.

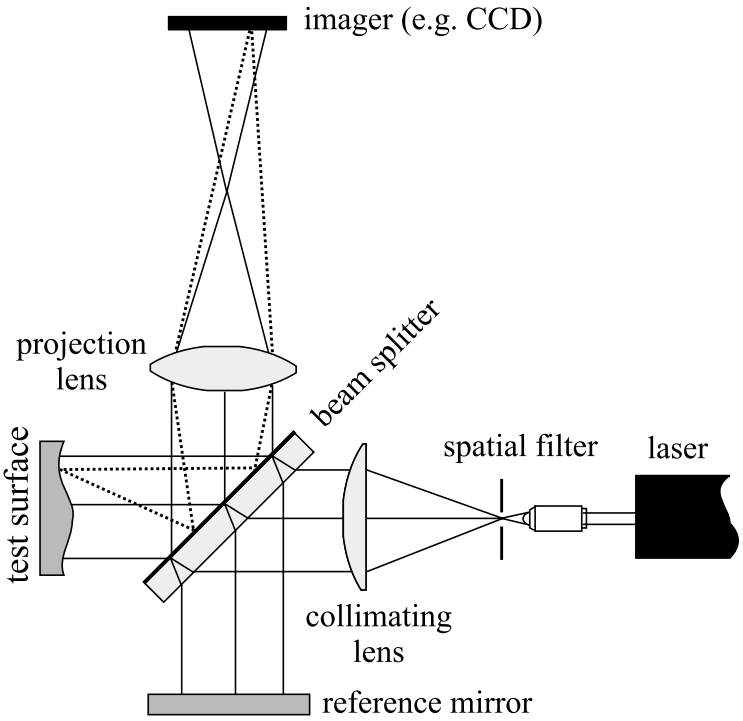

Figure 1.5 *Twyman-Green interferometer*

interferogram (fringe pattern). An accurate quantitative approach takes a considerable computational effort. The analysis of an interferogram poses an ill-defined mathematical problem where the intensity at each point is a function of several unknown variables, among which the phase.

The analytical description of the fringe pattern consists of a sinusoidal signal fluctuation in two-dimensional space, as in equation (1.9):

$$I(x, y) = B(x, y) + C(x, y)\cos[\varphi(x, y)] + N(x, y) \tag{1.9}$$

where $I(x,y)$ is the intensity at each point, $B(x,y)$ is the background illumination contribution, $C(x,y)$ is the fringe contrast, $\varphi(x,y)$ is the phase and $N(x,y)$ is the imager noise [1.5].

Finding the phase distribution is not straightforward and the reconstruction method differs for interferograms generated by different instrument configurations. In the shearing interferometer, for instance, the wavefront under test is superimposed on a sheared image of itself.

In general, one is only able to estimate phase from equation (1.9) within very specific conditions, since it represents an ill-posed mathematical problem. The main drawbacks in estimating the phase from an interferogram are:

- coexistent functions on the intensity map (background, contrast and noise) partially screen the phase function $\varphi(x,y)$;
- phase periodicity only allows its unique determination to within $|2\pi|$, termed phase ambiguity;
- the sign of the phase is masked by the even nature of the cosine function.

Methods to overcome the above problems do exist, but are either calculation intensive or require multiple exposures [1.6], rendering real-time analysis unfeasible. For modern optical testing three main branches exist: vertical scanning interferometry (VSI) [1.7], multiple-wavelength interferometry (MWI) and phase-shifting interferometry (PSI) [1.8]. The first uses a white-light (broad-band) source and can handle rough surfaces, with accuracy in the nanometer range. MWI requires two or more wavelengths and is able to measure steep but smooth surfaces with sub-nanometer accuracy. PSI is the most popular solution, although limited to measuring surfaces that are smooth and with low local variations ($< \lambda/4$ variations). It uses a highly coherent and intense source (usually a He-Ne laser) so that the resulting fringe pattern has a high contrast, therefore favoring high accuracy.

Usual practical inconveniencies of interferometers include their inability to operate with pulsed sources, sensitivity to vibration, double-pass compensation and the usage of a mechanical scan stage in some cases. Also, they often require several high quality optical components, a reference beam and a highly coherent source. The latter can introduce undesired interference fringes from several optical surfaces in the system.

1.2.2 Irradiance methods

Irrandiance methods are based on the measurement of the radiant-flux densities on different planes. The principle is that there exists a relationship between the irradiances at these planes that provides information on the wavefront profile. There are two well-known techniques based on different physical bases and mathematical protocols: phase retrieval and curvature sensing.

Phase retrieval

Amplitude and phase of a wave at the input plane affect the projected intensity. Traditional phase retrieval is based on the assumption that the far-field as well as the near-field intensity patterns are known, but not the phase distribution (wavefront) at the input plane [1.9, 1.10]. Irradiance is then measured at both the input and the far-field plane.

A initial guess is made about the wavefront at the input plane, which along with the input intensity are used to estimate the far-field intensity pattern. This is compared to the measured far-field pattern and, based on the deviation between the two patterns, the inverse calculation is performed in order to estimate a better trial input wavefront; this now serves as the input for a second iteration. This process undergoes several iterations until the calculated far-field irradiance matches the measured one, thus presumably identifying the correct input wavefront.

Despite its practical success for most common cases, there is to date no mathematical proof that a particular far-field irradiance distribution can be uniquely attributed to a specific wavefront [1.11]. Besides the non-uniqueness of the solution, the complex propagation algorithm and iteration approach often take too long to allow real-time diagnostics.

Curvature sensor

Another widely used method was proposed in the late 1980s, and is called the curvature sensing[2] method [1.12, 1.13]. It is a variation on the

2. It is called curvature sensing because it involves measuring the local second-order derivative of the wavefront, by using a relationship between irradiance and phase that contains second-order derivatives of the wavefront, through the Laplacian operator ∇^2 [1.14].

traditional phase retrieval method.

The principle relies on measuring almost simultaneously the irradiances at two closely spaced planes located at the same distance but at opposite sides of a reference plane where a pupil is present. Knowing the irradiances and the pupil edge geometry one is able to solve a second-order differential equation whose solution is the wavefront function.

The proposed iteration algorithms to solve this equation are not trivial and rather time consuming. As regards the setup, this technique requires movable parts: either the detector should be moved from one measuring plane to the other or some adjustable optics to re image each plane on the detector should be used.

1.2.3 Geometrical methods

The concepts for these methods benefit from the simplest and plainest approach to optics: ray optics [1.15]. In some cases, however, diffraction needs to be approached as a secondary phenomenon. Although fundamentally useful, the most primitive method in this category is the knife-edge test, followed by more elaborate ones as the Ronchi test, Hartmann method, laser ray tracing and pyramidal sensing. Recently, a combination of curvature and Hartmann methods in a single setup has been proposed too.

Knife-edge test

The Foucault knife-edge test is meant to test spherical surfaces with a point source and has been used since the late 19th century [1.16]. Observe the setup to test a concave mirror in Figure 1.6. The mirror is illuminated by a point source and an observation screen is placed outside the focus. A razor blade (knife edge) is placed near the focus and if it slowly intercepts the converging beam inside the focus from the bottom towards the line *AB*, the blade shadow on the initially bright screen will move from the top towards the image center. If, on the other hand, the interception is done outside focus, the shadow will move in the same sense and direction of the blade, i.e., from the bottom towards the center. In case the blade intercepts the line *AB* exactly at the focus point, the image changes suddenly from fully bright to completely dark.

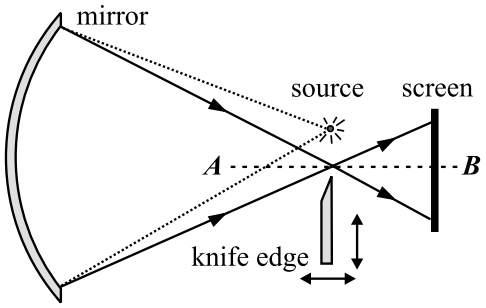

Figure 1.6 *The knife-edge setup to test a concave mirror.*

If the perfect mirror is replaced by an aberrated one, and the razor blade is inserted at the position of the main focus point, the whole set of reflected rays will no longer be simultaneously blocked. This is due to the fact that local deviations on the mirror surface introduce different focal points to different sets of rays. Therefore, an image of the mirror imperfections, called *shadowgram*, will be formed on the screen with bright and dark regions representing the unblocked and blocked sets of rays, respectively. This is a robust qualitative method, but rather complex to extract the magnitude of the aberrations in a quantitative way.

This method can be performed with white light and the setup is straightforward. Disadvantages are the poor sensing of aspherical surfaces with large asphericity and the inability to provide quantitative data.

Ronchi test

Although this test can be fully explained on the basis of diffraction and interferometry, it was originally devised as a geometrical test. The Ronchi test uses a screen with interleaving transmissive and opaque straight stripes, called Ronchi ruling [1.17]. To test a mirror, the ruling is suitably placed between the mirror and an imager, near the center of curvature of the mirror. A light source illuminates the mirror and what is observed through the ruling is the superposition of the ruling itself with its own image on the mirror surface. The superimposed images form a fringe pattern that is related to the aberrations on the mirror surface. The Ronchi setup can actually be regarded as a shearing interferometer, where a

number of diffraction orders laterally overlap. The fringe spacing is independent of the wavelength, which enables the use of white light.

The geometrical interpretation is valid as long as coarse gratings are used (about 20 lines/mm) [1.18]. The stripes can be thought of as a multiple knife-edge test and the distorted fringes can be explained by the fact that the aberrated surface under test reflects ray sets with different foci. It is known that a perfectly spheric mirror surface generates a pattern of equally spaced fringes on the detector. When the incident wavefront is not spheric, an associated fringe pattern is observed. Transverse aberrations with gradients orthogonal to the direction of the stripes can be sensed. In case the aberration is not radially symmetric two orthogonal fringe patterns are necessary. This is one of the limitations of this test, because either only one ruling is used and rotated or a beam is split towards two separate perpendicular rulings, or the two perpendicular rulings are superimposed, in which case a more complex fringe-analysis algorithm has to be used.

Hartmann test

The Hartmann method is a more direct technique based on a single mask and it offers straightforward data interpretation [1.19, 1.20]. It consists of an opaque mask with a grid of sub-apertures placed at a distance *D* from an imager. A wavefront at the mask is sampled into as many spots as the number of sub-apertures. The position of each spot at the imager is proportional to the respective local tilt of the wavefront.

Consider the case of a plane wavefront - parallel to the mask - as reference, then each spot centroid on the imager is located exactly under its respective sub-aperture, defining a regular grid of spots. Any aberrated wavefront, e.g. a tilted wavefront, causes the spots to depart from the reference positions, as can be seen in Figure 1.7. In other words, local wavefront tilts (transverse wavefront gradients) at the region of each sub-aperture cause the corresponding rays (orthogonal to the wavefront) to be displaced from the perpendicular direction. If one knows the spot centroid displacement and the distance *D* between the mask and detector, one can determine the local tilt directly.

One of the unfavorable characteristic of this method is the fixed number of sub-apertures that deprives the reconstruction algorithm of contiguous spatial information. But this can be overcome to a large extent by

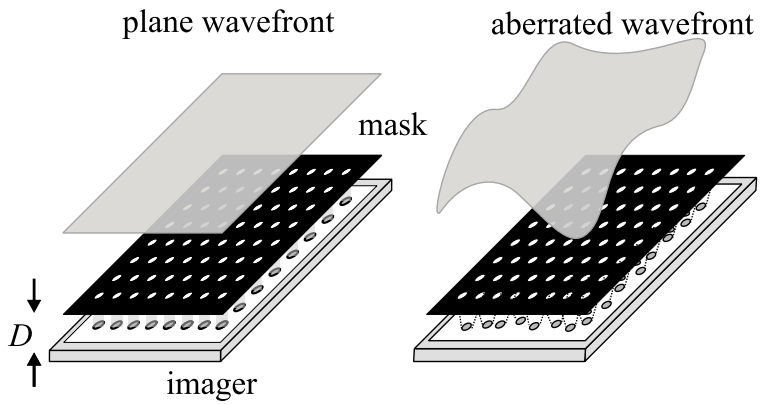

Figure 1.7 *Scheme of the Hartmann setup in the case of an incident plane wavefront and an aberrated one.*

substituting the mask with another one with a higher density of holes. Another limitation, mostly not significant, is the lack of information about the wavefront at the boundaries. In spite of that, this method is robust in the sense that there are no moving parts and that there is no requirement for monochromatic sources. Also, in principle, it can deal with both continuous and pulsed light sources, the recorded displacements are directly related to the wavefront shape and no intrinsic additional optics are required. Besides, the angular sensitivity, and hence the detection range of aberration magnitude, can be simply adjusted by changing *D*. A variation to this method is the Hartmann-Shack setup [1.21, 1.22], in which the Hartmann mask is substituted with an array of microlenses which have the advantages of gathering more light power per spot and covering the whole sensing area.

Laser ray tracing (LRT)

Laser ray tracing is an object scan technique based on the concept of software ray tracing analysis [1.23]. Instead of a light beam sampled into several spots, as in the Hartmann test, a single collimated laser beam scans a raster of points on the aberration plane, Figure 1.8. This plane is imaged on the surface of an imager. The scan is done with the aid of a moving mirror. The reference laser spot positions are recorded and subsequently

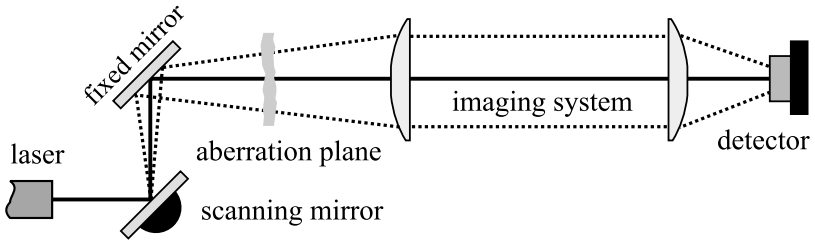

Figure 1.8 *Laser ray tracing setup.*

the test object is placed in position and a new scan is performed. The transmitted (or reflected) spots through (from) a test object are imaged onto the imager, the displacements are computed and the wavefront is calculated in basically the same way as in the Hartmann method.

The advantage of this technique is that it allows as many raster points as necessary, even if they generate spatially overlapping spots, since each spot position is recorded independently. Indexation becomes trivial and optical crosstalk in the detector is nonexistent. Also, the optical power is optimized. On the other hand, the need of a scanning device introduces jitter and results in a much longer overall acquisition time than a Hartmann sensor. Basically, the LRT is an active method, whereas the Hartmann technique is a passive method.

Pyramidal sensor

Another interesting geometrical solution is the pyramidal sensor [1.24]. A lens focuses the test beam (wavefront) into the central vertex of a pyramidal plate, which has four faceplates and splits the focal spot into four separate images projected on a screen, Figure 1.9. The pyramidal plate actually works as a four-fold knife-edge test in the sense that each of the four split images is equivalent to a particular image that would have been obtained if the knife edge had intercepted the focal point along different directions. Figure 1.9 illustrates the setup.

A plane wavefront is focused on the vertex of the pyramid as an Airy pattern and all spatially equivalent points (P_1, P_2, P_3 and P_4) in the four sub-images have the same intensity. Conversely, an aberrated wavefront results in a blurred spot on the vertex of the pyramidal plate, which will

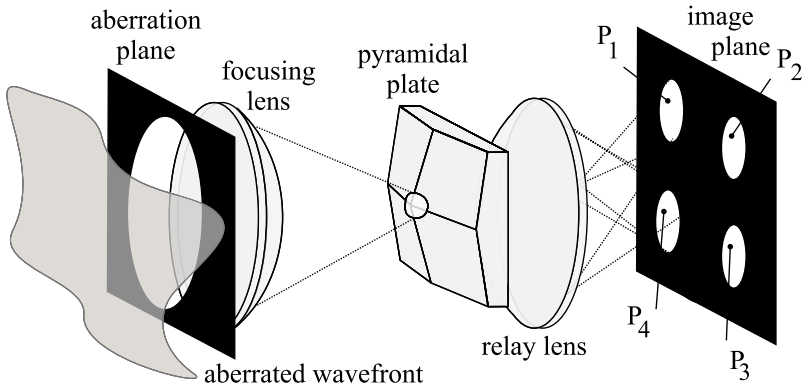

Figure 1.9 *Pyramidal sensor setup*

consequently care for different intensity values at each of the four equivalent points. Any local transverse aberration of the incoming wavefront can be calculated by a simple mathematical relationship between the intensities at the corresponding four points [1.25].

The number of sampling points in each sub-image is only limited by the number of pixels in the respective sub-region of the detector. The size of the sub-regions can be changed by modifying the relay lens. The vertex angle should be such that the sub-images do not overlap. This method is very sensitive to tilt.

Hybrid curvature and gradient sensor

The hybrid curvature and gradient sensor has not been experimentally demonstrated yet, but is a promising concept [1.26]. It employs one detection plane and an array of astigmatic lenses, resembling the Hartmann-Shack setup. The shape of the light spot at the focus point of such a lens has a clear dependence on the curvature of the test wavefront. Therefore, an array of foci enables one to obtain information on the local curvature as well as on the local tilts (gradients), resulting in a high spatial frequency resolution. This sensor would enable a good coupling to a curvature-based Adaptive Optics system.

1.3 Motivation and goals

In spite of the diversity of methods in the realm of wavefront sensing, each technique has a limited range of applications. In general, compactness, low cost, easy calibration, repeatability, high accuracy, large dynamic range and real-time measurement are desired features. Among the previously discussed methods, the Hartmann(-Shack) test is one of the most promising in achieving most of these characteristics for the following reasons:

- *compactness* - only a sampling plane (a mask with holes or a microlens array) and a detector are necessary;
- *low cost* - for the same reason as the previous. In addition, the detector can be the same as for all other methods, usually a CCD. For interferometers, the CCD needs a higher resolution[3] to provide adequate fringe data, whereas in the Hartmann test correct location of spot centroids can be achieved with a lower resolution;
- *repeatability* - no moving parts are present and the test is insensitive to vibrations if mask and detector are appropriately fixed;
- *high accuracy* - it is dependent on the number of sub-apertures (which is mostly a matter of choice for each application) and on the *wavefront sensitivity*, which is the minimum aberration magnitude detectable and is dictated by detector noise and the reconstruction algorithm. This figure is not as good as for the best interferometers (e.g. phase-shifting interferometry with sub-nanometer accuracy);
- *aberration magnitude range* - this can be adjusted by changing the distance between mask and detector, and eventually the mask pitch;
- *fast wavefront reconstruction* - because of the linear relationship between spot displacement and local tilt. Once the displacements are known, a linear system of equations can be promptly solved to find the appropriate coefficients necessary to reconstruct the wavefront (section 2.3), which substantially reduces the mathematical complexity inherent to other methods. The reconstruction time becomes mostly dependent on the detector response time and, when an imager is used, on the bitmap evaluation too.

3. The number of fringes depends on the type and range of aberrations. For instance, the higher the magnitude, the more fringes and the higher the resolution demanded [1.5].

Amid all these potential features the least developed is *real-time quantitative analysis*. Dynamic holography and sequential interferometric recording, for instance, do provide ways to analyze dynamic profile changes after data is processed, i.e., data is first acquired and later interpreted. In the Hartmann method, conventional imagers are used that are typically limited to frame rates ranging from 25 to 60 images per second. The images must be further processed to supply information on spot displacements. This limits the speed to less than 25 Hz, in low-cost systems. A trade-off between frame rate and cost can be made by using a dedicated CCD, but the need for an image-processing step and special hardware still remains.

The wavefront sensor described in this thesis essentially tackles the operational speed issue, while sticking to standard and potentially low-cost solutions and conceptual simplicity. The idea is to implement a sensor based on the Hartmann method that replaces the imager (CCD) by a customized integrated detector, which provides nearly direct information about the spot positions. This speeds up the whole detection and computational process. Such a sensor can then be especially employed where fast phase variations occur.

1.4 Integrated wavefront sensor

A couple of requirements have been set when devising the proposed device. It should be fast, compact and use standard technologies. As discussed previously, the Hartmann method seems to be most appropriate to meet these demands.

The chief idea is to substitute the imager chip with an integrated matrix of optical position-sensitive detectors (PSDs), as shown in Figure 1.10, such that each PSD is associated to a single spot sampled by the Hartmann-Shack raster. Therefore, the computer performs trivial calculations to find out the spots displacements, rendering image processing dispensable. In addition, the PSD signals are multiplexed on chip to reduce the number of output channels. Last but not the least, silicon Complementary Metal-Oxide-Semiconductor (CMOS) technology is used to offer a standard solution and to enable integration of control electronics and any further added circuitry.

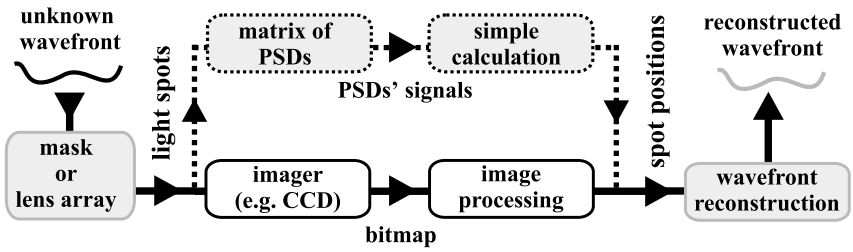

Figure 1.10 *Hartmann wavefront sensing diagram. Conventional (solid line) and proposed (dashed line) approaches.*

1.5 Organization of chapters

This thesis reports on the design, implementation and performance of a customized integrated wavefront sensor aimed at operating at higher speeds than conventional wavefront sensing methods while keeping conceptual simplicity. This introductory chapter reviewed the major characteristics of various wavefront determination methods and introduced the motivation for the chosen approach. The following chapters will detail several aspects for the project realization.

Chapter 2 offers a handful of concepts including wavefront description, reconstruction algorithm, photodetection mechanisms and optical position detection. It will establish a basis for subsequent discussions while giving some preliminary insight on the choices that were to be made in the realization of this wavefront sensor.

Chapter 3 addresses technological issues regarding the Complementary Metal-Oxide-Semiconductor (CMOS) process, sampling plane and optical detection.

Chapter 4 describes the overall implementation ranging from the position-sensitive detector pixels to the on-chip and off-chip electronics. Different architectures are compared and layout and simulation of various constituent stages of the chip are commented.

Chapter 5 is devoted to results and analysis. Capabilities and limitations of the wavefront sensor are presented and its performance in a closed-loop adaptive optical system is discussed.

Chapter 6 offers the concluding remarks and possibilities for further development.

The very last section of every chapter, except the last one, contains a concise summary of the contents addressed, where the most important points are highlighted. Its goal is to give some readers the opportunity to recall the thread of the discussions and to spare some other readers the burden of reading through all chapters.

1.6 Chapter summary

In this chapter we investigated several optical profile tests that rely on different methods: interferometric, irradiance and geometrical methods. We concluded that each has its particular strengths and limitations. We pointed out various desirable features expected from optical profiling, and for the tests, we indicated the lack of an inexpensive compact device capable of real-time operation and quantitative analysis. Therefore, we proposed the implementation of a Hartmann-method wavefront sensor that uses a matrix of integrated position-sensitive detectors (PSDs), compatible with standard Complementary Metal-Oxide-Semiconductor technology (CMOS), instead of a conventional imager. The idea is to have one PSD per sampled spot, such that direct information about the spot-centroid position is available, circumventing the need of image processing.

References

[1.1] D. Malacara, “Optical shop testing,” 2nd edition, John Wiley & Sons Inc., New York, 1992

[1.2] B. Saleh and M. Teich, “Fundamentals of Photonics,” 157-174, John Wiley & Sons Inc., New York, 1991

[1.3] E. Hecht, “Optics,” 3rd edition, 43-46, Addison Wesley Longman Inc., Reading, 1998

[1.4] D. Malacara, “Optical shop testing,” 2nd edition, Chapter 2, John Wiley & Sons Inc., New York, 1992

[1.5] M. Servin and M. Kujawinska, “Modern fringe pattern analysis in interferometry,” in Handbook of optical engineering, 373-375, edited by D. Malacara and B. Thompson, Marcel Dekker Inc., New York, 2001

[1.6] M. Servin and M. Kujawinska, “Modern fringe pattern analysis in interferometry,” in Handbook of optical engineering, 376-377, edited by D. Malacara and B. Thompson, Marcel Dekker Inc., New York, 2001

[1.7] A. Harasaki, J. Schmit and C. Wyant, “Improved vertical-scanning interferometry,” Applied Optics, vol.39, no.13, 2107-2115, 2000

[1.8] D. Malacara, M. Servín and Z. Malacara, “Interferogram analysis for optical testing,” Chapter 7, Marcel Dekker Inc., New York, 1998

[1.9] G. Vdovin, “Reconstruction of an object shape from the near-field intensity of a reflected beam,” Applied Optics, vol.36, no.22, 1997

[1.10] W. Southwell, “Wavefront analyzer using a maximum likelihood algorithm,” J. Opt. Soc. Am. A, vol.67, no. 3, 396-399, 1977

[1.11] J. Foley and R. Butts, “Uniqueness of phase retrieval from intensity measurements,” J. Opt. Soc. Am. A, vol.71, no. 8, 1008-1014, 1981

[1.12] F. Roddier, “Curvature sensing and compensation: a new concept in adaptive optics,” Applied Optics, vol.27, no.7, 1223-1225, 1988

[1.13] F. Roddier, “Variations on a Hartmann theme,” Optical Eng., vol.29, no.10, 1239-1242, 1990

[1.14] R. Tyson, “Principles of adaptive optics,” 2nd edition, p.171, Academic Press, San Diego, 1998

[1.15] E. Hecht, “Optics,” 4th edition, Chapter 5, Addison Wesley, San Francisco,

2002

[1.16] J. Ojeda-Castañeda, “Focault, wire and phase modulation tests,” in Optical shop testing, 2nd ed., 265-288, edited by D. Malacara, John Wiley & Sons Inc., New York, 1992

[1.17] A. Cornejo-Rodriguez, “Ronchi test,” in Optical shop testing, 2nd ed., Chapter 9, edited by D. Malacara, John Wiley & Sons Inc., New York, 1992

[1.18] J.D. Briers, “Interferometric testing of optical systems and components: a review,” Optics and laser technology, February, 1972

[1.19] I. Ghozeil, “Hartmann and other screen tests,” in Optical shop testing, 2nd ed., Chapter 10, edited by D. Malacara, John Wiley & Sons Inc., New York, 1992

[1.20] J. Hartmann, “Bemerkungen über den Bau und die Justirung von Spektrographen,” Zt. Instrumentenkd, vol.20, 47, 1900

[1.21] R. Shack and B.Platt, “Production and use of a lenticular Hartmann screen,” Proc. Spring Meeting of the Optical Society of America, MG23, 64, April 1971

[1.22] R. Tyson, “Principles of adaptive optics,” 2nd edition, 163-164, Academic Press, San Diego, 1998

[1.23] R. Navarro and E. Moreno-Barriuso, “Laser ray-tracing method for optical testing,” Optics Letters, vol.24, no.14, 951-953, 1999

[1.24] R. Ragazzoni and J. Farinato, “Sensitivity of a pyramidic wavefront sensor in close-loop adaptive optics,” Astronomy and Astrophysics, vol.350, no.2, L23-L26, 1999

[1.25] A. Riccardi, N. Bindi, R. Ragazzoni, S. Esposito and P. Stefanini, “Laboratory characterization of a Foucault-like wavefront sensor for Adaptive Optics,” Proc. SPIE, vol.3353, 941-951, Kona, March 1998

[1.26] C. Paterson and J. C. Dainty, “A hybrid curvature and gradient wavefront sensor,” Optics Letters, vol.25, no.23, 1687-1689, 2000

Theoretical concepts 2

As discussed in the previous chapter, the geometry of a component can be tested by measuring the probe wavefront transmitted through (or reflected by) it. Among several techniques to analyze this imprinted wavefront, we selected the Hartmann method as being the most promising, especially in achieving real-time analysis and because of its suitability for integration.

This chapter covers a number of theoretical concepts related to wavefront description, Hartmann technique, data handling, photodetection and spot-position sensing. It provides the building blocks for later discussions on technological aspects and sensor performance. Section 2.1 addresses the wavefront mathematical description; section 2.2 elaborates on the Hartmann method; section 2.3 indicates how the wavefront can be reconstructed from the Hartmann spots; section 2.4 has a discussion on photodetection and in section 2.5 spot position sensing layouts are commented upon. Although a wavefront sensor is an invaluable adaptive optical instrument, a brief introduction to an adaptive optical system is postponed to Chapter 5, where the coupling to the adaptive mirror becomes clear.

2.1 Wavefront description

In practice, we are only interested in the *patch* of the wavefront that propagates through the component under test. Henceforth we will refer to this patch as the wavefront, unless indicated otherwise.

Hence, assuming that the wavefront Ξ_W propagates along the *z-axis*, it can be described as a function of *x* and *y*:

$$\Xi_W = F(x, y) \ , \tag{2.1}$$

where *F(x,y)* is a function defined over the pupil plane of interest, as can be seen in Figure 2.1. In general propagation alters the wavefront profile, so if the detector is placed away from the aberration plane, the measured wavefront will not correspond to the original one. To avoid this situation the exit pupil must be imaged on the detector plane by means of a lens or a lens system.

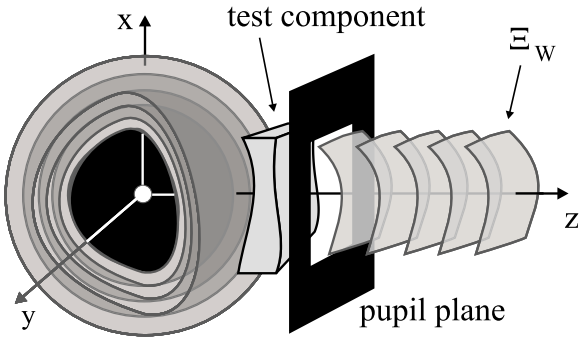

Figure 2.1 *Patch of a wavefront selected by a square aperture (pupil plane).*

There are two commonly used ways to express Ξ_W mathematically: discretely and continuously, also referred to as *zonal* and *modal*. In the first, values are attributed to *F(x,y)* at *N* discrete points (x_i,y_i), for $i = 0,\ldots, (N-1)$; one is then able to visualize the surface, for instance, by connecting

adjoining points. In the second, the function *F(x,y)* is described as a weighted sum of *M* modes, i.e., sub-functions $f_i(x,y)$, *i = 0,..., (M-1)*; each mode is defined all over the pupil plane.

Wavefronts are smooth and the zonal approach has the disadvantage of having many 'broken points' to describe it; it offers a broken interface to a real wavefront, whereas the modal method offers a smooth interface except at the pupil edges. Concise discussions on these two approaches are given below.

2.1.1 Zonal approach

If a grid with N *zones* is defined within a fixed area on the *x-y* plane, we can represent a surface by attributing a height value *z* per zone. The representation is a 'broken' surface whose approximation to a smooth surface enhances with an increasing number of zones. A surface with any spatial frequency can be correctly described when the number of zones approaches infinity. Figure 2.2 gives an example of the accuracy of wavefront description for different numbers of zones.

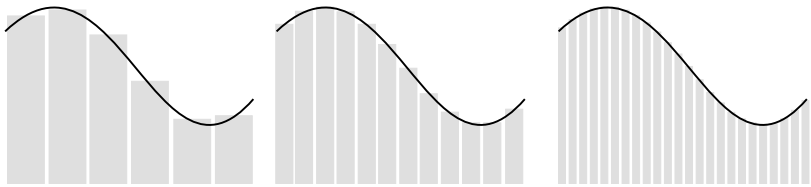

Figure 2.2 *A given wavefront, plotted with the solid line, is represented with 6 zones (left), 12 zones (middle) and 24 zones (right).*

2.1.2 Modal approach

In this case the wavefront is described as a superposition of two-dimensional smooth functions, called *spatial modes*. This means that a number of modes over a pupil can be adequately summed to generate a certain wavefront. A one-dimensional representation is given in Figure 2.3

Analytically, the modal representation can be written as in equation (2.2), where f_0 is a constant representing a plane wavefront. It is convenient

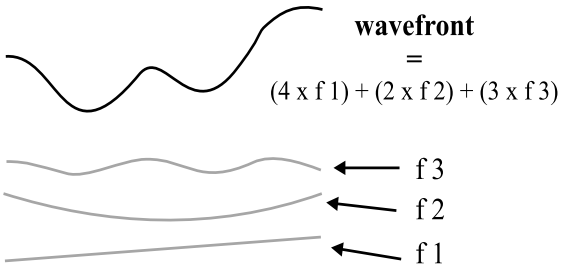

Figure 2.3 *Wavefront represented by three spatial modes.*

in the sense that a particular wavefront is approximated by a set of basis functions $\{f_i(x,y)\}$ when one attributes appropriate values to the coefficients C_i. The disadvantage of this representation is that high spatial frequencies require a very large number of terms to describe the wavefront correctly.

$$\Xi_W = \sum_{i=0}^{M-1} C_i \, f_i(x,y), \quad \forall (x,y) \supset pupil \quad . \tag{2.2}$$

Zernike polynomials

Most optical components, apertures and beams are circular, which demands an orthonormal set of functions over a circle. Of many possible orthogonal sets of functions, a very convenient set consists of Zernike polynomials, after the Dutch physicist F. Zernike, because the most frequent aberrations in optical systems can be represented by low-order Zernike polynomials, usually termed primary aberrations, as presented in Figure 2.4. The average of each term over the pupil is zero and any local amplitude is confined within the interval [-1,1]. Each term definition is conveniently associated with a known profile aberration occurring in optical shop testing. It is complete and defined over a unit-radius circle; each term of the series is normalized. The magnitude and measurement unit are carried by the respective Zernike coefficient C_i.

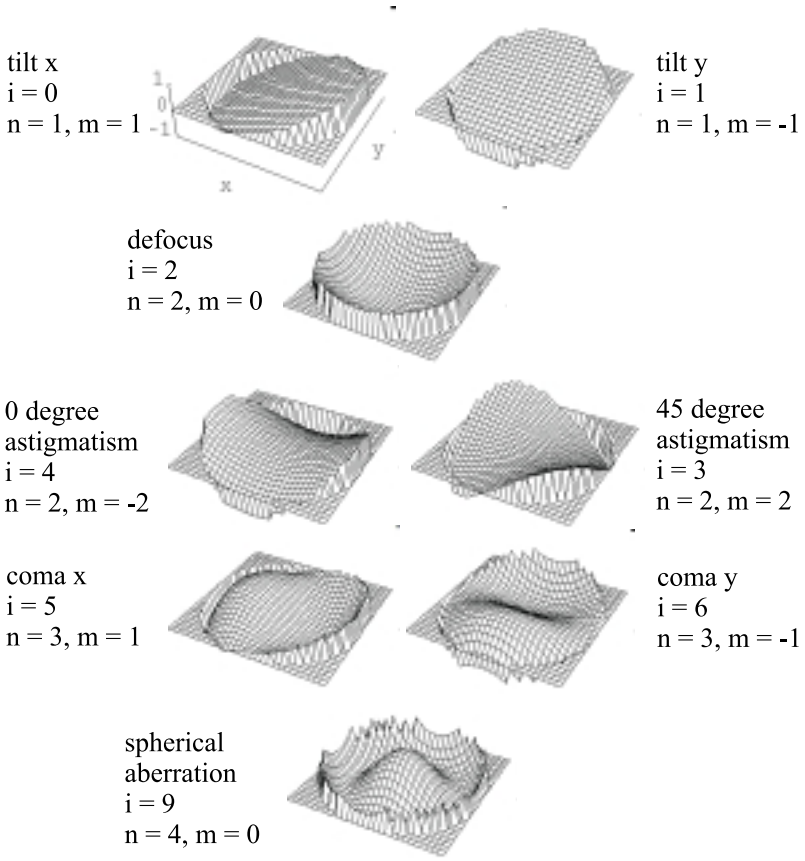

Figure 2.4 *Most often occurring Zernike terms and their indices.*

The circular nature of the polynomials favors a description in polar coordinates (ρ,θ). Thus, an arbitrary wavefront Ξ_W, with respect to a plane wavefront Ξ_{W0}, can be represented by Zernike polynomials $Z_i(\rho,\theta)$ as in equation (2.3)[2.1][1]:

1. Note that the plane wavefront can be considered as a constant term, which simply represents *piston* added to the wavefront as a whole.

$$\Xi_W - \Xi_{W0} = \sum_{i=0}^{\infty} C_i \; Z_i(\rho, \theta) \text{ , where} \tag{2.3}$$

$$Z_i(\rho, \theta) = Z_{n,m}(\rho, \theta) = \begin{cases} R_n^m(\rho)\cos m\theta, & m > 0 \\ R_n^m(\rho)\sin m\theta, & m < 0 \\ R_n^m(\rho), & m = 0 \end{cases} . \tag{2.4}$$

The indices *n* and *m* must obey a set of rules:

- *n* and *m* are integers, where $n > 0$;
- *n* and *m* have the same parity (they are either odd or even);
- for a given *n*, *m* ranges from -*n* to *n* with step 2.

The indices *n* and *m* can be variously associated to a single index *i* for a given number of polynomial terms `M`. The following routine in C defines the indexation we use:

```
int index_transform (M)
int *n, *m;
{
  int i, j, k;
  i = k = 0;

  while (i <= M)
   {for (j = k % 2; j <= k; j += 2)
     { n[i]=k;       m[i]=j;
       i++;
       if (j != 0)
        { n[i]=k;    m[i]=(-j);
          i++;
        }
     }
     k += 1;
   }
  j = i - 1;
  for (i=0; i < j; i++)
   { n[i] = n[i + 1]; m[i] = m[i + 1];
   }
return j;
}
```

For the first nine Zernike polynomials, the indices are given in Table 1:

i	0	1	2	3	4	5	6	7	8
n	1	1	2	2	2	3	3	3	3
m	1	-1	0	2	-2	1	-1	3	-3

Table 1: Relation between indices *i* and *(n,m)*.

The radial term in equations (2.4) is given by [2.1]:

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s\,(n-s)!}{s!\;[(n+m)/2-s]!\;[(n-m)/2-s]!}\,\rho^{n-2s}. \qquad (2.5)$$

Normalization as well as indexation of the polynomials can vary for different authors [2.1, 2.2, 2.3], therefore it is always advisable to check the description of each term before comparing results.

2.2 The Hartmann method

The concept of the Hartmann method was discussed in section 1.2.3. In short, an incoming wavefront is sampled by a mask with an array of apertures. Supposing the incoming light beam is perfectly collimated and impinges perpendicularly on the mask, then each sampled spot centroid is located exactly under the center of the respective sample hole. However, if the beam is convergent, divergent or if it features local deviations, the spots are displaced proportionally to the local wavefront slopes (tilts). An illustration of the cross sections of the Hartmann scheme for such cases is shown in Figure 2.5. The mathematical relations associated with the Hartmann method are also introduced in this section.

The concept of the Hartmann method is intrinsically geometrical, i.e., it can be explained by ray optics, where each portion of a wavefront is associated with a ray normal to it. A perfectly collimated beam is defined as having plane wavefronts, or alternatively, a set of parallel light rays. If, for instance, the wavefront has a concave profile towards the direction of propagation, the corresponding rays converge. This means that if there is aberration it can be translated as transverse ray aberration.

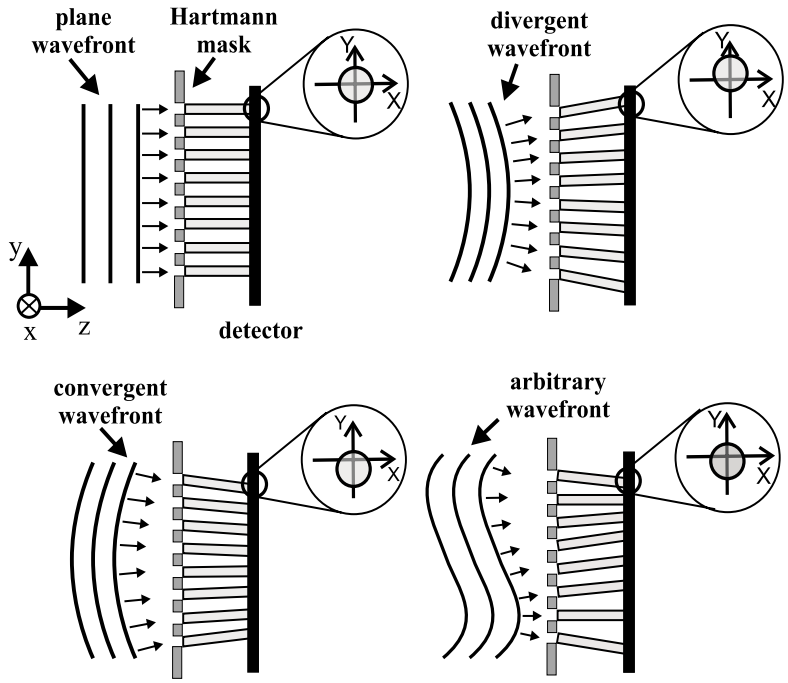

Figure 2.5 *Cross section of the Hartmann sensor. Spot centroid positions for different incoming wavefront profiles.*

The exact relation between transverse ray aberration and wavefront aberration, with respect to a reference sphere, is given by [2.4]; an approximation is given by equations (2.6), as long as the aberration magnitude is much smaller than the radius of curvature of the sphere.

$$\frac{\partial W}{\partial x} \approx \frac{\Delta x}{R}, \qquad \frac{\partial W}{\partial y} \approx \frac{\Delta y}{R} \quad . \tag{2.6}$$

In these equations the wavefront aberration function *W(x,y)* is defined at every point *(x,y)* over the aberration pupil as the optical path difference (*OPD*) between the aberrated wavefront Ξ_{ab} and a perfect sphere *S* of radius *R*, as shown in Figure 2.6. Δx and Δy are the ray aberrations on the plane of the center of curvature.

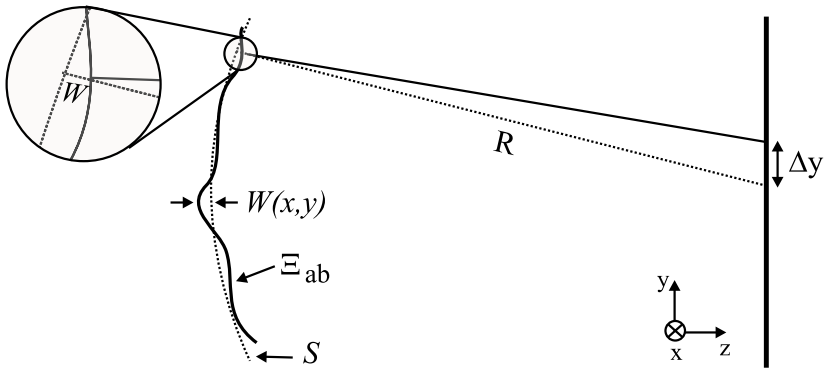

Figure 2.6 *Wavefront aberration W and transverse ray aberration* Δy.

This relationship is then directly applicable to the Hartmann method. Let a plane wavefront represent the reference and an aberrated wavefront be sampled by a Hartmann mask. This is shown in Figure 2.7 for a single aperture.

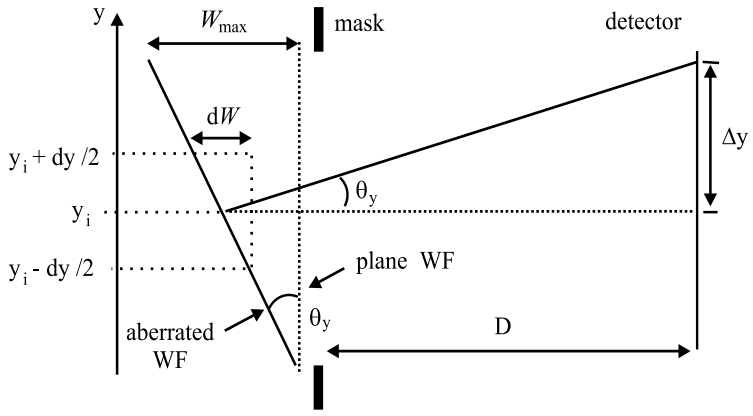

Figure 2.7 *Illustration of the relation between wavefront tilt in one direction and ray aberration for a single aperture of the Hartmann mask.*

Assuming that the wavefront has no aberration other than tilt at each aperture, the displacement in a given direction, measured from the reference position (x_i, y_i), is directly proportional to the corresponding tilt (slope) in that direction. The tilt can be decomposed in components x and y, as in equations (2.7), where i represents a given sub-aperture and $D \gg W_{max}$, where D is the distance between the mask and the detector plane and W_{max} is the maximum aberration magnitude over the aperture.

$$\left.\frac{dW}{dx}\right|_{(x_i, y_i)} = \tan\theta_{x_i} = \frac{\Delta x_i}{D} \quad , \qquad \left.\frac{dW}{dy}\right|_{(x_i, y_i)} = \tan\theta_{y_i} = \frac{\Delta y_i}{D} \quad . \tag{2.7}$$

In practice, the registered displacement on the detector plane is that of the centroid of the spot intensity distribution[2]. The displacements of the centroids of two arbitrary distributions are shown in Figure 2.8. The spot centroid coordinates represent its displacement from the center, as given by equations (2.8) and (2.9).

$$\Delta x = \frac{\iint H(x, y)\, x\, dxdy}{\iint H(x, y)\, dxdy}, \tag{2.8}$$

$$\Delta y = \frac{\iint H(x, y)\, y\, dxdy}{\iint H(x, y)\, dxdy}, \tag{2.9}$$

where $H(x,y)$ is the intensity distribution function.

2. The intensity distribution is in this case described by wave optics as the near-field image (Fresnel diffraction) on the detector resulting from light passing through a finite aperture. The far-field image (Fraunhofer diffraction) is a particular case of the Fresnel diffraction.

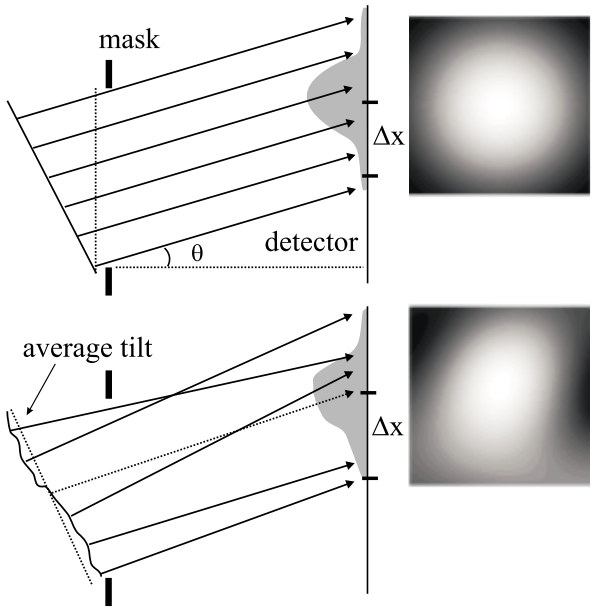

Figure 2.8 *Average tilt and spot intensity profile for one sampling aperture.*

Spatial frequency and aliasing

In the Hartmann method the wavefront is sampled, which can lead to all intricacies of sampled data: loss of detail, incorrect sensing of the magnitude or spatial frequency and the occurrence of high-frequency components as fake low-frequency terms.

A wavefront can be simply composed by a DC component (plane wavefront parallel to the mask) or by a combination of several spatial frequencies, which will be correctly sampled across the aperture or not depending on the density of sampling points along the corresponding direction and their finite size, where each sampling point (interval) corresponds to either a hole or a microlens.

For points uniformly spaced along an arbitrary direction ξ, the sampling frequency f_s along this direction is defined as the inverse of $pitch_\xi$. According to the Nyquist-Shannon theorem, only spatial frequencies lower than $f_s/2$ will be correctly sampled and it is safe to define $f_{max} = f_s/4$

ensuring that any frequency lower than this will be sampled with at least four points per cycle. If an input frequency component f is higher than $f_s/2$, it will be (under)sampled as a low-frequency component, which is called *aliasing*.

It is desirable not only to correctly sense frequencies within the test wavefront spatial bandwidth but also to guarantee that the amplitude of each component is properly taken into account. There is a parameter, termed the *geometrical* modulation transfer function (MTF_g) that represents the ratio of the sensed modulation to the input modulation and depends on the architecture of the sampling points. It is widely used for solid-state imagers [2.5], in which it is related to the size of the pixels. Since each pixel is a sampling region rather than a point, it gathers information from a scene by averaging over its light-sensitive area. This effect compromises the image contrast. The higher the MTF_g, the higher the contrast.

In the case of a wavefront sensor, each sub-aperture (hole, for example) is a finite sampling region where tilts are averaged. This results in a spot position representing an average rather than a punctual tilt value, consequently affecting the proper reconstruction of the wavefront. An expression for MTF_g in an arbitrary direction ξ, with holes spaced regularly with $pitch_\xi$, can be derived from the equation 5.33 in reference [2.5] for an imager and is given by equation (2.10):

$$MTF_g = sinc\ [\pi\, h_d\, f\,] , \tag{2.10}$$

where $sinc\ [x] = x^{-1} sin\ [x]$, h_d is the hole diameter and f is the input frequency along ξ. Since $pitch_\xi f_s = 1$, equation (2.10) can be rewritten as in (2.11).

$$MTF_g = sinc\left[\pi\, \frac{h_d}{pitch_\xi}\ \frac{f}{f_s}\right]. \tag{2.11}$$

Supposing the spatial frequencies present in the input wavefront are such that $f_{max} = f_s/4$, the minimum MTF_g expected is given by equation (2.12):

$$MTF_g^{min} = sinc\left[\frac{\pi}{4}\frac{h_d}{pitch_\xi}\right]. \quad (2.12)$$

Since $h_d \le pitch_\xi$, the argument of the *sinc* function is always smaller than $\pi/4$. Therefore, for a fixed pitch value, the smaller the hole, the higher the geometrical modulation transfer function. But, the smaller the hole, the more aliasing to the input signal [2.5], if any frequency component is higher than $f_s/2$.

In conclusion, for a fixed pitch, the smaller the hole, the higher the MTF_g (desirable), but also the higher the susceptibility to frequency aliasing (undesirable). For each particular spatial bandwidth of the input wavefronts, a trade-off between MTF_g and aliasing is established by two parameters: array pitch and hole size.

Also, the hole should be sufficiently large to prevent large secondary diffraction orders and to allow enough optical power through it. The hole size is directly associated with the spot size on the detector plane, therefore the smallest size for the hole might also be limited by the detector characteristics.

An array of adjoining microlenses collects more light than a Hartmann mask and each local tilt is averaged over a large area (i.e. microlens area). Holes, on the other hand, leave a substantial part of the field not sampled, therefore some information on the wavefront shape is lost.

If the detector plane is placed out of the focal plane of the microlens array, the detection of the spot-centroid position depends not only on the local tilt, but also on the spatial intensity distribution of the spot (Fresnel diffraction). This might introduce an error in the wavefront reconstruction. The same kind of error is present in a setup with a Hartmann mask, unless the detector plane is located at a distance $D >> (h_d/2)^2/\lambda$ from the plane of the mask, where h_d is the mask-aperture diameter and λ is the wavelength used[3]. For $D >> (h_d/2)^2/\lambda$ Fraunhofer diffraction predominates and renders an intensity pattern mostly independent of the intensity distribution at the mask plane, except for a lateral scaling factor [2.6].

3. The dimensionless Fresnel number $F_n \equiv (h_d/2)^2/\lambda D$ determines whether the Fresnel ($F_n \ge 1$) or the Fraunhofer ($F_n \ll 1$) diffraction dominates.

Local tilt range

The maximum tilt magnitude per aperture is related to the spatial dynamic range $||DR||_S$ of the detector and the distance between the latter and the mask. $||DR||_S$ is the maximum spatial interval the spot centroid is allowed to move. It is based on two requirements: two neighboring spots should not overlap and each spot should remain in a reliable position-detection interval; it is given for each detector (CCD, PSD matrix, etc.), and depends on the detector architecture and the density and size of spots. Thus, the maximum local tilt is given by equation (2.13),

$$tilt_{max} = \frac{\|DR\|_S}{2D}, \tag{2.13}$$

where D is the distance between the mask and the detector. The minimum correctly detectable tilt, on the other hand, is dependent on the position resolution δ_R of the sensor system, which in its turn depends on the position detection mechanism, centroid calculation algorithm, detector noise, fluctuation of the light source and system speckle effects. The minimum tilt is given by equation (2.14).

$$tilt_{min} = \frac{\delta_R}{D}. \tag{2.14}$$

It can be seen that both the maximum and minimum local tilts can simply be changed by changing the distance D. Increasing the distance favors applications where good tilt sensitivity is required, at the expense of lower spatial dynamic range, whereas decreasing this distance favors the detection of large tilts, with a coarser sensitivity. Figure 2.9 shows a diagram illustrating the previous discussions.

2.3 Wavefront reconstruction

The Hartmann mask located at a distance D from the detection plane samples an unknown wavefront that causes each spot i to displace Δx_i and Δy_i from a given reference position (x_i , y_i). The local tilts (slopes) along the x and y direction are given by $S_i^x = \Delta x_i/D$ and $S_i^y = \Delta y_i/D$, respectively. A mathematical method to evaluate the shape of the sampled wavefront from the local tilts needs to be used. There are two widespread ways of

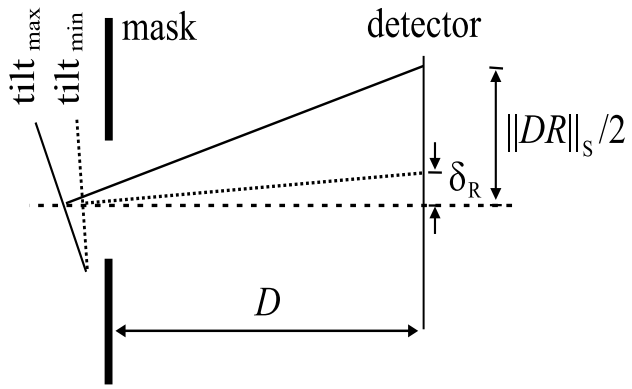

Figure 2.9 *Relation between tilt, spatial dynamic range and position detection resolution for one aperture.*

reconstructing the wavefront from information on the local tilts and they are called *zonal* and *modal* reconstructors. Both methods can use least-squares approximation to find the most appropriate solution, but their inherent characteristics are distinct.

The reconstruction of a wavefront from a set of displacements basically means solving a system of linear equations, which in fact is very conveniently tackled with matrix algebra. The nature of the elements in these equations is determined by the type of reconstructor chosen, and will be approached later. The relevant point so far is that one has to deal with a matrix equation of the form:

$$\boldsymbol{s} = \mathbf{B}\,\boldsymbol{c}\,, \tag{2.15}$$

where $\boldsymbol{s}$ is a vector whose components are the recorded tilts and **B** is called the reconstruction matrix that contains information on how tilts are related either to the local phase or to the spatial modes; **c** is a vector of unknown constants that are either the phase values at local zones or coefficients of spatial modes, depending on the type of reconstructor. The constant character of the **c** elements guarantees the linearity of the system [2.7].

If the dimension of vector **s** is lower than that of vector **c**, there are more unknowns than equations and this under determined system has either no

solution or various solution vectors, which can be found by singular value decomposition (*SVD*) [2.8]. In case **s** is dimensionally equal to **c**, then **B** is a square matrix and can be easily inverted, provided that it is neither singular nor nearly singular[4]. In this case the system admits a unique solution that represents the best fit to ***s*** [2.9, 2.10]. The most often encountered case in wavefront reconstruction belongs to a third kind, where **s** has a higher dimension than **c** and the system has more equations than unknowns. In a strict sense, this system does not have a solution and is said to be overdetermined. In spite of this, a clever method can be used to estimate the best solution vector $\mathbf{c}_f$ that is the closest to satisfying the system of linear equations, as in equation (2.16) [2.11].

$$\boldsymbol{s} \approx \mathbf{B}\,\boldsymbol{c}_f\,. \tag{2.16}$$

The method used to determine $\mathbf{c}_f$ is the least-squares approximation that seeks a numerical solution that minimizes the sum of the squares of the differences between the vector elements on left and right sides of equation (2.16). In other words, it determines which solution vector best fits **s** through **B** [2.12], which is equivalent to solving the vector **c** by the following equation:

$$[\mathbf{B}^T\mathbf{B}]^{-1}\,\mathbf{B}^T\,\boldsymbol{s} = \boldsymbol{c}\,, \tag{2.17}$$

where $\mathbf{B}^T$ is the transpose and $[\mathbf{B}^T\mathbf{B}]^{-1}\mathbf{B}^T$ is the *pseudo-inverse* of the reconstruction matrix **B**. Equation (2.17) is valid on the condition that $\mathbf{B}^T\mathbf{B}$ is invertible. If this condition is satisfied, direct inversion methods can be used as, for instance, the Gaussian elimination method with pivoting strategy. If, on the other hand, $\mathbf{B}^T\mathbf{B}$ is nearly singular or singular, alternative methods as iterative methods or *SVD* should be considered [2.7, 2.8, 2.9].

4. A system of linear equations is nearly singular when some of the equations are close to being linearly dependent. They are vulnerable to becoming linearly dependent, at some step of computations, due to inherent numerical round-off errors.

Zonal reconstructor

In this method the aperture is mapped as a regular grid of phase points, whose values are evaluated from the measured tilts at sampling points that do or do not spatially coincide with the grid points [2.13]. The reconstruction matrix **B**, in this case referred to as *geometric matrix*, is defined by the way tilts at the sampling points are associated with the grid of phase points. To illustrate the zonal concept, the Hudgin configuration [2.14] is shown in Figure 2.10.

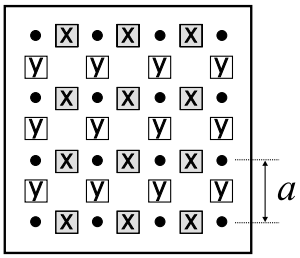

Figure 2.10 *Hudgin zonal estimate configuration.*

The black dots represent the points at which information on the phase φ_{ij} is wanted. In this particular layout only tilt in either the x or y direction is extracted from each spot displacement. Therefore, the centers of the squares (i.e. the centers of the sub-aperture) represent the loci at which either tilts S_{ij}^{x} or S_{ij}^{y} are calculated from the measured displacements.

If NxN phase points are aimed, then $2N(N-1)$ tilts must be registered by the detector. The relationship between tilts and local phase is given in equations (2.18) and (2.19), where a is the phase grid pitch [2.15].

$$S_{ij}^{x} = \frac{(\varphi_{(i+1),j} - \varphi_{i,j})}{a} \qquad \begin{matrix} i = 1, N-1 \\ j = 1, N \end{matrix} \quad , \tag{2.18}$$

$$S_{ij}^{y} = \frac{(\varphi_{i,(j+1)} - \varphi_{i,j})}{a} \qquad \begin{matrix} i = 1, N \\ j = 1, N-1 \end{matrix} \quad . \tag{2.19}$$

From these equations and (2.15), the *2N*(*N-1*) vector ***s*** can be formed as well as the rectangular matrix **B**, with *2N*(*N-1*) rows and N^2 columns. The elements of this matrix are either *1*, *-1* or *0*, depending on how each element of **s** is related to the elements of **c**, which are the phase values to be found at the specified points. Since each tilt is only calculated by a difference between two elements of **c**, matrix **B** is sparse.

Furthermore, $\mathbf{B}^T\mathbf{B}$ is singular (non invertible) because any constant phase added to the elements of **c** will be a solution of this linear system [2.7, 2.15]. For the Hudgin configuration, this can be overcome by adding an extra row to **B** with all elements equal to *1* and an extra element to **s** equal to *0* [2.15, 2.16], which will force the sum of phase values to *0* and yield a non singular $\mathbf{B}^T\mathbf{B}$ matrix, thus enabling the use of (2.17).

Since the zonal method does not require any function to be fit to the data, it is more general and does not depend on the type of optical aperture in question. However, it only provides phase information at specific points, and interpolation to points lying in between can be rather inaccurate.

Also, if enough phase points are considered, it gives a clear picture of the unknown wavefront even if steep local variations are present, but provides no two-dimensional function, in analytic form, to describe the surface.

The only degree of freedom in this reconstructor is the number of phase points to be evaluated, i.e., the number of zones considered, which is limited by the geometry chosen and the number of sub-apertures present.

Modal reconstructor

In contrast to the previous method, the modal reconstructor is based on the assumption that there exists a set of two-dimensional smooth basis functions defined over the aperture of interest, which can describe any incoming smooth-varying wavefront when appropriately combined. This set of functions can be Taylor polynomials [2.3, 2.17], Legendre polynomials [2.15], complex exponentials [2.18], Zernike polynomials, or simply the response functions of a deformable mirror, in the case of a closed-loop system with a wavefront sensor and an adaptive corrector.

The reconstruction matrix **B**, in the modal approach, contains the first

derivatives of the basis functions at the sampling points, i.e., the local tilts associated with each function.

If Zernike polynomials, equation (2.4), are chosen as the basis functions, then the wavefront aberration function in cartesian coordinates *W(x,y)* (representing the deviation from a reference wavefront) can be expressed as in equation (2.20):

$$W(x,y) = \Xi_W - \Xi_{Wref} = \sum_{k=0}^{\infty} C_k \, Z_k(x,y) \,, \tag{2.20}$$

where Ξ_W is the aberrated wavefront and Ξ_{Wref} is the reference wavefront. It is, however, unrealistic to use an infinite series to represent the aberration function and a practical problem suggests a truncated set with a finite number *M* of terms.

From equation (2.7) it is seen that displacements Δx_i and Δy_i, and therefore the local tilts S_i^x and S_i^y, are related to the local derivatives of *W(x,y)* at the center of each aperture labeled *i*, which yields equations (2.21) and (2.22).

$$S_i^x = \left.\frac{dW}{dx}\right|_{(x_i, y_i)} = \sum_{k=0}^{M-1} C_k \left.\frac{dZ_k}{dx}\right|_{(x_i, y_i)} \,, \tag{2.21}$$

$$S_i^y = \left.\frac{dW}{dy}\right|_{(x_i, y_i)} = \sum_{k=0}^{M-1} C_k \left.\frac{dZ_k}{dy}\right|_{(x_i, y_i)} \,. \tag{2.22}$$

For a sensor with *N* sub-apertures, these equations can be written in the matrix form suggested by (2.15) and illustrated in equation (2.23).

(2.23)

$$\begin{bmatrix} S_0^x \\ S_0^y \\ S_1^x \\ S_1^y \\ \vdots \\ S_i^x \\ S_i^y \\ \vdots \\ S_{N-1}^x \\ S_{N-1}^y \end{bmatrix} = \begin{bmatrix} \frac{dZ_0}{dx}\Big|_0 & \frac{dZ_1}{dx}\Big|_0 & \frac{dZ_2}{dx}\Big|_0 & \frac{dZ_3}{dx}\Big|_0 & \cdots & \frac{dZ_{M-2}}{dx}\Big|_0 & \frac{dZ_{M-1}}{dx}\Big|_0 \\ \frac{dZ_0}{dy}\Big|_0 & \frac{dZ_1}{dy}\Big|_0 & \frac{dZ_2}{dy}\Big|_0 & \frac{dZ_3}{dy}\Big|_0 & \cdots & \frac{dZ_{M-2}}{dy}\Big|_0 & \frac{dZ_{M-1}}{dy}\Big|_0 \\ \frac{dZ_0}{dx}\Big|_1 & \frac{dZ_1}{dx}\Big|_1 & \frac{dZ_2}{dx}\Big|_1 & \frac{dZ_3}{dx}\Big|_1 & \cdots & \frac{dZ_{M-2}}{dx}\Big|_1 & \frac{dZ_{M-1}}{dx}\Big|_1 \\ \frac{dZ_0}{dy}\Big|_1 & \frac{dZ_1}{dy}\Big|_1 & \frac{dZ_2}{dy}\Big|_1 & \frac{dZ_3}{dy}\Big|_1 & \cdots & \frac{dZ_{M-2}}{dy}\Big|_1 & \frac{dZ_{M-1}}{dy}\Big|_1 \\ \vdots & & & & \ddots & & \vdots \\ \frac{dZ_0}{dx}\Big|_i & \frac{dZ_1}{dx}\Big|_i & \frac{dZ_2}{dx}\Big|_i & \frac{dZ_3}{dx}\Big|_i & & \frac{dZ_{M-2}}{dx}\Big|_i & \frac{dZ_{M-1}}{dx}\Big|_i \\ \frac{dZ_0}{dy}\Big|_i & \frac{dZ_1}{dy}\Big|_i & \frac{dZ_2}{dy}\Big|_i & \frac{dZ_3}{dy}\Big|_i & & \frac{dZ_{M-2}}{dy}\Big|_i & \frac{dZ_{M-1}}{dy}\Big|_i \\ \vdots & & & & \ddots & & \vdots \\ \frac{dZ_0}{dx}\Big|_{N-1} & \frac{dZ_1}{dx}\Big|_{N-1} & \frac{dZ_2}{dx}\Big|_{N-1} & \frac{dZ_3}{dx}\Big|_{N-1} & & \frac{dZ_{M-2}}{dx}\Big|_{N-1} & \frac{dZ_{M-1}}{dx}\Big|_{N-1} \\ \frac{dZ_0}{dy}\Big|_{N-1} & \frac{dZ_1}{dy}\Big|_{N-1} & \frac{dZ_2}{dy}\Big|_{N-1} & \frac{dZ_3}{dy}\Big|_{N-1} & & \frac{dZ_{M-2}}{dy}\Big|_{N-1} & \frac{dZ_{M-1}}{dy}\Big|_{N-1} \end{bmatrix} \begin{bmatrix} C_0 \\ C_1 \\ \vdots \\ C_k \\ \vdots \\ C_{M-2} \\ C_{M-1} \end{bmatrix}$$

s (2N x 1) = **B** (2N x M) **C** (M x 1)

Vector **s** has *2N* elements, which are *x* and *y* slope pairs at each aperture *i*; vector **c** of modal coefficients has dimension *M*; reconstruction matrix **B** is *2N* x *M* large and its row-vector components $\mathbf{b}_i$ are of the form given in the following equation:

$$\boldsymbol{b}_i = \left[\frac{dZ_0}{d\xi}\Big|_i, \frac{dZ_1}{d\xi}\Big|_i, \ \ldots\ , \frac{dZ_k}{d\xi}\Big|_i, \ \ldots\ , \frac{dZ_{M-1}}{d\xi}\Big|_i \right], \quad for \quad \xi = x \ or \ y \ . \quad (2.24)$$

They must be linearly independent to guarantee the inversion of $\mathbf{B}^T\mathbf{B}$ [2.19, 2.20]. This condition is satisfied since the vector elements are derivatives

of Zernike polynomials that are linearly independent themselves. The feasibility of the inversion operation relies on the choice of appropriate functions.

The inversion of the MxM matrix $\mathbf{B}^T\mathbf{B}$ can be performed, for instance, by Gaussian elimination [2.12] with the least-squares method, equation (2.17), if the number of registered slopes $2N$ is larger than the number M of reconstruction modes. When the modal coefficients are known, the calculation of $W(x,y)$ from equation (2.20) is straightforward. Notice, however, that what is fitted by solving the matrix equation is not the wavefront itself but the local tilts associated with the spot displacements. In fact, the derivatives of the Zernike terms weighted by the solution coefficients C_k represent the best fit to the tilts associated with the measured spot displacements.

The pseudo-inverse matrix $[\mathbf{B}^T\mathbf{B}]^{-1}\mathbf{B}^T$ only needs to be calculated once for a given configuration, i.e., for a given number of tilt measurements, modes and geometrical distribution of apertures. Thereafter, the computational effort is reduced to measuring displacements, calculating the associated tilts and evaluating the best-solution coefficients for arbitrary incoming wavefronts. The number of modes necessary to describe a wavefront is usually smaller than the number of phase points in the zonal reconstructor. Consequently, the vector $\mathbf{c}$ for the modal method has fewer elements than vector $\mathbf{c}$ for the zonal one, which implies that the modal reconstructor renders faster computations.

The reconstructed wavefront is usually not an exact reproduction of the test wavefront. The causes of the reconstruction errors are manifold: spot position detection noise, an inappropriate set of functions, inadequate number of modes, error propagation through matrix algebra, density of sampling points and spatial distribution of apertures and their sizes.

For a given detector and a wavefront with a certain spatial frequency distribution over a known aperture, a compatible set of functions should be chosen [2.17]. Of the candidate functions, a set of orthogonal functions over each particular aperture is less error prone [2.12]. Also, the less modes M used to correctly describe the wavefront, the lower the reconstruction error. For example, if only an incoming tilted flat wavefront is expected, then the use of only Zernike terms Z_0 and Z_1 is optimal. This is because the root-mean-square phase error is proportional to the trace of matrix $[\mathbf{B}^T\mathbf{B}]^{-1}$ (sum of its diagonal elements) [2.15].

It has been shown by Wang and Da Silva [2.12] that both least-squares matrix inversion and Gram-Schmidt orthogonalization [2.11] are well conditioned (yield numerical stability) as long as enough sampling points are available and distributed conveniently over the aperture. A larger sampling density allows the use of more reconstruction modes, but as more modes are introduced the variances of the individual modes increase. Hence, a convenient trade-off should be found for each particular system.

It has been shown that a regular orthogonal grid of sampling points contribute to a higher numerical stability than a radial grid[5], for instance [2.12]. If the density of points is made larger as the distance from the center increases, the variance for the higher-order Zernike modes is minimized [2.21].

Modal reconstructors can be confronted with either *undermodeling* or *undersampling*. Where the first is not detrimental, the latter can lead to undesired results. Undermodeling occurs when enough sampling points are available for M modes but only a few modes $m < M$ are sought. Either one can handle only the m projections out of the full set of M modes, or a reconstruction matrix containing only the chosen modes can be used. Although higher-order modes will appear as part of the m modes, the second alternative results in a lower residual error, mainly because the modes are the best-fit solutions rather than a projection. Undersampling is simply the opposite: more modes are calculated than the number of sampling points affords, resulting in aberration aliasing [2.22]. In a simplistic analogy, it is like trying to fit a line to one data point, or a parabola to two data points.

An appropriately chosen modal reconstructor is less prone to being affected by noise in the tilt measurement than a zonal reconstructor, and it is superior to the latter especially when only a fixed number of modes are necessary. It is also faster and more flexible since more degrees of freedom are involved: set of functions, number of modes, number of points and sub-aperture geometry. Moreover, it favors data interpolation because it readily provides a functional form to the wavefront instead of the discrete phase data supplied by the zonal reconstructor.

5. grid formed by equally spaced sub-apertures placed along lines radially connecting the center of the mask to its outer edge.

2.4 Photodetection

With the present technology any attempt in measuring phase in the optical spectrum relies fundamentally on measuring the light-intensity distribution (see Chapter 1). In the Hartmann method, phase is obtained from the displacement of light spots, which in fact form a two-dimensional light-intensity-distribution pattern. Therefore, by detecting photons over a certain area we can obtain a phase spatial profile (wavefront). This section examines the mechanism of the photoelectric effect and some photodetector structures in silicon. Furthermore, we examine noise sources inherent to photodetection.

2.4.1 Photodetection mechanism

In the case of the internal photoelectric effect, where photons with energy $E_{ph}=h\nu=hc/\lambda$ impinge on a semiconductor, there is a chance that they will be absorbed depending on the material properties and its thickness. Absorbed photons can generate electron-hole pairs, which can either immediately recombine, recombine after some errant course or be definitely separated by the presence of an electric field, which is the most convenient case because then carriers can be more effectively collected.

The adequacy of a material for photodetection can be understood by considering its valence and conduction bands, which are usually separated by an energy gap [2.23]. Thermal processes or the absorption of photons can excite valence electrons to make a transition to the conduction band, leaving empty quantum states, holes, in the valence band, therefore constituting electron-hole pairs. In electrical conductors, this energy gap is either inexistent or very small (<0.1eV). This means that the number of carriers in the conduction band is already large and photon absorption results in thermalization[6] and subsequent recombination. Another circumstance holds for insulators, where the bandgap energy is large (>3eV). Then, the energy of a photon is not enough to cause the transition of an electron to the conduction band, in which case no photodetection happens. Semiconductors are, on the other hand, very appropriate materials for photodetection because they usually feature an energy gap that can be

6. An (photo-)excited electron in the conduction band decays to the bottom of the band by releasing energy in the form of lattice vibrations.

surpassed by electrons with sufficient energy given by the absorbed photons, resulting in a number of electrons contributing to electrical conduction. For this purpose, the *minimum* photon energy E_{ph} should be equal or larger than the bandgap energy E_g, depending on whether the gap is direct or indirect, respectively. Silicon (Si), for instance, has an indirect band gap, so a photon ($E_{ph} > E_g$) first induces a transition of the electron from the valence band to a higher energy state in the conduction band, which subsequently achieves the minimum-energy condition via thermalization (emission of a phonon).

As stated before, the electron and hole of a photo-generated pair are spatially driven away from each other by the presence of an electric field across a depletion region. In semiconductors, the depletion layer can result from a metallurgical junction, the so called *p-n junction*, from an external voltage, or from both, as depicted in Figure 2.11.

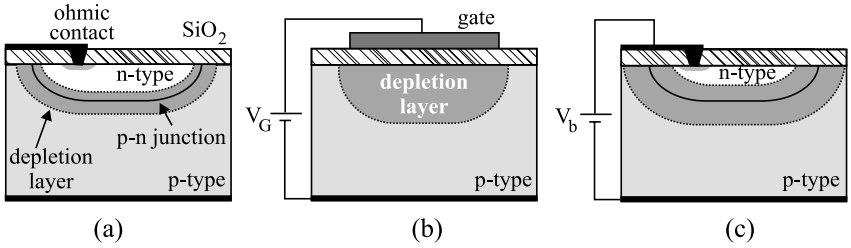

Figure 2.11 *The depletion layer in (a) floating p-n junction, (b) mono-doped semiconductor with external voltage and (c) reverse biased p-n junction.*

In the case of a *p-n junction*, even in the absence of an external potential, a depletion region exists around the interface of the *p-doped* and *n-doped* regions; it extends to each side by a distance inversely proportional to the doping concentration of each region. As a consequence an electric field appears across the junction. Electrons generated within the depletion region can drift to the *n* region and be collected at the ohmic contact, where they are read out. Also, photo-generated electrons in the *p* region in the neighborhood of the depletion region might reach it by random diffusion and end up being collected (Figure 2.12). The diffusion length specifies how far on average a carrier can go without recombining; this length

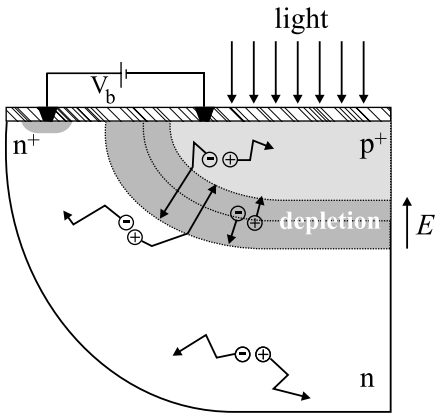

Figure 2.12 *Electron-hole pairs generated in different regions of a p-n structure.*

decreases as the doping concentration increases. The junction can also operate in reverse bias when a positive voltage V_b is applied to the *n* region (taken that the *p* region is connected to a voltage lower than V_b, usually ground) (Figure 2.11c); the depletion region is enlarged in both directions contributing to the collection of more carriers. The width W_d of the depletion layer for an abrupt junction is given by equation (2.25) [2.24]:

$$W_d = \sqrt{\frac{2\varepsilon}{q}\,\frac{N_A + N_D}{N_A N_D}\,(\phi_{bi} + V_b)}\ , \qquad (2.25)$$

where ε is the dielectric constant of the semiconductor[7], q is the elementary charge, N_A and N_D are concentrations of impurity atoms in the *p*- and *n-type* regions, respectively, and ϕ_{bi} is the built-in junction potential that can be calculated by:

7. ε is the product of the adimensional relative dielectric constant of the material and $\varepsilon_0 = 8.85 \times 10^{-12}$ As/Vm (or as often encountered $\varepsilon_0 = 8.85 \times 10^{-14}$ F/cm).

$$\phi_{bi} = \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) , \tag{2.26}$$

where n_i is the intrinsic carrier density of the semiconductor[8] and k is the Boltzmann constant. The thickness of the depletion region in each side of the junction, x_n and x_p, can be obtained by the following relations:

$$W_d = x_n + x_p \tag{2.27}$$

and

$$x_n N_D = x_p N_A . \tag{2.28}$$

When an external voltage V_G is applied to a gate separated from a grounded *mono-doped* semiconductor substrate (Figure 2.11b) by a thin oxide layer, a depletion region of depth x_d (see equation (2.29) for a *p-type* substrate) is created in the bulk as a result of the charges present at the gate contact:

$$x_d = \sqrt{\frac{2\varepsilon}{q N_A} \phi_S} , \tag{2.29}$$

where ϕ_S is the potential at the semiconductor surface under the gate oxide, that represents the voltage across the depleted region. In a ideal case, the relation between this potential and the applied voltage V_G can be approximated by the following relation [2.25, 2.26]:

$$V_G = \frac{\sqrt{2\varepsilon q N_A \phi_S}}{C_{ox}} + \phi_S , \tag{2.30}$$

where C_{ox} is the capacitance of the oxide layer. This depleted region can be thought of as a potential well, where photo-generated electrons will be stored and subsequently transferred to an adjoining well or to a read-out node. This is basically the working mechanism of a charge-coupled device (*CCD*).

8. In silicon $n_i = 9.65 \times 10^9$ cm^{-3}, at room temperature (300K).

2.4.2 Silicon photodetector efficiency

The overall quantum efficiency η of a photodetector is defined as the probability that photons generate electron-hole pairs that are collected before recombination. It is wavelength dependent, through the absorption coefficient α of the material, number of doped regions, their location in the bulk Si, doping levels, voltage bias and diffusion length of the carriers L_d (L_p or L_n, for holes or electrons, respectively). η also depends on the optical power reflectance $\Re$ of the Si surface plus possible dielectric layers on top of it. A theoretical expression for the quantum efficiency is given by [2.24]:

$$\eta = (1-\Re)\left(1 - \frac{e^{-\alpha_{Si} W_d}}{\alpha_{Si} L_d + 1}\right). \qquad (2.31)$$

The responsivity R_λ of a detector relates the photo-generated current I_p to the incident optical power P_{opt}, resulting in a number to the inherent sensitivity of the detector - equation (2.32)- provided that the detector is not saturated (see equation (2.45)). The spectral response is simply a measure of the responsivity as a function of wavelength.

$$R_\lambda = \frac{I_p}{P_{opt}} = \frac{\eta e}{h\nu} = \frac{\eta\,\lambda[\mu m]}{1.24} \qquad [A/W] \quad . \qquad (2.32)$$

Figure 2.13 shows the optical power transmittance from air to Si as a function of the wavelength for various thicknesses of a covering silicon dioxide layer, when normal incidence is considered [2.27]. It is observed that the transmittance is considerably affected by reasonably small variations of the oxide thickness d_{ox}, due to interference. For a particular wavelength, one can maximize transmittance by using an anti-reflective coating.

Once a number of photons N_0 traverses the surface of silicon, their absorption is governed by its absorption coefficient α_{Si}, which varies with the wavelength λ and represents the inverse of the distance over which the number of transmitted photons is reduced by a factor e, i.e., about 63% of the photons are absorbed before reaching the depth $1/\alpha_{Si}$ (penetration depth). If the photon energy is larger than the band gap, across-the-gap transitions occur, upon photo-absorption, resulting in electron-hole pairs.

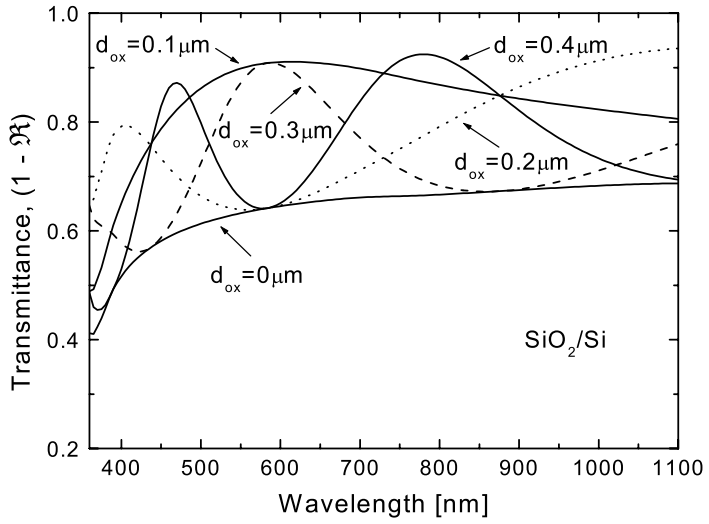

Figure 2.13 *Optical power transmittance from air to the surface of Si for different thicknesses d_{ox} of the oxide layer.*

Then the number of photons, *N*, transmitted over a depth *z* is given by $N = N_0 \exp(-\alpha_{Si} z)$. The graph in Figure 2.14 shows the penetration depth for different wavelengths in Si. Longer-wavelength photons are absorbed deeper in the material, provided that their energy $E_{ph}=hc/\lambda$ is larger than E_g (1.12 eV for Si). Photons with very short wavelengths are absorbed so close to the surface that most photo-generated carriers diffuse to the surface and recombine there.

2.4.3 Silicon photodetector structures

A number of photodetector structures can be implemented in silicon, mostly dependent on the junctions and doped regions. Each type of structure has a particular footprint that results in a different detector performance.

A slice or thin film of intrinsic semiconductor with two contacts held at different voltages has an electric field set up across the material and a current that flows as a result of impinging photons. This can be viewed as a change in the conductivity of the semiconductor and is therefore called

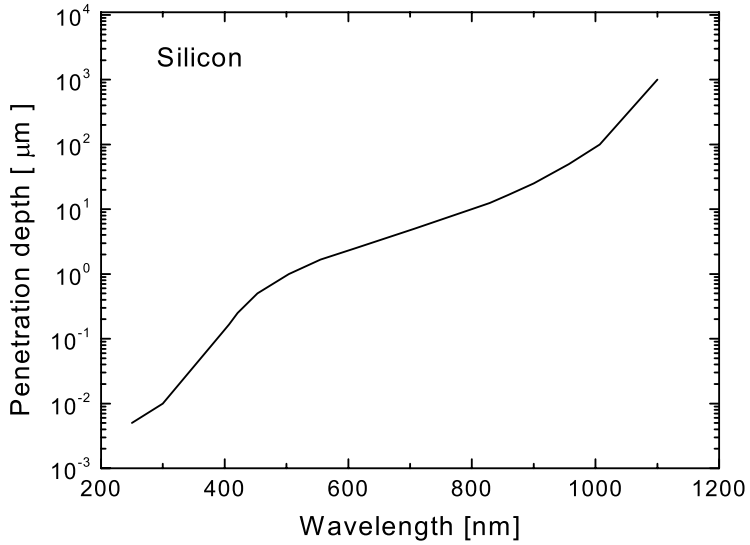

Figure 2.14 *Penetration depth (1/α) in silicon.*

photoconductivity. Photoconductors are per definition simple to fabricate but yield more noise and higher response times (1 to 10ms) than other detectors. In spite of this, they are to date still largely employed in specific applications, especially where no integrated arrays are demanded.

An optimal photodetector structure for a particular application might require a careful recipe and special processing of the Si wafer with several non-standard steps, which usually involves high costs and the need for easily adaptable and suitably equipped semiconductor facilities. Consequently, most Si photodetectors comply with the possibilities of standard foundries, at most introducing minor process modifications. Some of the most common structures are presented in the simplified schematics of Figure 2.15.

Schottky photodiode

The Schottky photodiode is formed by the interface of the doped semiconductor with a thin metal layer (10 - 20 nm) (Figure 2.15a). It is a unipolar device in the sense that the electrons and holes do not both contribute to the current, but only one type of them, depending on the

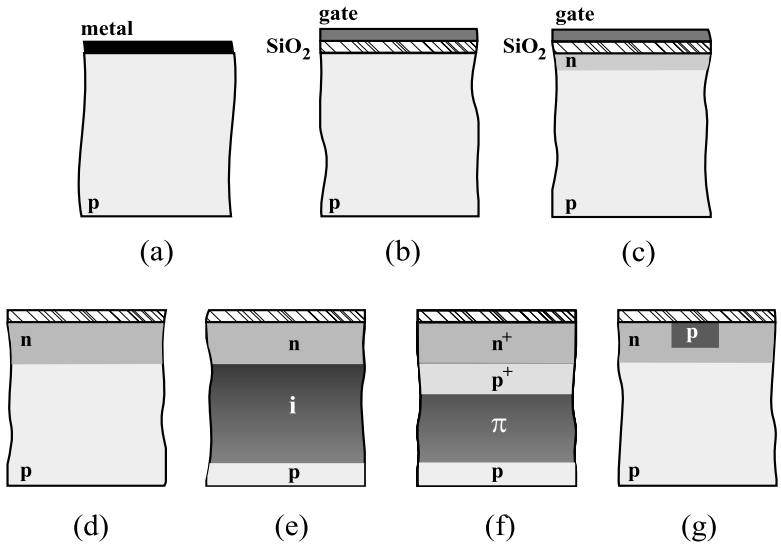

Figure 2.15 *Schematics of the vertical layer structure of several photodetectors: a) Schottky (metal-semiconductor) photodiode, b) SCCD cell, c) BCCD cell, d) p-n junction photodiode, e) p-i-n photodiode, f) avalanche photodiode and g) vertical phototransistor.*

dopant material. It is very appropriate for short wavelengths when front illuminated (light impinging on the metal side), because the depletion layer is formed just under the *metal/Si* interface.

The Schottky photodiode is a fast device that can attain internal response times in the order of picoseconds, leading to bandwidths of the order of tens of GHz. The quantum efficiency is usually lower than that for other structures and the device usually needs to be cooled to suppress the large dark current [2.28, 2.29]. Nevertheless, it can simply be built as a PtSi (Platinum silicide) film deposited on top of either *p-* or *n-type* Si.

CCD photocell

The surface CCD cell (SCCD) is basically a MOS capacitor and consists of

doped *p-type* Si, for instance, with oxide on top and either a metal or poly-Si gate, where a voltage is applied (Figure 2.15b). As explained at the end of section 2.4.1, when the pixel is subjected to a gate voltage, a potential well is formed under it, where photo-generated electrons are stored and later transferred or collected. Nowadays, however, an enhanced version called buried-channel CCD (BCCD) is used in most image-sensing applications. For a *p-type* substrate it has a shallow *n-buried* channel just under the gate oxide to avoid the influence of surface states on the charge transfer (Figure 2.15c). It allows higher operation frequencies and features lower noise than the SCCD, but it supports lower full-well capacities [2.5] and requires a fabrication process that is a bit more complex.

Junction photodiode

The most common structure is the junction photodiode, which is simply a *p-n* junction located at some distance from the surface (Figure 2.15d). If it is illuminated from the front side, the deeper the junction, the higher the efficiency in collecting longer wavelengths (provided that $E_{ph} > E_g$). It can be operated in zero bias or reverse bias. In the latter case a thicker depletion layer is formed, a higher quantum efficiency is achieved and the junction capacitance is reduced.

The intrinsic response time of a junction photodiode is limited by the diffusion time; the time it takes for electrons generated outside the depletion layer to reach it.[9] For very high speed applications, depending on the bias and the depletion width, the drift time should also be taken into account. The extrinsic response time is determined by the junction capacitance (proportional to the ratio of the junction area to the width of the depletion layer) and by the series resistance between the edge of the depletion layer and the ohmic contact. The overall bandwidth is determined by the external circuit as well. There is always a compromise between collection efficiency and response time.

Connecting several alternately doped junctions together poses an architectural alternative to 'reverse bias', thus improving η. Interestingly enough, not connecting them together and providing each layer with an external contact yields a rather useful application, namely an intrinsic color

9. carriers generated inside the depletion region move by drift, and their speed is much higher than that of those moving by diffusion.

sensor (i.e. with no color filters) [2.30, 2.31, 2.32].

The popularity of the Si junction photodiode can be mainly attributed to its relative low cost, wavelength working range (~ 350nm - 1100nm) and the availability of *p-n* junctions in standard Si processes such as Bipolar, CMOS and BiCMOS.

p-i-n photodiode

A direct extension of the junction photodiode is the *p-i-n* structure, with a thick intrinsic layer - lightly doped in fact - inserted between the *p* and *n* regions (Figure 2.15e). Once the thickness of the depletion layer in each region is inversely proportional to its doping concentration, the lightly doped *i* region can be designed such that it is fully depleted. Therefore it features a lower junction capacitance than an ordinary *p-n* junction and consequently faster extrinsic response times. Although an increase in the carrier drift transit time is observed, this is largely compensated by the reduction of the ratio between diffusion length and drift length as compared to a junction photodiode [2.33]. Besides, a larger quantum efficiency is achieved due to the much broader collecting region. The *p-i-n* photodiode can be operated under no bias or under moderate reverse voltages.

Avalanche photodiode

In its simpler form, the avalanche photodiode (*APD*) can be a mere junction device. More often it is a modified *p-i-n* structure called a *reach through* with a *np–π-p*, the π layer being a lightly doped sandwiched *p-type* region (Figure 2.15f). *APD*s require a strong highly stabilized reverse bias applied to it [2.34]. Electrons and/or holes in the depletion region are accelerated by the strong electric field. A number of them collide with the crystal lattice on the acceleration path, therefore losing energy; those which can accelerate enough to gain energies above the band gap manage to generate secondary electron-hole pairs by impact ionization, which generate further electron-hole pairs and the process repeats itself in an avalanche fashion, providing the detector with an internal gain that in Si can amount to 200 [2.35].

The initialization of avalanche multiplication occurs at a specific breakdown voltage that depends on the doping concentrations and the geometry of the device [2.34]. Noise is often higher than in the gainless

photodetectors. To keep the noise level down, preferably only one type of carrier (electron or hole) should take part in the multiplication process, minimizing unnecessary collisions. This situation is advantageous in Si, since electrons have a much higher ionization probability than holes, at room temperature. Integration of electronic circuitry along with an *APD* is not straightforward due the high reverse voltage requirement. A careful design is necessary to prevent premature breakdown at *APD* junction edges.

Phototransistor

The vertical phototransistor is a variation of the simple junction structure and can be interpreted as being a junction photodiode directly feeding a transistor, as in Figure 2.16. If the base is floating, it is a transistor whose

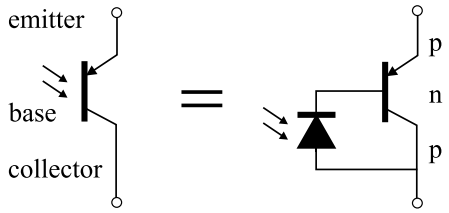

Figure 2.16 *A bipolar phototransistor represented as a photodiode feeding the base current of an ordinary transistor.*

base current is the photocurrent from the base-collector junction photodiode, and any current other than this acts as leakage current. Therefore, the emitter area is usually small compared to the base-collector area to minimize the photocurrent contribution from the upper junction [2.36, 2.37] (Figure 2.15g). The phototransistor can be biased by using a constant background illumination. The base current is amplified with a gain that is a function of the light intensity and varies uniformly with the bias voltage [2.38]. This undoubtedly compromises the linearity of the device, which can only be partly compensated by external circuitry at the expense of noise [2.39]. Dark current is also amplified by the same gain.

Moreover, the larger the gain, the slower the phototransistor response due to the Miller effect[10]. Lateral phototransistor structures have also been reported [2.40], with the disadvantage of larger dark current due to the lateral current flow close to the silicon-to-oxide interface.

2.4.4 Noise and equivalent circuit

Noise determines the minimum amount of light that can be detected. It is an inherent part of photodetection, and can be introduced by the light source, the photodetection mechanism or the technological process and layout employed. Also, external circuitry connected to the photodetector might contribute to it. Noise in semiconductor devices represents a whole field of research by itself and a thorough discussion of its mechanisms is out of the scope of this thesis. Although not exhaustively, the following paragraphs address the most important noise contributions in junction photodetectors.

The rate of photons impinging on a photodetector is not constant and varies according to the Poisson distribution around a mean value of photons. This causes fluctuations in the electron-hole pair generation, translated as noise in the collected photocurrent. Besides, not all photons generate electron-hole pairs, whose total number is not only due to photon absorption but also to thermal generation and band transitions induced by lattice defects and traps, generally more numerous close to the surface. This means that even when the photodetector does not receive photons, a noisy current is present within the device. Additionally, the finite conductance of the depletion layer and its series resistance to the ohmic contact contribute to thermal noise. To sum it up, these effects can be categorized into a number of noise components present in the photocurrent that is read out:

- *photon shot noise* - it results from the degree of randomness of photon arrival at the detector. In most practical cases, from the detector perspective photon noise can be translated as *photoelectron noise*, which

10. The base-collector junction capacitance C_j is subjected to voltage variations on both sides because of the transistor amplification mechanism. If the base voltage changes by ΔV_{base}, then the collector voltage changes by $\beta\Delta V_{base}$, which means that the voltage across the capacitor will change by $(1+\beta)\Delta V_{base}$. This can be viewed as a capacitance that is $(1+\beta)$ times larger, i.e. $C_j' = (1+\beta)\, C_j$.

is the fluctuating number of photo-generated electron-hole pairs resulting from the fluctuating number of arriving photons. It can be estimated as being proportional to the square root of the photocurrent;

- *1/f noise* - also called flicker noise. Several different theories, some with conflicting conclusions, attempt to explain its origins. It is suspected that it derives from either carrier fluctuations (due to $Si\text{-}SiO_2$ interface traps), or mobility fluctuations, or even both [2.41]. The fact that it is nearly proportional to the inverse of the frequency is in consensus. It is also known that structures with a larger area are less prone to its effects because then the spatial fluctuations are averaged out; it is more significant in lateral shallow devices (e.g. MOS transistors). Because there is no general agreement on the causes of this noise, only speculative expressions exist, which renders it suspicious for modelling;
- *dark-current shot noise*- dark current is the current read-out when the device is not exposed to light. It is highly sensitive to defects (electron trapping and detrapping), depends on the technology used, changes considerably with temperature and also has an indirect dependence on the reverse bias [2.42], through the depletion layer width. It represents an offset to the signal due to reverse leakage across the junction plus a generation-recombination term, which in Si is dominating. Dark-current shot noise is associated with fluctuations of the dark current value;
- *thermal noise* - present in any resistive material and also called Johnson noise; it increases directly with temperature. It arises from the random motion of carriers in resistive electrical materials.

For modelling a junction photodiode, an acceptable equivalent circuit for most practical cases is that of Figure 2.17, where uncorrelated noise sources can also be added quadratically as a single source that yields the total noise contribution. The photo-generated current is represented by a current source and modeled as:

$$I_p = e\eta\Phi \,, \tag{2.33}$$

where e is the electron charge, η is the quantum efficiency ($0 \le \eta \le 1$) and Φ is the photon flux (photons/s) impinging on the active detector area A and given by:

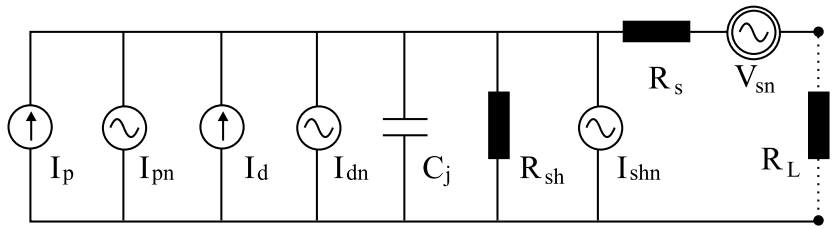

Figure 2.17 *Equivalent circuit of a junction photodiode.*

$$\Phi = \int_A \phi \, dA_j \,, \tag{2.34}$$

where ϕ is the photon flux density.

Dark current I_d is given as a contribution of the reverse leakage current I_l and the generation-recombination current in the depleted region I_{gr}, as in equations (2.35), (2.36) and (2.37).

$$I_d = I_l + I_{gr} \,, \tag{2.35}$$

$$I_l = A_j \, e \, n_i^2 \left[\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right] \,, \tag{2.36}$$

$$I_{gr} = A_j \, e \, n_i \left(\frac{W_d}{2\tau_g} \right) \,, \tag{2.37}$$

where A_j is the junction area, $D_{n,p}$ is the minority-carrier diffusion constant, $L_{n,p}$ is the minority-carrier diffusion length, τ_g is the generation lifetime [2.43] and W_d is the depletion layer width.

The term I_{pn} represents the photoelectron noise and I_{dn} the noise associated with the dark current. These are shown in equations (2.38) and (2.39), where B is the total bandwidth of the sensor/read-out circuit ensemble.

$$I_{pn} = \sqrt{2\,e\,I_p\,B} \quad , \tag{2.38}$$

$$I_{dn} = \sqrt{2\,e\,I_d\,B} \quad . \tag{2.39}$$

The capacitance C_j is the junction capacitance given by $C_j = \varepsilon_{Si}(A_j / W_d)$. The parallel resistance R_{sh} represents the finite conductance of the depletion layer and is usually fairly high (10MΩ to 10GΩ). In practice it is measured as the slope (*dV/dI*) of the photodiode I-V curve at V=0. The series resistance R_s is the resistance of the undepleted region between the edge of the depletion layer and the metal contact; it usually has a value ranging from a few Ohms (Ω) to hundreds of Ohms. Both these resistances have thermal noise associated with them, the former in the form of a noise current source I_{shn} in parallel with R_{sh} and the latter in the form of a noise voltage source V_{sn} in series with R_s, which are given by the following equations:

$$I_{shn} = \sqrt{\frac{4kTB}{R_{sh}}} \quad , \tag{2.40}$$

$$V_{sn} = \sqrt{4kTBR_s} \quad . \tag{2.41}$$

When a load resistor R_L is connected to the output terminals of the photodiode, thermal noise is added to the system and the bandwidth B can be calculated according to three approximating cases [2.44], where C_p is the parasitic capacitance of the output terminal:

- $R_L >> R_s$, $R_{eq} \sim R_L \,||\, R_{sh}$ and $B = (1/2\pi)[(R_{sh} \,\|\, R_L)(C_j + C_p)]^{-1}$;
- $R_L < R_s$, $R_{eq} \sim R_s \,||\, R_{sh} \sim R_s$ and $B = (1/2\pi)[R_s\, C_j]^{-1}$;
- $R_L \sim R_s$, $B = (1/2\pi)[(R_s R_L)(C_j C_p)]^{-1/2}$.

An important figure of merit in photodetection is the signal-to-noise ratio *SNR*, which is given by the ratio of the total measured current I_{ph}, with the dark current already subtracted from it, to the quadratic sum of the

uncorrelated noise contributions. In practice, this is given by the mean measured current (or voltage) squared divided by its variance[11] σ_I^2, as in equation (2.42).

$$SNR = \frac{(I_{mean})^2}{\sigma_I^2} \tag{2.42}$$

The threshold input power at which noise dominates the signal is termed the noise equivalent power *NEP* and corresponds to *SNR*=1. It can be calculated as $NEP=\sigma_I / \mathsf{R}_\lambda$, where R_λ is the responsivity. It is usually further normalized to the square root of the product of the photodiode bandwidth *B* and area A_j, due to the quadratic dependence of noise on these parameters. The detectivity *D** is the reciprocal of the normalized *NEP*; the larger its value, the higher the detection performance. It is useful for comparing devices of different sizes.

The photodiode optical dynamic range $\|DR\|_o$ indicates the span of light levels under which it operates, ranging from *NEP* to saturation optical power P_{sat}, and is given by equation (2.43).

$$\|DR\|_o = 10 \log\left(\frac{P_{sat}}{NEP}\right) \quad [dB] \quad . \tag{2.43}$$

$$P_{sat} = \Phi_{sat}\, h\nu = \frac{I_{psat}}{e\eta}\, h\nu \quad . \tag{2.44}$$

The photo-generated saturation current I_{psat} for a p^+n photodiode is:

11. In a system with random noise described by Gaussian distribution, the variance σ^2 over signal *S* is given by: $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(S_i - \overline{S})^2$, where N is the number of measurements, $\overline{S}$ is the mean signal value and σ is the standard deviation.

$$I_{psat} = \frac{A_j\, e\, N_A}{2\left(\frac{1}{\mu_n} + \frac{1}{\mu_p}\right)} \frac{V}{W_d} \quad , \tag{2.45}$$

where N_A is the acceptor atom concentration, $\mu_{n,p}$ is the carrier mobility and (V/W_d) is the junction field.

2.5 Optical position-sensitive detectors (PSDs)

With the Hartmann method wavefronts can be reconstructed if the displacements of light spots are known. Traditionally conventional imagers (mostly charge-coupled devices - *CCD*s) are used to obtain these displacements. It is a two-step process in which the *CCD* first records the image of all spots, whose displacements from a reference grid are then computed by means of a data-reduction algorithm. The whole transfer and acquisition time plus the lengthy image processing time renders such method inappropriate for real-time analysis. A way around this inconvenience is to have *one optical position-sensitive detector PSD per light spot*, whose output is directly proportional to the spot centroid position on its surface. In this section a number of position-sensing geometries are introduced.

Position-sensitive devices can be divided into two main categories: *lateral-photoeffect PSD*s and *multipixel PSD*s. They differ in operation principle and layout. The former is usually based on a single large contiguous photosensing area (*LEP* - lateral-effect photodiode) whereas the latter relies, as the name suggests, on an cluster or array of several photosensing elements.

2.5.1 Lateral-effect PSD

The lateral-effect photodiode *LEP* is simply a large-area *p-n junction* with either reverse or zero voltage applied across the junction [2.45, 2.46]. Also two contacts are present at opposite sides of one sensing layer (e.g. the *n-layer*). When a light spot much smaller than the detector area impinges on its surface, a potential difference will exist between the illuminated position at the *n-layer* and any other point at a certain lateral distance from this position. This difference is termed the lateral photovoltage and causes

electrons to drift from the illuminated area towards the oppositely located electrodes. The current difference between these electrodes provides one-dimensional spot position information.

In many cases two-dimensional information is required and therefore four contacts can be placed at each edge line of the *LEP*. The potential difference increases radially from the illuminated focus and the current will be divided between four electrodes providing either *x* or *y* information per opposite electrode pair. In this configuration the coordinates are interdependent, resulting in considerable position non-linearity. This can be circumvented by placing electrode pairs on two different layers.

2.5.2 Multipixel PSDs

There are several possible spatial arrangements of pixels to sense the position of a light spot. They can differ in the number of pixels necessary and the architecture. The implementation of the layouts usually gives one the flexibility to choose from different kinds of photosensitive elements within the requirements of a given technology. For one-dimensional sensing two simple structures illustrate the concept of multipixel *PSD*s: the *bi-cell* and the *stripe PSD*.

The *bi cell* is formed by, two photosensitive elements, with dimensions *a/2* x *a*, placed side by side (Figure 2.18a). The detector response x_{BC} to a spot-centroid position *x* is expressed by equation (2.46)[2.47].

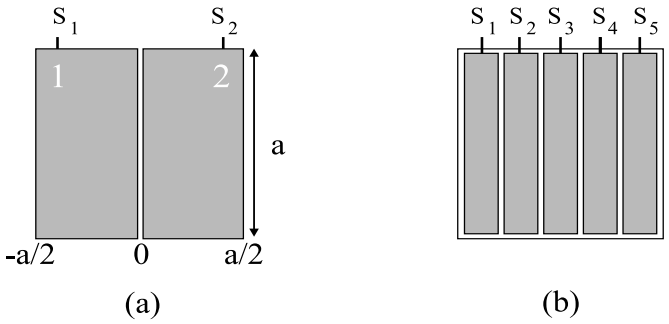

Figure 2.18 *One-dimensional PSDs: a) bi cell and b) stripe.*

$$x_{BC} = \frac{\int_0^a dy \int_0^{a/2} H(x,y)dx - \int_0^a dy \int_{-a/2}^0 H(x,y)dx}{\int_0^a dy \int_{-a/2}^{a/2} H(x,y)dx}, \qquad (2.46)$$

which can be simplified using the notation of equation (2.47):

$$x_{BC} = \frac{S_2(x) - S_1(x)}{S_2(x) + S_1(x)} \quad , \qquad (2.47)$$

where x_{BC} is called the '*pseudo-position function*' given by the *bi-cell* signals S_1(x) and S_2(x), which are functions of the spot-centroid position *x*. It is normalized so that the maximum response range is always [-1,1]. The linearity of the pseudo-position function x_{BC} depends on the cell size, the gap between the cells and on the intensity distribution, given by *H(x,y)*. In a *bi cell* the position information of a spot centroid is only achieved while the spot illuminates both cells.

The *stripe PSD* is an extension from a simple *bi cell* to a larger number of photosensing elements placed side by side (Figure 2.18b). Each element has a signal S_i and the pseudo-position function x_{stripe} is given by a combination of all signals. The form of this function depends on the centroiding method used (Section 2.5.4), ratio of the spot size to the stripe width, spot intensity profile, number of stripes and gap between cells. For the same lateral dimension of the PSD, smaller spots can be used with the *stripe* layout than with the *bi-cell* one, yielding a larger displacement range.

The two-dimensional multipixel *PSD*s are modified versions of the one-dimensional structures but are based on the same principles. A straightforward modification of the *bi cell* is the *quad cell*, in which four photodiodes are positioned as in Figure 2.19a. This is quite the same as two *bi cells* perpendicular to each other. The normalized pseudo-positions x_{BC} and y_{BC} are analogous to the two-cell counterpart and given by:

$$x_{QC} = \frac{(S_B + S_C) - (S_A + S_D)}{\sum_i S_i}, \text{ and} \qquad (2.48)$$

$$y_{QC} = \frac{(S_A + S_B) - (S_C + S_D)}{\sum_i S_i} \quad . \qquad (2.49)$$

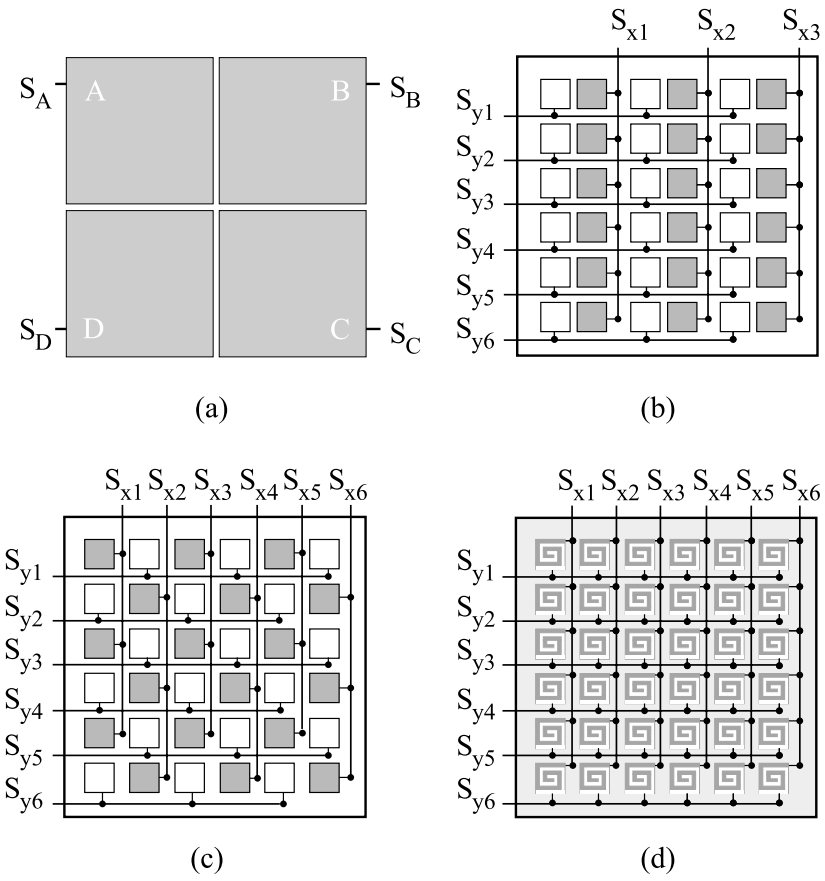

Figure 2.19 *2-D multipixel PSDs: a) quad cell, b) alternating, c) chessboard-like and d) spiral structures.*

The *alternating* structure is depicted in Figure 2.19b. It is a matrix of photodetectors, in which the ones indicated as gray in a column are connected to an *x* sensing line and the ones indicated as white in a row are connected to a *y* sensing line. All photodiodes on a sensing line contribute to its end-point signal, as in the one-dimensional *stripe* structure. The number of photodiodes is equal for sensing displacements in either *x* or *y* direction, but an *x* sensing line has twice as many photodiodes as a *y* sensing line and twice as few end points. This means that position resolution in the *x* direction is half of that in the *y* direction.

This mismatch in coordinate resolution is solved by the chessboard-like structure [2.48, 2.49], as shown in Figure 2.19c. There all gray pixels in a column are connected to an *x* sensing line and all white pixels in a row are connected to a *y* sensing line. Despite its advantage over the previous structure, *x* sensing interprets white photodiodes as gaps and *y* sensing interprets gray photodiodes as gaps. Therefore, for a spot size comparable to the size of one cell has a step-like pseudo-position function, which also depends on the complementary coordinate, i.e., the position *x* measured when the spot is located at (x_1, y_1) is different from that when the spot is at coordinate $(x_1, y_1 + \textit{pixel size})$. This effect is minimized by larger spots or a larger pixel density.

Yet, for the same spot size and number of cells, an enhancement to the performance of the chessboard-like PSD is possible with the spiral structure [2.48, 2.50]. Each cell of the structure consists of two juxtaposed spiral photosensitive elements (depicted as gray and white), where one is connected to the *x* sensing line and the other to the *y* sensing line (Figure 2.19d). Therefore, each *cell* has two *pixels* and contributes to sensing both coordinates. In this case, it features equal resolution and minimum gap for each coordinate, enabling linearity of the pseudo-position function for smaller spots than in the case of the chessboard-like structure, thus rendering a larger spatial dynamic range.

2.5.3 PSD figures of merit

Position-sensitive detectors can be mainly evaluated by their pseudo-position function linearity and their position resolution. These two figures are dependent on the layout and the type of photosensitive element chosen. They have limitations as regards spot intensity and photodetector noise.

The more linear the *PSD* position response is, the more straightforward the calculation of the real coordinate value. For a perfectly linear device, the relation between the real position and the normalized pseudo-position (*PSD* response) is $x = a\, x_{PSD}$, where a is the slope that contains information on the spot size. Consider a *bi cell* with laterally large cells, compared to the spot size, and a perfectly square light spot moving across the cells (Figure 2.20). Position information is only available while the spot

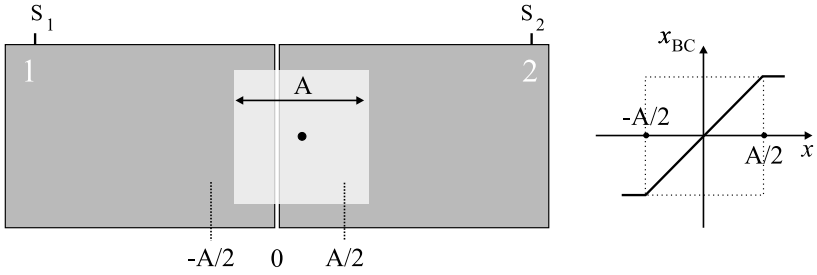

Figure 2.20 *A large bi cell and its response x_{BC} (pseudo-position) as a function of a square spot centroid position.*

illuminates both cells, otherwise, like for a point-like spot, a two-step function is observed. Thus, if the spot side measures A, the centroid position can only be measured if it lies within the interval [*-A/2, A/2*]. In this case, the slope a is simply $(A/2)^{-1}$. This perfect linearity is kept as long as the cells have a finite lateral size larger than A, otherwise edge effects occur, i.e., some boundaries of the spot do not impinge on the detector active area when it moves away from the center, which leads to a non-linear response. Nevertheless, the non-linearity due to edge effects is extracted by normalizing the differential signal. On the other hand, if the spot is not a perfect square the normalized response departs from linear even for spots smaller than the cell size.

For multipixel *PSD*s, like the chessboard, a step-like response function appears if the spot is smaller than or comparable to the lateral cell size and/ or the gap size. In general, the larger the cell density, the smaller the spot sizes the device is able to handle appropriately. Non-linear devices are not

useless but require either an analytical-function approximation or a look-up table. In a nutshell, the most important parameters affecting linearity are:

- finite size of the detector;
- spatially discrete nature of the photo-sensitive structure;
- spot intensity profile;
- density of cells;
- gap between cells.

Another important figure of merit is the *rms* position resolution δ_r^{rms} of the *PSD*, which depends on the device *rms* response error, σ_{PSD}^{rms}. The position resolution is defined as the root-mean-square position displacement that can be measured by the device. Due to photodetector noise, the *PSD* response for a particular spot position can only be determined to within a certain accuracy. This accuracy depends on the spot position, as can be understood from a *bi cell*. If both photodetectors are perfectly identical and subjected to the same conditions, and the spot is distributed evenly over both cells, then they are equally illuminated and their noise levels are equal. When the spot is moved almost completely to the right, the noise levels for the right cell for photon shot noise and photoelectron noise are different from those for the left cell. It is interesting to notice that the smaller the spot, the better the resolution because smaller spots yield a larger response swing for a given centralized spot-position range.

The position signal-to-noise ratio SNR_P is calculated as follows:

$$SNR_P = 10 \log\left(\frac{x_{PSD}(b) - x_{PSD}(a)}{\sigma_{PSD}^{rms}} \right)\Bigg|_{(a,\,b)} \quad [dB] \quad , \qquad (2.50)$$

where σ_{PSD}^{rms} is the root-mean-square PSD-response error for spot centroid displacements within the position interval (a,b).

The spatial dynamic range $||DR||_S$ of a PSD is defined as the maximum range within which the spot centroid can be displaced while still yielding a bi-univocal[12] relationship between the real coordinate and the pseudo-coordinate. As an example, the bi-univocality of a *bi-cell* response is

violated when, for instance, a spot with a well defined boundary does not illuminate one of the cells, in which case the normalized position response is truncated to either +1 or -1, if the spot is smaller than the *bi-cell* lateral size. $||DR||_S$ is chiefly dependent on the ratio of the spot size to the cell size. Sometimes a linear spatial dynamic range $||DR||_S^{lin}$ is defined as the range over which the response is linear.

2.5.4 Centroiding methods for multipixel PSDs

In a Hartmann sensor, each light spot requires one position-sensitive detector and the total number of signal wires can be excessively large. A possibility to minimize it is to integrate with each PSD a centroid computing circuitry as for example the winner-take-all algorithm [2.51, 2.52]. This, however, leads to rather complex circuitry on chip and adds circuitry inaccuracy to the sensor inaccuracy. Yet another option, which is in fact simpler, is based on current division by means of a resistive network [2.53, 2.54]. It attempts to reproduce the idea of the lateral-effect photodiode, but in a discrete fashion.

A resistive line is shown in Figure 2.21, where each end point from a

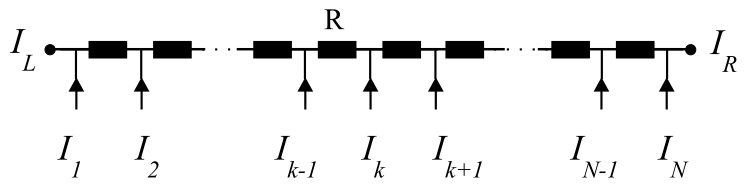

Figure 2.21 *Resistive line with N resistors R.*

sensing line is connected to a node between resistors. The photocurrents from all photodiodes appended to sensing line *k* will contribute to the current on node *k*; part of this current flows to the right and part to the left. All other currents on the other nodes do the same. As a result, weighted sums are obtained at the right-most and left-most nodes as in equations (2.51) and (2.52):

12. A bi-univocal relationship between *a* and *b* indicates that if $a = f(b)$, then each value of *b* yields a single value of *a* and vice versa.

$$I_R = \frac{1}{N} \sum_{i=1}^{N+1} (i-1)\, I_i \tag{2.51}$$

and

$$I_L = \frac{1}{N} \sum_{i=1}^{N+1} (N+1-i)\, I_i \quad . \tag{2.52}$$

The pseudo-position function is then obtained as for a *bi cell*, as seen below in equation (2.53).

$$x_{PSD} = \frac{I_R - I_L}{I_R + I_L} = \frac{\sum_{i=1}^{N+1} [2(i-1) - N]\, I_i}{N \sum_{i=1}^{N+1} I_i} \quad . \tag{2.53}$$

The advantages of the resistive line are its simplicity and the need of only two outputs, independent of the number of sensing lines. For two dimensions, independent resistive lines can be used to sense the *x* and *y* coordinates of the spot centroid. The resistive line, however, should have a rather accurate and equal value of resistances between nodes, in order to avoid an asymmetrical PSD response. An important disadvantage is the introduction of thermal noise to the output signal.

2.6 Chapter summary

This chapter provided the necessary background material on the various subjects needed for the implementation and operation of the sensor. First the fundamentals of wavefront description were addressed, where the Zernike polynomials were presented as a convenient set of orthogonal functions for the description of aberrations over a circular aperture. Then the Hartmann method was investigated more closely, followed by a discussion on how a wavefront can be numerically reconstructed from the measured light spots resulting from the sampled wavefront. The wavefront

reconstruction basically means solving a system of linear equations by means of matrix algebra, where the role of the unknown vector in the surface description depends on the reconstruction approach used: zonal or modal.

The section on photodetection reviewed some basic parameters used to characterize a photodetector and listed the most common photodetectors in silicon, each with its own footprints. The choice of the detector depends on the application requirements and on the specific type of silicon technology selected. The section continued with a word on possible noise sources affecting the performance of a photodetector. Noise is present everywhere from the light source to the circuit appended to the detector. If one understands the cause of noise, one can minimize it through proper design. However, the cause in some cases is not well understood, *1/f noise*, and for other cases a remedy is not available, *photon shot noise*. When modelling a photodetector, one needs a combination of theoretical and empirical results.

A section on optical position-sensitive detectors (PSDs) was included in which both one- and two-dimensional structures were discussed together with the important figures of merit to be observed in a PSD. The idea is to detect the position of the centroid of a light spot. Basically, a convenient signal swing and linearity of the PSD response over a wide position range are very welcome features. Finally, we commented on the current-division centroiding method for multipixel PSDs, which attempts to reproduce the idea of the lateral-effect photodiode (LEP) in a discrete manner. It is simple to implement, but accurate and uniform resistors should be used to guarantee position response linearity.

References

[2.1] M. Born and E. Wolf, “Principles of Optics,” 6th ed., 523-527, Pergamon Press, Oxford, 1986

[2.2] R. J. Noll, “Zernike polynomials and atmospheric turbulence,” J. Opt. Soc. Am., vol.66, no.3, 207-211, 1976

[2.3] D. Malacara and S. DeVore, “Interferogram evaluation and wavefront fitting,” in Optical shop testing, 2nd ed., p.465, edited by D. Malacara, John Wiley & Sons Inc., New York, 1992

[2.4] J. L. Rayces, “Exact relation between wave aberration and ray aberration,” Optica Acta, vol. 11, 85-88, 1964

[2.5] A. J. P. Theuwissen, “Solid-state imaging with charge-coupled devices,” 148-154, Kluwer Academic Press, Dordrecht, 1995

[2.6] E. Hecht, “Optics,” 3rd edition, Addison Wesley Longman Inc., Reading, 1998

[2.7] R. Tyson, “Principles of adaptive optics,” 2nd ed., 252-257, Academic Press, San Diego, 1998

[2.8] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, “Numerical Recipes in C - The art of scientific computing,” 2nd ed., 34-35, 59-70, 676-681, Cambrigde University Press, Cambridge, 1992

[2.9] M. Ruggiero and V. da Rocha Lopes, “Cálculo numérico - Aspectos teóricos e computacionais,” *in Portuguese*, 91-145, McGraw-Hill, São Paulo, 1988

[2.10] A. Jennings, “Matrix computation for engineers and scientists,” 10-11, Wiley and Sons, London, 1977

[2.11] G. Williams, “Linear algebra with applications,” 212-213, Allyn and Bacon Inc., Newton, 1984

[2.12] J. Y. Wang and D. E. Silva, “Wave-front interpretation with Zernike polynomials,” Applied Optics, vol.19, no.9, 1510-1518, 1980

[2.13] R. Tyson, “Introduction to adaptive optics,” 99-105, SPIE Press, Bellingham, 2000

[2.14] R. H. Hudgin, “Wave-front reconstruction for compensated imaging,” J. Opt. Soc. Am., vol.67, no.3, 375-378, 1977

[2.15] W. H. Southwell, “Wave-front estimation from wave-front slope measurements,” J. Opt. Soc. Am., vol.70, no.8, 998-1006, 1980

[2.16] J. Hermann, “Least-squares wave front errors of minimum norm,” J. Opt. Soc. Am., vol.70, no.1, 28-35, 1980

[2.17] J. Liang, “A new method to precisely measure the wave aberrations of the human eye with a Hartmann-Shack wavefront sensor,” Naturwissenschaftlich-Mathematischen Gesamfakultät der Ruprecht-Karls Universität Heidelberg, PhD Thesis, Heidelberg, 1991

[2.18] K. Freischlad and C. Koliopoulos, “Modal estimation of a wave front from difference measurements using discrete Fourier transform,” J. Opt. Soc. Am., vol.3, no.11, 1852-1861, 1986

[2.19] R. Tyson, “Principles of adaptive optics,” 2nd ed., 264-267, Academic Press, San Diego, 1998

[2.20] W. H. Southwell, “What’s wrong with cross coupling in modal wave-front estimation?” Proc. SPIE, vol.365, 97-104, 1982

[2.21] R. Cubalchini, “Modal wave-front estimation from phase derivative measurements,” J. Opt. Soc. Am., vol.69, no.7, 972-977, 1979

[2.22] J. Hermann, “Cross coupling and aliasing in modal wavefront estimation,” J. Opt. Soc. Am., vol.71, no.8, 989-992, 1981

[2.23] B. E. Saleh and M. C. Teich, “Fundamental of photonics,” 544-545, John Wiley and sons, New York, 1991

[2.24] S. M. Sze, “Physics of semiconductor devices,” 2nd edition, 74-79, John Wiley and sons, New York, 1981

[2.25] Y. Tsividis, “Operation and modeling of the MOS transistor,” 2nd ed., 62-71, WCB/McGraw-Hill, Boston, 1999

[2.26] A. J. P. Theuwissen, “Solid-state imaging with charge-coupled devices,” 10-25, Kluwer Academic Press, Dordrecht, 1995

[2.27] W. J. Kindt, “Geiger mode avalanche photodiode array,” Faculty of Information Technology and Systems, PhD Thesis, Delft University Press, Delft, 1999

[2.28] P. N. Dennis, “Photodetectors - an introduction to current technology,” 101-102, Plenum Press, New York, 1985

[2.29] B. E. Saleh and M. C. Teich, “Fundamental of photonics,” 664-665, John

Wiley and sons, New York, 1991

[2.30] R. F. Wolffenbuttel, “Integrated silicon colour sensors,” PhD Thesis, Delft University Press, Delft, 1988

[2.31] M. Bartek, P. Genissen, P. Sarro, P. French and R. Wolffenbuttel, “An integrated silicon colour sensor using selective epitaxial growth,” Sensors and Actuators A, vol.41A-42A, 123-128, 1994

[2.32] G. N. Lu, M. B. Chouikha, G. Sou and M. Sedjil, “Colour detection using a buried double p-n junction structure implemented in the CMOS process,” Electronics Lett. vol.32, no.6, 594-596, 1996

[2.33] B. E. Saleh and M. C. Teich, “Fundamental of photonics,” p.660, John Wiley and sons, New York, 1991

[2.34] A. Biber, “Avalanche photodiode image sensing in standard silicon BiCMOS technology,” PhD Thesis, Swiss Federal Institute of Technology, Zürich, 2000

[2.35] R. H. Kingston, “Optical sources, detectors and systems,” 109-120, Academic Press, San Diego, 1995

[2.36] F. de la Moneda, E. Chenette and A. van der Ziel, “Noise in phototransistors,” IEEE Trans. Elect. Dev., vol. ED-18, no.6, 340-346, 1971

[2.37] M. Puig Vidal, M. Bafleur, J. Buxo and G. Sarrabayrouse, “A bipolar photodetector compatible with standard CMOS technology,” J. Solid-State Electronics, vol.34, no.8, 809-814, 1991

[2.38] P. Bhattacharya, “Semiconductor optoelectronic devices,” 2nd ed., 401-405, Prentice Hall, New Jersey, 1997

[2.39] S. Donati, “Photodetectors - devices, circuits and applications,” 146-174, Prentice Hall PTR, New Jersey, 2000

[2.40] R. W. Sandage and J. Connelly, “Producing phototransistors in a standard digital CMOS technology,” IEDM Technical Digest, 369-372, 1996

[2.41] Y. Tsividis, “Operation and modeling of the MOS transistor,” 2nd ed., 422-424, WCB/McGraw-Hill, Boston, 1999

[2.42] S. Donati, “Photodetectors - devices, circuits and applications,” p.144, Prentice Hall PTR, New Jersey, 2000

[2.43] S. M. Sze, “Semiconductor devices - Physics and technology,” 2nd ed.,

109-112, John Wiley & Sons, Willard, 2002

[2.44] S. Donati, “Photodetectors - devices, circuits and applications,” 127-131, Prentice Hall PTR, New Jersey, 2000

[2.45] D. J. Noorlag, “Lateral-photoeffect position-sensitive detectors,” PhD Thesis, Delft University Press, Delft, 1982

[2.46] D. J. Noorlag and S. Middelhoek, “Two-dimensional position-sensitive photodetector with high linearity made with standard IC-technology,” IEE J. Solid-State and Elect. Devices, vol.3, 75-82, 1979

[2.47] G. A. Tyler and D. L. Fried, “Image-position error associated with a quadrant detector,” J. Opt. Soc. Am., vol. 72, no.6, 804-808, 1982

[2.48] D. W. de Lima Monteiro, G. Vdovin and P. M Sarro, “Various layouts of analog CMOS optical postion-sensitive detectors,” Proc. SPIE, vol. 3794, 134-142, 1999

[2.49] A. Mäkynen, H. Benten, T. Ruotsalainen, T. Rahkonen and J. Kostamovaara, “CMOS position-sensitive photodetectors (PSDs) for integrated sensor systems,” Proc. SPIE, vol. 3100, 89-100, 1997

[2.50] D. W de Lima Monteiro, G. Vdovin and P. M. Sarro, “Position-sensitive detectors for a wavefront sensor,” Proc. 2nd Annual Workshop on Semiconductor Advances for Future Electronics - SAFE, 287-294, 1999

[2.51] J. Lazzaro, S. Ryckebusch, M. Mahowald and C. Mead, “Winner-take-all architecture of O(N) complexity,” Caltech report CS-TR-21-88, 1988

[2.52] L. G. Johnson and S. M. Jalaleddine, “MOS implementation of winner-take-all network with application to content-addressable memory,” Electronic Lett., vol.27, no.11, 957-958, 1991

[2.53] A. Dutta and Y. Hatanaka, “Design and performance of the mesh-type high-speed silicon optical position-sensitive devices (MEPSDs),” IEEE Trans. on Electr. Devices, vol.38, no.3, 498-504, 1991

[2.54] M. Tartagni and P. Perona, “Computing centroids in current-mode technique,” Electronic Lett., vol.29, no.21, 1811-1813, 1993

Technological aspects 3

We chose to develop an integrated wavefront sensor based on the Hartmann method and a number of technological issues need to be considered for its implementation. The Hartmann sensor requires two essential elements: a sampling plane and a light-sensitive detector. Conventionally, either an opaque mask with holes or a microlens array can be used to sample the wavefront. The detector can be an off-the-shelf imager (e.g. CCD) or a customized detector, which is in our case an integrated matrix of optical position-sensitive detectors. The performance of the wavefront sensor is closely related to the technology used to fabricate each of the elements. Optimal performance often requires dedicated processes, which results in high costs. Therefore, by choosing a standard technology, like Complementary Metal-Oxide-Semiconductor (CMOS), we need to take into account its limitations, as mainstream semiconductor processes are primarily meant for electronic circuitry and not for light detection.

This chapter includes a number of issues concerning the technology chosen for each element of the Hartmann sensor. Section 3.1 presents the motivation for choosing standard silicon CMOS technology. Section 3.2 is about the possibilities to manufacture the sampling plane and Section 3.3 focuses on the options available in CMOS for the detector and gives an overview of several photodetector parameters related to the technology.

3.1 Standard silicon CMOS

CMOS is the acronym for the Complementary Metal-Oxide-Semiconductor process, the most widely used technology today. It owes this popularity to its low-power low-voltage capabilities, the availability of both NMOS and PMOS field-effect transistors on chip and the small chip area required for one transistor, compared to the footprint of an equivalent bipolar transistor. The small transistor size favors Very Large Scale Integration (VLSI), which to date prompts tens of millions of logic transistors per square centimeter. Digital circuitry is the workhorse of this technology and represents the cornerstone of computer systems. Analog circuitry in CMOS has achieved increasing importance over the years and features elements ranging from simple switches to operational amplifiers as part of larger Systems on a Chip (SoC). The ample market interest in this technology promotes continuous investigation and technology improvement. Moreover, more and more efforts are spent on developing reliable device models and tools for accurate circuit simulations.

Besides being widely available, a standard CMOS process is generic, supports several photo-sensitive structures, favors the integration of multiple electronic functions with a high yield and is currently the cheapest of the competing technologies. Therefore, it stands as a reasonable choice for the fabrication of the proposed integrated Hartmann wavefront sensor. Bipolar and CCD technologies would be candidates as well, and they are compared to CMOS in the following paragraphs.

3.1.1 CMOS versus Bipolar

Bipolar technology was historically the first in having more than one transistor integrated on a chip, it accommodates both digital and analog circuitry and is usually employed where very high analog performance is required.

The fabrication of Bipolar devices demands a few more process steps than CMOS and the size of a circuit block is usually larger than that of an equivalent CMOS circuit. Also, the power consumption of a Bipolar circuit in stand-by mode is larger than that of a corresponding CMOS circuit and its digital efficiency is poorer [3.1]. Nonetheless, Bipolar circuits are still faster and able to drive larger currents. Besides, Bipolar technology also supports photo-sensitive structures on chip.

Considering the restricted chip area, if a matrix of position-sensitive detectors (PSDs) is to be integrated along with electronic circuitry, it is important to maximize the photo-sensing area and to minimize the circuit area. This favors CMOS if signal multiplexing, amplification, data conversion and process are ever to be integrated along with the PSD matrix. Besides, the potential for technology improvement is greater for CMOS than for Bipolar technology.

BiCMOS is a hybrid technology that combines the advantages of Bipolar and CMOS. It offers higher analog performance and improved speed over CMOS, and lower power dissipation and higher density than Bipolar. In spite of its flexibility it accounts for less than 5% of all fabricated integrated circuits [3.2], and its fabrication costs are obviously higher than for each separate technology [3.3, 3.4].

3.1.2 CMOS versus CCD

CCD technology is robust for imaging applications and offers several charge-transfer architectures and specialized structural layouts, often making use of up to 15 lithographic masks. In general it is optimized for photodetection and features very low noise and a very high fill factor[1]. CCD manufacturing tackles a number of technological issues in order to achieve high charge-transfer efficiency and to minimize crosstalk and noise. Some of these issues are overlapping gates, buried channels, vertical anti-blooming strategy and multiple clock voltages (reaching up to 20V in some cases). They increase the CCD fabrication complexity and cost. Also, because of the serial nature of the charge-transfer process, a conventional CCD does not feature random access to each pixel [3.5, 3.6]. Typically, functional circuitry cannot be integrated reliably on chip and when it is, the performance is poor, unless a combined CCD/CMOS process is meant, which clearly introduces relevant extra costs.

Consider an array of light spots on the surface of a conventional CCD. Charge packets accumulated at isolated pixel clusters (those illuminated by the spots) need to go through ‘dummy’ pixels all the way to reach the output node, limiting the read-out speed. The proposed matrix of PSDs is supposed to improve the read-out speed with respect to that of conventional

1. The fill factor is represented by the ratio of the photosensitive area to the total array area.

solid-state imagers and can rather benefit from the concept of CMOS imagers, whose key features are random pixel address and direct signal transfer to the output line. If we stick to a standard technology for the fabrication of the wavefront sensor, CMOS is a more sensible option due to its wide availability and, more importantly, its capacity to accommodate multi-functional circuitry on chip. Standard CMOS, however, is optimized for electronics and not for imaging and it is also prone to higher noise levels than CCD technology. When very low noise is demanded, modifications to the standard mainstream process should be considered.

3.2 The sampling plane

The sampling plane breaks the wavefront into a number of spots, whose positions are used to calculate the wavefront profile. The Hartmann mask and the microlens array are two possible sampling planes (Figure 3.1). The former is cheap and straightforward to fabricate but blocks a considerable amount of the incident light. The latter involves more expensive and elaborate fabrication methods; when closely packed it features high fill factor[2], but results in stray-light contributions and optical crosstalk.

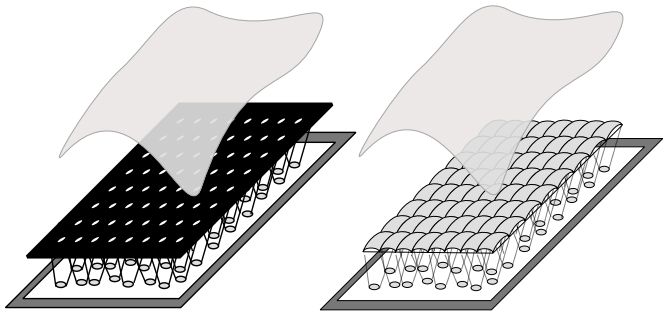

Figure 3.1 *Sampling plane with a) Hartmann mask and b) microlens array.*

2. For a sampling plane, the fill factor represents the ratio of the effective sampling area to the total area.

3.2.1 Hartmann mask

The Hartmann mask is the easiest and cheapest solution to the sampling plane. It is simply an opaque mask with apertures (holes). To match the matrix of position-sensitive detectors, the mask must have the same geometry and pitch as those of the PSDs on the array. The near-field image (Fresnel diffraction) of each sub-aperture on the detection plane can be estimated theoretically and depends on the distance between the mask and the detector, on the wavelength, and on the sub-aperture shape and size [3.7]. The higher the precision of the sub-aperture diameter and edge, the more symmetrical the spot shape, which results in a more symmetrical PSD response to the spot position. Also, the precision of the sub-apertures' grid is important for correct alignment of the mask with the PSD matrix. The disadvantage of the mask is that part of the incident light is blocked, which represents a shortcoming in low-light level applications.

A way to manufacture a Hartmann mask is to etch a grid of holes through a metal layer deposited on a glass substrate, which results in a 1-μm precision. If the pitch is 1000μm, this represents a 99.9-% relative sub-aperture precision. When the input wavefront contains most spatial frequencies smaller than half the sampling frequency (section 2.2), small sub-apertures yield a better geometrical modulation transfer function, but their diameter should be much larger than the wavelength to be measured in order to prevent significant diffraction effects. The ultimate choice of spot diameter and shape is dictated by the type of PSD, as will be seen in section 4.3.2.

The use of Liquid-Crystal Television (LCTV)[3] as a substitute for the Hartmann mask has recently been proposed. One approach is conceptually similar to the Laser Ray Tracing (LRT) method, as explained in section 1.2.3. An LCTV is programmed to have a single sub-aperture at a time, i.e., the liquid-crystal plane is divided into clusters of pixels, all of which block the incident light, except for one, which defines the sub-aperture. The sub-aperture sequentially changes its position over the LCTV screen (clusters are sequentially made transparent) [3.8]. A lens, placed behind the LC plane, images the spot on the image sensor. As for the LRT method, there is no crosstalk between spots and each spot can have a deflection over the

3. The LCTV is a nematic liquid-crystal spatial light modulator.

total area of the imager, increasing the spatial dynamic range and, therefore, the maximum tilt magnitude. Also, arbitrary sub-aperture positions, shapes and sizes can be chosen, by arranging the liquid-crystal pixels in different ways. In contrast to LRT, no moving parts are present. The LCTV, however, can only be used for static measurements. In addition to its slow operation, this approach does not offer advantages over the Hartmann mask for use with the matrix of position-sensitive detectors proposed, where each PSD is spatially fixed. Another approach consists of using an LCTV as an array of programmed lens-effective clusters, instead of simple holes [3.9]. Its main advantage is the possibility to change the array configuration and sub-aperture size, none of which applies to the proposed PSD matrix. A disadvantage of LCTV is that a diffraction pattern due to spaces between liquid-crystal pixels may affect position sensing. Additionally, contrast between ON and OFF crystals is not optimal.

3.2.2 Microlens array

There are basically two types of microlens arrays: spaced and close-packed arrays. In general, spaced microlenses have better optical quality than a close-packed array and an opaque material can be deposited on the interstitial areas between the lenses to prevent that stray light degrades the array performance. In fact, they can be thought of as a Hartmann mask with a microlens located at each hole. The proposed position-sensitive detectors require a certain effective spot size for efficient operation, as will be seen in the next chapter. Therefore, the PSD matrix has to be placed out of the focal plane of the microlens array, such that the spots become as large as required.

For a wavefront sensor, the spaced microlenses approach has no practical advantage over the Hartmann mask because the amount of light transmitted is equal in both cases and microlenses are much more expensive and also more prone to introduce variations in the spot intensity profile, because of microlens-profile variations over the array.

Close-packed microlens arrays offer 100% fill factor and the simplest structure is formed by superimposing two 90º rotated cylindrical microlens arrays. A monolithic close-packed structure has a single plane with adjoining microlenses. In both cases, the edges have much larger aberrations than the lens centers and the edge lines contribute significantly

to optical crosstalk, which can be overcome to a large extent if apodization methods are used, at the expense of extra steps and higher costs [3.10]. Nevertheless, an important advantage over the Hartmann mask is its more efficient light transmission.

A number of techniques exists for microlens array fabrication. Some of the most used are planar graded index by ion diffusion [3.11], photothermal swelling [3.12, 3.13], hot pressing, laser ablation [3.14], mechanical profiling [3.11], photoresist melting [3.15] and photoresist engraving [3.16, 3.17]. Most of these techniques are proprietary and encompass a number of controlled steps to guarantee good optical quality. The photoresist technique has been enhanced since it was first proposed and has been used by many companies. It offers the best compromise between quality, speed and cost of manufacture [3.18]. A microlens array is, nevertheless, much more expensive than a Hartmann mask. Microlens array replication is possible but the manufacture of the custom master copy is costly. Inexpensive microlens arrays stamped on plastic by a stainless-steel ball have been reported [3.19], however, the large lens-profile variations result in spot-profile variations that would lead to considerable centroid-estimation errors on the PSD.

Alternative methods, for example tunable microlens arrays involving different technologies such as liquid-crystal spaced microlens arrays [3.20] and electro-optic wafers ($LiNbO_3$ or $Sr_{0.3}Ba_{0.7}Nb_2O_6$) [3.21] have also been proposed, but have not been completely developed yet. The former still induces considerable aberrations to the spot and the latter has only been considered theoretically.

We have recently succeeded in fabricating contiguous orthogonal and hexagonal micromirror array structures in silicon by means of anisotropic etching with KOH:H_2O (Potassium hydroxide aqueous solution) based on the technology suggested by D. Kendall [3.22]. The structures served as a template for a close-packed 300-μm pitch PDMS[4] microlens array on a glass substrate. It offers a promising alternative to the Hartmann mask especially because its fabrication is as inexpensive and simple as the Hartmann mask we made (see section 4.2), with the advantage of yielding more light per spot. The orthogonal array was not tested with the implemented detector because the pitches were not compatible. Figure 3.2

4. PDMS - PolyDiMethyl Siloxane.

shows the cross section and photograph of part of the hexagonal microlens array. A grid of square openings is etched on an oxide layer previously deposited on silicon. Then the bulk silicon is anisotropically etched through these openings, with a KOH solution, until pyramidal pits are achieved. The oxide layer is then stripped and the wafer surface is etched further with KOH. This thins the wafer and transforms the pits into smooth spherical wells. A long enough etching time causes the wells to overlap laterally resulting in a close-packed two-dimensional array of micromirrors[5].

3.3 The Detector

The photodetectors in a position-sensitive detector matrix (section 2.5) must comply with standard CMOS capabilities. The photodetector structures in a n-well[6] process are introduced in this section; some pixel architectures too. Mismatch and crosstalk are discussed and the effects of dielectric layers and junction depth on the quantum efficiency are indicated. Also, we discuss the influence of the adoption of a state-of-the-art technology on photodetectors.

3.3.1 Photodetectors in CMOS

In a standard n-well CMOS process, several photosensitive structures are possible as junction photodiodes and phototransistors. A photogate may be implemented too, depending on certain characteristics of the process chosen. Additionally, an avalanche photodiode can be realized in CMOS technology but this does not allow read-out electronics on chip.

The photodiodes are the n^+*/p-epilayer*, p^+*/n-well*, and *n-well/p-epilayer* junctions, indicated in Figure 3.3. The phototransistors are a vertical p^+*/n-well/p-epi* structure and a lateral p^+*/n-well/*p^+ phototransistor, both shown in Figure 3.4. Design variations of these two structures have been reported [3.23]. For the vertical phototransistor, the *base/collector* junction is where the photocurrent - which is also the base current- is generated. The *emitter/*

5. The fabrication and the characterization of these micromirror/microlens arrays were done in collaboration with G. Vdovin, O. Akhzar-Mehr, M. Loktev and P. Sarro.
6. Equivalent photodetector structures are possible in a p-well process.

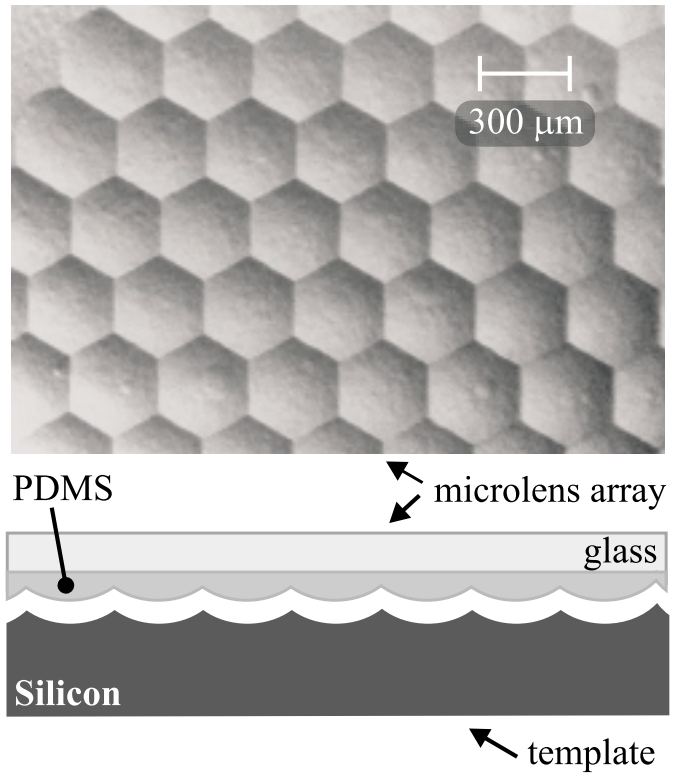

Figure 3.2 *Implemented low-cost microlens array.*

base junction is small enough, and is preferably covered, to minimize its effect on the main photo-generated current. The lateral phototransistor is fully implemented inside the *n-well* and only the area between the two diffusions should be exposed, in order to avoid contributions from the two vertical *pnp* phototransistors.The dark current associated with this structure is usually higher than with the vertical one. Still a third phototransistor structure is possible, namely, the lateral n^+*/p-epi/n-well* (not shown), but no array is advised because of the lack of reliable isolation.

For use in a array, the p^+*/n-well* photodiode is the least sensitive to substrate noise and crosstalk with neighboring pixels, because each junction is isolated within its own n-well. However, dense arrays cannot afford the huge separation required between two wells and if electronic

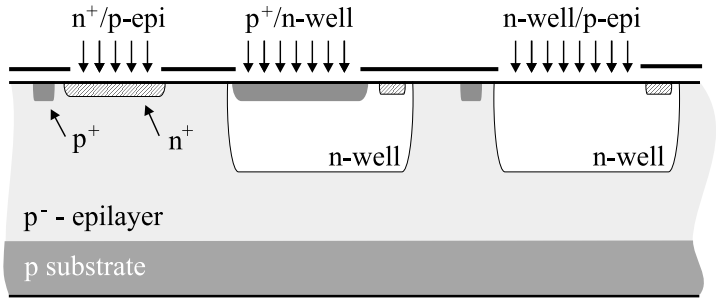

Figure 3.3 *Three junction photodiodes in a n-well CMOS process: n+/p-epi, p+/n-well and n-well/p-epi.*

circuitry is to be added to each photodiode, the minimum distance between the well and the drain/source of a surrounding NMOS transistor must also be sufficiently large to avoid latchup[7]. Equally, *n-well/p-epilayer* junctions and phototransistors impart a large pitch on an array. This favors the *n^+/p-epilayer* junction where dense matrices are desired, with the disadvantage of higher crosstalk and lower sensitivity for longer wavelengths than the *n-well/p-epilayer* junction.

In addition to the previous photo-sensitive structures, a photogate can be implemented, which is actually a MOS capacitor, like the SCCD structure discussed in section 2.4.3. A voltage applied to the (photo)gate creates a depletion region under it, where photo-generated electrons are stored. There is one important issue: the gate material must be highly transmissive for the wavelengths used, i.e., preferably no silicide or metal should be used on the gate [3.24, 3.25]. To increase light transmission, opening windows can be designed through the photogate, but they can only be used if the pixels are large enough.

Yet an avalanche photodiode can be implemented with an *n^+/p-epilayer* junction and an overlapping *n-well* guard ring, as shown in Figure 3.4 [3.26]. The lightly doped *n-well* guard ring is necessary to prevent

7. Latchup occurs when parasitic transistors (connected as a thyristor) are formed between adjoining structures causing short-circuit currents in the bulk. Its occurrence is considerably prevented by using a lightly doped *p-epilayer* on top of the substrate [3.1].

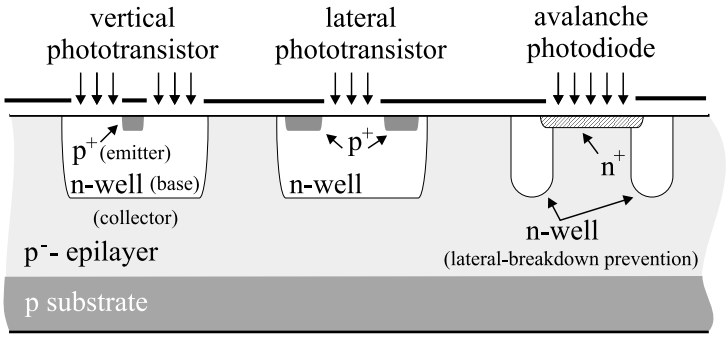

Figure 3.4 *Parasitic phototransistors in CMOS and a possible avalanche photodiode structure.*

premature breakdown at the lateral edges of the main diode. Avalanche multiplication depends on the doping concentration, and too large concentrations ($>10^{18}$ cm^{-3}) result in breakdown predominantly by tunneling instead of avalanche, compromising the device gain. Therefore, sub-micron technologies, which are characterized by high doping concentrations, are not appropriate [3.27]. The breakdown voltage is 82V in a 2μm process and since the n^+ diffusion is connected to such a high voltage, the read-out circuitry must withstand it, which is not feasible in the framework of standard CMOS.

3.3.2 Pixel architectures

For an array of photosensitive elements, it makes sense to provide each pixel with at least one switch to allow signal multiplexing on a data bus. Pixel categories are generally labeled as passive and active, depending on the absence or presence of an amplifying transistor, respectively [3.28].

Passive pixel

A passive pixel is simply a photodiode with the output connected to the data line via a MOS transistor, as a switch. When the transistor is switched on, the charge accumulated at the junction capacitance of the photodiode is transferred to the data line, where a charge-to-voltage conversion takes place. For an *n^+/p-epilayer* junction, the layout of this pixel is very compact with the extended drain of the NMOS transistor serving as the photodiode.

Figure 3.5 shows the layout and the diagram of such a pixel.

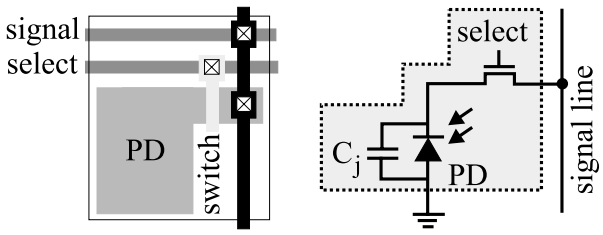

Figure 3.5 *Layout and diagram of a passive pixel with an n^+/ p-epilayer photodiode and an NMOS transistor as switch.*

The main advantages of passive pixels are the high fill factor, large voltage swing and layout simplicity. The main disadvantage is that the photodiode junction capacitance adds to the long capacitive line and results in a longer response time and excess noise charge[8].

Active pixel

In addition to the address switch, as in the previous architecture, the active-pixel concept uses an amplifying transistor per pixel [3.32, 3.33]. A diagram of this structure is shown in Figure 3.6a. The photodiode is appended to a sense node, which is connected both to the gate of the amplifying transistor M2 and to the source of the reset transistor M1. M3 is the pixel address transistor.

For $V_{out} > (V_{load} - V_{t4}) > 0$, the load transistor M4 is in saturation and can be represented as a current source with a load current given by equation (3.1):

$$I_{load} = 0.5\mu_n C_{ox}(W/L)_{M4}\,(V_{load}-V_{t4})^2, \qquad (3.1)$$

8. This is called kTC noise and it is derived from the thermal noise of the channel resistance of the transfer transistor when charging a capacitance. The noise charge in electrons is $q_{noise} = (\sqrt{kTC})/e$.

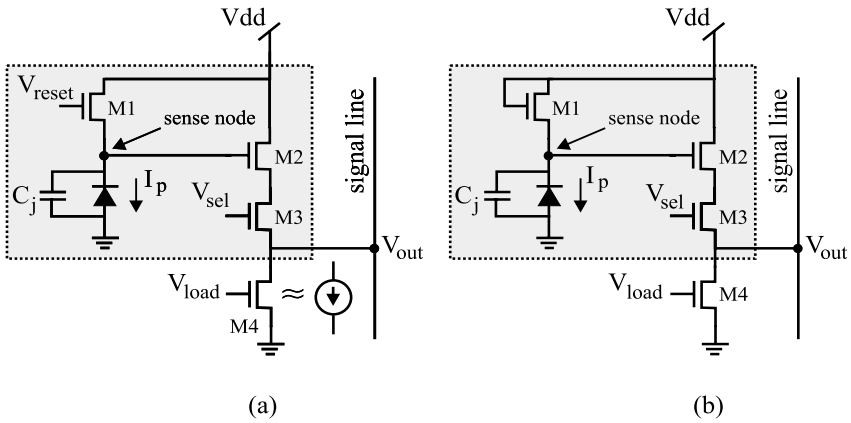

Figure 3.6 *Active pixel circuits: a) linear and b) logarithmic*

where μ_n is the electron mobility, C_{ox} the gate-oxide capacitance, $(W/L)_{M4}$ the aspect ratio of M4 and V_{t4} the threshold voltage of M4. This current forces M2 into the saturation regime too, which results in the following relation to V_{out}:

$$V_{out} = V_{sn} - V_{t2} - \sqrt{\frac{2I_{load}}{\mu_n C_{ox}(W/L)_{M2}}} \quad , \tag{3.2}$$

where V_{sn} is the sense-node voltage, V_{t2} is the threshold voltage of M2 and $(W/L)_{M2}$ is the aspect ratio of M2.

When V_{reset} is applied to M1, it resets the sense node to a voltage equal to $V_{sn} = V_{reset} - V_{t1}$, where V_{t1} is the threshold voltage of the reset transistor. When M1 is turned off, the photocurrent starts discharging the junction capacitor C_j and the voltage at the sense node is lowered to V_{sn}' by an amount ΔV_p proportional to the photocurrent I_p and the integration time τ_{int}:

$$\Delta V_p = \frac{I_p \, \tau_{int}}{C_j} \, . \tag{3.3}$$

Therefore, V_{out} varies linearly with the photocurrent, provided that C_j remains constant with the sense node-voltage V_{sn}. Random access in the time domain is not possible because the integration time should be kept constant all over the array. For different transistors in the array, variations in the threshold voltages V_{t1} and V_{t2} cause fixed-pattern noise (FPN). By reading the sense-node voltage after reset and then again after integration, one can subtract the signal voltage from the reset voltage, eliminating the contributions of the threshold voltages. This cancellation method is called correlated double sampling (CDS). It also cancels the effect of kTC noise (see footnote 8) introduced by the reset transistor.

A modification of the linear active pixel is the logarithmic pixel, shown in Figure 3.6b. The gate of the reset transistor M1 is always connected to the supply voltage and the operation of M1 is pushed into the subthreshold regime by the small photocurrent from the photodiode [3.34]. In weak inversion the *I*-*V* characteristic of M1 is given by equation (3.4):

$$I_p = I_{ds} = I_0 \exp\left[\frac{e}{nkT} (V_{dd} - V_{sn} - V_{t1}) \right], \tag{3.4}$$

where n is the subthreshold slope and I_0 is a process-dependent parameter that is proportional to the gate oxide capacitance and the aspect ratio of M1. Rearranging the terms in equation (3.4), we obtain the expression for the sense-node voltage:

$$V_{sn} = (V_{dd} - V_{t1}) - \frac{nkT}{e} \ln\left(\frac{I_p}{I_0} \right). \tag{3.5}$$

The sense-node voltage varies logarithmically with the photocurrent, which provides the pixel with a very high dynamic range, but a restricted voltage swing. Because there is no reset operation, it allows real random access in both the space and the time domains. As for the linear active-pixel architecture, an array of logarithmic pixels is subjected to FPN. However, a main disadvantage of the logarithmic pixel is that there is no use in applying CDS, as there is no sense-node voltage as reference. Several reset solutions for FPN cancellations in log pixels have been proposed [3.35, 3.36], at the expense of increased pixel area because of the extra transistors

required.

In a linear active pixel, a photogate can substitute the photodiode [3.37] (Figure 3.7a). When a voltage is applied to the photogate, photo-generated charge accumulates under it. Then, when a voltage is applied to the transfer gate, the charge is transferred to a floating node n^+, where it is sensed. A partial overlap of the transfer gate with the photogate ensures a better transfer efficiency, but many standard CMOS processes only have one

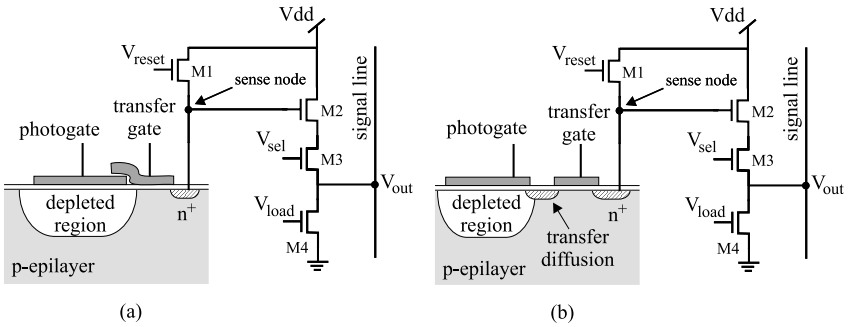

Figure 3.7 *Active pixel with a) a photogate with overlapping gates and b) a photogate with a n^+ transfer diffusion.*

polysilicon layer. In this case a transfer diffusion can be used between the gates (Figure 3.7b).

An alternative to the photogate is the pinned photodiode, in which a carefully designed p^+ layer is created on top of an *n* region (Figure 3.8). The shallow p^+ is *pinned* to the substrate voltage. It stabilizes the value of the transfer capacitance and reduces dark current. Photo-generated carriers are accumulated and transferred to the floating diffusion *n+*, which is connected to the sense node. Also, it is possible to apply a voltage to the *n* region so that the two depletion layers formed around p^+/n and *n/p-epi* are enlarged and touch each other, eliminating majority carriers in the *n* region [3.38]. The voltage at which this happens is termed the pinned voltage, V_p. It can be found by solving the following equations[9]:

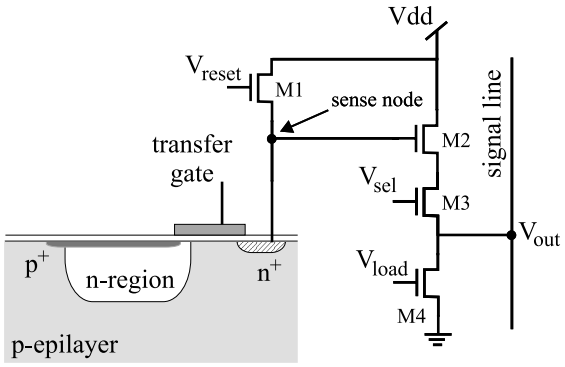

Figure 3.8 *Pinned photodiode. The p^+ layer is 'pinned' to the epitaxial layer voltage.*

$$x_{n1} + x_{n2} = W_n, \tag{3.6}$$

$$x_{n1} = \sqrt{\frac{2\varepsilon}{e}\frac{N_{A1}}{N_D(N_{A1}+N_D)}(\phi_{bi1}+V_{b1})}, \tag{3.7}$$

$$x_{n2} = \sqrt{\frac{2\varepsilon}{e}\frac{N_{A2}}{N_D(N_{A2}+N_D)}(\phi_{bi2}+V_{b2})}, \tag{3.8}$$

$$V_{b1} = V_{b2} = V_p, \tag{3.9}$$

where the sub-indices *1* and *2* and related to the upper and bottom junctions, respectively. W_n is known from the process and given by the total width of the *n* region minus the *p* region width. V_p should be lower than the transfer-gate voltage to guarantee a reliable transfer. It is possible

9. Equations (3.6) and (3.7) are derived from equations (2.25) through (2.28).

to satisfy this condition with the foregoing equations by choosing appropriate doping concentrations (N_D, N_{A1} and N_{A2}). This, however, usually implies modifications to the standard CMOS process. In contrast to the photogate, the pinned photodiode pushes the charge accumulation away from the surface (as for the BCCD, see section 2.4.3), which reduces the dark current. A disadvantage of the pinned photodiode, however, is that its saturation charge is about twice as low as that of a photogate. This is because, for a pinned photodiode, the transfer-gate voltage has to be altered, instead of a photogate voltage, to enable charge transfer from the collection site to the sense node.

3.3.3 Pixel mismatch and crosstalk

For a specific process, the smaller the photosensitive area and the transistor-gate area, the higher the mismatch. The mismatch in photodiode arrays can occur because of variations in the doping profile, dark-current fluctuations and cosmetic defects on the silicon surface and on the dielectric layers. MOSFET mismatch is mainly due to threshold-voltage and channel-size variations. Both photodiodes and transistors are affected by temperature variations across the chip. When the photodiode area and transistor channel are large, most of these effects are averaged out, reducing mismatch. Crosstalk between pixels depends on the separation distance and on the type of isolation used.

Channel size and temperature are reasonably constant over the chip [3.36], which indicates that transistors are mainly affected by threshold-voltage variations, which depend on the doping profile, on the non-uniformity of the oxide capacitance and on the source-to-bulk voltage.

Dark-current fluctuations occur due to charge trapping in intermediate energy states created by lattice defects, mismatches and impurities. These are most common at interfaces, such as the top Si/SiO_2 interface and lateral interfaces of Si with the isolation, for instance, LOCOS (local oxidation of silicon) or trench isolation. Several annealing steps in standard CMOS contribute to a reduction in the defect density. Additionally, the *<100>* wafer-surface orientation is generally chosen over the *<111>* orientation because it has ten times less interface traps [3.29]. The dark-current fluctuations are proportional to the square-root of the average dark-current, which increases with temperature. Around room temperature, the dark

current approximately doubles for every 8ºC increase in temperature [3.30, 3.39]. At 300K the dark-current density has typical values over 1000pA/cm^2 for a standard CMOS process. An optimized process can reduce these to a range around 50 to 200pA/cm^2, which is still high compared to the 10pA/cm^2 of high-end CCDs[10][3.40, 3.41].

Electrical crosstalk is caused by charge leakage from a pixel to its neighbors. If the pixels are located too close to each other and the diffusion length is sufficiently large, generated charges in one pixel move by diffusion through the substrate and may end up being collected by another pixel. This effect is minimized when isolation is present between pixels, e.g. LOCOS or trench isolation. Moreover, if a lightly doped epilayer is used, noise coupling becomes independent of pixel separation, provided that the distance between pixels is at least four times larger than the epitaxial-layer depth [3.30, 3.31]. Also, guard rings can be designed around each pixel to minimize crosstalk, at the expense of increased pixel area.

3.3.4 Quantum efficiency

The quantum efficiency of a photodiode indicates how well it collects the incident light and is generally expressed as a percentage. It is determined by intrinsic and extrinsic parameters. Intrinsic parameters are the type of semiconductor, junction depth, depletion width and carrier diffusion length. Extrinsic parameters are, for instance, dielectric coatings and effective photosensitive area (fill factor).

At a particular wavelength, the quantum efficiency varies for each of the junction photodiodes in CMOS. The shallower the junction, the more efficient it is in sensing shorter wavelengths. In the visible spectrum (400-700nm), blue light is more efficiently collected by a shallower junction (p^+/*n-well* and n^+/*p-epi*) and red light by a deeper junction (*n-well/p-epi*). The wider the junction, the higher the chance that carriers are collected. In a standard process, the junction depth and the doping profile are fixed, and one can only change the depletion layer width by applying a reverse bias voltage (see equation (2.25)). Also, a longer carrier diffusion length improves the chance that a charge generated outside the depletion region

10. The dark-current densities of CCDs normally range from 10pA/cm^2 to higher than 1000pA/cm^2, depending on the manufacturer and mode of operation.

will eventually reach it and be collected. Lower doping profiles result in longer diffusion lengths.

An extrinsic parameter directly affecting the quantum efficiency is the stack of dielectric layers that may be present on top of the photodiode. They act as a thin-film stack and influence the optical power transmittance for each wavelength independently. As shown in the graph of Figure 2.13, when no dielectric layer is present, the *air-Si* interface represents a great transmission loss. We can achieve a much better and still smooth transmittance over a wide range of the spectrum if a single 0.1-μm oxide layer is present on top of the photodiode. A properly designed anti-reflective coating can optimize transmittance even further, but this represents a post-processing step. If a combination of dielectric layers is present, the thickness of most of these layers is not accurately controlled and each process batch is prone to result in a different combination of thicknesses. Therefore, it is important to know what the tolerances are and to calculate the transmittance for each combination of thicknesses within the given tolerance margins. In some cases, the removal of all dielectric layers from the top of the Si surface would offer a better choice. However, from the practical point of view, this is not feasible because a thermal oxide layer would grow there over time, compromising the transmittance anyway.

Another parameter that influences the quantum efficiency is the effective photosensitive area (fill factor) of an array of photodiodes, which depends on the density of the array and on the pixel architecture chosen. Obviously, a low-density and passive-pixel array features the largest active area.

3.3.5 Conservative versus state-of-the-art CMOS processes

Although breakthroughs in the minimum gate length of transistors allow an increasingly higher density of circuits on a chip, they often result in disadvantages for photodetectors. Several modifications to the technology are necessary as the feature size decreases. For instance, junctions become shallower, doping concentrations higher and metal layers and dielectric stacks more numerous.

Junctions closer to the surface reduce the quantum efficiency for longer

wavelengths. Higher doping levels create thinner depletion regions around the junctions, reducing even further the quantum efficiency, which is also compromised by a shorter carrier lifetime and diffusion length [3.24, 3.29].

With an increase in the number of metal layers, more insulating layers are necessary. If the dielectric-layer stack is removed from the top of the photodetector, high side walls are left, which block off-normal light incidence. In CMOS image sensors, a countermeasure is the use of microlenses.

In sub-micron technologies, the transistor threshold voltage does not scale down proportionally to the supply voltage. For active pixels it means a more limited voltage swing and consequently a reduced dynamic range.

For large photodetectors, the perimeter-to-area ratio is smaller and the edge effects contribute less to the dark-current fluctuations. [3.42], i.e., LOCOS- or trench-isolation defects induced on the edges become less important, but they can still be relevant at low light levels. The dielectric side walls also become less critical. When the dielectric stack is left on top of the light-sensitive area, each layer has an inaccuracy in its thickness and the transmittance for a particular wavelength through the stack may vary considerably between two wafers and especially from one fabrication batch to another.

3.4 Chapter summary

Chapter 3 was dedicated to the technological issues involved in the fabrication of the Hartmann wavefront sensor. Although not optimized for imaging applications, CMOS was the standard technology of choice for several reasons, the most important of which are its availability and the possibility of integrating multiple electronic functions on chip. Two alternative ‘standard’ options for the development of this sensor would be the Bipolar and the CCD technologies. However, if more and more minimum-sized mixed-signal electronic circuitry is to be integrated in newer versions of the sensor, neither of these two would be a judicious choice. In a combined CCD/CMOS process, low noise and imaging robustness of CCDs can be added to CMOS functionality, but relevant extra costs are to be expected.

With regard to the sampling plane of the sensor, the Hartmann mask and a microlens array were considered. The former is fundamentally cheaper

but blocks a considerable amount of incident light, whereas the latter is usually more expensive but when in a close-packed structure, it features a higher light-collection efficiency. Several technologies for lens arrays that are either currently used or investigated were listed. We recently managed to fabricate an inexpensive microlens array that represents a promising alternative to the mask we used.

This chapter also discussed a number of feasible photodetector structures in standard CMOS technology, such as junction photodiodes and phototransistors. A photogate may also be implemented, depending on certain characteristics of the process chosen. Additionally, an avalanche photodiode can be realized but it renders the implementation of read-out electronics on chip unfeasible.

Passive- and active-pixel architectures were investigated. The chief advantages of passive pixels are their high fill factor, large voltage swing and layout simplicity. Their drawbacks are a longer response time and excess noise charge. In the case of active pixels, the photosensitive site is buffered from the signal line and the response is faster. There are more transistors per pixel and a more limited voltage swing, however; threshold-voltage variations result in fixed-pattern noise and additional circuitry must be used to cancel it. A variation on the linear active pixel is the logarithmic pixel which, besides the random spatial access also offered by its linear counterpart, permits asynchronous operation. Nonetheless, fixed-pattern noise cancellation is a dreadful task and the voltage swing is extremely small.

In terms of pixel size, the smaller the photosensitive area and the transistor channel, the higher the signal mismatch. Most of the variations causing mismatch are averaged out when a large-area photodiode is used. The dark-current density in standard CMOS is roughly one order of magnitude higher than that of conventional CCDs. An optimized CMOS process can feature a dark-current density in the range between 50 and 200 pA/cm^2. However, this is still much higher than the 10 pA/cm^2 obtained by specialized CCD imagers.

The quantum efficiency of a photodiode is determined by intrinsic and extrinsic parameters. Intrinsic parameters are the type of semiconductor, junction depth, depletion width and carrier diffusion length. However, there is little room to play with these parameters in standard CMOS. The most common extrinsic parameters are the stack of dielectric layers

possibly left on the photodiode and the pixel effective area (fill factor). One can tailor these parameters to a certain extent, by adding post-processing steps and by design.
A remarkable observation presented in this chapter is that the rapid scaling of CMOS processes often result in disadvantages for photodetection. The tendency to have devices closer to the surface reduces the quantum efficiency for longer wavelengths; the tendency to use a larger number of metal and dielectric layers results in blockage of off-normal light incidence on small pixels, while the usage of a lower voltage supply restricts the output signal swing of pixels.

References

[3.1] M. Ohring, “Reliability and failure of electronic materials and devices,” 71-72, Academic Press, San Diego, 1998

[3.2] H. Veendrick, “Deep-submicron CMOS ICs - From basics to ASICs,” vi-vii, Kluwer, Deventer, 1998

[3.3] A. R. Alvarez, “BiCMOS technology and applications,” 2nd ed., Chapter 1, Kluwer Academic Publishers, Boston, 1993

[3.4] A. Moini, “Vision chips,” 256-260, Kluwer Academic Press, Boston, 2000

[3.5] A. J. P. Theuwissen, “Solid-state imaging with charge-coupled devices”, Chapter 4, Kluwer Academic Press, Dordrecht, 1995

[3.6] J. R. Janesick, “Scientific charge-coupled devices,” 387-418, SPIE Press, Bellingham, 2001

[3.7] E. Hecht, “Optics,” 3rd edition, 476-500, Addison Wesley Longman Inc., Reading, 1998

[3.8] S. Olivier, V. Laude and J. P. Huignard, “Liquid-crystal Hartmann wavefront sensor,” Applied Optics, vol.39, no.22, 3838-3846, 2000

[3.9] J. Rha and M. Giles, “Implementation of an adaptive Shack-Hartmann sensor using a phase-modulated liquid crystal spatial light modulator,” Proc. SPIE, vol.4493, 80-87, San Diego, 2001

[3.10] J. D. Mansell, D. R. Neal, “Apodized microlenses for Hartmann wavefront sensing,” Proc. 2nd International Workshop on Adaptive Optics for Industry and Medicine, 203-208, Durham, 1999

[3.11] D. Daly, “Microlens arrays,” 26-28, Taylor and Francis, London, 2001

[3.12] N. F. Borrelli, D. Morse, R. Bellman and W. Morgan, “Photolytic technique for producing microlenses in photosensitive glass,” Applied Optics vol.24, no.16, 2520-2525, 1985

[3.13] M. Kufner and Stefan Kufner, “Micro-optics and lithography,” 81-89, VUB Press, Brussels, 1997

[3.14] S. Mihailov and S. Lazare, “Fabrication of refractive microlens arrays by excimer laser ablation of amorphous Teflon,” Applied Optics, vol.32, no.31, 6211-6218, 1993

[3.15] Z. D. Popovic, R. Sprague and G. Neville Connell, “Technique for

monolithic fabrication of microlens array," Applied Optics, vol.27, no.7, 1281-1284, 1988

[3.16] G. Artzner, "Microlens arrays for Shack-Hartmann wavefront sensors," Optical Eng., vol.31, no.6, 1311-1321, 1992

[3.17] M. C. Hutley, "Refractive lenslet arrays," in Micro-optics - elements, systems and applications, 127-152, edited by H. Herzig, Taylor & Francis, London, 1997

[3.18] D. Daly, R. Stevens, M. Hutley and N. Davies, "The manufacture of microlenses by melting photoresist," Measurement Science and Technology, vol.1, no.8, 759-766, 1990

[3.19] G. Y. Yoon, T. Jitsuno, M. Nakatsuka and S. Nakai, "Shack Hartmann wavefront measurement with a large f-number plastic microlens array," Applied Optics, vol.35, no.1, 188-192, 1996

[3.20] S. Masuda, S. Fujioka, M. Honma, T. Nose and S. Sato, "Dependence of optical properties on the device and material parameters in liquid crystals microlenses," Jpn. J. Appl. Phys. - Part I, vol.35, no.9A, 4668-4672, 1996

[3.21] M. Kulishov, "Tunable electro-optic microlens array - I. Planar geometry," Applied Optics, vol.39, no.14, 2332-2337, 2000

[3.22] D. L. Kendall, W. Eaton, R. Manginell and T. Digges Jr., "Micromirror arrays using KOH:$H_2$0 micromachining of silicon for lens templates, geodesic lenses, and other applications", Opt. Eng., vol.33, no.11, 3578-3588, 1994

[3.23] R. W. Sandage and J. Connelly, "Producing phototransistors in a standard digital CMOS technology," IEDM Technical Digest, 369-372, 1996

[3.24] H. Wong, "Technology and device scaling considerations for CMOS imagers," IEEE Trans. Electr. Dev., vol.43, no.12, 2131-2142, 1996

[3.25] H. Wong, R. Chang, E. Crabbé and P. Agnello, "CMOS active pixel image sensor fabricated using a 1.8-V, 0.25-μm CMOS technology," IEEE Trans. Electr. Dev., vol.45, no.4, 1998

[3.26] A. Biber, "Avalanche photodiode image sensing in standard silicon BiCMOS technology," PhD Thesis - ETH Nr.13544, Swiss Federal Institute of Technology, Zürich, 2000

[3.27] A. Biber, P. Seitz and H. Jäckel, "Avalanche photodiode image sensor in standard silicon BiCMOS technology," Sensors and Actuators A, vol.90,

82-88, 2001

[3.28] E. R. Fossum, “CMOS image sensors: electronic camera-on-a-chip,” IEEE Trans. Elect. Dev., vol.44, no.10, 1689-1698, 1997

[3.29] S. M. Sze, “Semiconductor devices - Physics and technology,” 2nd ed., 504-505, John Wiley & Sons, Willard, 2002

[3.30] N. Marston, “Solid-state imaging: a critique of the CMOS sensor,” PhD Thesis, University of Edinburgh, 1998

[3.31] M. J. Loinaz, D. K. Su and B. Wooley, “Experimental results and modeling techniques for switching noise in mixed-signal integrated circuits,” Digest of Technical Papers, IEEE Symposium on VLSI circuits, 40-41, 1992

[3.32] E. R. Fossum, “Active pixels sensors - are CCDs dinosaurs?” Proc. SPIE, vol.1900, 2-14, 1993

[3.33] S. K. Mendis, S. Kemeny, R. Gee, B. Pain, Q. Kim and E. Fossum, “Progress in CMOS active pixel image sensors,” Proc. SPIE, vol.2172, 19-29, 1994

[3.34] D. Scheffer, B. Dierickx and G. Meynants, “Random addressable 2048x2048 active pixel image sensor,” IEEE Trans. Electr. Dev., vol.44, no.10, 1716-1720, 1997

[3.35] L. G. McIlrath, V. Clark, P. Duane, R. McGrath and W. Waskurak, “Design and analysis of a 512x768 current-mediated active pixel array image sensor,” IEEE Trans. Electr. Dev., vol.44, no.10, 1997

[3.36] Y. Hagihara, “Logarithmic-converting CMOS area image sensors with an FPN cancel circuit,” *in Japanese*, Eizou Jouhou Media Gakkai Publication (Image Information Media Society), vol.55, no.7, 1039-1044, 2001

[3.37] S. Mendis, S. Kemeny and E. Fossum, “CMOS active pixel image sensor,” IEEE Trans. Electr. Dev., vol.41, no.3, 452-453, 1994

[3.38] T. Lulé, S. Benthien, H. Keller et al., “Sensitivity of CMOS based imagers and scaling perspectives,” IEEE Trans. Electr. Dev., vol.47, no.11, 2110-2122, 2000

[3.39] S. Smith et al., “A single-chip CMOS 306x244-pixel NTSC video camera and a descendant coprocessor device,” IEEE J. Solid-state circ., vol.33, no.12, 2104-2111, 1998

[3.40] J. R. Janesick, “Scientific charge-coupled devices,” 605-648, SPIE Press,

Bellingham, 2001

[3.41] J. R. Janesick, “Dueling detectors - CMOS or CCD? The choice depends on the application,” OE Magazine - SPIE, 30-33, no.2, 2002

[3.42] S. G. Chamberlain, “Photosensitivity and scanning of silicon image detector arrays,” IEEE J. Solid-State Circ., vol.SC-4, no.6, 1969

Sensor implementation 4

The implementation of a customized integrated wavefront sensor in the framework of standard CMOS technology is a challenging task that demands one to make sensible choices in terms of process parameters, photosensitive structures and level of integration. Before a first complete prototype is fabricated, a number of test structures must be made and their appropriateness in a complete system evaluated. In the previous chapter, important technological considerations were tackled and this chapter addresses the choices that were made.

In Section 4.1 the process used is characterized, then there is a concise explanation about the fabrication of the Hartmann mask in Section 4.2, followed by a thorough description of all implemented position-sensitive detectors (PSDs) and their results in Section 4.3. Section 4.4 ends the chapter with a discussion about the implementation of the CMOS wavefront sensor, its layout, schematics, detector and electronic parts and the interface to the computer.

4.1 Technology

The process we used to implement all test structures and the wavefront sensor is the standard CMOS process available at DIMES (Delft Institute for MicroElectronics and Sub-micron technology). It is a 1.6-μm n-well

single-polySi-layer double-metal-layer process with LOCOS isolation. A lightly doped *p-epilayer* (10 ± 2μm) is present on top of a 525–μm *<100> p-substrate*. This process supports the fabrication of mixed-signal (analog/ digital) circuitry and no modification whatsoever has been introduced to the standard process line. Some remarks on this specific process as compared to other standard CMOS processes are presented in the following paragraphs.

For photodetection in the visible-spectrum range this conservative process is more suitable than sub-micron processes: its junctions in the bulk are deeper, the doping concentrations are lower and consequently the inherent quantum efficiency is higher (see section 3.3.5). Table 4.1 gives some approximate junction-depth values, depletion layer extensions to the *n*- and *p*-side of the junctions, the total depletion widths without reverse-bias voltage, the junction capacitances and the built-in potential across the junctions. The doping profiles across n^+*/p-epilayer* and p^+*/n-well/p-*

junction	depth ± 3σ [nm]	x_n [nm]	x_p [nm]	width [nm]	C_j ± 3σ [fF/μm²]	ϕ_{bi} [V]
n^+*/p-epilayer*	350 ± 50	<< x_p	510	510	0.19 ± 0.04	0.98
p^+*/n-well*	500 ± 50	205	<< x_n	205	0.57 ± 0.11	0.96
n-well/epilayer	~ 2900	90	890	980	0.078 ±0.016	0.66

Table 1: Junction depths, approximated depletion widths, junction capacitances and built-in voltages in the process used.

epilayer vertical sections are displayed in Figure 4.1. If a photogate is implemented in the process we used, it should be of the 'transfer-diffusion' type (section 3.3.2), because overlapped gates are not possible due to the absence of a second polysilicon layer.

The two metal layers available allow a reasonable system interconnection complexity, but a third metal layer would be useful to shield the circuitry from light. In all implemented structures, the full dielectric stack was left on top of the photo-sensitive elements. These layers are shown in Figure 4.2. Each layer is shown along with its nominal thickness and tolerance. One can define the location of gate oxide or local

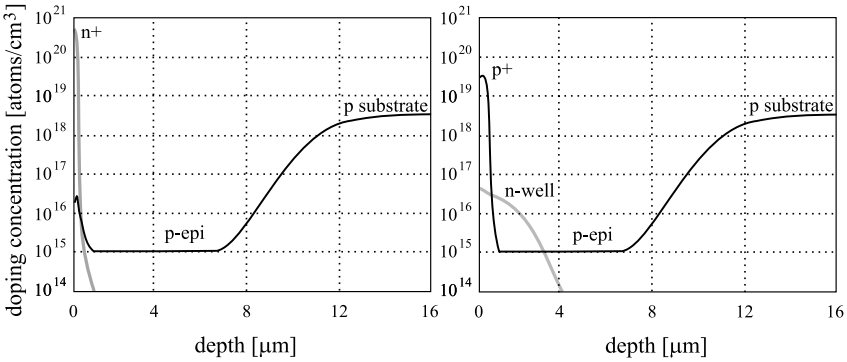

Figure 4.1 *Doping profiles across the n^+/p-epilayer/ substrate and p^+/n-well/p-epilayer/substrate at the end of the process*

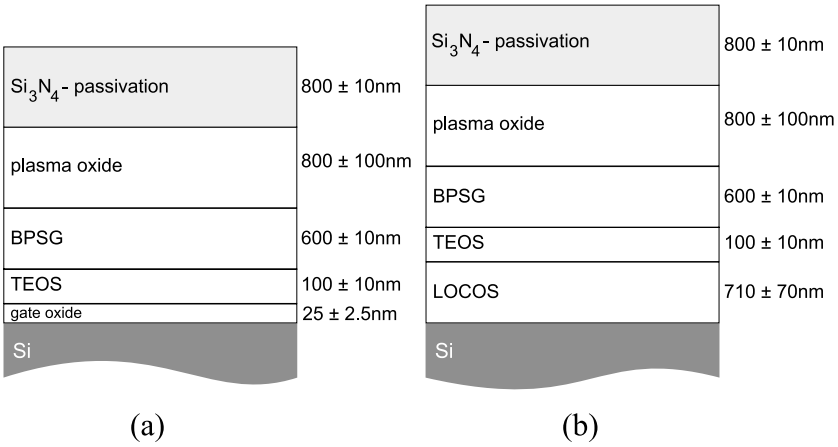

Figure 4.2 *Dielectric layers left on top of the photo-sensitive elements in regions where there is either (a) gate oxide or (b) LOCOS. (Si_3N_4 - silicon nitride, BPSG - Boron phosphosilicate glass, TEOS - Tetra ethylorthosilicate ($Si(OC_2H_5)_4$), LOCOS - local oxidation of silicon.)*

oxidation of silicon (LOCOS) by using an oxide mask. Several combinations of the dielectric layers are possible by applying other masks. Besides the oxide mask (OD)[1], standard masks that can be used are the metal-contact masks (CO and COS) and the bond pad-contact mask (CB).

When etching the dielectric layers, the metal layers are used as etch stops and the etch time does not need to be very accurately controlled. This might represent a problem if the process steps are kept standard, because on top of the photodetector active region there should not be any of these metal layers and when etching some of the dielectric layers one has no proper control over the thicknesses of the other layers left. Therefore, the repeatability of the spectral transmittance through the dielectric stack from one fabrication batch to another is compromised.

Figure 4.3 shows the etched layers for each etch mask and the respective approximate tolerances for the standard CMOS process at DIMES.

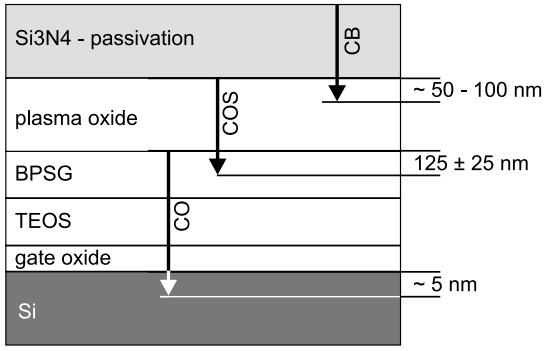

Figure 4.3 *Etched layers and tolerances when etch masks are applied (CO, COS and CB).*

4.2 Hartmann mask

For the proposed Hartmann wavefront sensor we used a mask with sub-apertures. At any stage, the mask can be substituted with a customized microlenses array with no modification to the CMOS detector.

1. the mask label depends on the foundry; definitions adopted by DIMES are used in this thesis.

To fabricate the mask, a 1-μm thick aluminum layer was first deposited on a 500-μm glass substrate and then dry etched (RIE[2]) to achieve the desired sub-aperture pattern and an accurate definition of the openings. The precision of the position of the sub-apertures is better than 1μm. It is a 1-cm^2 mask with an 8x8 orthogonal array of circular holes (ϕ =450μm) with a 1000-μm pitch, as seen in the photograph shown in Figure 4.4.

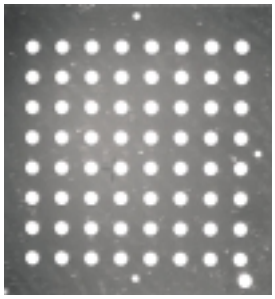

Figure 4.4 *Photograph of the manufactured Hartmann mask: aluminum pattern on a glass substrate.*

4.3 Postion-sensitive detectors (PSDs)

To achieve a fast Hartmann wavefront sensor, we replace the conventional imager by a matrix of optical position-sensitive detectors, so that the image processing step is circumvented. Some possible PSD structures were discussed in section 2.5. The 1D structures we fabricated were 4-, 6- and 10-stripe structures and a lateral-photoeffect photodiode (LEP). They were meant as starting points for the aimed 2D PSDs. The spot-position response of the stripe structures is very similar to that, in one coordinate, of the 2D devices implemented later, and we will only discuss the latter. The implemented 2D PSDs consist of three chessboard-like structures, with different photosensing elements, a juxtaposed spiral structure and a quad cell.

2. RIE - Reactive Ion Etching.

4.3.1 Lateral-photoeffect photodiode (LEP)

The theoretical advantage of this kind of PSD is the high response linearity, due to the 100-% fill factor, which is almost independent of the spot size if it remains within the boundaries of the device. In practice, however, the signal difference between two electrodes, placed at opposite edges of the device, is only significant if the sheet resistance R_{sq} of the LEP layer is large enough. It has been suggested that R_{sq} should be larger than 10KΩ/sq and very uniform for a good LEP [4.1]. In standard CMOS, R_{sq} is much lower than this. We used the *n-well* as the sensitive layer because it has the largest sheet resistance (R_{sq} = 600Ω/sq). This value, however, is not well controlled and the spread is significant for a large area. This layer was pinched with a p^+ implantation on top to increase the sheet resistance, however the spread on the value remains. The structure is 3 x 3 mm and is depicted in Figure 4.5.

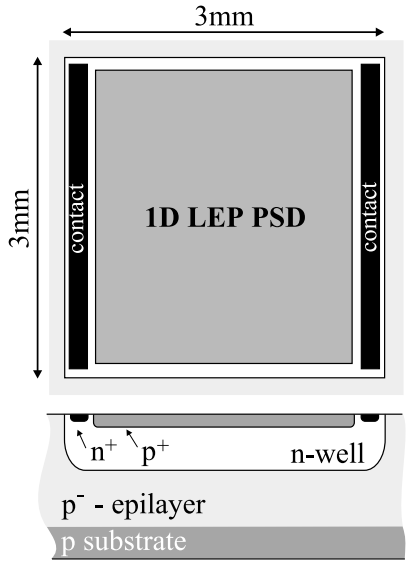

Figure 4.5 *One-dimensional lateral-effect photodiode with n-well as the resistive layer.*

We scanned the surface with a 5-μW spot ($\phi \sim 800\mu m$) and the output signal could not be distinguished from noise. Other layers, such as n^+ and p^+, are even less practical because they have lower sheet resistances, R_{sq}=55Ω/sq and R_{sq}=75Ω/sq, respectively. When the spot intensity cannot be increased, a possible solution to this problem is to modulate the input signal and to use a filter to read the LEP response.

4.3.2 Multipixel structures

Four 5x5-pixel 2D structures and a quad cell (QC) were implemented. Their responses were simulated and compared to the experimental results for several spot diameters. All structures, except the quad cell, are connected to two linear resistive arrays (1-KΩ resistors), as explained in section 2.5.4, and their characteristics are given in Table 2.

structure		chessboard-like (5x5)			spiral (5x5)
photosensitive element		photodiode	photodiode	phototransistor	photodiode
layout		single n-well	multiple n-wells	vertical pnp	single n-well
structure lateral size [μm]		2120x2040	2196x2116	2056x2056	2064x2064
cell lateral size [μm]		400 x 400	400 x 400	400 x 400	390 x 390
*pixel size** [μm]		400 x 400	400 x 400	400 x 400	340 x 340
pitch x direction [μm]		860	898	828	419
pitch y direction [μm]		820	858	828	419
X coordinate array	*# photo elements*	13	13	13	25
	fill factor [%]	48	45	49	41
Y coordinate array	*# photo elements*	12	12	12	25
	fill factor [%]	44	41	45	41

Table 2: Characteristics of the multipixel PSD structures implemented in the framework of the standard CMOS process available at DIMES.

*. See Figure 4.9.

Chessboard-like PSDs

Three chessboard-like PSD structures (CBRD) with different photosensing elements are addressed. All have 25 cells, 13 of which contribute to sensing the spot displacement along the *x* direction and the other 12 cells contribute

to sensing the displacement along the *y* direction. The chessboard-like layout can be seen as two independent arrays spatially interlaced. The resistive array is integrated on chip and consists of non-silicided polysilicon resistors[3].

To study the influence of the structure layout on the device response we built a geometrical model of the structure. When simulating the response of the chessboard-like structure we scanned its surface with a spot function and calculated the contribution of each cell to the signal at both ends of an ideal resistive array. The approach was purely geometrical and did not take into account quantum efficiency, noise or crosstalk between cells. Figure 4.6 shows the *x-coordinate* normalized response of the chessboard-like structure for four different effective radii of spots with a uniform circular intensity profile. The radii values are normalized to the pixel lateral size. This profile is used throughout the simulations, unless indicated otherwise.

When the spot is smaller than a cell, the response is step-like, which does not provide accurate information on the spot-centroid position. A very small spot results in a PSD response imprinted with the geometrical arrangement of the pixels - compare Figure 4.6 for a spot with r = 0.1 cell with Figure 4.7. A somewhat larger spot, comparable to the cell size, results in a smoother response, but still limits the accuracy because of the cell-size gaps between two *x*-sensing cells, which is due to the presence of the complementary *y*-sensing cells. A smooth and more accurate response can be obtained for spots equal to or larger than two pixels. However, in practice, the larger the spot, the closer the magnitude of the signals at opposite ends of the resistive array and the shorter the total signal range. Therefore, if the spot is too large, the differential signal (equation (2.53)) becomes comparable to the photo-detection noise level and the position resolution becomes very poor.

The three implemented structures have the same geometrical layout but different photosensitive-cell structures, as described below:

- *structure 1*: single n-well photodiodes (SNW). The p^+*/n-well* junction is used as the active photodiode. The p^+ implants define the square cells and they all share a common *n-well*;

3. The sheet resistance of the polysilicon layer is R_{sq}=25±6Ω/sq.

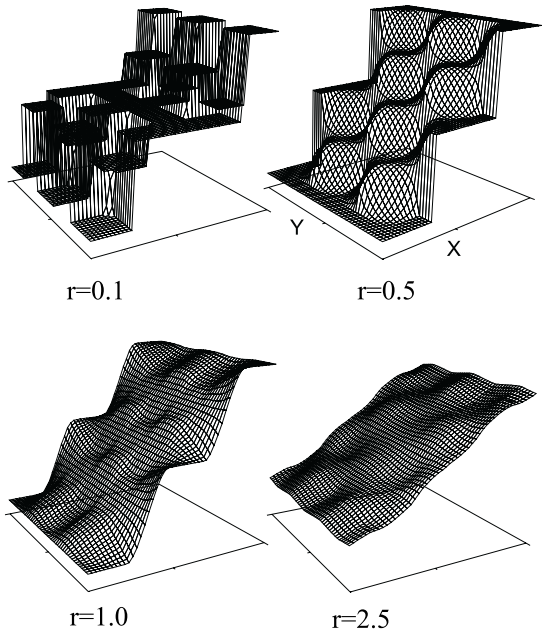

Figure 4.6 *Response of the chessboard-like structure, in the x direction, for a uniform circular profile of the spot intensity with four different radii, normalized to the pixel lateral size.*

- *structure 2*: multiple n-wells photodiodes (MNW). Again the p^+*/n-well* junction is used as the active photodiode, however, in this structure each cell consists of a p^+ implant inside an *n-well* that defines the cell boundaries. It imparts a somewhat larger separation between cells, but is less sensitive to crosstalk;
- *structure 3*: vertical pnp phototransistor (PTr). The phototransistor-based PSD is made with cells that have a p^+*/n-well/p-epilayer* structure. Multiple *n-wells* are necessary. The emitter (p^+ implant) is defined in the center of each cell and represents no more than 4% of the cell area. It is covered with two metal layers. In this architecture, all collectors (*p-epilayer*) are compulsorily connected together. To minimize lateral crosstalk, we designed a guard ring around each pixel. Gain

variations between different pixels is expected, because the *n-well*, which is actually the transistor base, is not as highly homogeneous as in the case of a true bipolar transistor.

A photograph of one of the implemented chessboard-like structures is shown in Figure 4.7. The picture was retouched to enhance contrast and to enable visualization of the two interlaced photodetector arrays: one array to sense the *x* displacement and the other to sense the *y* displacement of the spot.

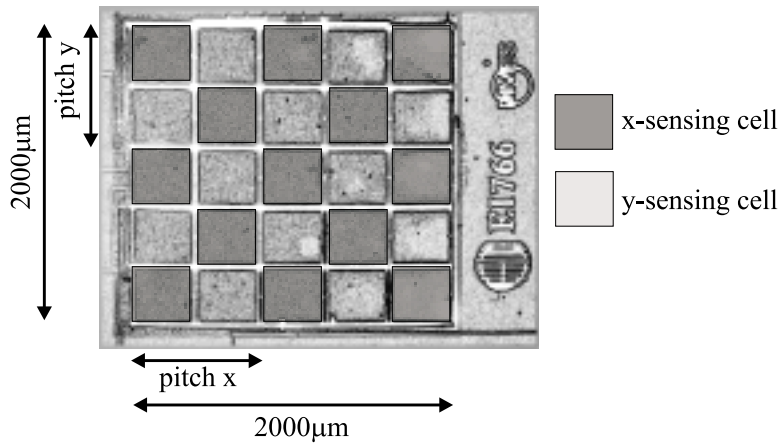

Figure 4.7 *Photograph of the implemented single n-well chessboard-like PSD.*

Spiral PSD

One spiral PSD structure was implemented. The structure consists of 25 cells, each with two spiral-like juxtaposed photodiodes. One photodiode in a cell belongs to the *x*-coordinate sensing array and the other photodiode belongs to the *y*-coordinate sensing array. The resistive array was made external to the chip and consists of commercial 1-KΩ resistors with a nominal-value variation of 1%. The simulation conditions to the response of this geometrical structure are the same as for the chessboard-like structure. Figure 4.8 shows the *x*-coordinate response when we scan the spiral PSD with different spot sizes. For very small spots, the device

response reveals the discrete (pixelated) nature of the structure. Larger spots result in a smoother and more linear response.

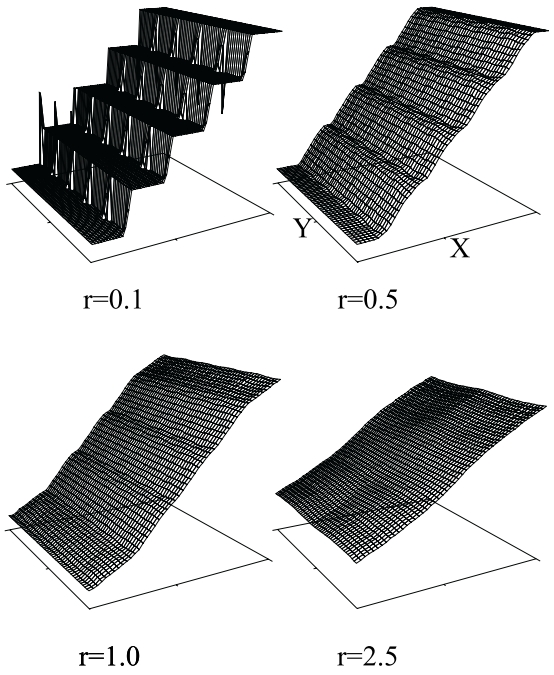

Figure 4.8 *Response of the spiral PSD structure, in the x direction, for a uniform circular profile of the spot intensity with four different effective radii, normalized to the pixel lateral size.*

The photodiode junction used for the spiral PSD is the p^+*/n-well* shallow junction. All photodiodes in the array are inside a single *n-well* to yield the highest possible fill factor. The gap between the two juxtaposed photodiodes inside a cell is covered with a metal layer to prevent electron-hole-pair generation at the edges of the photodiodes and therefore minimizes crosstalk. Around each cell there is an n^+ ring to ensure a uniform potential in the *n-well* over the array. In this layout, the number of

photodetectors per *coordinate* array is nearly twice as high as in the case of the chessboard-like layout. Moreover, the fill factor of the spiral PSD is only slightly lower than those of the chessboard-like PSDs. Figure 4.9 shows a photograph of the spiral PSD structure that we realized at DIMES.

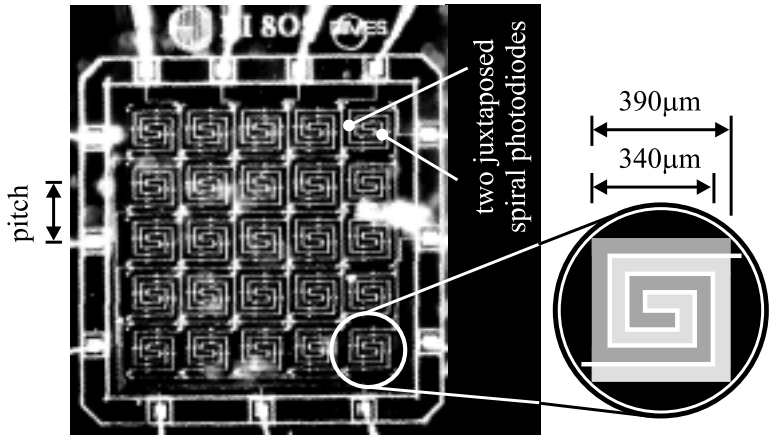

Figure 4.9 *Photograph of the spiral PSD. Each cell has two juxtaposed photodiodes.*

Results of the chessboard-like and spiral PSDs

The chessboard-like (CBRD) and spiral PSD structures were tested with different spot sizes. We used non-modulated light and different effective light-spot diameters to scan the surface, ranging from 1 to 5 times the cell size with a 5-μW intensity and a 0.63-μm wavelength. The measured responses are compared to the simulations. In all cases we are interested in the linearity of the response when the spot moves along one direction. The measured mismatch between the responses along the *x* and *y* coordinates was less than 10% for the chessboard-like structures and less than 4% for the spiral structure.

The responses of the chessboard structures for a spot moving along the *x* direction are displayed in Figure 4.10. Both the simulated response and those for the three chessboard PSDs with different photodetectors are shown. For an effective spot size comparable to the pixel size, the

experimental results confirmed the expectation that the chessboard structures have a step-like response, to which no linearity at all can be attributed. As the spot size increases, it appears that the response linearity keeps improving, but this is not quite the case, as discussed later after Figure 4.11.

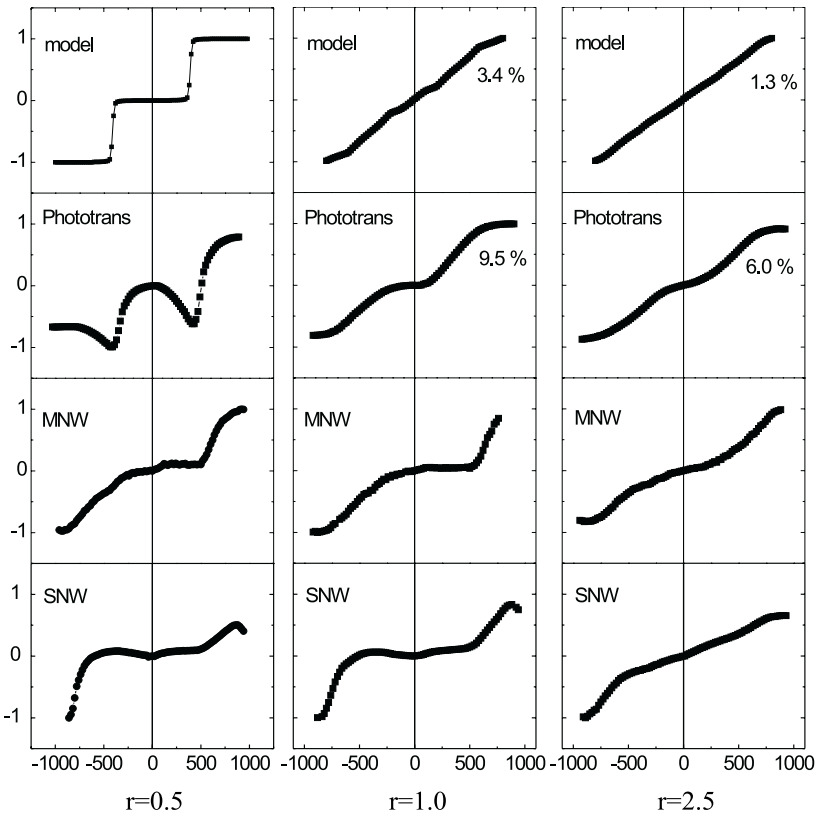

Figure 4.10 *Normalized position response of 5x5-element chessboard-like structures for different spot radii: a) 0.5 pixel, b) 1 pixel and c) 2.5 pixels. The non-linearity of the simulated and phototransistor PSDs are indicated.*

We observed that the position signal-to-noise ratio (SNR_P) is comparable for the three chessboard PSDs, however, the linearity of the PSD responses differs considerably. This cannot be directly attributed to the type of photodetector used, but rather to the significant spatial variations of the polysilicon-array resistance. Resistance variations across the array induce an asymmetrical position response with respect to the center of the array. Not only the zero response can be shifted from the geometrical center, but also the response slope can vary across the array. The responses of the three PSDs were extracted from structures on the same chip.

Figure 4.11 shows the graph of the response non-linearity versus the effective spot size for the modelled chessboard PSD. It is interesting to note that the non-linearity at first decreases, then increases and then decreases again as the spot size is made larger. This can be explained by different

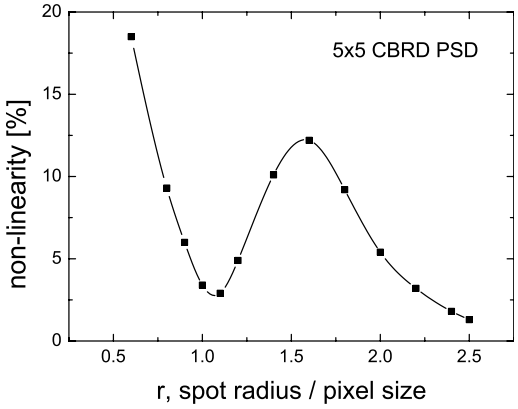

Figure 4.11 *Response non-linearity of the chessboard-like PSDs for various spot sizes.*

factors: the discrete nature of the structure and the finite size of the detector and of the spot. In the case of small spots, the linearity is dominantly limited by the pixelated characteristic of the structure, namely by the pixel size and pitch. Also, in the case of a small spot, the response varies more significantly between two cross sections along the PSD, that is, the coordinate interdependence is more critical. Figure 4.12 shows the PSD

response for a small spot and large spot along two different chessboard sections.

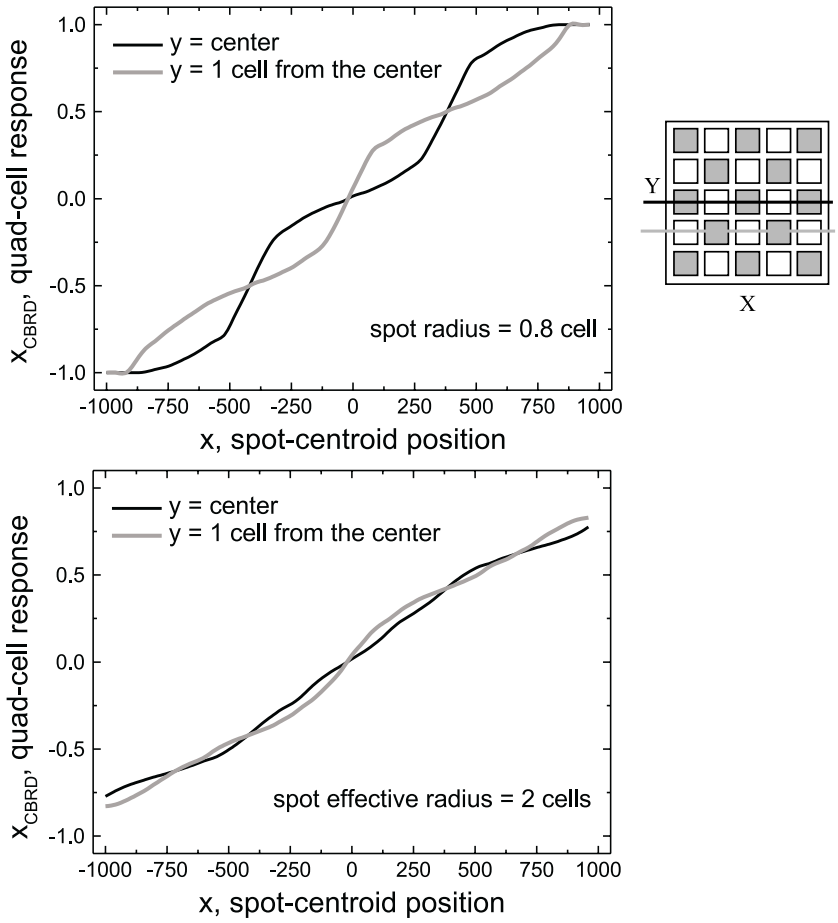

Figure 4.12 *Chessboard response along the x direction for a small spot and a large spot at different y positions, as indicated*

Also interesting to point out is that contrary to an initial guess, it was found that electrical crosstalk between neighboring pixels smoothens the PSD position response for small spot sizes. For the PSD, lateral crosstalk is interpreted as a larger spot. In the case of large spots, comparable to the PSD size, the linearity is mainly affected by edge effects[4]. Hence, an ideal spot should be just as large as the pixel pitch, yielding a large spatial dynamic range $||DR||_S$. Spots larger than the PSD theoretically improve the linearity, but in practice, they are not desirable because they increase crosstalk with neighboring PSDs in a Hartmann sensor, and because the larger the spot, the more difficult it is to detect the signal difference between the two end points of the resistive array, i.e., the total response range decreases as the spot size increases (Figure 4.13). In any case, the best solution to linearity would be a higher density of pixels, that is, if the resistor array did not represent a limitation. But the more resistors, the more imbalance - due to the sum of the individual tolerances of the resistors -, the larger the total area occupied by them, and the higher the thermal noise associated with them.

The absence of a consistent and predictable linearity, mainly because of poor resistance tolerances, rendered the chessboard approach with integrated polysilicon resistors unattractive as a PSD to be used in a matrix for a Hartmann sensor.

For the same number of *cells* and the same device size, the spiral PSD offers a higher *pixel* density than the chessboard PSD. The modelled and measured spiral PSD responses for a spot moving along the *x* direction are displayed in Figure 4.14. In contrast to the chessboard PSDs, for an effective spot size comparable to the pixel size, the spot-position response of the spiral structure features a reasonably good linearity. The high symmetry of the response is accomplished by the high precision of the external resistors used, in contrast to the poor precision of the integrated resistors used in the chessboard case. Figure 4.15 shows how the spiral-PSD non-linearity changes with the spot radius. This graph confirms the statement that the most appropriate spot size should be about as large as the pixel pitch.

4. Edge effects occur when part of the effective beam is out of the PSD surface and does not contribute to the signal anymore. The spot-centroid position is then estimated from an asymmetric (degenerated) spot.

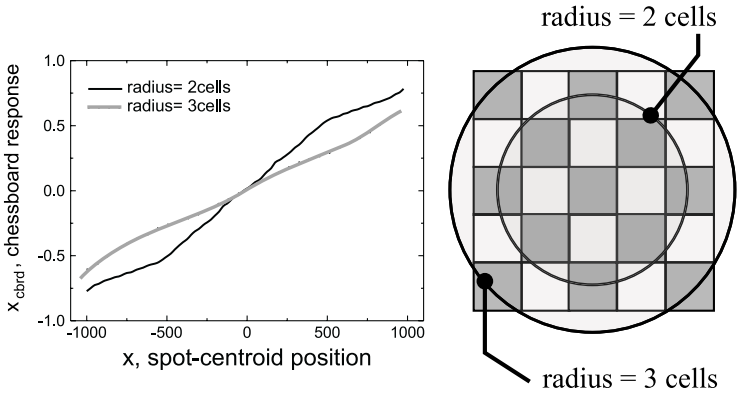

Figure 4.13 *Non-normalized simulated chessboard response for a spot four times and six times larger than the cell respectively. The larger the spot, the lower the signal range.*

We simulated the response linearity of the chessboard and the spiral PSDs for structures with more cells in the *x* direction than the structures we had implemented. A 1-pixel effective radius was used. The conclusion was that the linearity, due to the geometrical layout, has improved with a further increase in the number of cells and reached a saturation value of 99.7% for the chessboard structure and 99.8% for the spiral structure when 20 or more cells were used in one direction. In other words, the linearity improves until the total number of cells equals 400; further increase in the number of cells does not contribute to a better linearity. This limit cannot be further improved because there is always a minimum 'ripple' in the signal due to the discrete structure of the PSD. Therefore, for each expected spot-intensity profile, an optimal number of cells can be found, such that the geometrical contribution to the PSD response non-linearity is minimized.

An integrated matrix of spiral PSDs to be used as a Hartmann sensor demands that the resistors are integrated, which means that the response of each PSD is subjected to resistor imbalance - because of the poor tolerances of the resitive layers -, and the response of a single PSD cannot be taken as the representative of the whole matrix; each PSD might have a different

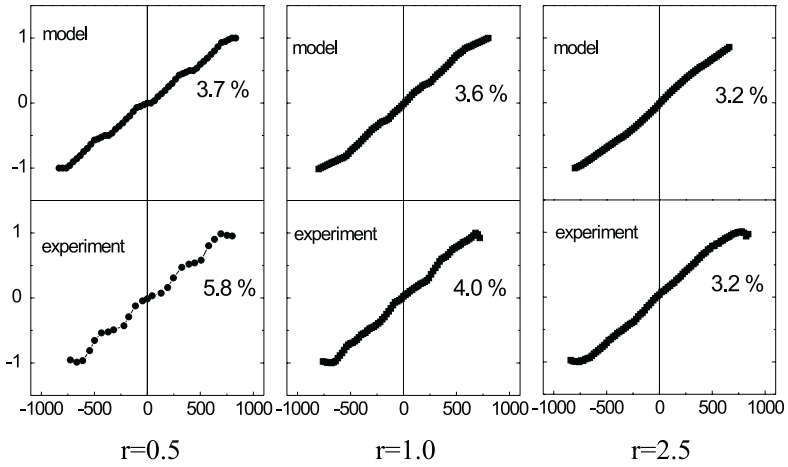

Figure 4.14 *Normalized position response of 5x5-element spiral structure for different spot radii: (a) 0.5 pixel, b) 1 pixel and c) 2.5 pixels. The non-linearity of the simulated and measured PSDs are indicated.*

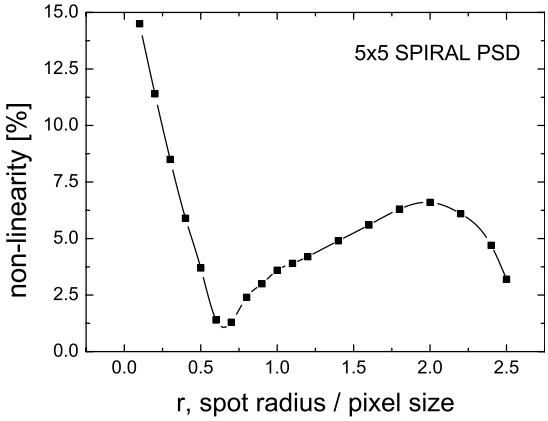

Figure 4.15 *Response non-linearity of the spiral PSD for various spot sizes.*

response curve. If a look-up table is made in a pre-calibration step, it ought to be available for every PSD. Besides, offsets due to pixel mismatch in a PSD cannot easily be extracted in the current-division read-out scheme. A way to minimize the resistor imbalance is possibly the implementation of an array of active resistors, namely MOS transistors operating in the ohmic region. The MOS transistor resistance can be adjusted by design (*W/L ratio*) and by the applied gate-to-source voltage [4.2].

Quad cell

The quad cell (QC) is a two-dimensional position-sensitive detector[5] with a very well-defined geometrical center, simple layout and reduced number of output nodes. Its simple layout gives more flexibility to lateral scaling and guarantees a higher fill factor than for the chessboard-like and spiral PSD versions. The presence of only four outputs per quad cell renders the use of resistive arrays unnecessary, which reduces the contribution of thermal noise and guarantees a more symmetric response.

We implemented a quad cell with a 600 x 600 μm cross section. It is based on double-junction photodiodes as photosensitive elements, as depicted in Figure 4.16. Each double-junction photodiode has a parallel

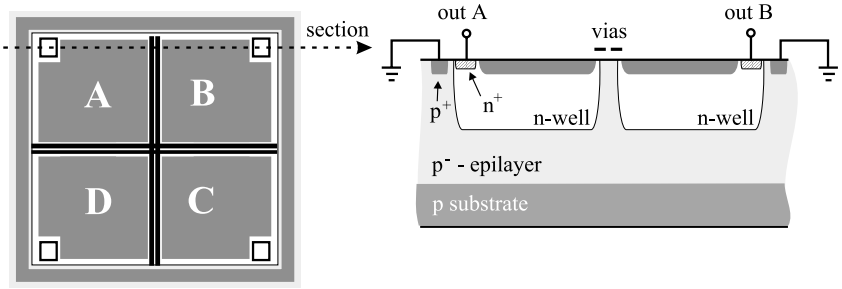

Figure 4.16 *Layout of a quad cell with double-junction photodiodes.*

shunt resistance R_{sh}= 3 GΩ, estimated as suggested in 2.4.4, and a series resistance, R_s, that lies between 5 Ω and 40 Ω. This value depends on the

5. The quad-cell PSD is sometimes referred to as a quadrant detector.

n-well sheet resistance, which is not very well controlled. The 13-μm gap between the cells is covered with a metal layer to avoid crosstalk between the photodiodes.

The response of a quad cell is only linear over the whole range if the spot features a uniform square profile. For uniform circular spots the response in only linear at the very central region of the quad cell. The response linearity depends on the spot intensity profile and on its effective radius. Figure 4.17 shows the quad-cell position response along the x direction for various sizes of a spot with a uniform circular profile. The

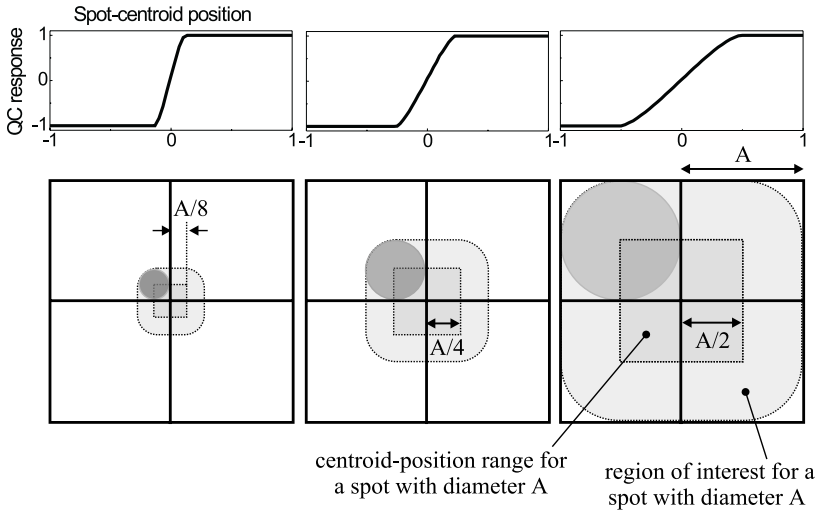

Figure 4.17 *Quad-cell position response in the x direction for different sizes of a spot with a uniform circular intensity profile.*

spatial dynamic range of a quad cell, $||DR||_S$, for a given circular spot, is equal to the spot diameter if the gap between cells is negligible. Small spots can be used for center tracking, provided that they are larger than the gap between cells, and larger spots are necessary when one wants to acquire centroid displacement information, as is the case with the Hartmann sensor.

There are some rules of thumb that must be checked for optimal position

sensing: the effective spot radius should not be larger than the lateral size of one cell of the detector, the smaller the spot the better the resolution around the quad-cell center and the shorter the spatial dynamic range. For spots with a uniform intensity distribution larger than the quad-cell total size, the response curve does not represent the spot-centroid position, as seen in Figure 4.18. We observe that the best linearity and maximum range

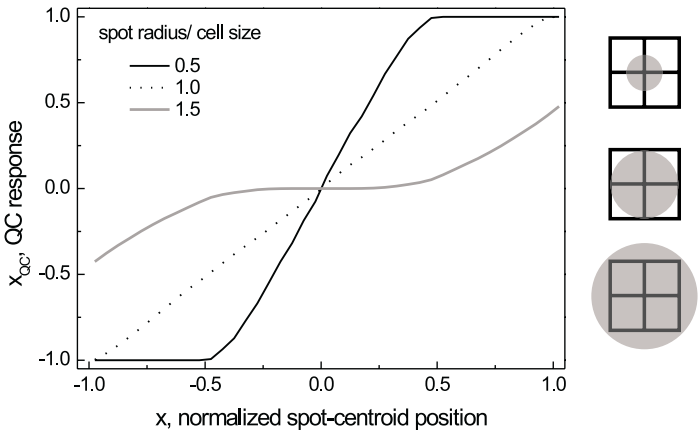

Figure 4.18 *Quad-cell response usefulness. Only spots with effective radii smaller or equal to a cell lateral size convey useful position information.*

is achieved for a spot as large as the quad cell. However, to minimize optical crosstalk between neighboring PSDs, in the Hartmann sensor, and to maximize the he amount of light impinging on each quad cell, it is convenient to limit the maximum size of the spot to half the quad-cell lateral size. This size for the circular spot yields the maximum centroid-displacement range for which the spot remains completely on the quad-cell surface, i.e., it offers a good balance between position resolution, range and photodetection, besides limiting the possibility for crosstalk. A non-truncated spot profile yields a larger centroid-position range than a truncated spot, but also leads to higher optical crosstalk for quad cells located close to each other.

Although the quad-cell response for a circular spot is not linear, it is possible to find an analytical function to approximate the response. A convenient approximation to the response curve is the normalized Boltzmann-sigmoidal function, in which the parameter s_x determines the slope of the sigmoidal curve, as given by equation (4.1):

$$x_{QC} = A_2 + \frac{(A_1 - A_2)}{1 + \exp(x/s_x)}, \tag{4.1}$$

where, for a spot with well-defined boundaries and whose maximum size is equal to the quad-cell lateral size, A_1=-1 and A_2=1. Whereas a spot larger than the quad-cell size or with non-truncated edges yields $|A_1|$ and $|A_2|$ less than unity. An equivalent expression holds for y_{QC}.
For the same spot effective radius, *r*, but different spot profiles, the response curve changes. The expressions for some intensity profiles, *I(x,y)*, with centroids located at coordinates (x_c, y_c) are given from equation (4.2) to equation (4.6).
Uniform square spot:

$$I(x,y) = \begin{cases} 1, & -r \le (x - x_c) \le r \quad \text{and} \quad -r \le (y - y_c) \le r \\ 0, & |x - x_c| > r \quad \text{and} \quad |y - y_c| > r \end{cases}. \tag{4.2}$$

Uniform circular spot:

$$I(x,y) = \begin{cases} 1, & (x - x_c)^2 + (y - y_c)^2 \le r^2 \\ 0, & (x - x_c)^2 + (y - y_c)^2 > r^2 \end{cases}. \tag{4.3}$$

Gaussian spot:

$$I(x,y) = e^{-[(x - x_c)^2 + (y - y_c)^2]/\sqrt{2}r^2}. \tag{4.4}$$

Sinc^2-circular spot[6]:

6. The *sinc(x)* function is defined as *sin(x)/x*.

$$I(x,y) = \left(\frac{\sin\left((1/r)\sqrt{(x-x_c)^2 + (y-y_c)^2} \right)}{(1/r)\sqrt{(x-x_c)^2 + (y-y_c)^2}} \right)^2 . \quad (4.5)$$

Sinc^2-rectangular spot:

$$I(x,y) = \left(\frac{\sin((x-x_c)/r)\sin((y-y_c)/r)}{(1/r)^2 (x-x_c)(y-y_c)} \right)^2 . \quad (4.6)$$

Figure 4.19 shows different spot intensity profiles and Figure 4.20 shows the quad-cell response for each of them, where the effective radii are equal to half a cell size. The most linear response is achieved with the

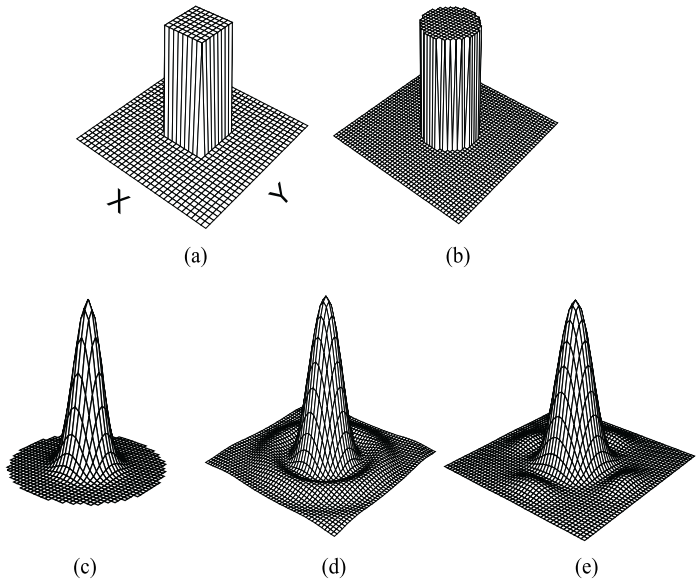

Figure 4.19 *Spot intensity profiles: a) uniform square b) uniform circular, c) Gaussian, d) Sinc^2 circular and e) Sinc^2 rectangular.*

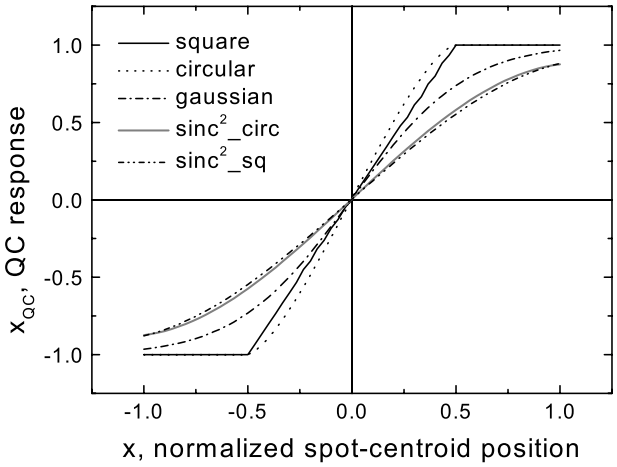

Figure 4.20 *Quad-cell response for different spot intensity profiles with effective radii equal to half a cell lateral size.*

uniform square profile. The value of the sigmoidal parameter s_x as a function of the effective radius for various spot profiles is displayed in Figure 4.21. A uniform square profile can be approximated better by a linear fit and is not included in this graph.

For an arbitrary spot profile, the quad cell can be calibrated by scanning the spot on its surface and then fitting the most appropriate mathematical function. If the intensity profile remains constant, the response can be very accurate over the whole displacement range. Figure 4.22 shows the response of the implemented quad cell for the indicated spot profile. The spot is radially symmetric, its intensity is 7 μW, and the light is not modulated. For this response the best fitting sigmoidal parameter, s_x, normalized to the cell size is 0.22 (~ 66 μm), with $|A1|=|A2|=0.96$.

Noise at the edges of the response curve is higher because when the centroid of the spot is at one edge, the opposite photodiodes are scarcely illuminated and contribute to higher detection noise than when the spot is perfectly centralized. Therefore, if the whole range is considered, i.e., the interval [-300μm, 300μm], the position signal-to-noise ratio, SNR_p, is limited because of the noise values at the edges and can be calculated by means of equation (2.50). For the response presented in Figure 4.22, SNR_p

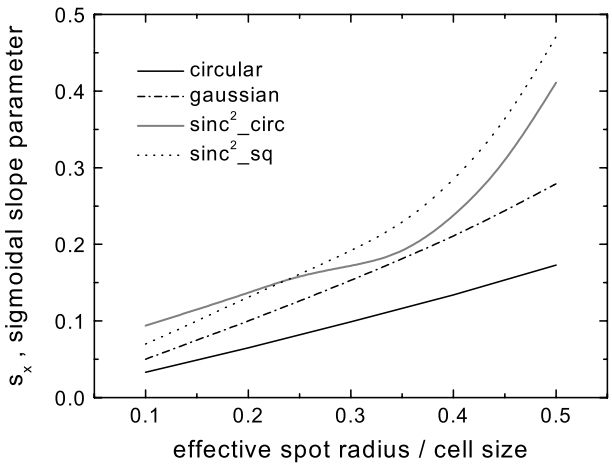

Figure 4.21 *Sigmoidal slope parameter, normalized to the cell size, as a function of the effective spot radius for various spot intensity profiles.*

is equal to 26 dB. For the more central interval [-150μm, 150μm] the SNR_p is close to 26.5 dB. In a quad cell the offset values of each pixel are easily subtracted and the response curve of one quad cell can be considered as representative of the response of any other quad cell in a matrix under the same conditions. Figure 4.23 shows the error associated with a linear fit and sigmoidal fit to the response over the whole range and in the central region of the quad cell. Over the full range, the *rms* deviations between the response and fitted curves are 15% and 1%, for the linear and sigmoidal fits, respectively. When only the central region is fitted, the deviations are 2.6% and 0.6%, for the linear and sigmoidal fits, respectively.

4.4 Wavefront sensor

This first prototype of a CMOS-based wavefront sensor consists of a Hartmann mask with 8x8 circular sub-aperture (ϕ =450μm) with a 1000-μm pitch positioned at 6 cm from the detector plane. The detector is a custom 1-cm^2 CMOS chip, with an orthogonal matrix of 600 x 600 μm quad cells (1000-μm pitch), a distributed demultiplexer and a sampling-

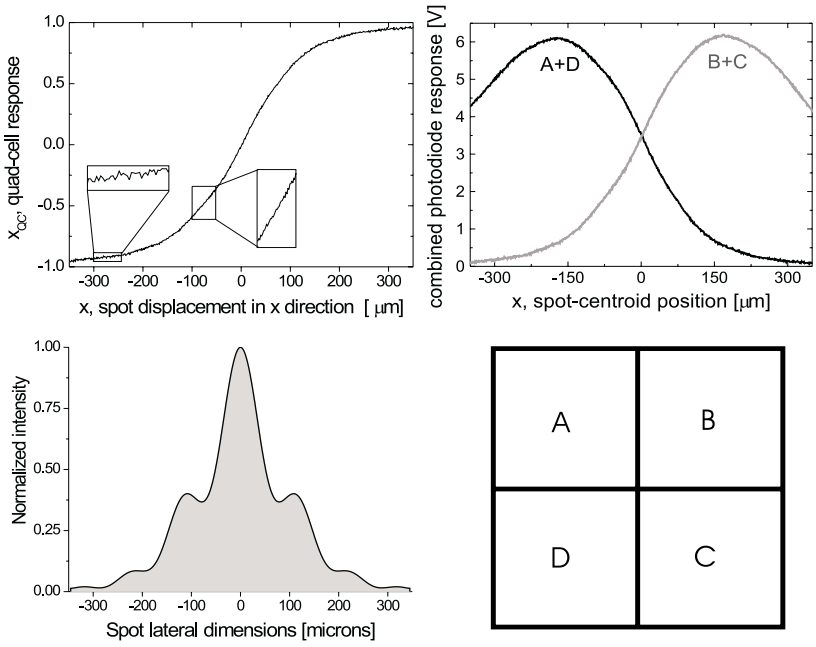

Figure 4.22 *Response of the implemented quad cell in the x direction for a 7-μW spot with the spot profile shown. The combined photodiode response is also displayed.*

rate controller. The chip is mounted on a Printed-Circuit Board (PCB), where signal amplification takes place. The PCB is interfaced to the computer via a commercial data-acquisition board. A diagram illustrating the concept is shown in Figure 4.24. This section discusses the implementation of the various components present on chip and briefly covers the interface to the computer.

4.4.1 The quad-cell pixel

Each quad cell (QC) consists of four double-junction photodiodes. To minimize the number of outputs from the chip, we implemented an analog demultiplexer integrated on chip, such that selected photodiode signals are

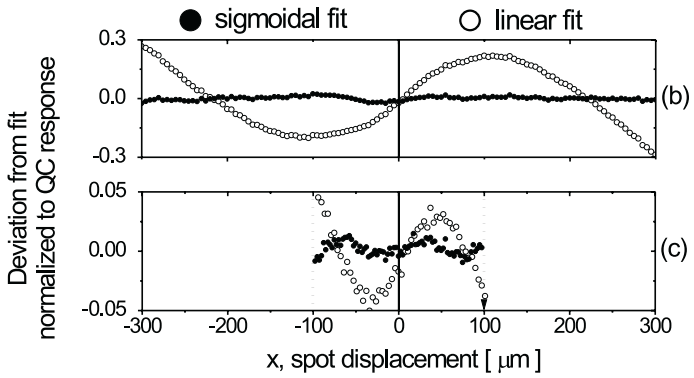

Figure 4.23 *Error associated with the sigmoidal fitting and linear fitting in the whole range and central region of the quad cell.*

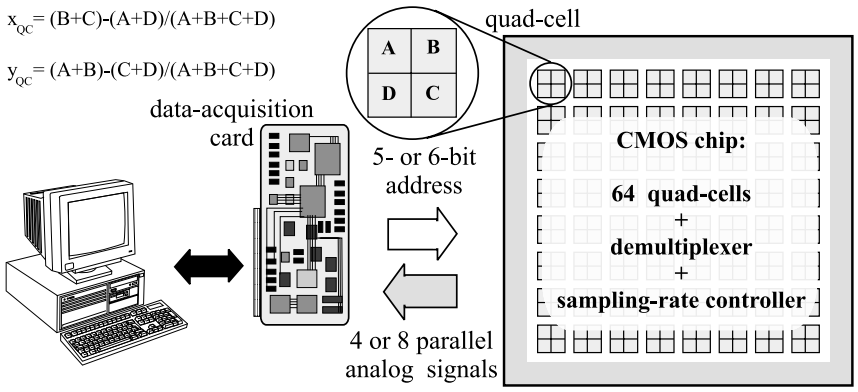

Figure 4.24 *Diagram of the data flow for the implemented Hartmann wavefront sensor.*

available sequentially at the parallel output lines. For this purpose, each pixel must consist of a photodiode provided with a switch circuit. Two main pixel architectures could be used: an active-pixel and a passive-pixel architecture. Quad-cell test structures with both architectures were fabricated.

Figure 4.25 shows the schematics for each of these pixels. The active pixel is basically the same as that of Figure 3.6a, where the pixel address transistor is substituted with a complementary switch set with attached dummy switches. An inverter is also present in every pixel to operate the complementary switch set and an extra block is added to allow pixel discharge.

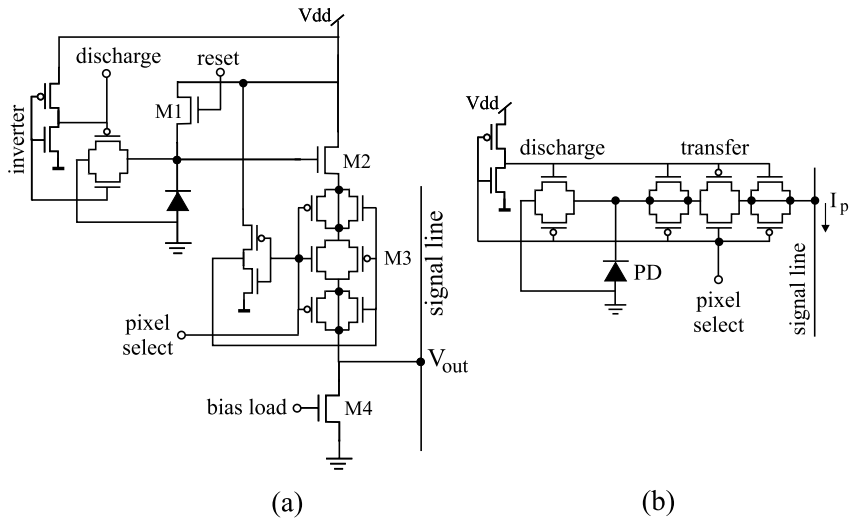

Figure 4.25 *Schematics of a) active-pixel and b) passive-pixel architectures of implemented test quad cells.*

The NMOS/PMOS complementary switches were used to increase the signal-voltage swing and the two attached ‘dummy’ switches favor charge redistribution, attenuating signal oscillation when M3 is turned on. The discharge switch is not relevant to pixel operation and was only included for test reasons. The passive pixel is similar to the one in Figure 3.5, where the single address switch is replaced by a complementary switch set with an inverter. A discharge switch is also added, such that whenever the address switch is off, the pixel is discharging (continuous offline discharge). In both architectures the pixel circuitry was interconnected by

means of the first metal layer and shielded from direct light incidence by means of the second metal layer. A simplified schematics of a quad cell is illustrated in Figure 4.26.

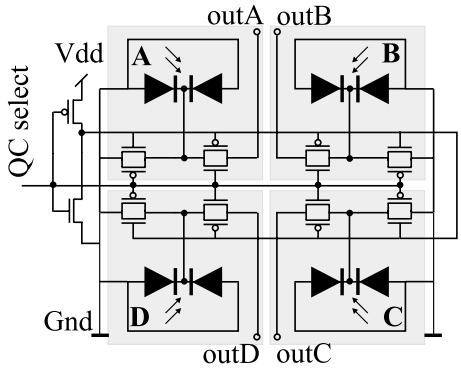

Figure 4.26 *Quad-cell schematics with double-junction photodiodes and switch sets.*

The normalized quad-cell response for quad cells with either type of architecture were equivalent. The non-logarithmic active pixels work in integration mode and although spatial random access is possible, temporal random access is not. The passive-pixel layout with continuous offline discharge enables true random access in both space and time because the pixels only effectively start transferring their photocurrents when they are addressed. However, a minimum boundary to the pixel access time is imposed either by the pixel capacitance or by the switches, which have a settling time. A logarithmic active pixel has not been considered at this first evaluation step because of the difficulty in extracting its fixed-pattern noise and because of its low output-voltage swing.

The non-normalized response of the active-pixel quad cell presented a somewhat higher non-linearity at the edges because of extraordinary behavior of the buffer transistor M2 induced by the change in illumination at its surroundings, but this was readily overcome by normalization. We chose to use the passive-pixel type of quad cell for the first prototype of the Hartmann wavefront sensor because it is simpler to implement and requires

less control lines than the active-pixel counterpart. Figure 4.27 shows a micro-photograph of one passive-pixel quad cell integrated on the wavefront sensor chip.

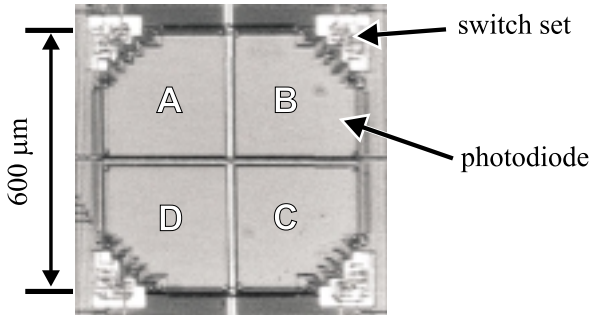

Figure 4.27 *Micro-photograph of one of the 64 passive-pixel quad cells integrated on chip. Each of the four pixels in the quad cell consists of a double-junction photodiode and an attached switch set.*

4.4.2 The CMOS chip

The CMOS chip was fabricated at DIMES with the 1.6-µm standard process. It comprises 256 passive pixels clustered four-by-four, so that sixty-four spatially isolated quad cells are formed. There is a 380-µm gap between quad cells, such that electrical crosstalk between neighboring quad cells is minimized. The pixels can be randomly addressed both in space and time. Figure 4.28 shows the different circuit blocks integrated on chip.

The chip was designed in such a way that all four pixels from a certain quad cell are addressed simultaneously. This minimizes the effect of temporal artifacts on the light-spot position information. That is, for a given quad cell, if the signal of pixel B is collected at a time τ after the signal of pixel A is collected, and if the spot moves meanwhile, the accuracy of the registered position will be affected. The quad cells are divided into two access groups: inner and outer quad cells, which can either be addressed independently or simultaneously, i.e., uncoupled or coupled control modes, respectively. The choice depends on the number of parallel outputs aimed

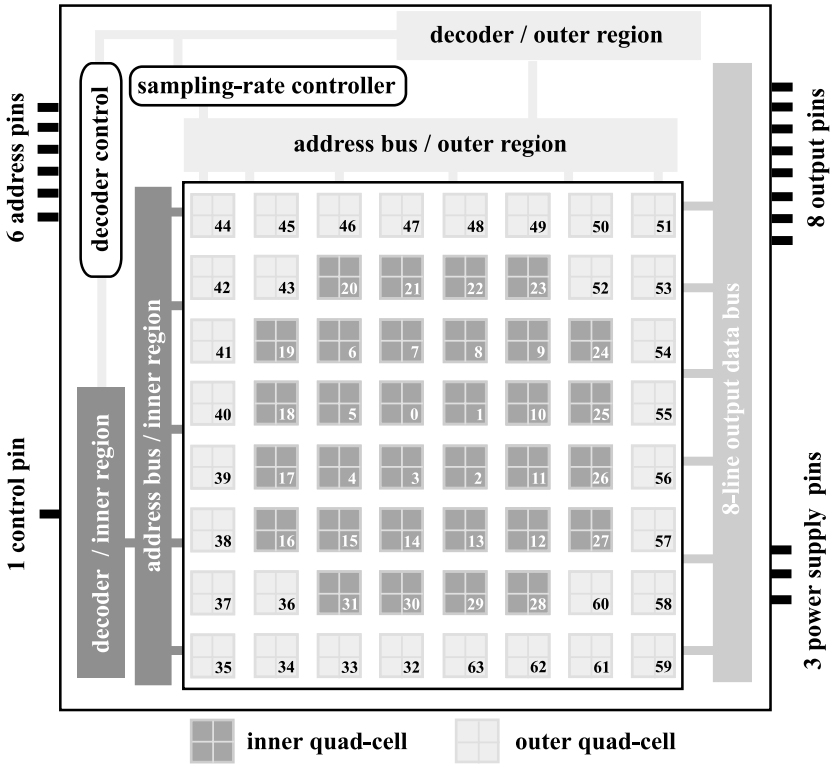

Figure 4.28 *Block layout of structures on chip.*

and determines the read-out rate. For the former case, the demultiplexer needs a 6-bit word as input address and in the latter case a 5-bit word (Figure 4.29). Depending on the digital level of the control pin, the sampling-rate controller determines whether one or two quad cells are addressed at a time. The distributed demultiplexer consists of several digital and analog units. The decoder control decodes the input address and operates either the inner-region decoder or both. The decoders control the analog switch set in each pixel. An 8-line output data bus either carries 4 or 8 signals at a time, depending on the control mode. A photograph of the implemented wavefront sensor, with a detail of the upper-right corner is shown in Figure 4.30.

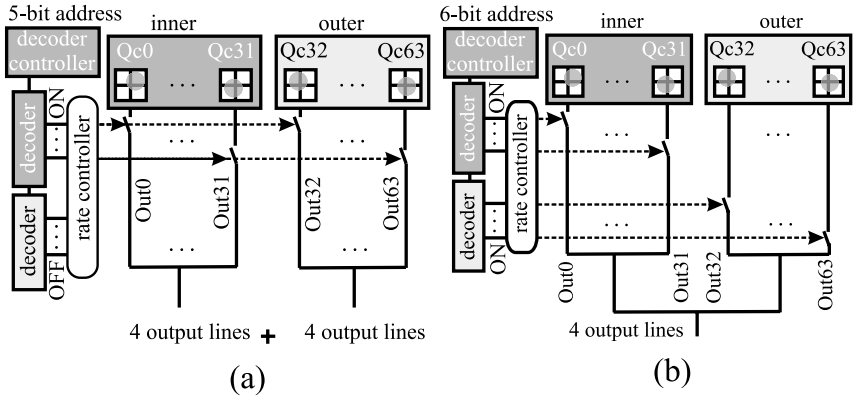

Figure 4.29 *(a) Coupled and (b) uncoupled control modes. Each drawn line from a given quad cell represents four parallel signal lines been addressed simultaneously. The dotted lines are the digital control lines to address the analog switches.*

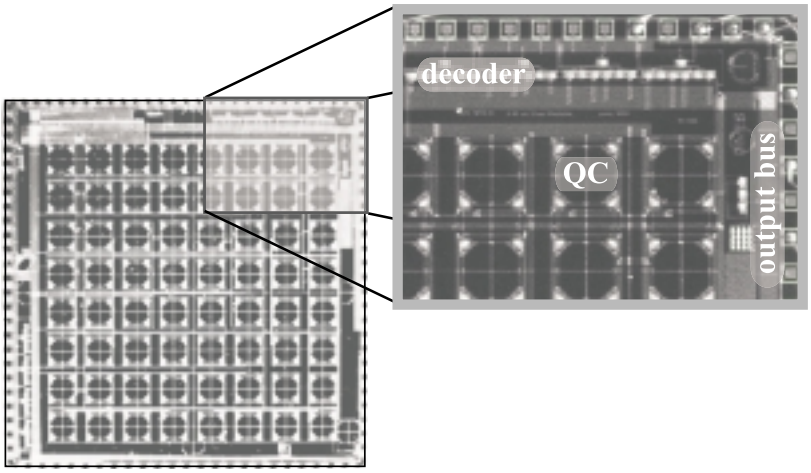

Figure 4.30 *Photograph of the implemented chip.*

4.4.3 Chip-level electronics

In this first prototype chip, the electronic circuitry present on chip consist of a distributed demultiplexer and a sampling-rate controller. The circuit blocks were located as much as possible along the chip edges, around the aimed photosensitive area, such that illumination influences the circuitry operation as little as possible. The blocks were also surrounded with guard rings in order to minimize the interference of stray carriers from the substrate. Whenever feasible, a second metal layer was used for light shielding.

The demultiplexer is denominated '*distributed demultiplexer*' because it consists of several independent blocks spread over the chip. It has a digital part and an analog part. The digital part is actually a 1-to-64 router, where the HIGH input level is routed to one of the 64 outputs depending on the input address (6-bit word). All other outputs are kept at a LOW digital level. The analog part consists of pixel-level analog switches, which either transfer the analog photocurrents to the output bus or discharge them to the ground.

The 1-to-64 router can be subdivided in a decoder controller and two main decoder blocks: an inner-region decoder and an outer-region decoder. The former addresses quad cells 0 to 31 and the latter quad cells 32 to 63 (see Figure 4.28). These two decoders are commanded by the controller, which is a simple two-channel demultiplexer coupled to two four-channel demultiplexers. The schematics of the two-channel demultiplexer is depicted in Figure 4.31 along with the block diagram of a four-channel demultiplexer. Depending on the value of input "A", it turns on either the 4-channel demultiplexer that controls the inner decoder or the 4-channel demultiplexer that controls the outer decoder, which are connected to output lines *out1* and *out2*, respectively. The design was made such that it guarantees the integrity of the HIGH and LOW voltage levels as they are passed on within the circuit. Each decoder consists of four sub-decoders, whose architecture is based on two-channel demultiplexers. A block diagram of one of these decoders is shown in Figure 4.32.

For operation in either coupled or uncoupled mode, a sampling-rate controller was included on chip. Its task is to transfer either four or eight signals simultaneously to the output pins. It is a batch of 32 switches which are either turned on or off simultaneously by the voltage applied to the

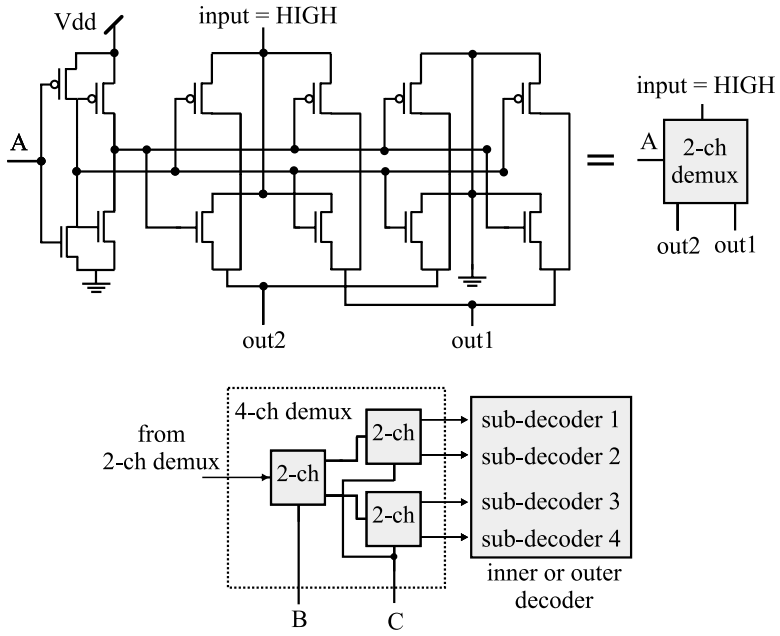

Figure 4.31 *Schematics of the decoder control unit.*

control pin. When the control pin is at a LOW digital level, the inner quad cells are addressed by the inner decoder and the outer quad cells by the outer decoder. When the control pin is set to the HIGH digital level, the switch sets of the outer quad cells are disconnected from the outer decoder and connected to the inner decoder, such that one inner quad cell and one outer quad cell are addressed simultaneously by each channel of the inner decoder. The operational schematics of one switch of the controller is depicted in Figure 4.33. Photographs of the decoder controller, sampling-rate controller and sub-decoder are shown in Figure 4.34.

4.4.4 External electronics

The chip is mounted on a Printed-Circuit Board (PCB), where the signal is amplified before going to a multifunctional data-acquisition card (DAq) that both sends the control signal to the chip and samples the analog voltages, which are then digitized and input to the computer for further processing.

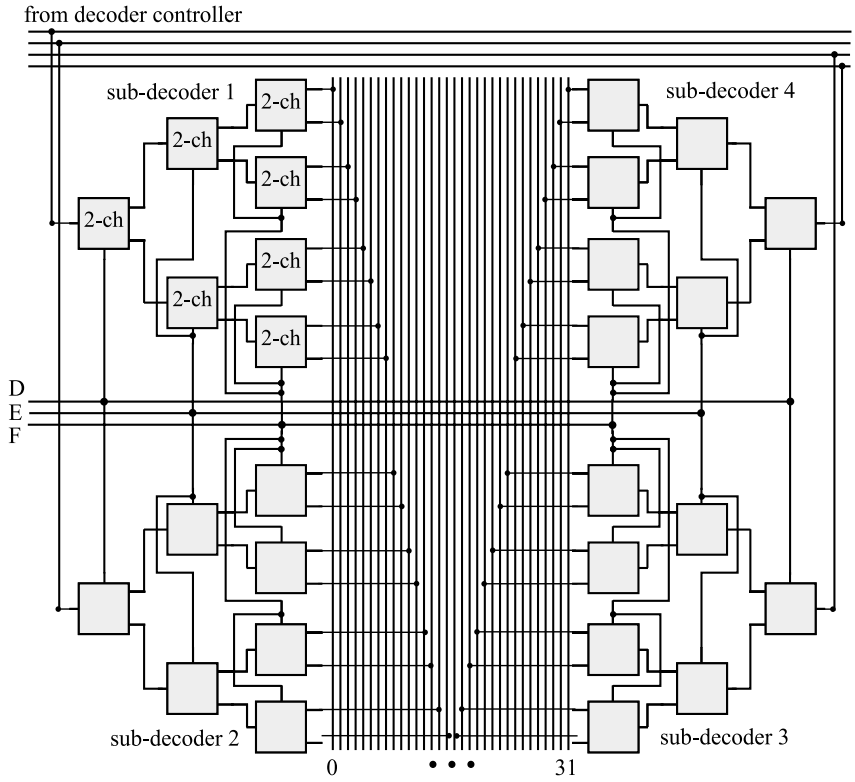

Figure 4.32 *Block diagram of one of the decoders. The control bus has 32 lines connected to the quad-cell pixels.*

The PCB is shown in Figure 4.35, where one sees the chip positioned in the middle and connected with gold bond wires to the contact pads. The photocurrents from the chip are amplified by means of commercial transimpedance amplifiers, connected in resistive feedback mode. The PCB is interfaced to the DAq via a 20-pin connector for flat cable.

The data-acquisition card we used to interface the data to the computer is the PCI6023E from National Instruments. PCI was chosen over ISA for its higher transfer rate. This model has a 12-bit resolution analog-to-digital converter, features 8 digital I/O ports and 16 single-ended analog inputs, with a maximum sampling rate of $2x10^5$ samples per second. However, this

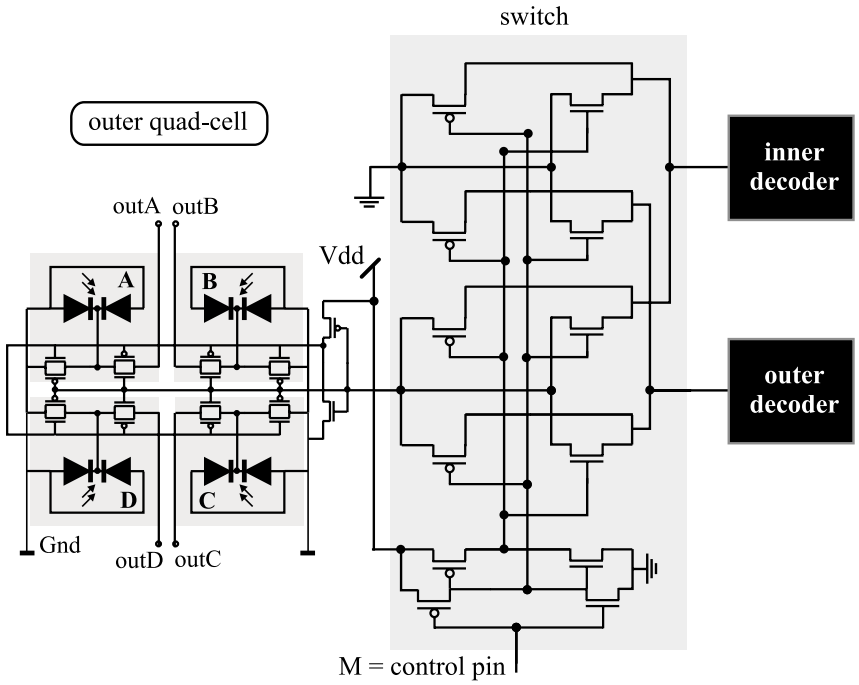

Figure 4.33 *Schematics of one switch of the sampling-rate controller and its connection to an outer quad cell and the decoders.*

is only achieved for a single-channel readout with the data-integrity control disabled in its software driver. The card multiplexes the input signals and converts them serially to digital signals. The performance of the wavefront sensor when using this interface card will be discussed in the next chapter.

4.5 Chapter summary

Chapter 4 dealt with the development of the Hartmann wavefront sensor. It presented an overview of the standard CMOS technology we used and included the implementation and the results of several independent position-sensitive structures realized prior to the implementation of the complete sensor prototype. The chapter described the development of the

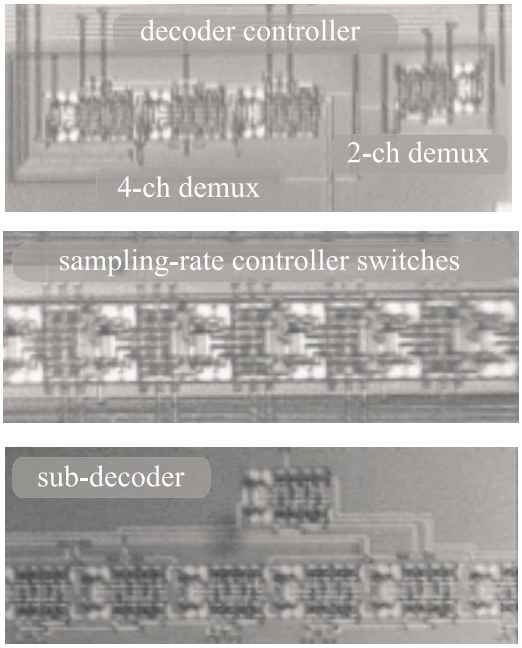

Figure 4.34 *Photographs of the integrated decoder controller, sampling-rate controller and sub-decoder.*

wavefront sensor prototype, including a note on the fabrication of the Hartmann mask and a detailed review of the implementation of the elements integrated on chip as well as a description of the hardware used as the interface to the computer.

All detector structures were fabricated at DIMES in a standard CMOS n-well process with 1.6-μm design rules. Compared to state-of-the-art CMOS technologies, this conservative process offers advantages to photo detection as, for instance, deeper junctions and lighter doping concentrations, favoring a high intrinsic quantum efficiency.

A Hartmann mask with sixty-four 450-μm circular apertures in a orthogonal grid was designed and fabricated with aluminum, which was first deposited on a glass substrate and then etched at the aperture locations. It has a high geometrical precision of 1 μm.

Since the cornerstone of the proposed wavefront sensor is a matrix of

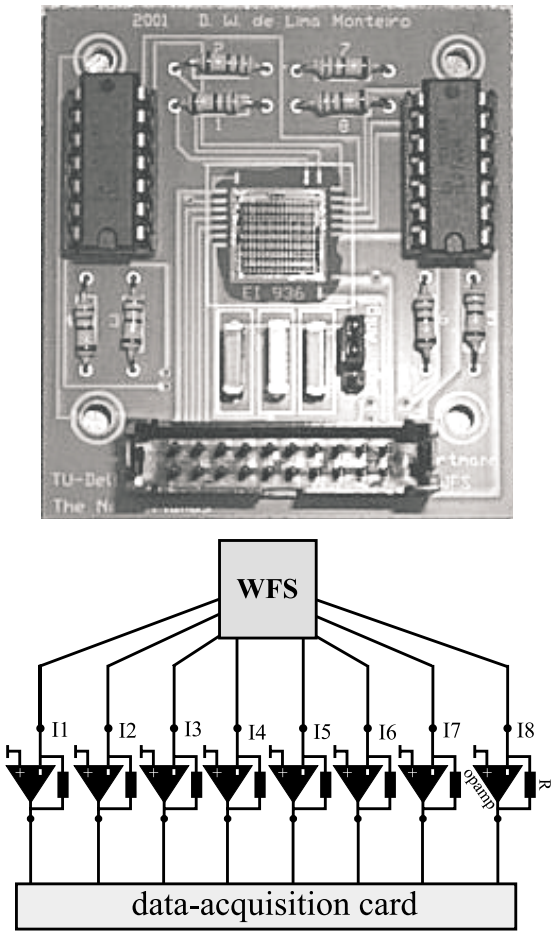

Figure 4.35 *CMOS chip mounted on a Printed-Circuit Board (PCB) with the amplification stages.*

position-sensitive detectors (PSDs), before implementing the whole wavefront sensor we developed a number of PSD structures in the chosen CMOS technology in order to evaluate their performance. The first targeted characteristics were high linearity, response uniformity and a large spot-displacement range. The fabricated one-dimensional lateral-effect

photodiode was disregarded because of its poor differential response. Among the two-dimensional multipixel PSD structures, we implemented three chessboard-like structures (different photosensitive elements), one spiral structure (photodiodes) and a quad cell (photodiodes). The first two PSD types were read out with the current-division technique, where the chessboard structures had integrated resistive arrays and the spiral had external discrete resistors.

For the chessboard PSDs the differences in the linearity of the responses were attributed to the variations of the read-out array resistance rather than to the different types of photodetectors used. Resistance variations across a single array induced asymmetrical position responses with respect to the center of the devices. Modelling indicated that if accurate resistors were used, an optimal compromise among response linearity, dynamic range and signal swing would be obtained for a spot with a diameter approximately equal to the structure pitch. Simulation also showed that some crosstalk between cells in the chessboard PSD has a beneficial effect, that is, it partly conceals its discrete nature and smoothens the response. One obvious drawback of the chessboard PSD is that each of the two superimposed 1-D cell arrays represents huge gaps to the other array. Therefore, a spot moving along different lines along the surface of the device results in different responses. This effect is minimized for a larger spot-to-pixel ratio. To keep the spot-displacement range the same, one can improve the spot-to-pixel ratio by increasing the density of cells, which in practice means more resistors and hence more response imbalance.

The spiral structure is a novel PSD structure that, for the same number of cells and device size as the chessboard, features a double pixel density per coordinate and, hence, a pitch that is about twice as small. Therefore, the optimal spot can have about half the diameter of that of a chessboard PSD. Because of this higher pixel density it offers a much more uniform response along different sections on the device area. The measured linearity and uniformity of the response of this structure agreed well with the results of the numerical model, which can be attributed to the high precision of the external resistor array used for the current division. Despite all advantages this structure has over the previous one, integration of a matrix of spiral PSDs on chip would require a large number of integrated resistors with the disadvantages of resistor imbalance and an increased chip area.

A simpler approach is the quad cell with a well-defined geometrical center, simple layout, high fill factor and the presence of only four output nodes, which renders the use of resistive arrays unnecessary. Although its response is only linear for spot displacements around the very center of the device, the response curve can be approximated in most cases by a sigmoidal curve. Smaller spots can be used for center tracking and larger spots are necessary when one wants to acquire centroid displacement information, as is the case with the Hartmann sensor. There are some observations for position sensing that must be taken into account: the effective spot radius should not be larger than the lateral size of one cell of the detector, and the smaller the spot, the better the resolution around the quad-cell center and the shorter the spatial dynamic range. For maximum light levels per quad cell an effective spot radius equal to half the lateral cell size offers a good compromise between position resolution, displacement range, photodetection and optical crosstalk.

Our first prototype of the proposed CMOS-based wavefront sensor consists of a Hartmann mask positioned at 6 cm from the chip, which has 64 quad cells that can be randomly addressed. For this first approach, passive-pixel quad cells were chosen for their simplicity. An integrated analog demultiplexer decodes the address provided by the computer and transfers the corresponding quad-cell signals to the data bus. The diagrams of the various circuit stages integrated on chip were displayed and commented upon throughout this chapter.

A printed-circuit board was designed to accommodate the CMOS chip and the operational amplifiers used to convert and amplify the input photocurrents into signal voltages, which are sampled by a commercial data-acquisition card, which performs analog-to-digital conversion and feeds the data to the computer, in our case a 750-MHz Pentium III computer running Linux.

References

[4.1] D. J. Noorlag, “Lateral-photoeffect position-sensitive detectors,” PhD Thesis, Delft University Press, Delft, 1982

[4.2] K. Bult, G. J. Geelen, “An inherently linear and compact MOST-only current division technique,” IEEE J. Solid-state Circ., vol.27, no.12, 1730-1735, 1992

References

Measurement results 5

In the preceding chapter we discussed the implementation of the wavefront sensor in detail. The goal was to develop a custom CMOS-chip sensor that is capable of real-time operation. This sensor relies on an integrated two-dimensional array of optical position-sensitive detectors (PSDs) for sensing the displacements of spots resulting from an arbitrary wavefront sampled on a regular grid. It represents an alternative to conventional imagers used with the Hartmann(-Shack)[1] wavefront sensing technique. Among various PSD structures, quad cells were chosen to be integrated in a matrix with random addressing capability. The choice was supported by the implementation simplicity and high fill factor per PSD besides the uniformity and the symmetry of the position response.

This chapter discusses the results obtained from the fabricated sensor and is divided into two main sections. Section 5.1 is devoted to the sensor performance as a whole and to the influence of the various parts involved in the wavefront detection. Section 5.2 introduces the concept of Adaptive Optics as well as presents the preliminary results of an adaptive optical system with the sensor coupled in closed-loop to a deformable mirror.

1. Recall that the only aspect in which the Hartmann-Shack method differs from the Hartmann one is its use of a microlens array instead of a perforated mask.

5.1 Wavefront sensor

This section analyzes the performance of the various parts of the sensor. First, some considerations on the choice of the distance between the Hartmann mask and the sensor chip are discussed, followed by an assessment of the requirement to correct mask alignment. Later in this section the characteristics of position-sensitive detectors are revealed and the sensor test setup is introduced. At last, the wavefront sensor performance is evaluated, where the role of the reconstruction algorithm is addressed, and the results on wavefront-reconstruction accuracy and operational frequency are presented.

5.1.1 Hartmann mask and spot profile

As seen in Chapter 4, the most linear quad cell (QC) response is achieved with a uniform square spot profile. However, in order to obtain a rather compact device, one should situate the Hartmann mask somewhat close to the sensor chip, at which distance the spot-intensity profile on the chip is determined by near-field diffraction effects (Fresnel diffraction)[2] and a square aperture on the mask yields a non-uniform profile with square symmetry, to which the quad-cell response is no longer best matched by a linear fit.

We evaluated numerically the quad-cell response for different spot profiles resulting from masks with 450-μm circular and square apertures placed at various distances from the sensor. The responses were fitted with sigmoidal curves and the fitting errors over the full and central quad-cell ranges were computed. Figure 5.1 shows the error as a percentage of the total response range for different distances from the mask to the detector. As regards fitting, these graphs suggest that the choice of the aperture geometry depends on the distance between the Hartmann mask and the detector. Limiting the maximum distance to 16 cm and using the full quad-cell range, it is better to use the circular apertures almost over the whole range of distances considered, except from 7 cm to 12 cm, where the spot profile resulting from a square aperture offers a better fitted quad-cell

2. For apertures with 450-μm greatest widths and a wavelength λ=633nm, near-field spot-intensity distributions prevail up to a distance of about 10 cm (section 2.2 / footnote 3).

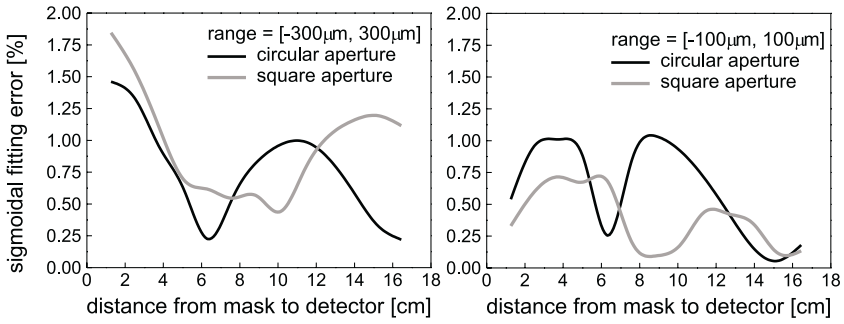

Figure 5.1 *Fitting error as a function of the distance between the Hartmann mask and the detector: (a) full range and (b) central range.*

response. When the spot displacements are limited to the central region of the quad cell [-100μm, 100μm], the previous choices only change for distances shorter than 5 cm, at which a mask with square apertures proves to be a better option. A mask with circular apertures placed 6 cm away from the sensor chip seems to be a reasonable choice, resulting in a fitting error of only 0.2% fitting error while the mask remains at a distance close enough to the detector to allow a compact mask-to-sensor assembly.

When a matrix of PSDs is used instead of a camera, one should consider the alignment of the detector with the mask. Translational misalignment is interpreted as a global tilt of the wavefront and, to a certain extent, has no effect on the response characteristic of the quad cells. The unwanted effect is that it limits the maximum spot-displacement range. Rotational misalignment, on the other hand, can be a problem. If we consider a rotation angle about an axis that passes through the mask and detector centers, then rotation introduces both translation and rotation of each quad cell with respect to its corresponding aperture, as seen in Figure 5.2.

These misalignments limit the spot displacement range and introduce a coordinate estimation error because the quad-cell coordinate system differs from the one of the mask. The external spots are subjected to more severe truncation of the displacement range than central spots. The errors Δx_{err} and Δy_{err} between the real (Δx and Δy) and the sensed displacements (Δx_{QC}, Δy_{QC}) are given by equations (5.1) and (5.2):

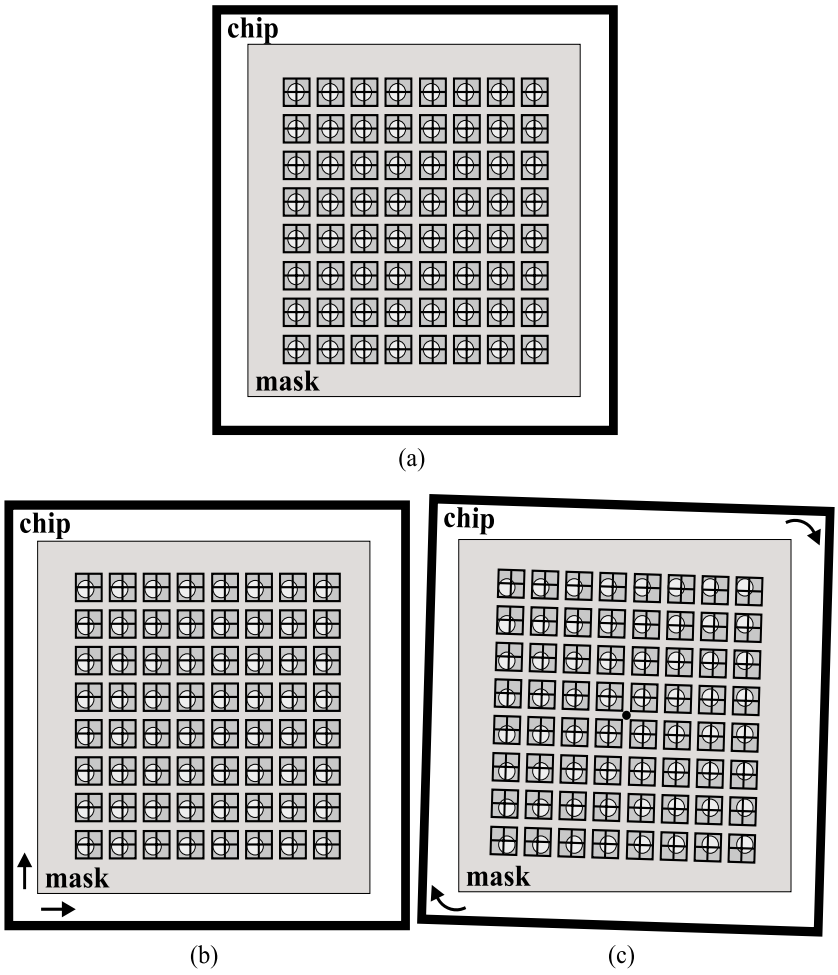

Figure 5.2 *Detector alignment with respect to the mask: (a) aligned, (b) translated, (c) rotated about the central axis.*

$$\Delta x_{err} = (1 - \cos\theta)\,\Delta x - \Delta y \sin\theta\,, \tag{5.1}$$

$$\Delta y_{err} = (1 - \cos\theta)\,\Delta y + \Delta x \sin\theta\,. \tag{5.2}$$

Since every element in the detector matrix is a position-sensitive device, we can use their own responses to finely align the detector to the mask by means of a reference light beam.

5.1.2 Detector characteristics

As seen in the previous section, a mask-detector distance of 6 cm is the most convenient when a mask with 450-μm circular apertures is used. In this case, the spot intensity profile on each quad cell yields a position response that can be fitted with minimum error by a sigmoidal function (equation (4.1)). However, not only data fitting plays a role in the determination of the quad-cell position resolution, but also noise in the signal is crucial.

For proper data fitting one must calibrate the quad cells in the matrix. Once the pixels consist of large photodiodes with complementary switches, their responses over the array match very uniformly. The Hartmann mask was fabricated with high accuracy and the intensity profiles of the light spots over the array also match closely. Therefore, the position response curve of one quad cell can be taken as the representative of any other quad cell in the matrix. Noise levels might differ because of non-uniform illumination but the average response curve remains about the same.

To calibrate the quad-cell response we used a spatially filtered and collimated He-Ne laser beam (λ=633nm). The PSD matrix was mounted on *x-y* translation stages and positioned 6 cm from the Hartmann mask. The quad-cell response in the *x*-direction, for instance, was recorded by keeping the illuminated Hartmann mask fixed and by moving the quad cells along the *x*-direction. The output-data lines of the sensor were connected to commercial operational amplifiers with 3.2-MΩ feedback resistors for current-to-voltage conversion. First, with the light blocked, the offset voltages were recorded and then the pixel signals were measured, from which the offset values were subtracted. This eliminates most of the amplifier offset and dark fixed-pattern noise (FPN) in the pixels. In the CMOS process we used, the dark current resulting from the double junction photodiodes is about 1800 pA/cm^2 at room temperature.

The position response was measured for different spot intensities along the *x*-direction. Figure 5.3 shows the variation of the response characteristics with intensity for a spot with an approximately 400-μm

effective diameter. The lower the intensity, the more significant the noise in the response and the worse the position resolution, as indicated in the figure for each response. The impact of the quad-cell position resolution on the wavefront accuracy will be addressed in section 5.1.4.

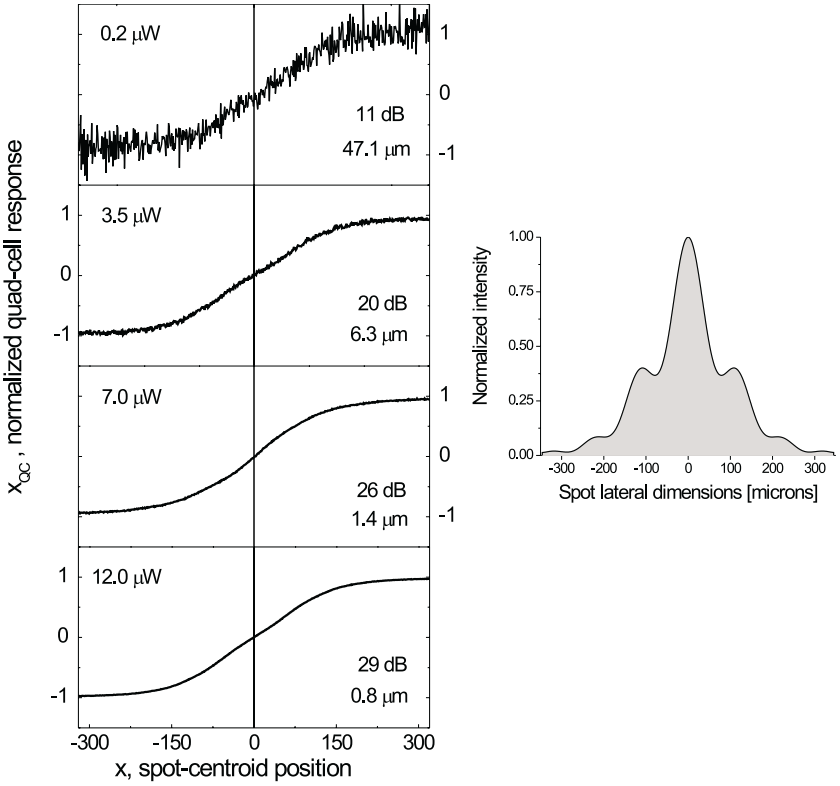

Figure 5.3 *Quad-cell response curve along the x-direction for different light power values of a radially symmetric spot with a 400-μm effective diameter. The position signal-to-noise ratio and the position resolution are indicated in each case.*

Dark-current shot noise for our setup is 0.16 pA/pixel. In the low-noise case (e.g. 12–μW spot) the response was registered in the *x*- and *y*-directions and along different quad-cell sections. All response curves over the full range, with the same spot intensity, matched to within less than 2%. In the central region [-100μm, 100μm] they matched to less than 0.5%.

The higher the quantum efficiency, the lower the spot power that will still yield a satisfactory position resolution. To obtain the quantum efficiency at λ=633 nm, we illuminated a test photodiode[3] on chip with a laser spot with known optical power and then registered the output current. The total quantum efficiency η of the photodiode, at this wavelength, is about 40%, which results from equation (2.32). This figure might change from one chip to another, but especially from one fabrication batch to another because of the variations in thickness of the dielectric layers left on top of the photodiode. The layer stack works as a thin-film stack - refer to Figure 4.2- and influences the spectral transmittance, consequently influencing the amount of light that reaches the Si surface and therefore the photodiode quantum efficiency.

We also measured the spectral response of the test photodiode from 400 to 700 nm. The measurements were performed with a broad-spectrum light source and a monochromator. A 150-W Xe lamp with a spectrally flat characteristic in the wavelength range 400nm-700nm was used as the light source; the monochromator used was the imaging spectrometer Triax 180, with a wavelength accuracy of ± 0.3 nm. The photodiode response is shown in Figure 5.4. The same figure has a plot of the associated transmittance through the film stack for various wavelengths. By varying the thickness of the dielectric layers within their known process-related tolerances in the transmittance model, we numerically sought the stack combination that best matched the peaks obtained from the photodiode response. The best match within the given tolerances has a maximum spectral peak deviation of 13 nm and corresponds to a 791-nm Si_3N_4 layer on top of a 1572-nm stack formed by different oxide layers, making the simplifying assumption that these layers have the same refraction index and absorption coefficient.

3. the test photodiode has a layout and connections identical to a photo-diode in a quad-cell pixel, the only difference is that its output is directly connected to a test bond pad on the chip instead of going through a switch.

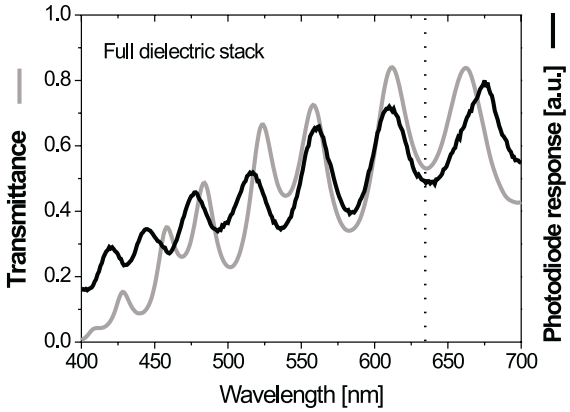

Figure 5.4 *Transmittance and spectral response of the photodiodes with a full dielectric stack on top.*

Refer to Figure 4.2 for the possible nominal flow-chart values.

The important observation here is that at λ=633 nm, the response is off the peak, i.e., it is at a local minimum and the corresponding transmittance is about 50%. It can be concluded that the dielectric layers are critical to the photodetector quantum efficiency and that for this particular fabrication batch it would have been better to remove all dielectric layers from the top of the photodiodes. An optimal combination of the standard layers can be found for the 633-nm wavelength, which results in circa 85% transmittance, as indicated in Figure 5.5; nevertheless the thicknesses of these layers are not accurately controlled in a standard CMOS process, which might cause them to deviate substantially from the theoretical prediction.

5.1.3 Optical setup

The optical setup used to test the fabricated wavefront sensor is shown in Figure 5.6. The setup was built not only to test the sensor, but also to make it possible to couple the sensor with a deformable mirror in an adaptive optical closed-loop system later on, as will be discussed in Section 5.2. A 5-mW He-Ne laser (λ=633 nm) is spatially filtered and expanded to 11 mm. There are three optically conjugated planes: the aberration plane,

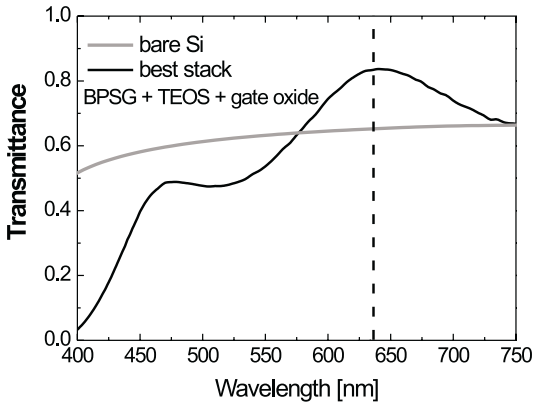

Figure 5.5 *Optimal transmittance for the 633-nm wavelength, when only BPSG, TEOS and the gate oxide are left on top of the photodiode. The dotted line represents the transmittance of the bare silicon surface.*

mirror *M1* and the Hartmann mask. The conjugation is achieved by means of the lens systems A and B. The relationships between the indicated distances and the focal lengths of the various lenses are given by equations (5.3) and (5.4), respectively [5.1].

$$L_2 = \frac{f_2(f_1^2 + f_1 f_2 + f_2 L_1)}{f_1^2}. \tag{5.3}$$

$$L_4 = \frac{f_4(f_3^2 + f_3 f_4 + f_4 L_3)}{f_3^2}. \tag{5.4}$$

The lens pairs can form afocal systems to expand or shrink the beam diameter, according to the following equations:

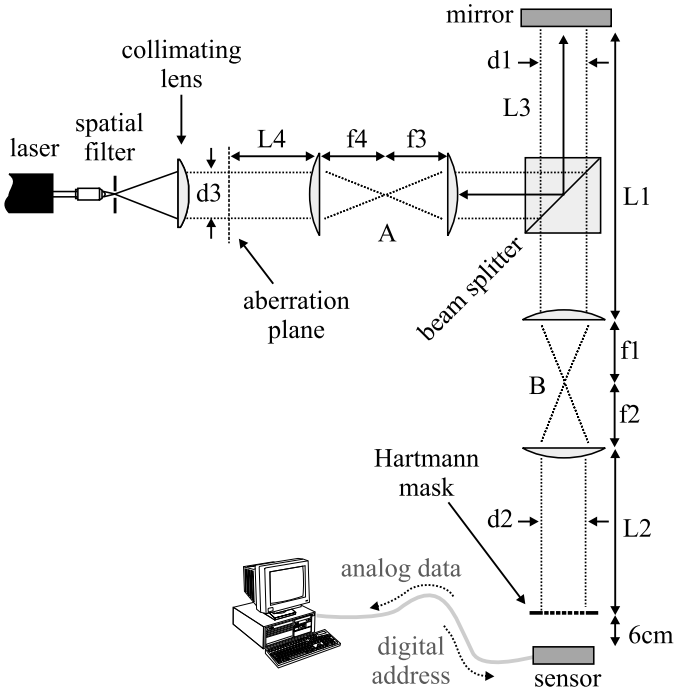

Figure 5.6 *Optical setup to test the wavefront sensor.*

$$d_2 = d_1 \frac{f_2}{f_1} . \tag{5.5}$$

$$d_3 = d_1 \frac{f_4}{f_3} . \tag{5.6}$$

The 11-mm diameter was preserved such that the 64 apertures of the Hartmann mask were illuminated, as shown in Figure 5.7. Aberrations introduced at the aberration plane are seen by the sensor as if they were introduced at the plane of the Hartmann mask, where the wavefront is sampled. Rotation and translation stages equipped with micro-screws

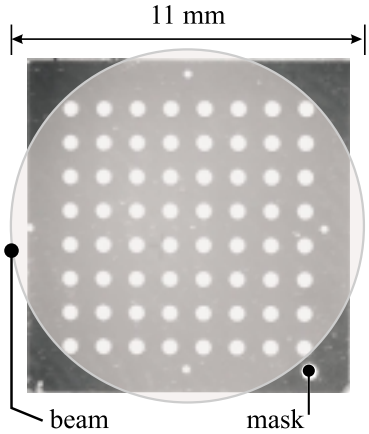

Figure 5.7 *The laser beam illuminating the Hartmann mask.*

enabled the alignment of the CMOS chip to the mask. Coarse alignment was done visually by observing the spots on the chip surface when the mask was illuminated by a reference beam; fine alignment was done with the aid of a virtual grid on the computer screen, to which we could align virtual spots obtained from the position responses of the quad cells.

5.1.4 Sensor performance

The wavefront sensor performance can be mainly characterized by the accuracy with which it can read a wavefront and by its operational frequency. This section first addresses the wavefront-reconstruction method used, the sensor model and the estimated error budget; then it introduces the results obtained with the fabricated device.

Wavefront reconstruction

In the Hartmann method, an arbitrary wavefront is sampled into light spots, whose centroid positions are the only data available for estimating the wavefront aberrations. For the wavefront reconstruction we use the modal approach, in which the first derivatives of Zernike polynomials are used as the set of basis functions. The number of polynomials is referred to as the reconstruction order. A maximum of 128 slopes can be measured with this

sensor, with two orthogonal tilts at each quad-cell site. We use Gaussian elimination with partial pivoting with the least-squares approximation to solve equation (2.15) and to find the vector of Zernike coefficients [5.2, 5.3].

A numerical model of the sensor allows us to assess how the reconstruction error varies with noise in the quad cells. The various steps of this model are introduced below:

- random spot-centroid positions are generated on the surface of the quad cells and stored; they obey a uniform random distribution, lie within a 150-μm radius about each quad-cell center and correspond to a reference wavefront;
- an arbitrary wavefront is represented by a set of random positions on each quad-cell surface, according to a uniform distribution;
- random position fluctuations around the latter positions are introduced. Each fluctuation is determined by a Gaussian distribution whose mean is the arbitrary position value generated, and the variance is the *rms* detector resolution squared. Therefore the fluctuations around the reference position values correspond to detector noise;
- the noisy position values are translated into quad-cell responses based on equation (4.1). An equivalent expression holds for *y*. The parameters in the sigmoidal-fit expression can be extracted from experimental data or from a numerical evaluation of the quad-cell response for different spot profiles;
- the displacements from the reference positions are computed and the tilts are calculated;
- the wavefront is reconstructed with an error associated with the initial position noise that propagates within the algorithm.

These steps can be evaluated many times for different reference wavefronts and various incident wavefronts to provide a reliable idea of how the sensor performs for different noise levels. As explained, the wavefront is reconstructed based on the measured spot displacements; the accuracy of these measurements is limited by the resolution of the position-sensitive detectors, namely the quad cells in our case.

The position resolution is determined by noise in the photodetector and read-out circuitry. Noise in the displacement vector propagates within the

reconstruction algorithm, which induces inaccuracy in the determination of the Zernike coefficients and therefore causes a wavefront reconstruction error. This error was modelled as a function of the root-mean-square position resolution of the detector for λ=633 nm, as shown in Figure 5.8.

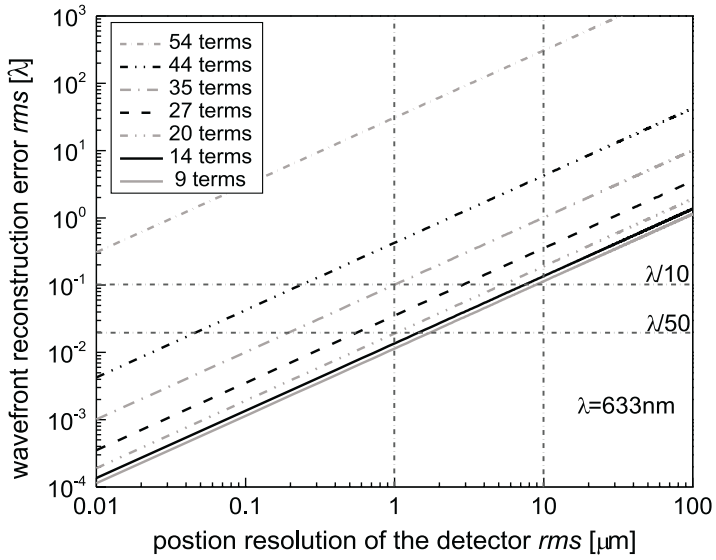

Figure 5.8 *The rms wavefront reconstruction error, for various reconstruction orders, as a function of the position resolution of the detector.*

The higher the reconstruction order, the more significant the error. When the position resolution of the device is 1μm, reconstruction orders up to 35 Zernike terms yield reconstruction errors below λ/10, whereas when the position resolution is 10 μm, not even 9 Zernike terms result in a better result than λ/10.

Wavefront accuracy

The wavefront accuracy can be defined as the minimum *rms* variation in the wavefront that the sensor can detect. This can be experimentally found by measuring static wavefronts a number of times. The accuracy is then calculated from the *rms* deviations among the reconstructed wavefronts.

As seen before, the accuracy is dependent on the detector noise and on the reconstruction algorithm. For 7-μW spots[4], with the profile given in Figure 5.3, the quad cells feature a 1.4-μm position resolution, which means that to achieve an accuracy better than λ/10 one can use a maximum numerically predicted reconstruction order that can be used lies between 27 and 35.

To estimate the wavefront accuracy of the implemented sensor, we sampled and reconstructed *n+1* times *N* arbitrary static wavefronts. Each time a wavefront was reconstructed, we computed its *rms* deviation from the previous wavefront was computed and attributed a label $\delta w(i,j)$, where *i* runs from 1 to *n* for any wavefront *j* in a set of N. The wavefront accuracy Δw and its corresponding standard deviation Δw_{rms} were then calculated from equations (5.7) and (5.8).

$$\Delta w = \frac{1}{Nn} \sum_{j=1}^{N} \sum_{i=1}^{n} \delta w(i,j) \, . \tag{5.7}$$

$$\Delta w_{rms} = \left[\frac{1}{Nn-1} \sum_{j=1}^{N} \sum_{i=1}^{n} (\delta w(i,j) - \Delta w)^2 \right]^{\frac{1}{2}} . \tag{5.8}$$

This procedure was done for different reconstruction orders and the results are plotted against the simulated results in Figure 5.9. The error bars represent the standard deviation from the mean value. It is noted that the simulated curve lies within the range of the experimental results. The spread of values increases with the reconstruction order, which is due to the

4. This values results from a 1.6-mW beam with 11mm in diameter.

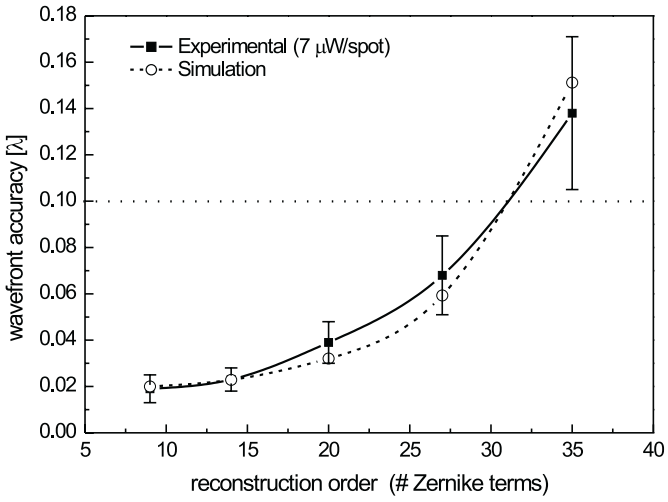

Figure 5.9 *Experimental and simulated wavefront accuracies of the sensor for different reconstruction orders.*

more significant noise propagation in the reconstruction algorithm. The number of Zernike terms used to reconstruct the wavefront should not exceed 30 if one wants to obtain an accuracy better than λ/10. This curve might change for different spot intensities, spot profiles and number of pixels actively used in the detection.

We observed that the accuracy is practically identical when 44, 54 and 64 quad cells, in the configurations of Figure 5.10, are used. This favors the use of the 44-QC configuration, once it yields a higher operational frequency.

When the device is subjected to lower light levels, taking the average of position data can improve the accuracy, at the expense of operational speed. Figure 5.11 shows the influence of averaged experimental data on the wavefront accuracy. It shows how the accuracy changes with the number of averaged position samples for two different spot intensities, 2μW and 7μW, when the 44-QC configuration and 9 Zernike terms are used.

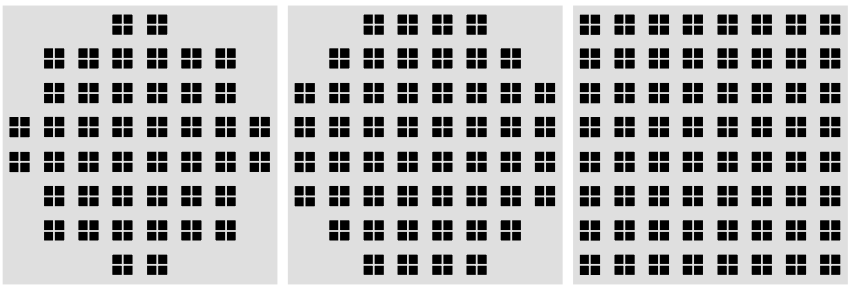

Figure 5.10 *Different configurations with different numbers of pixels used to sense the incoming wavefront (44, 52 and 64 quad cells).*

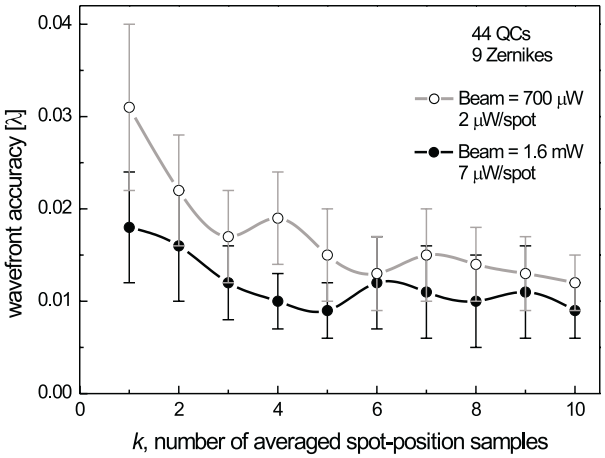

Figure 5.11 *Wavefront accuracy as a function of averaged position data for two beam intensities.*

For each quad cell q [5], position coordinates x_q and y_q are averaged k times before they are used in the reconstruction. The subsequent procedure to estimate the wavefront accuracy was described in the paragraph

5. The index q varies from 0 to 43 in this case.

preceding equation (5.7). The 2-μW spots with position data averaged over three samples yield about the same wavefront accuracy as the 7-μW spots with no averaging.

Operational frequency

The operational frequency is determined by the sensor read-out frequency, the interface hardware and the data-processing time. It represents the frequency at which samples can be acquired and reconstructed.

The sensor read-out frequency is the rate at which samples are available at the chip output pins. It depends on the pixel access time and on the number of parallel data outputs. The photodiode capacitance is determined by the sum of the capacitances of the two junctions actively used in the photodetection, which is 50 pF. However, in our case, the pixel access time is mostly limited by the switch settling time, which is the interval necessary to redistribute the charge previously accumulated in the capacitances of the source and drain of the switch. This redistribution process causes unwanted initial oscillations in the pixel signal. The switch settling time is 10 μs, and therefore, to guarantee stability, the data signal should only be read out 10 μs after the switch has been addressed. The number of parallel outputs can be either 4 or 8, uncoupled or coupled mode, respectively (Figure 4.29). The coupled mode yields a sensor read-out frequency of 3.125 kHz when all 64 quad cells are used.

The interface hardware between the chip and the computer represents a critical step to the overall frequency performance. There are two separate pieces of hardware used, a custom Printed Circuit Board (PCB) for current-to-voltage conversion and amplification and a multifunctional data acquisition board (DAq).

The role of the custom PCB is to convert and amplify the input photo currents into voltages. In the simplest approach, generating an amplified voltage, V_{out}, from an input current I_p implies in having a large load resistance R_L connected to the terminals of the photodiode, so that $V_{out} = R_L I_p$. This configuration significantly limits the bandwidth of the system, because the bandwidth is inversely proportional to the load resistance (section 2.4.4). Bandwidth improvement can be obtained by using an operational amplifier with a feedback resistor R_f (transimpedance operation), which yields a system bandwidth several orders of magnitude

larger than a simple load-resistor configuration [5.4]. This is achieved because when keeping the same output voltage as before, which means having $R_f = R_L$, the input resistance R_{in} of the operational amplifier is given by $R_{in} = R_L / (A + 1)$, where A is the amplifier gain. Another advantage of using an operational amplifier is that it has a large linear-response range, which is compatible to the inherent linearity of the responses of the photodiode [5.5].

The next stage is the data-acquisition board, which must provide the chip with the control digital address, acquire the parallel analog voltages from the chip, convert them to the digital domain and deliver them time multiplexed to the computer. We used a data-acquisition card (PCI6033E, section 4.4.4) driven by a register-level C program in order to enable direct coupling to the other pieces of C code used for the sensor alignment, spot-displacement calculation and wavefront reconstruction. The computer used has a 750-MHz Pentium III processor running Linux.

A timing diagram indicating the delays introduced by the data-acquisition card and the software routines is depicted in Figure 5.12. The

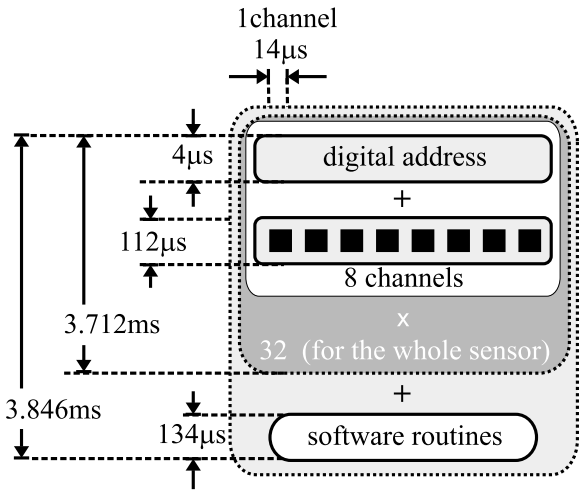

Figure 5.12 *Timing diagram for the data-acquisition card and software routines.*

acquisition cycle is as follows: the DAq provides the sensor chip with a particular 5-bit word and 8 analog signals are amplified and transferred to the DAq; the analog voltages are serially converted to digital format with a 12-bit resolution and transferred to the computer; another 5-bit word is sent to the chip until all 256 pixels are read. The acquisition time for one channel with data integrity enabled and in multiple-channel scanning mode is 14μs, hence the total time required to read all 64 quad cells is 3.712 ms. When all signals are acquired, the spot-centroid coordinates are computed and the wavefront is reconstructed. When 64 quad cells (128 displacement coordinates) and 9 Zernike terms are chosen, these software routines need 134μs. In this case, the resulting operational frequency is 260 Hz, which is primarily limited by the acquisition time of the DAq. Substitution of this card by a faster one will improve the frequency.

Figure 5.13 shows the operational frequency with the data-acquisition card used when different numbers of quad cells are read out. Since 64 quad

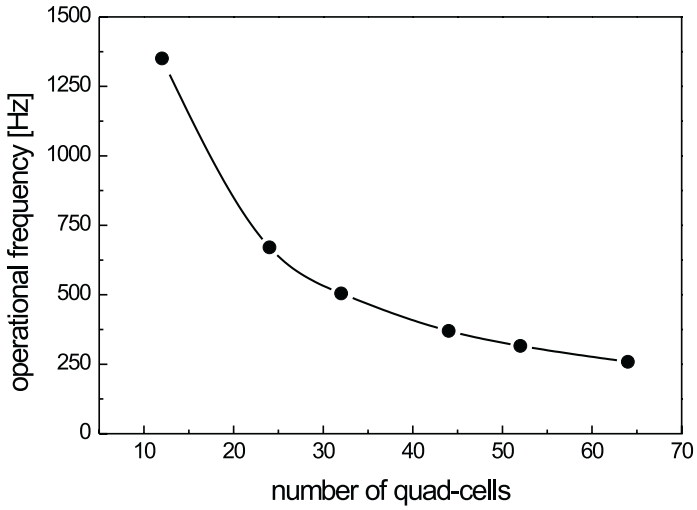

Figure 5.13 *Operational frequency for different numbers of quad cells actively used.*

cells yield nearly the same wavefront accuracy as 44 quad cells it is now clear that it is more advantageous to use the latter configuration for it renders roughly 1.5 times faster operation than the former configuration, namely 370 Hz. The pixel configurations for 12, 24 and 32 quad cells are illustrated in Figure 5.14.

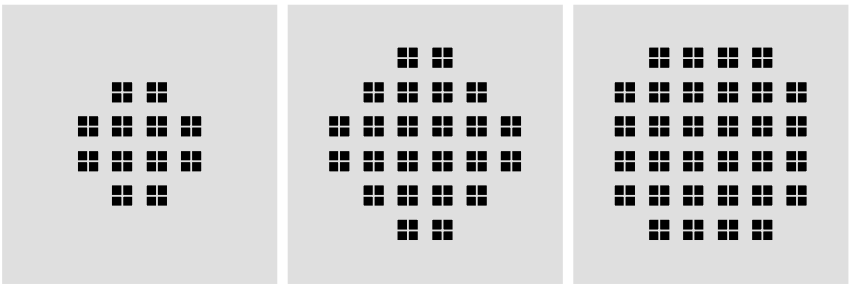

Figure 5.14 *Different configurations with different numbers of pixels used to sense the incoming wavefront (12, 24 and 32 quad cells).*

5.2 Adaptive optical system

Adaptive Optics (AO) consists in the real-time detection and correction of aberrated wavefronts [5.6]. A typical adaptive optical system is built with three main elements: a wavefront sensor [5.7], a data processor and an adaptive corrector. AO has been used for decades in Astronomy to compensate for wavefront distortions caused by a turbulent atmosphere. In parallel, considerable military efforts have been invested in this field. The astronomical and military sectors, however, have been mostly interested in high performance rather than cost. In the past few years, the interest in the research of low-cost AO components has increased, chiefly because of potential commercial, industrial and medical applications, such as free-space optical link in dense urban areas, extra- and intra-cavity laser beam optimization, dynamic surface-profile testing and ophthalmology [5.8, 5.9].
When real-time analysis and compensation of wavefront aberrations are desired, fast AO elements are needed. A challenge is posed by devising fast

and yet inexpensive components. Very often the wavefront sensor limits the speed performance of the AO system. The Hartmann method offers a straightforward way to obtain information about the wavefront aberration; the custom sensor we implemented brings an alternative to the speed issue besides being compatible to a standard technology.

We built a closed-loop AO system with the fabricated wavefront sensor and a micromachined deformable mirror [5.11] as the adaptive corrector (section 5.2.1). The data processor is a 750-MHz Pentium computer running a feedback-control algorithm (section 5.2.2). A diagram of the setup is shown in Figure 5.15 and its implementation in Figure 5.16. The

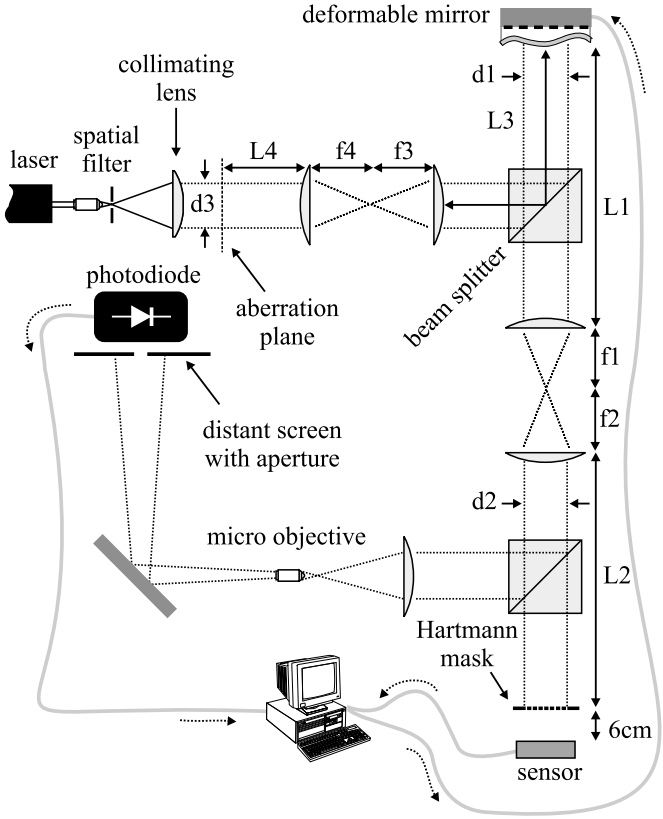

Figure 5.15 *Adaptive optical system setup.*

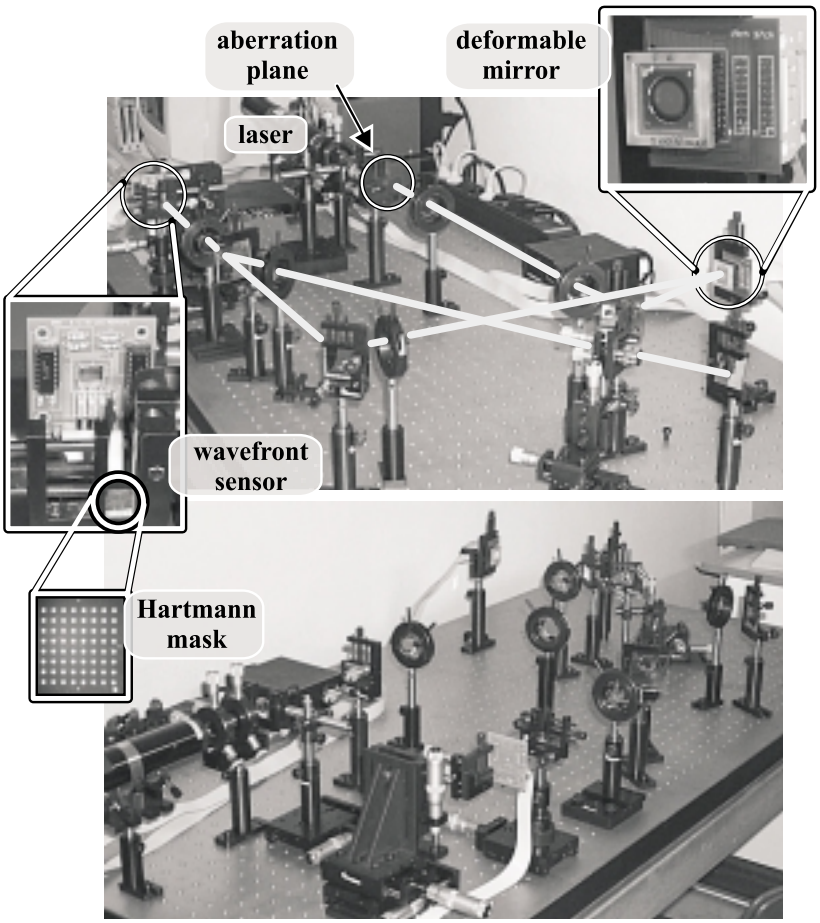

Figure 5.16 *Photographs of the adaptive optical system setup.*

beam is spatially filtered and expanded to 11 mm. The plane of the mirror is conjugated to the aberration plane and to the wavefront-sensor plane. The wavefront is reflected by the mirror and heads towards the sensor, where it is sampled. Part of the beam is split towards a micro-objective, which projects its far-field distribution to a distant screen with a small aperture in its center. Behind the screen there is a photodiode used for the beam optimization (section 5.2.2).

5.2.1 Deformable mirror

A deformable mirror alters an incoming wavefront by changing the shape of its surface. Figure 5.17 illustrates this concept for two different initial wavefronts. An adaptive mirror whose surface is described by the continuous function S(*x*,*y*) transforms a plane wavefront $\Xi_0(x,y)$, propagating in the *z* direction, into a wavefront $\Xi_{out}(x,y)$ [5.10]:

$$\Xi_{out}(x, y) = -2\, S(x, y)\,. \tag{5.9}$$

Likewise, an arbitrary wavefront $\Xi_{in}(x,y)$ can be transformed into a plane wavefront by applying the following deformation to the mirror surface:

$$S(x, y) = -\Xi_{in}(x, y)/2\,. \tag{5.10}$$

Micromachined-Membrane Deformable Mirrors (MMDMs) are compact low-cost devices with good optical quality [5.11]. They have been successfully employed in a wide range of applications such as laser-pulse compression [5.12, 5.13], retinal imaging [5.14, 5.15], confocal microscopy [5.16], intra-cavity laser optimization [5.17], aberration generation [5.18, 5.19] and dynamic focusing [5.20], among others. These

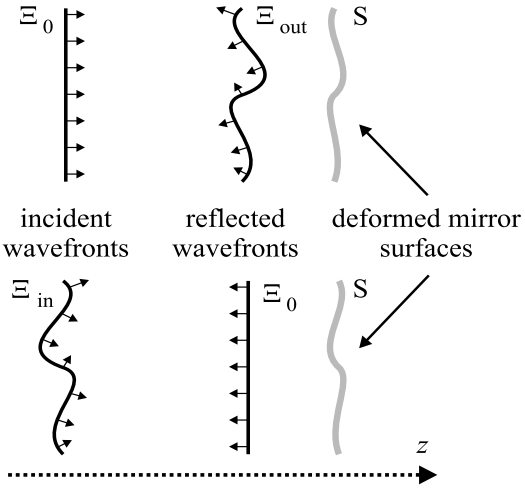

Figure 5.17 *Wavefront shaping with a deformable mirror.*

adaptive mirrors are fabricated by OKO Technologies in cooperation with DIMES [5.21]. They basically consist of a thin metal-coated Si_3N_4 membrane positioned above an electrode structure, as shown in Figure 5.18a. By applying voltages to the electrodes one can deform the membrane by means of localized electrostatic forces.

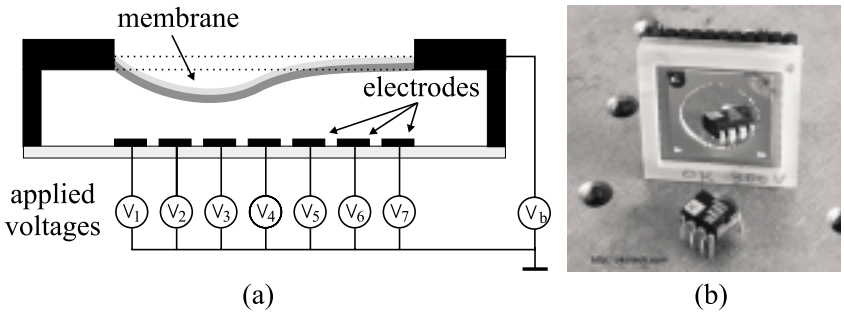

Figure 5.18 *Schematics of the micromachined membrane deformable mirror and photograph.*

We used a 37-electrode circular adaptive mirror with a 15-mm diameter and a maximum membrane deflection of 8μm at the center. It is aluminum coated, supports a maximum 0.03-W/mm^2 optical load (λ=633nm CW), features less than 1% scattering at the surface and operates up to 1000 Hz. Figure 5.18b shows a photograph of this mirror.

The initial mirror surface is very close to flat, which only permits deflection in the direction of the electrode structure by means of electrostatic attraction. In order to enable the mirror to correct bidirectional aberrations its surface is biased, i.e., an equal voltage is applied to all actuators, so that the mirror surface assumes a near parabolic shape and local deflections become possible in both directions [5.22]. Figure 5.19 presents the layout of the electrode structure and illustrates the concept of biased operation.

5.2.2 System operation

To operate the system we need to follow two steps before we start the feedback loop. The system first needs to be calibrated to establish a reference beam and then a basis must be generated. The elements of the

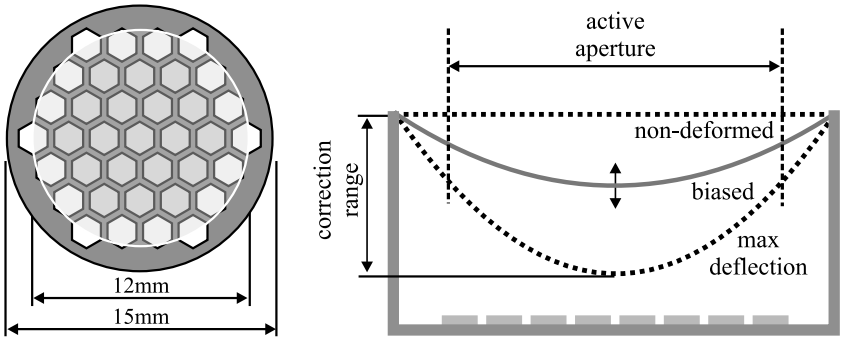

Figure 5.19 *Mirror hexagonal electrode structure and concept of biased operation of the mirror membrane.*

basis are the sensor responses to the wavefronts resulting from the beam reflected by each particular shape of the mirror surface corresponding to the actuation of individual electrodes.

Calibration

The aim of calibration is to provide the sensor with a reference wavefront that is represented, from the sensor's point of view, by a set of spot-centroid coordinates, to which other sets of spots positions, associated with arbitrary incoming wavefronts, will be compared. This reference wavefront is not necessarily flat and incorporates possible static aberrations present throughout the optical setup. To find a good reference beam we use a downhill simplex control algorithm [5.23] to obtain the best possible combination of voltages to be applied to the mirror to achieve a good far-field distribution of the beam reflected by the mirror. The far-field intensity profile targets a distant screen which is perforated in the center. A photodiode is positioned behind the small aperture on the screen and its signal provides the algorithm with an indication of whether a certain combination of voltages improves the optical throughput through the small aperture.

In this calibration step, the mirror works in feedback loop with this single photodiode. The mirror is set initially to a biased surface and the far-field distribution is presumably not optimized. The optimization process starts searching for good sets of control voltages in the whole practical

range of voltages. Then, a fine-adjustment mode looks within these sets for voltage values that provide the smoothest and fastest convergence to a maximum photodiode response. With this optimization method the wavefront *rms* deviation from a reference sphere was found to be about $\lambda/8$ [5.22], in which an astigmatic term from the initial mirror figure prevailed. Besides the iteration procedure we always check if the solution results in a far-field diffraction pattern as close to the Airy disk and rings as possible. After optimizing the beam, the sensor stores the average values of spot positions over a large number of readings of the reference wavefront. This is done to eliminate the contribution of detector noise from the reference-position values.

Basis generation

The feedback control is mainly used to discover which voltages should be applied to the adaptive mirror, based on the sensor response, in order to restore the reference wavefront from an aberrated one. In the feedback-loop operation the same reconstruction algorithm as in Section 5.1.4 is used, but here the reconstruction matrix **B** and the vector of coefficients **c** are different. In the present case, the basis is given by the feedback matrix **M**, where the former Zernike derivatives at the centers of the apertures are substituted with the measured spot displacements corresponding to characteristic shapes of the mirror obtained by actuating each actuator independently. Also, the elements of the vector of coefficients **c** are replaced by the control signals $\mathbf{c}_s$ of the mirror. If the wavefront sensor is operated with 64 quad cells, the feedback matrix **M** has 128 x 37 elements and the control vector $\mathbf{c}_s$ has 37 elements. This means that we are dealing with an overdetermined system of linear equations and this confirms that the closest solution can be found by solving equation (5.11).

$$[\mathbf{M}^T\mathbf{M}]^{-1}\,\mathbf{M}^T\,\boldsymbol{s} = \boldsymbol{c}_s\,. \tag{5.11}$$

The mirror surface deflection $\delta d(x,y)$ is proportional to the square of the voltage distribution $V(x,y)$ over the electrode array [5.22, 5.24]. The system of equations we solve is linear, which implies that the total deflection of the mirror is assumed to be a linear combination of the deflections caused by the actuation of individual electrodes. Hence, the unknown quantities in the

system, namely the elements of $\mathbf{c}_s$, consist of the square of the voltages to be applied to the mirror:

$$\boldsymbol{c}_s = [\, V_1^2, V_2^2, \ \ldots, V_k^2, \ \ldots, V_{ch}^2 \,]^T, \tag{5.12}$$

where *ch* is the number of mirror channels.

The reference beam is obtained by applying the optimal set of voltages to the mirror. To generate the basis we apply an additional pre-defined voltage increment to each mirror actuator independently; this results in different deviations introduced to the reference wavefront. The mirror modes can be defined as the particular shapes the mirror surface assumes when the additional voltage is applied to each electrode. The sensor reads the spots displacements corresponding to each of the *influenced* wavefronts, resulting in sets of linearly independent responses that form the basis matrix. During system operation, an arbitrary sensor response can be decoded as a linear combination of the mirror modes (influence functions), and the feedback algorithm finds out which voltages should be applied to the mirror actuators[6] in order to compensate for wavefront distortions, i.e., to bring the spots to their reference positions.

Feedback loop

Now that the sensor is calibrated and the basis is generated the feedback loop can be started, as depicted in Figure 5.20. The sensor samples the wavefront, the displacements of the spots from the reference positions are computed and the feedback algorithm tries to find a set of voltages that -when applied to the mirror- will compensate the incoming wavefront in such a way that the spot positions will come closer to the reference ones. In other words, the system improves the wavefront quality by minimizing the displacement of the associated spots from their corresponding reference positions.

The closed-loop system works in feedback mode, i.e., after an iteration the spot displacements are checked again and alterations to the previous voltages are introduced if necessary. The solution to the system of linear

6. The control voltages to be applied to the mirror are the square root of the values found for the elements of the vector $\mathbf{c}_s$.

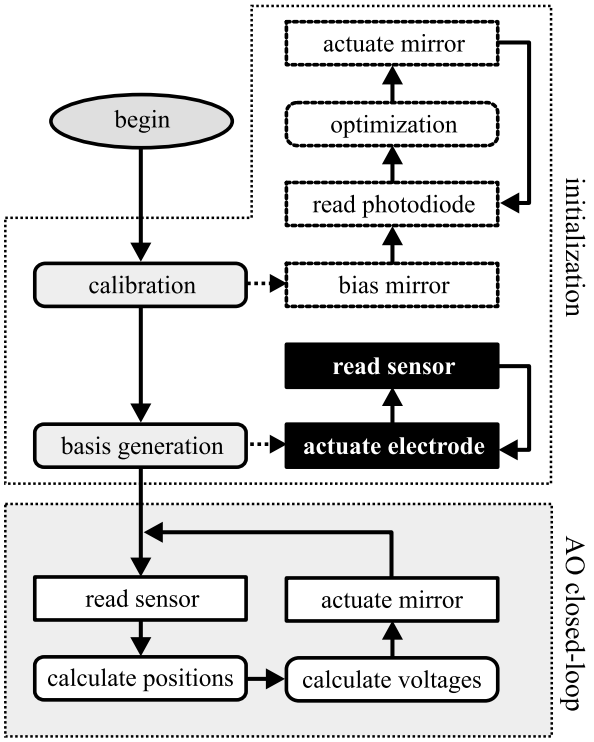

Figure 5.20 *Adaptive optical system operation.*

equations, solved in the feedback algorithm, is bound to diverge if we apply the calculated magnitudes to the control voltages. Therefore, we need to introduce a feedback coefficient that reduces the voltage magnitudes by a given factor. Consequently, the incoming wavefront is only partially compensated after a single iteration. If properly chosen, the feedback coefficient ensures convergence to an optimal correction, which is further limited by sensor noise and fluctuations of the mirror membrane.

5.2.3 Performance

Static correction by the system was tried with compensation for defocus introduced at the aberration plane. The optical aperture is 11 mm in

diameter. The wavefront sensor first reads the reference beam, then defocus is introduced in the system and the sensor reads its magnitude compared to the reference beam; based on the information provided by the sensor the mirror is actuated; the sensor reads the compensated wavefront, which enables the mirror to further correct for residual wavefront deviations from the reference beam. After three iterations in the loop, the initial 0.6-λ *rms* defocus aberration over the full aperture was reduced to 0.08λ over the full aperture and to 0.03λ over 80% of the aperture. The wavefront before and after compensation are shown in Figure 5.21.

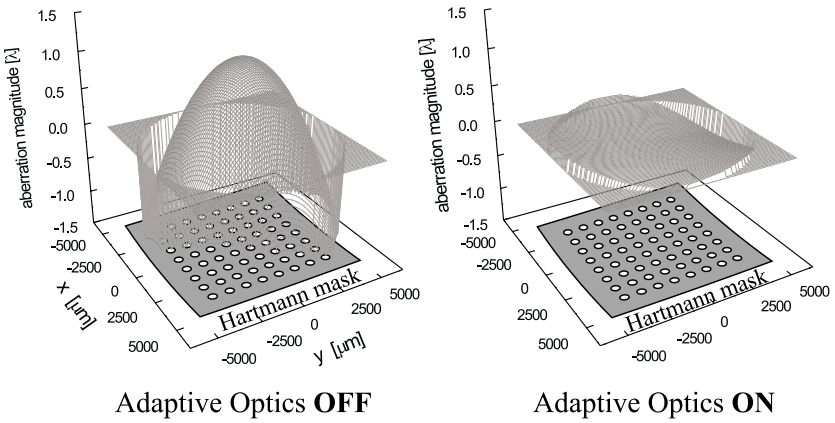

Figure 5.21 *Compensation of a wavefront with defocus aberration.*

Dynamic operation was verified by inserting random time-dependent aberrations with low tilt values and by monitoring the far-field image of the beam reflected by the deformable mirror while in feedback loop with the sensor. We observed a rather stable far-field distribution at the distant plane when the loop was running and a fuzzy distribution when the wavefront sensor was blocked. The beam quality was maintained for the 64-, 52- and 44-QC configurations of the sensor. The quality decreased for the 32-QC and lower configurations, which is explained by the fact that the beam was kept with the 11-mm diameter and, therefore, no information about the

periphery of the wavefront was registered. This can be partially compensated by using an afocal system to shrink the beam to fit the central quad cells better.

The closed-loop frequency of this adaptive optical system is defined as the inverse of the time elapsed in the following cycle: the sensor is addressed, the wavefront reflected by the mirror is sensed, the sensor data

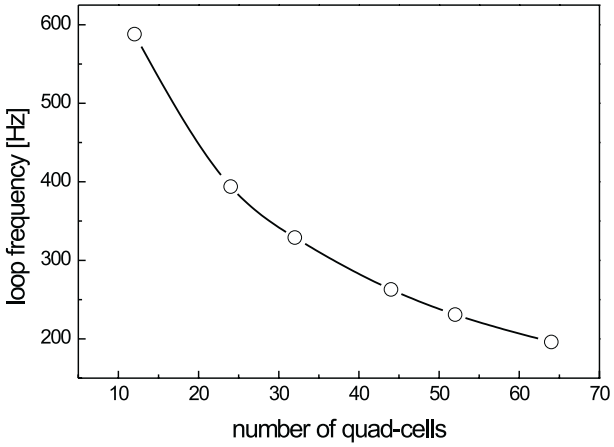

Figure 5.22 *Closed-loop frequency of the AO system.*

is processed, the voltages to be applied to the mirror are calculated, the mirror is actuated and the cycle begins again. This frequency depends on the operational frequency of the sensor, on the mirror settling time and on the number of influence functions, which correspond to the number of actuators, used in the feedback algorithm. The 37-actuator mirror we used has a settling time of 1 ms. The sensor frequency, as discussed before, is mainly limited by the data-acquisition card used in these experiments. The control and feedback algorithms run in a 750-MHz Pentium III, under Linux, and are not of concern compared to the other elements of the system. Figure 5.22 presents the loop frequency for different numbers of quad cells used actively in the sensing chip. The frequency is about 200 Hz when 64 quad cells are used, but it is improved to 260 Hz for 44 quad cells, yielding basically the same wavefront-detection accuracy.

5.3 Chapter summary

This chapter presented the results obtained with the implemented wavefront sensor and discussed an adaptive optical system in which the sensor was coupled to a deformable mirror.

For a compact device, close proximity of the Hartmann mask to the detector is desired, which renders the spot-intensity pattern at the detection plane dependent on the distance as a consequence of Fresnel diffraction. For a mask with 450-μm circular apertures, we found an optimal distance of 6 cm, at which each resulting spot yields a quad-cell position response that can be optimally fit by a sigmoidal curve. The Hartmann mask needs to be well aligned to the detector to minimize systematic errors and to obtain the maximum available spot-displacement range.

The total quantum efficiency for each pixel in the detector at $\lambda = 633$nm is 40%. The efficiency could be improved if the thicknesses of the dielectric layers were accurately controlled, which is usually not the case in standard CMOS, or if an custom antireflective coating were employed. The spectral response of the pixel in the visible range was presented.

In the Hartmann method, an arbitrary wavefront is sampled into light spots, whose centroid positions are used to estimate the wavefront aberrations by means of a reconstruction algorithm. We used the modal reconstruction approach that allows us to approximate an arbitrary aberration by finding the appropriate coefficient values to a set of orthogonal functions, in our case the Zernike polynomials. Noise in the position response propagates through the reconstruction algorithm and results in a reconstruction error. Based on a numerical model of the sensor, we evaluated the error associated with the resolution of the quad cells for different numbers of Zernike terms. The conclusion was that the quad-cell resolution must be better than 10 μm if at least the first nine lower-order terms are to be used.

For 7-μW spots we experimentally found that the quad-cell resolution is 1.4μm, leading to a wavefront accuracy better than $\lambda/10$ for up to 30 Zernike terms. This result was in close agreement with the numerical predictions. It is worth mentioning that we obtained practically the same accuracy with 64 quad cells and with a special configuration of 44 quad cells, the latter being more advantageous as it renders a higher operational frequency. Less intense spots than the ones above can still effect a good

wavefront-reconstruction accuracy if the spot-displacements are averaged. This is done at the expense of operational frequency, however.

The operational frequency of the sensor depends on the sensor read-out frequency, on the interface hardware and on the software routines. The sensor itself can be read out at 3.125kHz. The software routines, including sensor control, data handling, and wavefront reconstruction, consume less than 5% of the total cycle time if a reasonably fast computer is used. The critical part was the data-acquisition card we used, which pushed the frequency down to 260 Hz when all pixels in the sensor were actively used, and to 370 Hz when 44 quad cells were used.

One section discussed a closed-loop adaptive optical system using the implemented wavefront sensor. Adaptive Optics is a growing field of research with a wide spectrum of applications and consists in real-time detection and correction of distorted wavefronts. A typical adaptive optical system contains three main elements: a wavefront sensor, a data processor and a deformable mirror. There is a trend towards realizing inexpensive and fast systems. The implemented CMOS sensor coupled to a low-cost deformable mirror, whose technology was developed some years ago at DIMES, fits this trend. We assembled a feedback-loop system and achieved both static and dynamic wavefront correction.

References

[5.1] E. Hecht, “Optics,” 3rd edition, 167-172, Addison Wesley Longman Inc., Reading, 1998

[5.2] G. E. Forsythe and C. B. Moler, “Computer solutions of linear algebraic systems,” Chapter 17, Prentice-Hall, 1967

[5.3] G. Williams, “Linear algebra with applications,” Chapter 8, Allyn and Bacon, Newton, 1984

[5.4] S. Donati, “Photodetectors - devices, circuits and applications,” 146-153, Prentice Hall PTR, New Jersey, 2000

[5.5] J. H. Huijsing, “Operational Amplifiers,” Kluwer Academic Publ., 2000

[5.6] R. Tyson, “Principles of Adaptive Optics,” 2nd edition, 1-3, Academic Press, San Diego, 1998

[5.7] J. Geary, “Introduction to Wavefront Sensors,” SPIE Press, 1995

[5.8] C. Paterson, I. Munro and J. C. Dainty, “A low cost adaptive optics system using a membrane mirror,” Optics Express, vol.6, no.9, 175-185, 2000

[5.9] G. Love, “Adaptive Optics for Industry and Medicine,” World Scientific Publ. Co., Durham, 2000

[5.10] G. Vdovin, “Adaptive mirror micromachined in silicon,” PhD Thesis, Delft University Press, Delft, 1996

[5.11] G. Vdovin, S. Middelhoek and P. Sarro, “Technology and applications of micromachined silicon adaptive mirrors,” Opt. Eng., vol.36, no.5, 1382-1390, 1997

[5.12] E. Zeek, K. Maginnis, S. Backus, U. Russek, M. Murnane, G. Mourou, H. Kapteyn and G. Vdovin, “Pulse compression by use of deformable mirrors,” Optics Lett., vol.24, no.7, 493-495, 1999

[5.13] R. Bartels, S. Backus, E. Zeek, L. Misoguti, G. Vdovin, I. Christov, M. Murnane and H. Kapteyn, “Shaped-pulse optimization of coherent emission of high-harmonic soft X-rays,” Nature, vol.406, no.7, 164-166, 2000

[5.14] S. C. McQuaid, E. Seibel, R. Burstein, T. Furness III, “Three-dimensional virtual retinal display system using a deformable membrane mirror,” SID 02 Digest vol.50.4, 1-4, 2002

[5.15] E. Fernández, I. Iglesias and P. Artal, "Closed-loop adaptive optics in the human eye," Optics Letters, vol.26, no.10, 746-748, 2001

[5.16] O. Albert, L. Sherman, G. Mourou and T. Norris, "Smart microscope: an adaptive optics learning system for aberration correction in multiphoton confocal microscopy," Optics Letters, vol.25, no.1, 52-54, 2000

[5.17] G. Vdovin and V. Kiyko, "Intracavity control of a 200-W continuous-wave Nd:YAG laser by a micromachined deformable mirror," Optics Letters, vol.26, no.11, 798-800, 2001

[5.18] R. K. Tyson and B. W. Frazier, "Microelectromechanical system programmable aberration generator for adaptive optics," Applied Optics, vol.40, no.13, 2063-2067, 2001

[5.19] L. Zhu, P. Sun, D. Bartsch, W. Freeman and Y. Fainman, "Wave-front generation of Zernike polynomial modes with a micromachined membrane deformable mirror," Applied Optics, vol.38, no.28, 6019-6026, 1999

[5.20] L. Zhu, P. Sun and Y. Fainman, "Aberration-free dynamic focusing with a multichannel micromachined membrane deformable mirror," Applied Optics, vol.38, no.25, 5350-5354, 1999

[5.21] G. Vdovin and P. M. Sarro, "Flexible mirror micromachined in silicon," Applied Optics, vol.34, no.16, 2968-2972, 1995

[5.22] G. Vdovin, "Optimization-based operation of micromachined deformable mirrors," Proc. SPIE, vol. 3353, 902-909, 1998

[5.23] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, "Numerical Recipes in C - The art of scientific computing," 2nd ed., 408-412 Cambrigde University Press, Cambridge, 1992

[5.24] L. Zhu, P. Sun, D. Bartsch, W. Freeman and Y. Fainman, "Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation," Applied Optics, vol.38, no.1, 168-176, 1999

Conclusions **6**

The research addressed in this thesis is aimed at the development of an inexpensive integrated wavefront sensor capable of real-time operation. Preconditions were that it should be simple, compact and compatible with standard IC technologies, while offering prime performance both as a diagnostic tool for wavefront profiling and as a part of an adaptive optical system.

To achieve this we should choose a simple wavefront-sensing method, an optical-detector scheme that operates fast enough and a widely accessible standard fabrication technology. We chose the Hartmann method, a matrix of position-sensitive detectors and a Complementary Metal-Oxide Semiconductor process (CMOS), respectively. However, at the same time that these choices realize the simplicity, cost and speed requirements, they have inherent limitations that have to be taken into account when regarding performance.

The Hartmann method, which is based on the measurement of wavefront tilts, is straightforward and offers good flexibility. Since it is simply based on the direct correspondence of spot displacements to local wavefront tilts, a wavefront can be reconstructed by a linear decomposition of these tilts over a set of basis functions. The choice of detectable spatial frequencies is determined by the density of sampling apertures; the maximum aberration magnitude and sensitivity are selected by the distance

from the Hartmann mask to the detector plane. The main disadvantage of this method is that, besides blocking a considerable amount of light, the wavefront is reconstructed based on the information provided at fixed discrete points, which incidentally do not supply any information about the wavefront edges. When the Hartmann mask is replaced by a microlens array, more light is gathered but local tilts are averaged over the area of each lens, filtering some spatial frequencies anyway.

With the Hartmann method, a matrix of position-sensitive detectors (PSDs) represents an alternative to conventional imagers. Each PSD senses a spot displacement and renders the relevant information to the wavefront reconstruction directly. This circumvents the need for an image processing step, inherently associated with the output of an imager (e.g. CCD). The drawback of this approach is that once the number of integrated PSDs and their grid architecture are chosen, the geometry of the Hartmann mask is fixed as well. Also, the size and the type of each PSD impose limitations on the spot size and on the maximum displacement range; however, this is also true when an imager is used.

As regards technology, standard CMOS is a continuously improving, largely available and cost-effective fabrication process, and it offers the possibility of integration of multiple analog and digital functions on chip. It is, however, targeted at microelectronics circuitry and not optimized for photodetection, in contrast to CCD technology. Also, state-of-the-art process improvements concentrate on scaling down the minimum feature size, which implies technological modifications not favorable to photodetection and challenges analog applications. Therefore, the use of conservative (*supra*-micron) CMOS processes is advised whenever possible.

The wavefront sensor we developed consists of a high-precision Hartmann mask with 64 sub-apertures in an orthogonal grid and a 1 x 1 cm CMOS chip with one quad cell per aperture of the mask. The quad cells are rather large (600 x 600 μm) and therefore feature insignificant mismatch over the array. They have random-address capability and their signals are transferred to the output data bus by means of an integrated analog demultiplexer that decodes an input digital address provided by the computer.

The responses of the quad cells to the spots positions serve as the input data to a modal-reconstruction algorithm that approximates the sampled

wavefront with a set of Zernike polynomials. The number of Zernike terms limits the wavefront spatial frequencies that can be correctly reconstructed; for this sensor, the algorithm we use converges if less than 128 terms are used. In addition, position-response noise propagates through the reconstruction algorithm, therefore the more Zernike terms, the larger the error introduced to the reconstructed wavefront. This limits even further the maximum number of terms to be used in the wavefront approximation. For 7-μW spots with a 300-μm effective diameter, the position resolution of the quad cells is 1.4-μm, which results in a *rms* wavefront accuracy of λ/50 (λ=633nm) when the 9 lower-order Zernike terms are used. Wavefronts can be reconstructed with an accuracy better than λ/10 with up to 30 Zernike terms. It is worth mentioning that when we used 44 quad cells, the accuracy results were practically the same as with the complete set of 64 quad cells. Lower light levels yield a poorer position resolution, which can be compensated for to a certain extent by averaging the coordinate signals over time, at the expense of the operational frequency.

The sensor can be operated at a 3.125-kHz frequency, which is in our case mainly restrained by the settling time of the analog switches. Less than 150 μs are necessary to handle the sensor raw data and to reconstruct the wavefront. However, we used a 400-dollar commercial data-acquisition card to interface the sensor signal to the computer, which provides the chip with the digital address, samples the sensor output signals, performs the analog-to-digital conversions and feeds the data into the computer. This interface card brings the operational frequency down to 370 Hz when 44 quad cells are actively used in the wavefront-sensing process. This speed issue is solely attributed to off-chip hardware and can be solved, at a first instance, by either employing a more adequate data-acquisition card or by designing a custom card compatible with the sensor.

A wavefront sensor is an invaluable element in an adaptive optical system. Adaptive Optics consists of the real-time detection and correction of distorted wavefronts and has an expanding spectrum of applications, extending from astronomy to ophthalmology and from laser-beam optimization to optical communications. A typical adaptive optical system contains three main elements: a wavefront sensor, a data processor and a deformable mirror. The CMOS sensor we developed and the low-cost deformable mirror are in tune with the trend to implement inexpensive and fast systems. We built a feedback loop system and were able to perform

both static and dynamic wavefront correction. When 44 quad cells were actively used, the loop frequency was limited to 260 Hz by the data-acquisition card, which can be improved if a more adequate card is used. A summary of both the wavefront-sensor and adaptive-optical-system parameters is presented in Table 1 at the last page of this chapter.

This first prototype of a CMOS-based integrated wavefront sensor fulfilled the preliminary requirements imposed at the beginning of this research project. It is simple to understand and operate, compact, fully compatible with the most widely used technology today (CMOS) and capable of real-time operation. As regards performance, each particular application has its own requirements; the sensor in its current form should be used with light sources with wavelengths in the visible spectrum and shall perform well for applications involving enough light (>0.5mW beams) and there where lower-order wavefront aberrations are expected (<30 Zernike terms).

Future versions of this sensor can be implemented in many different directions and will probably be specialized for a given application. In general, modifications will come with trade-offs either in terms of cost or complexity. Some suggestions for possible further actions are listed below:

- improvement of the wavefront accuracy for low-light-level applications (e.g. astronomy and ophthalmology);
- capability to handle higher spatial frequencies;
- feasibility of operation with pulsed radiation;
- more functionality integrated on chip;
- operation in the infra-red part of the spectrum;

Improving the wavefront accuracy for lower light-level applications means that the signal-to-noise ratio has to be improved. This can be done by enhancing the quantum efficiency and minimizing the noise. The Hartmann mask can be substituted by a microlens array to gather more light per spot. In addition, due to the poor thickness tolerance of the dielectric layers in CMOS, these layers should be removed from the photosensitive area and an anti-reflective coating should be applied. Passive-pixel quad cells can be replaced by a combination of active-pixel quad cells and a correlated double-sampling circuitry. Also, direct cooling of the device with a Peltier element can be used to reduce the dark current, but this

demands appropriately sealed packaging to avoid surface fogging. Another alternative is to move from standard CMOS to a custom design in either an CMOS process optimized for imaging or a combined CMOS/CCD process.

To be able to detect higher spatial frequencies accurately, we need to use more sampling points, which means that the design should be changed to accommodate a higher density of quad cells. However, for a fixed chip size, increasing the number of quad cells limits the area of each quad cell and consequently the spot-displacement range. Additionally, a larger number of quad cells results in a longer time to read the wavefront sensor signals.

The present design does not allow operation with a pulsed light source because the photodiode starts accumulating the signal only when the pixel is addressed. With a pulsed source, all pixels should store the instantaneous signal and later transfer them sequentially to the output node(s). A step towards it consists in eliminating the off-line discharge switches in each pixel.

The sensor can also be made more compact and autonomous by integrating operational amplifiers, analog-to-digital conversion circuitry and centroid-calculation circuitry on chip. Besides increasing the level of integration, this reduces the capacitances associated with the bond pads and transmission lines.

Operation in the infrared can be suitable, for instance, for optical communications. Either an approach using thermal detection with bolometers or the use of a dedicated semiconductor technology can be considered.

In conclusion, the developed wavefront sensor was successful in attaining the project goals initially proposed and can be further modified in a variety of ways to attend to particular interests.

Integrated wavefront sensor			
technology	analog CMOS 1.6μm	*detector η*	40%
chip area	10 x 10 mm	*dark-current density*	1800 pA/cm^2 @300K
optical aperture	∅ ≤ 11mm , depends on the QCs used	*dark-current shot noise*	0.16 pA/pixel @300K
PSD type	quad cell	*spectral response*	from 400 to 700nm see Figure 5.4
grid	orthogonal, 1-mm pitch	*capacitance*	50 pF / photodiode
# quad cells	64 (256 pixels)	*settling time*	10 μs, analog switches
pixel type	passive, double-junction photodiode	*reconstruction algorithm*	modal, Zernike polynomials
pixel area	~ 0.3 x 0.3 mm	*max. # slopes*	128 (x and y directions)
random access	yes	*maximum local slope*	can be adjusted, see eq.(2.13) and Figure 2.9
on-chip electronics	analog demultiplexer sampling-rate controller	*intrinsic frequency*	~ 3 kHz (64 QCs)
# parallel outputs	4 or 8	*extrinsic frequency*	currently limited by DAq card, (Figure 5.13)
off-chip electronics	signal amplification data-acquisition card Pentium III 750 MHz	*QC position resolution*	1.4 μm for a 7-μW spot, sigmoidal fit, see Figure 5.3
light source	He-Ne laser λ=633nm CW 5mW	*wavefront accuracy*	better than λ/10 for up to 30 Zernike terms, see Figure 5.8/ Figure 5.9
Adaptive optical system			
adaptive corrector		deformable mirror, 37-ch, ϕ = 15 mm	
system operational mode		feedback loop	
reference-beam optimization		simplex algorithm, far-field mirror image	
wavefront restoration control		modal, mirror influence functions	
closed-loop frequency		see Figure 5.22, limited by DAq card.	

Table 1: Summary of wavefront-sensor and adaptive-optical-system parameters.

Summary

This thesis addresses the design, implementation and performance of an integrated Hartmann-Shack wavefront sensor suitable for real-time operation and compatible with standard CMOS technology. The contents of the six constituent chapters are summarized below.

Chapter 1 - Introduction

Chapter one investigates several optical profile tests that rely on different methods, each with its particular strengths and limitations. We pointed out various desirable features expected from optical profiling, and for the tests, we indicated the lack of an inexpensive compact device capable of real-time operation and quantitative analysis. We propose the implementation of a Hartmann wavefront sensor that uses a matrix of integrated position-sensitive detectors (PSDs), compatible with standard Complementary Metal-Oxide-Semiconductor technology (CMOS). A wavefront is sampled into a number of light spots, whose displacements are proportional to local tilts of the wavefront. The idea is to have one PSD per sampled spot, such that direct information about the spot-centroid position is available. This scheme renders faster operation than when a conventional imager is used because it circumvents the image-processing step.

Chapter 2 - Theoretical concepts

This chapter provides the necessary background material on the various subjects discussed throughout this thesis. First the fundamentals of wavefront description are addressed, where the Zernike polynomials are presented as a convenient set of orthogonal functions for the description of aberrations over a circular aperture. Then, the Hartmann method is investigated more closely, followed by a discussion on how a wavefront can be numerically reconstructed from the measured light-spot positions resulting from the sampled wavefront.

There is a section on photodetection that reviews some basic parameters used to characterize a photodetector and that lists the most common photodetectors in silicon. Possible noise sources affecting the performance of a photodetector are also concisely discussed as well as a photodiode equivalent model.

Both one- and two-dimensional optical position-sensitive detectors (PSDs) are considered along with their important figures of merit. Basically, a convenient signal swing and linearity of the PSD response over a wide position range are desired features. Finally, we comment on the current-division centroiding method for multipixel PSDs, which attempts to reproduce the idea of the lateral-effect photodiode (LEP) in a discrete way.

Chapter 3 - Technological aspects

Chapter three is dedicated to the technological aspects concerning the fabrication of the Hartmann wavefront sensor. Although not optimized for imaging applications, CMOS was the standard technology of choice for several reasons, among which the most important were its availability and possibility of integration of multiple electronic functions on chip. Two alternative ‘standard’ options for the development of this sensor would be the Bipolar and the CCD technologies. However, if more and more minimum-sized mixed-signal electronic circuitry is to be integrated in newer versions of the sensor, none of these two is a judicious choice.

As the sampling plane of the sensor, the Hartmann mask and a microlens array are considered. The former is fundamentally cheaper but blocks a considerable amount of incident light, whereas the latter is usually more expensive but when it is a close-packed structure, it features a higher

light-collection efficiency. We recently managed to fabricate an inexpensive microlens array that represents a promising alternative to the mask we used throughout this research project.

This chapter also discusses a number of feasible photodetector structures in standard CMOS technology. Passive- and active-pixel architectures are also investigated. The quantum efficiency of a photodiode is determined by intrinsic and extrinsic parameters, discussed in the text. Intrinsic parameters offer little flexibility to play with them in standard CMOS. On the other hand, extrinsic parameters can, to a certain extent, be tailored by the addition of post-processing steps and by design.

The rapid scaling of CMOS processes often results in disadvantages for photodetection. The tendency to have devices closer to the surface reduces the quantum efficiency for longer wavelengths; the tendency to use a larger number of metal and dielectric layers results in blockage of off-normal light incidence on small pixels, while the usage of a lower voltage supply restricts the output signal swing of pixels.

Chapter 4 - Sensor implementation

Chapter four presents the development of the Hartmann wavefront sensor. It features an overview of the standard CMOS technology we used and includes the implementation and the results of several independent position-sensitive structures realized before the implementation of the complete sensor prototype. The chapter describes the developed wavefront sensor prototype, including a note on the fabrication of the Hartmann mask and a detailed review of the implementation of the elements integrated on chip as well as a description of the hardware used as the interface to the computer.

All detector structures were fabricated in a standard 1.6-μm n-well CMOS process available at DIMES. A Hartmann mask with sixty-four 450-μm circular apertures in a orthogonal grid was designed and fabricated.

As a preliminary step in the development of the wavefront sensor we designed a number of PSD structures in the chosen CMOS technology in order to evaluate their performance. The chosen PSD was the quad cell with a well-defined geometrical center, simple layout, high fill factor and the presence of only four output nodes, which renders the use of resistive

arrays unnecessary. Although its response is only linear for spot displacements around the very center of the device, in most cases, the response curve can be approximated by a sigmoidal curve.

Our first prototype of the proposed CMOS-based wavefront sensor consists of a Hartmann mask positioned at 6 cm from the chip, which has 64 quad cells that can be randomly addressed. In this first approach, passive-pixel quad cells were chosen for their simplicity. An integrated analog demultiplexer decodes the address provided by the computer and transfers the corresponding quad-cell signals to the data bus. The diagrams of the various circuit stages integrated on chip are displayed and commented throughout this chapter.

A printed circuit board was designed to accommodate the CMOS chip and the operational amplifiers used to convert and amplify the input photocurrents into signal voltages, which are sampled by a commercial data-acquisition card, which performs analog-to-digital conversion and feeds the data to the computer.

Chapter 5 - Measurement results

This chapter presents the results of the implemented wavefront sensor and discusses an adaptive optical system, in which the sensor is coupled to a deformable mirror.

The effects of mask-to-detector distance and alignment are discussed. The pixel spectral response and operational speeds are also presented. As regards the wavefront reconstruction algorithm, the effect of the quad-cell position noise on the reconstruction accuracy is estimated. Experimental results on the wavefront-reconstruction accuracy are in close agreement to the numerical predictions. Faint spots can result in a good accuracy if the spot displacements are averaged. This is done at the expense of operational frequency, however.

The operational frequency of the sensor depends on the sensor read-out frequency, on the interface hardware and on the software routines. The sensor itself can be read out very fast. However, the data-acquisition card reduced the frequency by a factor ten.

We also show the feasibility of realizing a closed-loop adaptive optical system in feedback mode. We performed both static and dynamic wavefront correction.

Chapter 6 - Conclusions

Chapter six offers the conclusions concerning the research presented throughout this thesis.

The developed wavefront sensor proved to meet all the initially proposed prerequisites: it is capable of real-time operation, compact, simple and compatible with standard technology. In terms of performance, each particular application has its own requirements and the developed sensor, which should be used with light sources in the visible spectrum, shall perform well for applications involving enough light (>0.5 mW beams) and there where lower-order wavefront aberrations are expected (<30 Zernike terms). The sensor can be modified in a variety of ways to attend to particular applications.

Summary

Samenvatting

Dit proefschrift behandelt het ontwerp, de implementatie en prestatie van een geïntegreerde Hartmann-Shack golffront sensor die geschikt is voor real-time werking en compatibel is met standaard CMOS technologie. De inhoud van de zes hoofdstukken zijn hieronder samengevat.

Hoofdstuk 1 - Introductie

Hoofdstuk één onderzoekt verschillende optische-profieltests die gebaseerd zijn op verschillende methoden, elke met hun eigen sterke en zwakke kanten. We gaven aan welke gewenste resultaten we verwachten van de resulterende optische profielen, en wat betreft de test constateerden we dat er gebrek is aan een goedkoop compact apparaat dat real-time kan functioneren en kwantitatieve analyses oplevert.

Wij stellen de implementatie van een Hartmann golffrontsensor voor die gebruik maakt van een matrix van geïntegreerde positie-gevoelige detectoren (PSDs) die compatibel zijn met standaard "Complementary Metal-Oxide-Semiconductor technologie" (CMOS). Een golffront wordt gesampled in een aantal lichtpunten, wiens verplaatsingen proportioneel zijn aan de lokale hoeken van het golffront.

Het idee is om één PSD te hebben voor elk gesampled punt, zodat er directe informatie over de middelpuntspositie beschikbaar is. Door deze benadering is de sensor sneller dan een conventionele afbeelder omdat er

de beeldverwerkingsstap kan worden overgeslagen.

Hoofdstuk 2 - Theoretische concepten

Dit hoofdstuk geeft de nodige achtergrondinformatie over de verscheidene onderwerpen die door dit proefschrift heen behandeld zijn. Allereerst worden de basisprincipes van de golffrontsensor behandeld, waar de Zernike polynomen worden geïntroduceerd als een handige set van orthogonale functies voor het beschrijven van golffront aberraties over een cirkelvormige opening. Vervolgens wordt er dieper ingegaan op de Hartmann-methode, waarna we bespreken hoe een golffront numeriek gereconstrueerd kan worden vanuit de gemeten posities van lichtpunten die resulteren uit het gesamplede golffront.

Er is een gedeelte over fotodetectie, waarin een aantal basisparameters worden besproken die gebruikt worden om een fotodetector te karakteriseren en waarin de meest gebruikelijke fotodetectoren van silicium worden genoemd. Mogelijke ruisbronnen die de prestaties van een fotodetector kunnen beïnvloeden, worden nader beschreven, evenals een equivalent circuit voor een fotodetector.

Zowel één- als twee-dimensionale optische positie-gevoelige detectoren (PSDs) worden beschouwd samen met hun meest belangrijke parameters. De gewenste eigenschappen waar we ons op richten zijn een bruikbaar signaalbereik en een lineariteit van de PSD respons voor een groot positiebereik. Tot slot geven we commentaar op de stroom-splitsings ‘centroiding’ methode voor multipixel PSDs, welke het idee van de laterale-effect fotodiode (LEP) op een discrete manier probeert na te bootsen.

Hoofdstuk 3 - Technologische aspecten

Hoofdstuk 3 is gewijd aan de technologische aspecten van het fabriceren van de Hartmann golffront sensor. Hoewel CMOS niet is geoptimaliseerd voor afbeeldingsapplicaties, was het wel de standaardtechnologie waar onze voorkeur naar uitging om verschillende redenen; de belangrijkste zijn haar beschikbaarheid en de mogelijkheid van het integreren van verschillende elektronische functies op één chip. Twee alternatieve 'standaard' opties voor de ontwikkeling van deze sensor zouden Bipolair en CCD technologieën zijn. Echter, omdat men de mogelijkheid wil hebben

om in nieuwere versies van de sensor meer elektronische circuits van minimale grootte waarin analoog en digitaal signalen worden gebruikt, te integreren, zijn geen van deze twee een geslaagd alternatief.

Wat betreft de 'sampling plane' van de sensor wordt het gebruik van ofwel een Hartmann-masker ofwel een microlens array overwogen. De eerste is fundamenteel goedkoper, maar houdt een behoorlijk groot gedeelte van het inkomende licht tegen, terwijl de tweede normaal gesproken duurder is maar in een structuur met hoge dichtheden juist een hogere lichtverzamelings-efficiëntie heeft. Recentelijk zijn we erin geslaagd om een goedkope microlens-array te fabriceren dat een veelbelovend alternatief zal vormen voor het masker dat we gedurende het onderzoeksproject hebben gebruikt.

Dit hoofdstuk behandelt ook een aantal mogelijke fotodetector structuren in standaard CMOS technologie. Passieve- en actieve-pixel architecturen worden ook onderzocht. De quantum-efficiëntie van een fotodiode wordt bepaald door intrinsieke en extrinsieke parameters. Intrinsieke parameters bieden weinig speelruimte in standaard CMOS. Aan de andere kant kunnen extrinsieke parameters, tot op zekere hoogte, worden aangepast door de toevoeging van 'post-processing' stappen en door een ander ontwerp.

De voortdurende schaling van CMOS processen is vaak nadelig voor fotodetectie. De neiging om devices dichter bij het oppervlak te hebben vermindert de quantum-efficientië voor langere golflengtes; het gebruik van een groter aantal metaal- en diëlektrische lagen blokkeert de niet-normale lichtsinval voor kleine pixels en het gebruik van een lagere voedingspanning beperkt het bereik van uitvoersignaal van pixels.

Hoofdstuk 4 - Sensor implementatie

Hoofdstuk vier gaat over de ontwikkeling van de Hartmann golffrontsensor. Het bevat een overzicht van de standaard CMOS-technologie die we hebben gebruikt en ook de implementatie en resultaten van verschillende onafhankelijke positie-gevoelige structuren die we realiseerden vóór de implementatie van het complete sensorprototype. Het hoofdstuk beschrijft het ontwikkelde golffrontsensor prototype, waaronder een opmerking over de fabricatie van het Hartmann-masker en een gedetailleerd beschrijving van de implementatie van de op één chip

geïntegreerde elementen en ook een beschrijving van de hardware die gebruikt is als interface met de computer.

Alle detectorstructuren zijn gefabriceerd in een standaard 1.6 μm n-well CMOS proces beschikbaar bij DIMES. Een Hartmann masker met 64 450-μm cirkelvormige openingen in een orthogonaal grid werd ontworpen en gefabriceerd.

Als een voorbereidende stap in de ontwikkeling van de golffrontsensor hebben we een aantal PSD structuren ontwikkeld met de gekozen CMOS technologie om zo hun prestaties te evalueren. De gekozen PSD was de quadcell met een goed-gedefinieerd geometrisch middelpunt, simpele layout, hoge vulfactor en de aanwezigheid van slechts vier uitvoer punten, hetgeen het gebruik van weerstandsarrays overbodig maakt. Hoewel de respons alleen lineair is voor puntsverplaatsingen rondom het echte midden van het quadcell, kan men de responscurve in de meeste gevallen benaderen met een sigmoïdale kromme.

Ons eerste prototype van de voorgestelde CMOS-gebaseerde golffrontsensor bevat een Hartmann masker dat 6cm van de chip is gepositioneerd. Deze heeft 64 quadcells die willekeurig benaderd kunnen worden. In deze eerste benadering werd voor ‘passieve-pixel’ quadcells gekozen omwille van hun eenvoud. Een geïntegreerde analoge demultiplexer decodeert het adres zoals die word gegeven door de computer en zet de corresponderende quadcell-signalen op de data bus. De stages van de circuits geïntegreerd op chip worden getoond m.g.v. diagrammen en beschreven in dit hoofdstuk.

Een ‘printed-circuit board’ is ontworpen voor het behuizen van de CMOS chip en de operationele versterkers die de fotostromen omzetten in spanningssignalen en versterken; deze signalen worden dan gesampled door een commerciële data-acquisitie kaart, die de analoog-naar-digitaal omzetting regelt en de informatie doorstuurt naar de computer.

Hoofdstuk 5 - Meetresultaten

Dit hoofdstuk geeft de resultaten van de geïmplementeerde golffrontsensor en behandelt een adaptief optisch systeem, waarin de sensor gekoppeld is aan een vervormbare spiegel.

De effecten van masker-naar-detector afstand en uitrichting worden beschreven. Daarnaast worden de spectrale respons van de pixel en

operationele snelheden gegeven. Voor het golffront-reconstructie-algoritme schatten we welk effect ruis in de positie van de quadcell heeft op de nauwkeurigheid van de reconstructie. De experimentele resultaten van de nauwkeurigheid van de golffront-reconstructie zijn in goede overeenstemming met de numerieke voorspellingen.

Vage punten kunnen resulteren in een goede nauwkeurigheid als de puntsverplaatsingen worden gemiddeld. Dit wordt echter gedaan ten koste van operationele snelheid.

De operationele frequentie van de sensor hangt af van de uitlees frequentie, van de interface hardware en van de software routines. De sensor zelf kan heel snel worden uitgelezen. Echter, de data-acquisitie kaart reduceert deze frequentie met een factor tien.

We laten ook de haalbaarheid zien van een gesloten-lus teruggekoppeld adaptief optisch systeem. We hebben zowel statische als dynamische golffront correctie bestudeerd.

Hoofdstuk 6 - Conclusies

Hoofdstuk zes geeft de conclusies betreffende het onderzoek, welke op verschillende plaatsen in dit proefschrift worden gegeven.

De ontwikkelde golffrontsensor bleek aan alle in het begin voorgestelde eisen te voldoen: hij kan real-time functioneren, is compact, eenvoud en compatibel met standaardtechnologie. Wat betreft prestaties, heeft elke specifieke applicatie zijn eigen eisen en de sensor in de huidige vorm, die moet gebruikt worden met lichtbronnen in het zichtbare spectrum, zal goede prestaties leveren voor applicaties waar voldoende licht aanwezig is (>0.5 mW licht bundels) en waar lagere-orde golffront aberraties worden verwacht (<30 Zernike-termen). De sensor kan op een aantal manieren worden aangepast aan de behoeften van specifieke applicaties.

Samenvatting

Acknowledgments

The transition from Physics to Engineering, and more specifically from intermetallic magnetic alloys to adaptive optics and microelectronics was not the easiest task. To succeed in this new field I needed the solicitude, patience and support of many other people, who in many ways and at various stages contributed to the completion of the project presented in this thesis. I express my gratitude to all those who have cooperated with me and in the following paragraphs I mention some of them.

First of all I thank my late grandfather, Vô Lusitano, who was always ready to listen and motivate me in the pursuit of my goals throughout the overlapping parts of our lives, and my grandmother, Vó Alahyma, for her caring love. I also thank my parents, Deni and Terêsa, who provided me with all the best they could and who realized how important a foreign language and a good education can be. I thank my father for being an example of determination and my mother for being in tune with my thoughts. I thank my brother, Jean Davies, for his humor and respect for me, and apologize to him for having been so absent due to my dedication to my own studies. I am grateful to all relatives who cared about me.

I am thankful to my wife, Ruth Stefanie, who gave up her valuable job in Brazil and supported our coming to The Netherlands, exchanging warmth for coldness. She has managed to run our family, in spite of postponing her

career and of being away from her parents, brother and some friends in demanding times. A great deal of my thanks I give to our daughter, Moira Ellen, who has brought a lot of joy to our lives and who is full of happiness and inspiring creativity. Her participative character and unconditional love has made me able to manage in hard times.

I thank Dr. Gleb Vdovin for being my daily mentor. He has been straight to the point from our first e-mail contact in 1995 to the time of writing this thesis, which he carefully read and commented. He proposed the topic of this project and swiftly introduced me to adaptive optics and microelectronics. He taught me how to think in engineering and pragmatic terms and developed trust in my competence. The absence of strict consultation hours and his keen mind contributed together with numerous discussions and insights to the realization of this project.

I also thank Professor Paddy French for the informal work environment he established and his friendly and understanding guidance. He is a serious and hard-working professional with an easygoing character, whose leadership style I appreciate: demanding but cooperative and respectful.

Despite the short notice, Professor Middelhoek and Dr. Reinoud Wolffenbuttel created slots in their agendas, at Gleb's request, and received me with a warm welcome during my first informal visit to Delft in 1996. I thank them for the uplifting feeling I had when I left Delft, although at that time I still pondered whether I, as a physicist, would fit an engineering scenario.

Shortage of proper housing facilities in Delft and its surroundings is an increasing problem for many, and especially for those coming from abroad with a family. My wife and I thank our friends Jacolien Roos and Robbert van den Nieuwenhof, who made their house ours for over three months upon our arrival in The Netherlands. They also provided us with very nice company and helped us with lots of practicalities, which included filling in countless forms in Dutch.

I am grateful to all staff of the Electronic Instrumentation Laboratory for having helped me when I needed it, and a couple of these people I need to mention in particular. Ger de Graaf and Piet Trimp got me through many

new tools of the trade and gave me an invaluable help in internal and external issues. Inge Egmond and Evelyn Sharabi patiently solved most of my doubts concerning administration and Willem van der Sluys carefully assessed all financial matters with a sympathetic smile. At early stages of this project, when I barely understood the jargon of electronics and silicon technology, Wilko Kindt and José Higino Correia offered many hints about sensors and basic electronics. Equally, I thank Serhat Sakarya for having shared his expertise with linux and C programming in the beginning, even when he could hardly understand how someone could not know what a command like `outportb(base+1)`was. He was good company during a couple of conferences we attended together and even under the pressure to finish his own thesis he kindly translated the summary of this thesis to Dutch. Maureen Meekel, Jeroen Bastemeijer and Jeff Mollinger helped me with software in general, the early stages of design and with the connections with the Delft Technology Center.

I thank my room-mates for the cooperative atmosphere and exchange of remarks and thoughts. Ulrike Dauderstädt was always ready to help when I arrived. Fredrik Creemer was the room-mate with whom I had the longest contact and who gave many hints about the Dutch way. We had many conversations about a large range of topics. Dilaila Criado offered me the chance to get a glimpse of the fast-spoken Cuban Spanish. Mikhail Loktev is an experienced and talented person who gave me invaluable help in developing programming skills and optical setups. Gabriel Craciun is a solicitous colleague with many experiences in life. Oleg Soloviev is an obstinate and inquisitive individual, whose charisma and sharp humor favored a very good fellowship.

My thanks are extended to José Gerardo Rocha, who cooperated in some numerical simulations, Ventzeslav Iordanov for his involvement in the photodiode characterization and Ourang Akhzar-Mehr, for having agreed to carry out the etch steps of the microlens-template project, which was delayed for three years due to lack of time. I also thank Armando Vilaça, whose doubts during his Master's project made me review my understanding of several topics. He was extremely useful in helping in the development of the closed-loop adaptive optical system.

Acknowledgments

In the beginning, useful hints were provided by Fabiano Fruett and Daniel Rocha, with whom and Heino Bukkens and Paul van der Meer I had many delicious home dinners after enjoyable squash matches. Some colleagues at work offered outstanding cooperation and friendship. Vladimir Kutchoukov is a determined and dedicated friend who many times came with nice suggestions and solutions, despite having managed to crash my computer a couple of times. Dafina Tanase is an eloquent, objective and joyful lady, with whom I have exchanged many experiences about life and about changing research fields. Miodrag (Miki) Djurica is a competent and easygoing friend, who despite the hectic pace of his life was always very solicitous in helping me with circuit simulations and design issues. Although we do not meet very often, when we do, we enjoy hours of good conversation. Graça Minas is a sympathetic and friendly person who besides the many talks in the coffee breaks provided my family and me with pleasant pizza evenings. With Manel Gasulla I exchanged several work experiences and also hard balls in squash games. He and his family, Mari and Rita, also contributed to several enjoyable moments with my family.

Many thanks to Professor Albert Theuwissen for his useful hints and for the very well organized and clear course on CMOS image sensors, which structured my knowledge on the subject. I thank Professor Lina Sarro too, for her interest, involvement with the project and nice conversations.

I am grateful to professors Carlos Larica and Klinger Barbosa Alves for having supported my decision to go abroad for a Phd in microelectronics, even though, after dedicated assistance, it represented one professional less in the field of magnetic materials.

I thank my very good friend Edward A. Baptista for being a true friend for over fifteen years, despite the fact that for about ten years either he, or I, or both have been working thousands of kilometers away from Brazil and from each other.

Thank you Fábio and Elaine Tagliaferro, for the very good time we had together in The Netherlands and for the friendship we developed and maintained, having matched so well in character, goals and attitudes. I also thank our friends Mauro, Liana and Daniel Ferreira for the many weekends

in common. Mauro, as a theoretical physicist and eclectic researcher, offered a competent listening ear to my research babble. We also spent many hours discussing books of many kinds.

I thank all members of the Delft Institute for MicroElectronics and Submicron technology (DIMES), especially Hugo Schellevis for the data he provided about the CMOS process and Wim van der Vlist for the wire bonding of the chips.

My father-, mother- and brother-in-law, Norberto, Micaela and Friedemann Berger, respectively, have taken me on as part of their family, with whom a very pleasant relationship exists and I do appreciate it. With my mother-in-law I have always had meaningful talks about life style, Nature and well being; soothing topics that make me feel good.

Evenly important are our friends, Ton, Margje, Marjon, Jacolien and Mário Roos; Russell Konkie Halmeyer; Robbert and Anne van den Nieuwenhof. They have considered us as true part of their family, with whom we spent many happy holidays and weekends. They have always assisted us, both in joyful and difficult times and I am thankful to them for having encouraged me in the journey towards the completion of this thesis. With all of them I had moments of reflection, but in particular with Russell I shared many moments of fine-tuned discussion about the actual essence of life and the meanings of feelings and reasoning.

Delft University of Technology financially supported me for three years. The Dutch Technology Foundation (STW) has financially taken up this project (DOE.5375) from January 2001 on. I thank the Program Officer of the foundation, Dr. Leo Korstanje, and his assistant, Ms. Annemiek Peeters, for their involvement and organization.

I am thankful to Mirjam Nieman for having revised the language of the manuscript and for her good-humored remarks. She is a delicate person with a very professional and human attitude, whose observations certainly improved the readability of some parts of this thesis.

This part of the thesis was one of the most enjoyable to write, especially because it gave me the opportunity to think over the importance of the

many people who contributed in so many ways to the realization of this goal. Sometimes it is said that the realization of a thesis is a rather individualistic job. In my case I wonder how could it be so with so many people in the background! I thank you all.

List of publications

Publications and presentations

- D. W. de Lima Monteiro, G. Vdovin and P. M. Sarro, “Technological aspects of a custom CMOS sensor for Adaptive Optics,” to be published in the Proc. of the 4th Photonics Prague Conference, Prague - Czech Republic, 2002. *Oral presentation*.
- D. W. de Lima Monteiro, G. Vdovin and M. Loktev, “Customized CMOS imager as wavefront sensor,” Proc. of Semiconductor Sensors Conference - SeSens, Veldhoven - The Netherlands, 821-824, 2001. *Flash oral presentation and poster presentation (Best Flash Presentation Award)*.
- D. W, de Lima Monteiro, G. Vdovin and M. Loktev, “IC compatible optical sensor for Adaptive Optics,” Proc. IEEE International Microwave and Optoelectronics Conference, Belém - Brazil, 513-516, 2001. *Oral presentation*.
- D. W. de Lima Monteiro, G. Vdovin, J. Rocha, V. Iordanov, M. Loktev and P. M. Sarro, “Customized CMOS wavefront sensor,” Proc. SPIE vol. 4493, 88-99, San Diego - USA, 2001. *Oral presentation.*
- D. W. de Lima Monteiro, A. Vilaça, G. Vdovin, M. Loktev and P. M. Sarro, “Integrated Hartmann-Shack wavefront sensor,” Proc. of the 3rd

Adaptive Optics Conference for Industry and Medicine, 179-183, Albuquerque - USA, 2001. *Oral presentation*.

- D. W. de Lima Monteiro, G. Vdovin and M. Loktev, “CMOS integrated wavefront sensor,” Proc. of the 11th Transducers/ Eurosensors XV, 532-535, Munich - Germany, 2001. *Oral presentation*.
- D. W. de Lima Monteiro and G. Vdovin, “Custom imager as wavefront sensor,” Compilation of the IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, 114-117, Lake Tahoe - USA, 2001. *Flash oral presentation and poster presentation (Best Poster Award).*
- D. W. de Lima Monteiro and G. Vdovin, “High speed integrated wavefront sensor,” Proc. of the Sensor Technology Conference, 37-42, Enschede - The Netherlands, 2001. *Oral presentation.*
- D. W. de Lima Monteiro, G. Vdovin, “CMOS integrated Hartmann-Shack wavefront sensor,” Adaptive Optics Symposium, Murcia - Spain, 2000. *Poster presentation.*
- D. W. de Lima Monteiro, G. Vdovin, P. M. Sarro, “Position-sensitive detectors for a wavefront sensor,” Proc. 2nd Annual Workshop on Semiconductor Advances for Future Electronics - SAFE, 287-294, Mierlo - The Netherlands, 1999. *Poster presentation*.
- D. W. de Lima Monteiro, G. Vdovin and P. Sarro, “One- and two-dimensional CMOS position-sensitive detectors,” Topical meeting on Adaptive Optics, US Air Force, Albuquerque - USA, 1999. *Oral presentation*.
- D. W. de Lima Monteiro, G. Vdovin and P. Sarro, “Various layouts of analog CMOS optical position-sensitive detectors,” Proc. SPIE, vol.3794, 134-142, Denver - USA, 1999. *Oral presentation*.
- D. W. de Lima Monteiro, G. Vdovin and P. Sarro, “Integration of a Hartmann-Shack wavefront sensor,” Proc. of the Adaptive Optics Workshop for Industry and Medicine, 215-220, Durham - UK, 1999. *Oral presentation*.

About the author

Davies William de Lima Monteiro was born in Volta Redonda - Rio de Janeiro, Brazil, on January 24th, 1972. He received his secondary education at "Escola Técnica Federal do Espírito Santo" from 1986 to 1990. He then studied at "Universidade Federal do Espírito Santo", where he received his BSc and MSc degrees in Physics in 1994 and 1997, respectively. The Master's degree was obtained on the preparation and characterization of meta-stable Fe-Cu magnetic alloys. In 1998 he joined the Electronic Instrumentation Laboratory as a Ph.D. Student, where he had the opportunity to deal with a whole new spectrum of research activities. His research focused on the implementation of an adaptive optical element in the framework of a standard CMOS technology, namely an integrated wavefront sensor. The various preliminary considerations, the implementation and the performance of the sensor resulted in this thesis.

About the author